Oncology, Nuclear Medicine and Transplantology
(eISSN 3105-8760)

Review Article

Predicting Cancer Before It Begins: An AI-Enabled Framework for Precision Assessment of Individual Susceptibility to Environmental Carcinogens

Cecilia Oyinlade Ogundare[1], Abdul-Basid Yakubu[2], Mary Adebimpe Adewoye[3], Esther Uyoyooghene Olokede[4]

[1]Department of Chemical Sciences, College of Basic Science, Lagos State University of Science and Technology, Ikorodu, Nigeria
[2]School of Natural Sciences, University of Chester, England, United Kingdom
[3]Department of Applied Health Science, Indiana University Bloomington, IN, USA
[4]Department of Medical Laboratory Science, Faculty of Basic Medical Sciences, University of Benin,Edo State, Nigeria

Received: Sep 28, 2026
Accepted: Nov 07, 2026

Corresponding author's email:
olokedesther@gmail.com

Citation: Ogundare CO, Yakubu AB, Adewoye MA, Olokede EU. Predicting Cancer Before It Begins: An AI-Enabled Framework for Precision Assessment of Individual Susceptibility to Environmental Carcinogens. Oncology, Nuclear Medicine and Transplantology. 2026;2(4):onmt030. https://doi.org/10.63946/onmt/19943

Copyright: By the author(s).

License: A non-exclusive license by the publisher. Published by Australasia Publishing Group LLP (Astana, Kazakhstan) on behalf National Research Oncology Center.

Open Access: This article is an open access article distributed under the terms and conditions of the CC-BY Creative Commons Attribution license https://creativecommons.org/licenses/by/4.0



ABSTRACT

Environmental carcinogens remain significant contributors to cancer development; however, conventional risk assessment has largely focused on population-level exposure without adequately accounting for variation in individual susceptibility. Individuals exposed to comparable environmental hazards can nevertheless face different cancer risks, as susceptibility is influenced by genetic and epigenetic variation, metabolic capacity, immune function, age, lifestyle, comorbidities, and cumulative exposure histories. The growing availability of environmental, clinical, molecular, and longitudinal health data offers new opportunities to characterize these differences. Artificial intelligence (AI), including machine learning and multimodal methods, has been proposed as a means of modelling complex patterns across such heterogeneous data, although comparative evidence that it outperforms conventional statistical approaches remains limited. This narrative review synthesizes evidence on environmental carcinogenesis, host susceptibility, and AI-based risk modelling, and proposes a conceptual seven-stage framework that integrates exposure and contextual information, host susceptibility, biological response, and longitudinal data to estimate an individual's absolute risk of a specified cancer over a defined time horizon. The framework explicitly separates primary prevention, which acts on exposure before carcinogenic processes are established, from early detection, which acts on preclinical disease, and is illustrated with a worked example of occupational aromatic amine exposure, NAT2 acetylator status, and bladder cancer. We present no new model or performance data; instead, we summarize primary studies of AI in exposure modelling and cancer risk prediction and identify the cohorts and data resources that could support model development and validation. Key challenges include exposure misclassification, data quality, algorithmic bias, limited interpretability, privacy concerns, the distinction between prediction and causation, insufficient external validation, and implementation barriers. AI should therefore complement, not replace, established epidemiological and mechanistic approaches. With rigorous prospective validation, transparent reporting, and appropriate ethical and clinical safeguards, AI-enabled susceptibility assessment may support a transition toward more individualized, risk-informed environmental cancer prevention.

Keywords: Environmental Carcinogens; Cancer Susceptibility; Artificial Intelligence; Precision Prevention; Primary Prevention; Exposome; Risk Prediction; Environmental Health

Introduction

Cancer development arises from a complex interaction among inherited susceptibility, biological processes, and environmental exposures accumulated over the course of life. Environmental and lifestyle determinants, such as tobacco smoke, air pollution, occupational carcinogens, ultraviolet and ionizing radiation, and specific chemical agents, are well-established contributors to cancer risk [1,2]. A considerable proportion of cancer cases is attributable to modifiable risk factors and is therefore potentially preventable through the modification or avoidance of those factors, underscoring the importance of primary prevention within broader cancer control efforts [1,3]. Nevertheless, environmental carcinogenesis is not governed by exposure alone. Individuals encountering the same carcinogenic agent may develop different biological responses because susceptibility is shaped by genetic background, epigenetic regulation, metabolic capacity, immune function, age, lifestyle, comorbidities, and the timing, duration, and combination of exposures [4,5]. Consequently, the detection of an environmental hazard does not imply an identical level of cancer risk among all exposed individuals [6].

Conventional environmental risk assessment has largely relied on population-level estimates based on exposure distributions and epidemiological associations. These approaches remain fundamental to the identification of carcinogenic hazards and the development of regulatory and public health measures, but they offer less resolution when the aim is to determine risk at the individual level [7,8]. Human exposure is increasingly understood as a dynamic and multidimensional phenomenon, encompassing numerous environmental agents encountered across different environments and life stages. The exposome concept reflects this broader view by accounting for the cumulative environmental influences experienced throughout an individual's lifetime and their interactions with endogenous biological processes [9,10]. Capturing this complexity presents a considerable analytical challenge, as susceptibility is unlikely to be adequately characterized by any single exposure measure or biomarker [11].

Precision prevention provides a complementary strategy by focusing on variation in individual risk and the adaptation of preventive measures to those differences [12–14]. Such an approach requires the combined consideration of environmental, clinical, molecular, behavioral, and longitudinal information rather than treating these domains independently. AI has been proposed as one analytical means of achieving this integration. Validated AI models have predicted future breast, lung, and pancreatic cancer from mammograms, low-dose computed tomography, and electronic health record trajectories, in some cases with better discrimination than established clinical risk models [15–17]. However, most of these models use imaging or clinical data rather than environmental exposure information, many operate closer to early detection than to primary prevention, and important concerns remain regarding transportability, interpretability, bias, and translation into preventive settings [18,19]. Representative primary studies are summarized in Table 3.

Two distinctions are central to this review. The first is between primary prevention, which seeks to avoid or reduce carcinogenic exposure before carcinogenic processes are established, and early detection (secondary prevention), which seeks to identify preclinical or early-stage cancer when treatment is more effective [20,21]. The second is between prediction, which identifies who is at elevated risk, and causal inference, which establishes why that risk exists and whether changing an exposure would alter it [22]. Working definitions of these and related terms are provided in Table 2.

Against this background, this narrative review examines how AI might integrate multidimensional environmental and host-related data to characterize individual susceptibility to environmental carcinogens. It proposes a conceptual framework, not a validated model, that connects exposure and contextual information, host susceptibility, biological responses, and longitudinal data to estimate individual risk that could inform primary prevention and, where evidence supports it, risk-adapted early detection. The review also sets out how the framework differs from existing exposome, precision-prevention, and AI risk-prediction frameworks, and addresses the methodological, ethical, and translational challenges that must be overcome before such approaches can be incorporated into preventive practice.

Review Methodology

Search Strategy

A narrative literature review was undertaken to examine the relationship between environmental carcinogen exposure, individual susceptibility to cancer, AI, and precision cancer prevention. PubMed/MEDLINE, Scopus, Web of Science, Embase, IEEE Xplore, and Google Scholar were searched for literature published between 1 January 2020 and 31 August 2026 (the date of the last search). The core search string, adapted to the syntax of each database,

combined four concept blocks: (i) environmental or occupational exposure and the exposome; (ii) cancer, neoplasm, or carcinogenesis; (iii) individual or cancer susceptibility, gene–environment interaction, biomarkers, risk prediction, or precision prevention; and (iv) artificial intelligence, machine learning, deep learning, or multimodal methods. Supplementary searches paired the AI block with "cancer risk prediction", "external validation", "explainable", "dataset shift", "algorithmic bias", and "federated learning". For Google Scholar, results were screened by relevance until no further eligible records were identified. Reference lists of key studies and review articles were also screened.

Eligibility Criteria and Study Selection

Records were eligible if they (i) were published in English in peer-reviewed journals; (ii) addressed environmental or occupational carcinogen exposure, host susceptibility to cancer, exposome or biomarker research relevant to cancer, or the development, validation, interpretation, or governance of AI or machine-learning models for exposure assessment or cancer risk prediction; and (iii) were original studies, systematic or scoping reviews, meta-analyses, methodological guidance, or reports from recognized cancer or environmental health agencies. Records were excluded if they were conference abstracts without full text, opinion pieces without original data or methodological content, studies of AI applied only to cancer diagnosis, treatment selection, or post-diagnosis prognosis, or studies outside human health. Titles and abstracts were screened independently by two of the authors, and potentially relevant full texts were assessed by at least two authors; disagreements were resolved by discussion. Because the aim was conceptual synthesis rather than exhaustive retrieval, no formal risk-of-bias assessment or quantitative synthesis was performed, and record counts are not reported in a flow diagram.

Foundational sources published before 2020 were added purposively, outside the database search, where they provided the original description of an established concept, method, or guideline: the exposome concept [9], the key characteristics of carcinogens [23], principles of screening [21], phases of biomarker development [24], and landmark gene–environment and exposure-modelling studies [4,25,26]. Authoritative agency sources were likewise included, notably the IARC Monographs Preamble and the World Cancer Report [1,27].

Synthesis and Framework Development

The evidence was synthesized narratively according to biological relevance, predictive capability, methodological limitations, and translational potential, with attention to exposure assessment, population characteristics, data quality, model validation, interpretability, and the distinction between predictive association and causation. The framework was developed iteratively by mapping the stages of the exposure–disease continuum, from exposure to internal dose, host modification, biological response, and preclinical and clinical disease, onto the data domains and analytical requirements identified in the literature, and onto established guidance for prediction-model development and reporting [19,28]. The framework is therefore a conceptual proposal derived from this synthesis; it has not been empirically tested. The limitations of this review are described in Section 8.7.

From Environmental Exposure to Individual Cancer Susceptibility

Environmental carcinogenesis arises from exposures encountered across multiple settings and stages of life. Individuals may encounter carcinogenic agents through ambient and household air pollution, tobacco smoke, occupational environments, contaminated food and water, pesticides, industrial chemicals, and different forms of radiation. Importantly, these exposures rarely occur in isolation [29]. Exposure intensity, duration, route, timing, and co-exposure patterns may vary substantially between individuals and can change throughout the life course. Two features of carcinogenic exposure are particularly relevant to individual risk estimation. Dose–response refers to the relationship between the amount of exposure, often expressed as cumulative exposure, and the resulting cancer risk; latency refers to the interval, frequently several decades for solid tumors, between exposure and clinically detectable cancer [25,30]. Evidence from cancer research increasingly emphasizes the importance of considering multiple environmental and occupational exposures and their interactions rather than evaluating hazards independently [31].

The Exposome Perspective

The exposome provides a framework for capturing this complexity by encompassing the full range of environmental exposures encountered throughout the life course and the biological effects associated with them [9,32]. This perspective extends beyond the evaluation of individual carcinogens to include chemical and physical agents as well as lifestyle, occupational, and contextual exposures that may interact over time. The dynamic and heterogeneous nature of environmental exposure makes comprehensive characterization technically demanding. However, advances in high-throughput analytical techniques and molecular profiling have

created new opportunities to identify exposure-related biological signatures and examine associations between complex exposure patterns and cancer risk [10,33].

A recent scoping review indicated that exposome research in cancer remains relatively limited, with many studies continuing to rely on predefined exposure categories rather than comprehensive profiling of environmental chemicals [34]. Several existing resources could nevertheless support the development and validation of the models proposed here. Large prospective cohorts with linked exposure, genetic, and outcome data include UK Biobank, which combines questionnaire, residential air-pollution, genotype, and proteomic data with cancer registry linkage [35]; the European Prospective Investigation into Cancer and Nutrition (EPIC), with stored biospecimens and dietary and lifestyle data [36]; the Agricultural Health Study, with detailed pesticide-exposure histories [37]; and the All of Us Research Program, with electronic health record, genomic, and wearable data from a diverse population [38]. Exposome consortia such as HELIX provide harmonized exposure and molecular data that, although not primarily designed for cancer outcomes, offer methods and data infrastructure that could be extended [39]. Exposome-wide analyses in UK Biobank have shown that measured exposures explain a substantial proportion of variation in mortality and in the incidence of several age-related diseases, independently of polygenic risk [40].

Biological Determinants of Individual Susceptibility

Environmental exposures do not elicit uniform biological responses across individuals. Genetic variation can alter toxicokinetics, including the absorption, distribution, metabolism, and excretion of carcinogenic agents, as well as toxicodynamic responses. These variations can affect carcinogen bioactivation, detoxification, and the repair of carcinogen-induced damage, while epigenetic changes may mediate interactions between environmental factors and cellular regulatory processes [41]. The key characteristics of carcinogens, a framework now used in IARC evaluations, include genotoxicity, impaired DNA repair, epigenetic alterations, oxidative stress, chronic inflammation, and immunosuppression among the properties commonly exhibited by established human carcinogens [23,27].

These mechanisms function within interconnected biological networks rather than in isolation. Genetic and epigenetic differences may alter metabolic responses to environmental exposures, while changes in metabolism can affect the production or persistence of reactive intermediates. In turn, DNA damage and compromised repair mechanisms may contribute to genomic instability, while persistent inflammatory and immune alterations can modify the cellular environment in which carcinogenesis occurs [42].

The Conceptual Gap

The complexity of these interactions reveals an important limitation of conventional approaches to environmental cancer risk assessment. Although population-level epidemiology remains essential for identifying carcinogenic hazards and estimating risk, translating population-based associations into individualized susceptibility requires the concurrent assessment of exposure history, biological characteristics, and responses [43]. The central challenge is therefore not merely to establish whether an environmental factor is carcinogenic, but to determine which individuals may be more susceptible, under what exposure conditions, and through which biological mechanisms. Current exposome research offers a promising basis for addressing this challenge while also demonstrating the difficulty of integrating such multidimensional evidence [44].

This gap provides a rationale for exploring an AI-enabled approach. Flexible machine-learning methods may be useful for detecting nonlinear relationships and interactions across environmental and biological variables, although in clinical prediction they have not consistently outperformed well-specified regression models [45]. Through the integration of heterogeneous exposure, host, molecular, and longitudinal datasets, AI may facilitate a shift from assessing isolated risk factors toward a more comprehensive characterization of individual environmental cancer risk [46]. Figure 1 illustrates the pathway from external exposure to internal dose, host modification, biological response, and cancer, and indicates where primary prevention and early detection respectively act.

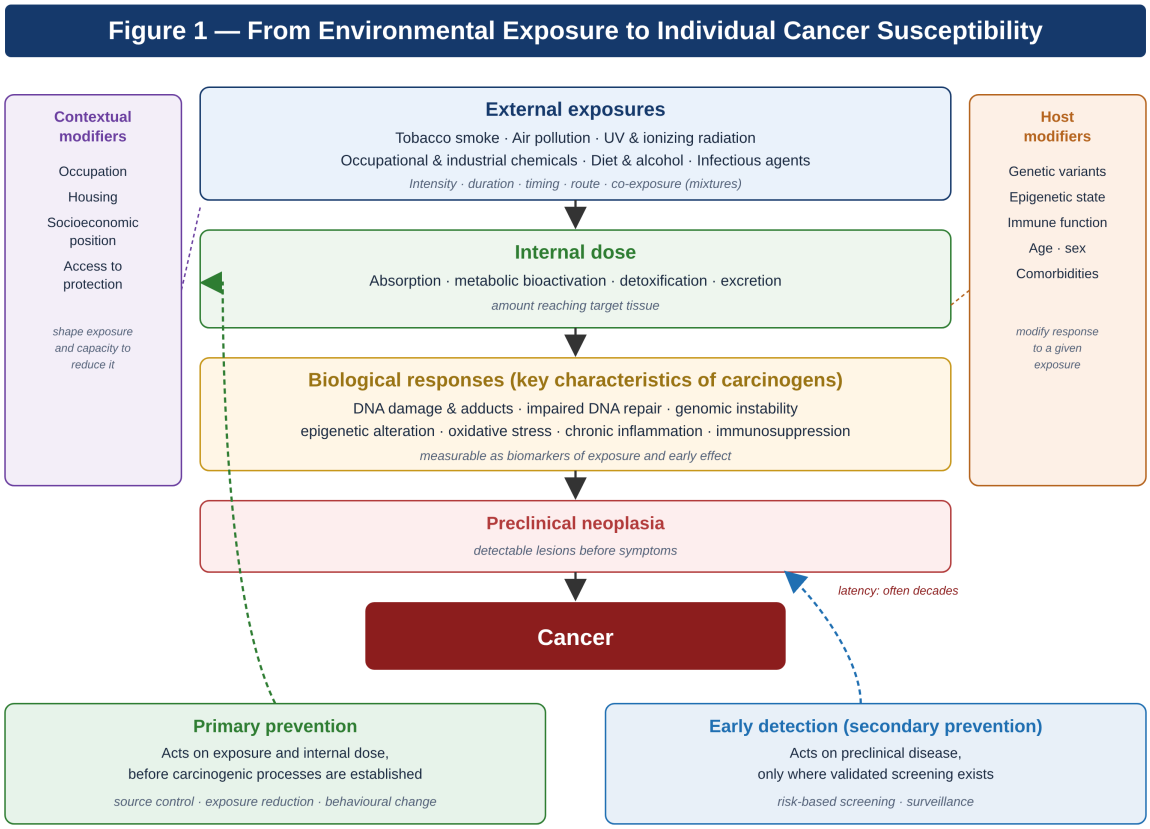


Figure 1. Conceptual pathway from environmental exposure to individual cancer susceptibility. Original figure created by the authors.

External exposures, modified by contextual factors such as occupation, housing, and socioeconomic conditions, give rise to an internal dose that depends on absorption, metabolic bioactivation, and detoxification. Host factors, including genetic variants, epigenetic state, immune function, age, sex, and comorbidities, modify this process. Biological responses corresponding to the key characteristics of carcinogens may progress, after a latency period that is often several decades long, to preclinical and then clinically manifest cancer. Primary prevention acts on exposure and internal dose; early detection acts on preclinical disease.

AI-Enabled Integration of Multidimensional Risk Data

Assessing individual susceptibility to environmental carcinogens requires the integration of information that varies considerably in scale, structure, and biological relevance. Environmental datasets may describe exposure patterns across different locations and time periods, whereas genomic, molecular, clinical, and behavioral data reflect characteristics of the individual [43]. Conventional analytical methods can evaluate these domains and pre-specified interactions between them, but may be less practical when the number of candidate interactions is large. AI, including machine learning, deep learning, and multimodal methods, has been proposed as a means of bringing these sources together. The aim is not merely to increase predictive complexity, but to develop a more complete representation of how environmental exposures interact with the biological and contextual factors that shape individual susceptibility [5,40].

Environmental Exposure Data

Accurate characterization of environmental exposure represents the foundation of an individualized susceptibility model. Traditional assessment methods frequently depend on measurements collected at specific locations or on self-reported exposure histories, which may not adequately reflect the intensity, duration, timing, or variability of exposure experienced by an individual [47]. Machine-learning exposure models are among the most mature AI applications in environmental health. An ensemble of neural network, random forest, and gradient-boosting models estimated daily fine particulate matter (PM2.5) concentrations across the contiguous United States at 1 km resolution, with a cross-validated R^2 of 0.86 for daily estimates [26], and a random forest model built from more than 50,000 groundwater measurements was used to map the global probability of arsenic exceeding 10 μg/L, estimating that 94–220

million people are potentially exposed [48]. Combined with residential histories, occupational records, remote-sensing observations, and personal or wearable sensors, such models may allow exposure burden to be estimated with greater spatial and temporal resolution [49].

This capability is especially important when multiple environmental agents are encountered concurrently. Instead of assigning individuals to a single exposure category, computational models can incorporate mixtures of environmental agents and identify patterns associated with particular environments or periods of the life course. The reliability of these estimates, however, remains closely tied to the quality and representativeness of the underlying exposure data. Area-level exposure models also estimate ambient concentrations rather than personal dose, and AI cannot overcome systematic gaps, measurement errors, or inaccuracies in the information from which predictions are generated [50].

Host Susceptibility Data

Exposure history alone is insufficient to capture the extent to which an individual may be susceptible to environmental carcinogens. Host characteristics can influence the absorption, metabolism, detoxification, repair, and elimination of carcinogenic agents and, consequently, modify their biological effects. Potentially relevant variables include genetic variants, epigenetic profiles, metabolic characteristics, immune status, age, sex, family history, and comorbidities [4,51].

Machine-learning methods may assist in identifying interactions among these variables that are difficult to specify in advance within conventional statistical models. A particular genetic or metabolic characteristic, for instance, may have limited significance when considered independently but may contribute more meaningfully to susceptibility when evaluated in the context of a specific exposure profile [5]. One potential advantage of AI therefore lies in modelling relationships among susceptibility factors rather than treating each variable as an isolated predictor. Such associations nonetheless require biological interpretation and independent validation before they can be considered clinically meaningful [28].

Molecular and Biological Response Data

Another important component of the framework is the characterization of biological responses associated with environmental exposure. Carcinogenic agents can induce molecular changes involving DNA damage, oxidative stress, inflammation, epigenetic regulation, cellular metabolism, and immune function. These effects can be examined using circulating biomarkers and high-dimensional molecular techniques, including transcriptomics, proteomics, metabolomics, and epigenomic profiling [10,33].

Combining molecular measurements with exposure histories may help distinguish the presence of an environmental hazard from the biological response it elicits. This distinction is important because detection of a contaminant does not necessarily imply that all exposed individuals have undergone the same degree of biological perturbation [41]. It is also useful to separate biomarkers of exposure and early biological effect, such as carcinogen–protein or DNA adducts, which can inform primary prevention, from markers of preclinical neoplasia, which belong to early detection. AI may facilitate the identification of combinations of molecular features associated with particular exposure patterns or susceptibility characteristics. Before such biomarkers are incorporated into predictive models, however, their reproducibility, specificity, temporal stability, and clinical relevance must be established through staged validation [24,52].

Lifestyle and Contextual Data

Individual susceptibility is also shaped by behavioral and social circumstances. Factors such as smoking, alcohol consumption, diet, physical activity, occupation, socioeconomic conditions, and relevant medication use may influence both the extent of environmental exposure and the resulting biological response. Incorporating these variables may reduce the likelihood that an AI model attributes an observed association entirely to an environmental factor when part of that relationship is related to correlated behavioral or social determinants [53].

Contextual information can also strengthen the practical interpretation of estimated risk. For example, a person residing in an area with persistent air pollution may have limited capacity to reduce exposure without changes at the environmental or occupational level. Individualized risk assessment should therefore avoid assigning responsibility for environmental risk solely to the person being assessed. Predictions generated by AI need to be interpreted within the broader social, occupational, and environmental circumstances in which exposure occurs [54]. In the proposed framework, lifestyle and contextual data are characterized in the first stage, together with environmental exposure, and enter the risk model as inputs and potential effect modifiers (Section 5.1).

AI Approaches for Multidimensional Risk Integration

Different AI methods may contribute complementary capabilities. Machine-learning algorithms can model nonlinear relationships among environmental, host, molecular, and clinical variables

and may support risk stratification when appropriately trained and validated. In lung cancer, for example, an ensemble machine-learning model using only age, smoking duration, and pack-years had higher sensitivity than existing risk models at equivalent specificity in external validation [55], and a model using routine clinical and laboratory data identified non-small cell lung cancer 9–12 months before diagnosis more accurately than a modified PLCOm2012 model [56]. Deep learning may be useful for high-dimensional or unstructured data such as imaging and longitudinal records [15,16]. Greater model complexity, however, does not necessarily translate into greater clinical utility: a systematic review of clinical prediction studies found no average discrimination benefit of machine learning over logistic regression in comparisons at low risk of bias [45]. Any proposed AI model should therefore be benchmarked against a well-specified conventional model.

Explainable AI is especially relevant to susceptibility assessment because a clinically useful model should provide more than a simple classification of high or low risk. Ideally, it should help identify the exposure characteristics, biological features, or contextual factors contributing to the estimated risk while also conveying the uncertainty associated with that estimate. Model-explanation techniques may therefore improve transparency, facilitate scientific interpretation, and help clinicians determine whether model outputs are biologically and clinically plausible [57].

Multimodal AI is a further option because environmental cancer susceptibility is inherently multidimensional. Rather than evaluating exposure, genomic, molecular, and clinical information separately, multimodal models can in principle learn relationships across these domains. Their use, however, introduces further challenges, including data harmonization, missing information, differences in measurement scales, model transportability, and the risk of algorithmic bias [43]. Table 3 summarizes representative primary studies. Among the studies identified, none integrated individual-level exposure histories, host susceptibility markers, and biological response data in a single validated cancer risk model; the validated cancer models relied mainly on imaging or clinical records, and the exposure models produced area-level estimates. This gap motivates the framework proposed in Section 5.

From Data Integration to an Individualized Susceptibility Profile

The integration of these data streams provides the basis for characterizing individual susceptibility to environmental carcinogens. Environmental data can describe where, when, and to what extent exposure may have occurred, while host characteristics provide information about factors that may modify an individual's response. Molecular measurements can provide additional evidence of biological responses associated with exposure, and lifestyle, clinical, and contextual information can help account for factors that influence both exposure and susceptibility [5].

Within the proposed approach, AI functions as an integrative analytical layer that examines relationships across these domains rather than evaluating each source independently. The resulting analysis combines exposure characteristics with biological and contextual information to generate an individualized estimate of risk. Importantly, this estimate should distinguish established evidence, such as agents classified as carcinogenic to humans by IARC, from associations identified only by the model, and should communicate the uncertainty associated with the estimated risk [19,27]. Rather than representing susceptibility as a function of a single exposure or isolated biological characteristic, the proposed approach considers it as an outcome of interacting environmental, biological, behavioral, and contextual factors. The data domains and their contributions are summarized in Table 1.

Table 1. Data domains for AI-enabled assessment of individual susceptibility to environmental carcinogens.

Data domain	Examples	What it contributes
Environmental exposure	Air pollution, chemicals, radiation, occupation, geography	Exposure burden
Life-course / contextual	Residential and occupational history, duration, socioeconomic position	Cumulative and lagged exposure
Genetic	Variants in carcinogen-metabolism and DNA-repair genes, polygenic scores	Host predisposition
Epigenetic	DNA methylation, chromatin changes	Exposure-responsive regulation

Data domain	Examples	What it contributes
Molecular / biological response	Transcriptomics, proteomics, metabolomics, adducts	Biological response to exposure
Clinical	Age, sex, comorbidities, family history	Clinical context
Lifestyle	Smoking, diet, alcohol, physical activity	Modifying factors and confounders
Longitudinal	Repeated exposure, biomarker and clinical measures	Change over time; dynamic reassessment

Model outputs, namely the absolute risk estimate with its uncertainty interval and the factors contributing to it, are not a data domain; they are produced in Stages 4 and 5 of the framework (Section 5) and can be fed back as prior information at reassessment (Stage 7).

Table 2. Working definitions of key terms used in this review.

Term	Working definition as used in this review
Hazard	An agent or exposure with the intrinsic capacity to cause cancer (e.g., an IARC Group 1 carcinogen).
Exposure	Contact of an individual with a hazard, described by intensity, duration, timing, route, and frequency.
Internal dose	The amount of a hazard reaching target tissue after absorption, metabolism, and detoxification.
Risk	The probability that an individual develops a specified cancer over a defined time horizon (absolute risk), or relative to a reference (relative risk).
Susceptibility	The extent to which host factors modify the risk associated with a given exposure (effect modification), rather than an exposure-independent trait.
Latency	The interval between exposure and clinically detectable cancer, often several decades for solid tumors.
Dose–response	The relationship between the amount of (often cumulative, lagged) exposure and the resulting risk.
Prediction	Identifying who is at elevated risk, without necessarily explaining why.
Causation	Establishing that changing an exposure would change risk; required for effective prevention.
Primary prevention	Action on exposure or internal dose before carcinogenic processes are established.
Early detection	Identification of preclinical or early-stage cancer (secondary prevention), where validated tests exist.

Table 3. Representative primary studies of AI in environmental exposure modelling and cancer risk prediction.

Study	Domain / data	Task and main result	Relevant stage
Di et al. 2019	Environmental exposure; satellite, land-use, meteorological data	Ensemble ML estimated daily PM2.5 across the USA at 1 km; cross-validated R^2 0.86 (daily), 0.89 (annual)	Stage 1
Podgorski & Berg 2020	Environmental exposure; geospatial data, >50,000 measurements	Random forest mapped global groundwater arsenic >10 μg/L; ~94–220 million people potentially exposed	Stage 1
Argentieri et al. 2025	Exposome; UK Biobank exposures, proteomics	Exposome explained more variation in mortality and lung/heart/liver disease than polygenic risk	Stages 1–2

Study	Domain / data	Task and main result	Relevant stage
García-Closas et al. 2005	Gene–environment; NAT2/GSTM1 genotype, occupational data	NAT2 slow acetylation increased bladder cancer risk and modified the smoking/occupational effect	Stage 2
Callender et al. 2023	Cancer risk; UK Biobank, NLST, PLCO	Parsimonious ensemble ML selected lung-screening eligibility; higher sensitivity than existing models at equal specificity	Stages 4, 6
Gould et al. 2021	Cancer risk; routine clinical and laboratory data	ML identified NSCLC 9–12 months pre-diagnosis; AUC 0.86, exceeding modified PLCOm2012	Stages 4, 6
Mikhael et al. 2023	Cancer risk; single low-dose chest CT	Sybil predicted 6-year lung cancer risk; 1-year AUC 0.86–0.94 across external sets	Stage 4
Yala et al. 2021	Cancer risk; screening mammograms	Mirai predicted 5-year breast cancer risk; C-index 0.76–0.81, exceeding Tyrer-Cuzick	Stage 4
Placido et al. 2023	Cancer risk; EHR disease-code trajectories	Deep learning predicted pancreatic cancer within 36 months; AUROC up to 0.88	Stage 4
Christodoulou et al. 2019	Methods; systematic review of clinical prediction models	No average discrimination benefit of ML over logistic regression at low risk of bias	Stages 4–5

AUC, area under the receiver operating characteristic curve; C-index, concordance index; LDCT, low-dose computed tomography; NSCLC, non-small cell lung cancer; PM2.5, fine particulate matter; R^2, coefficient of determination. Studies were selected to illustrate the types of evidence available at each framework stage and do not constitute a systematic sample.

Building an Individual Susceptibility Framework

The framework comprises seven stages (Figure 2). Its target output, or estimand, is the absolute risk that a person develops a specified cancer, for example urothelial bladder cancer, over a defined time horizon, typically 5 or 10 years and accounting for the competing risk of death, given that person's exposure history, host characteristics, and biological response profile [58]. Absolute risk is preferred because preventive and screening decisions depend on it; relative risk, compared with a similarly exposed reference profile, can be reported alongside it to show how much the individual's host factors modify the exposure effect. Exposure enters the model as time-varying cumulative exposure, lagged to reflect latency, so that dose–response relationships can be estimated rather than assumed.

Exposure and Contextual Characterization

The first stage establishes a profile of the individual's environmental exposure burden and the context in which it occurs. This requires characterization of exposure intensity, duration, timing, route, and frequency, together with estimation of cumulative exposure across occupational, residential, dietary, and other relevant microenvironments [59]. Because the most recent years of exposure are unlikely to contribute to cancers diagnosed within the prediction horizon, cumulative exposure should be lagged, commonly by 5–10 years or more, to reflect latency. Co-exposure patterns and chemical mixtures should also be considered, as human exposure typically involves multiple agents acting concurrently [60]. In the same stage, lifestyle factors such as smoking, alcohol, diet, and physical activity, and contextual factors such as occupation, socioeconomic position, and housing, are recorded. These factors are needed both to adjust for confounding and to identify which exposures are modifiable at the individual level and which require source-level intervention. AI-based exposure models may contribute high-resolution estimates from monitoring systems, occupational records, and geospatial datasets [26,48].

Biological Susceptibility Profiling

The second stage focuses on host determinants that may modify the biological consequences of environmental exposure. Genetic variation can influence carcinogen bioactivation, metabolic clearance, DNA damage recognition and repair, and downstream cellular responses, whereas epigenetic alterations may modify transcriptional regulation without altering the underlying DNA sequence [44]. Metabolic phenotypes, immune status, demographic characteristics, and relevant physiological states may further contribute to interindividual heterogeneity in response [61]. In this framework, susceptibility is defined operationally as the extent to which host factors modify the risk associated with a given exposure, that is, as effect modification on the absolute or relative risk scale,

rather than as an intrinsic attribute independent of exposure.

Biological Response Assessment

The third stage evaluates evidence of exposure-related biological perturbation. Relevant indicators may include biomarkers of internal dose and early biological effect, such as carcinogen–DNA or protein adducts, together with markers of oxidative stress, inflammation, epigenetic change, and metabolic dysregulation [62]. These measurements occupy an intermediate position between environmental exposure and disease. For primary prevention, their role is to confirm that an exposure has reached biologically relevant targets and to monitor whether exposure reduction lowers them. Markers that indicate preclinical neoplasia, such as tumor-derived circulating DNA, serve early detection instead and are treated as a separate output (Section 5.6). Their incorporation into susceptibility models should be guided by established analytical validity, biological plausibility, reproducibility, and evidence of relevance to the exposure or disease process [24,52].

AI-Based Risk Integration

The fourth stage constitutes the principal analytical component of the framework. AI models could integrate exposure and contextual information (Stage 1), host susceptibility determinants (Stage 2), and biological response data (Stage 3) to estimate individual absolute risk. Depending on data dimensionality and availability, machine-learning methods such as gradient-boosted survival models or neural networks may be used to identify nonlinear dose–response relationships and higher-order interactions [40,55]. Robust development should include internal validation, assessment of calibration as well as discrimination, comparison against a conventional regression model with pre-specified interactions, and independent external testing to assess transportability across populations and settings [19,28,63].

Explainable Risk Profiling

A clinically and environmentally meaningful system should not reduce susceptibility to a single numerical output. The fifth stage therefore emphasizes model interpretability and explainability. The resulting model should ideally identify the exposure characteristics, host factors, molecular features, or biomarkers contributing most substantially to the estimated risk while also communicating the uncertainty associated with the prediction [64]. Such information can help distinguish potentially modifiable determinants from relatively fixed susceptibility characteristics and enhance the transparency of algorithmic outputs for researchers, clinicians, and other decision makers. Explainability is therefore not merely a technical consideration but an important requirement for biological interpretation and responsible translation [65].

Preventive Action

The sixth stage translates the risk profile into action, distinguishing two pathways. The primary-prevention pathway, which is the principal aim of the framework, applies when modifiable exposures contribute substantially to estimated risk. Actions include reduction or elimination of environmental and occupational exposures at source, workplace controls, smoking cessation and other behavioral support, and preventive counseling [66,67]. The early-detection pathway applies only where an established screening or surveillance test exists for the cancer concerned and there is evidence that earlier detection improves outcomes in people at the estimated level of risk; lung cancer screening with low-dose computed tomography, for which risk-based eligibility has been evaluated, is one example [55,68]. The purpose is not to classify individuals as predetermined to develop cancer, but to identify modifiable pathways through which prevention may be implemented earlier and in proportion to assessed risk.

Dynamic Reassessment

Susceptibility should ultimately be regarded as dynamic rather than static. Changes in environmental exposure, biomarkers, lifestyle, occupational circumstances, clinical status, or other relevant determinants may alter an individual's risk profile over time. Periodic reassessment can therefore allow the framework to update risk estimates as new exposure, biological, and longitudinal information becomes available [69]. This longitudinal structure differentiates the proposed framework from conventional models that typically provide a one-time estimate, and positions AI as an adaptive analytical tool for precision environmental health [70]. Figure 2 illustrates the complete framework.

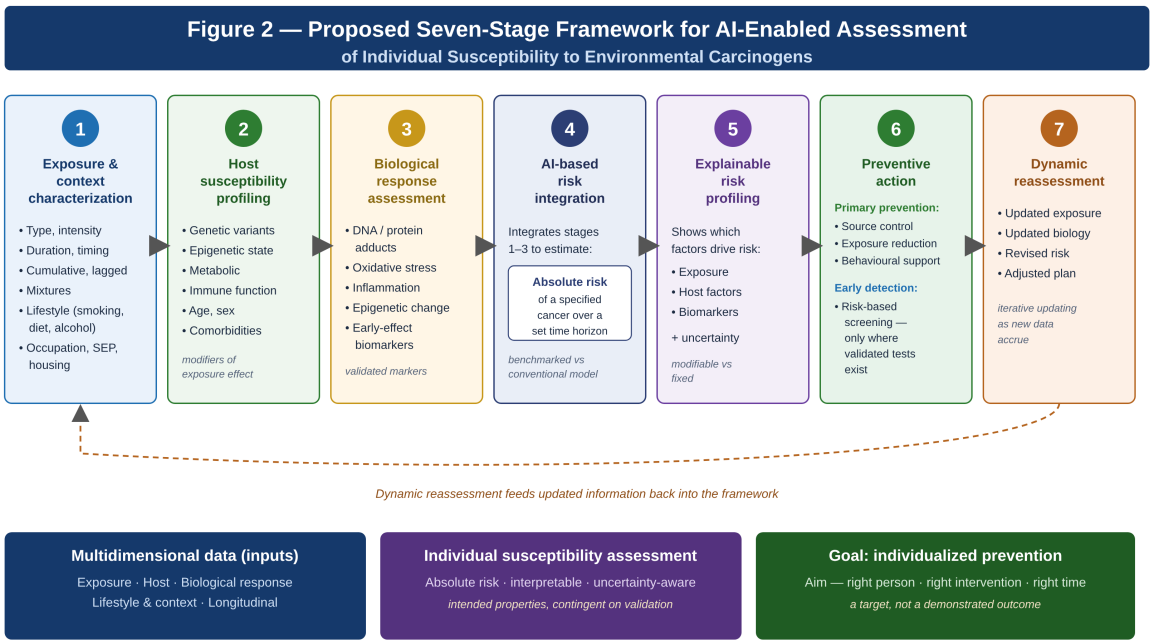


Figure 2. Proposed seven-stage framework for AI-enabled assessment of individual susceptibility to environmental carcinogens. Original figure created by the authors.

The framework progresses from exposure and contextual characterization, host susceptibility profiling, and biological response assessment to AI-based risk integration, which produces an absolute risk estimate for a specified cancer and time horizon, and explainable risk profiling. These outputs inform preventive action, which separates a primary-prevention pathway from an early-detection pathway used only where validated tests exist. Dynamic reassessment updates the risk profile as new information becomes available. The properties listed in the lower panel are design goals that depend on validation; they are not demonstrated properties of an existing system.

Worked Example: Aromatic Amines, NAT2 Acetylator Status, and Bladder Cancer

A single carcinogen–cancer pair illustrates how the stages fit together. Occupational exposure to aromatic amines, encountered historically in dye, rubber, and chemical manufacturing, is an established cause of urothelial bladder cancer and shows the long latency typical of solid tumors [30]. Susceptibility is modified by the N-acetyltransferase 2 (NAT2) gene: individuals with the slow-acetylator phenotype detoxify these compounds less efficiently and have a higher bladder cancer risk for a given exposure, an interaction supported by pooled epidemiological analyses [4]. In Stage 1, cumulative occupational exposure is reconstructed from job history and a job-exposure matrix and lagged to reflect latency, alongside the major competing cause, tobacco smoking. In Stage 2, NAT2 acetylator status is determined by genotyping. In Stage 3, biomarkers of internal dose or early effect may indicate whether exposure has reached the urothelium. In Stage 4, a model estimates the person's absolute 10-year bladder cancer risk, allowing the exposure effect to differ by acetylator status rather than assuming a single population effect. In Stage 5, the explanation identifies cumulative amine exposure and slow-acetylator status as the principal contributors. In Stage 6, because the dominant driver is a modifiable occupational exposure, action centres on primary prevention through workplace exposure control, with early detection considered only if a validated surveillance test is available for high-risk workers. In Stage 7, the estimate is updated as exposure changes or new biomarker data accrue. This example also marks the boundaries of present evidence: although the exposure–gene interaction is established, a fully integrated, prospectively validated AI model for this pair does not yet exist, and building one would require the cohorts and validation steps described above.

Novelty and Relationship to Existing Frameworks

The components used here, the exposome, gene–environment interaction, precision prevention, and AI-based risk prediction, are individually well established, so it is important to state what the

proposed framework adds. Exposome frameworks comprehensively characterize environmental exposure and its biological embedding but are generally oriented toward exposure measurement and etiological discovery rather than individual-level, decision-oriented risk estimation [9,10]. Precision-prevention frameworks adapt preventive action to individual risk but have focused mainly on genetic and clinical risk factors, with environmental exposure handled coarsely [12,13]. Validated AI risk models predict specific cancers with high discrimination but rely largely on imaging or clinical records and seldom incorporate individual environmental exposure histories or their interaction with host susceptibility [15–17].

The proposed framework differs in four respects. First, it places quantitative, lagged environmental exposure, rather than genetic or clinical risk alone, at the centre of an individual risk estimate, and it defines that estimate as an absolute risk for a specified cancer and time horizon. Second, it treats host factors explicitly as modifiers of the exposure effect, operationalizing susceptibility as effect modification rather than as a standalone score. Third, it builds the distinction between primary prevention and early detection into the output of the model, routing modifiable exposures to source-level action and reserving screening for cancers with validated tests, which most existing AI risk models do not do. Fourth, it is explicitly longitudinal, updating estimates as exposure and biology change. The framework is thus best understood as an integration of these existing strands around a decision-oriented, exposure-centred estimand, rather than as a new algorithm; it remains a conceptual proposal requiring empirical evaluation.

From Risk Prediction to Precision Cancer Prevention

The value of an individual susceptibility framework depends on its capacity to translate risk information into meaningful and evidence-based cancer prevention. Identifying an individual as potentially susceptible should therefore not constitute the endpoint of the analytical process [71]. Instead, susceptibility estimates should provide a basis for proportionate preventive action while minimizing unnecessary medicalization and interventions unsupported by sufficient evidence. The transition from risk prediction to prevention can operate across individual, clinical, and population-health domains [72].

Primary Prevention

The most direct application of susceptibility information is the prevention or reduction of harmful exposures before carcinogenic processes are established. Where a specific environmental or occupational exposure contributes substantially to an individual's estimated risk, reducing or eliminating that exposure represents the central opportunity for primary prevention [2]. Depending on the exposure context, relevant measures may include workplace exposure controls, environmental remediation, modification of residential conditions, dietary change, or other evidence-based strategies. Consistent with the hierarchy of controls, preventive efforts should prioritize control or elimination of the environmental source where feasible, rather than placing disproportionate responsibility on individuals to modify behaviors in response to hazards they may have limited capacity to avoid [73].

Risk-Adapted Early Detection

For individuals with evidence of elevated susceptibility, risk-adapted early detection may provide an additional pathway, but only where clinically validated monitoring or screening strategies exist and where earlier detection has been shown to improve outcomes. This may include established screening programs or appropriately defined follow-up for populations with demonstrably increased risk [74]. An AI-derived susceptibility estimate should not by itself determine the intensity or frequency of clinical surveillance; recommendations for screening or biomarker monitoring should rest on evidence that earlier detection or intervention produces meaningful benefit at the relevant level of risk [54,68]. Risk-based selection for lung cancer screening, where machine-learning models have been used to determine eligibility, illustrates both the promise and the evidentiary requirements of this pathway [55].

Risk-Adapted Interventions

Integrating environmental exposure profiles with biological determinants of susceptibility may allow preventive strategies to be tailored more closely to individual circumstances. Rather than applying uniform interventions to all individuals exposed to a particular carcinogen, risk-adapted prevention could account for differences in exposure magnitude, exposure history, biological response, and modifiable determinants of susceptibility [75]. For instance, where an individual's profile is driven by a modifiable occupational exposure, preventive measures may appropriately emphasize engineering controls, personal protective measures, or other workplace interventions [76]. A different profile may indicate the relevance of alternative strategies. Such adaptation should remain evidence based, transparent, and proportionate to the strength and certainty of the underlying evidence.

Population Health Integration

Precision prevention should not be construed as a substitute for population-level environmental health interventions. Environmental carcinogens frequently affect large and geographically dispersed populations, and source-level exposure reduction can prevent disease irrespective of whether individual susceptibility has been characterized [77]. Population-based prevention therefore seeks to reduce environmental hazards across communities, whereas precision prevention aims to identify individuals who may exhibit greater biological vulnerability within those populations [44].

These approaches are consequently complementary rather than mutually exclusive. Population-level regulation, environmental remediation, occupational protection, and source-directed exposure control remain foundational components of environmental cancer prevention [78]. Individualized susceptibility assessment can add resolution by identifying differences in exposure burden, host biology, and biological response that may support more tailored strategies. Within this context, the proposed framework positions AI as a potential analytical bridge between population environmental health and individualized prevention [79].

Challenges, Limitations and Ethical Considerations

Despite the potential of AI to integrate environmental, biological, and clinical information into individualized susceptibility profiles, substantial methodological, ethical, and translational barriers must be addressed before these approaches can reliably inform preventive decision-making [80]. The principal challenges include deficiencies in data quality and completeness, limitations in model development and validation, uncertainty surrounding causal interpretation, algorithmic bias and inequitable performance, privacy and data-governance concerns, and the potential clinical consequences of inaccurate predictions [81].

Data Limitations

The reliability of an AI-derived susceptibility estimate is fundamentally contingent on the quality, completeness, and representativeness of the data from which it is generated. Environmental exposure histories are frequently incomplete, particularly when exposure occurs across multiple geographic locations, occupational settings, or developmental stages over the life course [82]. Exposure estimates may also depend on indirect measures or surrogate indicators derived from geographic proxies, occupational classifications, environmental monitoring systems, or self-reported histories, thereby introducing uncertainty into exposure characterization and cumulative dose estimation. Longitudinal biological and clinical datasets may similarly be affected by missing observations, measurement inconsistencies, temporal variability, and differences in sampling or analytical procedures across data sources [83]. Heterogeneity in population characteristics and data-generating processes can further complicate multimodal integration and constrain model generalizability when training cohorts are narrowly defined or poorly representative of the populations in which the model is ultimately deployed. Although AI can detect complex patterns within high-dimensional datasets, it cannot rectify systematic measurement error, exposure misclassification, or deficiencies inherent in the underlying data [84].

AI Model Limitations

The application of AI introduces additional methodological challenges related to model complexity, robustness, and generalizability. Highly flexible models may overfit training data and exhibit substantial reductions in performance when evaluated in independent populations or under different environmental and clinical conditions [19]. Differences between the distributions of development data and those encountered during real-world deployment, commonly described as dataset shift, can further compromise predictive performance and model calibration [18]. Prediction errors may also have asymmetric consequences: false-positive classifications could lead to unnecessary investigations, interventions, or psychological distress, whereas false-negative predictions may create unwarranted reassurance and delay appropriate preventive or diagnostic evaluation. Model interpretability is therefore particularly important when algorithmic outputs have potential implications for health-related decisions. Although explainable AI techniques can identify influential variables and provide insights into model behavior, interpretability alone does not establish clinical validity, biological plausibility, or causal relevance [85].

Causality Versus Prediction

A fundamental conceptual limitation lies in the distinction between predictive association and causal inference. An AI model may identify statistical relationships linking an environmental exposure, host characteristic, biomarker, and cancer outcome without demonstrating that the exposure is causally responsible for the observed disease risk [22]. Predictive performance may instead derive from correlated variables, proxy measures, residual confounding, selection effects, or complex statistical dependencies that do not correspond to underlying carcinogenic

mechanisms. This distinction has direct practical consequences, because only causal determinants will reduce risk when modified: a model may stratify risk accurately using a non-causal marker, yet acting on that marker would not prevent cancer. AI-derived susceptibility estimates should therefore be interpreted as complementary to, rather than substitutes for, epidemiological, mechanistic, and experimental evidence, and preventive action should preferentially target exposures with established causal support, such as agents evaluated as carcinogenic by IARC [27,86].

Bias and Health Inequity

Algorithmic bias represents a further concern because AI systems can reproduce, amplify, or obscure biases embedded within their training data and data-generating processes. Environmental exposures are not randomly distributed across populations and may differ according to geographic location, occupation, socioeconomic conditions, housing quality, environmental regulation, and access to protective resources [87]. When such variation is inadequately represented in development datasets, model performance may differ across demographic, geographic, socioeconomic, or exposure-defined groups, potentially reinforcing existing health inequities. Accordingly, fairness evaluation should extend beyond aggregate performance metrics to examine discrimination, calibration, sensitivity, specificity, and predictive value across relevant population subgroups. Particular attention should also be given to whether apparently accurate models rely on variables that function as proxies for structural disadvantage or unequal exposure conditions [88].

Privacy and Ethical Considerations

Individual susceptibility assessment may require the integration of highly sensitive information, including genomic profiles, residential and geospatial histories, occupational exposure records, electronic health records, and longitudinal molecular or biomarker measurements. The convergence of these data streams increases the potential for re-identification, unauthorized access, secondary use, and unintended disclosure, thereby heightening the importance of robust data governance and privacy protections [89]. Ethical implementation should incorporate secure data infrastructures, appropriate access controls, transparent policies governing data collection and use, and meaningful informed consent. Governance frameworks must also establish how susceptibility information is accessed, communicated, retained, and potentially shared with third parties, particularly when the clinical significance of an AI-derived prediction remains uncertain. Safeguards are additionally required to minimize stigmatization, discrimination, or inappropriate attribution of responsibility to individuals for environmental risks that may arise from structural, occupational, or community-level conditions [90].

Clinical Validation and Implementation

Translation of AI-enabled susceptibility assessment into preventive practice requires evaluation that extends well beyond technical performance within the original development cohort. A credible translational pathway should progress from model derivation and internal validation to rigorous external validation in independent populations, followed by prospective evaluation under real-world conditions, consistent with established reporting and appraisal standards for prediction models [19,91]. Assessment should encompass discrimination, calibration, subgroup performance, interpretability, clinical utility, transportability, and the consequences of both false-positive and false-negative classifications, rather than relying exclusively on conventional performance metrics. Where the intended application involves preventive intervention, studies should also establish whether incorporation of the model into clinical or public health workflows leads to improved decision-making and meaningful health outcomes. Ultimately, the relevance of an AI system depends not only on its capacity to predict risk accurately, but also on whether its deployment supports appropriate, equitable, interpretable, and evidence-based preventive decisions [92].

Limitations of This Review

This review has limitations. As a narrative rather than systematic review, it did not use a registered protocol, duplicate screening of all records, formal risk-of-bias assessment, or quantitative synthesis, so selection of the cited evidence may be subject to reviewer judgment and potential bias. The proposed framework is conceptual and has not been implemented or empirically validated; its feasibility, calibration, and clinical utility are therefore unknown. The worked example and the cohorts and primary studies cited are illustrative rather than comprehensive, and the evidence summarized in Table 3 is drawn predominantly from high-income settings, which may limit generalizability. Finally, the field is evolving rapidly, and findings published after the last search date are not reflected.

Future Directions: Toward Dynamic and Preventive Oncology

The future of environmental cancer prevention is likely to move beyond static assessments of exposure and disease risk toward continuously updated models of individual susceptibility. Advances in multimodal AI may facilitate the integration of environmental measurements, clinical characteristics, multi-omics profiles, imaging data, and longitudinal biomarkers within unified analytical frameworks. Concurrently, wearable technologies and distributed environmental sensing systems may generate increasingly granular information on real-time exposure, allowing susceptibility estimates to incorporate changing environmental conditions and exposure dynamics rather than relying predominantly on retrospective exposure histories [93].

Longitudinal biomarker profiling represents another important avenue. Repeated measurement of molecular and physiological responses may enable the detection of emerging biological perturbations and their incorporation into dynamically updated susceptibility models [94]. More sophisticated computational approaches, including individualized mechanistic simulations and digital twins, could eventually permit scenario-based modeling of how alterations in exposure patterns, host characteristics, or biological states might influence future cancer risk. However, translating these approaches into preventive oncology will require rigorous validation of their predictive performance, biological validity, clinical utility, and generalizability [95].

As these systems increasingly depend on the integration of sensitive, high-dimensional personal data, privacy-preserving computational strategies will become critical [96]. Federated learning may enable model training across multiple institutions without requiring direct centralization of individual-level datasets, while complementary privacy-preserving methods could support secure exchange and integration of distributed environmental, clinical, and molecular information. Such approaches may expand opportunities for collaborative model development while reducing the risks associated with centralized data repositories [97].

Ultimately, prospective prevention studies will be essential for establishing whether AI-derived susceptibility estimates can improve clinically and epidemiologically meaningful outcomes rather than merely enhance predictive performance [43]. Future research should therefore progress beyond static risk stratification toward dynamic susceptibility monitoring, in which environmental exposure, biological response, and individualized risk estimates are iteratively updated as new information becomes available [96]. Such a paradigm could provide the basis for a more adaptive model of precision cancer prevention, in which modifiable susceptibility is identified sufficiently early to support proportionate, evidence-based intervention before carcinogenic processes advance to clinically manifest disease.

Conclusion

Environmental carcinogenesis cannot be adequately characterized by determining simply whether an individual has encountered a carcinogenic agent. Cancer susceptibility emerges from the dynamic interplay among exposure characteristics, host biological determinants, molecular responses, temporal factors, and the environmental and social contexts in which exposure occurs. Although conventional population-level risk assessment remains essential for identifying carcinogenic hazards and guiding regulatory and public health action, it provides limited resolution for explaining why individuals exposed to similar environmental conditions may experience different biological responses and disease risks. Addressing this gap requires integration of environmental exposure histories with genomic, epigenetic, metabolic, immunological, clinical, behavioral, and longitudinal information. AI offers a potential analytical framework for integrating these heterogeneous, high-dimensional datasets and developing a more comprehensive representation of individual susceptibility. The framework proposed here conceptualizes susceptibility as a dynamic process, estimates an individual's absolute risk of a specified cancer over a defined time horizon, and separates primary prevention from early detection so that modifiable exposures are addressed at source. Its intended value lies not merely in improving risk prediction, but in clarifying how environmental exposures interact with individual biology over time and in identifying modifiable determinants of susceptibility. The framework is nonetheless a conceptual proposal: it has not been implemented or validated, and its translation into practice remains constrained by incomplete exposure histories, heterogeneous data sources, measurement uncertainty, dataset shift, algorithmic bias, limited interpretability, privacy concerns, and the gap between prediction and causation. Rigorous model development, external validation, prospective evaluation, transparent reporting, and appropriate ethical and clinical safeguards will therefore be essential. AI should

function as a complementary analytical tool rather than a replacement for established epidemiological, mechanistic, experimental, and public health approaches. With these safeguards, AI-enabled susceptibility assessment could support a more dynamic and individualized model of risk-informed prevention while complementing population-level strategies aimed at reducing carcinogenic exposures and addressing the environmental and structural conditions that affect entire populations.

Acknowledgements

Author Contributions: Conceptualization: C.O.O.; Methodology: C.O.O.; Investigation: C.O.O., A.-B.Y., M.A.A., and E.U.O.; Writing – Original Draft Preparation: M.A.A. and E.U.O.; Writing – Review and Editing: C.O.O., A.-B.Y., M.A.A., and E.U.O.; Visualization: E.U.O.; Project Administration: E.U.O.
All authors contributed to the manuscript in accordance with their respective roles, critically reviewed the manuscript, and read and approved the final version for publication.
Acknowledgments: Not applicable.
Funding: Not applicable.
Conference presentation: Not applicable.
Disclosure of AI Use: ChatGPT (GPT), developed by OpenAI, was used during the preparation of this manuscript to assist with language refinement, editing, and improving the clarity and readability of the text. The authors critically reviewed and revised all AI-assisted outputs and take full responsibility for the accuracy, originality, integrity, and final content of the manuscript. ChatGPT was not used as an author, and all scientific interpretations, analyses, and conclusions remain the responsibility of the authors.

References

1. Wild CP, Weiderpass E, Stewart BW, editors. World Cancer Report: Cancer Research for Cancer Prevention. Lyon: International Agency for Research on Cancer; 2020.
2. Rondelli R, Leardini D, Bossù G, Masetti R, Mastronuzzi A. Environmental risk factors for cancer in childhood, adolescence, and young adults: A narrative review. Glob Pediatr. 2026 Sep 1;17:100351. doi:10.1016/j.gpeds.2026.100351
3. Fink H, Langselius O, Vignat J, Rumgay H, Rehm J, Martinez RX, et al. Global and regional cancer burden attributable to modifiable risk factors to inform prevention. Nat Med. 2026 Apr;32(4):1306–15. doi:10.1038/s41591-026-04219-7. PubMed PMID: 41634393.
4. García-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, Hein DW, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. Lancet. 2005;366(9486):649–59. doi:10.1016/S0140-6736(05)67137-1
5. Motsinger-Reif AA, Reif DM, Akhtari FS, House JS, Campbell CR, Messier KP, et al. Gene-environment interactions within a precision environmental health framework. Cell Genomics. 2024 Jul 10;4(7):100591. doi:10.1016/j.xgen.2024.100591. PubMed PMID: 38925123; PubMed Central PMCID: PMC11293590.
6. Kasperczyk E, Tarhonska K, Jablonska E. Epigenetic Changes Induced by Carcinogenic Agents or Occupational Exposure with Sufficient Evidence for Bladder Cancer. Cancer Prev Res (Phila). 2025 Aug 1;18(8):439–51. doi:10.1158/1940-6207.CAPR-24-0450. PubMed PMID: 40304368; PubMed Central PMCID: PMC12314524.
7. National Research Council. Science and Decisions: Advancing Risk Assessment. Washington, DC: The National Academies Press; 2009. doi:10.17226/12209
8. Shetty BR, Pai BJ, Salmataj SA, Naik N. Assessment of Carcinogenic and non-carcinogenic risk indices of heavy metal exposure in different age groups using Monte Carlo Simulation Approach. Sci Rep. 2024 Dec 5;14:30319. doi:10.1038/s41598-024-81109-3. PubMed PMID: 39638837; PubMed Central PMCID: PMC11621557.
9. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005;14(8):1847–50. doi:10.1158/1055-9965.EPI-05-0456

10. Vermeulen R, Schymanski EL, Barabási AL, Miller GW. The exposome and health: Where chemistry meets biology. Science. 2020;367(6476):392–6. doi:10.1126/science.aay3164
11. Gudi-Mindermann H, White M, Roczen J, Riedel N, Dreger S, Bolte G. Integrating the social environment with an equity perspective into the exposome paradigm: A new conceptual framework of the Social Exposome. Environ Res. 2023 Sep 15;233:116485. doi:10.1016/j.envres.2023.116485
12. Rebbeck TR, Burns-White K, Chan AT, Emmons K, Freedman M, Hunter DJ, et al. Precision prevention and early detection of cancer: fundamental principles. Cancer Discov. 2018;8(7):803–11. doi:10.1158/2159-8290.CD-17-1415
13. Pearson TA, Vitalis D, Pratt C, Campo R, Armoundas AA, Au D, et al. The Science of Precision Prevention. JACC Adv. 2023 Dec 13;3(1):100759. doi:10.1016/j.jacadv.2023.100759. PubMed PMID: 38375059; PubMed Central PMCID: PMC10876066.
14. Vineis P. Prevention, precision prevention and precision medicine. Exposome. 2025 Oct 25;5(1):osaf011. doi:10.1093/exposome/osaf011. PubMed PMID: 41209440; PubMed Central PMCID: PMC7618338.
15. Yala A, Mikhael PG, Strand F, Lin G, Smith K, Wan YL, et al. Toward robust mammography-based models for breast cancer risk. Sci Transl Med. 2021;13(578):eaba4373. doi:10.1126/scitranslmed.aba4373
16. Mikhael PG, Wohlwend J, Yala A, Karstens L, Xiang J, Takigami AK, et al. Sybil: a validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. J Clin Oncol. 2023;41(12):2191–200. doi:10.1200/JCO.22.01345
17. Placido D, Yuan B, Hjaltelin JX, Zheng C, Haue AD, Chmura PJ, et al. A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. Nat Med. 2023;29(5):1113–22. doi:10.1038/s41591-023-02332-5
18. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med. 2021;385(3):283–6. doi:10.1056/NEJMc2104626
19. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378
20. Crosby D, Bhatia S, Brindle KM, Coussens LM, Dive C, Emberton M, et al. Early detection of cancer. Science. 2022;375(6586):eaay9040. doi:10.1126/science.aay9040
21. Wilson JMG, Jungner G. Principles and Practice of Screening for Disease. Public Health Papers No. 34. Geneva: World Health Organization; 1968.
22. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. Chance. 2019;32(1):42–9. doi:10.1080/09332480.2019.1579578
23. Smith MT, Guyton KZ, Gibbons CF, Fritz JM, Portier CJ, Rusyn I, et al. Key Characteristics of Carcinogens as a Basis for Organizing Data on Mechanisms of Carcinogenesis. Environ Health Perspect. 2016;124(6):713–21. doi:10.1289/ehp.1509912
24. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst. 2001;93(14):1054–61. doi:10.1093/jnci/93.14.1054
25. Darby S, Hill D, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ. 2005;330(7485):223. doi:10.1136/bmj.38308.477650.63
26. Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, et al. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ Int. 2019;130:104909. doi:10.1016/j.envint.2019.104909

27. International Agency for Research on Cancer. Preamble to the IARC Monographs on the Identification of Carcinogenic Hazards to Humans (amended January 2019). Lyon: IARC; 2019.
28. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925–31. doi:10.1093/eurheartj/ehu207
29. Silva DB da, Pianovski MAD, Carvalho Filho NP de. Environmental pollution and cancer. J Pediatr (Rio J). 2025 Mar 1;Climate Change and Childhood Health101:S18–26. doi:10.1016/j.jped.2024.09.004
30. International Agency for Research on Cancer. Chemical Agents and Related Occupations. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 100F. Lyon: IARC; 2012.
31. Ekpanyaskul C. Update on Occupational Cancer for Better Cancer Prevention and Control. Asian Pac J Cancer Prev. 2024;25(12):4465–76. doi:10.31557/APJCP.2024.25.12.4465. PubMed PMID: 39733440; PubMed Central PMCID: PMC12008330.
32. Vineis P, Barouki R. The exposome as the science of social-to-biological transitions. Environ Int. 2022 Jul 1;165:107312. doi:10.1016/j.envint.2022.107312
33. Vogli M, Jeong A, Yu Z, Vonk JM, Ibi D, Kronberg J, et al. The impact of environmental exposures on DNA methylation in the EXPANSE project. eBioMedicine. 2026 Jan 1;123:106084. doi:10.1016/j.ebiom.2025.106084
34. Young AS, Mullins CE, Sehgal N, Vermeulen RCH, Kolijn PM, Vlaanderen J, et al. The need for a cancer exposome atlas: a scoping review. JNCI Cancer Spectr. 2025 Jan 3;9(1):pkae122. doi:10.1093/jncics/pkae122. PubMed PMID: 39700422; PubMed Central PMCID: PMC11729703.
35. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
36. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002;5(6B):1113–24. doi:10.1079/PHN2002394
37. Alavanja MCR, Sandler DP, McMaster SB, Zahm SH, McDonnell CJ, Lynch CF, et al. The Agricultural Health Study. Environ Health Perspect. 1996;104(4):362–9. doi:10.1289/ehp.96104362
38. All of Us Research Program Investigators. The "All of Us" Research Program. N Engl J Med. 2019;381(7):668–76. doi:10.1056/NEJMsr1809937
39. Vrijheid M, Slama R, Robinson O, Chatzi L, Coen M, van den Hazel P, et al. The human early-life exposome (HELIX): project rationale and design. Environ Health Perspect. 2014;122(6):535–44. doi:10.1289/ehp.1307204
40. Argentieri MA, Amin N, Nevado-Holgado AJ, Sproviero W, Collister JA, Keestra SM, et al. Integrating the environmental and genetic architectures of aging and mortality. Nat Med. 2025;31(3):1016–25. doi:10.1038/s41591-024-03483-9
41. Reddy SR, Bangeppagari M, Lee SJ. Immune–Epigenetic Effects of Environmental Pollutants: Mechanisms, Biomarkers, and Transgenerational Impact. Curr Issues Mol Biol. 2025 Sep;47(9):703. doi:10.3390/cimb47090703
42. Chen C, Wang Z, Qin Y. Connections between metabolism and epigenetics: mechanisms and novel anti-cancer strategy. Front Pharmacol. 2022 Jul 22;13:935536. doi:10.3389/fphar.2022.935536. PubMed PMID: 35935878; PubMed Central PMCID: PMC9354823.
43. Zhao S, Yu P, Xu R, Li S. The role of artificial intelligence in cancer epidemiology: Challenges and opportunities. Intell Oncol. 2026 Jul 1;2(3):100070. doi:10.1016/j.intonc.2026.100070

44. Mafe AN, Büsselberg D. Cancer and Environmental Xenobiotics: Mechanisms, Controversies, and Innovations. J Xenobiot. 2026 Feb;16(1):2. doi:10.3390/jox16010002
45. 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
46. Cheng CH, Shi SS. Artificial intelligence in cancer: applications, challenges, and future perspectives. Mol Cancer. 2025 Oct 30;24:274. doi:10.1186/s12943-025-02450-3. PubMed PMID: 41168799; PubMed Central PMCID: PMC12574039.
47. Goddard FGB, Ban R, Barr DB, Brown J, Cannon J, Colford JM Jr, et al. Measuring Environmental Exposure to Enteric Pathogens in Low-Income Settings: Review and Recommendations of an Interdisciplinary Working Group. Environ Sci Technol. 2020 Aug 19;54(19):11673. doi:10.1021/acs.est.0c02421. PubMed PMID: 32813503.
48. Podgorski J, Berg M. Global threat of arsenic in groundwater. Science. 2020;368(6493):845–50. doi:10.1126/science.aba1510
49. Olawade DB, Wada OZ, Ige AO, Egbewole BI, Olojo A, Oladapo BI. Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions. Hyg Environ Health Adv. 2024 Dec 1;12:100114. doi:10.1016/j.heha.2024.100114
50. Falakdin P, Sabzevari S, Bos P, van Engelen J, Kennedy M, Peters S, et al. Modeling external exposure to pesticides in human populations: Developing an exposure scenario generator. Ecotoxicol Environ Saf. 2026 Jun 15;318:120201. doi:10.1016/j.ecoenv.2026.120201
51. Kasperczyk E, Tarhonska K, Jablonska E. Genotoxicity Induced by Carcinogenic Agents or Occupational Exposure with Sufficient Evidence for Bladder Cancer. J Clin Med. 2025 Jan;14(13):4492. doi:10.3390/jcm14134492
52. Ahmad A, Imran M, Ahsan H. Biomarkers as Biomedical Bioindicators: Approaches and Techniques for the Detection, Analysis, and Validation of Novel Biomarkers of Diseases. Pharmaceutics. 2023 May 31;15(6):1630. doi:10.3390/pharmaceutics15061630. PubMed PMID: 37376078; PubMed Central PMCID: PMC10303887.
53. Tavares MS, Rodrigues CT, de Menezes SLS, Moreno AM. Health at Risk: Air Pollution and Urban Vulnerability—Perspectives in Light of the 2030 Agenda. Green Health. 2025 Dec;1(3):21. doi:10.3390/greenhealth1030021
54. Benoussaïd T, Coll I, Charreire H, Makni I, Costes M, Elessa Etuman A. Reassessing air pollution exposure: How daily mobility and activities shape individual risk in greater Paris. Comput Environ Urban Syst. 2025 Dec 1;122:102340. doi:10.1016/j.compenvurbsys.2025.102340
55. Callender T, Imrie F, Cebere B, Pashayan N, Navani N, van der Schaar M, et al. Assessing eligibility for lung cancer screening using parsimonious ensemble machine learning models: a development and validation study. PLoS Med. 2023;20(10):e1004287. doi:10.1371/journal.pmed.1004287
56. Gould MK, Huang BZ, Tammemagi MC, Kinar Y, Shiff R. Machine learning for early lung cancer identification using routine clinical and laboratory data. Am J Respir Crit Care Med. 2021;204(4):445–53. doi:10.1164/rccm.202007-2791OC
57. Carriero A, de Hond A, Cappers B, Paulovich F, Abeln S, Moons KG, et al. Explainable AI in healthcare: to explain, to predict, or to describe? Diagn Progn Res. 2025 Dec 5;9:29. doi:10.1186/s41512-025-00213-8. PubMed PMID: 41345739; PubMed Central PMCID: PMC12679721.
58. Gail MH, Pfeiffer RM. On criteria for evaluating models of absolute risk. Biostatistics. 2005;6(2):227–39. doi:10.1093/biostatistics/kxi005
59. Hosseini B, Hall AL, Zendehdel K, Kromhout H, Onyije FM, Moradzadeh R, et al. Occupational Exposure to Carcinogens and Occupational Epidemiological Cancer Studies

in Iran: A Review. Cancers. 2021 Jan;13(14):3581. doi:10.3390/cancers13143581

60. Simonnet-Laprade C, Bayen S, Le Bizec B, Dervilly G. Data analysis strategies for the characterization of chemical contaminant mixtures. Fish as a case study. Environ Int. 2021 Oct 1;155:106610. doi:10.1016/j.envint.2021.106610
61. Yang Q, Cai Y, Guan Y, Wang Z, Guo S, Qiu S, et al. Metabolic phenotypes: Molecular bridges between health homeostasis and disease imbalance. Comput Struct Biotechnol J. 2025 Oct 27;27:4710–9. doi:10.1016/j.csbj.2025.10.057. PubMed PMID: 41245891; PubMed Central PMCID: PMC12615335.
62. Bustaffa E, Aprea C, Barbieri A, Benassi A, Borghini A, Caini S, et al. From Environmental Evidence to Biomarker Selection: A Structured Decision-Support Process for Human Biomonitoring Studies in Contaminated Sites. Toxics. 2026 Jul 15;14(7):616. doi:10.3390/toxics14070616. PubMed PMID: 42515181; PubMed Central PMCID: PMC13416600.
63. Strandberg R, Jepsen P, Hagström H. Developing and validating clinical prediction models in hepatology – An overview for clinicians. J Hepatol. 2024 Jul 1;81(1):149–62. doi:10.1016/j.jhep.2024.03.030
64. Borra S, Dey N, Fong S, Sherratt RS, Shi F. Explainability and Trust in Deep Learning for Cancer Imaging: Systematic Barriers, Clinical Misalignment, and a Translational Roadmap. Cancers. 2026 Apr 24;18(9):1361. doi:10.3390/cancers18091361. PubMed PMID: 42122156; PubMed Central PMCID: PMC13163020.
65. Fayaz M, Hagsong K, Danish S, Dang LM, Sadeghi-Niaraki A, Moon H. Demystifying Artificial Intelligence: A Systematic Review of Explainable Artificial Intelligence in Medical Imaging. Sensors. 2026 Jan;26(7):2131. doi:10.3390/s26072131
66. AbdulRaheem Y. Unveiling the Significance and Challenges of Integrating Prevention Levels in Healthcare Practice. J Prim Care Community Health. 2023 Jul 14;14:21501319231186500. doi:10.1177/21501319231186500. PubMed PMID: 37449436; PubMed Central PMCID: PMC10350749.
67. Espina C, Ritchie D, Feliu A, Canelo-Aybar C, D'Souza E, Mitrou PN, et al. Developing evidence-based cancer prevention recommendations: Methodology of the World Code Against Cancer Framework to create region-specific codes. Int J Cancer. 2026 Jan 1;158(1):9–18. doi:10.1002/ijc.70068. PubMed PMID: 40741905; PubMed Central PMCID: PMC12588551.
68. US Preventive Services Task Force. Screening for lung cancer: US Preventive Services Task Force recommendation statement. JAMA. 2021;325(10):962–70. doi:10.1001/jama.2021.1117
69. Sulsky SI, Greene T, Gentry PR. A framework for integrating evidence to assess hazards and risk. Crit Rev Toxicol. 2024 May 27;54(5):315–29. doi:10.1080/10408444.2024.2342447. PubMed PMID: 38808643.
70. Fernandes Prabhu D, Gurupur V, Stone A, Trader E. Integrating Artificial Intelligence, Electronic Health Records, and Wearables for Predictive, Patient-Centered Decision Support in Healthcare. Healthcare. 2025 Oct 30;13(21):2753. doi:10.3390/healthcare13212753. PubMed PMID: 41228120; PubMed Central PMCID: PMC12607345.
71. Krewski D, Saunders-Hastings P, Baan RA, Barton-Maclaren TS, Browne P, Chiu WA, et al. Development of an Evidence-Based Risk Assessment Framework. ALTEX. 2022;39(4):667–93. doi:10.14573/altex.2004041. PubMed PMID: 36098377; PubMed Central PMCID: PMC10080579.
72. Palmieri L, Donfrancesco C, Elonheimo H, Marcozzi B, Cardinale A, Damiano C, et al. From prediction to prevention: strengths and limitations of risk calculators. Eur J Public Health. 2025 Oct 27;35(Suppl 4):ckaf161.122. doi:10.1093/eurpub/ckaf161.122. PubMed Central PMCID: PMC12556819.
73. Cannarella R, Howard S, Calogero AE, Heindel JJ. Reducing exposure to non-persistent

endocrine-disrupting chemicals and their impact on human health: Evidence from dietary, behavioral, and residential interventions. Ecotoxicol Environ Saf. 2026 Jun 15;318:120272. doi:10.1016/j.ecoenv.2026.120272

74. Plans-Beriso E, Babb-de-Villiers C, Petrova D, Barahona-López C, Diez-Echave P, Hernández OR, et al. Biomarkers for personalised prevention of chronic diseases: a common protocol for three rapid scoping reviews. Syst Rev. 2024 Jun 1;13:147. doi:10.1186/s13643-024-02554-9. PubMed PMID: 38824585; PubMed Central PMCID: PMC11143646.
75. Abravan A, Fornacon-Wood I, Kingston R, Topping D, Price G. Environmental determinants of cancer outcomes: a scoping review. Radiother Oncol. 2026 Apr 1;217:111408. doi:10.1016/j.radonc.2026.111408
76. Teym A, Moges M, Adane B, Bewket Y, Zeleke TK, Mengistu AK, et al. Utilization of Personal Protective Equipment and Its Determinants Among Health Care Professionals in Public Hospitals in Resource-Limited Settings. SAGE Open Nurs. 2026 May 15;12:23779608261452766. doi:10.1177/23779608261452766. PubMed PMID: 42152947; PubMed Central PMCID: PMC13180112.
77. Ren J, Yan G, Yang L, Kong L, Guan Y, Sun H, et al. Cancer chemoprevention: signaling pathways and strategic approaches. Signal Transduct Target Ther. 2025 Apr 18;10(1):113. doi:10.1038/s41392-025-02167-1
78. Aranda-Merino N, Gómez de las Heras C, Ramos-Payán M, Fernández-Torres R, Bello-López MA, Gómez-Ariza JL. The Environmental Contribution to Health and Disease: Analytical Tools for Monitoring and Diagnosis. Environments. 2026 Aug;13(8):416. doi:10.3390/environments13080416
79. Sorrentino E, Mazzeo A, Mascolo C, Chiara MV, Rosati S, Maiuro L. From Reactive to Predictive One Health: AI-Enabled Frameworks for Integrated Zoonotic Surveillance and Governance. Int J Environ Res Public Health. 2026 Jul;23(7):850. doi:10.3390/ijerph23070850
80. Khude H, Shende P. AI-driven clinical decision support systems: Revolutionizing medication selection and personalized drug therapy. Adv Integr Med. 2025 Dec 1;12(4):100529. doi:10.1016/j.aimed.2025.100529
81. Ajisafe OM, Ogebule LA, Egbon E, Olawuyi OF, Olasilola OR, Olawade DB. Mitigating algorithmic bias in AI-powered toxicology: Frameworks for explainable and equitable predictions in human health and environmental safety. Toxicol Lett. 2026 Jun 1;421:111924. doi:10.1016/j.toxlet.2026.111924
82. Abdelsalam MK, Nwokolo SC, Iwuji PC, Meyer EL, Ahia CC, Egdair IMM. Hazardous waste risk detection powered by artificial intelligence and machine learning for resilient monitoring and mitigation of potential environmental threats. Results Eng. 2026 Mar 1;29:109655. doi:10.1016/j.rineng.2026.109655
83. Hoyo C, Martin CL, Allen T, Skaar D, Murphy SK. Environmental exposures and epigenetic alterations in common chronic diseases: insights and challenges. Environ Epigenetics. 2026 Jan 1;12(1):dvag014. doi:10.1093/eep/dvag014
84. Fahim YA, Hasani IW, Kabba S, Ragab WM. Artificial intelligence in healthcare and medicine: clinical applications, therapeutic advances, and future perspectives. Eur J Med Res. 2025 Sep 23;30:848. doi:10.1186/s40001-025-03196-w. PubMed PMID: 40988064; PubMed Central PMCID: PMC12455834.
85. Ennab M, Mcheick H. Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions. Front Robot AI. 2024 Nov 28;11:1444763. doi:10.3389/frobt.2024.1444763. PubMed PMID: 39677978; PubMed Central PMCID: PMC11638409.
86. Ye Y, Pandey A, Bawden C, Sumsuzzman DM, Rajput R, Shoukat A, et al. Integrating artificial intelligence with mechanistic epidemiological modeling: a scoping review of opportunities and challenges. Nat Commun. 2025 Jan

10;16:581. doi:10.1038/s41467-024-55461-x. PubMed PMID: 39794317; PubMed Central PMCID: PMC11724045.

87. Hasanzadeh F, Josephson CB, Waters G, Adedinsewo D, Azizi Z, White JA. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. NPJ Digit Med. 2025 Mar 11;8:154. doi:10.1038/s41746-025-01503-7. PubMed PMID: 40069303; PubMed Central PMCID: PMC11897215.
88. Osório de Barros G, Condeixa J. Reducing Regional Inequalities in Healthcare Systems Through Data, Interoperability and Artificial Intelligence: An Equity-Oriented Systemic Framework. Systems. 2026 Sep;14(9):1162. doi:10.3390/systems14091162
89. Kim Y, Oh S, Kim G. Convergence of Integrated Sensing and Communication (ISAC) and Digital-Twin Technologies in Healthcare Systems: A Comprehensive Review. Signals. 2025 Dec;6(4):51. doi:10.3390/signals6040051
90. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. Heliyon. 2024 Feb 29;10(4):e26297. doi:10.1016/j.heliyon.2024.e26297
91. Wu Q, Qi Z, Li Q, Hao C, Tang S. Systematic evaluation of machine learning models for clinical risk prediction on real-world hospital datasets. iScience. 2026 Apr 8;29(5):115654. doi:10.1016/j.isci.2026.115654. PubMed PMID: 42063551; PubMed Central PMCID: PMC13127387.
92. Nastoska A, Jancheska B, Rizinski M, Trajanov D. Evaluating Trustworthiness in AI: Risks, Metrics, and Applications Across Industries. Electronics. 2025 Jan;14(13):2717. doi:10.3390/electronics14132717
93. Fuse Y, Murphy SN, Ikari H, Takahashi A, Fuse K, Kawakami E. Artificial intelligence in clinical data analysis: A review of large language models, foundation models, digital twins, and allergy applications. Allergol Int. 2025 Oct 1;74(4):499–513. doi:10.1016/j.alit.2025.06.005
94. Lv J, Yu F, Wang H, Zhang Y, Wang Y, Yan L, et al. Longitudinal metabolomic profiling unveils dynamic biomarkers for predicting immune recovery in HIV-1 infection. Clin Transl Med. 2025 Oct 15;15(10):e70502. doi:10.1002/ctm2.70502. PubMed PMID: 41088882; PubMed Central PMCID: PMC12521791.
95. Shen S, Qi W, Liu X, Zeng J, Li S, Zhu X, et al. From virtual to reality: innovative practices of digital twins in tumor therapy. J Transl Med. 2025 Mar 19;23:348. doi:10.1186/s12967-025-06371-z. PubMed PMID: 40108714; PubMed Central PMCID: PMC11921680.
96. Amin Z, Anjum A, Khan A, Ahmad A, Jeon G. Preserving Privacy of High-Dimensional Data by l-Diverse Constrained Slicing. Electronics. 2022 Jan;11(8):1257. doi:10.3390/electronics11081257
97. Saidi R, Moulahi T, Zidi S, Mahfoudhi S. A privacy-aware federated learning framework for collaborative healthcare IoT security. Mach Learn Appl. 2026 Sep 1;25:100944. doi:10.1016/j.mlwa.2026.100944