

Review

Artificial Intelligence Approaches for Prediction and Detection of Immune-Related Adverse Events with Immune Checkpoint Inhibitor Cancer Therapy: A Narrative Review

Eman Nayaz Ahmed ^{1,*}, Mohamed S. Ahmed ²  and Ali H. Mushtaq ³ ¹ College of Medicine, Alfaisal University, Riyadh 11533, Saudi Arabia² Larner College of Medicine, University of Vermont, Burlington, VT 05405, USA³ Department of Internal Medicine, Cleveland Clinic, Cleveland, OH 44195, USA

* Correspondence: nayazah.eman@gmail.com

Abstract

Immune checkpoint inhibitors (ICIs) have translated the scope of cancer therapy, but their immune-restorative mechanism can lead to immune-related adverse events (irAEs), which can affect several organs with varying severity. Early identification of patients at risk of irAEs is prudent to inform clinical decisions in precision oncology and remains a challenge as current clinical and biomarker methods still lack predictive accuracy. Artificial Intelligence (AI) presents a promising strategy for improving early detection, risk stratification as well as monitoring of irAEs. This narrative review summarizes the current AI modalities for detecting and predicting irAEs risk occurring with ICI therapy, including clinical Machine Learning models, radiomics-based approaches, natural language processing (NLP) systems and the integration of modalities with multimodal AI frameworks. Clinical Machine Learning models demonstrate moderate predictive performance whereas radiomics-derived modeling appears promising for pneumonitis. NLP and language models have achieved higher accuracy for retrospective irAEs detection. Multimodal AI applications offer theoretical potential through diverse data integration that captures the complex biology of irAEs; however, current evidence is limited. All these modalities face limitations of inadequate sample sizes, retrospective design, heterogeneous outcome definitions, class imbalance, and insufficient external validation. AI-based models have significant potential for personalized immunotherapy monitoring but require prospective multicenter validation, standardized datasets and clinically interpretable frameworks prior to implementation. Future advances in multimodal modeling can also enable precise prediction and early detection of irAEs, ultimately improving the safety and effectiveness of cancer immunotherapy.

Keywords: immune checkpoint inhibitors; immune-related adverse events; artificial intelligence; machine learning; radiomics; natural language processing; multimodal AI; cancer immunotherapy; precision oncology; predictive modeling

Academic Editors: Paola D. Vermeer
and Raul Urrutia

Received: 9 June 2026

Revised: 16 July 2026

Accepted: 22 July 2026

Published: 27 July 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Immune checkpoint inhibitors (ICIs) are a breakthrough in the field of oncology for improving the survival of cancer patients. However, these immunomodulators have subsequently given rise to immune-related adverse events (irAEs), occurring in around 40% of treated patients. irAEs toxicity can vary from pneumonitis to dermatologic disorders with severe grade (≥ 3) in 19.7% cases [1]. These toxicities interrupt therapy cycles, affect clinical outcomes, and worsen the quality of life [2]. Consequently, improving the prediction and

early identification of irAEs has become the major goal of precision oncology, motivating the development of biomarkers and Artificial Intelligence (AI)-based approaches for individualized risk assessment. Numerous biomarkers and circulating tumor markers have been investigated for early irAEs prediction; however, their performance remains inconsistent across studies [3]. More recently, AI has emerged as a promising approach for improving irAEs prediction and individualized risk stratification. The ability to integrate high-dimensional data, recognize nonlinear relationships, process structured and unstructured data is highly relevant to the biologic dynamic and uniqueness of irAEs. In this review, we draw the landscape of AI models developed for irAEs prediction and detection and highlight the present challenges in evidence before clinical implementation.

1.1. Advent of Immune Checkpoint Inhibitors and the Emergence of Immune-Related Adverse Events

The treatment of advanced malignancies has been reshaped by ICIs; monoclonal antibodies that block inhibitory immune pathways exploited by tumors to evade immune surveillance. Modern ICI era began with ipilimumab, an anti-cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) antibody and extended towards programmed cell death receptor 1 and ligand (PD-1, PD-L1) antibodies successful in checkpoint blockade across various malignancies [4–6]. Although combination blockade improves efficacy in selected malignancies, it is associated with substantially higher rates of severe immune-related adverse events than monotherapy [7]. This trade-off captures a defining feature of immunotherapy, where immune disinhibition that restores antitumor T-cell activity, also disrupts self-tolerance.

These toxicities, termed as irAEs, resemble autoimmune disorders, can affect any organ and vary in severity according to the checkpoint inhibitor used [5,6,8]. As ICI use expands, it imposes a clinical burden rather than a niche toxicity problem.

Real-world pharmacovigilance data for irAEs show skin reactions, pneumonitis, enterocolitis, and thyroiditis as most frequent events [9]. Severe and fatal irAEs remain uncommon but clinically significant as they progress rapidly. While fatal irAEs remain uncommon, myocarditis carries one of the highest-case-fatality rates among immune-related toxicities [10,11].

irAEs toxicity can occur during initial therapy or months after ICI discontinuation, and chronic sequelae often involve endocrine, rheumatologic, pulmonary, neurologic, and other systems. This temporal heterogeneity limits static risk models and motivates longitudinal analysis of electronic health record (EHR) time-series using recurrent or transformer-based models. Likewise, the organ-specific expression of irAEs, such as pneumonitis, colitis, myocarditis, and hypophysitis, supports the use of computer vision models where imaging or digital histology contributes to prediction [8,12]. The biological complexity underlying these diverse clinical presentations also explains why accurate risk prediction remains challenging using conventional clinical variables alone, providing a rationale for AI-based multimodal prediction approaches.

1.2. Mechanisms Underlying Immune-Related Adverse Events

The pathogenesis of irAEs remains incompletely defined, but its central feature is a therapeutic breach of peripheral immune tolerance. Checkpoint therapy can therefore restore antitumor immunity while simultaneously releasing autoreactive lymphocyte populations that were previously kept in check [13,14].

T-cell dysregulation is a dominant mechanism, where ICI-associated colitis demonstrated tissue-resident memory CD8⁺ T-cells releasing interferon-gamma [15] and weakening of regulatory T cell control by anti-CTLA4 therapy [13,14]. Higher pretreatment activated CD4⁺ T-cells with greater T-cell receptor diversity, which was associated with

subsequent severe irAEs, indicating that baseline immune architecture influences toxicity risk [16]. A second axis involves pro-inflammatory amplification particularly seen with interleukin-6 increasing irAEs risk in several cohorts. Epitope spreading could sustain inflammation as tissue injury exposes additional self-antigens. Humoral immunity in cases of thyroiditis, hypophysitis, and bullous pemphigoid has been associated with autoantibodies [8,13].

Cross-reactive immune recognition may help explain why some toxicities localize to specific organs. In fatal ICI-associated myocarditis, shared T-cell receptor clonotypes were identified in myocardium, skeletal muscle, and tumor [17]. Host and environmental modifiers further shape susceptibility. The gut microbiome is among the best-studied examples, with severe irAEs during combined CTLA-4 and PD-1 blockade associated with distinct microbial features, including enrichment of *Bacteroides intestinalis*, loss of *Ruminococcaceae* and mucosal interleukin-1 β activity [18,19]. Genetic susceptibility such as variations in human leukocyte antigen (HLA) and other immune-regulatory pathways may influence irAEs risk [13,14].

These mechanisms help explain why irAEs are biologically diverse, temporally dynamic, and difficult to predict from any single feature. Most importantly, their complexity provides the biological rationale for AI approaches capable of integrating multimodal clinical, molecular, imaging and longitudinal data, which is an emerging strategy highlighted in this review.

2. Traditional Approaches to Predicting irAEs

Current methods to evaluate adverse events and monitor for their occurrence rely on a wide array of routine laboratory tests, including complete blood count, liver, thyroid and kidney function tests, coagulation profiles, virology panels, immune cell profiling, cytokines, and imaging modalities such as chest x-rays, ultrasound, and computed tomography scans [20]. Brahmer et al. further outlined organ-specific evaluation and management recommendations for colitis, pneumonitis, diabetes, hypothyroidism, hepatitis, nephritis, and other irAEs [21]. However, these approaches remain limited by low predictive accuracy and lack of applicability. No established biomarker has been proven for routine clinical practice, and studies often lack sensitivity and specificity analyses, which limits their potential as biomarker-based predictors [20,22].

In practice, traditional prediction is therefore closer to risk recognition and surveillance than true individualized forecasting. Baseline clinical assessment remains essential because ICI-toxicity is influenced by patient age, sex, cancer type, prior autoimmune disease, performance status, comorbid cardiopulmonary disease, prior thoracic radiation, concomitant medications, treatment class, and whether patients receive monotherapy or combination CTLA-4/PD-1 blockade [21]. These variables help clinicians identify patients who may require closer follow-up, but they do not reliably determine which patients will develop a specific toxicity, when the event will occur, or how severe it will become.

Guideline-based monitoring is also organ-specific rather than predictive. Current recommendations emphasize baseline history, physical examination, medication review, laboratory testing, and symptom-directed evaluation during therapy, with particular attention to dermatologic, gastrointestinal, hepatic, endocrine, pulmonary, renal, neurologic, rheumatologic, and cardiac toxicities [21]. Laboratory surveillance can detect biochemical hepatitis, thyroid dysfunction, nephritis, adrenal insufficiency, and other subclinical abnormalities, while imaging is used when symptoms suggest pneumonitis, colitis complications, or myocarditis. This approach improves recognition and management, but it remains reactive because abnormalities often become useful only after tissue injury has begun.

Biomarker-based prediction has been studied extensively, including cytokines, autoantibodies, routine blood counts, neutrophil-to-lymphocyte ratio, eosinophils, T-cell repertoire features, HLA alleles, soluble immune mediators, microbiome signatures, and organ-specific markers [20,22,23]. These studies support the biologic plausibility of risk stratification, but most biomarkers remain exploratory because cohorts are small, assays differ across studies, toxicity phenotypes are heterogeneous, and few markers have been validated prospectively with clinically useful sensitivity and specificity. As a result, no single biomarker has become a standard pretreatment test for irAEs prediction.

Traditional statistical models and claims-based approaches have attempted to combine baseline demographic, clinical, and medication variables to estimate severe irAEs risk. For example, Kalinich et al. used a population-level insurance claims database to predict severe irAEs requiring hospital admission among patients receiving ICIs, demonstrating that routinely collected data can support risk modeling at scale [24]. However, such approaches are constrained by coding accuracy, missing clinical detail, delayed capture of symptoms, and limited information on laboratory trajectories, imaging findings, and patient-reported symptoms. These limitations provide a strong rationale for the use of AI approaches capable of integrating longitudinal and multimodal data.

3. Artificial Intelligence in Oncology

Artificial Intelligence has enhanced the detection and characterization of imaging relevant to cancer diagnosis, treatment planning, and response assessment, thereby supporting clinical decision-making in oncology [25]. AI utilizes real data along with self-operating algorithms to produce solutions, mainly through Machine Learning (ML) and subsequent Deep Learning (DL) models [26]. DL models are particularly powerful because they learn hierarchical feature representations from complex real-world data through multiple algorithmic layers, enabling their application for complex healthcare tasks [27]. Algorithms may be predictive, i.e., trained to produce solutions from learned datasets or, generative, i.e., designed to develop solutions not explicitly present in the training datasets [26]. AI-based approaches for irAEs can be broadly categorized based on their clinical application, including pretreatment or early-treatment risk prediction and stratification, as well as post hoc detection, phenotyping and surveillance systems applied to structured, imaging, or unstructured data. This allows for future complex multimodal integration and risk stratification within the model that ultimately improves irAEs prediction.

4. AI Models and Approaches for Predicting irAEs

4.1. Clinical Data-Driven Machine Learning Modeling

The use of ML models has been studied to personalize therapeutic decisions, weigh risk-benefit profiles and has been primarily focused on pre or early treatment risk analysis to predict irAEs occurrences and mortality. Several studies have developed models that identify one or more of these adverse events using routinely accessible clinical, laboratory and patient-reported data. According to these studies, predictive performance varies by the type of toxicity and data modality. Studies predicting thyroid and cardiovascular irAEs, show poor to moderate discriminative power with an area under the receiver operating characteristic curve (AUROC) of 0.6–0.7. However, performances are influenced by cohort type and size, endpoint definitions as well as the ongoing challenge to detect less frequent yet clinically significant irAEs [28,29]. Heterogeneous predictive utility can be seen in studies using biomarkers, which can be attributed to the clinical and biological variability of the population [30,31]. In a mortality and risk stratification study, which integrated multidimensional clinical features, the ML model achieved an AUROC of 0.844 (95% CI:

0.826–0.862) in the development cohort and 0.786 (95% CI: 0.754–0.818) in the validation cohort utilizing 12 biomarkers alone. It must be noted that the study produced highly reliable results due to inclusion of structured validation cohorts beyond cross-validation, and this emphasizes importance of robust validation [32]. Similar response was seen in a checkpoint inhibitor-related pneumonitis (CIP) prediction model showing moderate to good discriminative power (AUROC 0.81; 95% CI 0.58–0.90); however, performance remained sensitive to cohort size and selected features [33].

Structured data driven models are attractive for clinical implementation because they leverage routinely collected clinical information, making them readily accessible and relatively inexpensive. For example, a model incorporating electronic patient-reported outcomes (ePRO) achieved high accuracy for detecting both incident and existing irAEs (AUROC 0.93), highlighting the value of real-time symptom monitoring [34]. Similarly, a large real-world study using routine EHR demonstrated moderate discrimination across multiple outcomes, including pneumonitis, hepatitis, colitis, and 1-year overall survival (AUROC 0.729–0.755). The authors further evaluated the model in a simulated trial scenario, where the model-guided prioritization identified a hypothetical patient cohort with over 50% lower toxicity incidence and higher estimated 1-year overall survival (68% vs. 52%) against random selection, illustrating the potential utility of structured data models for patients stratification [35]. Several studies also reported complementary performance metrics apart from AUROC, including classification accuracies of up to 97% [34] and nomogram-based C-indices approaching 0.90 [36]. However, these findings should be interpreted cautiously given the methodological limitations discussed below.

Across the several ML models (seen in Table 1), several patterns can be observed where generalizability is consistently constrained by common methodological weaknesses. Mortality prediction models generally demonstrate higher discriminative power than toxicity prediction models, likely because survival represents a more frequent and objectively defined endpoint than the heterogeneous clinical manifestations of irAEs. On the contrary, models predicting individual toxicities, particularly rare irAEs, typically achieve moderate discrimination, reflecting challenges related to class imbalance, variable endpoint definitions, and limited pretreatment biological features. Among individual toxicities CIP currently appears to be the best-supported application of ML-based prediction, with multiple independent studies demonstrating consistently moderate to high predictive performance across different algorithms and cohorts. In contrast, prediction of other organ-specific irAEs, including thyroid dysfunction and inflammatory arthritis, remains supported by fewer studies with less consistent validations. Importantly, even studies using large cohorts and real-world datasets [30,32,35,37,38] do not exceed the AUROC performance range, suggesting that increasing sample size alone may be insufficient to overcome fundamental limitations. Furthermore, the near absence of independent external validation and the use of heterogeneous, non-standardized endpoints for adverse event severity severely limit clinical reproducibility and translational applicability. While increasing cohort sizes in real-world datasets helps stabilize model performance, sample size alone remains insufficient to overcome these fundamental validation gaps [39]. When comparing predictive performance across studies, differences in model accuracy are driven less by algorithmic sophistication than by feature richness, objective endpoint clarity, and validation rigor. Models that integrate dynamic, longitudinal data, such as ePRO or serial laboratory trends consistently outperform static baseline models by capturing early physiological shifts [34,40]. Furthermore, models predicting objective, hard endpoints like overall survival or mortality achieve superior discrimination compared to those predicting subjective or graded organ toxicities, where clinical ascertainment and billing code noise introduce significant misclassification. Con-

sequently, the optimal clinical use case for structured, clinical data-driven ML models lies in pretreatment risk stratification and broad mortality forecasting, where routinely accessible EHR and laboratory variables can reliably flag high-risk patient profiles prior to initiating immunotherapy.

Table 1. Clinical ML Models Predicting ICI-Related Toxicities and Survival Outcomes.

Study	Model Type	Population	AI/Statistical Model	Input Features	irAEs/Outcome Assessed	Predictive Performance	Top Performers (Model/Features)	Key Findings <i>n</i> = Number of Study Participants
Kim et al. 2021 [28]	Toxicity Prediction	Mixed solid tumors	RF, LR, EN, SVM ML models	Clinical features	Thyroid-related irAEs (hypo/hyperthyroidism)	AUROC 0.77 (95% CI 0.648–0.883)	- - RF model Smoking history, hypertension, opioid use, heart disease	<i>n</i> = 187 Retrospective cohort; No EV
Zhou et al. 2022 [30]	Toxicity Prediction	Advanced NSCLC	LASSO, GLM, SVM, Decision Tree, RF, XGBoost, GBM ML models	Pretreatment blood-test biomarkers	Any grade irAEs	Median AUROC ~0.57–0.58	- - XGBoost and LASSO model PLR, CRP, TSH, NLR, ALB, ALP, LDH	<i>n</i> = 1320 Biomarkers alone insufficient for clinical use (low predictive power). Poor discrimination despite large multicenter cohort. No EV
Hazim et al. 2025 [31]	Toxicity Prediction	Mixed cancers	XG Boost and LR	71 clinical features: baseline PFTs, ICI type, CBC values, monocyte and WBC count, PD-L1 expression	High vs. low grade CIP	XGBoost AUROC = 0.74	- baseline DLCO, rest of input features	<i>n</i> = 170 Retrospective small size single-center. Limited generalizability, overfitting risk. No EV
	Survival/Outcome	Mixed cancers with confirmed CIP		Oxygen dependence, baseline PFTs, lymphocyte count, CBC /CMP labs, PD-L1 expression	Death from pneumonitis	XGBoost AUROCs 0.75–0.85; LR AUROCs 0.85–0.89	- All input features	
Wu et al. 2023 [32]	Survival/Outcome	NSCLC, transitional bladder cell carcinoma and RCC	Supervised ML (XGBoost, RF, LASSO-LR, SVM, KNN)	235 baseline clinical variables → 12 LASSO-selected (CRP, PD-L1, ECOG, ALP, ALB, hemoglobin, WBC, dNLR, pulse, metastatic sites, liver metastasis, cancer type)	Mortality (primary), PFS, irAEs, treatment/study discontinuation, grade 3–5 TRAEs	XGBoost AUROC = 0.79 RF AUROC = 0.786 LR Lasso AUROC = 0.781 AUROC 0.707–0.79	- - XGBoost model overall RF (best calibrated and overall clinical balance model)	<i>n</i> = 2538 Retrospective pooled trials. Imputation bias risk; No EV
Gong et al. 2025 [33]	Toxicity Prediction	Mixed cancers	EN logistic prediction model	Clinical characteristics and treatment variables and laboratory biomarkers	CIP prediction	AUROC 0.81 (95% CI 0.79–0.84 validation; 0.55–0.90 test)	- - Underlying diseases, lung disease history, NSCLC, KPS, body temperature, CD4%, hemoglobin, prior therapy, ICI type	<i>n</i> = 190 Small single-center retrospective cohort. Moderate discrimination. No EV. Class imbalance (48 vs. 142)
Iivanainen et al. 2021 [34]	Toxicity Prediction	Advanced cancers	XGBoost (compared with LR, EN, SVM, LightGBM, RF)	ePRO, age, sex, time from treatment initiation; total 75 days before diagnosis	Presence and early onset of irAEs (0–21 days before diagnosis)	Presence model: AUROC 0.99, Accuracy 0.97, Onset model: AUROC 0.93, Accuracy 0.96,	- - Presence: diarrhea, joint pain, dizziness, age, treatment duration; Onset: fever, chest pain, stomach pain, dizziness, nausea, diarrhea	<i>n</i> = 34 Small sample. single-center; ~40% irAEs prevalence. Possible overfitting (repeated observations from same patients). No EV
Lippenszky et al. 2024 [35]	Toxicity Prediction	Mixed solid tumors	RF	EHR-derived features: demographics, ICD diagnosis codes, smoking history, laboratory values, BMI, O ₂ saturation, medications, procedures, prior chemotherapy, ICI regimen	irAEs risk prediction within 1 yr of ICI initiation (hepatitis, colitis, pneumonitis)	Pneumonitis AUROC 0.739; Hepatitis AUROC 0.729; Colitis AUROC 0.755	- - Colitis model Outcome-specific; included smoking status, lung disease codes, oxygen saturation, BMI (pneumonitis); liver enzymes (hepatitis); chemotherapy, CTLA-4 therapy, albumin, blood counts (colitis)	<i>n</i> = 2230 Retrospective single-center. Heterogenous cancer type and treatment. OS as proxy for ICI effectiveness. Lost to follow-up excluded
	Survival/Outcome			ALB, neutrophils, lymphocytes, LDH, ALP, ALT trends, packed cell volume, oxygen saturation, metastatic disease codes, age, chemotherapy exposure	1-yr OS after ICI initiation	AUROC 0.752 (95% CI 0.706–0.796)	- ALB, LDH, lymphocyte/neutrophil measures, oxygen saturation, metastatic disease burden, age, chemotherapy exposure	
Jia et al. 2023 [36]	Toxicity Prediction	Mixed solid tumors	Multivariable LR and nomogram	Baseline ILD, emphysema, CRP, AEC (Nomogram A); pulmonary metastasis, ILD, emphysema, CRP, AEC, SII (Nomogram B)	CIP (any grade or grade ≥ 2)	Nomogram A C-index: 0.827 (training), 0.860 (validation); Nomogram B C-index: 0.873 (training), 0.904 (validation)	- Nomogram models outperformed individual predictors	<i>n</i> = 460 Retrospective single-center study. Small number of grade ≥ 3 CIP cases; No EV
Tao et al. 2026 [37]	Toxicity Prediction	Lung cancer	Multivariable Cox based Nomogram	Sex, BMI, radiotherapy, chemotherapy, CRP, CD4/CD8, WBC, ALB/GLB, PLR, LMR, NPR, PAR, CLR	1-,1.5-,2-year CIP risk prediction	AUROC 0.733 (1 y), 0.756 (1.5 y), 0.764 (2 y); CV mean AUROC 0.733 C-index CV 0.712	- BMI, WBC	<i>n</i> = 2082 Retrospective Single-center. 22.6% CIP incidence (likely high-risk cohort, broader CIP definition or true real-world incidence). No EV

Table 1. Cont.

Tran et al. 2024 [38]	Toxicity Prediction	Mixed ICI cohort	Supervised ML (LR and RF)	Binary EHR codes (ICD, RxNorm, LOINC, CPT) pre/post ICI	ICI-induced inflammatory arthritis	LR AUROC: 0.77 RF AUROC: 0.81 Best reduced feature model: 0.81	RF (best AUROC 0.81)	<i>n</i> = 2451 Single-center EHR study. Low PPV due to rare outcome. Possible label noise from chart adjudication. No EV
--------------------------	------------------------	------------------	------------------------------	---	--	---	----------------------	---

Abbreviations: AEC, Absolute Eosinophil Count; ALB, albumin; ALB/GLB, albumin-to-Globulin Ratio; ALP, Alkaline Phosphatase; ALT, Alanine Aminotransferase; AUROC, area under the receiver operating characteristic curve; BMI, Body Mass Index; CD4/CD8, Cluster of Differentiation 4/Cluster of Differentiation 8 T-cell Ratio; CIP, Checkpoint Inhibitor Pneumonitis; CLR, C-Reactive Protein-to-Lymphocyte Ratio; CPT, Current Procedural Terminology; CRP, C-Reactive Protein; CV, cross-validation; dNLR, Derived Neutrophil-to-Lymphocyte Ratio; DLCO, Diffusing Capacity of the Lung for Carbon Monoxide; ECOG, Eastern Cooperative Oncology Group Performance Status; EHR, electronic health record; EN, Elastic Net; ePRO, Electronic Patient-Reported Outcomes; EV external validation; GBM, Gradient Boosting Machine; GLB, Globulin; GLM, Generalized Linear Model; ICD, International Classification of Diseases; ICI, immune checkpoint inhibitor; ILD, interstitial lung disease; irAEs, Immune-related adverse event; KPS, Karnofsky Performance Status; LASSO, least absolute shrinkage and selection operator; LDH, Lactate Dehydrogenase; LMR, Lymphocyte-to-Monocyte Ratio; LOINC, logical observation identifiers names and codes; LR, Logistic Regression; NLR Neutrophil-to-Lymphocyte Ratio; NPR, neutrophil-to-platelet ratio; NSCLC, non-small cell lung cancer; OS, overall survival; PAR, platelet-to-albumin ratio; PD-L1, programmed death-ligand 1; PFS, Progression-Free Survival; PFTs, Pulmonary Function Tests; PLR, Platelet-to-Lymphocyte Ratio; RCC, renal cell carcinoma; RF, random forest; RxNorm Standardized Drug Nomenclature System (U.S. National Library of Medicine); SII, Systemic Immune-Inflammation Index; SVM, Support Vector Machine; TSH, Thyroid-Stimulating Hormone; TRAE, Treatment-related Adverse Events; WBC, White Blood Cell Count.

4.2. Radiomics Feature and Imaging Based Modeling

Radiomics prediction of irAEs is relatively emerging given the limited evidence and small cohort sizes, designed for risk prediction prior or during early treatment. Radiomics has especially been studied to identify, predict and differentiate CIP through several CT features [41]. Given the structural and localized nature of pulmonary toxicity, the optimal clinical use case for radiomics and imaging-based AI is the early detection, baseline risk assessment, and diagnostic differentiation of CIP from routine baseline or early-treatment chest CT scans. Current studies show promising potential for CIP detection with AUROC values exceeding 0.86 with some cohorts surpassing 0.90 [42–44]. Early studies have also identified a range of quantitative pretreatment radiomic features associated with subsequent CIP risk highlighting the benefits of imaging features for prediction. Although no single radiomic feature has been consistently validated, predictive models commonly incorporate measures of lung texture heterogeneity and other higher-order textural features extracted from baseline CT imaging. However, variability in feature selection and extraction methods has limited reproducibility and hindered identification of universally applicable radiomic signatures [45,46]. Other studies combining radiomic signatures with clinical variables show improved accuracy of irAEs prediction, amplifying the multimodal approach [47,48]. Tan et al. further adopted DL techniques and 5-fold cross-validation to adequately predict CIP from baseline CT and clinical data owing to an AUROC of 0.918 and accuracy of 0.920, further supporting the feasibility of automated toxicity assessment [49].

Emerging models are heavily based on CIP prediction where studies show promising discrimination power in cohorts (highlighted in Table 2) although true clinical representation remains uncertain. There are hindrances that limit the formation of mature models, attributed to retrospective design increasing risk of overfitting and limiting external validity. Variability in imaging acquisition, reconstruction or feature extraction models also challenge cross-institutional reproducibility [41].

Table 2. Imaging-Based Radiomics Models Predicting ICI-Related Toxicity.

Study	Population	AI Model	Imaging	Toxicity	Performance	Radiomics Features	Top Predictive Features	Key Findings <i>n</i> = Number of Study Participants
Qiu et al. 2022 [42]	NSCLC	LASSO feature selection and rad score and LR nomogram	Chest CT	Classification of CIP vs. RP	Radiomics: 0.891 training, 0.901 validation Nomogram: AUROC 0.953 training, 0.947 validation	837 Radiomics features (first-order, GLCM, GLRLM, GLSZM, NGTDM, GLDM)	Wavelet-based first-order, GLCM, GLRLM, GLSZM, GLDM and NGTDM features (rad score = 11 selected features)	<i>n</i> = 126 Retrospective, classification study; Highly reproducible texture; Clinical features (bilateral changes and sharp borders) enhanced performance AUROC 0.95. Nomogram defined by rad score, bilateral changes and sharp border
Cheng et al. 2022 [44]	Lung Cancer	LR classifier (vs. RF and linear SVM)	Chest CT	CIP vs. RP	Internal 10-fold CV = 0.937 External/independent ICI and RT test cohort = 0.896	Intensity Histogram and GLCM & BoW features	Nine discriminative BoW visual-word patterns derived from 9 × 9 image patches	<i>n</i> = 73 Retrospective classification study. BoW texture features superior to conventional histogram and GLCM features.
Colen et al. 2018 [45]	Advanced solid tumors	Handcrafted radiomics and MMR feature selection and unsupervised anomaly detection	Baseline chest CT	ICI pneumonitis	Accuracy 100% (LOOCV (<i>p</i> = 0.0033)); No reported AUROC	1860 features (Histogram and GLCM texture features)	Skewness, Angular Variance of sum of squares	<i>n</i> = 32 Pilot study. High risk of overfitting. Exploratory feasibility study
Speiler et al. 2022 [46]	LLC syngeneic mouse model	Feature selection and binary LR	Pretreatment CT and MRI	Lung inflammation (CD45 infiltration surrogate for ICI pneumonitis)	CT model AUROC 0.787; MRI model AUROC 0.834 (two-fold internal cross-validation)	184 features (92 CT; 92 MRI); (geometric, first-histogram, second-order co-occurrence and thirds order joint probability features)	CT average gray, CT histogram kurtosis, and CT co-occurrence matrix entropy, MRI histogram kurtosis	<i>n</i> = 19 Preclinical study. Pretreatment radiomics along with NLR and GM-CSF predicted subsequent lung inflammation
Huigen et al. 2025 [50]	Melanoma and Lung Cancer	Radiomics and Biomarker & statistical analysis	FDG PET/CT	ICI-induced liver injury	0.8	110 PyRadiomics features (shape, intensity, texture)	SUV Kurtosis, NGTDM Contrast (<i>p</i> = 0.041), SUV Skewness, and Shape Flatness.	<i>n</i> = 16 Exploratory PCA study (radiomic biomarker discovery). Features suggest greater metabolic complexity and variability in liver FDG uptake.
Cousin et al. 2025 [51]	Advanced NSCLC/mixed tumors	Handcrafted CT radiomics and clinical feature using GLM and RF classifiers	Chest CT	CIP prediction	0.71 (Radiomic lobe GLM model)	Handcrafted radiomics (first-order and texture features)	NA	Combined model achieved moderate CIP prediction (AUC 0.75). Performance IV using nested CV only.
Yadav et al. 2025 [52]	Advanced NSCLC stage III-IV	RF radiomics classifier	Contrast enhanced chest CT	CIP prediction	0.59	First-order, Shape, GLCM, GLRLM, GLSZM, GLDM, NGTDM (tumoral and peritumoral CT radiomics)	NGTDM and GLCM texture features	Large single-center cohort. Radiomics failed to achieve clinical accuracy for CIP prediction despite extensive peritumoral and tumoral extraction

Abbreviations: AUROC, area under the receiver operating characteristic curve; BoW, Bag of Visual Words; CIP, Checkpoint Inhibitor Pneumonitis; CT, computed tomography; CV, cross-validation; FDG, Fluorodeoxyglucose; GLCM, Gray-Level Co-occurrence Matrix; GLDM, Gray-Level Dependence Matrix; GLM, Generalized Linear Model; GLRLM, Gray-Level Run Length Matrix; GLSZM, Gray-Level Size Zone Matrix; GM-CSF, Granulocyte-Macrophage Colony-Stimulating Factor; ICI, immune checkpoint inhibitor; IV, Internal Validation; LASSO, Least Absolute Shrinkage and Selection Operator; LLC, Lewis Lung Carcinoma; LOOCV, Leave-One Out Cross-Validation; LR, Logistic Regression; MRI, Magnetic Resonance Imaging; mRMR, Minimum Redundancy Maximum Relevance (feature selection method); NA, Not available; NGTDM, Neighborhood Gray Tone Difference Matrix; NLR, Neutrophil-to-Lymphocyte Ratio; NSCLC, Non-Small Cell Lung Cancer; PCA, Principal Component Analysis; PET/CT, Positron Emission Tomography/computed tomography; Rad score, Radiomics score; RF, Random Forest; RP, Radiation Pneumonitis; RT, radiotherapy; SUV, Standardized Uptake Value; SVM, Support Vector Machine.

4.3. NLP-Based Modeling

Natural language processing (NLP) systems enable rapid annotation of embedded words among abundant, clinical notes datasets enabling future toxicity risk predictions from unstructured clinical documentation [53]. These post hoc detection models, as detailed in Table 3, are designed with vector representations of the keywords manually extracted from charts and then fed into the training classifier. They are primarily designed for retrospective detection and phenotyping rather than true prospective risk prediction. An early study by using NLP-based DL model demonstrated improved performance, achieving an F1-score (harmonic mean of precision and recall) of 0.75 and AUROC of 0.85 for any irAEs detection, outperforming basic keyword filters and shallow ML models [54]. To gain insight on drug-irAEs causality, NLP based augmented curation approach was used on 9290 patients and compared with ICD diagnoses code filters. The Augmented curation model identified at least 70% of the irAEs that could be allocated by structured diagnoses coding, highlighting limitations of diagnoses code-based detection [55]. More recent irAEs detection with large language model (LLM) retrieval—augmented generation showed high sensitivity (98.1%),

specificity (95.7%) for selected irAEs such as hepatitis, myocarditis and pneumonitis irAEs, along with a 98.9% faster review speed compared to manual adjudication [56]. LLM-GPT models show strong performance in irAEs extraction, although results are dependent on prompt and annotation quality and error analysis depicted the need for textual clarification and pairing of irAEs with ICIs for model clarity [57].

Table 3. NLP-Based Models for Detection of ICI-Related Toxicity.

Study	Population	Study Design	Data Type	AI Model	Toxicity Predicted	Performance	Key Findings
Gupta et al. 2021 [54]	Mixed advanced cancers on ICI	Retrospective single-center	EHR clinical notes (unstructured)	Keyword filtering and word embeddings and ML classifiers)	Any irAEs and selected subtypes	AUROC = 0.85; F1 = 0.75	Single-center study. Subtype models built (skin, endocrine, colitis) but detailed per-organ metrics not fully reported numerically.
Barman et al. 2024 [55]	9290 ICI-treated patients	Retrospective EHR study	Unstructured EHR	AC	Myocarditis	Improved detection vs. ICD codes (no AUROC/F1 reported)	Phenotyping comparison study. No classifier performance metrics.
Sun et al. 2024 [56]	Hospitalized ICI-treated patients (multi-cancer)	Single-center retrospective with EV EV	EHR notes and manual adjudication	LLM with retrieval-augmented generation	Encephalitis	Sen 94.7%; Sp 93.7%	Capture improvement vs. ICD codes.
					Pneumonitis	Sen ~95–97%; Sp > 90%	Used steroid/ICI discontinuation proxies.
					Severe cutaneous adverse reactions	Sen > 95%; Sp > 90%	No gold-standard adjudicated metrics reported.
					Myocarditis	Sen > 90%; Sp > 90%	Rare event; smaller positive case count.
Bejan et al. 2026 [57]	ICI-treated patients (EHR and clinical trial datasets)	Retrospective multi-dataset validation	Clinical notes and trial reports	irAEs GPT (GPT-3.5/GPT-4/GPT-4o)	Gastrointestinal irAEs (incl. colitis)	F1 ≈ 0.81–0.85	Performance varied by LLM version; GPT-4o best performance
					Hematologic irAEs	F1 = 1.0	Small positive sample size.
					Hepatitis	F1 ≈ 0.32	Lower-performing category.
					Rash	F1 ≈ 0.31	Moderate-to-low performance; class imbalance.

Abbreviations: AC, augmented curation; AUROC, area under receiver operating characteristic curve; EHR, electronic health record; F1-score, Harmonic mean of precision and recall; GPT, generative pre-trained transformer; ICI, immune checkpoint inhibitor; irAEs, immune-related adverse event; LLM, Large Language Model; ML, Machine Learning; Sen, sensitivity; Sp, specificity.

While there exist large volumes of unstructured clinical data, identifying the accurate embedding terms or standard dictionary for irAEs would facilitate model maturity and utility for irAEs subtype classifications. Fewer studies have reported usable performance metrics for each irAEs, however there has been a clear significant evolution of methods from word embeddings towards GPT systems. Further, the addition of AI assistance during note taking could offer cleaner advantage for both the physician and model to annotate accurately [58]. Recent LLMs show higher significance in sensitivity and specificity than traditional NLP design. This explains an early yet promising outlook for the use of language models for irAEs detection. NLP models currently lack reproducibility due to non-universalized irAEs criteria across institutions, incomplete reporting of per-toxicity metrics (sensitivity, F1), and severe class imbalance. Hence, no mature unified clinical benchmark dataset exists for NLP based irAEs detection severely affecting its translational applicability. Crucially, while recent LLMs and retrieval-augmented generation architectures demonstrate remarkable retrospective sensitivity and significantly faster review speeds than manual chart adjudication or ICD codes alone, their clinical validation remains

narrow [56]. At present, the optimal clinical use case for NLP and language models is automated post hoc cohort identification, retrospective toxicity phenotyping, and large-scale pharmacovigilance surveillance from unstructured clinical narratives, rather than prospective real-time risk prediction.

4.4. Multimodal AI Modeling

The occurrence of immune toxicity is a systemic, multi-organ, heterogeneous phenomena shaped by tumor biology, microenvironment, host immune response, treatment exposure and genetic predisposition. Single modality systems target limited features of irAEs or fail to reflect the spectrum of immune dysregulation. This gives a strong rationale for a multimodal AI paradigm that integrates clinical, radiomics, structured and unstructured data streams to improve predictive robustness and clinical stratification of irAEs. A detailed analysis of the studies is provided in Table 4.

Table 4. Multimodal AI Models for ICI Related Toxicity Prediction.

Study	Population	Data Type	AI Model	Toxicity Predicted	Performance	Key Findings <i>n</i> = Number of Study Participants
Du et al. 2025 [48]	NSCLC	Contrast enhanced chest CT	Radiomics clinical LR model	CIP	0.935 (training), 0.905 (IV), 0.923 (EV)	<i>n</i> = 238 Top features: Age, baseline ILD, baseline emphysema, radiomics signature. Multicenter combined model outperformed clinical only and radiomics model. Strong performance seen in EV. Moderate to high methodological quality.
Tan et al. 2022 [49]	Lung cancer (NSCLC/SCLC)	Pretreatment chest CT imaging and clinical variables	3D CNN and multimodal fusion, transfer and contrastive learning	ICI-related pneumonitis	AUROC 0.918; Accuracy 0.920 (internal CV)	<i>n</i> = 48 Internal 5-fold CV only. Performance improved with multimodal fusion, two-stage transfer learning, and contrastive learning. High risk of bias (small case-control cohort; no EV)
Cousin et al. 2025 [51]	Advanced NSCLC/mixed tumors	Chest CT	Handcrafted CT radiomics and clinical feature using GLM and RF classifiers	CIP prediction	AUROC = 0.75	Combined radiomics and clinical modeling achieved better AUROC
Jing et al. 2020 [59]	Pan-cancer (ICI-treated cohorts)	TCGA Multi-omics and validation cohort and pharmacovigilance data	Binary regression/statistical ML biomarker model	Any irAEs	Validated LCP1/ADPGK biomarker model; no deployable clinical AUROC reported	Discovery-level study. Pharmacovigilance reporting bias. Limited EV.
Ayoub et al. 2024 [60]	Mixed solid tumors	ECG, EHR, laboratory variables	Multimodal deep neural network fusion	Composite ICI-related myocarditis and MACE	AUROC 0.717; NPV 0.98 for joint fusion model	Narrow endpoint (cardiac irAEs only). Limited EV
Frank et al. 2025 [61]	Real-world pan-cancer ICI cohort	Multimodal EHR clinical data (exact modalities not fully specified)	ML multimodal risk model (ensemble/tabular fusion)	Any-grade irAEs risk	Not reported	Pre-publication abstract. No architecture or validation details.

Abbreviations: 3D CNN, Three-Dimensional Convolutional Neural Network; AUROC, area under the receiver operating characteristic curve; CIP, Checkpoint Inhibitor Pneumonitis; CT, computed tomography; CV, cross-validation; ECG, electrocardiogram; EHR, electronic health record; EV external validation; GLM, Generalized Linear Model; ICI, immune checkpoint inhibitor; ILD, Interstitial Lung Disease; irAEs, immune-related adverse event; LCP1/ADPGK, Lymphocyte Cytosolic Protein 1 and ADP-Dependent Glucokinase; LR, Logistic Regression; MACE, Major Adverse Cardiovascular Events; ML, Machine Learning; NPV, negative predictive value; NSCLC Non-Small Cell Lung Cancer; RF Random Forest; SCLC Small Cell Lung Cancer; TCGA, The Cancer Genome Atlas.

In a large analysis from real-world and multi-omics data for predictors of irAEs, multimodal approach shows performance gains when integrating multiple data streams. Authors used stepwise linear regression showing improved prediction accuracy with CD8+ T-cells and TCR diversity (unexplained variance reduction from 0.59 to 0.44) and T-cell activation genes (univariate variance reduction to 0.17). The validation cohort confirmed these genes as promising predictive biomarkers of irAEs (AUROC 0.8) [59]. Multimodal DL techniques are significant as they can handle complex phenotypes when single sources are not reliable inference. DL approaches are complex to achieve and require extensive tuning and architectural optimization. Tan et al. developed a successful radiomics and clinical data DL model with improved predictive performance for toxicity risk using hybrid ML approaches from existing datasets [49].

ICI-related myocarditis and major adverse cardiac events prediction, through fusion of electrocardiographic data and EHR, outperformed baseline models achieving a high negative predictive value of 0.98 (AUROC of 0.717), indicating fusion model usage in irAEs screening [60]. The RADIOHEAD Cohort study developed a genomic integrated model

combining clinical notes, transcriptomic, and HLA data that showed modest performance for predicting colitis risk among severe irAEs (AUROC 0.68), demonstrating a threefold increase in HLA-driven colitis risk (Odds ratio 2.8) [61]. A review on AI models' predictive effect averaged at AUROC 0.83 where sensitivity and specificity improved upon combining structured, unstructured text and biomarker levels, supporting the theoretical advantage of multimodal fusion [62].

DL models focusing on transformer and graph-based architecture are significant as they handle missing modalities and complex feature interaction [63]. Advances in imaging and radiomics may further improve diagnostic accuracy of irAEs [64]. Multimodal models demonstrate superior theoretical and empirical potential for irAEs prediction and detection over single modality as they eliminate bias amplification. Currently, applicability is limited due to inadequate, variegated, non-standardized data, lacking external validation, predisposing it to increasing complexity and reduced model maturity. A key methodological concern is the lack of harmonization of data across institutions as well as during layering of the model input. Multimodal systems may be sensitive to missing data, creating imbalance that leads to bias and restricts generalizability across training datasets. Increasing architectural complexity further challenges interpretability, clinical trust and hinders regulatory adoption. These challenges highlight the issues of multimodal AI in oncology, where analytical processes have outpaced clinical translation. Ultimately, the optimal clinical use case for multimodal AI frameworks is comprehensive, systemic irAEs risk prediction, as integrating diverse data streams, such as pretreatment pathomics, radiomics, clinical notes, and multi-omics, is uniquely capable of capturing the complex, multi-organ biological interplay and tumor microenvironment risk states missed by uni-modal assessments. Figure 1 delineates a predictive pipeline of multimodal AI framework that depicts the core elements required to achieve clinical utility.

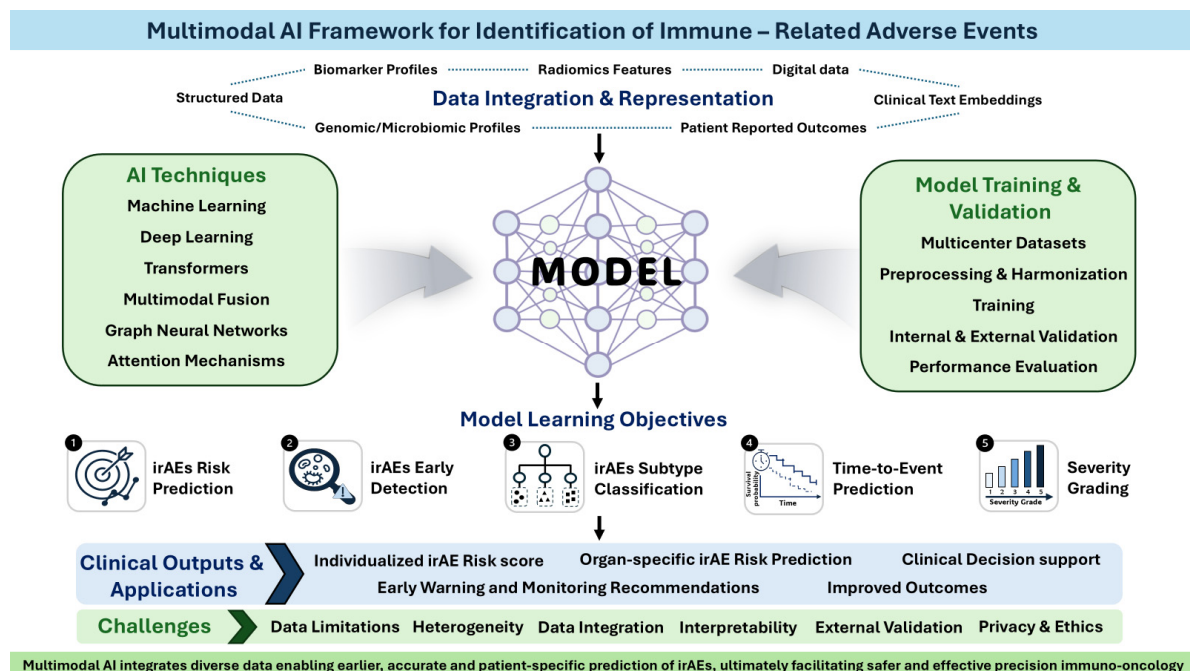


Figure 1. Overview of proposed multimodal AI predictive pipeline. The framework integrates diverse heterogeneous data modalities ranging from structured to unstructured data, radiomics and biomarkers. The data streams are processed through modality-specific encoders and unified using feature-fusion network to generate clinically predictive outputs. Key elements required to enrich model learning as well as validation have been described. Despite significant potential for enhanced clinical decision support, current limitations, albeit resolvable, hinder applicability.

5. Challenges and Limitations

A central challenge in AI-based irAEs prediction is the scarcity and fragmentation of high-quality data. Compared with AI studies of tumor detection or treatment response, irAEs prediction studies are fewer, often single-center, retrospective, and limited by small numbers of severe or organ-specific events. This is especially problematic because clinically important toxicities such as myocarditis, myositis, neurologic syndromes, and severe pneumonitis are uncommon but significant. Models trained on small event counts may appear promising internally while remaining unstable when applied to different populations.

Outcome heterogeneity further limits comparison across studies. Some models predict any irAEs, while others focus on grade 3 or higher toxicity, hospitalization, treatment interruption, corticosteroid use, pneumonitis, colitis, hepatitis, or endocrine events. These endpoints are not interchangeable. Mild rash, biochemical thyroiditis, and fulminant myocarditis differ in biology, timing, treatment, and clinical consequences. In addition, irAEs ascertainment may rely on clinician documentation, billing codes, manual chart review, NLP, or pharmacovigilance reports, each of which introduces different sources of misclassification.

Data quality and transferability are also major barriers. EHR contains missing laboratory values, inconsistent symptom documentation, variable follow-up intervals, and institution-specific treatment pathways. Imaging models depend on scanner protocols, segmentation methods, acquisition timing, and radiomics pipelines. Patient-reported outcome models can capture early symptoms but may be affected by adherence, digital access, and reporting fatigue [34]. Multimodal models are conceptually appealing, but aligning clinical notes, laboratory, imaging, genomic data, microbiome data, and medication timelines across sites is technically and operationally difficult.

Interpretability remains essential because irAEs prediction affects treatment selection, monitoring intensity, and patient counseling. Black-box risk scores may be difficult for oncologists to trust if they do not identify the clinical factors driving a prediction or distinguish modifiable surveillance needs from fixed baseline risk. Broader AI oncology concerns also apply, including algorithmic bias, dataset shift, privacy, regulatory oversight, and the possibility that models may perform worse in groups underrepresented in training data [26,27]. Without external validation and workflow integration, an AI model may add alert burden rather than improve care.

6. Future Directions

While general improvements such as prospective multicenter cohorts and standardized adverse event definitions remain field-wide imperatives, future research must move beyond generic recommendations to address priorities unique to each AI modality [65].

6.1. Clinical Machine Learning and Longitudinal Modeling

Future structured clinical ML efforts must pivot from static baseline prediction to dynamic time-series modeling. Because irAEs can manifest weeks or months after treatment initiation, sequential architectures, including long short-term memory (LSTM) networks and transformer-based temporal models, should continuously ingest evolving laboratory trajectories, medication changes, and real-time ePROs [66]. LSTM models have demonstrated superior sensitivity (95.2%) and F1-scores (0.97) compared to traditional classifiers in adverse drug reaction signal detection from longitudinal data [67]. However, most existing longitudinal AI models rely on single train–test partitions and exhibit poor reporting quality, with a third at high risk of bias [66]. Transportability, calibration, and temporal validation must be prioritized over internal discrimination metrics before clinical translation. Calibration has been termed the “Achilles heel” of predictive modeling, as

poorly calibrated models can render decision support tools clinically harmful. The TRIPOD-AI and PROBAST-AI frameworks provide structured guidance for transparent reporting and risk-of-bias assessment. Deployed models must also incorporate data drift detection and remediation mechanisms, as shifts in patient populations or coding conventions can degrade performance over time [68,69].

6.2. Radiomics and Organ-Specific Pipelines

Imaging research should prioritize organ-specific predictive pipelines, with CIP serving as the primary benchmark for clinical translation [70]. Priorities include harmonized imaging protocols, clearer semantic feature reporting, and evaluating deep foundation imaging models [71]. Furthermore, studies must establish dedicated cohorts for overlapping radiotherapy and immunotherapy (ICI + RT) exposure, applying probability recalibration prior to deployment in new clinical sites.

6.3. Natural Language Processing and Large Language Models

To transition NLP and LLMs from retrospective detection to prospective clinical utility, the field urgently requires standardized clinical benchmark datasets, unified irAEs annotation dictionaries, and expansion into outpatient EHR [72]. Future NLP pipelines must focus on broader toxicity coverage (including rare irAEs), advanced retrieval strategies capable of processing lengthy longitudinal charts, and determining causal drug-irAEs attribution rather than mere event mention [73]. Rigorous algorithmic fairness audits across diverse patient populations also remain essential.

6.4. Multimodal Integration and Actionable Decision Support

For multimodal AI, the field must move beyond generic “data integration” claims toward clinically feasible fusion of routine EHR, imaging, pathology, and selective multi-omics tailored to prespecified use cases. Recent multicenter studies demonstrate that integrating pretreatment H&E pathomics with clinical biomarkers can significantly improve CIP risk prediction (AUROC~0.91–0.93) by capturing tumor immune microenvironment states missed by single inputs [74]. Ultimately, clinical implementation depends on data harmonization, explainability, and developing actionable decision support frameworks that output specific surveillance recommendations (e.g., early pulmonary evaluation or cardio-oncology referral) to genuinely guide risk-adapted cancer immunotherapy.

7. Conclusions

Artificial Intelligence has displayed promising potential for prediction of ICI-related toxicities. Early model architectures have demonstrated feasibility of anticipating irAEs, particularly pneumonitis, using radiomics. Structured ML models currently contain the strongest body of evidence whereas NLP approaches have rapidly advanced through LLMs. Multimodal AI has the greatest long term clinical potential but remains the least mature for routine deployment. Across modalities, predictive performance improves with data integration; however, generalizability remains constrained by heterogeneous datasets and limited prospective validation. These models can significantly assist oncologists in delineating predictive pipelines for personalized monitoring, including time-to-event and survival outcome predictions, as well as early therapeutic intervention strategies. Future progress will depend on improved predictive biomarkers, standardization of structured data, integration of multi-omics datasets, and rigorous prospective validation. Together, these advances are essential to establish a clinically robust, gold-standard AI model for the prediction and timely detection of irAEs, ultimately improving the safety and effectiveness of ICI use in cancer immunotherapy.

Author Contributions: Conceptualization, E.N.A. and A.H.M.; Methodology, Writing—Original Draft Preparation, Writing—Review and Editing, E.N.A., M.S.A., A.H.M.; Visualization, E.N.A.; Supervision and Project Administration, A.H.M. and E.N.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Jayathilaka, B.; Mian, F.; Franchini, F.; Au-Yeung, G.; IJzerman, M. Cancer and treatment specific incidence rates of immune-related adverse events induced by immune checkpoint inhibitors: A systematic review. *Br. J. Cancer* **2025**, *132*, 51–57. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sullivan, R.J.; Weber, J.S. Immune-related toxicities of checkpoint inhibitors: Mechanisms and mitigation strategies. *Nat. Rev. Drug Discov.* **2022**, *21*, 495–508. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, J.; Ma, Y.; Lin, H.; Wang, J.; Cao, B. Predictive biomarkers for immune-related adverse events in cancer patients treated with immune-checkpoint inhibitors. *BMC Immunol.* **2024**, *25*, 8. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hodi, F.S.; O'Day, S.J.; McDermott, D.F.; Weber, R.W.; Sosman, J.A.; Haanen, J.B.; Gonzalez, R.; Robert, C.; Schadendorf, D.; Hassel, J.C.; et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N. Engl. J. Med.* **2010**, *363*, 711–723. [\[CrossRef\]](#) [\[PubMed\]](#)
- Larkin, J.; Chiarion-Sileni, V.; Gonzalez, R.; Grob, J.J.; Cowey, C.L.; Lao, C.D.; Schadendorf, D.; Dummer, R.; Smylie, M.; Rutkowski, P.; et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N. Engl. J. Med.* **2015**, *373*, 23–34. [\[CrossRef\]](#) [\[PubMed\]](#)
- Martins, F.; Sofiya, L.; Sykietis, G.P.; Lamine, F.; Maillard, M.; Fraga, M.; Shabafrouz, K.; Ribi, C.; Cairoli, A.; Guex-Crosier, Y.; et al. Adverse effects of immune-checkpoint inhibitors: Epidemiology, management and surveillance. *Nat. Rev. Clin. Oncol.* **2019**, *16*, 563–580. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ramos-Casals, M.; Brahmer, J.R.; Callahan, M.K.; Flores-Chávez, A.; Keegan, N.; Khamashta, M.A.; Lambotte, O.; Mariette, X.; Prat, A.; Suárez-Almazor, M.E. Immune-related adverse events of checkpoint inhibitors. *Nat. Rev. Dis. Primers* **2020**, *6*, 38. [\[CrossRef\]](#) [\[PubMed\]](#)
- Postow, M.A.; Sidlow, R.; Hellmann, M.D. Immune-Related Adverse Events Associated with Immune Checkpoint Blockade. *Engl. J. Med.* **2018**, *378*, 158–168. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gougis, P.; Jochum, F.; Abbar, B.; Dumas, E.; Bihan, K.; Lebrun-Vignes, B.; Moslehi, J.; Spano, J.P.; Laas, E.; Hotton, J.; et al. Clinical spectrum and evolution of immune-checkpoint inhibitors toxicities over a decade—A worldwide perspective. *EClinicalMedicine* **2024**, *70*, 102536. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, D.Y.; Salem, J.E.; Cohen, J.V.; Chandra, S.; Menzer, C.; Ye, F.; Zhao, S.; Das, S.; Beckermann, K.E.; Ha, L.; et al. Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. *JAMA Oncol.* **2018**, *4*, 1721. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nielsen, D.L.; Juhl, C.B.; Nielsen, O.H.; Chen, I.M.; Herrmann, J. Immune Checkpoint Inhibitor-Induced Cardiotoxicity: A Systematic Review and Meta-Analysis. *JAMA Oncol.* **2024**, *10*, 1390. [\[CrossRef\]](#) [\[PubMed\]](#)
- Johnson, D.B.; Nebhan, C.A.; Moslehi, J.J.; Balko, J.M. Immune-checkpoint inhibitors: Long-term implications of toxicity. *Nat. Rev. Clin. Oncol.* **2022**, *19*, 254–267. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dougan, M.; Luoma, A.M.; Dougan, S.K.; Wucherpennig, K.W. Understanding and treating the inflammatory adverse events of cancer immunotherapy. *Cell* **2021**, *184*, 1575–1588. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, L.; Huang, Y.; Xue, R.; Li, G.; Li, L.; Liang, L.; Lai, K.; Huang, X.; Qin, Y.; Zheng, Y. T cell-mediated mechanisms of immune-related adverse events induced by immune checkpoint inhibitors. *Crit. Rev. Oncol./Hematol.* **2025**, *213*, 104808. [\[CrossRef\]](#) [\[PubMed\]](#)
- Luoma, A.M.; Suo, S.; Williams, H.L.; Sharova, T.; Sullivan, K.; Manos, M.; Bowling, P.; Hodi, F.S.; Rahma, O.; Sullivan, R.J.; et al. Molecular Pathways of Colon Inflammation Induced by Cancer Immunotherapy. *Cell* **2020**, *182*, 655–671.e22. [\[CrossRef\]](#) [\[PubMed\]](#)

16. Lozano, A.X.; Chaudhuri, A.A.; Nene, A.; Bacchiocchi, A.; Earland, N.; Vesely, M.D.; Usmani, A.; Turner, B.E.; Steen, C.B.; Luca, B.A.; et al. T cell characteristics associated with toxicity to immune checkpoint blockade in patients with melanoma. *Nat. Med.* **2022**, *28*, 353–362. [[CrossRef](#)] [[PubMed](#)]
17. Johnson, D.B.; Balko, J.M.; Compton, M.L.; Chalkias, S.; Gorham, J.; Xu, Y.; Hicks, M.; Puzanov, I.; Alexander, M.R.; Bloomer, T.L.; et al. Fulminant Myocarditis with Combination Immune Checkpoint Blockade. *N. Engl. J. Med.* **2016**, *375*, 1749–1755. [[CrossRef](#)] [[PubMed](#)]
18. Andrews, M.C.; Duong, C.P.M.; Gopalakrishnan, V.; Iebba, V.; Chen, W.S.; Derosa, L.; Khan, M.A.W.; Cogdill, A.; White, M.G.; Wong, M.; et al. Gut microbiota signatures are associated with toxicity to combined CTLA-4 and PD-1 blockade. *Nat. Med.* **2021**, *27*, 1432–1441. [[CrossRef](#)] [[PubMed](#)]
19. Verheijden, R.J.; Van Eijs, M.J.M.; Paganelli, F.L.; Viveen, M.C.; Rogers, M.R.; Top, J.; May, A.M.; van de Wijert, J.H.; Suijkerbuijk, K.P. Gut microbiome and immune checkpoint inhibitor toxicity. *Eur. J. Cancer* **2025**, *216*, 115221. [[CrossRef](#)] [[PubMed](#)]
20. Liang, Y.; Maeda, O.; Ando, Y. Biomarkers for immune-related adverse events in cancer patients treated with immune checkpoint inhibitors. *Jpn. J. Clin. Oncol.* **2024**, *54*, 365–375. [[CrossRef](#)] [[PubMed](#)]
21. Brahmer, J.R.; Lacchetti, C.; Schneider, B.J.; Atkins, M.B.; Brassil, K.J.; Caterino, J.M.; Chau, I.; Ernstoff, M.S.; Gardner, J.M.; Ginex, P.; et al. Management of Immune-Related Adverse Events in Patients Treated With Immune Checkpoint Inhibitor Therapy: American Society of Clinical Oncology Clinical Practice Guideline. *J. Clin. Oncol.* **2018**, *36*, 1714–1768. [[CrossRef](#)] [[PubMed](#)]
22. Les, I.; Martínez, M.; Pérez-Francisco, I.; Cabero, M.; Teijeira, L.; Arrazubi, V.; Torrego, N.; Campillo-Calatayud, A.; Elejalde, I.; Kochan, G.; et al. Predictive Biomarkers for Checkpoint Inhibitor Immune-Related Adverse Events. *Cancers* **2023**, *15*, 1629. [[CrossRef](#)] [[PubMed](#)]
23. Pires Da Silva, I.; Ahmed, T.; McQuade, J.L.; Nebhan, C.A.; Park, J.J.; Versluis, J.M.; Serra-Bellver, P.; Khan, Y.; Slattery, T.; Oberoi, H.K.; et al. Clinical Models to Define Response and Survival With Anti-PD-1 Antibodies Alone or Combined With Ipilimumab in Metastatic Melanoma. *J. Clin. Oncol.* **2022**, *40*, 1068–1080. [[CrossRef](#)] [[PubMed](#)]
24. Kalinich, M.; Murphy, W.; Wongvibulsin, S.; Pahalyants, V.; Yu, K.H.; Lu, C.; Wang, F.; Zubiri, L.; Naranbhai, V.; Gusev, A.; et al. Prediction of severe immune-related adverse events requiring hospital admission in patients on immune checkpoint inhibitors: Study of a population level insurance claims database from the USA. *J. Immunother. Cancer* **2021**, *9*, e001935. [[CrossRef](#)] [[PubMed](#)]
25. Cheung, H.M.C.; Rubin, D. Challenges and opportunities for artificial intelligence in oncological imaging. *Clin. Radiol.* **2021**, *76*, 728–736. [[CrossRef](#)] [[PubMed](#)]
26. Kolla, L.; Parikh, R.B. Uses and limitations of artificial intelligence for oncology. *Cancer* **2024**, *130*, 2101–2107. [[CrossRef](#)] [[PubMed](#)]
27. Lotter, W.; Hassett, M.J.; Schultz, N.; Kehl, K.L.; Van Allen, E.M.; Cerami, E. Artificial Intelligence in Oncology: Current Landscape, Challenges, and Future Directions. *Cancer Discov.* **2024**, *14*, 711–726. [[CrossRef](#)] [[PubMed](#)]
28. Kim, W.; Cho, Y.A.; Kim, D.C.; Jo, A.R.; Min, K.H.; Lee, K.E. Factors Associated with Thyroid-Related Adverse Events in Patients Receiving PD-1 or PD-L1 Inhibitors Using Machine Learning Models. *Cancers* **2021**, *13*, 5465. [[CrossRef](#)] [[PubMed](#)]
29. Heilbroner, S.P.; Few, R.; Mueller, J.; Chalwa, J.; Charest, F.; Suryadevara, S.; Kratt, C.; Gomez-Caminero, A.; Dreyfus, B. Predicting cardiac adverse events in patients receiving immune checkpoint inhibitors: A machine learning approach. *J. Immunother. Cancer* **2021**, *9*, e002545. [[CrossRef](#)] [[PubMed](#)]
30. Zhou, J.G.; Wong, A.H.H.; Wang, H.; Tan, F.; Chen, X.; Jin, S.H.; He, S.S.; Shen, G.; Wang, Y.J.; Frey, B.; et al. Elucidation of the Application of Blood Test Biomarkers to Predict Immune-Related Adverse Events in Atezolizumab-Treated NSCLC Patients Using Machine Learning Methods. *Front. Immunol.* **2022**, *13*, 862752. [[CrossRef](#)] [[PubMed](#)]
31. Hazim, A.; Guiance, I.R.; Shreve, J.; Ruan, G.; McGlothlin, D.; LeMahieu, A.; Haemmerle, R.; Mcconn, K.; Godby, R.C.; Kottschade, L.; et al. Clinical characteristics, outcomes, and predictive modeling of patients diagnosed with immune checkpoint inhibitor therapy-related pneumonitis. *Cancer Immunol. Immunother.* **2025**, *74*, 194. [[CrossRef](#)] [[PubMed](#)]
32. Wu, Y.; Zhu, W.; Wang, J.; Liu, L.; Zhang, W.; Wang, Y.; Shi, J.; Xia, J.; Gu, Y.; Qian, Q.; et al. Using machine learning for mortality prediction and risk stratification in atezolizumab-treated cancer patients: Integrative analysis of eight clinical trials. *Cancer Med.* **2023**, *12*, 3744–3757. [[CrossRef](#)] [[PubMed](#)]
33. Gong, L.; Gong, J.; Sun, X.; Yu, L.; Liao, B.; Chen, X.; Li, Y.S. Identification and prediction of immune checkpoint inhibitors-related pneumonitis by machine learning. *Front. Immunol.* **2023**, *14*, 1138489. [[CrossRef](#)] [[PubMed](#)]
34. Iivanainen, S.; Ekstrom, J.; Virtanen, H.; Kataja, V.V.; Koivunen, J.P. Electronic patient-reported outcomes and machine learning in predicting immune-related adverse events of immune checkpoint inhibitor therapies. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 205. [[CrossRef](#)] [[PubMed](#)]
35. Lippenszky, L.; Mittendorf, K.F.; Kiss, Z.; LeNoue-Newton, M.L.; Napan-Molina, P.; Rahman, P.; Ye, C.; Laczi, B.; Csernai, E.; Jain, N.M.; et al. Prediction of Effectiveness and Toxicities of Immune Checkpoint Inhibitors Using Real-World Patient Data. *JCO Clin. Cancer Inform.* **2024**, *8*, e2300207. [[CrossRef](#)] [[PubMed](#)]
36. Jia, X.; Zhang, Y.; Liang, T.; Du, Y.; Li, Y.; Mao, Z.; Xu, L.; Shen, Y.; Liu, M.; Niu, G.; et al. Comprehensive nomogram models for predicting checkpoint inhibitor pneumonitis. *Am. J. Cancer Res.* **2023**, *13*, 2681–2701. [[PubMed](#)]

37. Tao, D.; Lei, H.; Sun, L.; Wang, L.; Zhou, W.; Wang, Y.; Wu, Y. Nomogram-based prediction of checkpoint inhibitor pneumonitis in lung cancer patients. *BMC Cancer* **2026**, *26*, 373. [[CrossRef](#)] [[PubMed](#)]
38. Tran, S.D.; Lin, J.; Galvez, C.; Rasmussen, L.V.; Pacheco, J.; Perottino, G.M.; Rahbari, K.J.; Miller, C.D.; John, J.D.; Theros, J.; et al. Rapid identification of inflammatory arthritis and associated adverse events following immune checkpoint therapy: A machine learning approach. *Front. Immunol.* **2024**, *15*, 1331959. [[CrossRef](#)] [[PubMed](#)]
39. Li, J.; Dan, K.; Ai, J. Machine learning in the prediction of immunotherapy response and prognosis of melanoma: A systematic review and meta-analysis. *Front. Immunol.* **2024**, *15*, 1281940. [[CrossRef](#)] [[PubMed](#)]
40. Katsura, H.; Suga, Y.; Kimoto, S.; Makihara, R.A.; Ota, H.; Toi, H.; Takata, N.; Ikesue, H. Effectiveness of electronic patient-reported outcomes in patients with cancer undergoing immunotherapy: A real-world retrospective study. *Int. J. Clin. Oncol.* **2026**, *31*, 1215–1225. [[CrossRef](#)] [[PubMed](#)]
41. Huang, J.; Chen, X.; Xia, B.; Ma, S. Advances in CT features and radiomics of checkpoint inhibitor-related pneumonitis: A short review. *Front. Immunol.* **2023**, *14*, 1082980. [[CrossRef](#)] [[PubMed](#)]
42. Qiu, Q.; Xing, L.; Wang, Y.; Feng, A.; Wen, Q. Development and Validation of a Radiomics Nomogram Using Computed Tomography for Differentiating Immune Checkpoint Inhibitor-Related Pneumonitis From Radiation Pneumonitis for Patients With Non-Small Cell Lung Cancer. *Front. Immunol.* **2022**, *13*, 870842. [[CrossRef](#)] [[PubMed](#)]
43. Chen, X.; Sheikh, K.; Nakajima, E.; Lin, C.T.; Lee, J.; Hu, C.; Hales, R.K.; Forde, P.M.; Naidoo, J.; Voong, K.R. Radiation Versus Immune Checkpoint Inhibitor Associated Pneumonitis: Distinct Radiologic Morphologies. *Oncologist* **2021**, *26*, e1822–e1832. [[CrossRef](#)] [[PubMed](#)]
44. Cheng, J.; Pan, Y.; Huang, W.; Huang, K.; Cui, Y.; Hong, W.; Wang, L.; Ni, D.; Tan, P. Differentiation between immune checkpoint inhibitor-related and radiation pneumonitis in lung cancer by CT radiomics and machine learning. *Med. Phys.* **2022**, *49*, 1547–1558. [[CrossRef](#)] [[PubMed](#)]
45. Colen, R.R.; Fujii, T.; Bilen, M.A.; Kotrotsou, A.; Abrol, S.; Hess, K.R.; Hajjar, J.; Suarez-Almazor, M.E.; Alshawwa, A.; Hong, D.S.; et al. Radiomics to predict immunotherapy-induced pneumonitis: Proof of concept. *Investig. New Drugs* **2018**, *36*, 601–607. [[CrossRef](#)] [[PubMed](#)]
46. Spieler, B.; Giret, T.M.; Welford, S.; Totiger, T.M.; Mihaylov, I.B. Lung Inflammation Predictors in Combined Immune Checkpoint-Inhibitor and Radiation Therapy-Proof-of-Concept Animal Study. *Biomedicines* **2022**, *10*, 1173. [[CrossRef](#)] [[PubMed](#)]
47. Mu, W.; Tunali, I.; Qi, J.; Schabath, M.B.; Gillies, R.J. Radiomics of ¹⁸F Fluorodeoxyglucose PET/CT Images Predicts Severe Immune-related Adverse Events in Patients with NSCLC. *Radiol. Artif. Intell.* **2020**, *2*, e190063. [[CrossRef](#)] [[PubMed](#)]
48. Du, Y.; Zhang, S.; Jia, X.; Zhang, X.; Li, X.; Pan, L.; Li, Z.; Niu, G.; Liang, T.; Guo, H. Radiomics Biomarkers to Predict Checkpoint Inhibitor Pneumonitis in Non-small Cell Lung Cancer. *Acad. Radiol.* **2025**, *32*, 1685–1695. [[CrossRef](#)] [[PubMed](#)]
49. Tan, P.; Huang, W.; Wang, L.; Deng, G.; Yuan, Y.; Qiu, S.; Ni, D.; Du, S.; Cheng, J. Deep learning predicts immune checkpoint inhibitor-related pneumonitis from pretreatment computed tomography images. *Front. Physiol.* **2022**, *13*, 978222. [[CrossRef](#)] [[PubMed](#)]
50. Huigen, C.M.C.; Coukos, A.; Latifyan, S.; Nicod Lalonde, M.; Schaefer, N.; Abler, D.; Depeursinge, A.; Prior, J.O.; Fraga, M.; Jreige, M. Function of ¹⁸F-FDG PET/CT radiomics in the detection of checkpoint inhibitor-induced liver injury (CHILI). *EJNMMI Rep.* **2025**, *9*, 27. [[CrossRef](#)] [[PubMed](#)]
51. Cousin, F.; Louis, T.; Frères, P.; Guiot, J.; Occhipinti, M.; Bottari, F.; Vos, W.; Hustinx, R. Machine Learning Model Integrating CT Radiomics of the Lung to Predict Checkpoint Inhibitor Pneumonitis in Patients with Advanced Cancer. *Technol. Cancer Res. Treat.* **2025**, *24*, 15330338251344004. [[CrossRef](#)] [[PubMed](#)]
52. Yadav, M.; Woo, W.; Chae, Y.K.; Lee, J.; Kim, P.H.; Lee, S.; Um, T.; Lee, S.; Chuchuca, M.J.A.; Djunadi, T.A.; et al. Radiomics Models to Predict Tumor Response and Pneumonitis in Non-Small Cell Lung Cancer Patients Treated with Immunotherapy. *J. Clin. Med.* **2025**, *14*, 4330. [[CrossRef](#)] [[PubMed](#)]
53. Whittington, C.; Feinman, T.; Zelman Lewis, S.; Lieberman, G.; Del Aguila, M. Clinical practice guidelines: Machine learning and natural language processing for automating the rapid identification and annotation of new evidence. *J. Clin. Oncol.* **2019**, *37*, 77. [[CrossRef](#)]
54. Gupta, S.; Belouali, A.; Shah, N.J.; Atkins, M.B.; Madhavan, S. Automated Identification of Patients With Immune-Related Adverse Events From Clinical Notes Using Word Embedding and Machine Learning. *JCO Clin. Cancer Inform.* **2021**, *5*, 541–549. [[CrossRef](#)] [[PubMed](#)]
55. Barman, H.; Venkateswaran, S.; Santo, A.D.; Yoo, U.; Silvert, E.; Rao, K.; Raghunathan, B.; Kottschade, L.A.; Block, M.S.; Chandler, G.S.; et al. Identification and Characterization of Immune Checkpoint Inhibitor-Induced Toxicities From Electronic Health Records Using Natural Language Processing. *JCO Clin. Cancer Inform.* **2024**, *8*, e2300151. [[CrossRef](#)] [[PubMed](#)]
56. Sun, V.H.; Heemelaar, J.C.; Hadzic, I.; Raghu, V.K.; Wu, C.Y.; Zubiri, L.; Ghamari, A.; LeBoeuf, N.R.; Abu-Shawar, O.; Kehl, K.L.; et al. Enhancing Precision in Detecting Severe Immune-Related Adverse Events: Comparative Analysis of Large Language Models and International Classification of Disease Codes in Patient Records. *J. Clin. Oncol.* **2024**, *42*, 4134–4144. [[CrossRef](#)] [[PubMed](#)]

57. Bejan, C.A.; Wang, M.; Venkateswaran, S.; Bergmann, E.A.; Hiles, L.; Xu, Y.; Chandler, G.S.; Brondfield, S.; Silverstein, J.; Wright, F.; et al. irAE-GPT: Leveraging large language models to identify immune-related adverse events in electronic health records and clinical trial datasets. *EBioMedicine* **2026**, *127*, 106227. [[CrossRef](#)] [[PubMed](#)]
58. Tierney, A.A.; Gayre, G.; Hoberman, B.; Mattern, B.; Ballesca, M.; Kipnis, P.; Liu, V.; Lee, K. Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation. *NEJM Catal. Innov. Care Deliv.* **2024**, *5*, CAT-23. [[CrossRef](#)]
59. Jing, Y.; Liu, J.; Ye, Y.; Pan, L.; Deng, H.; Wang, Y.; Yang, Y.; Diao, L.; Lin, S.H.; Mills, G.B.; et al. Multi-omics prediction of immune-related adverse events during checkpoint immunotherapy. *Nat. Commun.* **2020**, *11*, 4946. [[CrossRef](#)] [[PubMed](#)]
60. Ayoub, C.; Appari, L.; Pereyra, M.; Farina, J.M.; Chao, C.J.; Scalia, I.G.; Mahmoud, A.K.; Abbas, M.T.; Baba, N.A.; Jeong, J.; et al. Multimodal Fusion Artificial Intelligence Model to Predict Risk for MACE and Myocarditis in Cancer Patients Receiving Immune Checkpoint Inhibitor Therapy. *JACC Adv.* **2025**, *4*, 101435. [[CrossRef](#)] [[PubMed](#)]
61. Frank, A.; Brunovlenskaia-Bogoiavlenskaia, A.; Vasileva, T.; Savchenko, M.; Bolshakov, E.S.; Bolshakova, A.; Kust, S.A.; Wang, I.; Markova, K.L.; Bezlepikina, S.; et al. 1094 Pre-treatment predictive modeling of immune-related adverse event risk in immune checkpoint blockade therapy: A multi-modal machine learning approach from a real-world setting (RADIOHEAD Cohort Study). In *Regular and Young Investigator Award Abstracts*; BMJ Publishing Group Ltd.: London, UK, 2025; pp. A1245–A1246. [[CrossRef](#)]
62. Abdeldjouad, F.Z.; Brahami, M.; Sabri, M. Evaluating the Effectiveness of Artificial Intelligence in Predicting Adverse Drug Reactions among Cancer Patients: A Systematic Review and Meta-Analysis. In *2023 IEEE International Conference Challenges and Innovations on TIC (I2CIT)*; IEEE: New York, NY, USA, 2023; pp. 1–8. [[CrossRef](#)]
63. Waqas, A.; Tripathi, A.; Ramachandran, R.P.; Stewart, P.A.; Rasool, G. Multimodal data integration for oncology in the era of deep neural networks: A review. *Front. Artif. Intell.* **2024**, *7*, 1408843. [[CrossRef](#)] [[PubMed](#)]
64. Das, J.P.; Ma, H.Y.; De Jong, D.; Prendergast, C.; Baniasadi, A.; Braumuller, B.; Giarratana, A.; Khonji, S.; Paily, J.; Shobeiri, P.; et al. The evolving role of multimodal imaging, artificial intelligence and radiomics in the radiologic assessment of immune related adverse events. *Clin. Imaging* **2025**, *125*, 110571. [[CrossRef](#)] [[PubMed](#)]
65. Dhiman, P.; Ma, J.; Andaur Navarro, C.L.; Speich, B.; Bullock, G.; Damen, J.A.; Hooft, L.; Kirtley, S.; Riley, R.D.; Van Calster, B.; et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: A systematic review. *BMC Med. Res. Methodol.* **2022**, *22*, 101. [[CrossRef](#)] [[PubMed](#)]
66. Carrasco-Ribelles, L.A.; Llanes-Jurado, J.; Gallego-Moll, C.; Cabrera-Bean, M.; Monteagudo-Zaragoza, M.; Violán, C.; Zabaleta-del-Olmo, E. Prediction models using artificial intelligence and longitudinal data from electronic health records: A systematic methodological review. *J. Am. Med. Inform. Assoc.* **2023**, *30*, 2072–2082. [[CrossRef](#)] [[PubMed](#)]
67. Cao, M.; Chi, Y.; Yu, J.; Yang, Y.; Meng, R.; Jia, J. Adverse drug reaction signal detection via the long short-term memory model. *Front. Pharmacol.* **2025**, *16*, 1554650. [[CrossRef](#)] [[PubMed](#)]
68. Haue, A.D.; Hjaltelin, J.X.; Holm, P.C.; Placido, D.; Brunak, S.R. Artificial intelligence-aided data mining of medical records for cancer detection and screening. *Lancet Oncol.* **2024**, *25*, e694–e703. [[CrossRef](#)] [[PubMed](#)]
69. Hurkmans, C.; Bibault, J.E.; Brock, K.K.; van Elmpt, W.; Feng, M.; David Fuller, C.; Jereczek-Fossa, B.A.; Korreman, S.; Landry, G.; Madesta, F.; et al. A joint ESTRO and AAPM guideline for development, clinical validation and reporting of artificial intelligence models in radiation therapy. *Radiother. Oncol.* **2024**, *197*, 110345. [[CrossRef](#)] [[PubMed](#)]
70. Wang, S.; Wang, K.; Lin, J. Predictive value of machine learning for radiation pneumonitis and checkpoint inhibitor pneumonitis in lung cancer patients: A systematic review and meta-analysis. *Sci. Rep.* **2025**, *15*, 20961. [[CrossRef](#)] [[PubMed](#)]
71. Zhou, C.; Zhou, J.; Lv, Y.; Batuer, M.; Huang, J.; Zhong, J.; Zhong, H.; Qin, G. The impact of the novel CovBat harmonization method on enhancing radiomics feature stability and machine learning model performance: A multi-center, multi-device study. *Eur. J. Radiol.* **2025**, *184*, 111956. [[CrossRef](#)] [[PubMed](#)]
72. Golder, S.; Xu, D.; O'Connor, K.; Wang, Y.; Batra, M.; Hernandez, G.G. Leveraging Natural Language Processing and Machine Learning Methods for Adverse Drug Event Detection in Electronic Health/Medical Records: A Scoping Review. *Drug Saf.* **2025**, *48*, 321–337. [[CrossRef](#)] [[PubMed](#)]
73. Matheny, M.E.; Yang, J.; Smith, J.C.; Walsh, C.G.; Al-Garadi, M.A.; Davis, S.E.; Marsolo, K.A.; Fabbri, D.; Reeves, R.R.; Johnson, K.B.; et al. Enhancing Postmarketing Surveillance of Medical Products With Large Language Models. *JAMA Netw. Open* **2024**, *7*, e2428276. [[CrossRef](#)] [[PubMed](#)]
74. Boehm, K.M.; Khosravi, P.; Vanguri, R.; Gao, J.; Shah, S.P. Harnessing multimodal data integration to advance precision oncology. *Nat. Rev. Cancer* **2022**, *22*, 114–126. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.