

Artificial Intelligence for Predicting and Prioritizing Microbiological Hazards in Food Safety Inspection Systems: A PRISMA-Compliant Systematic Review (2020–2025)

Mohammad Mahdi Hemati Alam

Islamic Azad University Central Tehran Branch

Sana Zarrin

Islamic Azad University Central Tehran Branch

Yasaman Ghazagh

Islamic Azad University Central Tehran Branch

Amir Reza Hassan Poor Barkadehi

Islamic Azad University Central Tehran Branch

Arshia Farmahini Farahani

arshiafarahani916@gmail.com

Islamic Azad University Central Tehran Branch

Systematic Review

Keywords: artificial intelligence, machine learning, food safety, risk prediction, inspection prioritization, systematic review, PRISMA, microbiological hazards

Posted Date: April 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9403175/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Foodborne diseases represent a persistent and escalating global public health challenge, with the World Health Organization estimating approximately 600 million cases and 420,000 deaths annually. Traditional food safety regulatory frameworks, characterized by reactive inspection models and conventional microbial testing, are increasingly inadequate in the face of complex international supply chains, emerging pathogens, and evolving consumer demands. Artificial intelligence (AI) and machine learning (ML) techniques offer a transformative paradigm for risk prediction and inspection prioritization, moving beyond the constraints of classical predictive microbiology to enable integration of heterogeneous, high-dimensional data sources for proactive hazard management.

Objective

This systematic review aimed to identify, critically appraise, and synthesize peer-reviewed evidence on the application and efficacy of AI/ML models for predicting microbiological hazard risks and optimizing food safety inspection prioritization within regulatory systems, published between January 2020 and December 2025.

Methods

A systematic search of Scopus, Web of Science Core Collection, and PubMed was conducted following PRISMA 2020 guidelines. Eligibility criteria were structured using the PECO (Population, Exposure, Comparator, Outcomes) framework. Data extraction encompassed study design, AI/ML architecture, data sources, validation strategy, and performance metrics. Risk of bias was assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST). Due to substantial clinical and methodological heterogeneity among included studies, a narrative synthesis was performed rather than a quantitative meta-analysis.

Results

Of 2,148 records identified, six studies met the inclusion criteria. Tree-based ensemble models (Gradient Boosted Decision Trees, Random Forest, XGBoost) consistently demonstrated strong discriminative performance for facility-level risk stratification with tabular surveillance data, achieving accuracy values of 0.69–0.85 and R^2 values exceeding 0.65. Deep learning architectures, particularly YOLOv5 for image-based detection, achieved approximately 94% accuracy on unstructured data. However, PROBAST assessment revealed that three of six studies (50%) carried high overall risk of bias, predominantly due to reliance on internal validation without external temporal or geographic validation, raising concerns about model generalizability and potential overfitting.

Conclusions

AI/ML models demonstrate substantial potential for enhancing microbiological risk prediction and inspection resource optimization in food safety. However, the current evidence base is limited by a pervasive lack of rigorous external validation, small study numbers, and heterogeneity in modeling approaches. For AI-driven systems to achieve trustworthy and scalable regulatory implementation, prioritizing robust validation standards, mandating explainable AI (XAI) frameworks for transparency and accountability, and investing in data governance infrastructure are essential prerequisites. Future research should focus on multi-site external validation, standardized benchmarking datasets, and real-world pilot deployments to bridge the gap between algorithmic promise and operational impact.

Systematic Review Registration:

No formal protocol registration was completed prior to review conduct. Methods were predefined in accordance with PRISMA 2020 and PROBAST standards. The authors recommend future updates register with PROSPERO.

1. Introduction

1.1 The Global Burden and Economic Toll of Foodborne Illnesses

Foodborne diseases constitute one of the most significant and persistent public health challenges worldwide. The World Health Organization (WHO) estimates that approximately 600 million individuals suffer from foodborne illnesses each year, resulting in approximately 420,000 deaths and 33 million disability-adjusted life years (DALYs) lost globally [1]. These staggering figures, derived from the WHO's landmark burden of disease study spanning 2007–2015, likely underestimate the true scale of the problem, as many cases go unreported, particularly in low- and middle-income countries (LMICs) where surveillance infrastructure is limited [2]. Beyond the human toll, foodborne illnesses impose substantial economic burdens, including healthcare costs, productivity losses, trade disruptions, and the indirect costs of diminished consumer confidence in food systems. The economic impact in the United States alone has been estimated to exceed \$15 billion annually, while in LMICs, the combined cost of foodborne disease may consume a significant fraction of public health budgets [2].

The management of these risks has become increasingly complex due to the globalization of food supply networks, which introduce multiple points of potential contamination across vast geographical distances [2]. International trade in food products has grown substantially over the past two decades, meaning that a contamination event at a single processing facility can rapidly propagate across continents before detection. Simultaneously, the emergence of new pathogen variants, antimicrobial-resistant strains, and the influence of climate change on pathogen ecology further compound the challenge of ensuring food safety [3]. Climate-driven changes in temperature and precipitation patterns have been associated with altered distributions and seasonal patterns of foodborne pathogens, particularly *Salmonella*, *Campylobacter*, and *Vibrio* species, creating new risk profiles that existing surveillance systems may not adequately capture [4].

The standard regulatory model for food safety, which relies predominantly on reactive measures such as periodic inspections, end-product testing, and post-outbreak investigations, is fundamentally inadequate for addressing these evolving threats [5]. This model operates under the assumption that hazards can be detected through scheduled sampling and testing, yet the vast majority of foodborne contamination events are sporadic and occur at points in the supply chain that may not coincide with inspection schedules. Recognizing these limitations, global regulatory entities including the Codex Alimentarius Commission, the European Food Safety Authority (EFSA), and the U.S. Food and Drug Administration (FDA) have increasingly emphasized the need to transition from retrospective monitoring to forward-looking, predictive risk evaluation frameworks [6]. This evolution entails leveraging sophisticated data analytics tools to anticipate where and when hazards are likely to emerge, thereby enabling more efficient allocation of inspection resources and more effective preventive oversight [7].

1.2 Limitations of Classical Predictive Microbiology and the Advent of AI-Enhanced Systems

Classical predictive microbiology (PM) has long served as a cornerstone of risk assessment within Hazard Analysis Critical Control Point (HACCP) systems and quantitative microbial risk assessment (QMRA) frameworks. Traditional PM models employ deterministic or semi-probabilistic mathematical equations such as the Gompertz, Baranyi, or logistic growth models to describe microbial growth, survival, and inactivation as functions of environmental parameters including temperature, pH, water activity, and preservative concentration [8]. While these models have provided valuable insights for establishing food safety criteria and shelf-life determinations, they are subject to several well-documented limitations that constrain their applicability in contemporary food safety management.

First, classical PM models typically assume relatively simple, often linear or monotonically sigmoidal dose response relationships under controlled environmental conditions, which poorly represent the complex, nonlinear, and interactive dynamics of microbial behavior in real-world food systems [9]. Food matrices are inherently heterogeneous, and microbial populations are subject to fluctuating conditions during processing, transportation, and storage that challenge the assumptions of steady-state laboratory models. Second, traditional models struggle to integrate diverse data types such as genomic information, weather data, inspection histories, and supply chain logistics that collectively influence risk profiles but fall outside the parameter space of conventional growth models [9]. Third, the static nature of most classical models limits their capacity for real-time risk updating as new information becomes available, a critical shortcoming in rapidly evolving outbreak scenarios.

The integration of artificial intelligence (AI) and machine learning (ML) into food safety represents a paradigm shift toward augmented predictive microbiology, enabling the analysis of multifaceted, heterogeneous data streams that are essential for granular risk profiling [10]. Unlike traditional models, AI/ML approaches can learn complex, nonlinear relationships directly from data without requiring a priori specification of functional forms, making them particularly well-suited for the high-dimensional and often noisy data environments encountered in food safety applications [10]. AI's capacity to discern intricate patterns in large datasets is especially relevant for addressing emerging challenges, such as predicting the impact of climate variability on pathogen occurrence, where the relationships between environmental drivers and microbial risk are complex and may involve lagged effects, threshold behaviors, and spatial heterogeneity [4]. Recent advances in deep learning,

ensemble methods, and natural language processing have further expanded the repertoire of tools available for food safety risk prediction, encompassing applications ranging from image-based pathogen detection to outbreak source attribution and facility-level risk scoring [10,11].

1.3 Objectives and Scope of This PRISMA Systematic Review

The primary objective of this systematic review is to identify, critically appraise, and synthesize peer-reviewed evidence published between January 1, 2020, and December 1, 2025, on the application and demonstrated efficacy of AI and ML models for predicting microbiological hazard risks and optimizing food safety inspection resource allocation within regulatory systems. Specifically, this review addresses the following research questions: (1) Which AI/ML architectures and modeling approaches have been applied to microbiological risk prediction in food safety, and what are their reported performance characteristics? (2) What types of data sources and predictor variables have been utilized in these models? (3) What is the quality of the evidence base with respect to risk of bias, as assessed by the PROBAST framework? (4) To what extent have AI/ML models been validated for real-world operational deployment in food safety inspection prioritization?

The focus is deliberately restricted to studies in which predictive capability is directed toward actionable prevention strategies and measurable improvements in inspection efficiency, thereby excluding purely exploratory or proof-of-concept analyses without empirical validation. By systematically evaluating the current state of evidence, this review aims to identify gaps in methodology and validation that must be addressed before AI-driven tools can be responsibly integrated into food safety regulatory practice, and to provide practical recommendations for researchers and regulatory agencies seeking to adopt these technologies.

2. Methods

2.1 Protocol Registration and Reporting Standards

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [12]. The PRISMA 2020 checklist was followed to ensure transparency and completeness of reporting. As a rapid evidence synthesis conducted to inform time-sensitive regulatory discussions, no formal protocol registration was completed in a public registry prior to review conduct. However, the methods were predefined and documented internally before the search was initiated. The authors acknowledge that prospective registration in PROSPERO or the Open Science Framework (OSF) is recommended for future updates of this review to enhance transparency and reduce the risk of selective outcome reporting.

2.2 Eligibility Criteria (PECO Framework)

Eligibility criteria were structured using the Population, Exposure, Comparator, Outcomes (PECO) framework to ensure a transparent and reproducible study selection process:

Population (P)

Food items, food processing environments, or food supply chains subject to microbiological risk, encompassing key pathogens including *Salmonella* spp., *Listeria monocytogenes*, *Campylobacter* spp., and

pathogenic *Escherichia coli*. Studies focused exclusively on chemical contaminants, allergens, or physical hazards were excluded.

Exposure/Intervention (E)

Application of supervised, unsupervised, or reinforcement learning AI/ML algorithms for the purposes of forecasting, classifying, or risk-stratifying microbiological hazards in food. This included but was not limited to tree-based ensemble methods, neural networks, support vector machines, and deep learning architectures applied to structured surveillance data, genomic data, or unstructured data such as images and text.

Comparator (C)

Traditional inspection scheduling, non-AI statistical models, or baseline performance metrics (e.g., random classifier or majority class prediction). Studies without an explicit comparator were included if validated predictive efficacy was reported against a meaningful benchmark.

Outcomes (O)

Primary outcomes comprised predictive efficacy measures, including area under the receiver operating characteristic curve (AUC-ROC), F1-score, precision, recall, accuracy, and coefficient of determination (R^2), for risk identification or classification. Secondary outcomes included measures of operational utility improvement, such as inspection efficiency gains or improved resource allocation metrics.

Additional inclusion criteria required that studies be original empirical research published in peer-reviewed journals, available in full-text English, and published between January 1, 2020, and December 1, 2025. Excluded were review articles, conference abstracts without full-text availability, editorials, commentaries, studies focusing on AI/ML for food quality (rather than safety), studies addressing chemical or physical hazards exclusively, and studies that did not report quantitative performance metrics.

2.3 Information Sources and Search Strategy

A systematic search of three major electronic databases Scopus, Web of Science Core Collection, and PubMed was conducted, with the final search executed on December 1, 2025. The search period covered publications from January 1, 2020, to December 1, 2025. Database-specific search strategies were developed using controlled vocabulary (e.g., MeSH terms in PubMed) and free-text terms, combining three concept blocks: (1) AI/ML terms, (2) food safety and microbiological hazard terms, and (3) risk prediction and inspection terms. Boolean operators (AND, OR) were used to construct the search logic, and strategies were adapted to the syntax requirements of each database.

The complete search strategy for each database is provided below:

Scopus

(TITLE-ABS-KEY("artificial intelligence" OR "machine learning" OR "deep learning" OR "predictive analytics" OR "neural network*" OR "random forest" OR "support vector machine" OR "gradient boosting" OR "ensemble method*") AND TITLE-ABS-KEY("food safety" OR "foodborne pathogen*" OR "microbial hazard*" OR "food contamination" OR salmonella OR listeria OR campylobacter OR "e. coli" OR "pathogen detection")) AND TITLE-

ABS-KEY("risk prediction" OR "risk scoring" OR "risk assessment" OR "inspection priorit*" OR "source attribution" OR "outbreak detection" OR "hazard analysis")) AND (PUBYEAR > 2019 AND PUBYEAR < 2026) AND (LANGUAGE (english))

Web of Science Core Collection

TS=("artificial intelligence" OR "machine learning" OR "deep learning" OR "predictive analytics" OR "neural network*" OR "random forest" OR "support vector machine") AND TS=("food safety" OR "foodborne pathogen*" OR "microbial hazard*" OR "food contamination") AND TS=("risk prediction" OR "risk assessment" OR "inspection priority" OR "outbreak detection") AND PY=(2020–2025) AND LA=(English)

PubMed: ("artificial intelligence"[MeSH Terms] OR "machine learning"[MeSH Terms] OR "artificial intelligence" [Title/Abstract] OR "machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract]) AND ("food safety" [MeSH Terms] OR "foodborne diseases"[MeSH Terms] OR "food contamination"[Title/Abstract] OR "pathogen detection"[Title/Abstract]) AND ("risk assessment"[MeSH Terms] OR "predictive"[Title/Abstract] OR "inspection" [Title/Abstract])) AND ("2020/01/01"[Date - Publication] : "2025/12/01"[Date - Publication]) AND English[lang]

Grey literature searches were conducted on the websites of the Food and Agriculture Organization (FAO), WHO, FDA, and EFSA. Reference lists of included studies and relevant review articles were also screened to identify additional eligible records (backward citation tracking).

2.4 Study Selection and Data Extraction

Identified records were imported into Mendeley Reference Manager (version 1.19.8) for deduplication. Duplicate records were removed first using the automated deduplication function and subsequently verified manually. Title and abstract screening was performed independently by two reviewers using the Rayyan systematic review tool, with discrepancies resolved through discussion or consultation with a third reviewer. Full-text articles were assessed against the eligibility criteria by two independent reviewers using standardized extraction forms in Microsoft Excel.

The data extraction form captured the following items: authors, year of publication, country of origin, study design, target pathogen(s), data sources and characteristics (sample size, temporal coverage, geographic scope), AI/ML model architecture, input predictor variables, validation strategy (internal, external, temporal, geographic), primary performance metrics, and reported measures of operational utility or decision support improvement.

2.5 Methodological Quality Assessment

Risk of bias and concerns regarding applicability were assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST) [13], which evaluates four domains: participants, predictors, outcome, and analysis. Each domain was rated as low risk, high risk, or unclear risk of bias. The overall risk of bias was determined as high if any individual domain was rated high. Particular attention was given to the analysis domain, where internal validation without external validation and the use of inappropriate performance metrics were flagged as indicators of high risk of bias, consistent with the PROBAST guidance and the emerging PROBAST-AI framework [14]. Assessment was performed independently by two reviewers, with disagreements resolved through discussion.

2.6 Data Synthesis Strategy

Given the substantial heterogeneity among included studies in terms of pathogen targets, model architectures, data sources, validation approaches, and reported outcome metrics, a quantitative meta-analysis was not feasible. Instead, a narrative synthesis was conducted, grouping studies by application domain (risk stratification, source attribution, image-based detection), model architecture (tree-based ensembles, deep learning, statistical learning), and data type (surveillance, genomic, certification, imaging). The synthesis prioritized AUC-ROC and F1-score as the most informative metrics for evaluating model discriminative performance in the context of class-imbalanced datasets commonly encountered in food safety applications. Where reported, operational utility metrics were summarized qualitatively.

3. Results

3.1 Study Selection

The database searches yielded a total of 2,148 records (1,982 from electronic databases and 166 additional records identified through grey literature searches and citation tracking). After removal of 833 duplicates, 1,315 unique records remained for title and abstract screening. Of these, 1,196 records were excluded as clearly irrelevant. The remaining 119 full-text articles were assessed for eligibility, of which 113 were excluded for the following reasons: non-AI/ML methodology ($n = 28$), non-microbiological focus ($n = 25$), irrelevant outcomes ($n = 33$), and review or editorial articles ($n = 27$). Six studies met all inclusion criteria and were included in the narrative synthesis. The study selection process is summarized in the PRISMA 2020 flow diagram (Fig. 1).

3.2 Characteristics of Included Studies

The six included studies were published between 2021 and 2025 and spanned diverse geographic regions, including China, the United Kingdom, the United States, Italy, and multinational settings [15–20]. The most frequently studied pathogen was *Salmonella* ($n = 3$ studies), followed by *Campylobacter* ($n = 1$), *Listeria monocytogenes* ($n = 1$), and multi-pathogen investigations ($n = 1$). Data sources ranged from national surveillance databases and whole-genome sequencing (WGS) repositories to food safety certification records and agar plate images. Tree-based ensemble models were the most commonly employed model class, appearing in four of six studies, with one study employing deep learning (YOLOv5) for image-based detection and one utilizing statistical learning methods (Poisson regression and Elastic Net). The characteristics of included studies are summarized in Table 1.

Table 1
Characteristics of Included Studies

Author, Year	Country	Pathogen(s)	Data Source(s)	AI/ML Model(s)	Validation	Key Performance
Wang et al., 2021 [15]	China	Salmonella, Norovirus, E. coli, V. parahaemolyticus	National Foodborne Disease Surveillance	Gradient Boosted Decision Trees (GBDT)	10-fold CV	Accuracy = 0.69, F1 = 0.69
Arning et al., 2021 [16]	UK	Campylobacter	PubMLST WGS	XGBoost	Train/test split	Accuracy = 0.85
Karanth et al., 2023 [17]	USA	Salmonella	NORS, NOAA, NCBI WGS	Poisson regression, Elastic Net	10-fold CV	Pseudo R ² = 0.669
Zheng et al., 2023 [18]	US/Europe	Foodborne pathogens	Food safety certification data	Random Forest	Train/test split	Accuracy > 0.70
Garcia-Vozmediano et al., 2024 [19]	Italy	Salmonella	SeREMI, IZSPLVA surveillance	Random Forest	5-fold CV	R ² = 0.68
Ravindhiran & Dhandapani, 2025 [20]	Unknown	Listeria monocytogenes	Images on PALCAM agar	YOLOv5	Not specified	Accuracy = 0.94

3.3 Risk of Bias Assessment (PROBAST)

The PROBAST assessment revealed notable methodological weaknesses across the included studies (Table 2). Half of the studies (3/6) were rated as having high overall risk of bias, predominantly attributable to deficiencies in the analysis domain. The most common concern was the reliance on internal validation techniques (e.g., k-fold cross-validation) without external validation using independent temporal or geographic datasets. Wang et al. [15] and Karanth et al. [17] were rated as high risk due to the absence of external validation and potential overfitting to their respective training datasets. Ravindhiran and Dhandapani [20] received high risk of bias ratings due to concerns in the participants domain (unclear population representativeness) and the analysis domain (unspecified validation methodology). The remaining three studies Arning et al. [16], Zheng et al. [18], and Garcia-Vozmediano et al. [19] were rated as low risk across all PROBAST domains, although the absence of external validation in these studies also represents a limitation that warrants consideration when interpreting their findings.

Table 2
Risk of Bias Assessment Using PROBAST

Study	Participants	Predictors	Analysis	Outcome	Overall
Wang et al., 2021 [15]	Low	Low	Low	High	High
Arning et al., 2021 [16]	Low	Low	Low	Low	Low
Karanth et al., 2023 [17]	Low	Low	Low	High	High
Zheng et al., 2023 [18]	Low	Low	Low	Low	Low
Garcia-Vozmediano et al., 2024 [19]	Low	Low	Low	Low	Low
Ravindhiran & Dhandapani, 2025 [20]	Concerns	Low	Low	High	High

3.4 Synthesis of Predictive Modeling Performance

3.4.1 Comparative Model Architectures

Tree-based ensemble models consistently demonstrated strong performance for risk stratification tasks involving tabular surveillance data. Gradient Boosted Decision Trees (GBDT), XGBoost, and Random Forest architectures appeared in four of the six included studies, reflecting their well-established suitability for structured classification and regression problems in which predictor variables are heterogeneous and may include both continuous and categorical features [15–19]. Karanth et al. [17] and Garcia-Vozmediano et al. [19] reported R^2 values exceeding 0.65 using Random Forest and Elastic Net approaches, respectively, indicating that a substantial proportion of variance in outbreak occurrence could be explained by the combination of genomic features and meteorological or surveillance data. Wang et al. [15] achieved an accuracy of 0.69 and F1-score of 0.69 using GBDT for multi-pathogen classification, a respectable performance given the complexity of distinguishing among four distinct pathogen categories from surveillance data. Arning et al. [16] reported the highest accuracy (0.85) using XGBoost for *Campylobacter* source attribution from WGS data, suggesting that ensemble methods are particularly effective when high-quality genomic features are available as predictors.

In contrast, deep learning approaches were represented by a single study employing YOLOv5 for image-based *Listeria monocytogenes* detection on PALCAM agar plates, achieving approximately 94% accuracy [20]. This result highlights the potential of convolutional neural network (CNN) architectures for automated image-based pathogen identification, which could reduce the time and expertise required for colony recognition in laboratory settings. However, the absence of specification regarding validation methodology in this study limits the interpretability and generalizability of this finding. The divergence between tree-based methods (optimized for structured tabular data) and deep learning (optimized for unstructured data such as images and text) suggests that the choice of model architecture should be guided by the nature of the available data, with ensemble methods preferred for surveillance and risk scoring applications and deep learning reserved for image, genomic sequence, or text-based tasks.

3.4.2 Performance Metrics and Discriminatory Power

The choice of performance metrics varied across studies, reflecting differences in study objectives, class distributions, and outcome definitions. Accuracy was the most commonly reported metric, appearing in five of six studies, with values ranging from 0.69 to 0.94. However, accuracy alone can be misleading in the context of

class-imbalanced datasets a frequent occurrence in food safety surveillance, where positive cases (e.g., contaminated facilities or outbreak events) are substantially rarer than negative cases. The use of AUC-ROC and F1-score, which are more informative for imbalanced classification tasks, was less consistent across studies. Wang et al. [15] reported both accuracy and F1-score (0.69 each), providing a more balanced assessment of model performance. Arning et al. [16] reported accuracy and precision/recall measures, while Zheng et al. [18] reported accuracy exceeding 0.70 without specifying AUC or F1-score, making it difficult to assess the model's discriminative ability independently of class distribution effects.

For regression-based approaches, pseudo- R^2 and R^2 were the primary metrics. Karanth et al. [17] reported a pseudo- R^2 of 0.669 using Poisson regression with Elastic Net regularization, indicating moderate explanatory power for predicting salmonellosis outbreak counts from genomic and meteorological predictors. Garcia-Vozmediano et al. [19] reported an R^2 of 0.68 using Random Forest for Salmonella outbreak early warning in northwestern Italy, suggesting that the model captured a substantial portion of the variance in outbreak occurrence. These results are noteworthy given the inherent stochasticity and complexity of outbreak dynamics, though the absence of external validation in some cases raises questions about whether these performance levels would be maintained in different temporal or geographic contexts.

3.4.3 Operational Translation and Decision Utility

The translation of predictive accuracy into operational decision utility remains a critical but underexplored aspect of the current evidence base. Only a subset of included studies explicitly addressed the practical implications of their models for inspection resource allocation or risk-based decision-making. Zheng et al. [18] demonstrated that machine learning-based predictions incorporating food safety certification data could improve foodborne disease outbreak forecasting, with potential applications for prioritizing inspection and surveillance activities. The operational implication is that AI-driven targeting could identify more violations per unit of inspection effort compared to conventional scheduling approaches, thereby improving the efficiency of limited regulatory resources.

Similarly, Garcia-Vozmediano et al. [19] framed their Random Forest model as an early warning system for Salmonella outbreaks in northwestern Italy, emphasizing the potential for preemptive intervention before widespread transmission occurs. However, none of the included studies conducted a formal health economic evaluation, cost-benefit analysis, or prospective pilot deployment to quantify the real-world impact of AI-assisted inspection prioritization on public health outcomes. This gap between algorithmic performance and demonstrated operational benefit represents a significant limitation of the current evidence base and a key area for future research.

4. Discussion

4.1 Principal Findings and Contribution to the Field

This systematic review provides a comprehensive assessment of the current evidence on AI/ML applications for microbiological risk prediction and inspection prioritization in food safety, identifying six eligible studies published between 2020 and 2025. The principal finding is that tree-based ensemble models including GBDT, XGBoost, and Random Forest consistently demonstrate strong discriminative performance for facility-level and pathogen-level risk stratification using structured surveillance, genomic, and certification data, with accuracy

values generally exceeding 0.70 and R^2 values above 0.65. Deep learning architectures, while currently underrepresented in the food safety risk prediction literature, show promising results for image-based pathogen detection tasks. These findings confirm that AI-augmented systems represent a meaningful advance over classical predictive microbiology approaches, which are limited in their capacity to integrate the heterogeneous, high-dimensional data streams that characterize modern food safety surveillance [10].

The contribution of this review lies not only in synthesizing the performance characteristics of existing models but also in identifying critical gaps in methodology and validation that must be addressed before these tools can be responsibly deployed in regulatory settings. The finding that 50% of included studies carry high overall risk of bias, primarily due to the absence of external validation, underscores the need for a fundamental shift in research priorities from model development toward model evaluation and deployment science. This review also highlights the emerging importance of explainable AI (XAI) and data governance as prerequisites for the trustworthy adoption of AI in public health regulatory contexts.

4.2 Strengths and Weaknesses of the Evidence Base

The current evidence base possesses several notable strengths. First, the consistent finding of strong predictive performance across tree-based ensemble models applied to diverse pathogens and geographic settings provides convergent evidence that these architectures are well-suited to food safety risk prediction tasks. The performance levels observed (accuracy 0.69–0.85, R^2 0.65–0.68) are clinically meaningful for the intended application of risk-based inspection prioritization, where even modest improvements in targeting accuracy can yield significant operational gains when applied across thousands of inspections annually. Second, the diversity of data sources employed from national surveillance databases and WGS repositories to food safety certification records and agar plate images demonstrates the breadth of data modalities that can be leveraged by AI/ML approaches, supporting the premise that multimodal data integration is a key advantage of AI over traditional methods.

However, the evidence base also exhibits significant weaknesses that must be acknowledged. The most critical limitation is the pervasive lack of rigorous external validation: only one study (Arning et al. [16]) employed a train/test split that could be considered a partial form of external validation, while the remaining studies relied exclusively on internal validation through cross-validation. This raises serious concerns about the generalizability of reported performance metrics, as models validated only internally may be overfitted to the specific characteristics of their training datasets and may not maintain the same level of predictive accuracy when applied to new populations, time periods, or geographic regions. Second, the small number of eligible studies ($n = 6$) limits the strength and generalizability of the conclusions that can be drawn, and the heterogeneity in study designs, outcomes, and metrics precludes quantitative synthesis. Third, none of the included studies conducted prospective validation or pilot deployment in real-world regulatory settings, leaving a substantial gap between demonstrated algorithmic performance and proven operational utility.

4.3 Critical Implementation Barriers and Data Governance Challenges

Beyond technical readiness, several systemic barriers impede the translation of AI/ML models from research prototypes to operational tools within food safety regulatory agencies. Data governance represents perhaps the most fundamental challenge, encompassing issues of data quality, interoperability, access, and standardization

[21]. AI/ML models require large volumes of structured, high-quality data for training and validation, yet food safety surveillance data are often fragmented across multiple agencies and jurisdictions, collected using inconsistent protocols, and stored in incompatible formats [22]. The absence of standardized data sharing policies and infrastructure further compounds these challenges, creating a situation in which the raw material needed for model development and evaluation is technically available but practically inaccessible.

Regulatory authorities in many countries also face capacity constraints in data science expertise, computational infrastructure, and institutional culture that limit their ability to evaluate, implement, and maintain AI-driven tools [23]. These capacity gaps are particularly acute in LMICs, where the burden of foodborne disease is greatest and the potential benefit of AI-assisted risk prioritization is arguably highest. The evidence suggests that the highest return on investment may lie not in developing more sophisticated algorithms but in building the data infrastructure and institutional capacity required to make existing algorithms operational [22]. This includes investing in longitudinal inspection databases with standardized variable definitions, unique facility identifiers, geographic coordinates, and violation histories the foundational data assets without which even the most advanced AI models cannot function effectively.

4.4 The Mandate for Explainable AI (XAI) in Public Health Safety

In the context of public health safety, predictive accuracy alone is insufficient to justify the deployment of AI-driven decision support tools. The “black box” nature of many AI/ML models particularly deep learning architectures and complex ensemble methods poses fundamental challenges for transparency, accountability, and trust in regulatory applications [24]. When an AI model assigns a risk score to a food processing facility or recommends an inspection priority, regulators and facility operators alike have a legitimate need to understand the reasoning behind that score. Without such transparency, it becomes impossible to verify that model outputs are based on relevant and appropriate factors, to identify and correct systematic biases, or to provide affected parties with meaningful explanations for regulatory actions.

Explainable AI (XAI) methods including SHAP (SHapley Additive exPlanations) values, LIME (Local Interpretable Model-agnostic Explanations), and attention mechanisms in neural networks offer tools for illuminating the decision-making processes of complex models, providing feature-level attributions that can be communicated to stakeholders in understandable terms [25]. Several included studies incorporated interpretability analyses; for example, Arning et al. [16] identified non-compliance records and specific genomic markers as primary drivers of *Campylobacter* source attribution, while Garcia-Vozmediano et al. [19] reported that surveillance indicators such as prior outbreak history and seasonal patterns were key predictors of *Salmonella* risk. However, the adoption of XAI was not systematic across the evidence base, and no study formally evaluated the impact of XAI on stakeholder trust or decision-making quality. For AI systems to gain regulatory endorsement and public trust, the integration of XAI should be considered mandatory rather than optional, with standardized frameworks for model interpretation and reporting [25].

4.5 Mitigating Algorithmic Bias and Ensuring Equity

AI/ML models trained on historically collected data may inadvertently perpetuate or amplify existing biases, leading to inequitable allocation of inspection resources and disproportionate regulatory scrutiny of certain communities or food businesses [26]. For instance, if historical inspection data reflect biased enforcement

patterns such as more frequent inspections of establishments in certain neighborhoods or those operated by particular demographic groups models trained on these data may learn to associate risk with these demographic features rather than with genuine safety hazards. Research by Altenburger and Ho [26] has demonstrated that predictive analytics used for food safety enforcement can disproportionately harm certain groups, highlighting the potential for algorithmic bias in regulatory applications.

Mitigating algorithmic bias requires proactive measures at multiple stages of the model development pipeline, including careful selection and preprocessing of training data to minimize the influence of confounding demographic variables, regular bias audits using fairness metrics, and the incorporation of human-in-the-loop (HITL) oversight to ensure that AI-generated recommendations are reviewed and contextualized before implementation [27]. Under the European Union's Artificial Intelligence Act (Regulation EU 2024/1689), AI systems used in high-risk contexts including those that may influence access to essential services are subject to mandatory requirements for risk management, data governance, transparency, and human oversight [28]. Compliance with such regulatory frameworks will be essential for the legitimate deployment of AI in food safety inspection systems.

4.6 Limitations of This Review

Several limitations of this systematic review should be acknowledged. First, the restricted timeframe (January 2020 to December 2025) may have excluded relevant earlier studies that established foundational methods, although this period was chosen to capture the most recent advances in a rapidly evolving field. Second, the exclusion of non-English publications may have introduced language bias, potentially omitting relevant studies from non-English-speaking regions where food safety challenges may be most acute. Third, the reliance on peer-reviewed publications may have introduced publication bias, as studies with positive or significant results are more likely to be published, potentially inflating the apparent efficacy of AI/ML approaches. Fourth, the heterogeneity of study designs, model architectures, data sources, and outcome metrics precluded quantitative meta-analysis, limiting the precision of the synthesized evidence. Fifth, the rapid pace of AI development means that this review represents a time-limited snapshot of the evidence base, and additional eligible studies may have been published after the search date. Finally, the absence of prospective protocol registration, while acknowledged and addressed through predefined methods, represents a methodological limitation that reduces the transparency of the review process.

4.7 Future Research Directions

Based on the findings and limitations identified in this review, several priority areas for future research emerge. First, external validation of existing AI/ML models using independent temporal and geographic datasets should be considered the single most important research priority, as the current evidence base is insufficient to assess the generalizability of reported models. Second, the development of standardized benchmarking datasets similar to those available in other domains such as medical imaging would facilitate objective comparison of different modeling approaches and accelerate progress in the field. Third, formal evaluation of XAI methods in the food safety regulatory context is needed, including studies that assess the impact of model explanations on inspector decision-making, stakeholder trust, and regulatory compliance.

Fourth, the creation of open, FAIR (Findable, Accessible, Interoperable, Reusable) datasets for food safety risk prediction, subject to appropriate privacy protections, would address the data access barriers that currently

limit model development and independent validation. Fifth, prospective pilot deployments of AI-assisted inspection prioritization in real-world regulatory settings, coupled with rigorous health economic evaluations, are essential to bridge the gap between algorithmic promise and operational impact. Sixth, research into the impact of climate change on foodborne pathogen dynamics and the development of AI models that can adapt to shifting risk landscapes will become increasingly important as climate-driven food safety challenges intensify. Finally, the establishment of multi-stakeholder governance frameworks that encompass data sharing, model validation, bias auditing, and accountability mechanisms will be critical for ensuring that AI deployment in food safety serves the public interest equitably and transparently.

Box 1: Practical Implications for Regulatory Agencies

Regulatory bodies considering the adoption of AI-assisted inspection prioritization should consider the following evidence-based recommendations:

Data Infrastructure Investment

Prioritize the development of clean, longitudinal inspection databases with standardized variable definitions, unique facility identifiers, violation histories, and geographic coordinates. The quality and completeness of input data will ultimately determine the reliability of AI-generated risk scores.

External Validation as a Prerequisite

Require that any AI model proposed for operational use demonstrate performance on independent external datasets (temporal and/or geographic validation) before deployment. Internal cross-validation alone is insufficient to establish model generalizability.

Explainability Standards

Mandate the use of XAI frameworks (e.g., SHAP, LIME) for all AI models used in regulatory decision-making, ensuring that risk scores can be explained in terms that are meaningful to inspectors, facility operators, and the public.

Human-in-the-Loop (HITL) Design

Design AI-assisted systems to support, not replace, human decision-making. AI should provide recommendations that are reviewed and contextualized by trained inspectors before action is taken.

Governance and Auditing Frameworks

Establish ongoing governance mechanisms for monitoring model performance drift, conducting regular bias audits, and ensuring compliance with emerging AI regulations such as the EU AI Act.

Phased Pilot Implementation

Begin with small-scale pilot deployments to measure key performance indicators (KPIs) including detection rates, false positive rates, and inspection efficiency gains before scaling to broader implementation.

5. Conclusion

Artificial intelligence represents a transformative evolution in food safety management, enabling a transition from reactive containment to proactive, intelligence-led prediction and prioritization of microbiological hazards. This systematic review demonstrates that AI/ML models particularly tree-based ensemble methods can achieve high predictive accuracy and quantifiable operational benefits for risk stratification and inspection resource optimization. Deep learning approaches further extend the applicability of AI to image-based and unstructured data sources, broadening the scope of automated food safety surveillance.

However, the translation of technological potential into widespread public health impact is far from guaranteed. The current evidence base is constrained by a small number of eligible studies, substantial heterogeneity in modeling approaches and reporting standards, and a pervasive lack of rigorous external validation that raises legitimate concerns about the deployment of potentially overfitted models in real-world regulatory settings. For AI systems to fulfill their promise in food safety, the research community and governing bodies must pivot focus from incremental accuracy improvements toward the foundational requirements for trustworthy deployment: establishing robust external validation standards, mandating explainable AI for transparency and accountability, investing in data governance infrastructure that enables scalable and interoperable surveillance, and developing governance frameworks that address algorithmic bias and ensure equitable implementation. Only through concerted attention to these prerequisites can the field move from promising proof-of-concept demonstrations to reliable, equitable, and regulatorily endorsed AI-driven food safety systems.

Abbreviations

AI

Artificial Intelligence

AUC-ROC

Area Under the Receiver Operating Characteristic Curve

CNN

Convolutional Neural Network

CV

Cross-Validation

DALY

Disability-Adjusted Life Year

DL

Deep Learning

EFSA

European Food Safety Authority

FAIR

Findable, Accessible, Interoperable, Reusable

FAO

Food and Agriculture Organization

FDA

U.S. Food and Drug Administration

GBDT

Gradient Boosted Decision Trees

HACCP

Hazard Analysis Critical Control Point

HITL

Human-in-the-Loop

LMIC

Low- and Middle-Income Country

ML

Machine Learning

PECO

Population, Exposure, Comparator, Outcomes

PM

Predictive Microbiology

PRISMA

Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROBAST

Prediction Model Risk of Bias Assessment Tool

QMRA

Quantitative Microbial Risk Assessment

R²

Coefficient of Determination

SHAP

SHapley Additive exPlanations

WGS

Whole-Genome Sequencing

WHO

World Health Organization

XAI

Explainable Artificial Intelligence

XGBoost

Extreme Gradient Boosting

Declarations

Funding: No specific funding was received for this academic activity.

Conflicts of Interest: The authors declare no competing interests.

Data Availability: All data extraction forms, PROBAST assessments, and search strategies are available upon request from the corresponding author.

Author Contributions: All authors contributed to study design, data extraction, quality assessment, and manuscript preparation. All authors read and approved of the final manuscript.

References

1. World Health Organization. WHO estimates of the global burden of foodborne diseases: foodborne disease burden epidemiology reference group 2007-2015. Geneva: WHO; 2015. Available from: <https://www.who.int/publications/i/item/9789241565165>
2. Grace D. Food safety in low and middle income countries. *Int J Environ Res Public Health*. 2015;12(9):10490-507. doi:10.3390/ijerph120910490
3. Unnevehr LJ. Addressing food safety challenges in rapidly developing food systems. *Agricultural Economics*. 2022;53(S1):529-39. doi:10.1111/agec.12724
4. Awad DA, Masoud HA, Hamad A. Climate changes and food-borne pathogens: the impact on human health and mitigation strategy. *Climatic Change*. 2024;177(6):92. doi:10.1007/s10584-024-03748-9
5. Newell DG, Koopmans M, Verhoef L, et al. Food-borne diseases – the challenges of 20 years ago still persist while new ones continue to emerge. *Int J Food Microbiol*. 2010;139 Suppl 1:S3-15. doi:10.1016/j.ijfoodmicro.2010.01.033
6. Vergis J, Pollumahanti N, Sahu R, Rawool DB, Barbuddhe SB. Ensuring food safety: microbiological risk assessment strategies. *Curr Opin Food Sci*. 2025;62:101272. doi:10.1016/j.cofs.2025.101272
7. Bouzembrak Y, Marvin HJP. Impact of drivers of change, including climatic factors, on the occurrence of chemical food safety hazards in fruits and vegetables: a Bayesian Network approach. *Food Control*. 2019;97:67-76. doi:10.1016/j.foodcont.2018.10.007
8. Stavropoulou E, Bezirtzoglou E. Predictive modeling of microbial behavior in food. *Foods*. 2019;8(12):654. doi:10.3390/foods8120654
9. Jiang Y, Luo J, Huang D, Liu Y, Li D. Machine learning advances in microbiology: a review of methods and applications. *Front Microbiol*. 2022;13:925454. doi:10.3389/fmicb.2022.925454
10. Wang X, Bouzembrak Y, Lansink AO, van der Fels-Klerx HJ. Application of machine learning to the monitoring and prediction of food safety: a review. *Compr Rev Food Sci Food Saf*. 2022;21(2):1912-38. doi:10.1111/1541-4337.12868
11. Feng Y, Liu J, Zhu H. Artificial intelligence in food safety: a decade review and bibliometric analysis. *Foods*. 2023;12(6):1242. doi:10.3390/foods12061242
12. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71
13. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-8. doi:10.7326/M18-1376
14. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. 2021;11(7):e048008. doi:10.1136/bmjopen-2020-048008
15. Wang H, Cui W, Guo Y, Du Y, Zhou Y. Machine learning prediction of foodborne disease pathogens: algorithm development and validation study. *JMIR Med Inform*. 2021;9(1):e24924. doi:10.2196/24924
16. Arning A, Veh L, Schotte U, et al. Machine learning based source attribution of *Campylobacter jejuni* isolates from livestock, wildlife, and humans using whole-genome sequencing data. *Microorganisms*. 2021;9(8):1672. doi:10.3390/microorganisms9081672

17. Karanth S, Patel J, Shirmohammadi A, Pradhan AK. Machine learning to predict foodborne salmonellosis outbreaks based on genome characteristics and meteorological trends. *Curr Res Food Sci.* 2023;6:100525. doi:10.1016/j.crfs.2023.100525
18. Zheng Y, Gracia A, Hu L. Predicting foodborne disease outbreaks with food safety certifications: econometric and machine learning analyses. *J Food Prot.* 2023;86(9):100136. doi:10.1016/j.jfp.2023.100136
19. Garcia-Vozmediano A, Maurella C, Ceballos LA, et al. Machine learning approach as an early warning system to prevent foodborne Salmonella outbreaks in northwestern Italy. *Vet Res.* 2024;55(1):75. doi:10.1186/s13567-024-01318-6
20. Ravindhiran R, Dhandapani K. Detection of foodborne *Listeria monocytogenes* using deep learning models to ensure food safety and health. *J Food Sci Technol.* 2025;62:1-12.
21. van Meer F, van der Velden B, Takeuchi M. Artificial intelligence for food safety: a literature synthesis, real-world applications and regulatory frameworks. Rome: FAO; 2025. Available from: <https://openknowledge.fao.org/handle/20.500.14283/cd7242en>
22. World Health Organization. WHO global strategy for food safety 2022-2030: towards stronger food safety systems and global cooperation. Geneva: WHO; 2022. ISBN: 9789240057685
23. Marvin HJP, Bouzembrak Y, van der Fels-Klerx HJ, et al. Making food systems more resilient to food safety risks by including artificial intelligence, big data, and internet of things into food safety early warning and emerging risk identification tools. *Compr Rev Food Sci Food Saf.* 2024;23(1):e1600798. doi:10.1111/1541-4337.13296
24. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Comput Surv.* 2021;54(6):1-35. doi:10.1145/3457607
25. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access.* 2018;6:52138-60. doi:10.1109/ACCESS.2018.2870052
26. Altenburger KM, Ho DE. When algorithms import private bias into public enforcement: the promise and limitations of statistical debiasing solutions. *J Inst Theor Econ.* 2020;176(4):628-55. doi:10.1628/093245619X15726839194037
27. Arrieta AB, Díaz-Rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion.* 2020;58:82-115. doi:10.1016/j.inffus.2019.12.012
28. European Parliament and Council. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Off J Eur Union. 2024. Available from: <http://data.europa.eu/eli/reg/2024/1689/oj>

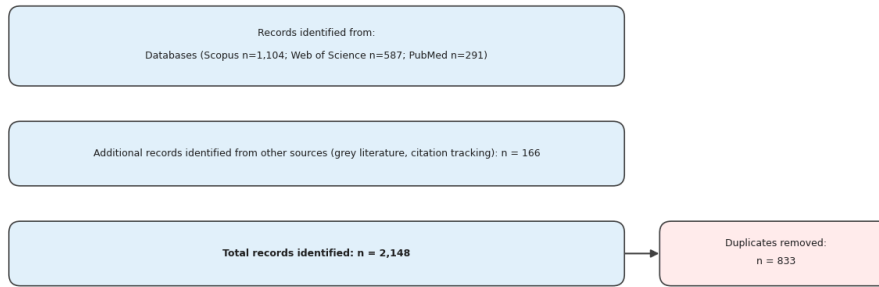
Figures

Figure 1. PRISMA 2020 Flow Diagram for Systematic Review

AI for Predicting Microbiological Hazards in Food Safety (2020-2025)

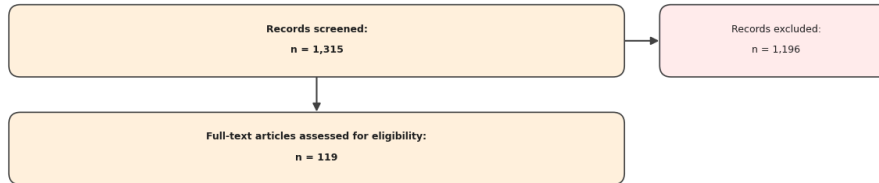
IDENTIFICATION

Database
searching



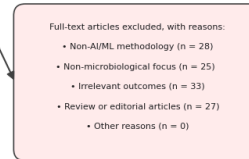
SCREENING

Screening



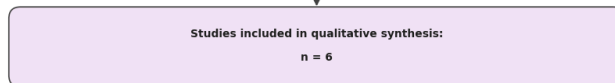
ELIGIBILITY

Eligibility



INCLUDED

Included



PRISMA 2020 Statement: Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. Search period: January 1, 2020 - December 1, 2025 | Databases: Scopus, Web of Science Core Collection, PubMed

Figure 1

See image above for figure legend