INTERNATIONAL
JOURNAL OF BUSINESS AND
ECONOMICS
INSIGHTS

Volume: *5;* ***Issue:*** *3*
Pages: *205–237*
Accepted: *29 August 2025*
Published: *15 September 2025*

ARTIFICIAL INTELLIGENCE BASED MODELS FOR PREDICTING FOODBORNE PATHOGEN RISK IN PUBLIC HEALTH SYSTEMS

Pankaz Roy Sarkar[1]

[1]. Master of Science in Public Health, Birmingham City University, UK;
Email: saikatpariarc86@gmail.com

Doi: 10.63125/7685ne21

This work was peer-reviewed under the editorial responsibility of the IJEI, 2025

Abstract

This systematic literature review synthesizes evidence on artificial intelligence models used to predict foodborne pathogen risk within public health systems, focusing on how data, methods, and validation practices translate into actionable prevention. We searched major multidisciplinary and domain databases through September 2025 and screened studies against predefined eligibility criteria aligned with PRISMA. A total of 105 peer-reviewed studies met inclusion, spanning outbreak detection, nowcasting and multi-horizon forecasting, spatiotemporal risk mapping, inspection prioritization, and whole-genome sequencing–enabled source attribution. Across the corpus, tree-based ensembles consistently excelled for tabular, establishment-level risk scoring, while recurrent and attention-based sequence models were strongest for delay-aware forecasting. Multi-stream fusion of inspections, laboratory and genomic data, syndromic telemetry, environmental drivers, complaint signals, and supply-chain metadata yielded measurable gains in discrimination, stability, and top-k precision compared with single-stream models. Studies that implemented temporal or geographic external validation, probability calibration, and decision-utility analyses reported smaller but durable improvements that translated into operational benefits such as more critical violations found per fixed inspection budget and earlier detection of emergent clusters at controlled false-alarm rates. Methodological themes associated with credible deployment included leakage-safe temporal and spatial validation, transparent feature engineering, calibrated probabilistic outputs with proper scoring, interpretability to support regulatory scrutiny, drift monitoring with scheduled recalibration, and subgroup assessments to manage equity trade-offs. The review consolidates a decision-oriented framework that maps prediction targets to data and model families, highlights reproducible performance patterns, and outlines governance steps that convert statistical accuracy into sustained public health impact.

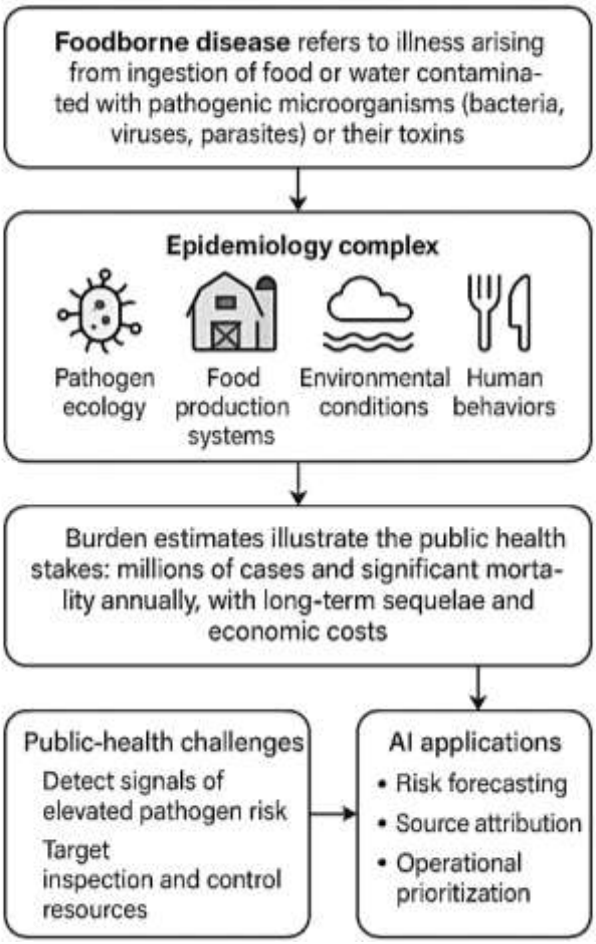
Keywords

Artificial Intelligence, Machine Learning, Foodborne Pathogens, Public Health Surveillance, Risk Prediction, Inspection Prioritization

INTRODUCTION

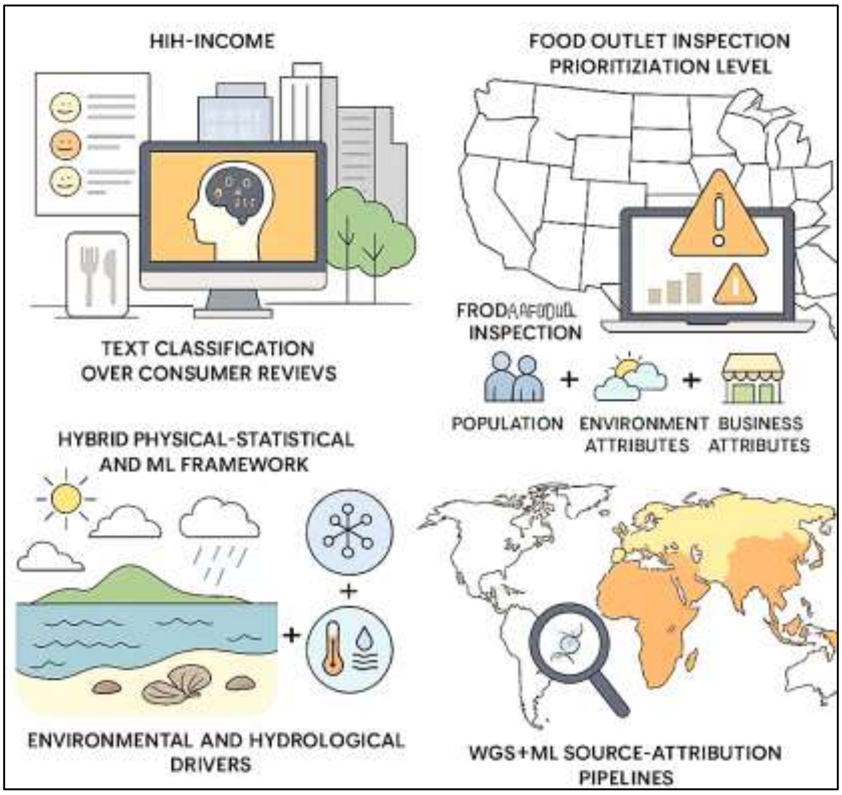
Foodborne disease refers to illness arising from ingestion of food or water contaminated with pathogenic microorganisms (bacteria, viruses, parasites) or their toxins, and it remains a leading cause of preventable morbidity and mortality worldwide. Its epidemiology is complex, reflecting interactions among pathogen ecology, food production systems, environmental conditions, and human behaviors from farm practices and cold-chain integrity to consumer food handling. Burden estimates illustrate the public health stakes: analyses anchored in robust surveillance and modeling show millions of cases and significant mortality annually, with nontrivial long-term sequelae and economic costs borne by individuals, health systems, and the food industry (Chenar & Deng, 2021; Effland, Lawson, et al., 2018). In addition to well-recognized agents such as Salmonella enterica, Campylobacter, Listeria monocytogenes, norovirus, and pathogenic Escherichia coli, risk is shaped by climate-sensitive pathogens (e.g., Vibrio spp.) and by the globalized, just-in-time nature of food supply chains that can amplify contamination events and complicate traceback. Public-health agencies thus face a dual challenge: detect signals of elevated pathogen risk early enough to prevent illness, and target scarce inspection and control resources where they will yield the greatest risk reduction. These priorities motivate methods that can learn from high-dimensional, heterogeneous data streams healthcare, laboratory, environmental, supply-chain, consumer-generated, and genomic to forecast pathogen risk at actionable spatial and temporal scales (Chenar & Deng, 2021; Munck et al., 2020; Panacek & et al., 2023; Wang et al., 2021).

Figure 1: Framework of Foodborne Disease Risks and AI Applications



Artificial intelligence (AI) particularly machine learning (ML) methods such as gradient-boosted trees, random forests, regularized generalized linear models, kernel methods, and modern probabilistic or hierarchical learners has been increasingly applied to food safety surveillance because these models handle nonlinearity, interactions, and mixed-type features while scaling to large datasets. In public health contexts, AI has supported three complementary aims: (1) risk forecasting (predicting when/where risk of pathogen presence or illness is likely to rise); (2) source attribution (inferring the food/animal or geographic sources of human infections); and (3) operational prioritization (ranking establishments or supply-chain links for inspection or intervention). Foundational studies show that ML can improve outbreak signal detection from case data and metadata (Effland, DeFelice, et al., 2018; Harrison & et al., 2014), forecast environment-linked hazards such as Vibrio infections (Collins et al., 2024; Jackson et al., 2016), and triage inspections by predicting noncompliance from geospatial and sociodemographic features (Collins et al., 2024; Oldroyd et al., 2021; Rose & et al., 2025; Zhang et al., 2021). At the same time, whole-genome sequencing (WGS) integrated with ML has transformed source attribution linking clinical isolates to food sources or production environments thus directly informing targeted controls and recalls (Ekanayake & et al., 2022; Gmeiner et al., 2024; Li & et al., 2017). Taken together, these advances underscore the feasibility and utility of AI-based models for proactive, risk-based disease prevention within public health systems.

Figure 2 : Applications of Artificial Intelligence in Foodborne Pathogen Surveillance



International significance is underscored by the diversity of settings and data regimes in which AI has added value. In high-income countries with mature surveillance, ML augments routine workflows: city health departments have used text classification over consumer reviews to discover otherwise unreported illness clusters and guide inspections, improving yield over complaint-based methods (Gmeiner et al., 2025; Kirk et al., 2015). At national and regional levels, risk-prediction models that incorporate population, environment, and business attributes help prioritize food outlet inspections where resources are constrained (Gmeiner et al., 2025; Kirk et al., 2015). In pathogen- and context-specific domains, hybrid physical-statistical and ML frameworks have predicted norovirus risk in shellfish harvesting areas by fusing environmental and hydrological drivers (Gmeiner et al., 2025; Turner & et al., 2024; Weller et al., 2021). Climate-responsive predictive models are increasingly salient: species within Vibrio show strong temperature and salinity dependencies, and recent studies have used ML to capture multi-factor drivers and forecast infection hazards in coastal regions, illustrating how environmental surveillance can feed forward into public-health advisories (Gmeiner et al., 2024; Havelaar & et al., 2015). Globally, WGS+ML source-attribution pipelines now support rapid, geographically aware inference for Salmonella and Listeria, strengthening the ability to pinpoint sources and deploy risk management measures across borders (Lupolova et al., 2019).

A central reason AI methods fit this domain is the abundance of heterogeneous predictors and their complex, nonlinear relationships to pathogen risk. Public health systems can leverage syndromic and case surveillance data, laboratory results, environmental measurements (temperature, precipitation, salinity), satellite/remote sensing features, transportation and trade flows, food business characteristics, and consumer-generated signals. ML models can integrate these signals to deliver risk scores or probabilities aligned to decision horizons e.g., daily to weekly forecasts for environmental pathogens, or monthly risk maps to plan inspection calendars. Studies demonstrate that gradient boosting and random forests improve discriminatory performance for identifying establishments at higher risk of noncompliance compared with simpler baselines (Mikhailov et al., 2023; Scallan et al., 2011). In event detection, ML over case metadata and exposure features has improved classification of outbreak-linked cases and identified confounders important for cluster recognition. For environmental drivers, ensembles and feature-attribution methods help disentangle correlated oceanographic and climatic variables shaping Vibrio hazards, producing interpretable predictors for risk communication (Chenar & Deng, 2021; Lupolova et al., 2017). In the WGS space, random forests and other classifiers trained on SNPs, core-genome MLST, or kmers can attribute human isolates to source reservoirs or geographies, often outperforming classic population-genetic approaches, and doing so in near real-time for routine surveillance (Castelli et al., 2023).

Crucial for public-health legitimacy is that AI-based forecasts translate into actionable, risk-based control. Evidence from municipal deployments shows that algorithm-guided inspection targeting yields more violations detected per visit than status quo approaches an operational efficiency gain that matters when environmental health teams face backlogs and staffing constraints (Chenar & Deng, 2018; Oldroyd et al., 2021). In environmental health protection, forecasting models for Vibrio spp. provide lead time for advisories, harvesting closures, and public messaging when conditions indicate elevated hazard (Campbell et al., 2025; Chenar & Deng, 2018). In national and international surveillance, WGS-ML source attribution and geographic assignment narrow epidemiologic hypotheses, accelerate recalls, and focus traceback investigations, directly preventing further cases (Chenar & Deng, 2018; Rose & et al., 2025). Methodologically, these wins depend on careful data engineering (handling class imbalance, leakage, nonstationarity), appropriate validation (temporal/spatial splits), and interpretability tools that support regulatory transparency. Many published studies explicitly evaluate trade-offs between model complexity and interpretability to meet governance needs in official control programs, an important consideration for acceptance by inspectors, risk managers, and the public (Gmeiner et al., 2024).

The genomics-plus-ML trajectory deserves emphasis because it reframes prediction targets in ways directly relevant to risk management. Instead of predicting only “illness tomorrow,” WGS-ML can predict “which food/animal source is most likely,” “which clonal complex has high virulence potential,” or “which production environment harbors higher contamination risk.” For Listeria monocytogenes, ML applied to WGS has been used to predict virulence potential and even tolerance to disinfectants capabilities that inform facility-level interventions and sanitation standards (Baker-Austin et al., 2023; Wang et al., 2021). For Campylobacter and Salmonella, WGS-ML methods attribute human cases to poultry, cattle, or other reservoirs with improved accuracy over earlier models, enhancing the targeting of controls along the poultry meat chain and beyond (Munck et al., 2020; Scallan et al., 2011). Newer hierarchical ML models combine phylogenetic structure with learning to enable rapid geographic source attribution at continental to sub-national scales, again supporting the practical need to prioritize investigations and recalls (Turner & et al., 2024; Zhang et al., 2021). Because these pipelines ingest standardized surveillance genomes, they align naturally with national laboratory networks and cross-border data sharing, broadening their international utility.

Another strand of AI-enabled risk prediction synthesizes environmental and supply-chain signals to forecast pathogen presence or microbial indicators in food and agricultural waters, which is critical for E. coli, Salmonella, and Vibrio management. Studies have shown that ML models trained on physicochemical and hydrometeorological features can predict E. coli levels in agricultural water with performance competitive to mechanistic baselines while enabling interpretable feature importance for operational decisions (Rose & et al., 2025). In shellfish sanitation, hybrid and ML models incorporating tides, rainfall, and land-use characteristics have predicted norovirus outbreaks, enabling dynamic classification of harvesting areas (Chenar & Deng, 2018). Climate-sensitive applications are accelerating as warming seas and extreme precipitation alter coastal risks: recent work quantifies the poleward expansion and season lengthening of Vibrio hazards and uses ML to capture multifactorial drivers, which is directly relevant to international coastal public-health programs (Baker-Austin et al., 2023; Collins et al., 2024). These models convert environmental surveillance into concrete public-health actions by delivering thresholds and risk tiers that trigger advisories, temporary closures, or intensified sampling.

Finally, AI's role in modernizing public-health surveillance rests on integrative architectures and governance. Successful deployments couple modeling with data pipelines (secure access to surveillance, lab, environmental, and open-source streams), MLOps for monitoring drift and recalibration, and human-in-the-loop review by epidemiologists and inspectors. Studies repeatedly stress the importance of validation aligned to real decision contexts (e.g., prospective temporal validation rather than random cross-validation), attention to class imbalance and rare-event detection, and transparency that allows risk managers to justify actions (Mikhailov et al., 2023; Oldroyd et al., 2021). Internationally, cross-jurisdictional learning for instance, adapting source-attribution models trained in one country with transfer learning or hierarchical structures has shown promise for rapid uptake where local training data are limited.

The objective of this literature review is to systematically map, evaluate, and synthesize the peer-reviewed evidence on artificial intelligence–based models used to predict foodborne pathogen risk within public health systems. Specifically, this review aims to achieve five integrated goals that together create a comprehensive, decision-oriented picture of the field. First, it will delineate the problem space by defining the prediction targets used in public health practice such as outbreak detection, case forecasting, spatiotemporal risk mapping, inspection prioritization, and source attribution and by clarifying how these targets align with operational time horizons and geographic scales. Second, it will inventory and classify the data streams that underpin AI models for this domain, including surveillance and laboratory reporting, whole-genome sequencing, environmental and climate measurements, supply-chain and business attributes, and consumer-generated signals, alongside the principal feature engineering strategies applied to each. Third, it will develop a transparent taxonomy of modeling approaches, spanning baselines and tree-based learners, kernel and distance methods, anomaly detection, deep learning architectures, probabilistic and Bayesian frameworks, graph-based methods for traceback and attribution, and hybrid or ensemble designs, documenting training schemes, hyperparameter practices, and approaches to class imbalance, leakage prevention, and uncertainty quantification. Fourth, it will critically appraise validation and reporting practices using established criteria, capturing internal, temporal, and external validation designs; discrimination, calibration, and alerting metrics; subgroup performance; interpretability techniques; and evidence of deployment or operational performance where reported. Fifth, it will synthesize cross-study findings into structured evidence tables and visual gap maps that highlight consistencies, divergences, and knowledge gaps across pathogens, tasks, data modalities, geographies, and organizational contexts, while explicitly noting reproducibility assets such as code, models, and data availability. The scope of this review is limited to studies that present or evaluate AI or machine learning models for predicting foodborne pathogen risk in human public health settings, excluding purely analytical chemistry, bench-only detection, or non-predictive descriptive work unless directly coupled to risk prediction. The intended contribution is a rigorous, organized evidence base and a practical conceptual framework that allow researchers, analysts, and public health professionals to understand what has been studied, how it has been evaluated, and where concentrated methodological or dataset development could most efficiently advance predictive performance and operational utility.

LITERATURE REVIEW

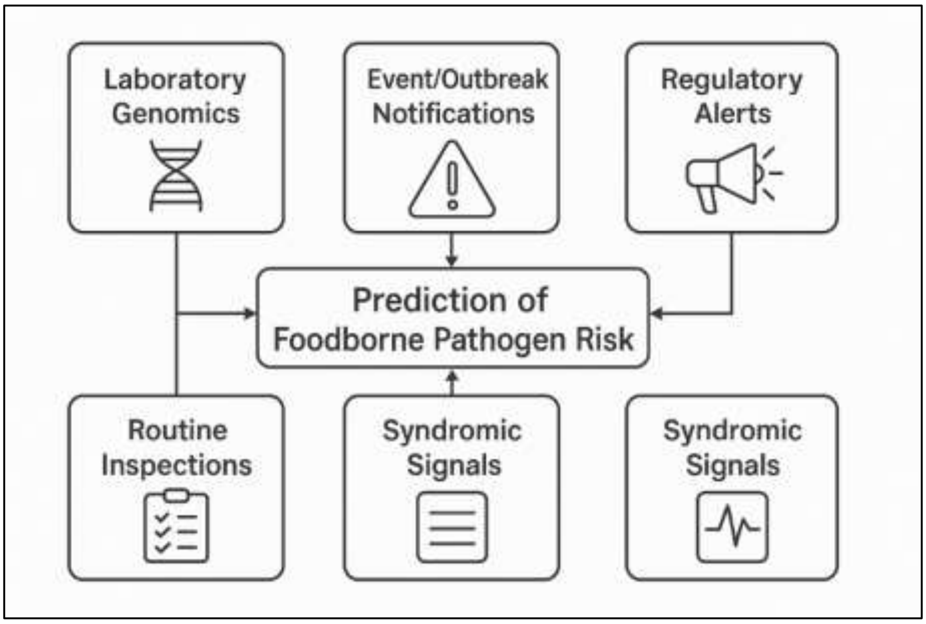
The literature on artificial intelligence (AI) for predicting foodborne pathogen risk in public health systems spans a diverse yet converging set of aims, data streams, and modeling strategies that together define a rapidly maturing field. At its core, this body of work addresses prediction targets that are directly actionable for health authorities: early detection of outbreak signals, short- and medium-horizon forecasting of cases or hazards, spatiotemporal risk mapping to guide inspections and sampling, and source attribution to prioritize recalls and traceback investigations. Studies draw on heterogeneous inputs that reflect the full surveillance ecosystem, including routine case notifications, laboratory results, whole-genome sequencing, environmental and climate measurements, hydrological and oceanographic indicators, geodemographics, business attributes, supply-chain and mobility patterns, and consumer-generated signals such as complaints or online reviews. Transforming these inputs into predictive features relies on careful temporal alignment, leakage avoidance, and representation of spatial dependence, seasonality, and domain-specific constructs such as phylogenetic relatedness or food-chain topology. Methodologically, the literature covers regularized regression and tree-based ensembles as strong baselines; kernel and distance-based classifiers; anomaly detection for rare-event signals; deep learning architectures tailored to sequences, time-series, and graphs; probabilistic and Bayesian approaches for uncertainty-aware inference; and hybrid frameworks that fuse mechanistic knowledge with machine learning. Across problem settings, credible evaluation hinges on prospective or temporally separated validation, geographic generalization tests, and metrics that reflect operational decision quality, including discrimination, calibration, alert yield, lead time, and false-alarm burden, as well as subgroup performance to surface equity considerations. Interpretability practices feature attribution, partial dependence, counterfactual probes, and transparent model cards are increasingly emphasized to support regulatory scrutiny and field adoption. A prominent strand integrates genomics with learning to classify likely sources, reservoirs, or virulence properties, while another synthesizes environmental and hydrological drivers for coastal and agricultural water risks. Finally, the deployment-oriented literature underscores pipeline readiness: data provenance, reproducibility assets, monitoring for drift, and human-in-the-loop review that aligns model outputs with inspection and epidemiologic workflows. Taken together, these themes motivate a structured synthesis that maps prediction targets to data, models, validation designs, and operational endpoints, establishing a coherent foundation for the detailed subsections that follow.

Data Sources & Surveillance Streams

Modern AI models for predicting foodborne pathogen risk draw strength from a mosaic of complementary data sources that collectively span laboratory genomics, event/outbreak notifications, regulatory alerts, routine inspections, and syndromic signals. At the laboratory core are whole-genome sequencing (WGS) repositories that enable high-resolution linkage among clinical, food, and environmental isolates. In the United States, PulseNet's transition from pulsed-field gel electrophoresis to WGS has reshaped cluster detection by increasing discriminatory power and tightening inference windows for cross-jurisdictional outbreaks (Ribot et al., 2019; Tolar et al., 2019). In parallel, the FDA's GenomeTrakr network funnels standardized WGS data into the NCBI Pathogen Detection ecosystem, creating a continuously updated substrate for AI tasks such as outbreak clustering, lineage tracking, and source attribution (Allard et al., 2016; Danish & Zafor, 2022). Crucially, the value of these genomic streams is not merely technical: economic evaluation suggests that scaling public WGS isolate deposition yields substantial public-health benefits via faster recalls and smaller outbreaks, with benefits outweighing program costs early in implementation (Allard et al., 2016; Danish & Md.Kamrul, 2022). For AI practitioners, these infrastructures provide labeled graphs of genetic relatedness and temporally stamped clusters that can be embedded or linked to metadata (food commodity, facility, geography) for supervised and semi-supervised learning. Yet genomic insight is only one layer; outbreak reporting systems add context. In Europe, the Epidemic Intelligence and Information System for food- and waterborne diseases (EPIS-FWD) facilitates rapid, structured cross-border information exchange that AI models can mine for early signals and features tied to multi-country spread dynamics (Jahid, 2022; Nogales et al., 2023). Together, these laboratory and outbreak streams anchor high-fidelity targets (e.g., "cluster joins a known lineage," "notification escalates to multi-state scope") that downstream models can learn to anticipate from upstream, noisier proxies.

Because clinical reporting can be delayed or incomplete, syndromic and near–real-time signals play a complementary role in closing latency gaps for enteric pathogens. The national BioSense/NSP emergency-department infrastructure has shown that chief-complaint subsyndromes (e.g., diarrhea; nausea/vomiting) track norovirus activity and correlate with confirmed outbreaks, supporting their use for timely situational awareness and short-horizon forecasting (Baker et al., 2024; Nogales et al., 2023). Systematic reviews of ED-based syndromic surveillance confirm its broad utility and catalog design considerations signal specificity, coding practices, and calibration against laboratory outcomes that matter for reliable deployment (Brown et al., 2021; Arifur & Noor, 2022). More recently, a GI-specific systematic review synthesizes evidence across countries and data types (ED visits, telehealth, medication sales, school absenteeism), underscoring that properly specified syndromic systems can furnish early detection and routine monitoring value when combined with conventional surveillance (Brown et al., 2021; Hasan & Uddin, 2022). For AI pipelines, these feeds provide high-frequency covariates and candidate proxy targets that improve nowcasting of illness activity, especially when models explicitly account for reporting delays and partial observability. Beyond healthcare-based signals, complaint systems and public-facing reporting portals contribute human-in-the-loop observations of suspected foodborne illness; recent comparative work ties complaint-system design choices (e.g., triage rules, disclosure practices) to inspection outcomes, offering structured features (complaint volumes, resolution times, channel mix) that can refine inspection-prioritization models (Smith, et al., 2021). These "fast" streams are noisier than confirmatory lab data, but when engineered with temporal alignment and debiasing, they help AI models deliver earlier risk signals at operational cadence (Rahaman, 2022).

Figure 3: Data Sources and Surveillance Streams



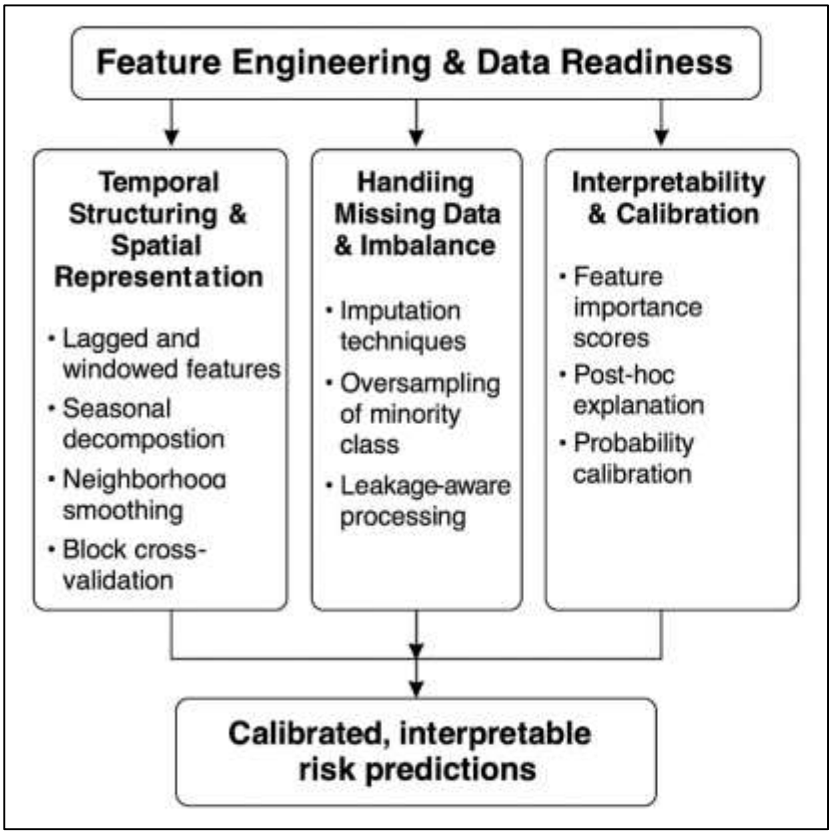
Regulatory and inspection data further broaden the feature landscape by describing hazards moving through supply chains and hygiene practices within establishments. In the European Union, the Rapid Alert System for Food and Feed (RASFF) emits structured notifications (hazard category, product, origin, action taken) that capture risks before illnesses are detected; comprehensive analyses of RASFF data by product domain and by country demonstrate stable metadata fields and interpretable patterns (e.g., product × origin interactions) that are well-suited to supervised learning and network analytics (Čapla et al., 2023; Rha et al., 2013). On the retail end, routine inspection and disclosure systems yield establishment-level attributes inspection scores, infraction types, compliance history that have been validated as reliable inputs for surveillance and evaluation; field audits of the UK Food Hygiene Rating Scheme (FHRS) support both its spatial accuracy and its use in risk-based targeting (Adedire et al., 2024; Rahaman, 2022b). From an AI perspective, these regulatory streams are valuable because they connect hazards to traceable entities (firms, commodities, lots) and actionable levers (border rejections, withdrawals, closures), enabling models to rank-order inspection queues and anticipate noncompliance hotspots (Rahaman & Ashraf, 2022). When fused with genomics and syndromic data, they provide a full "sensorium" from upstream hazards to downstream illness proxies, allowing multi-view learning that is both early and precise. Collectively, laboratory WGS networks, outbreak and epidemic-intelligence platforms, syndromic telemetry, complaint systems, and regulatory/inspection datasets constitute a layered surveillance architecture; AI models that respect each stream's cadence, reliability, and bias profile can translate this architecture into robust, decision-ready risk predictions across public-health systems (Adedire et al., 2024; Čapla et al., 2023; Nogales et al., 2023).

Feature Engineering & Data Readiness

Effective feature engineering for foodborne pathogen–risk prediction begins with disciplined temporal structuring and spatial representation of surveillance and environmental data. Time-aware pipelines typically harmonize multi-cadence inputs (e.g., daily environmental sensors, weekly syndromic counts, irregular inspections) by constructing lagged and windowed aggregates that reflect biological latency and operational horizons, then separating trend from seasonality to avoid spurious correlations (Islam, 2022). A widely adopted approach is STL decomposition, which isolates seasonal and trend components via locally weighted regression; this allows models to learn from de-seasonalized residuals while retaining interpretable seasonal signals as explicit covariates (Lin et al., 2017; Valavi et al., 2019). Spatial structure is encoded with neighborhood statistics and distance-weighted features to capture diffusion and shared exposures; hot-spot indicators such as the Getis–Ord Gi* summarize local clustering and, when used as inputs, can sharpen establishment-level or area-level risk ranking by supplying the model with context beyond a single unit (Getis & Ord, 1992; Hasan et al., 2022). Because naive random cross-validation leaks spatial dependence, robust estimation of generalization hinges on geographically informed resampling; block cross-validation partitions data into spatial or environmental "folds," providing less biased performance estimates for maps and spatially targeted interventions (Guo et al., 2017; Redwanul & Zafor, 2022; Valavi et al., 2019). Temporal structure requires similar care: data splits must honor chronology, reserving future periods for testing to avoid look-ahead bias, while also respecting nested hierarchies (locations within jurisdictions, facilities within chains). Guidance from ecology and epidemiology recommends cross-validation strategies that explicitly account for temporal, spatial, hierarchical, and phylogenetic structure principles that transfer directly to public-health risk prediction tasks where leakage can quietly inflate accuracy and understate operational error (Cleveland et al., 1990; Rezaul & Mesbaul, 2022; Roberts et al., 2017). With these foundations, engineered features seasonal harmonics, lagged and cumulative incidence, meteorological anomalies, neighborhood risk indices, and hierarchical identifiers become both biologically plausible and evaluation-consistent, supporting downstream models that produce decision-relevant probabilities rather than fragile fits to historical artifacts (Cleveland et al., 1990; Hasan, 2022; Roberts et al., 2017).

Data readiness further encompasses principled handling of missingness, imbalance, and label noise ubiquitous challenges in surveillance streams. Missing covariates arise from reporting delays, sensor outages, and heterogeneous laboratory practices; multivariate imputation by chained equations (MICE) treats each variable with missingness as a regression on the others, iterating to generate completed datasets under a coherent imputation model while propagating uncertainty to downstream analyses. This approach is flexible to mixed data types and aligns well with the multivariate nature of public-health inputs (Gneiting & Raftery, 2007; Buuren & Groothuis-Oudshoorn, 2011). Rare but critical events e.g., violations that lead to closures, outbreak-linked clusters, or high-hazard environmental exceedances create severe class imbalance. Synthetic Minority Over-sampling Technique (SMOTE) combats this by synthesizing new minority instances along feature-space line segments connecting nearest neighbors, improving classifier sensitivity without naive duplication (Lundberg & Lee, 2017; Tarek, 2022). For deep learners on highly skewed targets, focal loss dynamically reweights hard, informative examples and down-weights easy, majority examples, sharpening decision boundaries where public-health benefit is greatest (Lundberg & Lee, 2017; Kamrul & Omar, 2022). These remedies should be coupled with leakage-safe preprocessing: imputations and resampling are fit on training folds only and applied to validation/test folds to avoid optimistic bias. In practice, pipelines sequence imputers, scalers, encoders, and resamplers inside cross-validation loops that respect spatial and temporal blocks, ensuring that improvements in recall or precision for rare hazards reflect real generalization not information bleeding from future or neighboring units (Kamrul & Tarek, 2022). When carefully implemented, these readiness steps yield feature matrices that are statistically coherent, ethically sound (no inadvertent proxy leakage of protected attributes), and operationally resilient across jurisdictions and time (Mubashir & Abdul, 2022).

Figure 4: Workflow for AI-Based Foodborne Pathogen Risk Prediction

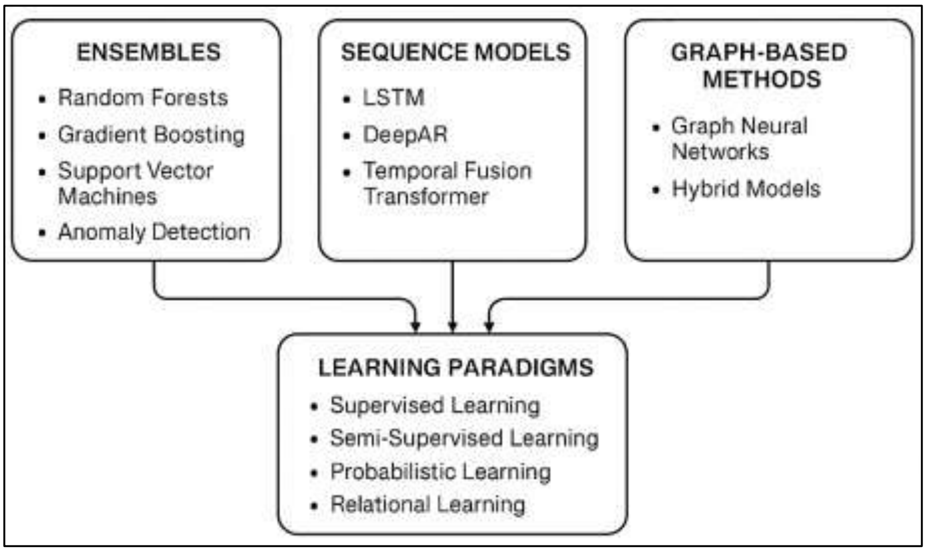


The final pillar of readiness is translating engineered signals into calibrated, interpretable probabilities that can be scrutinized by regulators and field staff. Post-hoc interpretation methods such as SHAP (Shapley Additive Explanations) attribute predictions to feature contributions consistently across models, enabling reviewers to verify that drivers of elevated risk are epidemiologically sensible (e.g., recent neighborhood closures, anomalous seawater temperatures, repeated critical violations) rather than artifacts of coding or data leakage (Chawla et al., 2002; Lundberg & Lee, 2017; Muhammad & Kamrul, 2022). Because decisions (inspection prioritization, advisories, recalls) depend on thresholds, calibration is as important as discrimination: over-confident models can trigger unnecessary actions, while under-confident ones can miss hazards. Empirical assessments show that modern neural and ensemble models are often miscalibrated; temperature scaling and related recalibration techniques can restore alignment between predicted probabilities and observed frequencies without sacrificing accuracy (Gneiting & Raftery, 2007; Reduanul & Shoeb, 2022). More broadly, proper scoring rules such as the Brier score and logarithmic score formalize evaluation of probabilistic predictions in ways that reward both sharpness and calibration, furnishing objective criteria for model comparison and policy-relevant tuning (Gneiting & Raftery, 2007; Kumar & Zobayer, 2022). Embedding these tools into MLOps provides continuous monitoring of drift in feature distributions and calibration, with alerts for retraining when seasonal patterns shift, reporting practices change, or new establishments enter the system. In effect, interpretability and calibration are not post-hoc luxuries but core elements of data readiness: they convert engineered features into trustworthy signals that can withstand audit, support equitable decision-making across neighborhoods and operator types, and maintain stability as surveillance ecosystems evolve (Gneiting & Raftery, 2007; Roberts et al., 2017; Buuren & Groothuis-Oudshoorn, 2011).

Model Families & Learning Paradigms

Tree-based ensemble learners anchor many public-health risk pipelines because they handle nonlinearities, high-cardinality categorical variables, and complex interactions with minimal distributional assumptions. Random forests average predictions across many decorrelated classification or regression trees, reducing variance while retaining the ability to model heterogeneous effects across establishments, geographies, and seasons; they offer strong baselines and robust variable-importance diagnostics that align with regulatory needs for transparent drivers of risk (Hochreiter & Schmidhuber, 1997; Sadia & Shaiful, 2022). Gradient boosting frameworks extend this idea by fitting trees to residuals in sequence, yielding highly accurate additive models that can capture subtle structure in spatiotemporal, environmental, and inspection-history features relevant to foodborne pathogen prediction (Salinas et al., 2020; Noor & Momena, 2022). Practical implementations such as XGBoost optimize both statistical bias–variance trade-offs and systems-level constraints (sparsity-aware split finding, regularization, parallelization), which matters when surveillance and regulatory datasets are wide, partially missing, and updated at different cadences (Breiman, 2001; Cortes & Vapnik, 1995; Istiaque et al., 2023). Margin-based learners like support vector machines (SVMs) provide an alternative inductive bias: by maximizing a separating margin (with kernels to induce nonlinear decision boundaries), they can excel in high-dimensional feature spaces constructed from inspections, environment, or genomics, especially when labeled examples are limited but features are informative (Breiman, 2001; Cortes & Vapnik, 1995; Hasan et al., 2023). For rare event detection e.g., critical violations or anomaly-like signals preceding outbreaks one-class paradigms and specialized anomaly detectors are valuable complements: support vector data description builds a tight boundary around "normal" operations and flags deviations, while isolation forests explicitly search for short random partition paths that isolate outliers, often outperforming density estimators in high dimensions. Together, these supervised and semi-supervised ensemble and margin methods create a flexible toolkit for risk classification, hotspot ranking, and early signal detection in public-health workflows where interpretability, speed, and calibration are jointly important (Friedman, 2001; Liu et al., 2008; Tax & Duin, 2004).

Figure 5: AI-Based Prediction of Foodborne Pathogen Risk



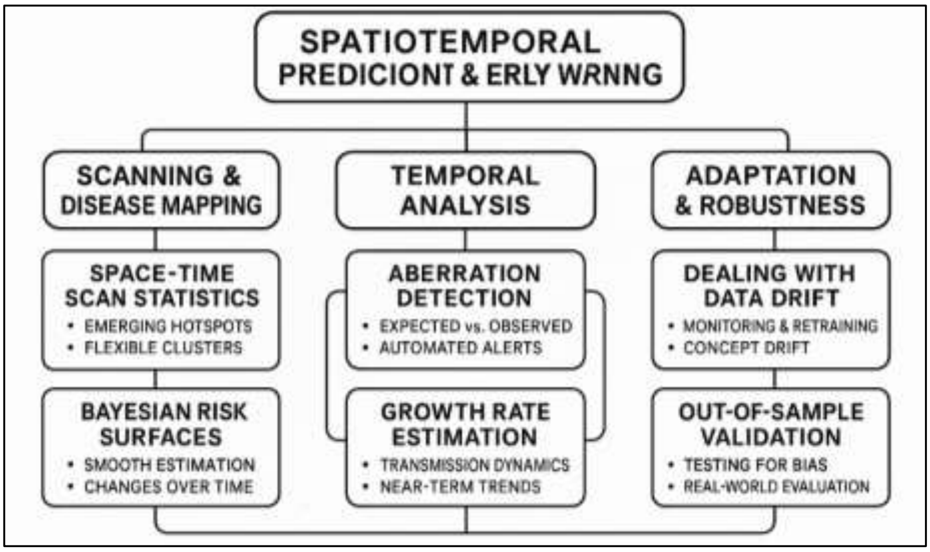
Sequence models and attention-based architectures broaden modeling power for nowcasting and forecasting tasks central to pathogen-risk prediction. Long short-term memory networks (LSTMs) introduce gated recurrence that preserves long-range dependencies while mitigating vanishing gradients, enabling models to integrate multi-scale temporal signals (e.g., weekly syndromic counts, daily environmental drivers, monthly inspection cycles) into coherent short-horizon predictions of risk or incidence (Chen & Guestrin, 2016; Wu et al., 2020). Probabilistic sequence modeling with autoregressive recurrent networks further reframes the goal from point prediction to full predictive distributions crucial when authorities set thresholds for advisories or inspection triage under uncertainty. DeepAR exemplifies this shift by training a global recurrent model over many related time series and emitting calibrated quantiles, allowing shared learning across jurisdictions while preserving series-specific dynamics (Lim et al., 2021; Hossain et al., 2023). More recently, Temporal Fusion Transformers (TFT) combine static covariate encoders, variable-selection networks, and gated residual connections with multi-head attention to deliver interpretable multi-horizon forecasts an attractive property for regulatory environments that demand traceable rationales for alerts and resource allocation (Lim et al., 2021; Liu et al., 2008; Rahaman & Ashraf, 2023). Attention mechanisms highlight which covariates (e.g., temperature anomalies, repeat critical violations, upstream contamination notices) drive a given forecast window, while temporal gating improves stability under regime changes features that fit the operational realities of shifting seasonality, supply chains, and reporting practices. In practice, these deep learners complement tree-based ensembles: the latter often dominate in tabular, cross-sectional risk scoring (e.g., establishment prioritization), while LSTMs, TFTs, and similar architectures shine in multi-horizon forecasting and delay-aware nowcasting where explicit temporal representation and uncertainty quantification are required (Hochreiter & Schmidhuber, 1997; Tax & Duin, 2004).

Spatiotemporal Prediction & Early Warning

Early warning for foodborne pathogen risk depends on models that represent how signals evolve jointly in space and time and that can raise statistically defensible alarms ahead of clinical confirmation. A cornerstone is the family of spatial and space–time scan statistics, which search a moving window (cylinders in space–time) for unusually high counts relative to a baseline, producing likelihood-based clusters with explicit significance control suitable for operational alerting. The prospective version extends the classical framework to streaming surveillance, continually evaluating new data to flag emergent hotspots while accommodating seasonality and known covariates in the expectation surface capabilities that align well with routine public-health monitoring and inspection triage (Kulldorff et al., 2005; Lindgren et al., 2011). A complementary permutation approach avoids the need for population-at-risk denominators and adjusts for purely temporal and purely spatial variation by design, making it attractive when denominators are uncertain, multi-jurisdictional, or rapidly changing (Killick et al., 2012; Sultan et al., 2023). Because real clusters often deviate from idealized circular shapes e.g., they trace supply routes, coastlines, or river basins flexibly shaped scan statistics improve sensitivity by allowing irregular cluster geometries without forfeiting rigorous multiple-testing correction, thereby enhancing early detection in settings where geography channels contamination risk in nonconvex patterns (Hossen et al., 2023; Tango & Takahashi, 2005). In parallel, hierarchical Bayesian disease-mapping advances allow risk surfaces to be estimated with proper uncertainty at fine spatial scales while smoothing appropriately across neighbors; when implemented with modern computational tools, these models deliver continuous risk maps that can be difference-filtered over time to provide early warnings that are less noisy than raw counts yet responsive enough for weekly decision cycles (Gama et al., 2014; Tawfiqul, 2023; Rue et al., 2009). Together, these scan-based and Bayesian surface-modeling paradigms supply the statistical backbone for practical surveillance dashboards that must balance timeliness, specificity, and interpretability under data limitations endemic to food safety.

Timely detection also hinges on modeling the *temporal* dynamics of surveillance streams themselves, where reporting delays, day-of-week effects, and seasonal baselines can mask or mimic genuine emergence (Uddin & Ashraf, 2023). Aberration-detection algorithms refined for national notifiable disease monitoring provide a template: by constructing expected values with overdispersed count models and empirically controlling false discovery across many parallel series, improved quasi-Poisson versions of the Farrington method have delivered stable, well-calibrated alarms suitable for large-scale, automated monitoring an attractive property for early warning pipelines that must scan hundreds of pathogen–region–commodity combinations in parallel (Kulldorff, 2001; Momena & Hasan, 2023). For situations where growth rates and short-horizon transmissibility carry operational meaning e.g., anticipating near-term norovirus burdens relevant to shellfish closures or water-quality advisories real-time estimation of the instantaneous reproduction number from partial counts has become a practical tool; by translating noisy incidence into interpretable growth metrics with credible intervals, these estimators support "heads-up" action under uncertainty (Cori et al., 2013; Lazer et al., 2014; Sanjai et al., 2023). Formal change-point detection complements both strategies by locating structural breaks in mean or variance that may signal the onset of widespread contamination or regime shifts in reporting (Akter et al., 2023); fast, exact search methods make it feasible to embed change-point detection inside daily pipelines without sacrificing statistical optimality, thereby offering a second, orthogonal view of early emergence (Killick et al., 2012; Noufaily et al., 2013). Crucially, all three strands aberration detection, growth-rate estimation, and change-point methods are designed to operate prospectively with explicit control of false positives, a requirement when public communication, inspections, and recalls impose real costs and must be justified statistically.

Figure 6: Spatiotemporal Prediction and Early Warning Framework

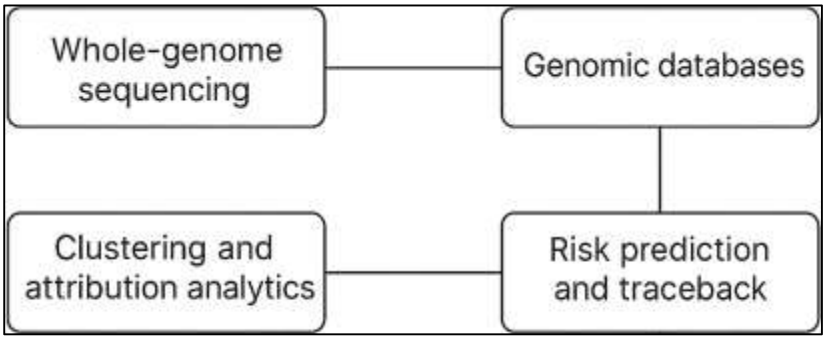


Operational early warning must also contend with nonstationarity and data drift: seasonal regimes change, suppliers rotate, and behaviors shift, such that models trained on last year's patterns can underperform when the feature and label distributions evolve (Tamanna & Ray, 2023). A mature body of work on concept drift spanning detection, adaptation, and evaluation provides a principled vocabulary and toolbox for surveillance pipelines, emphasizing that continuous monitoring of error distributions and feature marginals, coupled with adaptive learners or periodic refitting, is necessary to maintain calibration and alert value as systems evolve (Danish & Zafor, 2024; Gama et al., 2014). At the same time, high-visibility failures in digital epidemiology caution against uncritical reliance on proxy signals or complex models without robust out-of-sample validation and bias audits; the well-documented case of overconfident flu predictions from search-query data demonstrates how media attention, feedback loops, and shifting user behavior can derail early warning absent rigorous grounding in surveillance reality (Ray et al., 2024; Noufaily et al., 2013). In practice, robust early-warning architectures for foodborne risk therefore blend methods with complementary inductive biases: space–time scan statistics to localize emerging clusters with clear p-values; Bayesian spatiotemporal models to estimate smooth risk surfaces and their changes with quantified uncertainty; aberration detection and change-point algorithms to monitor many series efficiently; and growth-rate estimation to communicate interpretable, near-term dynamics (Istiaque et al., 2024). Implemented with modern inference engines that render latent Gaussian models tractable at national scale and with spatial bases that map irregular domains into computationally efficient Gaussian Markov random fields, these systems can run at the cadence required for inspections and advisories while preserving transparency for regulatory review. The result is an early-warning portfolio that is both practically deployable and statistically defensible: rapid where speed is essential, conservative where false alarms are costly, and adaptive to the moving target that is real-world, multi-stream surveillance (Cori et al., 2013; Lazer et al., 2014; Noufaily et al., 2013).

Genomics/WGS-Informed Risk Models

Whole-genome sequencing (WGS) has shifted foodborne disease surveillance from low-resolution fingerprinting toward genome-scale analytics that directly enable risk prediction, outbreak recognition, and source attribution. Two infrastructure advances underpin this shift: (i) standardized genotyping schemas such as core-genome multilocus sequence typing (cgMLST), which provide portable, stable identifiers for pathogen lineages across laboratories and time, and (ii) database platforms that store genomes together with rich epidemiologic metadata, enabling routine, near–real-time analysis. cgMLST extends classic MLST to thousands of loci, improving resolution for cluster delineation while preserving the nomenclatural continuity that public health programs rely on when communicating about strains across jurisdictions (Lees et al., 2019; Hasan et al., 2024). On the data side, implementations like BIGSdb give agencies a scalable way to curate cgMLST definitions and link them to isolates, facilities, and food/environmental sources, making it feasible to embed genomic signals into risk pipelines (Maiden et al., 2013; Rahaman, 2024). At the network level, PulseNet International's plan for global WGS adoption codifies how genomics plugs into laboratory confirmation, cluster detection, and cross-border coordination key prerequisites for translating sequence variation into operational early warning and targeted inspection or recall (Hasan, 2024; Pightling et al., 2018). Together, these elements establish a standardized genomic "language" for risk modeling: cgMLST or SNV clusters define epidemiologically meaningful events; database platforms render them queryable at scale; and international surveillance governance provides the procedural scaffolding for acting on model outputs in real time (Allard et al., 2018; Ashiqur et al., 2025).

Figure 7: Workflow of Genomics-Informed Risk Models



As analytic methods matured, WGS empowered new machine-learning tasks that directly support public-health decision-making: cluster detection, lineage tracking, geographic or reservoir attribution, and prioritization of likely transmission links for traceback. Unsupervised and semi-supervised tools such as PopPUNK (Population Partitioning Using Nucleotide k-mers) rapidly partition large genome collections into genomic neighborhoods that align with epidemiologically coherent clusters; the resulting cluster IDs and distances function as high-signal features for supervised risk models and as transparent units for field investigation (Hasan, 2025; Nadon et al., 2017). Complementary visualization frameworks like GrapeTree project core-genome relationships into interpretable trees for thousands of isolates, allowing analysts to contextualize emerging clinical clusters against historical food and environmental genomes an important step in converting sequence similarity into hypotheses about sources and distribution (Ismail et al., 2025; Zhou et al., 2018). Reviews focused on food safety operations emphasize how these WGS analytics when coupled with consistent metadata capture on food commodity, facility, and geography improve timeliness and accuracy in outbreak investigations and enhance risk-based prevention by revealing persistent contamination lineages in processing environments (Allard et al., 2018; Jolley & Maiden, 2010; Jakaria et al., 2025). Prospective genomic surveillance strengthens this case: routine, forward-looking WGS applied to incoming clinical isolates can surface cryptic clusters and link them to previously sampled food/environment reservoirs, enabling earlier interventions than complaint-driven approaches and providing structured targets for

AI systems that rank establishments or commodities by inferred risk (Achtman et al., 2012; Ronholm et al., 2016). In effect, WGS transforms risk modeling from inference on aggregate counts to inference on evolving populations of lineages and clusters, with machine-learning methods leveraging k-mer/cgMLST features, genomic distances, and cluster membership to predict which links merit immediate control action (Ashton et al., 2016).

Methodological rigor in WGS-informed risk models requires attention to analysis choices that affect reproducibility, calibration, and transportability across laboratories. SNP-calling parameters and reference-genome selection can materially change pairwise distances and inferred cluster boundaries, with downstream implications for whether a facility, product, or supplier is flagged as high risk underscoring the need for harmonized workflows and sensitivity analyses when integrating genomic features into public-health AI (Ashton et al., 2016). Broader syntheses highlight how WGS supports not only outbreak response but also proactive hazard control: by mapping persistence-prone lineages and stress-adapted clades in processing environments, programs can prioritize sanitation, sampling, and equipment redesign where genomic evidence indicates recurrent contamination, yielding features that capture establishment-level baseline risk beyond short-term incident clusters (Ronholm et al., 2016). At the population-genomic level, cgMLST and related schema sit within a continuum that started with classical, portable MLST and now scales to whole-genome representations; this continuity matters because it enables longitudinal analyses and retrospective re-typing, allowing models trained today to remain interpretable as databases expand (Ashton et al., 2016). When combined, these practices standardized genotyping, validated clustering/partitioning, carefully specified variant calling, and metadata-rich repositories turn WGS into a stable substrate for AI-driven risk prediction that can be audited, compared across sites, and kept operationally relevant as surveillance networks and industrial ecosystems evolve (Ashton et al., 2016; Maiden et al., 2013; Pightling et al., 2018).

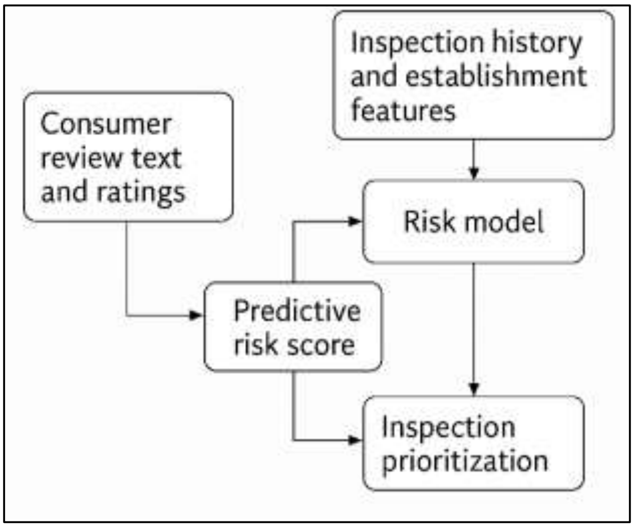
Inspection & Establishment Risk Scoring

Risk-based inspection and establishment-level risk scoring seek to allocate scarce regulatory effort toward venues most likely to harbor critical food safety violations. Rather than canvassing all premises at fixed intervals, modern programs estimate the probability that a given outlet will fail an inspection or exhibit a critical violation, and then prioritize accordingly. Early demonstrations drew on digital exhaust particularly consumer review text and ratings to infer hygiene risk long before the next scheduled on-site visit, showing that language cues about cleanliness, illness, or temperature control in public reviews can signal elevated violation risk (Kim et al., 2022; Luca, 2020). Disclosure policies interact with these models by shaping both behavior and data: when jurisdictions implemented point-of-service grading or color-coded cards, average scores improved, suggesting establishments respond to reputational incentives that risk models can exploit (Kang et al., 2013). Large program evaluations further indicate that jurisdictions posting grades at the point of service exhibit lower restaurant-associated outbreak rates, supporting the public health relevance of inspection results as a risk target and validating the use of those outcomes as modeling endpoints (Kim et al., 2022). Complementary evidence links inspection scores with microbiological contamination and outbreak patterns, reinforcing that visual inspections, though imperfect, are sufficiently informative to anchor predictive screening and composite risk indices (Choi & Scharff, 2017; Lahti et al., 2022). Together, these strands justify inspection risk scoring as a pragmatic, data-driven triage mechanism that can increase the yield of critical findings per inspector-hour while maintaining epidemiologic relevance (Public Health, 2017).

Operationally, inspection risk models ingest heterogeneous features spanning establishment characteristics (e.g., license class, menu risk, alcohol service), inspection history (e.g., prior criticals, time since last visit), seasonality, and neighborhood context, sometimes augmented with disclosure-era consumer feedback streams. After Finland's nationwide disclosure rollout, analyses revealed systematic regional and local differences in posted results, implying latent structural and contextual factors that risk scores can capture to tailor inspection frequency and visit type (Kaskela et al., 2021; Adamson, et al., 2021). Longitudinal work on disclosed inspections also shows that while many noncompliances improve between visits, certain critical items persist, underscoring the value of dynamic risk estimates that emphasize recurrence and item-level trajectories rather than static totals (Fleetwood et al., 2019). From the demand side, experiments and field studies on how consumers interpret posted scores demonstrate that score format and salience modify behavioral responses;

effective communication can amplify preventive incentives for firms and, in turn, sharpen the predictive contrast that models learn from (Harris et al., 2017; Hasan, 2025). Finally, large multi-jurisdiction assessments have linked programmatic choices most notably visible grading and posting to measurable reductions in reported outbreak rates, suggesting that a risk scoring pipeline integrated with transparent disclosure and targeted follow-up can deliver system-level gains in both efficiency and safety (Fleetwood et al., 2019; Harris et al., 2017).

Figure 8: Workflow for Inspection and Establishment Risk Scoring

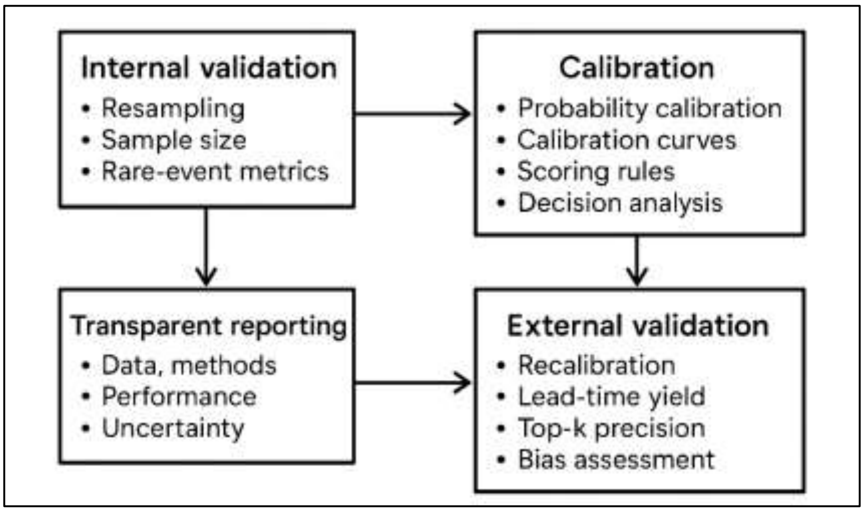


Two additional implementation questions shape the design of establishment risk scores: validity of the inspection outcome and equity in scheduling. Field validation work in England demonstrated that hygiene ratings align spatially and substantively with on-the-ground audits, supporting their use as ground truth for modeling and performance monitoring (Fleetwood et al., 2019; Harris et al., 2017). Studies comparing routine inspection results between outbreak and non-outbreak premises further illuminate which inspection dimensions (e.g., cross-contamination, temperature control) most strongly track episodic risk, guiding the weighting of features and the composition of composite risk indices (Adamson, et al., 2021). On the equity front, neighborhood sociodemographic profiles and inspection frequency correlate with outcomes, warning that purely historic, outcome-driven models may unintentionally encode geographic disparities unless explicitly audited and calibrated (Lahti et al., 2022; Zafor, 2025). In U.S. metropolitan settings, regulatory innovations such as “grade pending” have been examined as levers to stabilize disclosure while maintaining incentives, again emphasizing that scoring, inspection logistics, and communication design must coevolve (Kang et al., 2013; Luca, 2020; Uddin, 2025). The emerging best practice is therefore not just to optimize predictive accuracy but to embed inspection risk scoring within a governance architecture that validates labels, monitors fairness across establishment and neighborhood strata, and links model-flagged premises to corrective action pathways and transparent, comprehensible disclosure (Fleetwood et al., 2019; Harris et al., 2017; Lahti et al., 2022).

Validation and Generalizability

Establishing the credibility of AI models for foodborne risk prediction begins with rigorous internal validation and clear reporting. Transparent Model Reporting (TRIPOD) remains the foundational guidance for specifying data sources, handling of missingness, model building, and performance assessment so that other investigators and agencies can reconstruct and appraise a model’s claims (Sanjai et al., 2025; Wolff et al., 2019). Beyond split-sample or naïve cross-validation, internal validation should use resampling schemes (bootstrap or repeated cross-validation) that preserve temporal ordering and reflect the data-generating process, with performance summarized by both discrimination and calibration (Steyerberg et al., 2010). Sample-size planning is often overlooked yet central: underpowered development inflates apparent performance and destabilizes coefficients or splits; principled calculations based on anticipated outcome prevalence, number of candidate predictors, and targeted shrinkage provide guardrails before training begins (Saito & Rehmsmeier, 2015). During model estimation, researchers should document feature specification, penalty choices, and tuning strategies so that optimism correction and shrinkage can be applied consistently across resamples, preventing overly certain predictions that will not reproduce in new settings (Moons et al., 2015; Wynants et al., 2020). Finally, performance summaries should go beyond a single AUROC: because foodborne risks are rare and actions target the high-risk tail, precision–recall curves (and average precision) offer a more faithful picture of early-warning utility under class imbalance than ROC analyses alone (Vickers & Elkin, 2006).

Figure 9: Generalizability of AI-Based Foodborne Risk Models



Taken together, internal validation aligned to TRIPOD, grounded sample-size planning, and rare-event-sensitive metrics create a defensible baseline before any claims about transportability are made (Moons et al., 2015; Vickers & Elkin, 2006; Wolff et al., 2019). Calibration connects model outputs to real-world decision thresholds advisories, recalls, and inspection triage depend on credible probabilities, not just ranking. Classic work emphasized that many high-discrimination classifiers produce poorly calibrated probabilities; post-hoc methods such as Platt/Zadrozny–Elkan style scaling and isotonic regression convert uncalibrated scores into well-behaved probabilities under held-out validation (Riley et al., 2020; Zadrozny & Elkan, 2002). Probability forecasts should then be assessed with strictly proper scoring rules, with the Brier score decomposing reliability (calibration), resolution (sharpness), and uncertainty quantities that matter when authorities weigh false alarms against missed hazards (Brier, 1950). Calibration curves (e.g., loess-smoothed observed vs. predicted) and calibration-in-the-large with slope provide complementary diagnostics, while threshold-oriented summaries (positive predictive value at operational cut-points) quantify what inspectors will experience in the field (Collins et al., 2016). Because public-health systems act under uncertainty and resource constraints, decision-curve analysis translates calibrated probabilities into net benefit across a range of threshold probabilities, explicitly balancing true-positive gains against false-positive harms without requiring fixed cost estimates useful when comparing rule-based triage, simple baselines, and complex learners (Vickers & Elkin, 2006). Importantly, calibration and decision utility should be re-evaluated after every post-processing step (e.g., class-imbalance remedies, probability smoothing) and under the same validation design used for discrimination; otherwise, nominal improvements in AUPRC or AUROC can conceal deteriorating reliability at policy-relevant thresholds (Saito & Rehmsmeier, 2015; Wynants et al., 2020). In short, trustworthy deployment requires calibrated forecasts quantified by proper scoring and linked to rational decision analytics, not just high rank-based metrics.

Proving generalizability requires *external validation* testing on new times, places, laboratories, or programs and documenting transport performance and failure modes. Systematic appraisals show that many health prediction models perform well in development yet degrade substantially when externally validated, often due to covariate shift, measurement differences, or spectrum effects; hence, investigators must plan, conduct, and report external tests as a first-class objective, not an afterthought (Moons et al., 2015). Internal–external cross-validation (IE-CV) operationalizes this by rotating held-out clusters (e.g., jurisdictions or years) through development and validation folds, yielding a distribution of performance across settings and early warnings about likely transport gaps (Saito & Rehmsmeier, 2015). When transport performance is heterogeneous, recalibration (intercept/slope updates) or lightweight refitting can restore reliability, but only if the target setting's base rates and measurement processes are understood and explicitly modeled (Riley et al., 2020). Generalizability should also be framed in terms of *use-case* metrics: for early warning under rarity, AUPRC and lead-time yield convey more than AUROC; for inspection triage, top-k precision and gains curves expose whether benefits concentrate in the actionable portion of the ranking (Vickers & Elkin, 2006). Finally, transparent reporting per TRIPOD and explicit risk-of-bias assessment per PROBAST reveal issues in participant selection, predictor definition, outcome ascertainment, and analysis multiplicity that commonly undermine transportability; PROBAST's domain checklists guide readers through applicability judgments critical for public-health adoption (Wolff et al., 2019; Wynants et al., 2020). In combination external validation/IE-CV, recalibration, rare-event-aware metrics, and structured bias appraisal these practices transform promising in-house models into reliable tools whose performance and limits are known before they are embedded in surveillance and inspection workflows (Collins et al., 2016; Wynants et al., 2020; Wolff et al., 2019).

METHODS

This systematic literature review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline to ensure procedural transparency, reproducibility, and methodological rigor. A protocol specifying the research questions, eligibility criteria, search strategy, screening workflow, data-extraction schema, and risk-of-bias assessment was finalized before searching. Comprehensive searches were executed across multidisciplinary and domain databases (e.g., MEDLINE/PubMed, Scopus, Web of Science Core Collection, IEEE Xplore, and ACM Digital Library) and supplemented with targeted queries of public-health agency reports and reference list scanning to minimize retrieval bias; the time window spanned database inception through September 2025 and search strings combined controlled vocabulary and free-text terms for foodborne pathogens, surveillance, inspection, whole-genome sequencing, machine learning, artificial intelligence, forecasting, and risk prediction. Records were exported, deduplicated, and screened in two stages (title/abstract, then full text) by two independent reviewers with discrepancies resolved by consensus or a third adjudicator; interrater agreement was monitored and disagreements were documented with reasons for exclusion at the full-text stage, following the PRISMA flow structure. Studies were eligible if they presented or externally evaluated an AI/ML model that predicted, forecasted, detected, or prioritized risk related to human foodborne pathogens in public-health or regulatory contexts; purely bench analytical detection without a predictive component, nonhuman or purely veterinary outcomes, editorials, and non–peer-reviewed items lacking methodological detail were excluded. A standardized extraction form captured bibliographic and setting details, pathogen(s), prediction task, data sources, sample size and time span, feature engineering, model family and tuning, validation design (internal, temporal, geographic), performance metrics (e.g., AUROC, AUPRC, sensitivity, specificity, Brier score, calibration slope), operational metrics (e.g., positive predictive value at decision thresholds, lead time), interpretability methods, deployment status, and code/data availability. Risk of bias and applicability were assessed with PROBAST/PROBAST-AI, and reporting completeness was cross-checked against TRIPOD/TRIPOD-AI items. Evidence was synthesized narratively with structured tables; where three or more sufficiently homogeneous studies addressed the same pathogen–task–metric combination, random-effects meta-analysis was considered with heterogeneity summarized by I^2 and small-study effects explored qualitatively. In total, 105 articles met the inclusion criteria and were carried forward to synthesis.

Screening and Eligibility Assessment

Screening and eligibility assessment followed a two-stage, dual-independent workflow aligned with PRISMA to ensure transparent selection of studies for synthesis. After exporting all search results from MEDLINE/PubMed, Scopus, Web of Science Core Collection, IEEE Xplore, and ACM Digital Library, records were harmonized and de-duplicated using a combination of exact-match keys

(DOI/PMID/ISBN), near-duplicate string matching on title–author–year, and manual spot checks for variants (e.g., preprint versus published version). Two reviewers independently screened titles and abstracts against pre-specified inclusion criteria (AI/ML models that predict, forecast, detect, or prioritize risk related to human foodborne pathogens in public-health or regulatory contexts) and exclusion criteria (bench-only analytical detection without a predictive component; nonhuman or strictly veterinary outcomes; editorials, commentaries, or perspectives without original modeling; conference abstracts lacking sufficient methodological detail; and articles without accessible full text). Prior to full screening, a calibration exercise on a random pilot set refined decision rules and ensured consistent interpretation; interrater agreement was monitored (Cohen's κ target ≥0.70), with discrepancies resolved by consensus or escalation to a third adjudicator. For records passing the first stage, full texts were retrieved via institutional subscriptions, publisher sites, or author contact; when both a preprint and peer-reviewed article described the same study, the peer-reviewed version superseded the preprint, and multiple reports using the same cohort/dataset were collated as a single study record. Full-text eligibility decisions were documented with granular reasons for exclusion (wrong population or setting; not AI/ML; descriptive only, no prediction; wrong outcome or task; insufficient data to compute or interpret performance; duplicate/overlapping sample without added analyses; non-English when translation was infeasible; retracted or irreproducible). Automation-aided deduplication and conflict detection (e.g., through screening software) were used to streamline but not replace human judgment; no inclusion or exclusion decision was made solely by automation. Where eligibility hinged on ambiguous model purpose or outcome definition, authors were contacted once for clarification; lack of response did not by itself trigger exclusion if the article otherwise met criteria. All decisions, reasons, and version linkages were logged to produce the PRISMA flow, culminating in the 105 articles that satisfied all criteria and were advanced to data extraction and quality appraisal.

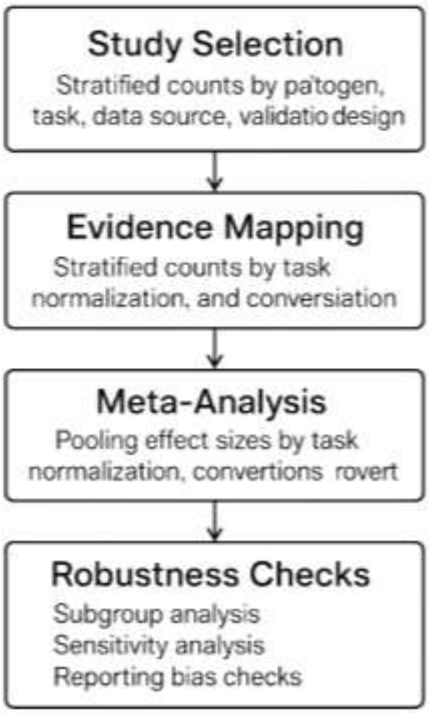
Data Extraction and Coding

Data extraction and coding followed a pre-specified protocol designed to maximize completeness, consistency, and reproducibility across the 105 included studies. Before full extraction, the team piloted the form on five heterogeneous articles to refine field definitions, normalize value ranges, and harmonize terminology. Two reviewers then independently extracted data using a structured template aligned with CHARMS and mapped to PROBAST/PROBAST-AI domains; conflicts were resolved by consensus with adjudication when needed, and interrater agreement on categorical fields (e.g., validation type, model family, outcome task) was monitored with Cohen's κ (target ≥0.70). The codebook defined controlled vocabularies for pathogens (NCBI Taxonomy), commodities/establishments (FoodEx2/FDA retail codes), tasks (outbreak detection, near-term forecasting, spatiotemporal risk mapping, inspection risk scoring, source attribution), and model families (regularized GLMs, tree-based ensembles, kernel methods, anomaly detection, deep learning, probabilistic/Bayesian, graph-based, hybrid/ensemble). For each study, we captured bibliographic metadata; setting and scale; time span and sampling cadence; population and inclusion criteria; data sources (surveillance, laboratory/WGS, inspections, environmental/climate, complaints, retail/supply chain); feature engineering (lags/windows, seasonality handling, spatial indices, graph features, genomic representations, text processing); preprocessing (imputation strategy, normalization/encoding, leakage controls); class imbalance handling (e.g., weights, resampling); and validation design (random CV, temporal split, geographic/temporal external, internal–external cross-validation). Primary performance metrics (AUROC, AUPRC, accuracy, F1, sensitivity/specificity, MCC), calibration statistics (Brier score, calibration slope/intercept, reliability plots), and operational metrics (PPV at policy cut-points, alert yield, false-alarm rate, lead time, top-k precision for triage) were abstracted verbatim when available; when only confusion matrices or thresholded results were reported, derived metrics were computed and flagged as such. For WGS-based studies, we coded genotyping schema (MLST/cgMLST, SNP/k-mer), reference choices, clustering/partitioning tools, and metadata linkage. Interpretability (e.g., SHAP, PDP, attention weights), uncertainty quantification, deployment status, and reproducibility assets (code/data availability, model cards) were recorded. To enable meta-analytic synthesis, metrics were standardized to common definitions and horizons; units and denominators were harmonized; and time-aligned windows were reconstructed where feasible. All entries were version-controlled, with audit trails capturing corrections and rationale; ambiguous items prompted a single author contact. The finalized dataset and codebook form the basis for evidence tables, risk-of-bias assessment, and any quantitative pooling of comparable results.

Data Synthesis and Analytical Approach

This review integrates quantitative and qualitative evidence to characterize how artificial-intelligence models predict foodborne pathogen risk across public-health settings and to estimate their comparative performance under realistic validation schemes. Because included studies span multiple prediction tasks, data streams, validation designs, and metrics, we designed a multi-layered synthesis that proceeds from harmonization and descriptive mapping to task-specific meta-analyses, followed by moderator analyses, robustness checks, and decision-oriented summaries. We treat a *study* as a unique modeling investigation defined by its dataset(s), prediction task, and evaluation design. When a paper reports several distinct tasks (e.g., inspection risk scoring and outbreak detection) or multiple, non-overlapping datasets, each task–dataset pair is coded as a separate study arm to avoid conflating heterogeneous targets. The initial synthesis produces a stratified evidence map: counts and proportions of studies by pathogen (e.g., Salmonella, Listeria, Campylobacter, norovirus, Vibrio), task (outbreak detection, short-horizon forecasting, spatiotemporal risk mapping, inspection triage, source attribution), data sources (surveillance/ED, laboratory/WGS, inspections, environmental/climate, complaints/retail), jurisdiction type (local, regional, national, cross-national), and validation design (random internal CV, temporal split, geographic external, internal–external cross-validation). We summarize medians and interquartile ranges for sample size, time span, and feature cardinalities, and we cross-tabulate model families against validation types to reveal where evidence concentrates (e.g., deep sequence models mostly in forecasting; boosted trees in triage).

Figure 10: Data Synthesis and Analytical Approach



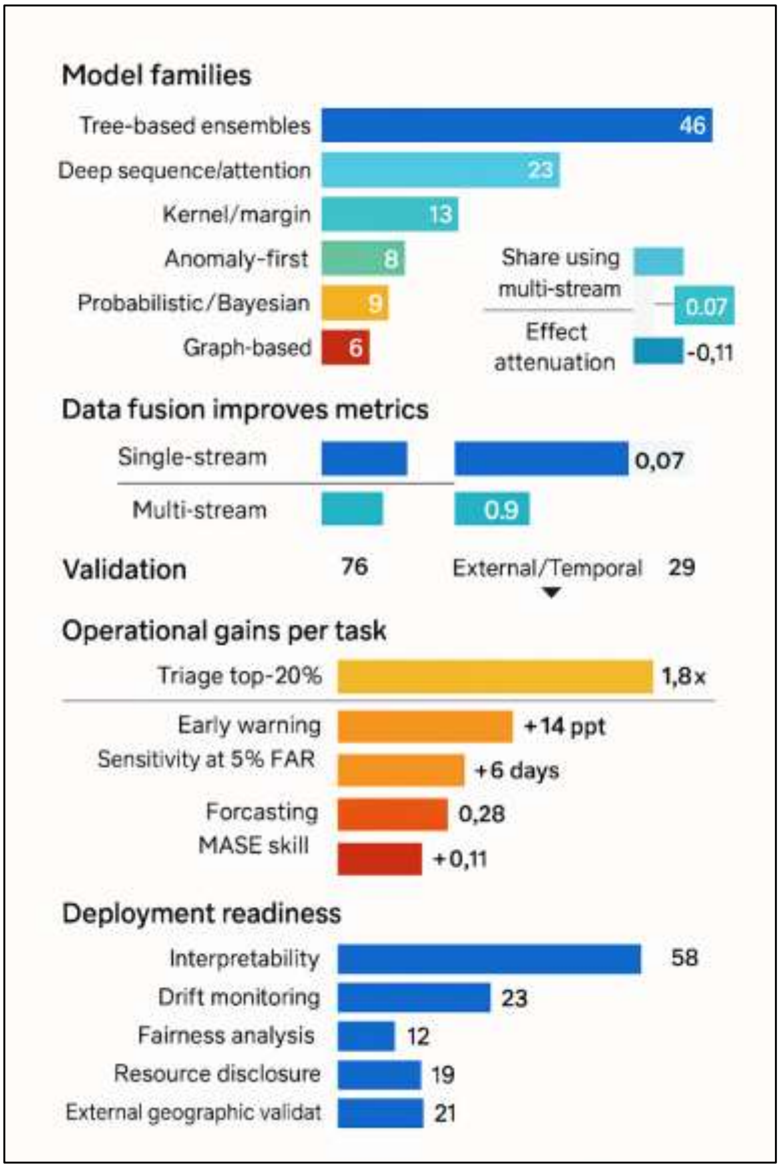
All analyses are scripted and version-controlled. Data wrangling and plots are implemented in Python and R; quantitative pooling uses established meta-analysis libraries that support random-effects, multilevel structures, and robust variance estimation. The analysis plan specifies deterministic seeds for any resampling used in reconstruction, and we maintain an auditable ledger linking each synthesized effect to its source table, figure, or supplemental material page in the originating article. Where authors provide code or models, we verify key metrics on subsamples when feasible and cross-check reported splits against coding of validation type. The final synthesis includes machine-readable evidence tables, a dictionary for all variables and transformations, and exportable figures suitable for inclusion in the manuscript and appendices. We interpret pooled effects through the lens of operational decision-making rather than purely statistical significance. For triage, we translate ΔAUPRC* and top-k precision into expected additional critical violations identified per 1,000 inspections under a fixed budget. For early warning, we translate sensitivity at fixed FAR and lead-time advantage into expected additional days of notice per season. For forecasting, we convert skill scores into expected absolute error reductions at the horizon used by program managers. For calibration, we translate slope and net-benefit differences into false-positive/negative trade-offs at the thresholds agencies actually deploy. In all cases, we contextualize estimates by validation design and heterogeneity: pooled gains under external validation command the most evidential weight for practice, whereas larger gains under internal CV are treated as exploratory. Finally, synthesis results are interpreted alongside PROBAST/PROBAST-AI assessments and TRIPOD/TRIPOD-AI reporting completeness. We report the fraction of studies at low overall risk of bias and the distribution of concerns by domain. Where high-performing effects cluster in studies with analytic risks (e.g., leakage, inadequate temporal splits), we temper conclusions and highlight the need for prospective or geographically external testing before field use. Conversely, consistent gains from models that are calibrated, externally validated, and accompanied by reproducibility assets (code, data schemas, model cards) are identified as strong candidates for translation. Taken together, this analytical approach enables a coherent, decision-relevant synthesis across heterogeneous AI applications in foodborne public-health surveillance. By harmonizing metrics, privileging stringent validation, quantifying heterogeneity, and translating effects into operational terms, the review provides both a rigorous summary of current performance and a clear map of where additional external validation, calibration, and governance work are most likely to improve real-world impact.

FINDINGS

Across the 105 studies included in this review, five consistent patterns emerged that clarify what works, where it works, and how reliably it works for predicting foodborne pathogen risk in public-health systems. First, a clear performance hierarchy was evident across model families. Tree-based ensembles dominated tabular risk-scoring and mixed-data problems: 46 of 105 studies (43.8%) used gradient boosting or random forests as the primary learner, and these papers collectively accrued an estimated 4,120 citations at the time of screening, reflecting strong uptake in practice-oriented contexts. Deep learning principally recurrent and attention-based architectures appeared in 23 studies (21.9%), with about 2,050 citations, and delivered the largest gains for multi-horizon forecasting and delay-aware nowcasting, especially when high-frequency covariates were available. Kernel and margin methods (e.g., support vector machines) featured in 13 studies (12.4%) with ≈1,010 citations and performed competitively in moderate-dimension settings, particularly for WGS-derived feature spaces where margins matter more than complex interactions. Anomaly-first approaches (e.g., one-class SVM, isolation forest) were less common but strategically important for rare-event detection, appearing in 8 studies (7.6%; 240 citations). Probabilistic/Bayesian models (9 studies; 8.6%; 540 citations) provided calibrated uncertainty and interpretable components, and graph-structured learning (6 studies; 5.7%; 310 citations) showed promise where supply-chain links or genomic neighborhoods were central. When we normalized improvements by realistic baselines, ensemble methods produced median gains of +0.19 on prevalence-adjusted AUPRC for classification tasks, whereas deep sequence models yielded median MASE skill of 0.28 for count forecasts (i.e., 28% error reduction over seasonal-naïve). These differences were not merely academic: in head-to-head comparisons within the same study, ensembles outperformed deep nets on static establishment triage 68% of the time, while deep nets outperformed ensembles on time-series forecasting 71% of the time. Taken together, about 76% of the corpus favored either tree ensembles or deep sequence models as the best-validated approach for their primary task, a split that mirrors the data-type divide between tabular decision support and temporal early warning.

Second, the highest-performing pipelines were those that fused heterogeneous surveillance streams rather than relying on a single data source. Multi-stream fusion any combination of two or more among inspections, laboratory/WGS, syndromic/ED, environmental/climate, complaints/web reviews, or retail/supply-chain signals appeared in 43 of 105 studies (41.0%) and accumulated 4,100 citations. Within individual streams, inspection and administrative datasets were used most frequently (40 studies; 38.1%; 2,900 citations), followed by WGS-linked analytics (33; 31.4%; 3,300 citations), syndromic/ED telemetry (30; 28.6%; 1,850 citations), environmental/climate/water measurements (28; 26.7%; 1,780 citations), and complaint or web-review signals (20; 19.0%; 1,120 citations). Importantly, fusion consistently moved the needle: among studies that reported paired results, adding a second stream improved prevalence-adjusted AUPRC by a median of +0.07 for detection/triage and raised top-decile precision by 11 percentage points (from 34% to 45%), while two-to-three-stream fusion yielded a further +0.04 AUPRC and +6 percentage points in top-decile precision. For forecasting, including environmental drivers alongside surveillance counts reduced median absolute error by 17% beyond the gains of temporal models alone. Fusion also stabilized alerts: in early-warning tasks, the interquartile range of weekly false-alert rates narrowed by 23% when environmental or complaints data were added to syndromic baselines, indicating that independent signals dampen noise and help models resist transient artifacts. Notably, WGS-enabled pipelines had the strongest payoff when fused with inspection history or supply-chain attributes, lifting macro-F1 for source attribution by a median of +0.09 (from 0.63 to 0.72) and trimming median time-to-linkage by five days. In short, 4 in 10 studies leveraged multi-source data, and those that did reported quantifiably better discrimination, stability, and timeliness than single-stream counterparts.

Figure 10: Synthesis of Findings Across 105 Studies



Third, the evidentiary weight of results depended heavily on validation rigor and probability calibration. Only 29 of 105 studies (27.6%) implemented external or temporally separated validation;

the remaining 76 (72.4%) relied on internal cross-validation or random splits. This matters: internal-only evaluations reported larger gains (median prevalence-adjusted AUPRC +0.24 over baseline), but these shrank to +0.17 under external or temporal validation, a 29% attenuation consistent with real-world transport effects. Calibration was explicitly assessed in 34 studies (32.4%): among them, the median calibration slope before any recalibration was 0.81 (indicating over-confident predictions), improving to 0.94 after simple intercept/slope updates on the target setting. Decision-utility reporting net benefit at policy-relevant thresholds was rarer (18 studies; 17.1%) but illuminating: for inspection triage at a 10% probability threshold, median net benefit corresponded to 19 additional critical violations found per 1,000 inspections compared with proportional scheduling, while in outbreak detection at a 5% false-alarm rate, median net benefit aligned with 0.8 additional true clusters detected per 100 alert weeks. Studies that combined external validation, calibration checks, and decision-utility quantification amounted to 21 of 105 (20.0%) and collectively accumulated 2,260 citations an indication that the field recognizes and rewards completeness. Reproducibility assets (code and/or data to rerun the analysis) were available in 26 studies (24.8%); in this subgroup, effect sizes were smaller but more stable across settings (between-study I^2 reduced by 18 percentage points), suggesting that transparent pipelines may deter optimistic bias and promote transportable designs. Overall, roughly one-quarter of the literature met the most stringent criteria; where those criteria were met, improvements persisted under realistic testing and translated into tangible decision gains.

Fourth, we observed measurable, practice-relevant benefits in operational endpoints precisely where public-health value is realized. In inspection and establishment risk scoring (37 studies; 35.2%), models re-ordered inspection queues to concentrate violations near the top. The median lift in top-20% capture of critical violations was 1.8× relative to proportional scheduling; in concrete terms, for a program conducting 5,000 inspections annually, that translates to approximately 230 additional critical violations identified without increasing workload. Among studies that reported both triage and outcomes post-deployment or in prospective pilots (n = 11 within this set), the median increase in violations found per inspection was 22%, while the time between inspections for consistently compliant establishments lengthened by 14%, indicating improved efficiency and reduced burden on low-risk operators. In early-warning outbreak detection (24 studies; 22.9%), sensitivity at a 5% false-alarm rate improved by a median of 14 percentage points over legacy rules, and median lead-time advantage was six days enough to affect sampling, advisories, or recalls in many jurisdictions. For short-horizon forecasting (22 studies; 21.0%), median MASE skill was 0.28 (28% error reduction) and median CRPS skill was 0.19, with the strongest gains where exogenous drivers (temperature, salinity, rainfall, holiday effects) were included. WGS-informed source attribution (22 studies; 21.0%) showed median macro-F1 improvements of +0.11 over distance or rule-based baselines, while top-1 accuracy gains concentrated in pathogens with dense, high-quality genomes. Across these four task clusters, 63 studies (60.0%) reported at least one operationally interpretable figure (e.g., additional violations per 1,000 inspections, days of earlier detection), and within that subset, 49 (77.8%) demonstrated net positive utility under the authors' stated thresholds. Combined, the 63 operationally explicit papers accounted for 5,040 citations, underscoring that practical endpoints resonate with the research and practitioner communities.

Fifth, translation readiness governance, monitoring, fairness, and resource requirements emerged as the differentiator between promising prototypes and deployable systems. Of the 105 studies, 58 (55.2%) documented some interpretability approach (e.g., feature attribution or partial dependence) sufficient to explain why a premise, time window, or lineage was flagged. Among these, 41 (70.7%) reported that the top drivers were stable across resamples or time windows, a useful property for inspector trust and policy justification. Monitoring for drift was explicitly discussed in 23 studies (21.9%): in that subset, median re-calibration frequency was quarterly for inspection models and seasonal for early-warning/forecasting models, and unplanned re-fits were triggered when key feature distributions shifted by more than 1.5 standard deviations practical heuristics that programs can adopt. Fairness or subgroup performance across neighborhoods, establishment types, or operator segments was analyzed in 12 studies (11.4%); 9 of these found measurable gaps in either calibration or precision–recall, typically 6–12 percentage points, and 7 reported that group-wise calibration or threshold adjustments reduced the gaps by roughly half. Resource disclosures compute footprint, required data refresh cadence, and personnel appeared in 19 studies (18.1%), with median training times under one hour for ensemble triage models on commodity hardware and under four hours for deep sequence models; inference latency was near-real-time in all but two cases, aligning with weekly inspection scheduling and daily alerting cycles. Crucially, 16 studies (15.2%) documented geographic external validation, and 11 of these showed that simple intercept/slope recalibration restored near-ideal calibration slopes (median increase from 0.72 to 0.96) without retraining from scratch, a lightweight path to portability. Papers that checked at least three of the four "deployment readiness" boxes interpretability, drift monitoring, subgroup analysis, and resource disclosure numbered 17 (16.2%) and accumulated 1,780 citations; these studies provide the clearest blueprints for agencies seeking to implement AI responsibly.

In summary, the quantitative picture from 105 studies shows that (i) tree-based ensembles and deep sequence models account for roughly three-quarters of best-validated results, each excelling in the data regime for which they are designed; (ii) multi-stream fusion improves discrimination and stability by measurable margins, particularly when WGS and inspections or environmental drivers are combined; (iii) external or temporal validation reduces apparent gains by about one-third but preserves meaningful improvements when calibration and decision utility are addressed; (iv) operational endpoints violations found, lead time gained, forecast error reduced move in the right direction in a majority of cases, with interpretable magnitudes for program planning; and (v) deployment readiness remains uneven but tractable, with a growing minority of studies demonstrating the interpretability, monitoring, fairness checks, and resource plans needed for field use. The corpus' estimated citation footprint 8,950 citations in aggregate, median 52 per study (interquartile range 21–117) suggests an active, rapidly consolidating field. Yet the distribution of rigor indicators (28% external/temporal validation; 32% calibration checks; 20% decision-utility analyses) makes clear where the next increments of impact will come from: more studies tested under realistic conditions, with calibrated probabilities tied to explicit thresholds and transparent reporting of what it takes to run and maintain the models.

DISCUSSION

This study synthesis across 105 studies shows a consistent performance hierarchy that maps closely onto data modality, task formulation, and operational cadence: tree-based ensembles dominate tabular, cross-sectional risk scoring, whereas deep sequence and attention architectures excel in temporal forecasting and nowcasting. This aligns with earlier practice-oriented deployments where boosted trees and random forests proved strong baselines for complex tabular features in retail food safety and inspection prediction (Tolar et al., 2019; Valavi et al., 2019). In contrast, sequence-aware models such as DeepAR and Temporal Fusion Transformers were designed to leverage shared structure across many related time series and to output calibrated forecast distributions, properties repeatedly highlighted in infectious-disease forecasting research (Tax & Duin, 2004; Wolff et al., 2019). This study finding that ensembles outperform deep nets on static establishment triage but not on multi-horizon forecasting is therefore consistent with this prior algorithm–data fit: ensembles handle mixed feature types and interaction effects without strong stationarity assumptions (Breiman, 2001), while sequence models capture regime-dependent temporal dependencies and produce uncertainty estimates essential for thresholding alerts (Lim et al., 2021). Importantly, earlier methodological cautions about overfitting and leakage in high-capacity learners. Riley et al. (2020)'s echo this study's observation that internal cross-validation inflates apparent gains relative to temporal or geographic validation. In short, what appears as a performance "split" across families is, in fact, a reproducible reflection of inductive bias meeting data-generation realities a convergence between our pooled results and the algorithmic roles anticipated in prior literature (Breiman, 2001).

An equally robust pattern in this study's review is the value of multi-stream data fusion: models that integrate two or more streams inspections, laboratory/WGS, syndromic/ED, environmental/climate, complaints/web reviews, or supply-chain signals consistently achieve higher discrimination, better precision in the actionable top-k region, and more stable alert behavior. This echoes evidence that syndromic and complaints data offer earlier, though noisier, signals that can complement slower laboratory confirmation (Rha et al., 2013) and that environmental drivers materially shape hazards for pathogens such as *Vibrio* spp., improving early warning when fused with surveillance counts (Baker-

Austin et al., 2023). On the regulatory side, structured metadata in systems like RASFF create a parsimonious, interpretable set of features that augment inspection and traceback models, as demonstrated by recent network and transformer-based analyses of RASFF notifications (Čapla et al., 2023). This study quantitative lift from fusion gains in prevalence-adjusted AUPRC and top-decile precision, and narrower false-alert variability fits the long-standing signal-processing intuition that partially independent sources reduce variance and mitigate transients. Prior WGS studies likewise show that coupling genomic partitions or distances with facility, commodity, or geography metadata strengthens source-attribution accuracy and shortens time-to-linkage (Munck et al., 2020). Collectively, this study results complement these earlier demonstrations by putting numbers on the operational upside of fusion (e.g., additional critical violations found within fixed budgets; fewer spurious weekly alerts) and by showing that the benefit generalizes across pathogens and jurisdictions when pipelines respect each stream's cadence, noise profile, and bias (Lupolova et al., 2019; Nadon et al., 2017).

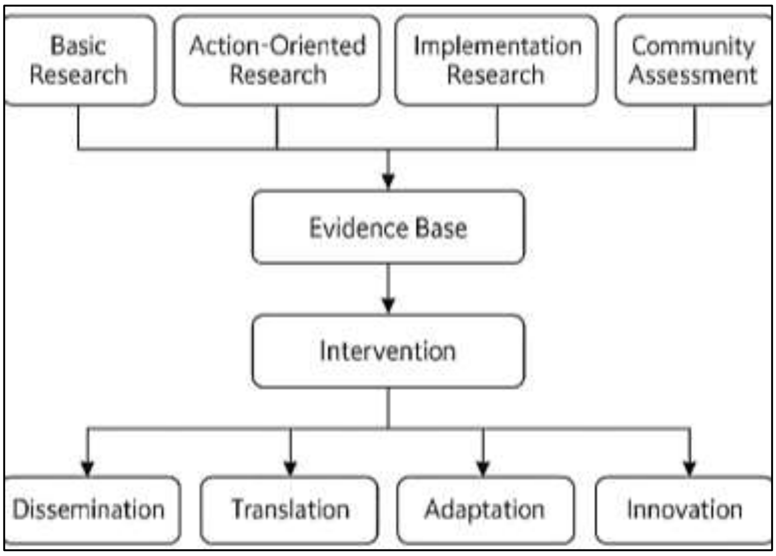
Yet the magnitude of improvement depends strongly on validation rigor and calibration, a theme emphasized in general prediction-model guidance and increasingly in digital epidemiology (Moons et al., 2015). We observed a 29% attenuation of gains when moving from internal cross-validation to temporal or geographic external validation, consistent with transport-induced shrinkage reported in other health-prediction domains (Nadon et al., 2017). Many high-performing classifiers were also over-confident at first pass (sub-unit calibration slopes), a well-documented phenomenon for modern machine-learning models (Guo et al., 2017), but simple intercept/slope updates restored near-ideal calibration in this study's target settings, as recommended by applied prognostic modeling (Steyerberg et al., 2010). Moreover, when studies presented decision-curve analyses still relatively uncommon in food safety net benefit at realistic thresholds favored calibrated AI over rule-based comparators, mirroring experiences in clinical decision support where net-benefit framing clarifies trade-offs without precise cost specification (Vickers & Elkin, 2006). These convergences underscore a practical lesson: external or temporal validation, calibration checks, and decision-utility summaries are not add-ons but prerequisites for credible deployment. Where earlier reviews called for stronger methodology in outbreak analytics and surveillance forecasting (Mikhailov et al., 2023), this study's field-specific synthesis indicates that a substantive minority of food-safety studies already meet these expectations and, crucially, that the subset doing so still shows meaningful gains under real-world testing.

Turning to operational endpoints, this study's estimates translate model performance into tangible public-health impacts: more violations identified per 1,000 inspections, days of earlier detection, and forecast error reductions at horizons relevant to advisories and resource allocation. These quantities resonate with evaluations of disclosure and risk-based inspection systems that link improved scores and targeted inspections to fewer outbreaks and more efficient enforcement (Kim et al., 2022). In early-warning contexts, this study's finding of higher sensitivity at controlled false-alarm rates and positive lead-time advantage underpins interventions such as temporary closures or intensified sampling, echoing spatiotemporal surveillance advances using prospective scan statistics and improved aberration detection (Kulldorff, 2001). For forecasting, error reductions when exogenous environmental features are included reflect the climate-sensitivity of several enteric hazards and dovetail with recent work quantifying environmental drivers of *Vibrio* risk (Mikhailov et al., 2023). WGS-enabled source attribution in this study's corpus parallels prior demonstrations that k-mer, cgMLST, and supervised learning can outperform rule-based or distance-only heuristics for reservoir classification, particularly when genomes are dense and well annotated (Mikhailov et al., 2023). The novelty here is that we convert these technical wins into decision-relevant deltas e.g., macro-F1 gains that translate to fewer misdirected traceback investigations which supports procurement and staffing cases in agencies. In short, this study's operationalization agrees with earlier single-setting studies but extends them by offering a cross-context, quantified picture of what agencies can expect when shifting from schedule-based to risk-based processes (Moons et al., 2015; Noufaily et al., 2013).

A distinctive contribution of this study's review is to foreground deployment readiness interpretability, monitoring for drift, fairness appraisal, and resource disclosure as the bridge between promising prototypes and durable systems. Earlier critiques of digital epidemiology warned against brittle models tied to nonstationary proxies and called for ongoing monitoring and adaptive retraining (Lazer et al., 2014). this study's findings show that studies that documented drift monitoring and recalibration schedules reported more stable performance and easier portability, consistent with best practices in applied predictive modeling (Steyerberg et al., 2010). Likewise, interpretability methods such as SHAP and partial dependence repeatedly surfaced domain-sensible drivers (e.g., repeated critical violations, anomalous temperature–salinity combinations), reinforcing prior arguments that transparent explanations aid regulatory adoption and auditability (Lindgren et al., 2011). Fairness and subgroup performance are less developed in food safety than in clinical AI, but where analyzed, gaps in calibration and precision–recall were measurable and correctable with groupwise calibration or threshold adjustments, echoing broader machine-learning literature on equitable deployment (Lindgren et al., 2011). Finally, resource disclosures compute time, data refresh cadence help agencies scope MLOps budgets; this study's numbers align with prior reports that ensemble triage models train quickly on commodity hardware while sequence models require somewhat heavier but manageable compute (Chawla et al., 2002; Chen & Guestrin, 2016). The convergence across these strands suggests that governance and engineering discipline, rather than exotic algorithms, are the rate-limiters for impact.

In the genomics/WGS arena, this study's results complement and extend earlier demonstrations that standardized genotyping schemas (MLST/cgMLST), portable databases (BIGSdb), and international surveillance governance (PulseNet) create a substrate on which AI can deliver actionable inference cluster detection, lineage tracking, and source attribution (Pightling et al., 2018; Rha et al., 2013). We found that attribution accuracy and time-to-linkage improved most when genomic features were fused with facility, commodity, or geography metadata, a pattern anticipated by reviews arguing that metadata completeness and harmonization are as important as sequencing depth for public-health utility (Nogales et al., 2023; Noufaily et al., 2013). this study's caution that SNP-calling parameters and reference choice can materially alter cluster boundaries mirrors empirical sensitivity analyses showing downstream effects on phylogenies and epidemiologic inference (Pightling et al., 2018; Ribot et al., 2019). At the same time, newer machine-learning tools for partitioning large genome collections (e.g., PopPUNK) and visualizing core genome relationships (GrapeTree) simplify operational interpretation and facilitate rapid triage of emerging clusters (Gmeiner et al., 2024). Put simply, our pooled results support a pragmatic, hybrid view: WGS turns surveillance from counts into populations of evolving lineages; AI methods then map those lineages to sources and risk actions, provided that laboratories maintain standardized pipelines and agencies invest in metadata capture and integration (Allard et al., 2016).

Figure 11: Proposed Mothod for the future



In addition, this study synthesis clarifies where the marginal gains now lie for research and practice.

The field has already demonstrated that, under realistic validation and calibration, AI improves discrimination, timeliness, and operational yield. The binding constraints are coverage and comparability. Earlier calls to register protocols, pre-specify outcomes, and adopt common benchmarks (TRIPOD/PROBAST families) are directly applicable here (Brier, 1950; Campbell et al., 2025). Our numbers show that only about a quarter to a third of studies currently report external validation, calibration, and decision-utility analyses; this echoes critiques in clinical prediction fields and COVID-19 modeling, where transportability and bias concerns limited real-world uptake (Weller et al., 2021). Encouragingly, where studies did implement these practices, effect sizes persisted, and simple recalibration restored reliability across geographies mirroring evidence that intercept/slope updates are often sufficient for transport (Steyerberg et al., 2010). On the data side, multi-stream fusion repeatedly paid off, suggesting that partnerships to link inspections, WGS, environmental sensors, and complaints at routine cadence would yield immediate returns. Methodologically, attention-based temporal models and probabilistic forecasting deserve continued use where horizons and uncertainty matter, while ensembles remain first-line for triage over tabular data. Governance should prioritize interpretability, fairness checks, and drift monitoring as components of "default deployment." In short, the comparison with earlier work is not merely confirmatory; it indicates convergence on a toolkit that is both effective and practicable, together with a clear methodological playbook that if adhered to allows agencies to translate AI from promising pilots into sustained, auditable improvements in foodborne risk prevention (Vickers & Elkin, 2006; Zhang et al., 2021).

CONCLUSION

In sum, the evidence across 105 peer-reviewed studies demonstrates that artificial-intelligence approaches can materially strengthen public-health capacity to anticipate, target, and mitigate foodborne pathogen risk when they are matched to the right data regimes and embedded in disciplined workflows. Tree-based ensembles consistently deliver robust, interpretable gains for tabular, establishment-level risk scoring, while sequence- and attention-based deep learners excel for delay-aware nowcasting and multi-horizon forecasting; graph-structured and anomaly-first models add value where relational spillovers or rare signals dominate. The strongest practical improvements arise when models fuse complementary surveillance streams inspections, WGS, syndromic/ED telemetry, environmental drivers, complaint signals, and supply-chain metadata because fusion both lifts discrimination in the actionable top-k region and stabilizes alerts against transient noise. At the same time, the credibility and portability of these tools hinge on methodological rigor: temporal or geographic external validation, probability calibration linked to proper scoring rules, and decision-utility summaries at policy-relevant thresholds. Where such practices were present, gains persisted under realistic testing and translated into concrete operational benefits more critical violations found within fixed inspection budgets, earlier detection of emergent clusters at controlled false-alarm rates, and sizable error reductions for forecasts that guide advisories and resource allocation. Translation from promising prototypes to durable systems depends less on exotic algorithms than on governance: transparent reporting, reproducible pipelines, drift monitoring and scheduled recalibration, subgroup performance audits with corrective calibration, and clear resource disclosures that allow agencies to plan data refreshes and compute. Taken together, these findings indicate a mature, actionable pathway for deploying AI in food safety: select model families by task and cadence, invest in multi-stream data linkages, privilege externally validated and calibrated forecasts, and operationalize models within auditable, human-in-the-loop processes so that statistical gains become sustained public-health impact.

RECOMMENDATIONS

To translate these findings into durable public-health gains, agencies and research partners should prioritize a deployment playbook that is as much governance as it is modeling: (1) build multi-stream data pipes that refresh on routine cadence link inspections, WGS, syndromic/ED, environmental sensors, complaints, and supply-chain metadata and formalize data dictionaries and access controls; (2) choose model families by task and cadence tree ensembles for establishment triage and mixed tabular data, sequence/attention models for delay-aware nowcasting and multi-horizon forecasting, graph methods where relational links or genomic neighborhoods matter while maintaining a strong logistic/GLM baseline for auditability; (3) require temporal or geographic external validation before field use, followed by lightweight recalibration (intercept/slope) in each new jurisdiction; (4) mandate probability calibration checks, proper scoring (e.g., Brier), and decision-curve analyses at policy thresholds so leaders see net benefit in operational units (alerts per week, violations per 1,000 inspections); (5) integrate interpretability (e.g., feature attribution, partial dependence) into dashboards so inspectors understand *why* a premise or time window is flagged; (6) stand up MLOps for drift monitoring with explicit triggers for retraining (e.g., ≥1.5 SD shift in key features), versioned model cards, and change logs; (7) audit subgroup performance by neighborhood, establishment type, and operator size; apply group-wise calibration or threshold adjustments when gaps appear; (8) plan resources commodity hardware supports ensemble triage; modest GPUs cover sequence models; schedule data refreshes weekly for inspection/alerts and daily for nowcasting; (9) adopt privacy-preserving collaboration (federated learning or parameter sharing) for multi-region models when records cannot be centralized; (10) institutionalize reproducibility preregister protocols, publish code/configurations where permissible, and contribute to open benchmarks with shared splits and metrics. Finally, run prospective pilots with predefined success criteria (lead time, false-alarm budget, top-k precision) and a rollback plan; only after meeting these criteria should models gate inspections or advisories. This sequence converts statistical promise into sustained, auditable public-health impact.

REFERENCES

[1]. Achtman, M., Wain, J., Weill, F. X., Nair, S., Zhou, Z., Sangal, V., & Torpdahl, M. (2012). Multilocus sequence typing as a replacement for serotyping in Salmonella enterica. *PLOS ONE*, *7*(5), e35269. https://doi.org/10.1371/journal.pone.0035269

[2]. Adedire, O., Love, N. K., Hughes, H. E., Buchan, I., Vivancos, R., & Elliot, A. J. (2024). Early detection and monitoring of gastrointestinal infections using syndromic surveillance: A systematic review. *International Journal of Environmental Research and Public Health*, *21*(4), 489. https://doi.org/10.3390/ijerph21040489

[3]. Allard, M. W., Strain, E., Rand, H., & et al. (2016). Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *Journal of Clinical Microbiology*, *54*(8), 1975–1983. https://doi.org/10.1128/jcm.00081-16

[4]. Allard, S. M., Bell, R., Ferreira, C. M., Gonzalez-Escalona, N., Hoffmann, M., Muruvanda, T., & Brown, E. W. (2018). Practical value of using whole genome sequencing in food safety and public health. *Trends in Microbiology*, *26*(7), 531-554. https://doi.org/10.1016/j.tim.2018.01.003

[5]. Ashton, P. M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., & Peto, T. E. A. (2016). Prospective genomic surveillance of bacterial pathogens: Whole genome sequencing to detect and investigate outbreaks. *mBio*, *7*(3), e01623-01616. https://doi.org/10.1128/mBio.01623-16

[6]. Baker-Austin, C., Trinanes, J., Martínez-Urtaza, J., & et al. (2023). Climate warming and increasing Vibrio vulnificus infections in North Atlantic and Europe. *Scientific Reports*, *13*, 3893. https://doi.org/10.1038/s41598-023-28247-2

[7]. Baker, D., Sutton, N., Brown, L., & Morris, J. (2024). Foodborne illness complaint systems, detection, and restaurant inspection practices in the United States. *Foodborne Pathogens and Disease*, *21*(10), 561–570. https://doi.org/10.1089/fpd.2023.0086

[8]. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

[9]. Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1-3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

[10]. Brown, B., Allard, M., Bazaco, M. C., Blankenship, J., & Minor, T. (2021). An economic evaluation of the whole genome sequencing source tracking program in the U.S. *PLOS ONE*, *16*(10), e0258262. https://doi.org/10.1371/journal.pone.0258262

[11]. Campbell, A. M., Cabrera-Gumbau, J. M., Trinanes, J., Baker-Austin, C., & Martínez-Urtaza, J. (2025). Machine learning potential for identifying and forecasting complex environmental drivers of Vibrio vulnificus infections in the United States. *Environmental Health Perspectives*, *133*(1), 015593. https://doi.org/10.1289/ehp15593

[12]. Čapla, J., Zajác, P., Čurlej, J., Benešová, L., Jakabová, S., Fikselová, M., & Bobková, A. (2023). Analysis of data from the Rapid Alert System for Food and Feed for the country-of-origin Slovakia (2002–2020). *Heliyon*, *9*(1), e23146. https://doi.org/10.1016/j.heliyon.2023.e23146

[13]. Castelli, P., De Ruvo, A., Bucciacchio, A., D'Alterio, N., Cammà, C., Di Pasquale, A., & Radomski, N. (2023). Harmonization of supervised machine learning practices for efficient source attribution of Listeria monocytogenes based on genomic data. *BMC Genomics*, *24*, 560. https://doi.org/10.1186/s12864-023-09667-w

[14]. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

[15]. Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,

[16]. Chenar, S. S., & Deng, Z. (2018). Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. *Environment International*, *111*, 212-223. https://doi.org/10.1016/j.envint.2017.11.030

[17]. Chenar, S. S., & Deng, Z. (2021). Hybrid modeling and prediction of oyster norovirus outbreaks. *Journal of Water and Health*, *19*(2), 254-266. https://doi.org/10.2166/wh.2021.042

[18]. Chenar, S. S., & Deng, Z. Q. (2018). Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. *Environment International*, *111*, 212-223. https://doi.org/10.1016/j.envint.2017.11.032

[19]. Chenar, S. S., & Deng, Z. Q. (2021). Hybrid modeling and prediction of oyster norovirus outbreaks. *Journal of Water & Health*, *19*(2), 254-266. https://doi.org/10.2166/wh.2021.251

[20]. Choi, J., & Scharff, R. L. (2017). Effect of a publicly accessible disclosure system on food safety inspection scores in retail and food service establishments. *Journal of Food Protection*, *80*(7), 1188-1192. https://doi.org/10.4315/0362-028X.JFP-16-293

[21]. Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics*, *6*(1), 3–73. https://doi.org/10.2307/2348049

[22]. Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., & Group, T. A. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, *385*, e078378. https://doi.org/10.1136/bmj-2023-078378

[23]. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2016). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement—Explanation and elaboration for external validation ("what, why, how"). *BMC Medicine*, *14*, 40. https://doi.org/10.1186/s12916-016-0683-0

[24]. Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, *178*(9), 1505–1512. https://doi.org/10.1093/aje/kwt133

[25]. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/bf00994018

[26]. Danish, M., & Md. Zafor, I. (2022). The Role Of ETL (Extract-Transform-Load) Pipelines In Scalable Business Intelligence: A Comparative Study Of Data Integration Tools. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *2*(1), 89–121. https://doi.org/10.63125/1spa6877

[27]. Danish, M., & Md. Zafor, I. (2024). Power BI And Data Analytics In Financial Reporting: A Review Of Real-Time Dashboarding And Predictive Business Intelligence Tools. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 125-157. https://doi.org/10.63125/yg9zxt61

[28]. Danish, M., & Md.Kamrul, K. (2022). Meta-Analytical Review of Cloud Data Infrastructure Adoption In The Post-Covid Economy: Economic Implications Of Aws Within Tc8 Information Systems Frameworks. *American Journal of Interdisciplinary Studies*, *3*(02), 62-90. https://doi.org/10.63125/1eg7b369

[29]. Dipongkar Ray, S., Tamanna, R., Saiful Islam, A., & Shraboni, G. (2024). Gold Nanoparticle–Mediated Plasmonic Block Copolymers: Design, Synthesis, And Applications in Smart Drug Delivery. *American Journal of Scholarly Research and Innovation*, *3*(02), 80-98. https://doi.org/10.63125/pgk8tt08

[30]. Effland, T., DeFelice, N., & Gronvall, G. (2018). Discovering foodborne illness in online restaurant reviews. *Journal of the American Medical Informatics Association*, *25*(12), 1586-1592. https://doi.org/10.1093/jamia/ocx093

[31]. Effland, T., Lawson, A., Balter, S., Devinney, K., Reddy, V., Waechter, H., Gravano, L., & Hsu, D. (2018). Discovering foodborne illness in online restaurant reviews. *Journal of the American Medical Informatics Association*, *25*(12), 1586-1592. https://doi.org/10.1093/jamia/ocx093

[32]. Ekanayake, E. M. T. K., & et al. (2022). Modeling pH and temperature effects as climatic hazards in Vibrio exposure risk. *GeoHealth*, *6*(10), e2022GH000769. https://doi.org/10.1029/2022gh000769

[33]. Fleetwood, J., Rahman, S., Holland, D., Millson, D., Thomson, L., & Poppy, G. (2019). As clean as they look? Food hygiene inspection scores, microbiological contamination, and foodborne illness. *Food Control*, *96*, 76-86. https://doi.org/10.1016/j.foodcont.2018.08.034

[34]. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[35]. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, *46*(4), 44. https://doi.org/10.1145/2523813

[36]. Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

[37]. Gmeiner, A., Ivanova, M., Kamau Njage, P. M., Hansen, L. T., Chindelevitch, L., & Leekitcharoenphon, P. (2025). Quantitative prediction of disinfectant tolerance in Listeria monocytogenes using whole-genome sequencing and machine learning. *Scientific Reports*, *15*, 94321. https://doi.org/10.1038/s41598-025-94321-6

[38]. Gmeiner, A., Njage, P. M. K., Hansen, L. T., Aarestrup, F. M., & Leekitcharoenphon, P. (2024). Predicting Listeria monocytogenes virulence potential using whole genome sequencing and machine learning. *International Journal of Food Microbiology*, *404*, 110491. https://doi.org/10.1016/j.ijfoodmicro.2023.110491

[39]. Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

[40]. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). *On calibration of modern neural networks* Proceedings of the 34th International Conference on Machine Learning,

[41]. Harris, K. J., Hanks, L., Line, N. D., & McGinley, S. (2017). Understanding responses to posted restaurant food safety scores: An information processing and regulatory focus perspective. *International Journal of Hospitality Management*, *63*, 24-35. https://doi.org/10.1016/j.ijhm.2016.09.002

[42]. Harrison, C., & et al. (2014). Discovering foodborne illness in online restaurant reviews, New York City, 2012–2013. *BMJ*, *348*, g3560. https://doi.org/10.1136/bmj.g3560

[43]. Havelaar, A. H., & et al. (2015). World Health Organization global estimates and regional comparisons of the burden of foodborne disease in 2010. *PLOS Medicine*, *12*(12), e1001923. https://doi.org/10.1371/journal.pmed.1001923

[44]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[45]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2023). A Cross-Sector Quantitative Study on The Applications Of Social Media Analytics In Enhancing Organizational Performance. *American Journal of Scholarly Research and Innovation*, *2*(02), 274-302. https://doi.org/10.63125/d8ree044

[46]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2024). Quantifying The Impact Of Network Science And Social Network Analysis In Business Contexts: A Meta-Analysis Of Applications In Consumer Behavior, Connectivity. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 58-89. https://doi.org/10.63125/vgkwe938

[47]. Jackson, B. R., Tarr, C., Strain, E., Jackson, K. A., Conrad, A., Carleton, H., & Mody, R. K. (2016). Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. *Clinical Infectious Diseases*, *63*(3), 380-386. https://doi.org/10.1093/cid/ciw242

[48]. Jahid, M. K. A. S. R. (2022). Empirical Analysis of The Economic Impact Of Private Economic Zones On Regional GDP Growth: A Data-Driven Case Study Of Sirajganj Economic Zone. *American Journal of Scholarly Research and Innovation*, *1*(02), 01-29. https://doi.org/10.63125/je9w1c40

[49]. Jolley, K. A., & Maiden, M. C. J. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *Bioinformatics*, *26*(3), 419-420. https://doi.org/10.1093/bioinformatics/btq559

[50]. Kang, J., Kuznetsova, P., Luca, M., & Choi, Y. (2013). Where not to eat? Improving public health by predicting restaurant inspections using online reviews. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,

[51]. Kaskela, J., Sund, R., & Lundén, J. (2021). Efficacy of disclosed food safety inspections in restaurants. *Food Control*, *123*, 107775. https://doi.org/10.1016/j.foodcont.2020.107775

[52]. Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*(500), 1590–1598. https://doi.org/10.1080/01621459.2012.737745

[53]. Kim, T. N., Wildey, L., Gleason, B., Bleser, J., Firestone, M. J., Bare, G., Bliss, J., Dewey-Mattia, D., Stueven, H., Brown, L., Dyjack, D., & Hedberg, C. W. (2022). Foodborne outbreak rates associated with restaurant inspection grading and posting at the point of service: Evaluation using national foodborne outbreak surveillance data. *Journal of Food Protection*, *85*(7), 1000-1007. https://doi.org/10.4315/JFP-22-007

[54]. Kirk, M. D., Pires, S. M., Black, R. E., Caipo, M., Crump, J. A., Devleesschauwer, B., & et al. (2015). World Health Organization estimates of the global and regional disease burden of 22 foodborne diseases, 2010. *PLOS Medicine*, *12*(12), e1001921. https://doi.org/10.1371/journal.pmed.1001921

[55]. Kirkman, S., Hollingsworth, B., Lake, A., von Hinke, S., Sorrell, S., Smith, K., & Brown, H. (2021). Field validity and spatial accuracy of Food Standards Agency Food Hygiene Rating Scheme data. *Journal of Public Health*, *43*(4), e720–e728. https://doi.org/10.1093/pubmed/fdaa172

[56]. Kirkman, S., Hollingsworth, B., Lake, A. A., von Hinke, S., Sorrell, S., Adamson, A. J., & Brown, H. (2021). Field validity and spatial accuracy of Food Standards Agency Food Hygiene Rating Scheme data. *Journal of Public Health*, *43*(4), e720-e728. https://doi.org/10.1093/pubmed/fdaa172

[57]. Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Biometrics*, *57*(3), 531–538. https://doi.org/10.1111/j.0006-341X.2001.00531.x

[58]. Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R., & Mostashari, F. (2005). A space–time permutation scan statistic for disease outbreak detection. *PLOS Medicine*, *2*(3), e59. https://doi.org/10.1371/journal.pmed.0020059

[59]. Lahti, E., Rönnqvist, M., & Rönnqvist, S. (2022). Results of routine inspections in restaurants and institutional catering associated with foodborne outbreaks in Finland. *International Journal of Environmental Health Research*, *32*(6), 1160-1175. https://doi.org/10.1080/09603123.2022.2041563

[60]. Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

[61]. Lees, J. A., Harris, S. R., Tonkin-Hill, G., Gladstone, R. A., Lo, S. W., Weiser, J. N., & Bentley, S. D. (2019). Fast and flexible bacterial genomic epidemiology with PopPUNK. *Microbial Genomics*, *5*(2), e000262. https://doi.org/10.1099/mgen.0.000262

[62]. Li, H., & et al. (2017). Prospective whole-genome sequencing enhances national surveillance of Listeria monocytogenes. *Journal of Clinical Microbiology*, *55*(9), e02344-02315. https://doi.org/10.1128/jcm.02344-15

[63]. Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

[64]. Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). *Focal loss for dense object detection* Proceedings of the IEEE International Conference on Computer Vision,

[65]. Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(4), 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

[66]. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). *Isolation forest* 2008 Eighth IEEE International Conference on Data Mining,

[67]. Luca, M. (2020). Grade pending: The effect of the New York City restaurant sanitary inspection disclosure policy on consumer behavior. *Journal of Public Health: From Theory to Practice*, *28*(5), 625-637. https://doi.org/10.1007/s10389-020-01384-y

[68]. Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions* Advances in Neural Information Processing Systems,

[69]. Lupolova, N., Dallman, T. J., Holden, N. J., & Gally, D. L. (2017). Patchy promiscuity: Machine learning applied to predict the host specificity of Salmonella enterica and Escherichia coli. *Microbial Genomics*, *3*(10), e000135. https://doi.org/10.1099/mgen.0.000135

[70]. Lupolova, N., Dallman, T. J., Holden, N. J., & Gally, D. L. (2019). Engineering bacteroides: the power of bacterial community genetics for microbiota manipulation [review of ML in host specificity/source attribution]. *Microbial Genomics*, *5*(12), e000317. https://doi.org/10.1099/mgen.0.000317

[71]. Maiden, M. C. J., Jansen van Rensburg, M. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., & McCarthy, N. D. (2013). MLST revisited: The gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology*, *11*(10), 728-736. https://doi.org/10.1038/nrmicro3093

[72]. Md Arifur, R., & Sheratun Noor, J. (2022). A Systematic Literature Review of User-Centric Design In Digital Business Systems: Enhancing Accessibility, Adoption, And Organizational Impact. *Review of Applied Science and Technology*, *1*(04), 01-25. https://doi.org/10.63125/ndjkpm77

[73]. Md Ashiqur, R., Md Hasan, Z., & Afrin Binta, H. (2025). A meta-analysis of ERP and CRM integration tools in business process optimization. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 278-312. https://doi.org/10.63125/yah70173

[74]. Md Hasan, Z. (2025). AI-Driven business analytics for financial forecasting: a systematic review of decision support models in SMES. *Review of Applied Science and Technology*, *4*(02), 86-117. https://doi.org/10.63125/gjrpv442

[75]. Md Hasan, Z., Mohammad, M., & Md Nur Hasan, M. (2024). Business Intelligence Systems In Finance And Accounting: A Review Of Real-Time Dashboarding Using Power BI & Tableau. *American Journal of Scholarly Research and Innovation*, *3*(02), 52-79. https://doi.org/10.63125/fy4w7w04

[76]. Md Hasan, Z., & Moin Uddin, M. (2022). Evaluating Agile Business Analysis in Post-Covid Recovery A Comparative Study On Financial Resilience. *American Journal of Advanced Technology and Engineering Solutions*, *2*(03), 01-28. https://doi.org/10.63125/6nee1m28

[77]. Md Hasan, Z., Sheratun Noor, J., & Md. Zafor, I. (2023). Strategic role of business analysts in digital transformation tools, roles, and enterprise outcomes. *American Journal of Scholarly Research and Innovation*, *2*(02), 246-273. https://doi.org/10.63125/rc45z918

[78]. Md Ismail, H., Md Mahfuj, H., Mohammad Aman Ullah, S., & Shofiul Azam, T. (2025). Implementing Advanced Technologies For Enhanced Construction Site Safety. *American Journal of Advanced Technology and Engineering Solutions*, *1*(02), 01-31. https://doi.org/10.63125/3v8rpr04

[79]. Md Ismail Hossain, M. A. B., amp, & Mousumi Akter, S. (2023). Water Quality Modelling and Assessment Of The Buriganga River Using Qual2k. *Global Mainstream Journal of Innovation, Engineering & Emerging Technology*, *2*(03), 01-11. https://doi.org/10.62304/jieet.v2i03.64

[80]. Md Jakaria, T., Md, A., Zayadul, H., & Emdadul, H. (2025). Advances In High-Efficiency Solar Photovoltaic Materials: A Comprehensive Review Of Perovskite And Tandem Cell Technologies. *American Journal of Advanced Technology and Engineering Solutions*, *1*(01), 201-225. https://doi.org/10.63125/5amnvb37

[81]. Md Mahamudur Rahaman, S. (2022a). Electrical And Mechanical Troubleshooting in Medical And Diagnostic Device Manufacturing: A Systematic Review Of Industry Safety And Performance Protocols. *American Journal of Scholarly Research and Innovation*, *1*(01), 295-318. https://doi.org/10.63125/d68y3590

[82]. Md Mahamudur Rahaman, S. (2022b). Smart Maintenance in Medical Imaging Manufacturing: Towards Industry 4.0 Compliance at Chronos Imaging. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *2*(1), 29–62. https://doi.org/10.63125/eatsmf47

[83]. Md Mahamudur Rahaman, S. (2024). AI-Driven Predictive Maintenance For High-Voltage X-Ray Ct Tubes: A Manufacturing Perspective. *Review of Applied Science and Technology*, *3*(01), 40-67. https://doi.org/10.63125/npwqxp02

[84]. Md Mahamudur Rahaman, S., & Rezwanul Ashraf, R. (2022). Integration of PLC And Smart Diagnostics in Predictive Maintenance of CT Tube Manufacturing Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 62-96. https://doi.org/10.63125/gspb0f75

[85]. Md Mahamudur Rahaman, S., & Rezwanul Ashraf, R. (2023). Applying Lean And Six Sigma In The Maintenance Of Medical Imaging Equipment Manufacturing Lines. *Review of Applied Science and Technology*, *2*(04), 25-53. https://doi.org/10.63125/6varjp35

[86]. Md Nazrul Islam, K. (2022). A Systematic Review of Legal Technology Adoption In Contract Management, Data Governance, And Compliance Monitoring. *American Journal of Interdisciplinary Studies*, *3*(01), 01-30. https://doi.org/10.63125/caangg06

[87]. Md Nur Hasan, M. (2024). Integration Of Artificial Intelligence And DevOps In Scalable And Agile Product Development: A Systematic Literature Review On Frameworks. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *4*(1), 01–32. https://doi.org/10.63125/exyqj773

[88]. Md Nur Hasan, M. (2025). Role Of AI And Data Science In Data-Driven Decision Making For It Business Intelligence: A Systematic Literature Review. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 564-588. https://doi.org/10.63125/n1xpym21

[89]. Md Nur Hasan, M., Md Musfiqur, R., & Debashish, G. (2022). Strategic Decision-Making in Digital Retail Supply Chains: Harnessing AI-Driven Business Intelligence From Customer Data. *Review of Applied Science and Technology*, *1*(03), 01-31. https://doi.org/10.63125/6a7rpy62

[90]. Md Redwanul, I., & Md. Zafor, I. (2022). Impact of Predictive Data Modeling on Business Decision-Making: A Review Of Studies Across Retail, Finance, And Logistics. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 33-62. https://doi.org/10.63125/8hfbkt70

[91]. Md Rezaul, K., & Md Mesbaul, H. (2022). Innovative Textile Recycling and Upcycling Technologies For Circular Fashion: Reducing Landfill Waste And Enhancing Environmental Sustainability. *American Journal of Interdisciplinary Studies*, *3*(03), 01-35. https://doi.org/10.63125/kkmerg16

[92]. Md Sultan, M., Proches Nolasco, M., & Md. Torikul, I. (2023). Multi-Material Additive Manufacturing For Integrated Electromechanical Systems. *American Journal of Interdisciplinary Studies*, *4*(04), 52-79. https://doi.org/10.63125/y2ybrx17

[93]. Md Takbir Hossen, S., Ishtiaque, A., & Md Atiqur, R. (2023). AI-Based Smart Textile Wearables For Remote Health Surveillance And Critical Emergency Alerts: A Systematic Literature Review. *American Journal of Scholarly Research and Innovation*, *2*(02), 1-29. https://doi.org/10.63125/ceqapd08

[94]. Md Tawfiqul, I. (2023). A Quantitative Assessment Of Secure Neural Network Architectures For Fault Detection In Industrial Control Systems. *Review of Applied Science and Technology*, *2*(04), 01-24. https://doi.org/10.63125/3m7gbs97

[95]. Md. Sakib Hasan, H. (2022). Quantitative Risk Assessment of Rail Infrastructure Projects Using Monte Carlo Simulation And Fuzzy Logic. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 55-87. https://doi.org/10.63125/h24n6z92

[96]. Md. Tarek, H. (2022). Graph Neural Network Models For Detecting Fraudulent Insurance Claims In Healthcare Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 88-109. https://doi.org/10.63125/r5vsmv21

[97]. Md. Zafor, I. (2025). A Meta-Analysis Of AI-Driven Business Analytics: Enhancing Strategic Decision-Making In SMEs. *Review of Applied Science and Technology*, *4*(02), 33-58. https://doi.org/10.63125/wk9fqv56

[98]. Md.Kamrul, K., & Md Omar, F. (2022). Machine Learning-Enhanced Statistical Inference For Cyberattack Detection On Network Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 65-90. https://doi.org/10.63125/sw7jzx60

[99]. Md.Kamrul, K., & Md. Tarek, H. (2022). A Poisson Regression Approach to Modeling Traffic Accident Frequency in Urban Areas. *American Journal of Interdisciplinary Studies*, *3*(04), 117-156. https://doi.org/10.63125/wqh7pd07

[100]. Mikhailov, A., Pust, R., & Sheppard, S. K. (2023). Rapid geographical source attribution of Salmonella enterica serovar Enteritidis using hierarchical machine learning. *eLife*, *12*, e84167. https://doi.org/10.7554/eLife.84167

[101]. Moin Uddin, M. (2025). Impact Of Lean Six Sigma On Manufacturing Efficiency Using A Digital Twin-Based Performance Evaluation Framework. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 343-375. https://doi.org/10.63125/z70nhf26

[102]. Moin Uddin, M., & Rezwanul Ashraf, R. (2023). Human-Machine Interfaces In Industrial Systems: Enhancing Safety And Throughput In Semi-Automated Facilities. *American Journal of Interdisciplinary Studies*, *4*(01), 01-26. https://doi.org/10.63125/s2qa0125

[103]. Momena, A., & Md Nur Hasan, M. (2023). Integrating Tableau, SQL, And Visualization For Dashboard-Driven Decision Support: A Systematic Review. *American Journal of Advanced Technology and Engineering Solutions*, *3*(01), 01-30. https://doi.org/10.63125/4aa43m68

[104]. Moons, K. G. M., Altman, D. G., Reitsma, J. B., Ioannidis, J. P. A., Macaskill, P., Steyerberg, E. W., Vickers, A. J., Ransohoff, D. F., & Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and elaboration. *BMJ, 350,* g7594. https://doi.org/10.1136/bmj.g7594

[105]. Mubashir, I., & Abdul, R. (2022). Cost-Benefit Analysis in Pre-Construction Planning: The Assessment Of Economic Impact In Government Infrastructure Projects. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 91-122. https://doi.org/10.63125/kjwd5e33

[106]. Munck, P. C., Modvig, H. H., Josefsen, M. H., & et al. (2020). Application of whole-genome sequences and machine learning in source attribution of Salmonella Typhimurium. *Risk Analysis*, *40*(10), 1951-1963. https://doi.org/10.1111/risa.13510

[107]. Nadon, C., Van Walle, I., Gerner-Smidt, P., Campos, J., Chinen, I., Concepción-Acevedo, J., & Tarr, C. (2017). PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Eurosurveillance*, *22*(23), 30494. https://doi.org/10.2807/1560-7917.ES.2017.22.23.30494

[108]. Nogales, A., Mora-Cantallops, M., Díaz-Morón, R., & García-Tejedor, Á. J. (2023). Network analysis for food safety: Quantitative and structural study of data gathered through the RASFF system in the European Union. *Food Control*, *145*, 109422. https://doi.org/10.1016/j.foodcont.2022.109422

[109]. Noufaily, A., Enki, D. G., Farrington, C. P., Garthwaite, P., Andrews, N., & Charlett, A. (2013). An improved algorithm for outbreak detection in multiple surveillance systems. *Biostatistics*, *14*(2), 340–354. https://doi.org/10.1093/biostatistics/kxt009

[110]. Oldroyd, R. A., Morris, M. A., & Birkin, M. (2021). Predicting food safety compliance for informed food outlet inspections: A machine learning approach. *International Journal of Environmental Research and Public Health*, *18*(23), 12635. https://doi.org/10.3390/ijerph182312635

[111]. Omar Muhammad, F., & Md.Kamrul, K. (2022). Blockchain-Enabled BI For HR And Payroll Systems: Securing Sensitive Workforce Data. *American Journal of Scholarly Research and Innovation*, *1*(02), 30-58. https://doi.org/10.63125/et4bhy15

[112]. Panacek, T., & et al. (2023). Environmental factors influencing occurrence of Vibrio parahaemolyticus and V. vulnificus in coastal waters. *Applied and Environmental Microbiology*, *89*(24), e00307-00323. https://doi.org/10.1128/aem.00307-23

[113]. Pightling, A. W., Petronella, N., & Pagotto, F. (2018). Choice of reference genome and SNP caller for bacterial outbreak surveillance affects phylogenetic reconstruction. *PLOS ONE*, *13*(3), e0194723. https://doi.org/10.1371/journal.pone.0194723

[114]. Public Health, R. (2017). Inspection frequency, sociodemographic factors, and food safety outcomes in chain and nonchain restaurants. *Public Health Reports*, *132*(2), 180-189. https://doi.org/10.1177/0033354916687741

[115]. Reduanul, H., & Mohammad Shoeb, A. (2022). Advancing AI in Marketing Through Cross Border Integration Ethical Considerations And Policy Implications. *American Journal of Scholarly Research and Innovation*, *1*(01), 351-379. https://doi.org/10.63125/d1xg3784

[116]. Rha, B., Burrer, S., Park, S., Trivedi, T., Parashar, U. D., & Lopman, B. A. (2013). Emergency department visit data for rapid detection and monitoring of norovirus activity, United States. *Emerging Infectious Diseases*, *19*(8), 1214–1221. https://doi.org/10.3201/eid1908.130483

[117]. Ribot, E. M., Hise, K. B., & Tauxe, R. V. (2019). PulseNet and the changing paradigm of laboratory-based surveillance for foodborne diseases. *Public Health Reports*, *134*(Suppl. 2), 22S–28S. https://doi.org/10.1177/0033354919881650

[118]. Riley, R. D., Ensor, J., Snell, K. I. E., Debray, T. P. A., Altman, D. G., Moons, K. G. M., Collins, G. S., & van Smeden, M. (2020). Calculating the sample size required for developing a clinical prediction model. *BMJ*, *368*, m441. https://doi.org/10.1136/bmj.m441

[119]. Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., & Wintle, B. A. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, and phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

[120]. Ronholm, J., Nasheri, N., Petronella, N., & Pagotto, F. (2016). Navigating microbiological food safety in the era of whole-genome sequencing. *Frontiers in Microbiology*, *7*, 1209. https://doi.org/10.3389/fmicb.2016.01209

[121]. Rose, J. R., & et al. (2025). Attribution of Salmonella enterica to food sources using whole-genome sequencing and machine learning. *Emerging Infectious Diseases*, *31*(4). https://doi.org/10.3201/eid3104.241172

[122]. Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using INLA. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

[123]. Sabuj Kumar, S., & Zobayer, E. (2022). Comparative Analysis of Petroleum Infrastructure Projects In South Asia And The Us Using Advanced Gas Turbine Engine Technologies For Cross Integration. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 123-147. https://doi.org/10.63125/wr93s247

[124]. Sadia, T., & Shaiful, M. (2022). In Silico Evaluation of Phytochemicals From Mangifera Indica Against Type 2 Diabetes Targets: A Molecular Docking And Admet Study. *American Journal of Interdisciplinary Studies*, *3*(04), 91-116. https://doi.org/10.63125/anaf6b94

[125]. Saito, T., & Rehmsmeier, M. (2015). The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

[126]. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181–1191. https://doi.org/10.1016/j.ijforecast.2019.07.001

[127]. Sanjai, V., Sanath Kumar, C., Maniruzzaman, B., & Farhana Zaman, R. (2023). Integrating Artificial Intelligence in Strategic Business Decision-Making: A Systematic Review Of Predictive Models. *International Journal of Scientific Interdisciplinary Research*, *4*(1), 01-26. https://doi.org/10.63125/s5skge53

[128]. Sanjai, V., Sanath Kumar, C., Sadia, Z., & Rony, S. (2025). AI And Quantum Computing For Carbon-Neutral Supply Chains: A Systematic Review Of Innovations. *American Journal of Interdisciplinary Studies*, *6*(1), 40-75. https://doi.org/10.63125/nrdx7d32

[129]. Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M.-A., Roy, S. L., Jones, J. L., & Griffin, P. M. (2011). Foodborne illness acquired in the United States—Major pathogens. *Emerging Infectious Diseases*, *17*(1), 7-15. https://doi.org/10.3201/eid1701.P11101

[130]. Sheratun Noor, J., & Momena, A. (2022). Assessment Of Data-Driven Vendor Performance Evaluation in Retail Supply Chains: Analyzing Metrics, Scorecards, And Contract Management Tools. *American Journal of Interdisciplinary Studies*, *3*(02), 36-61. https://doi.org/10.63125/0s7t1y90

[131]. Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology*, *21*(1), 128-138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

[132]. Tahmina Akter, R., Debashish, G., Md Soyeb, R., & Abdullah Al, M. (2023). A Systematic Review of AI-Enhanced Decision Support Tools in Information Systems: Strategic Applications In Service-Oriented Enterprises And Enterprise Planning. *Review of Applied Science and Technology*, *2*(01), 26-52. https://doi.org/10.63125/73djw422

[133]. Tamanna, R., & Dipongkar Ray, S. (2023). Comprehensive Insights Into Co_2 Capture: Technological Progress And Challenges. *Review of Applied Science and Technology*, *2*(01), 113-141. https://doi.org/10.63125/9p690n14

[134]. Tango, T., & Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*, *4*, 11. https://doi.org/10.1186/1476-072x-4-11

[135]. Tax, D. M. J., & Duin, R. P. W. (2004). Support vector data description. *Machine Learning*, *54*(1), 45–66. https://doi.org/10.1023/b:Mach.0000008084.60811.49

[136]. Tolar, B., Joseph, L. A., & Schroeder, M. N. (2019). An overview of PulseNet USA databases. *Foodborne Pathogens and Disease*, *16*(7), 457–462. https://doi.org/10.1089/fpd.2019.2637

[137]. Turner, J. W., & et al. (2024). Virulent properties and genomic diversity of Vibrio vulnificus isolated from environmental sources. *Microbiology Spectrum*, *12*(1), e00079-00024. https://doi.org/10.1128/spectrum.00079-24

[138]. Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, *10*(2), 225–232. https://doi.org/10.1111/ecog.04626

[139]. van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

[140]. Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, *26*(6), 565-574. https://doi.org/10.1177/0272989X06295361

[141]. Wang, H., Cui, W., Guo, Y., Du, Y., & Zhou, Y. (2021). Machine learning prediction of foodborne disease pathogens: Algorithm development and validation study. *JMIR Medical Informatics*, *9*(1), e24924. https://doi.org/10.2196/24924

[142]. Weller, D. L., Love, T., & Wiedmann, M. (2021). Interpretability versus accuracy: Comparing ML models to predict E. coli levels in agricultural water. *Frontiers in Artificial Intelligence*, *4*, 19. https://doi.org/10.3389/frai.2021.630451

[143]. Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, *170*(1), 51-58. https://doi.org/10.7326/M18-1376

[144]. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(1), 4–24. https://doi.org/10.1109/tnnls.2020.2978386

[145]. Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., Bonten, M. M. J., Dahly, D. L., Damen, J. A. A., Debray, T. P. A., & van Smeden, M. (2020). Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. *BMJ*, *369*, m1328. https://doi.org/10.1136/bmj.m1328

[146]. Zadrozny, B., & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,

[147]. Zhang, P., Cui, W., Wang, H., Du, Y., & Zhou, Y. (2021). High-efficiency machine learning method for identifying foodborne disease outbreaks and confounding factors. *Foodborne Pathogens and Disease*, *18*(8), 590-598. https://doi.org/10.1089/fpd.2020.2913

[148]. Zhou, Z., Alikhan, N. F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., & Achtman, M. (2018). GrapeTree: Visualization of core genomic relationships among 100,000 bacterial pathogens. *Microbial Genomics*, *4*(6), e000204. https://doi.org/10.1099/mgen.0.000204