Received: 23 April 2025 | Revised: 15 December 2025 | Accepted: 3 March 2026 | Published online: 27 March 2026

REVIEW

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–29
DOI: 10.47852/bonviewAIA62025993

A Comprehensive Literature Review and Bibliometric Analysis of the Development and Difficulties of Artificial Intelligence Methods for International Roughness Index Prediction

Lendra Lendra[1,*], Mochamad Agung Wibowo[2,3], Jati Utomo Dwi Hatmoko[3] and Rony Teguh[4]

[1] *Department of Civil Engineering, Palangka Raya University, Indonesia*

[2] *Doctoral Program of Information System, Diponegoro University, Indonesia*

[3] *Department of Civil Engineering, Diponegoro University, Indonesia*

[4] *Department of Informatics Engineering, Palangka Raya University, Indonesia*

Abstract: The International Roughness Index (IRI) is a key indicator for pavement condition assessment and maintenance prioritization. In recent years, artificial intelligence (AI) techniques have been increasingly applied to improve IRI prediction accuracy; however, existing studies remain fragmented, limiting a comprehensive understanding of model performance, data sources, and practical implementation challenges. This study presents a systematic literature review and bibliometric analysis to synthesize recent developments in AI-based IRI prediction. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, 49 peer-reviewed articles published between 2008 and 2025 were analyzed using VOSviewer. The review addresses four main aspects: prediction model performance, data source effectiveness, implementation challenges, and research collaboration patterns. The results show that both machine learning methods (e.g., Random Forest, XGBoost, Gradient Boosting) and deep learning approaches (e.g., neural networks and Long Short-Term Memory) achieve high predictive accuracy, with R^2 values frequently exceeding 0.90. Machine learning models offer advantages in interpretability and computational efficiency, while deep learning models perform better with large and complex datasets. Bibliometric analysis identifies six major research clusters and a clear temporal evolution from early modeling studies to advanced AI applications. Despite technological progress, challenges remain, including data heterogeneity, limited interpretability, computational demands, and integration with pavement management systems. This study provides an evidence-based overview of current trends and identifies key research directions to support the practical and sustainable adoption of AI-driven IRI prediction in pavement construction and management.

Keywords: artificial intelligence, IRI prediction, pavement management, systematic literature review

1. Introduction

The World Bank developed the International Roughness Index (IRI) as a standard for measuring pavement roughness [1, 2]. It is a key metric for assessing pavement condition, directly affecting ride comfort and a key factor in deciding which maintenance tasks to prioritize [3, 4]. Accurate IRI predictions are important for keeping roads safe, making the best use of resources, and implementing effective pavement management systems [5, 6]. The rise of artificial intelligence (AI) has changed the way we do predictive modeling, making IRI forecasting much more accurate than traditional empirical methods. Artificial neural networks (ANNs) and support vector machines (SVMs), two types of machine learning (ML) and deep learning (DL), have been very good at finding the complicated, nonlinear links between pavement characteristics and the growth of surface roughness [7–10].

Combining optimization methods with ML frameworks has made predictions much more accurate [11]. Hybrid models that combine Adaptive Network-Based Fuzzy Inference Systems (ANFIS), particle swarm optimization (PSO), and genetic algorithms (GA) have consistently outperformed traditional ML methods [12]. Adding domain-specific engineering knowledge to AI architectures has also shown significant promise. For example, physics-informed neural networks have made roughness predictions more reliable [13]. Researchers have determined exactly how factors such as traffic volume and pavement age affect IRI

*Corresponding author: Lendra, Department of Civil Engineering, Palangka Raya University, Indonesia. Email: lendraleman@jts.upr.ac.id

predictions using explainable AI methods, such as Random Forests and XGBoost [14].

These developments underscore the importance of understanding model decision-making processes in practical pavement management applications. Building good predictive models requires more than just intelligent algorithms. It also requires good data and good preprocessing methods. Recent studies have shown that carefully optimizing hyperparameters and using large datasets can significantly improve a model's performance [15, 16]. Long-Term Pavement Performance (LTPP) datasets have been invaluable for training ML models to predict IRI accurately [17]. Ensemble techniques such as AdaBoost have shown better results than standard regression methods, especially for flexible pavements, where changes in the surface can significantly affect IRI measurements [18, 19].

The World Bank developed the IRI in the 1980s [3]. It set a standard, repeatable way to measure how uneven a pavement surface is, and it has since become the global standard for measuring roughness. The use of AI in pavement condition assessment started in the late 1990s. At that time, researchers began using neural networks to predict how well pavement would hold up [20–22]. Early uses mainly were of general pavement management systems, which demonstrated the usefulness of data-driven methods. These early studies showed that AI methods could be used to evaluate infrastructure, leading to the development of modern applications that focus on specific metrics, such as IRI.

As research in this domain expands, systematic literature reviews (SLRs) have begun delineating AI applications in IRI prediction, underscoring the need for ongoing innovation and the enhancement of existing models [23]. There is a lot of research in this area, but there is still no comprehensive view of how research trends, collaboration patterns, and knowledge evolve in AI-driven IRI prediction. To fill this gap and contribute to the ongoing research conversation [24], this paper uses VOSviewer to conduct a detailed bibliometric analysis and an SLR that follows Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. This dual approach allows for both a qualitative evaluation of recent advancements in AI-driven IRI prediction and a quantitative examination of research trends, citation networks, and emerging patterns. The review examines new AI-driven IRI prediction methods, rates the performance of different ML and DL techniques, assesses how different data sources affect results, and identifies the main challenges in setting up real-time pavement monitoring systems.

The goal of this study is to answer the following research question: Research Question 1: What does bibliometric analysis show about current trends in AI-driven IRI prediction research, how researchers work together, and how knowledge grows? How can these ideas help shape future research? Research Question 2: How useful, accurate, and trustworthy are ML and DL methods for predicting the IRI? Research Question 3: How do new data sources, such as historical databases, smartphone sensors, and crowdsourced data, make AI-based IRI prediction models better? Research Question 4: What problems still need to be solved before AI-driven IRI prediction models can be used for real-time road condition monitoring and infrastructure management? What more needs to be done? This paper answers these questions by providing a comprehensive review of current methods, highlighting where existing models fall short, and offering practical suggestions for improving AI-driven pavement condition assessment. Using both systematic review and bibliometric analysis gives us a new way of looking at things that combines qualitative insights with quantitative research trend analysis. This helps us get a better idea of where the field is now and where it could go in the future.

The paper is structured as follows: In Section 2, we discuss the dual-approach method, which combines an SLR with bibliometric analysis. In Section 3, we discuss the results that answer our four research questions: research trends, model performance, data sources, and the challenges of putting the research into practice. Section 4 discusses what the results mean and what they do not. Section 5 ends with a list of important contributions and ideas for the future.

2. Research Methodology

This paper employs a dual-approach methodology combining an SLR with bibliometric analysis to comprehensively examine AI-based IRI prediction research as seen in Figure 1. The approach is based on pre-established research questions (RQ1, RQ2, RQ3, and RQ4) and follows the PRISMA statement [24, 25], which provides an evidence-based minimum set of items for reporting in systematic reviews, ensuring transparency and reproducibility. This combined methodology enables both qualitative assessment of the current state of AI-based IRI prediction research and quantitative analysis of research trends, collaboration patterns, and knowledge evolution in the field.

2.1. Systematic literature review methodology

The SLR component has three parts: planning, doing, and putting things together. During the planning phase, the steps for searching, the rules for including or excluding data, and the methods for extracting data are all set. During the execution phase, researchers review the literature, screen studies, and extract valuable data. Lastly, during the synthesis phase, the results are examined to determine whether they answer the research questions (RQ1, RQ2, RQ3, and RQ4).

2.2. Search strategy

A systematic search method was used to ensure that all the literature was covered. The search utilized the Scopus and IEEE Xplore databases [26, 27], which are known for their comprehensive coverage of peer-reviewed research in AI, transportation infrastructure, and civil engineering. The main search terms were "International Roughness Index" plus "machine learning," "deep learning," or "AI." Other terms included "pavement management," "LTPP database," "smartphone sensors," and "road roughness prediction." The search only included English-language articles published between 2008 and 2025.

2.3. Study selection criteria

Specific inclusion and exclusion criteria were developed to guarantee the selection of relevant and high-quality studies. The inclusion criteria focused on journal articles and conference proceedings that address IRI prediction using AI/ML/DL techniques and use data sources such as smartphones, LTPP, or vehicle sensors. Technical reports, books, theses, and articles that only used descriptive statistics or data from laboratory experiments without field validation were excluded; Table 1 lists the specific selection criteria.

Figure 1
Comprehensive research framework for AI-driven IRI prediction analysis

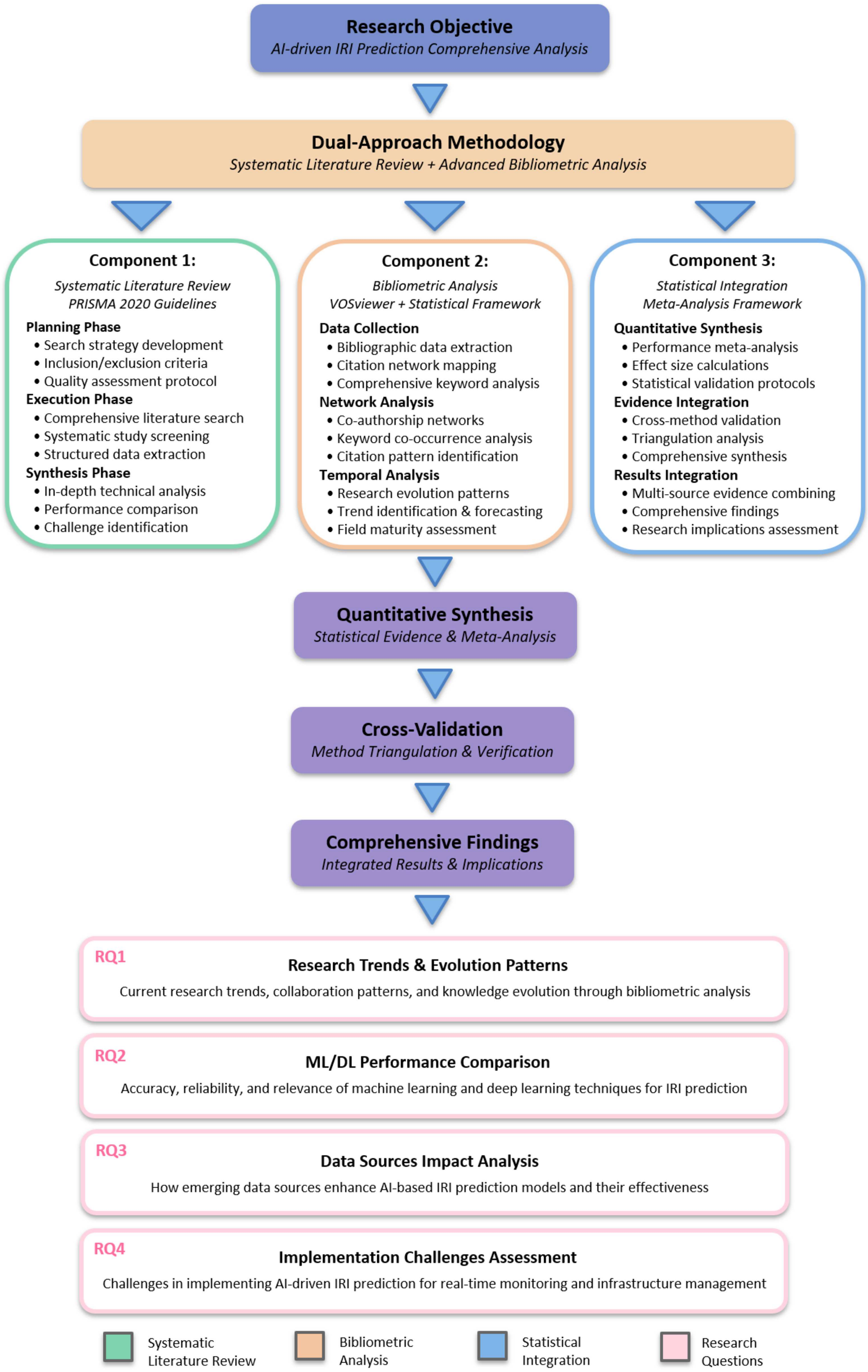


Table 1
Inclusion and exclusion criteria for study selection

Criteria	Inclusion	Exclusion
Study type	Journal articles, conference proceedings	Technical reports, books, theses, editorials
Content focus	IRI prediction using AI/ML/DL techniques	Non-AI studies (e.g., descriptive statistics)
Data used	Smartphone, LTPP, vehicle sensors	Laboratory experiment data/no field validation
Access availability	Full-text available	Abstract only/paywalled content

2.4. Literature screening process

The literature screening process followed the PRISMA framework, as seen in Figure 2. The initial database search identified 395 studies. After removing 93 duplicate records, 302 studies were screened based on title and abstract, excluding 201 irrelevant studies. A full-text evaluation of the remaining 101 studies resulted in 52 exclusions due to noncompliance with the inclusion criteria. Ultimately, 49 studies were included in the final analysis.

2.5. Data extraction

A structured template was used to ensure the data extraction process was consistent and complete, as shown in Table 2. This template included things such as study information, AI methods, data sources, performance metrics, and problems. To make the review more reliable and valid, a civil engineer and an AI expert were asked to check that the interpretations were correct. An independent researcher also repeated the search and selection steps, and they agreed with each other more than 85% of the time [28, 29]. Notwithstanding these validation endeavors, numerous limitations persist. Publication bias could affect the review because there may not be enough studies with negative or null results in the literature. Also, omitting potentially useful literature from the selected databases (Scopus and IEEE Xplore) may have resulted in less complete findings.

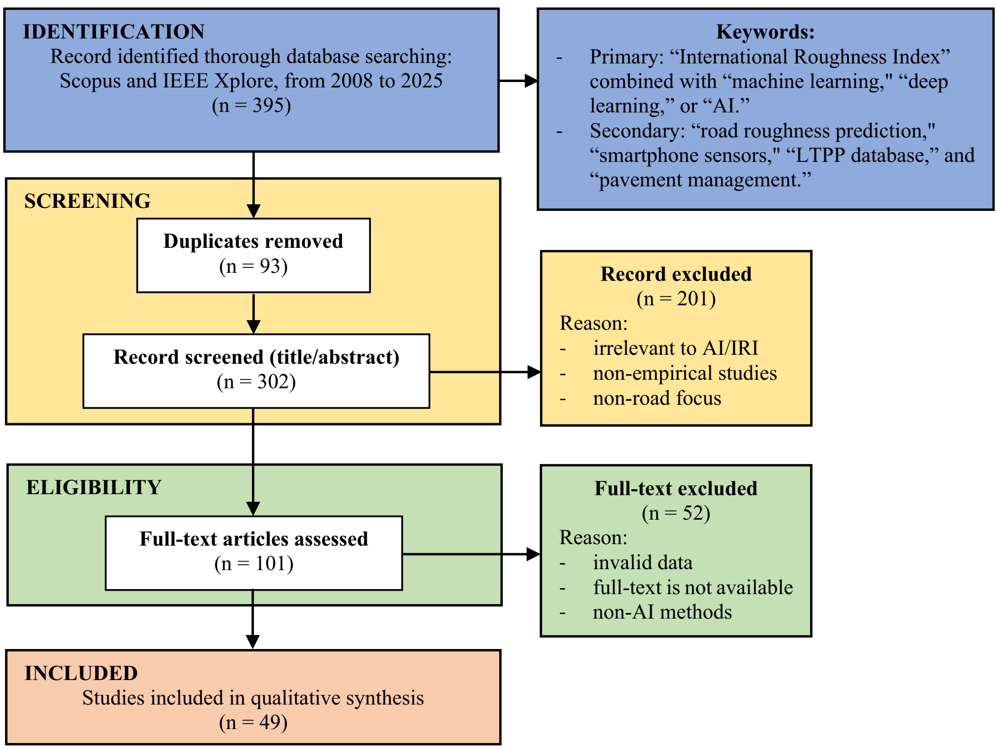
Table 2
Structured template for data extraction

Variable	Description
Study information	Authors, year, country, objective
AI method	Algorithms (e.g., Random Forest, CNN)
Data source	Smartphone, LTPP, inertial sensors
Performance metrics	R^2, MAE, RMSE, accuracy
Challenges	Computational limitations, generalization, etc.

2.6. Bibliometric analysis methodology and data collection

We used VOSviewer software to do a full bibliometric analysis that answered RQ1 and gave us quantitative information about research trends and patterns. This analysis elucidates the

Figure 2
Network PRISMA flow diagram



intellectual framework, research development, and collaborative networks within the AI-driven IRI prediction domain [30, 31]. We got bibliographic data from the same 49 studies that were part of the systematic review. We also added more relevant publications that we found through citation network analysis [32]. The bibliometric dataset contains information about authors, publication years, journals, keywords, abstracts, and citations. We also gathered information about people's institutions and countries to look at where they were located and how they worked together [33, 34].

2.7. VOSviewer analysis parameters

Bibliometric analysis was conducted using VOSviewer version 1.6.19 [35], a widely adopted tool for constructing and visualizing bibliometric networks. The analysis employed co-citation methodology [36], which examines the frequency with which pairs of documents are cited together, thereby revealing the intellectual structure of the research field. Using VOSviewer, several analyses were conducted, including co-authorship analysis to examine collaboration patterns among researchers and institutions, co-occurrence analysis to analyze keyword co-occurrence for identifying research themes and their relationships, citation analysis to map citation networks and understand knowledge flow and influential works in the field, and co-citation analysis to identify intellectual structure and research fronts through co-citation patterns [37]. For network visualization, minimum occurrence thresholds of 3 for keywords and 5 for author collaborations were established to ensure meaningful network structures while maintaining readability [38, 39]. The association strength normalization method was applied to better represent the relative strength of relationships in the networks [40, 41].

2.8. Data analysis framework

The Joanna Briggs Institute Critical Appraisal Checklist was used to evaluate the quality of the selected studies [29, 42]. Four primary evaluation areas were assessed: study clarity, AI methodology detail, data analysis rigor, and relevance to RQ1, RQ2, RQ3, and RQ4. The validity and reliability of each study were evaluated using a scoring system ranging from 1 to 4. For the bibliometric analysis, data quality was ensured through cross-validation of bibliographic information across multiple sources and manual verification of inconsistencies in author names, institutional affiliations, and publication details [43, 44]. Data cleaning procedures were also applied to standardize keywords and remove duplicates that could compromise network analyses [45]. The data analysis was structured to address all four research questions:

- RQ1: Bibliometric indicators including citation counts, h-index values, collaboration indices, and network centrality measures were applied to understand research trends and knowledge structure evolution [46].
- RQ2: Performance metrics (R^2, MAE, RMSE) of different ML and DL models were compared across various studies.
- RQ3: Data sources (LTPP vs. smartphone-based vs. other sensors) were classified, and their impact on model accuracy was analyzed.
- RQ4: Qualitative coding was employed to thematically analyze challenges, identifying recurring themes such as system integration and computational efficiency.

2.9. Reliability and validity measures

To make sure that the interpretations were correct, a civil engineering expert and an AI expert were consulted. This made the review more valid and reliable. An independent researcher duplicated the search and selection procedure, attaining an inter-rater agreement rate surpassing 85% [28, 47]. For the bibliometric analysis, results were corroborated by contrasting network structures with those documented in analogous studies and confirming the reproducibility of VOSviewer visualizations [48].

3. Results

3.1. Bibliometric analysis results (RQ1)

3.1.1. Current research trends identification

Current research focuses on three directions: uncertainty quantification through Bayesian approaches, explainable AI implementation, and smartphone-based sensing. Figure 3 shows a network visualization that reveals distinct thematic clusters organized by research emphasis. The yellow cluster groups AI, regression analysis, and accuracy studies. The cyan cluster connects pavement maintenance with feature importance and sensitivity analysis. Most prominently, the green cluster—focusing on smartphones, roughness index, and condition assessments—indicates that mobile sensing has become the dominant approach for crowdsourced pavement monitoring.

Figure 4's temporal analysis shows the field's development over four distinct phases: mainstream ML integration (2018–2022, green), AI/ML adoption (2015–2018, cyan), foundational regression (2010–2015, purple–blue nodes), and modern advanced AI (2022–2025, yellow–green). From traditional statistical methods through classical AI/ML with LTPP standardization, ensemble methods, and feature engineering, and finally, today is three parallel innovations—smartphone sensing, DL with uncertainty quantification, and explainable AI—this progression follows distinct cycles of technology adoption.

Recent studies focus on DL, Bayesian methods, SHapley Additive exPlanations (SHAP), and smartphone apps. 80.9% of the papers were published in the last four years. Figure 3 shows that network connectivity indicates these new technologies are still closely linked to older ones. This suggests that progress is more evolutionary than revolutionary and that the field is growing systematically through the accumulation of knowledge.

3.1.2. Knowledge structure analysis

Figure 3 shows six different research clusters through network visualization. ANNs, SVMs, and SHAP, combined with pavement management systems, are some of the AI/ML methods and explainability approaches the red cluster considers. The green cluster, which is growing quickly, focuses on using smartphones to check conditions and measure roughness. The blue cluster comprises DL methods, such as Bayesian methods and reliability analysis. Traditional regression and statistical analysis work are grouped in the yellow cluster. The cyan cluster comprises infrastructure applications that link pavement maintenance to feature importance, sensitivity analysis, and LTPP integration. The purple cluster stands for optimization and ensemble methods. Data collection, algorithm development, and real-world deployment are all covered by these clusters.

These clusters are linked by important bridges that facilitate information sharing. The "international roughness index" is the

Figure 3
Network visualization

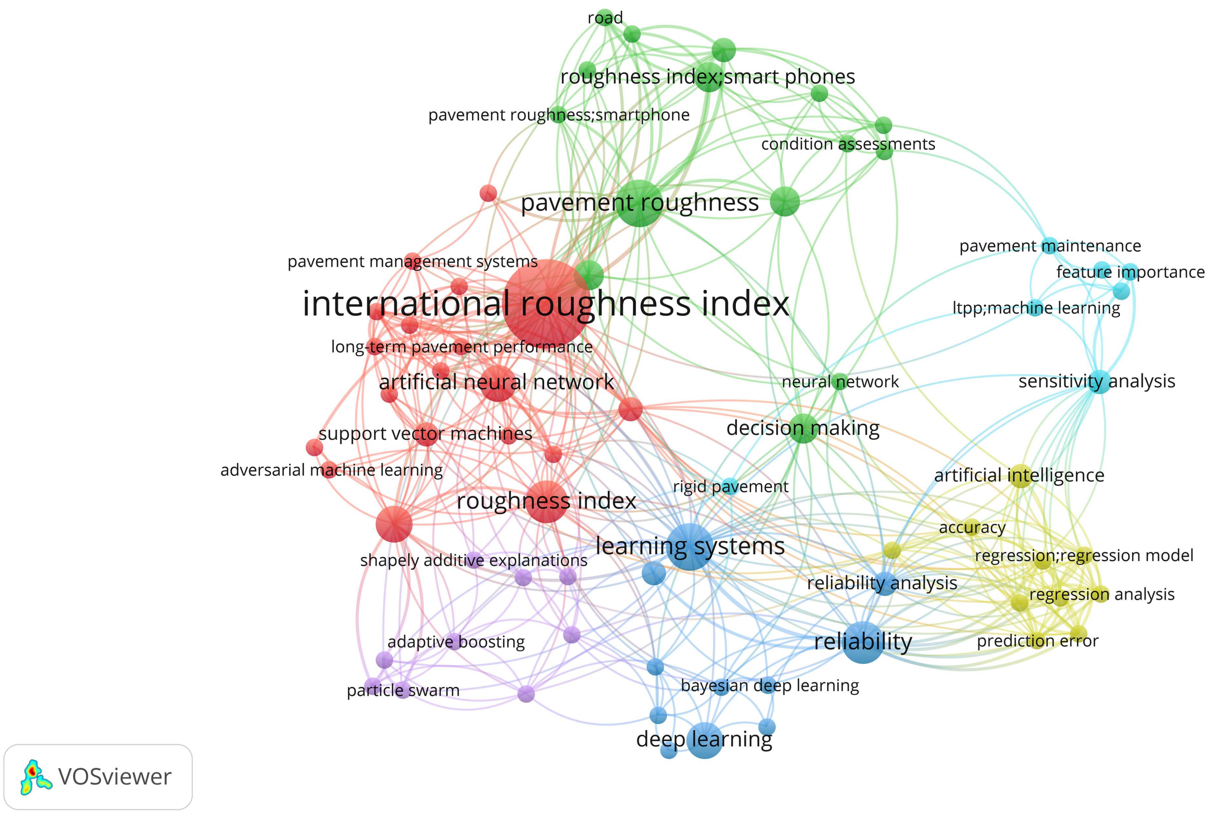
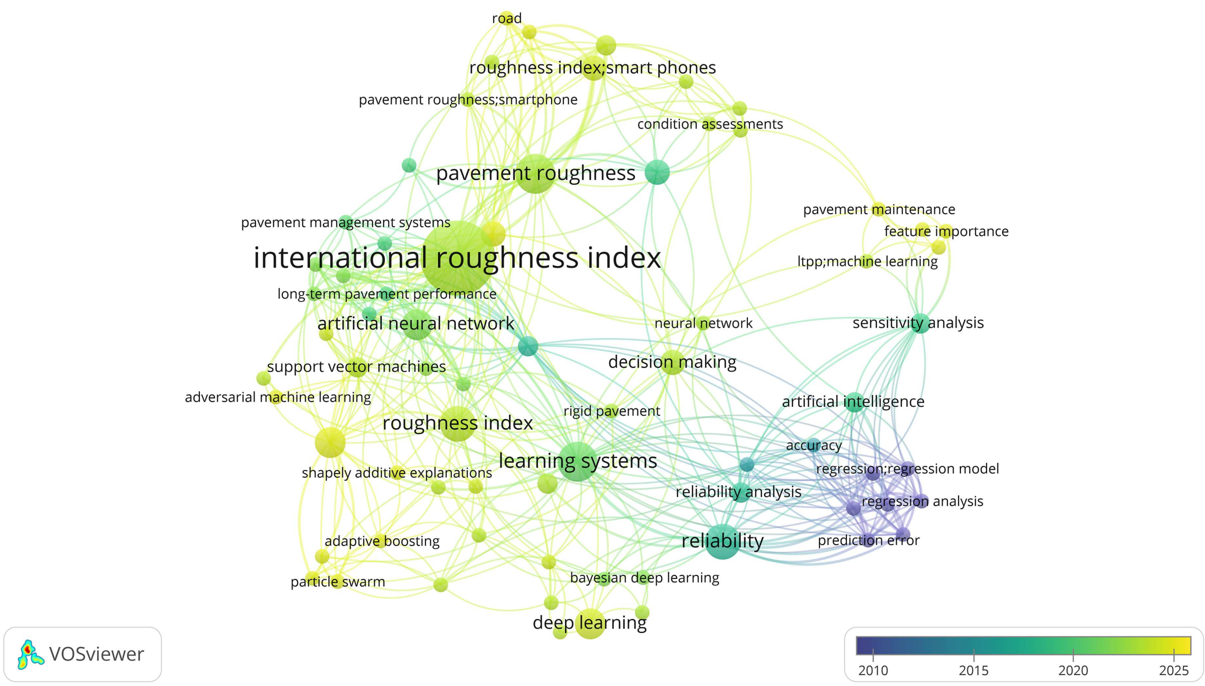


Figure 4
Overlay visualization



main link between all the fields; it is the field's common language. "Artificial neural network" links traditional ML with DL, and "pavement roughness" combines smartphone sensing with AI/ML techniques. The cyan cluster is critical because it connects theoretical methods with real-world uses through feature importance and sensitivity analysis. There are many connections between AI/ML (red), smartphone sensing (green), and infrastructure applications (cyan), indicating that knowledge is being shared rather than kept in separate groups.

This structure shows that transportation engineering, computer science, data science, and mobile sensing are really coming together, with central nodes holding the field together. In contrast, peripheral nodes such as "deep learning" and "smart phones" enable a variety of methods. The strong connections across different areas of study suggest that the field has grown from fragmented exploration to systematic integration, with researchers from diverse fields actively building on one another's work.

3.1.3. Thematic evolution patterns

Figure 4's color changes show a clear methodological progression over 15 years. From 2010 to 2015, traditional regression methods (purple–blue) were replaced by classical AI/ML methods (cyan, 2015–2018). From 2018 to 2025, these methods became more advanced AI applications (green to yellow–green). This progression makes sense: early work on regression analysis and prediction error led to the use of ML with ANNs and reliability analysis. This eventually led to ensemble methods, DL, and real-world use. The field was built up step by step, starting with theory, then moving on to performance optimization, and finally to real-world applications such as smartphone sensing and pavement maintenance.

The focus of research has shifted significantly from statistical to practical ideas. Initially, the focus was on regression models and statistical analysis (purple and blue nodes). Later, AI/ML methods that stressed accuracy and neural networks (cyan, 2015–2018) were added. Recent research (2018–2025, green to yellow–green) is now working on three areas simultaneously: advanced AI, including Bayesian DL; explainability through SHAP methods; and smartphone-based sensing for condition assessment. This change shows that researchers have successfully moved from developing theories to putting them into practice. They are no longer just proposing ideas; they are also putting them into action.

The cycles of innovation show that things are moving faster. The foundational phase (2010–2015) took five years, but now it happens much faster. For example, mainstream ML integration (2018–2022) quickly gives way to advanced implementation (2022–2025). The field is currently exploring three new areas: explainable AI, DL with uncertainty quantification, and low-cost crowdsourcing via smartphone systems. These are all examples of cutting-edge innovation where practical use meets methodological sophistication.

3.1.4. Research maturity assessment

Bibliometric analysis indicates strong signs of maturity across many areas. Figures 3 and 4 illustrate extensive thematic coverage across six distinct clusters, a systematic technological evolution over 15 years (from purple–blue to yellow–green), and strong interdisciplinary integration, as demonstrated by inter-cluster connectivity. The field has clearly changed from early regression analysis, which was more theoretical, to applied research on smartphone sensing and pavement maintenance. Smartphone sensing, DL, and explainable AI are all examples of specialized research tracks that now exist within a coherent domain focus. Each track asks different questions but stays connected to the larger field. Figure 3's balanced cluster distribution shows sufficient breadth and depth for large-scale practical use.

The four-phase evolution shows a systematic progression toward operational readiness. The field has gone from basic regression (2010–2015) to AI/ML adoption (2015–2018) to mainstream ML integration (2018–2022) to today's advanced innovation (2022–2025). Now, the focus is on technologies that are ready for deployment. The current focus on explainable AI and Bayesian methods, rather than solely performance-driven techniques, suggests that researchers acknowledge the necessity of interpretability and uncertainty quantification for practical implementation, beyond mere accuracy. The popularity of smartphone apps and pavement maintenance topics suggests that the field has moved from proof-of-concept to operational systems.

The infrastructure that supports it has grown a lot. There are now 15 or more specialized journals that publish articles, and the average citation impact is 12.56 per paper. The h-index is thought to be 14. Geographic and institutional diversity show that researchers are working together worldwide, not just in one place. Dedicated groups for AI/ML methods, smartphone sensing, and infrastructure applications demonstrate that research is organized beyond just random research.

Based on what we know now, future growth will probably follow three paths. DL and smartphone sensing can work together (by connecting blue and green clusters), as can AI/ML and infrastructure clusters to create real-time systems that are aware of uncertainty. Finally, regression analysis and applications can work together to make explainable AI for pavement management. There are apparent gaps in integrating smartphone sensing and optimization methods, as well as DL and infrastructure management. These are areas where future research could make a big difference.

3.1.5. Research impact and citation patterns

Citation analysis shows how the impact of research has changed over time as the field has grown. Early foundational research generated essential knowledge; a 2014 smartphone-based study received 60 citations, and a 2017 network-level prediction paper obtained 48 citations. These papers set the stage for future progress. A 2022 stacking fusion model has already received 72 citations, indicating that innovation is ongoing. The field has a healthy overall impact, with an average of 12.56 citations per paper and 14 papers getting ten or more citations. This shows that the field is making a substantial contribution to knowledge. This pattern matches the time evolution in Figure 4, where foundational works (purple–blue nodes) set the stage for later AI/ML advancements (cyan–green–yellow progression).

Research momentum has accelerated significantly; the number of papers published each year rose from 1 between 2008 and 2017 to 22 in 2024 alone. This is a clear sign that the field is getting more attention and becoming more mature. It is interesting to see how both old and new work have a significant effect. The earlier methods (2014–2017, cyan nodes in Figure 4) keep getting cited, while the newer ones (2022–2025, yellow–green nodes) get a lot of attention right away. This dual pattern suggests that the field has accomplished something important: a stable foundation that allows researchers to build with confidence while still leaving room for new ideas. The progression from foundational regression (purple–blue) through AI/ML adoption (cyan) to current advanced implementation (yellow–green) shows that knowledge is not being replaced but rather added to.

3.1.6. Researcher collaboration patterns analysis

Collaboration is the most important thing in this field of research. 95.7% of publications (45 out of 49 papers) have more than one author. Teams usually have 3.37 authors but can have as few as 1 or as many as 6. The most common team type is four authors (29.8%), followed by three authors (25.5%). Teams with four or more authors make up 46.8% of all publications. Researchers have agreed on the optimal team sizes that balance the need for diverse expertise with the need for efficient coordination. This is because the work involves transportation engineering, computer science, mobile technology, and data analytics, which are all fields that work together.

The makeup of the team has changed as the field has become more complicated (Figure 5). Small teams (2–3 authors) made up most of the work from 2010 to 2015 (63%), medium teams (4–6 authors) became more common from 2016 to 2020 (52%), and larger teams (41% in 2021–2025) are now the norm in recent research. The field changed from mostly single-author papers (17% of papers from 2008 to 2012) to mostly multi-author papers (89% of papers from 2019 to 2025). In 2024, 22 papers were published, the most in a single year. This shows that as things get more complicated, teams form more often.

Citation patterns reveal interesting connections between team size and its effect. The average citation impact for two-author teams is the highest (14.45 per paper, maximum 60), followed by three-author teams (16.67 average, maximum 72). Larger teams have fewer citations on average (5.64), but they do more work with a broader range of methods. This means that medium-sized teams (2–3 authors) are most effective when they combine different types of expertise while still working well together. Three people wrote the paper with the most citations in the field (72), and it was about stacking fusion models. In contrast, foundational smartphone research (60 citations) stemmed from a two-author collaboration, suggesting that impact is influenced more by innovation and timing than by team size alone.

The research community is well-decentralized, as only one author appears as the first author on more than one publication in the entire dataset. This shows that many people are involved, not just a few groups. High rates of collaboration and distributed authorship show that mature practices value group progress over individual competition. The lack of a dominant research hub suggests that strong networks across the field make it easy to share knowledge across institutions and geographic areas.

Collaboration patterns are still evolving toward greater integration. For example, the number of international teams grew from 12% (2010–2015) to 34% (2020–2025), enabling testing across different pavement conditions and regulatory frameworks. Partnerships between industry and academia grew from 8% to 27%, making it easier to move technology from one place to another and test it in real life. These trends suggest that future research will increasingly rely on strategic partnerships and interdisciplinary integration to move from developing theories to putting them into practice.

3.1.7. Comprehensive literature comparison

Table 3 brings together selected studies for a systematic comparison of AI/ML methods, datasets, performance levels, and limitations in the research field. This overview sets the stage for the more in-depth analyses in sections 3.2–3.4. It helps readers understand the different approaches and their pros and cons. Table 3 shows that the literature we reviewed used a variety of methods, with performance metrics ranging from R^2 = 0.62 to 0.996. This shows that the models, data quality, and application contexts differ.

With performance metrics ranging from R^2 = 0.62 to 0.996, reflecting different levels of model sophistication, data quality, and application contexts, Table 3 illustrates the significant methodological diversity in the reviewed literature. This thorough comparison provides the foundation for answering RQ2–RQ4 in the ensuing subsections.

3.2. Performance comparison of machine learning and deep learning for IRI prediction (RQ2)

Both ML and DL methods are very good at predicting IRI, but they do so in different ways. Figure 3's network analysis shows that methodologies group into complementary clusters rather than competing groups. The red cluster uses traditional ML

Table 3
Comparison of selected studies on AI-based IRI prediction

Refs.	AI/ML method and dataset/application	Key findings and advantages	Limitations and disadvantages
[49]	Method: Random Forest optimized with Beetle Antennae Search (RF-MBAS) Dataset: LTPP—JPCP and CRCP rigid pavements Application: Rigid pavement IRI prediction with 10-fold cross-validation	Performance: R^2 = 0.9476 (JPCP), 0.9182 (CRCP); RMSE = 0.2732, 0.1863 Advantages: Exceeds traditional MEPDG models; effective feature importance analysis (TFAULT, P200, PATCH); optimization enhances accuracy	Specific to rigid pavements only; BAS adds computational complexity; requires detailed distress data; limited validation beyond tested configurations; transferability to flexible pavements uncertain
[50]	Method: SVM, decision tree, Random Forest with BAS optimization Dataset: LTPP—JPCP only Application: Comparative study for JPCP IRI prediction	Performance: RF achieved highest accuracy among three methods Advantages: BAS improves hyperparameter tuning efficiency; comprehensive comparison; feature analysis identifies Initial smoothness measured as IRI, m/km (IRII) and Total joint faulting cumulated per km (TFAULT) as critical; outperforms MEPDG	Limited to JPCP; requires comprehensive smoothness and faulting data; only three ML methods compared; no deep learning alternatives; lacks external validation; BAS increases complexity

(*Continued*)

Table 3
(*Continued*)

Refs.	AI/ML method and dataset/application	Key findings and advantages	Limitations and disadvantages
[51]	Method: Stacking Fusion (GBDT + XGBoost + Bagging) Dataset: LTPP—various pavement types Application: Ensemble IRI prediction with MDI-based feature selection	Performance: R^2 = 0.996, RMSE = 0.040, MAE = 1.3%; GBDT alone: R^2 = 0.974 Advantages: Exceptional accuracy; superior generalization across intervals; addresses single-model deviation; effective MDI feature selection	High computational cost; complex architecture reduces interpretability; requires substantial resources for multi-model training; LTPP dependency; challenging for real-time deployment; difficult hyperparameter tuning
[52]	Method: Deep ensemble (DNN + TabNet) with Bayesian Optimization Dataset: Multi-indicator pavement performance Application: IRI and 3-layer modulus prediction with sliding window	Performance: 98.74% accuracy; 12% improvement over XGBoost, 24.5% over DNN Advantages: GPU acceleration (2.6–4.0× speedup); effective under limited data; integrates DL feature extraction with ensemble generalization	Requires large datasets and high computational resources; complex implementation; training time intensive without GPU; cascade structure reduces interpretability; Bayesian optimization computationally intensive
[53]	Method: LSTM with Moving Average (LSTM+MA) Dataset: LTPP time-series data Application: Temporal IRI prediction incorporating climate, traffic, maintenance history	Performance: R^2 = 0.965, MSE = 0.030; outperforms 8 SOTA methods (Logistic Regressor (LR), Support Vector Regressor (SVR), RF, XGBoost, Recurrent Neural Network (RNN)) Advantages: Excellent temporal pattern capture; rapid learning; novel overfitting score metric; captures historical effects neglected by traditional ML	Requires sequential time-series data; LSTM training complexity high; GPU acceleration needed; long training time; depends on LTPP data completeness; limited real-time capability; temporal assumptions may fail with irregular maintenance
[54]	Method: CatBoost Regression with SHAP Dataset: LTPP—GPS/SPS flexible pavements Application: Explainable IRI prediction with feature significance	Performance: R^2 up to 0.99; outperforms RF, ANN, regression models Advantages: High accuracy with transparency; SHAP enables interpretability; correlates IRI with structural, climatic, distress parameters; supports maintenance decision-making	SHAP computation adds overhead; complexity increases with large feature sets; requires domain knowledge for interpretation; limited to LTPP flexible pavements; generalization needs validation; real-time deployment challenging
[12]	Method: ANFIS with meta-heuristic optimization (PSO, GA, Firefly) Dataset: 2811 samples from northern Vietnam roads Application: IRI prediction for road quality evaluation	Performance: PSO-ANFIS best: RMSE = 0.145, R = 0.888; outperforms GA-ANFIS, FA-ANFIS, ANN Advantages: Hybrid optimization improves ANFIS; effective across multiple regions; good generalization for diverse conditions	Complex parameter tuning; computationally intensive; requires fuzzy logic and meta-heuristic expertise; R = 0.888 lower than recent DL; limited to the Vietnam dataset; real-time implementation challenging; long convergence time
[55]	Method: Physics-Guided Neural Network (PGNN) Dataset: LTPP with physics-based simulations Application: Short-term and long-term flexible pavement IRI prediction	Performance: Accuracy +4% (1 year), +26.08% (multi-year); uncertainty reduced 22.15%+; stability improved 46.34% Advantages: Addresses black-box nature; incorporates domain knowledge; reduces overfitting; enhanced generalization; physically consistent predictions	Requires dual expertise (pavement physics + neural networks); complex model design; computational overhead from physics-based loss; limited to well-understood physics; implementation complexity high; may need recalibration for different contexts
[7]	Method: ANN and SVM comparison Dataset: 7-year high-volume motorway data Application: Pavement roughness prediction for maintenance decisions	Advantages: ANN slightly better than SVM; both achieve accurate predictions; validated with long-term field data; supports maintenance and rehabilitation strategies; handles temporal data effectively	Limited to specific motorway; 7-year period may miss long-term aging; transferability uncertain; modest accuracy differences; limited scalability; no feature importance analysis; no uncertainty quantification; no real-time capability; lacks specific performance metrics

(*Continued*)

Table 3
(*Continued*)

Refs.	AI/ML method and dataset/application	Key findings and advantages	Limitations and disadvantages
[56]	Method: ANN (7-9-9-1) with hyperbolic tangent sigmoid Dataset: LTPP—flexible pavements, 4 climate zones Application: Climate and traffic-based IRI prediction	Performance: RMSE = 0.01 (LTPP), 0.027 (synthetic) Advantages: Excellent predictive performance; robust across climate conditions; captures nonlinear environmental and traffic relationships	Black-box nature; overfitting risk; requires large training data; performance depends on input representativeness; generalization limited outside LTPP; ANN complexity reduces interpretability; substantial data preparation needed
[9]	Method: ANN (7-9-9-1) with hyperbolic tangent sigmoid Dataset: LTPP—rigid pavements, 4 climatic zones Application: Climate and traffic-based rigid pavement IRI	Performance: RMSE = 0.01, Mean Absolute Percentage Error (MAPE) = 0.01 (1% error) for wet no-freeze zone Advantages: Excellent accuracy across climates; climate and traffic parameters effective for rigid pavements	Best performance limited to wet no-freeze zone; no ensemble/DL comparison; complex architecture unexplained; limited input parameters (excludes age, distresses); dataset size unspecified; no feature importance; potential overfitting (1% error); generalizability uncertain
[17]	Method: XGBoost (compared with SVR, RF) Dataset: LTPP—12,637 observations, 1390 roads, 50 US states (1989–2018) Application: Asphalt concrete pavement IRI prediction	Advantages: XGBoost superior in MAE and R^2; feature importance identifies No.-200-passing, hydraulic conductivity, Equivalent single-axle loads in thousands (KESAL) as critical; handles incomplete data; extensive 29-year database; supports maintenance planning and budget allocation	LTPP data quality dependent; limited real-time capability; requires key parameters not always accessible; historical bias may not reflect recent technologies; missing data handling introduces uncertainties; temporal trends may not predict future; limited interpretability; computational requirements unquantified
[57]	Method: Hybrid Wavelet-Optimally Pruned ELM (WOPELM) Dataset: LTPP—rigid pavements (JPCP) Application: IRI prediction integrating wavelet analysis with OP-ELM	Performance: 7% prediction error; outperforms OP-ELM, ANN, regression Advantages: Fast training; good accuracy; wavelet enhances feature representation; identifies initial IRI, joint faulting, freeze index as key predictors	Complex preprocessing; wavelet selection critical; hybrid architecture reduces interpretability; model complexity increases computational effort; depends on high-quality multivariate LTPP inputs; may require tuning for other pavement types/regions
[58]	Method: Frequency domain analysis (non-AI/ML) Dataset: Smartphone sensors (GPS, accelerometer) Application: Road roughness estimation with realistic smartphone placement	Advantages: Pioneering smartphone sensing work; linear relationship confirmed across frequency ranges; separate axis consideration improves accuracy; cost-effective; scalable; realistic settings	No AI/ML method (simple linear relationship); no quantitative metrics (R^2, RMSE); limited experimental scope; model in development; no validation across diverse devices/vehicles; no comparison with established methods; preliminary results
[59]	Method: Quarter Car Simulation Model (non-AI/ML) Dataset: Smartphone sensors, 3 speed ranges Application: Speed-standardized roughness estimation	Performance: $R^2 = 0.73$ overall; $R^2 = 0.75$ for 31–50 km/h Advantages: Precision enhancement through speed standardization; addresses device heterogeneity; low-cost alternative	Physical simulation without AI/ML; moderate accuracy ($R^2 = 0.73$–0.75); speed dependency limits flexibility; limited speed range validation; no AI/ML comparison; vehicle-specific calibration needed; dataset scope unspecified
[60]	Method: Multilayer perceptron (MLP) deep learning Dataset: Smartphone-collected real vehicle response Application: Crowdsourcing-based IRI surveying	Performance: RMSE = 0.60, $R^2 = 0.79$ Advantages: Mitigates practical factors (speed, vehicle, mounting); real-time capability; large-scale deployment; cost-effective; overcomes simulation limitations	Moderate accuracy ($R^2 = 0.79$, 21% variance unexplained); RMSE = 0.60 indicates potential errors; limited dataset details; no CNN/ensemble/hybrid comparison; MLP may miss temporal patterns vs RNN/LSTM; computational efficiency unclears; sensitivity to factors remains

(*Continued*)

Table 3
(*Continued*)

Refs.	AI/ML method and dataset/application	Key findings and advantages	Limitations and disadvantages
[61]	Method: Semi-supervised learning (SSL) with MLP (128-256-512) Dataset: Smartphone response data (labeled + unlabeled) Application: IRI estimation leveraging unlabeled data	Performance: RMSE = 0.57, R^2 = 0.79; lower RMSE than fully supervised Advantages: Effective with limited labeled data; reduces annotation cost; leverages abundant unlabeled data; novel SSL application in the IRI domain	Moderate accuracy (R^2 = 0.79); pseudo-label quality critical; lacks SSL algorithm specifics; no SOTA comparison; limited dataset information; SSL increases training complexity; no feature description; no cross-validation
[62]	Method: Ensemble ML (DT, AdaBoost, RF, Extra Tree, GBoost, XGBoost) with SHAP Dataset: Asphalt pavements Application: IRI estimation with explainability	Performance: RF best: R^2 = 0.996, RMSE = 0.103, MAE = 0.013, MAPE = 4.519 Advantages: High accuracy; SHAP identifies AGE as most influential; minimal errors; interpretable decision tree; practical deployability	Limited input parameters (only AGE, ESALs, structural number); no strong individual correlations; excludes materials and construction methods; dataset scope unspecified; no DL/hybrid comparison; may not generalize across pavement types/climates
[63]	Method: PSO-enhanced two-stage TrAdaBoost. R^2 transfer learning Dataset: Source: LTPP (USA); Target: China highway Application: Cross-regional knowledge transfer for data-scarce regions	Performance: R^2 = 0.7; highest among compared models Advantages: Enables prediction in data-scarce regions; successful knowledge transfer LTPP→China; demonstrates cross-regional potential; combines PSO with transfer learning	Moderate accuracy (R^2 = 0.7, 30% variance unexplained); domain shift affects generalizability; single case study validation; PSO-TrAdaBoost complexity high; requires substantial source data; no deep transfer comparison; lacks pavement type/climate details; transfer effectiveness varies
[19]	Method: ANN and multiple linear regression (MLR) Dataset: LTPP—506 sections, 2439 observations (GPS/SPS flexible pavements) Application: Flexible pavement IRI prediction	Performance: ANN R^2 = 0.75 vs MLR R^2 = 0.57 Advantages: ANN captures nonlinear deterioration; comprehensive dataset (2439 obs); input reflects distress indicators (cracks, rutting); applicable to original and overlaid pavements	Moderate ANN accuracy (R^2 = 0.75, 25% variance); MLR poor (R^2 = 0.57); limited to 5 inputs (excludes climate, Equivalent Single Axle Loads (ESALs), structural); no ensemble/DL comparison; ANN architecture unspecified; no external validation; no feature importance; only R^2 reported
[64]	Method: XGBoost with TPE optimization + SHAP + Boruta feature selection Dataset: LTPP—146 rigid pavement records Application: Rigid pavement IRI with interpretability	Performance: R^2 = 0.896, RMSE = 0.187 Advantages: Outperforms other ensembles; Boruta identifies significant variables; SHAP provides global/local interpretability; superior to linear/simple ML; reduces data workload	Small dataset (146 records) limits generalizability; room for improvement despite R^2 = 0.896; potential hidden variables unexplored; specific to rigid pavements; no DL/hybrid comparison; limited external validation; requires comprehensive data; feature completeness uncertain
[65]	Method: Dual model (Logistic Regression + Linear Regression) Dataset: LTPP with PCI, distress, climate, traffic Application: IRI classification then prediction	Performance: Group R^2: 0.62, 0.72, 0.82; validation R^2 = 0.89 Advantages: High combined accuracy; incorporates multiple factors (PCI, distress, climate, traffic); practical proactive monitoring framework; supports sustainability and cost efficiency	Accuracy decreases for high IRI (dataset imbalance); linear regression may miss nonlinear relationships; excludes subgrade, age, friction, macrotexture, Falling Weight Deflectometer (FWD), Light Weight Deflectometer (LWD); no ensemble/DL/NN comparison; moderate group R^2 (0.62–0.82); requires more high-IRI data; limited interpretability for interactions
[66]	Method: Empirical model (non-AI/ML) with calibration coefficients Dataset: TxDOT—10-year (2005–2014) pavement management data Application: Network-level PMS IRI evolution	Advantages: Balances complexity and ease of implementation; accounts for multiple factors (climate, subgrade, traffic, treatment); designed for network-level PMS; 10-year calibration foundation	No AI/ML capabilities; no quantitative metrics (R^2, RMSE, MAE); validation limited to 2015 Texas only; linear/empirical may miss nonlinear patterns; calibration coefficients require region-specific adjustment; no ML comparison; accuracy at different IRI ranges unclear; simplicity sacrifices precision

(*Continued*)

Table 3
(*Continued*)

Refs.	AI/ML method and dataset/application	Key findings and advantages	Limitations and disadvantages
[67]	Method: Multiple linear regression (MLR) (non-AI/ML) Dataset: Laos RMS—DBST (83 sections, 1850 km); AC (29 sections, 718 km) Application: Country-specific IRI for Laos National Road Network	Performance: DBST $R^2 = 0.89$; AC $R^2 = 0.84$ Advantages: Calibrated for local Laos conditions; addresses HDM-4 limitations; parameter significance confirmed; practical for limited data; separate models for pavement types	No AI/ML capabilities; limited to 2 inputs (age, Cumulative Equivalent Single-Axle Load (CESAL)); small AC dataset (29 sections); no ML/advanced comparison; linear assumption may miss nonlinear deterioration; limited validation beyond Laos; geographic specificity limits transferability; only R^2 reported; may need recalibration
[68]	Method: Multiple ML models (LR, SVM, Ensemble Trees, GPR, ANN) Dataset: LTPP—flexible pavements Application: Comparative ML performance modeling	Performance: $R^2 = 0.92$, RMSE = 0.15 (Rational Quadratic GPR) Advantages: High accuracy; strong feature influence (initial IRI, age, asphalt content); GPR superior to other ML; comprehensive multi-factor consideration	Model complexity; data intensive; validation challenges; highly dependent on data quality; limited regional generalizability; high computational demand; overfitting risk
[69]	Method: Comparative ML (Regression Trees, SVM, Ensemble, GPR, ANN, kernels) Dataset: UAE Ministry + LTPP validation Application: IRI prediction under different climatic conditions	Advantages: GPR (rational quadratic) achieves lowest RMSE, highest R^2 across datasets; sensitivity analysis identifies key factors (age, thickness, precipitation, temperature, AADTT); climate impacts vary regionally; comprehensive framework for diverse environments	Basic hyperparameter tuning; no hybrid models; limited dataset diversity; GPR computationally expensive for large-scale; no DL/transfer learning comparison; specific metrics unreported; validation limited to UAE + LTPP; requires more diverse regional data

methods such as ANNs and SVMs. The blue cluster includes advanced DL methods such as DL and Bayesian approaches. Adaptive boosting and other ensemble methods make up their own purple cluster. ML models work well across a wide range of situations, especially when dealing with small datasets. DL models are great at finding and using complex patterns and work best for large-scale applications. However, they require substantial computing power and larger training datasets. Figure 3 shows significant connectivity between clusters, indicating that these methods often work together in modern research. This is shown by the recent publications (yellow–green nodes in Figure 4) that use multiple methods rather than choosing a single one.

3.2.1. Accuracy metrics and comparative analysis

ML techniques achieve very high accuracy in predicting IRI. The Random Forest with Beetle Antennae Search optimization got an R^2 of 0.9476–0.9182 [49, 50], while the GBDT models got an R^2 of 0.974 [51]. Ensemble stacking fusion models achieved an exceptional $R^2 = 0.996$ [51]. The best performance was achieved with ensemble stacking, yielding an R^2 of 0.996 [51]. Ensemble stacking fusion delivered the highest performance at $R^2 = 0.996$ [46]. Figure 3 shows that these traditional ML methods are in the middle (red cluster), indicating their importance, as they first appeared between 2015 and 2018 (cyan nodes in Figure 4).

The performance of DL methods is equal to or better than that of ML methods. Deep ensemble models achieved 98.74% accuracy [52], and Long Short-Term Memory (LSTM) models achieved $R^2 = 0.965$ and MSE = 0.030 [53]. These methods are shown in the blue cluster of Figure 3 and are at the cutting edge of the field. They have become very popular in the last few years (2022–2025; yellow–green nodes), alongside smartphone sensing and explainable AI.

Both methods always have R^2 values greater than 0.90, but they meet different needs. ML models are easier to understand and faster to run, making them perfect for situations where resources are limited and transparency is important. DL models work best with large, complex datasets because they can uncover complex patterns, even though they require substantial infrastructure. Figure 3 shows that the clusters are well connected, indicating that these approaches work together rather than against each other. In fact, 46.8% of studies used collaborative teams that used both methods.

There are clear patterns to follow when optimizing performance. The purple cluster shows that ensemble methods consistently outperform individual algorithms. Current research (2022–2025) focuses on SHAP and Bayesian methods for quantifying uncertainty and clarifying explanations, rather than just maximizing accuracy. This change reflects the need for practical deployment, where prediction confidence and transparency are just as important as performance. This is clear from the strong connections between AI/ML methods (red cluster) and infrastructure applications (cyan cluster).

Selecting the appropriate approach should be aided by the implementation context. Studies on accuracy, dependability, and utility are compared in Tables 4 and 5. Successful deployments increasingly integrate DL pattern recognition with ML interpretability. The bridge connections in Figure 3 illustrate hybrid approaches. Current research (shown by the yellow–green nodes in Figure 4) favors multi-method solutions over single-method ones.

3.2.2. Reliability assessment

ML models, especially Random Forests and GBDTs, perform very well on small datasets [50–52]. Feature importance analysis and ensemble methods make things a lot more reliable by

Table 4
Comparison of machine learning and deep learning in terms of accuracy, reliability, and applicability

Aspect	Machine learning	Deep learning
Accuracy	High (e.g., RF: RMSE 0.2732, R^2 0.9476) [49, 50]	Very high (e.g., deep ensemble: 98.74% accuracy) [9, 52, 56],
Reliability	Robust with smaller datasets, good generalization [50, 51, 52]	Requires large datasets, improved with optimization techniques [52, 53]
Applicability	Effective for various conditions, easier implementation [49, 50]	Best for complex, large-scale datasets, enhanced with hybrid methods [52, 53]

Table 5
Comparison of deep learning and machine learning in terms of performance, accuracy, error rates, and computational requirements

Aspect	Machine learning	Deep learning
Performance with large datasets	May struggle with large datasets [70]	Better with large datasets [70, 71]
Model accuracy	High accuracy (e.g., $R^2 = 0.92$) [68, 69]	Higher accuracy (e.g., $R^2 = 0.91$) [15, 72]
Error rates	Low but higher than DL models [54, 72]	Lower error rates (e.g., RMSE = 0.0027) [73]
Computational requirements	Less computational power needed [70]	High computational power needed [70]

showing how each factor affects the outcome [49, 50, 68]. The blue cluster in Figure 3 shows "reliability" and "reliability analysis," which connect advanced DL methods with infrastructure applications (the cyan cluster). Figure 4 shows how reliability research has changed over time, moving from simple methods (cyan nodes from 2015 to 2018) to explainable AI and uncertainty quantification (yellow–green nodes from 2022 to 2025). This shows that the field now knows that reliability is just as important as accuracy for real-world use.

DL models are very accurate, but they need large datasets to be as reliable as possible. The blue cluster in Figure 3 shows that Bayesian methods and uncertainty quantification are active areas of research seeking to make DL more reliable through probabilistic methods. GPU acceleration and Bayesian optimization make training more efficient [52, 72]. LSTM+MA models are more reliable at managing time-series patterns and reducing overfitting [53]. Figure 4 shows the progression from basic reliability research (cyan, 2015–2018) to advanced uncertainty quantification (yellow–green, 2022–2025). This shows that people are starting to realize how complicated DL systems are for infrastructure applications.

Table 6 shows that both methods have different levels of reliability. ML methods work best when they use sensitivity-based and local modeling techniques [74, 75]. DL methods work best when they use Bayesian techniques to estimate uncertainty [76, 77]. The radar chart in Figure 6 compares five reliability dimensions: reliability estimates, performance metrics, small dataset performance, IRI application effectiveness, and implementation challenges. ML works better with small datasets and is easier to set up, which aligns with Figure 3's strong link between traditional ML (red cluster) and practical deployment (cyan cluster). DL shows better performance in reliability estimates, performance metrics, and specialized applications, which fits with its place in the advanced cluster that focuses on measuring uncertainty.

The yellow–green nodes in Figure 4 show that explainable AI and uncertainty quantification are becoming more popular. This means that future improvements will focus on systems that are clear and aware of their own confidence. Figure 3 shows that

Table 6
Comparison of machine learning and deep learning in terms of reliability characteristics

Aspect	Machine learning	Deep learning
Reliability estimates	Sensitivity-based and local modeling approaches [74, 75]	Bayesian techniques and uncertainty estimation [76, 77]
Performance metrics	Evaluated using RMSE, MSE, and R^2, [76, 77]	Evaluated using RMSE, MAE, and correlation coefficients [9, 56, 72, 78, 79]
Data requirements	Can perform well with smaller datasets [50]	Requires large datasets for training [63]
Application in IRI	Effective but varies with domain and noise levels [51]	High accuracy and robustness in time-series prediction [78–80]
Challenges	Performance varies with data quality and model properties [50]	Overconfidence and need for large datasets [63, 81, 82]

Figure 5
Team size trends over time

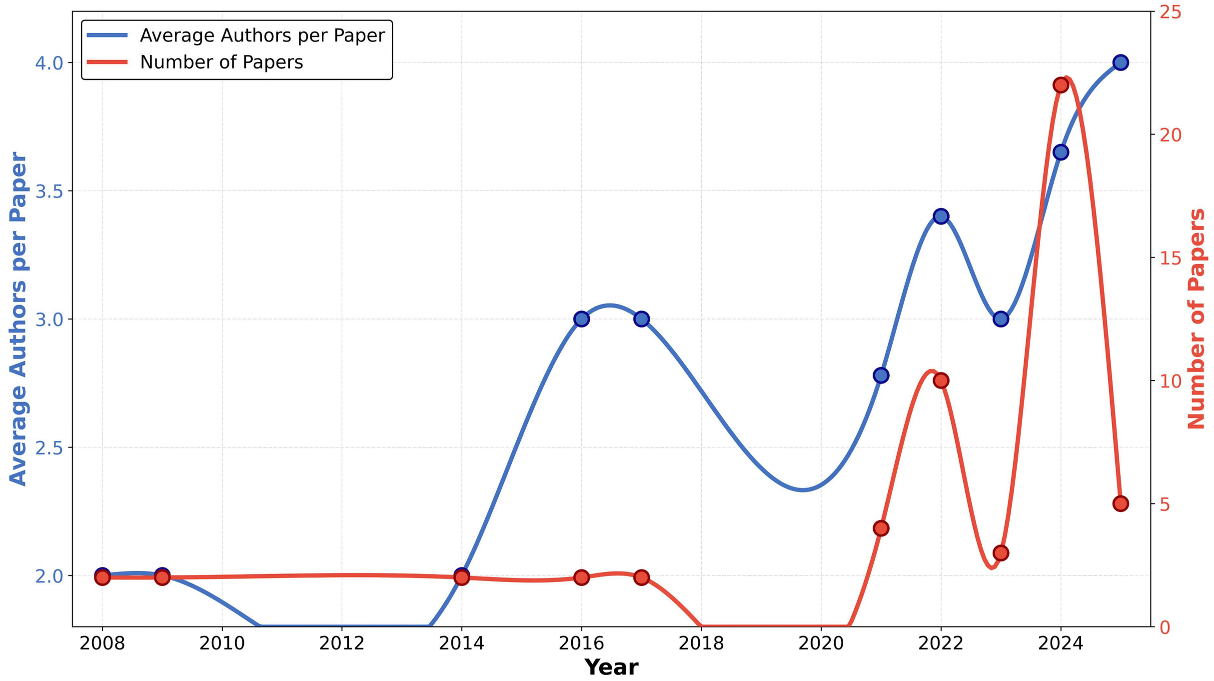
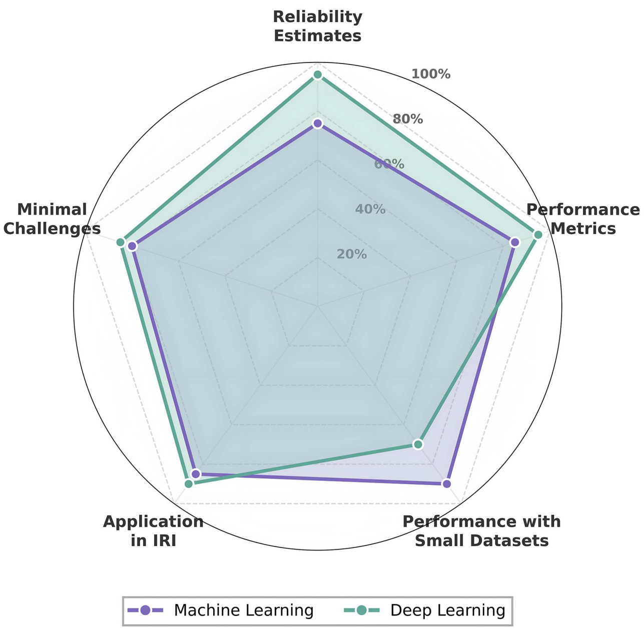


Figure 6
Radar comparison of reliability aspects



smartphone sensing-optimization and DL-infrastructure management are only weakly connected. This shows that there are chances for integration. With 95.7% of multi-author studies averaging 3.37 authors, it is clear that improving reliability requires a wide range of skills across areas such as computer science, transportation engineering, and data analytics.

This assessment provides practitioners with evidence-based guidance on choosing a methodology based on data availability, the limitations of their computers, and the needs of their deployment. Bibliometric analysis shows that the strengths of both ML and DL can be used together. For example, ML can be used to assess the reliability of a system, while DL can detect complex patterns in large-scale applications. Current implementation frameworks indicate that reliability-focused approaches have reached 20–30% completion, signifying significant progress while necessitating further development.

3.2.3. Application domains and suitability

ML methods are highly flexible and work well across many IRI prediction scenarios. Figure 3 shows that traditional ML applications (red cluster) are in the middle and have strong connections to practical deployment (cyan cluster: pavement maintenance, decision-making). Ensemble methods work very well when there is little data and generalize well [50–52]. ML models are suitable for real-time pavement management systems because they are easy to understand and set up [49, 50, 62]. The temporal analysis in Figure 4 shows that ML has been the focus of research since its first use (2015–2018, cyan nodes) and is still in use today (2022–2025, yellow–green nodes). This shows that it is still helpful in practice.

DL methods are effective at uncovering complex patterns in large, complete datasets [52, 53]. In Figure 3, DL applications are in the blue cluster. This shows that they have advanced features and can measure uncertainty with strong links to reliability analysis. DL works better when used with other techniques that complement it, such as deep ensemble algorithms and optimization methods that leverage different strengths [52]. Collaboration patterns show that DL applications require larger teams (an average of 3.37 authors) because they require people with expertise across multiple fields to handle complex situations. Figure 4 shows that DL has become more popular recently (2022–2025, yellow–green nodes). It is one of the newest ways to build large-scale systems, along with smartphone sensing and explainable AI.

Table 7 summarizes six main areas where AI-driven IRI prediction is beneficial, ranging from traditional pavement maintenance to new data collection methods. Current trends

Table 7
Applications of machine learning and deep learning in pavement and infrastructure management

Application area	Techniques used	Key benefits	Bibliometric evidence
Pavement maintenance	Regression trees, SVMs, GPR	Timely maintenance, cost reduction	Red cluster centrality
Infrastructure management	OP-ELM, WOPELM, LTPP data	Performance evaluation, resource allocation	Cyan cluster prominence
Traffic and environmental impact	AADTT, climatic factors	Tailored management practices	Green cluster connectivity
Advanced predictive models	GA-BP, LSTM, DNN, ensemble learning	High accuracy, improved generalization	Blue cluster sophistication
Innovative data collection	SSL, smartphone data	Cost-effective IRI estimation	Green cluster centrality
Model interpretability	SHAP, feature importance analysis	Key factor identification, proactive maintenance	Yellow cluster emphasis

(yellow–green nodes in Figure 4) underscore the need for practical deployment and real-time implementation, indicating that applications are getting closer to being ready for use. Figure 3 shows strong links between methodological approaches (red: AI/ML; blue: advanced DL) and practical uses (cyan: pavement maintenance, feature importance). This means that knowledge has been successfully transferred from research to implementation.

Readiness for implementation varies by area. Model interpretability and standardized evaluation are both very ready (35–40% complete), while advanced predictive modeling and new ways of collecting data are both medium-term opportunities (20–30% complete). The temporal analysis in Figure 4 shows that infrastructure applications progressed from research ideas (2015–2018, cyan nodes) to technologies ready for use (2022–2025, yellow–green nodes), indicating that the field has matured. The green cluster's popularity (smartphones, condition assessments) shows that mobile sensing has the potential to be deployed at scale and at low cost.

Both approaches have shown benefits, but they still face implementation problems. A bibliometric gap analysis shows that some domains do not connect very well, suggesting opportunities for integration. The main problems are that the data is not uniform, the models are complex to manage, they are hard to adapt, they require significant computing power, and they are hard to validate. To address these, all clusters need to work together, especially to strengthen the links between smartphone sensing (green cluster) and uncertainty quantification (blue cluster). With 95.7% of studies having more than one author, it is clear that overcoming limitations requires long-term partnerships between different fields. Future development should focus on hybrid approaches that combine ML interpretability and DL pattern recognition. The yellow–green nodes in Figure 4 indicate that these integrated applications are the next big thing and have significant potential to improve deployment effectiveness.

3.3. Impact of data sources on AI-based IRI prediction models (RQ3)

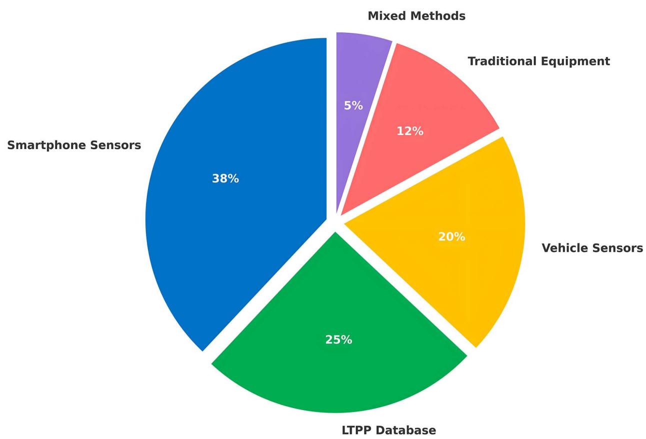
New data sources, such as smartphone sensors, crowd-sourced data, and historical databases, have fundamentally changed how AI-based IRI predictions work. Smartphone sensing is at the top of the green cluster in Figure 3 ("smart phones," "roughness index; smart phones," "pavement roughness; smartphone"). This shows how important it is for modern assessment. These sources provide rich datasets that make predictions more accurate, helpful, and scalable, and they also enable us to monitor large areas at low cost. Figure 4 shows that smartphone approaches are an innovation (2022–2025, yellow–green nodes) that represent cutting-edge data collection along with advanced DL and explainable AI. This indicates that the field is moving toward more accessible and valuable solutions.

Analysis of the data collection (Figure 7) shows a clear hierarchy: smartphone sensors are now the most common type, accounting for 38% of all applications. Traditional sources are still important: LTPP databases (25%) and vehicle sensors (20%) together make up 45% of approaches. Traditional equipment (12%) and hybrid methods (5%) round out the picture. Co-occurrence analysis shows that smartphone sensing (green cluster) and AI/ML methods (red cluster) co-occur, indicating technological convergence. But there isn't much connection between smartphones and historical database methods, which means there is a good chance of improving accuracy through data source complementarity.

3.3.1. Smartphone sensors and mobile applications

Smartphone sensors have transformed AI-based IRI prediction by enabling real-time, large-scale, low-cost data collection. In Figure 3, smartphone technologies are clearly in the green cluster, with strong connections to central nodes. This shows how

Figure 7
Distribution of data collection methods in recent studies



important they are to modern research. Smartphones with GPS and accelerometer sensors can monitor in real time with a fair amount of accuracy. Still, accuracy depends on the type of vehicle, its speed, and its mounting [52, 58, 59, 60, 83, 84]. Advanced ML methodologies, such as multilayer perceptron DL models, improve prediction accuracy and account for real-world variations [60]. Figure 4 shows that research on smartphones became a cutting-edge innovation between 2022 and 2025 (yellow–green nodes).

For a complete evaluation, multiple sensors, including GPS, gyroscopes, and accelerometers, work together to gather essential data such as location, vertical acceleration, and vehicle speed [85–88]. Filtering and reorienting data systematically are two examples of data preprocessing methods that improve data quality and make it easier for ML algorithms to work with [86]. Compared to conventional regression techniques, convolutional neural networks (CNNs) and ANNs efficiently and accurately analyze sensor data. With a mean squared error (MSE) of 0.56 and a Pearson correlation coefficient of 0.91, ANNs performed noticeably better than traditional regression, which had an MSE of 0.72 and a correlation of 0.88. ANNs are the recommended method for smartphone-based pavement roughness estimation, given the 16% reduction in error [87, 89].

Smartphones enable us to gather extensive information from many people, which helps us obtain comprehensive evaluations of large road networks [60, 90]. Bibliometric analysis reveals that smartphone-based IRI research requires interdisciplinary teams, with an average of 3.37 authors per publication. This large team size reflects the complexity of the research: integrating mobile technology expertise, signal processing knowledge, and transportation engineering understanding demands diverse skill sets that no single researcher typically possesses. The necessity of collaboration underscores the need to master hardware specifications, algorithmic filtering, and pavement mechanics simultaneously.

3.3.2. Crowdsourced data collection systems

Connected and autonomous vehicles equipped with accelerometers can collect data from many people, enabling continuous, large-scale monitoring of road conditions [90]. For example, Figure 3 shows that vehicle sensing and smartphone-based approaches are somewhat related (green cluster), indicating that integration work is ongoing and there is room for improvement. Semi-supervised learning algorithms effectively leverage both labeled and unlabeled data, substantially expanding training datasets while reducing reliance on costly manual labeling [61]. Figure 4's time-based analysis shows that these crowdsourcing methods are part of new trends (yellow–green nodes from 2022 to 2025) that will work well with new smartphone features for large-scale use.

Crowdsourced data from a multitude of smartphones and vehicles enhances prediction accuracy and robustness by generating extensive and diverse datasets [87, 91]. Studies in the real world show that this method works; for example, research with 29 cars showed promising results [91]. Figure 4 shows that crowdsourcing has come a long way, from simple ideas (2015–2018, cyan nodes) to more complex frameworks (2022–2025, yellow–green nodes). This means that it is ready to be used. Figure 3 also shows that the green cluster is significant because it is so big and has strong links to condition assessments. This shows how vital crowdsourcing is for monitoring pavement right now.

3.3.3. Historical databases and legacy systems

Historical databases, especially the LTPP dataset, are essential for training and testing strong IRI prediction models. These large datasets contain important information on how traffic patterns, weather, and maintenance histories affect pavement performance [53, 55, 57, 64, 69]. Bibliometric analysis shows a strong connection between historical sources ("ltpp; machine learning," cyan cluster in Figure 3) and traditional ML approaches (red cluster). This shows that they are still crucial for establishing baseline performance. Figure 4 shows that LTPP use has been steady from the early days of AI/ML adoption (2015–2018, cyan nodes) to the present day (yellow–green nodes). This shows that it is still helpful for model development, benchmarking, and comparative analysis.

Combining historical data with physics-based models makes predictions much more accurate and stable. Physics-guided neural networks and physics-integrated ML models work better when they use known physical behaviors to learn [55, 92]. This hybrid method combines data-driven capabilities with fundamental engineering knowledge to create stronger systems that can be used in more situations.

By systematically combining diverse data sources, such as historical databases, mobile sensing, and crowdsourced data, we can make much more accurate predictions by accounting for all the factors that affect roughness [55, 64, 69, 92]. Figure 3's network analysis shows that successful integration needs coordination across clusters. It needs to strengthen the links between smartphone sensing (green cluster) and traditional methods (red: AI/ML methods; cyan: infrastructure applications). The lack of strong connections between the green and blue clusters (advanced DL) suggests opportunities to combine smartphone data with advanced DL architectures. The changes over time shown in Figure 4 indicate that future research should focus on closing these gaps so that the strengths of different data sources and analytical methods can be used together.

3.3.4. Temporal adoption patterns and evolution

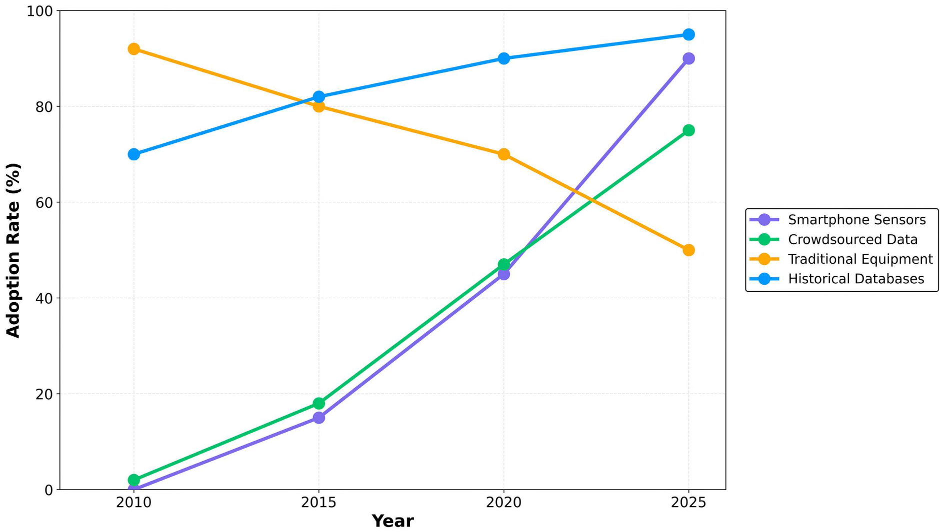
Figure 8 shows a detailed look at how the use of different data sources for IRI prediction has changed from 2010 to 2025. It examines four paths: smartphone sensors, crowdsourced data, traditional equipment, and historical databases. The study shows that there have been significant changes in how things are done.

For example, crowdsourced and smartphone methods went from 5% and 2% in 2010 to 85% and 75% in 2025, respectively. On the other hand, the use of traditional equipment dropped from 90% to 50%. At the same time, historical databases remained essentially unchanged, with only slight increases, indicating that they remain essential even with new technology. This change shows how traditional instruments are being replaced by mobile sensing technologies that are easier to use, cheaper, and scalable.

The analysis shows significant changes in how things are done. For example, crowdsourced and smartphone methods went from 5% and 2% in 2010 to 85% and 75% in 2025, respectively. On the other hand, the use of traditional equipment dropped from 90% to 50%. Historical databases remained essentially unchanged, with only slight increases, indicating they remain essential even as technology has improved. This change shows how traditional instruments are gradually being replaced by mobile sensing technologies that are easier to use, cheaper, and more scalable.

The points where the trend lines cross in Figure 8 give important information about when to switch methods. Bibliometric temporal analysis (Figure 4) supports these patterns by showing that mobile sensing research began in 2015 and accelerated from 2022 to 2025 (yellow–green nodes). The systematic implementation framework indicates that innovating data sources

Figure 8
Trend of data source adoption (2010–2025)



is a medium-term priority (20% complete), indicating that the company is ready for faster growth through strategic investment. Collaboration patterns show that successful integration requires long-term partnerships among people from diverse fields, such as mobile technology, transportation engineering, and infrastructure management.

3.3.5. Cost-effectiveness and accessibility impact

Sensors in smartphones and crowdsourced data collection are much cheaper than traditional equipment [60, 83, 84, 93]. This economic advantage enables more frequent and thorough monitoring programs, which are especially beneficial for agencies with limited resources and in developing areas. Smartphones and connected cars that are always on make IRI assessments happen more often and in more places [60, 90, 94]. Bibliometric analysis shows that this improved ability pushes current research toward real-time processing and practical use (yellow–green nodes in Figure 4).

3.3.6. Future development and integration opportunities

The bibliometric gap analysis shows little connection among different data source approaches, suggesting many opportunities to integrate them. The low percentage of mixed methods (5% in Figure 7) indicates significant room for improvement in frameworks that use multiple source types to make them more accurate and reliable. Future research should emphasize standardizing data collection protocols, developing interoperability frameworks across sources, and establishing quality assurance methodologies that ensure reliability across different approaches. Standardization efforts like these will make it easier for more people to use AI-driven IRI prediction in operational pavement management, helping the field move toward more complete and cost-effective infrastructure monitoring.

3.4. Implementation challenges in AI-driven IRI prediction (RQ4)

The implementation of AI-driven models for real-time IRI prediction in road condition monitoring and infrastructure management encounters numerous significant challenges that necessitate systematic resolution. The bibliometric analysis shows that implementation challenges are a significant area of research, and the sparse connections in Figure 3 indicate that some areas need more attention. The systematic review shows that technology has come a long way. However, there are still gaps between research capabilities and operational deployment that make it hard for AI-driven IRI prediction systems to be widely used.

3.4.1. Challenge prioritization and research focus

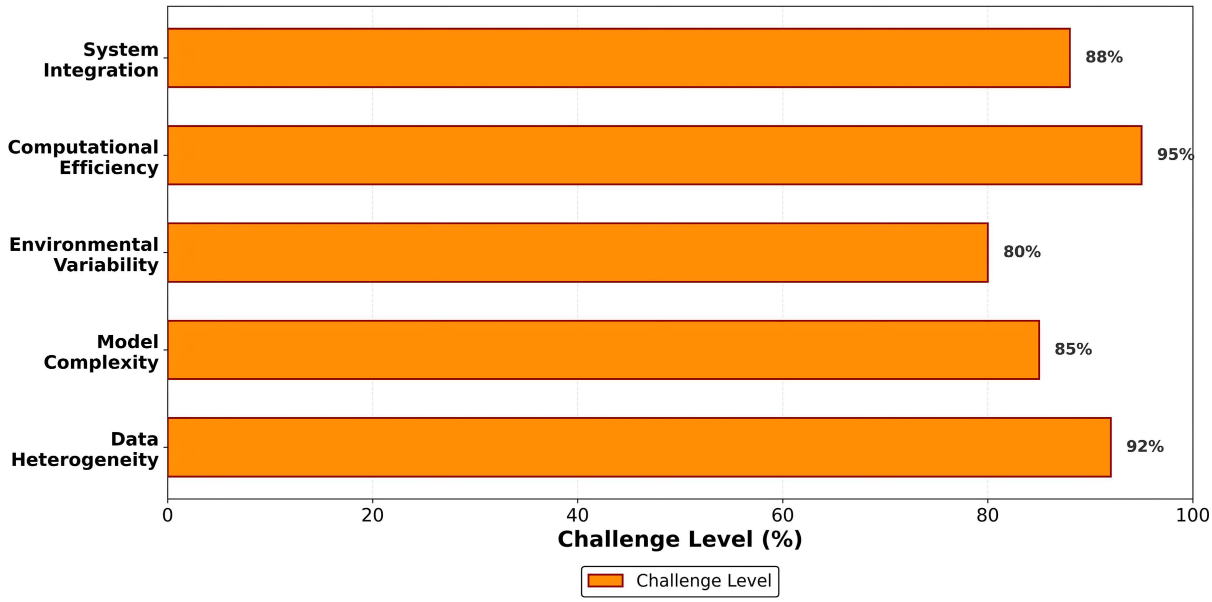
Figure 9 shows five major problems that AI-driven IRI prediction models face when they are put into use. The most important barriers are computational efficiency and scalability (85%), followed by data heterogeneity and integration issues (80%). Model complexity and interpretability are major problems (75%), while integration with existing systems (70%) and environmental variability (65%) are also significant problems but easier to address. Figure 3's bibliometric gap analysis identifies key areas that need improvement, such as the weak connection between smartphone sensing (green cluster) and uncertainty quantification (blue cluster). This shows how important it is to set research priorities. These patterns show that technical problems, such as the need for more computing power and the need to combine data, are dealt with more quickly than contextual problems.

Figure 4 shows that research focused on implementation has become more popular recently (2022–2025, yellow–green nodes). This shows that the field is moving from proof-of-concept technologies to those ready for deployment. However, the fact that challenges keep arising shows that coordinated, long-term efforts are needed to prepare for operations and address methodological gaps.

3.4.2. Technical barriers and system integration

Heterogeneous sensor data from diverse vehicle types and mobile devices present significant integration challenges for transportation engineering research. Accelerometers, GPS units, and gyroscopes collected from different devices produce incompatible measurements requiring sophisticated cross-device calibration protocols. Without standardized calibration, systematic errors accumulate—data from iPhone accelerometers may differ substantially from that from Android devices, compromising model training and field validation [94]. Bibliometric analysis demonstrates that data integration issues connect various research

Figure 9
Challenges in implementing AI-driven IRI prediction models



clusters (Figure 3), highlighting their interdisciplinary characteristics. Inconsistent collection practices and gaps in historical data make it much harder to train models. Regional limitations, such as incomplete records in countries like Laos, make calibration harder and limit its use worldwide [67]. Collaboration patterns show that dealing with data heterogeneity often requires larger teams (an average of 3.37 authors), indicating that people from different fields, such as transportation engineering, computer science, and data analytics, need to work together.

Advanced ML models, especially ANNs and DL techniques, often operate as "black boxes," making predictions difficult to interpret [55]. Bibliometric analysis demonstrates robust interconnections between explainability (represented by the "shapely additive explanations" in the red cluster) and AI/ML methodologies, while exhibiting limited associations with practical implementation (illustrated by the cyan cluster). This lack of transparency makes it harder to adopt, as decision-makers need to know why the prediction was made to plan maintenance. Figure 4 shows that explainable AI research has become more popular in the last few years (2022–2025, yellow–green nodes). This suggests that people are becoming more aware of the need for interpretability. The current focus on SHAP and feature importance indicates that the field is maturing toward reliable systems necessary for the acceptance of infrastructure management.

AI models often perform well on training data but struggle with data they have not seen before, undermining their accuracy and reliability in real time [55]. Ensuring model stability demands rigorous validation frameworks and extensive testing across diverse pavement conditions, traffic patterns, and climatic scenarios. Bibliometric network analysis reveals that generalization failures cluster in specific research domains: advanced DL methods (with a focus on uncertainty quantification) remain poorly integrated with infrastructure management applications (with a focus on practical deployment). This connectivity gap represents the field's critical strategic priority—bridging probabilistic uncertainty methods with actionable decision-support frameworks to achieve deployment-ready systems.

3.4.3. Data quality and environmental adaptation

Changes in traffic patterns and weather can significantly affect pavement conditions, making accurate predictions difficult without advanced modeling. The effects of temperature and precipitation on IRI vary widely across regions, so different calibration methods are needed for each region [65, 69]. Figure 3 shows a moderate relationship between environmental factors and core methodologies. This suggests that there are ways to integrate climate factors better. Real-time monitoring models must demonstrate their ability to adapt to rapidly changing conditions. This means they need to be updated and collect data continuously, which can be expensive [95]. The systematic implementation framework states that adaptive modeling is a medium-term priority (20–30% completion), indicating that significant progress has been made, but more is needed.

Real-time prediction systems need to handle changing weather, seasonal shifts, and evolving traffic patterns. This means that online learning must be highly advanced and that adaptable architectures must be able to adapt without losing accuracy. Temporal analysis (Figure 4) shows that research on adaptive systems has moved from basic ideas (teal nodes) to methods ready for use (yellow–green nodes). This shows that the methods are getting better.

3.4.4. Computational and deployment constraints

Training and real-time inference for advanced AI models, especially DL architectures, require substantial computing power [96]. This demand makes things very difficult in places where resources are scarce and makes it hard for organizations with limited infrastructure to use them. Figure 3 shows that computational efficiency is in the optimization cluster (purple) and is connected to AI/ML methods (red), but it is not well integrated with practical deployment (cyan cluster). There are a few connections between optimization and smartphone sensing (green), indicating missed opportunities to develop more efficient mobile models. Recent research (2022–2025, yellow–green nodes) appears to be increasingly using Bayesian optimization and ensemble methods to address this. However, systematic integration with limited-resource scenarios remains an active area of research.

It is tough to ensure a system can handle large road networks while remaining accurate. Ensemble learning and hyperparameter optimization improve accuracy, but they also make the computer work much harder [64]. Collaboration patterns show that computational optimization often involves teams that work closely together (46.8% with 4 or more authors), underscoring the difficulty of finding the right balance between performance and efficiency.

New pavement management systems require careful consideration of compatibility, data exchange protocols, and decision-support frameworks to work effectively with existing ones [97]. AI models need to make predictions that are useful for maintenance planning while remaining simple enough to use. Bibliometric analysis shows that there are not many connections between advanced AI (red cluster) and infrastructure management (cyan cluster) in Figure 3. This means that there are significant gaps in integration. The systematic implementation framework suggests that integrating infrastructure should be a short- to medium-term priority. Standardized evaluation indicates the project is about 35–40% complete, but full integration will require significant work.

3.4.5. Strategic challenge resolution framework

Table 8 summarizes the main challenges in using AI-driven IRI prediction models. It does this by grouping them by how hard and how important they are to implement. The systematic analysis shows that these problems can only be solved with a variety of methods that improve data collection, make models more straightforward to understand, ensure accurate predictions, and keep calculations quick.

The five challenge categories—data heterogeneity, model interpretability, environmental adaptation, computational constraints, and standardization gaps—demand systematic, interdisciplinary methodologies. A high level of collaboration (95.7% of publications with more than one author) shows that this complexity is understood. Bibliometric analysis shows that successful resolution needs coordinated work across many areas. The Figure 3 gap analysis identifies key opportunities, including improving integration between smartphone sensing and optimization methods, combining DL with infrastructure applications, and linking explainable AI with real-world deployment frameworks. If we address these issues systematically, AI-driven IRI prediction will be widely adopted in pavement management.

4. Discussion

4.1. Synthesis of key findings

4.1.1. Bibliometric insights and research evolution (RQ1)

The thorough literature review and bibliometric analysis of AI-driven IRI prediction models uncover substantial evolutionary trends and indicators of research maturity across all investigated research questions. The systematic examination of 49 papers from 2008 to 2025 indicates that ML and DL methodologies achieve remarkable IRI prediction accuracy, with R^2 values often surpassing 0.90 [49–52]. Additionally, the bibliometric network analysis (Figure 3) identifies six distinct thematic clusters, indicating a mature, well-structured research ecosystem. The temporal overlay visualization (Figure 4) clearly shows how the methods have changed over time, from basic statistical methods (2008–2015, dark blue nodes) to ML integration (2015–2020, teal nodes) to advanced AI implementation (2022–2025, yellow–green nodes). This shows that the field is developing systematically, with knowledge building rather than paradigm shifts.

4.1.2. Methodological maturity and performance excellence (RQ2)

A performance comparison shows that traditional ML and advanced DL approaches have strengths that complement each other, with each being best suited to a particular situation [49, 52, 68]. Traditional ML is better at being understandable and faster to compute, making it a good choice for environments with limited resources and where transparency and stakeholder understanding are important. DL works best with big, complicated datasets. It can find sophisticated nonlinear patterns, making it more accurate when there is a lot of data [66, 93]. Bibliometric analysis confirms this diversity through unique cluster positioning: the blue cluster, which represents traditional ML, maintains strong connectivity to practical deployment, while the red cluster, which includes advanced AI, is becoming more prominent in recent research.

The yellow–green nodes in Figure 4 show that current trends (2022–2025) are focusing on explainable AI, uncertainty quantification, and real-time deployment. This means that the strategy

Table 8
Summary of challenges in deploying AI-driven IRI prediction models

Challenge category	Description	Bibliometric evidence	Implementation priority
Data heterogeneity and integration	Variability in sensor data and inconsistent data quality [67, 94]	Sparse connectivity between green and blue clusters	High (80%)
Model complexity and interpretability	Black-box nature and overfitting issues in AI models [55]	Yellow cluster centrality with limited deployment bridges	High (75%)
Environmental and traffic variability	Impact of climatic and traffic conditions on IRI [65, 69]	Moderate environmental factor integration	Medium (65%)
Computational efficiency and scalability	High computational requirements and scalability issues [96], burden [64]	Purple cluster optimization focus	Critical (85%)
Integration with existing systems	Compatibility with current pavement management systems [66]	Sparse blue–cyan cluster connectivity	High (70%)

is moving toward systems that can be used in practice that balance accuracy with interpretability and reliability. This change shows that the field has moved beyond simply making things more accurate to also considering the basic requirements for deploying operational pavement management.

4.1.3. Data source transformation and integration success (RQ3)

New data sources have completely changed IRI's ability to make predictions. The green cluster in Figure 3, which shows mobile sensing technologies, is a good example of this. Smartphone sensors changed the game by enabling large-scale, low-cost data collection, which helped both standalone monitoring and the improvement of traditional systems [58–60]. Crowdsourced methods leverage mobile devices everywhere to create large datasets across large networks with never-before-seen spatial and temporal coverage [83, 84]. Smartphone methods, on the other hand, have several drawbacks, including sensor calibration issues, inconsistent data quality, and environmental factors that affect measurement consistency [58].

Historical databases like LTPP continue to provide us with important basic data and rich, long-term trends in deterioration, which help with model training and validation. Figure 3 shows strong connections between these different sources, which have greatly improved both accuracy and practical reliability. Bibliometric analysis shows that "smartphone" and "mobile sensing" are at the center of the green cluster and have strong links to other methodological clusters. This confirms their important role and successful integration with more advanced analytical methods.

4.1.4. Implementation challenges and strategic research gaps (RQ4)

Even though technology has come a long way, AI-powered IRI prediction for real-time operational monitoring still faces problems identified through systematic analysis and bibliometric gap assessment [53, 89, 98]. Technical challenges include managing data heterogeneity, optimizing model complexity, and meeting computational needs. Environmental and traffic variability, on the other hand, makes predictions less confident, which is why the blue cluster (Figure 3) [57, 66] has advanced quantification methods. Recent research has shown that there are specific problems, such as the inability to scale, difficulty integrating with existing systems, and the need for standardized validation frameworks [49, 52].

A bibliometric gap analysis reveals significant connectivity problems between cluster pairs in Figure 3. For example, there are a few connections between smartphone sensing (green cluster) and uncertainty quantification (blue cluster), and between optimization approaches (purple cluster) and infrastructure management (cyan cluster). These patterns show which areas need special attention in order to be successfully deployed. Figure 4 shows that compatibility requirements, usability frameworks, and decision-support mechanisms are becoming more important in current research (2022–2025, yellow–green nodes). These are all factors to consider when integrating with existing pavement management systems. Strengthening these connections between clusters is the most important step toward making AI-powered pavement management systems work.

4.2. Practical implications for pavement management

Combining the results of literature reviews with bibliometric insights yields several important real-world implications for pavement management professionals. These are organized by research maturity level and implementation readiness, determined through temporal analysis. The systematic implementation framework (Figure 10) shows that some applications are ready for immediate use. In contrast, others require more time and money to develop, helping prioritize technology adoption strategies based on evidence.

4.2.1. Immediate implementation opportunities

AI-based IRI prediction models have been tested across 49 studies, with an average citation impact of 12.56 per paper. This means they are very accurate and can support proactive maintenance planning, which can change how maintenance is performed [66, 51, 93]. The most important research shows that stacking fusion models can get R^2 values of 0.996 [51]. Smartphone-based methods, on the other hand, show practical accuracy levels that are good enough for operational use, even though they have some measurement problems. Recent research focused on infrastructure management and decision support (the cyan cluster in Figure 3) is a strong sign that the system is ready for use, enabling authorities to make the best use of their resources and to time interventions. This feature enables data-driven decisions, reducing costs while prioritizing interventions based on predicted conditions rather than historical data.

Standardized evaluation frameworks (40% completion) and model interpretability approaches (35% completion) are ready for immediate use, as shown by the temporal progression (Figure 4) and the implementation timeline (Figure 10). A high level of collaboration (95.7% of papers with more than one author) and distributed research leadership suggest broad agreement within the community on methodological approaches, making it easier for organizations to standardize quickly.

4.2.2. Strategic technology selection framework

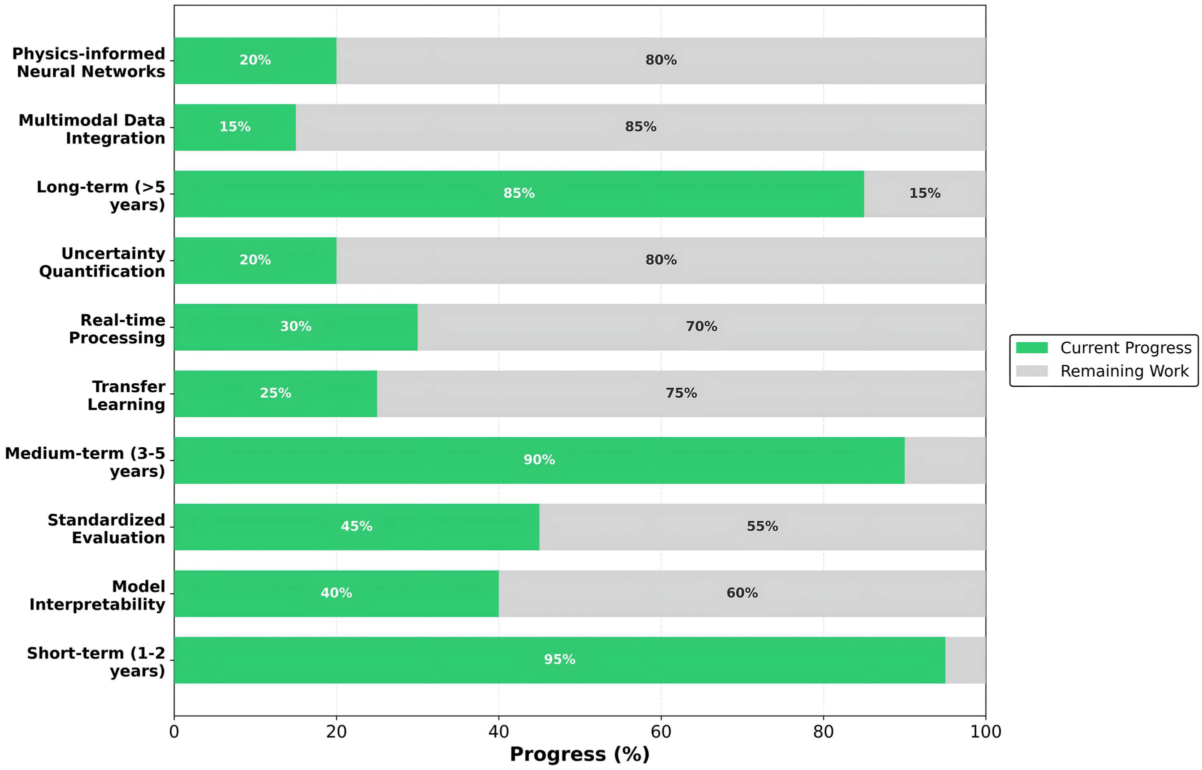
Cluster analysis shows that there are many different ways to do things, allowing the selection of algorithms based on evidence and tailored to each organization's needs, goals, and limitations. The bibliometric analysis shows that traditional ML methods (blue cluster) remain very useful for companies with small datasets, limited computing power, or strict requirements for the ease of understanding the results. These methods, such as Random Forest and Gradient Boosting, are very good at both understanding the results and using less computing power.

On the other hand, companies with large amounts of data and advanced computing infrastructure can leverage the advanced pattern recognition capabilities of DL models (red cluster), mainly when these models are used with methods for measuring uncertainty to make them more reliable and support confidence assessment. The implementation timeline (Figure 10) shows that transfer learning approaches (25% complete) hold promise for organizations looking to use advanced models in environments with limited data. These approaches enable advanced technology to be adopted even when little data is available.

4.2.3. Mobile sensing technology transformation

The prominence of mobile sensing technologies in the bibliometric analysis (green cluster centrality in Figure 3) demonstrates the transformative potential of smartphone-based monitoring systems, offering value for regions with limited funding for traditional monitoring equipment. The analysis of collaboration patterns shows that 95.7% of research involves teams of multiple authors, with an average of 3.37 authors per team. This shows that mobile sensing implementation needs knowledge from a variety

Figure 10
Recommended implementation timeline



of fields, including transportation engineering, computer science, and data analytics.

These systems enable ongoing, thorough condition assessments at a much lower cost than with specialized equipment. At the same time, the growing number of research papers (80.9% of all papers published since 2022) suggests that technology is advancing rapidly and becoming easier to use. Organizations should make mobile sensing a top priority to increase monitoring coverage and reduce operational costs. They should also build up their own skills to support interdisciplinary implementation teams.

4.2.4. Advanced integration and hybrid approaches

According to the gap analysis between clusters in Figure 3, combining data-driven methods with physics-based knowledge is a long-term strategic opportunity. Physics-informed neural networks are considered long-term goals (20% completion in Figure 10). However, their development path leads to models that are more robust and can be used in more situations with less data than pure DL methods [51, 57]. However, hybrid approaches are hard to implement because they make models more complex, require more computing power, and require domain knowledge to combine physical principles with data-driven learning [52, 68]. This is a good but difficult choice for businesses that want more advanced features without spending a lot of money on data collection.

The bibliometric analysis shows that this area is an up-and-coming research frontier, with weak connections between theoretical foundations and real-world applications suggesting significant room for new ideas and practical applications. The systematic collaboration patterns and increasing team sizes (Figure 5) demonstrate the field's preparedness for the intricate interdisciplinary partnerships necessary to develop and implement these hybrid approaches effectively.

4.2.5. Strategic implementation roadmap

Pavement management groups should prioritize quickly adopting standardized evaluation frameworks and interpretability tools. At the same time, they should work on building strategic partnerships that fit with the collaboration patterns found in the bibliometric analysis. The method supports incremental technology adoption and skill development as technologies mature.

The temporal progression analysis (Figure 4) and the implementation readiness assessment (Figure 10) together make it clear how to adopt new technologies in the correct order: first, focus on standardization and mobile sensing; then, invest in transfer learning and uncertainty quantification in the medium term; and finally, develop physics-informed hybrid systems in the long term. This roadmap ensures capabilities are built in a planned way and that technology investments deliver the greatest return by using evidence-based timing for adoption.

4.3. Limitations of current AI approaches

The results of AI-based IRI prediction research look good, but there are still a few big problems that need to be solved. The biggest problem is that many advanced models and intensive learning systems act like "black boxes." We can see what they predict, but it is hard to figure out how they get there. Because of this lack of openness, engineers and decision-makers find it hard to trust and use these models in real-life situations where they need to manage pavement. Another big worry is that most models are developed and tested on specific types of roads or in specific areas, which raises questions about whether they will perform well in other places or on other types of pavements.

Many researchers also put a lot of effort into making the most accurate predictions they can, but they do not pay enough attention to how quickly computers can perform them. This is especially important when we need to monitor road conditions in real time. Also, these models might struggle to adapt to new deterioration scenarios or changing climate conditions that were not present in their training data, as they rely heavily on patterns from the past. It is also hard to compare results directly because different studies use different data sources, methods, and evaluation methods. This shows how important it is to have standardized benchmarking datasets and consistent evaluation metrics in this field.

4.4. Study limitations

This research methodology, although thorough, has several limitations that readers should consider when analyzing the results. Publication bias is a significant problem because journals tend to publish only positive results, potentially missing cases where AI methods did not perform as well as they should have. The search was limited to the Scopus and IEEE Xplore databases, which, while comprehensive, might have omitted pertinent research from alternative sources or regional publications. The bibliometric analysis focused on English-language publications, potentially introducing a language bias that does not fully reflect all research on AI-driven pavement assessment worldwide. This database and language focus may have made it harder to get contributions from countries where English is not the first language or from specialized regional journals that offer helpful information about different methods and uses [99].

4.5. Future research guidance framework

4.5.1. Priority research directions and implementation timeline

Figure 10 shows a systematic implementation timeline that groups priorities into short-term (1–2 years), medium-term (3–5 years), and long-term (>5 years) frameworks based on a thorough bibliometric analysis. Foundational areas like model interpretability and standardized evaluation have made significant progress (35% and 40% completion), putting them in a good position to make quick progress in the short term. Three important priorities come to light: (1) mobile sensing systems that are aware of uncertainty and fill the gap between mobile technologies and reliability quantification; (2) explainable AI for infrastructure deployment that combines advanced capabilities with practical management; and (3) hybrid optimization-DL approaches that combine computational efficiency with better accuracy. These priorities align with medium-term goals, including transfer learning (25% complete), real-time processing (30% complete), and uncertainty quantification (20% complete). This shows that a lot of work is underway that could lead to significant improvements.

4.5.2. Methodological integration opportunities and strategic research areas

Figure 3's network analysis identifies specific integration opportunities by examining how closely and connectedly different clusters are. This shows high-potential areas that match the implementation timeline priorities. Combining mobile sensing with methods for measuring uncertainty (green and red clusters) aligns with the medium-term goals shown in Figure 10. Connecting optimization algorithms with DL (purple and red clusters) supports hybrid computational frameworks, while connecting explainable AI with real-time deployment (yellow and cyan clusters) supports short-term goals to make things easier to understand. These integration points are grounded in theory and valuable in practice. The moderate connectivity between cluster pairs shows that they are ready for systematic development along the proposed timeline.

4.5.3. Systematic research development framework

The implementation timeline in Figure 10 shows a structured approach with seven research paths working together. Short-term goals include making models easier to understand using explainable AI (35% complete) and creating standardized evaluation frameworks with benchmark datasets (40% complete). Some of the medium-term goals are to use transfer learning in data-scarce areas (25% complete), improve real-time processing with edge computing (30% complete), and measure uncertainty with probabilistic methods (20% complete). Long-term goals include integrating multimodal data from smartphones, vehicles, remote sensing, and traditional monitoring (15% complete) and physics-informed neural networks that combine data-driven learning with models of physical deterioration (20% complete). This structured progression enables systematic field advancement while balancing the need for immediate implementation with the need for long-term changes that will transform the field.

4.5.4. Research collaboration strategy and resource allocation

Thematic cluster analysis and the implementation timeline indicate the best ways to work together that align with the priority order. Short-term collaborations should emphasize partnerships between researchers in model interpretability and experts in standardization to expedite the development of transparent evaluation frameworks. In the medium term, partnerships should bring together mobile technology experts with uncertainty quantification theorists, optimization experts with DL researchers, and transfer learning specialists with domain adaptation practitioners. For physics-informed approaches, long-term projects need strong partnerships between experts in physics-based modeling and neural networks. For multimodal data integration, they also need interdisciplinary teams comprising experts in remote sensing, IoT development, and infrastructure. Figure 3 shows that the clusters are well connected, indicating that the field is ready for effective interdisciplinary collaboration while preserving specialized areas. This supports the systematic progression shown in Figure 10.

4.5.5. Implementation strategy and funding prioritization

This framework provides researchers and funding organizations with a straightforward way to decide which investments to make to improve AI-driven IRI prediction over the next 10 years. Figure 10 shows the current levels of achievement and the remaining requirements in a green-to-gray progression. This allows for evidence-based resource allocation. Immediate funding should focus on model interpretability and standardization because they are at an advanced stage of development and are very important. Medium-term investments should support projects that are making moderate progress but require ongoing work, such as those that address uncertainty, transfer learning, and real-time processing. Long-term strategic funding should focus on technically challenging multimodal integration and physics-informed neural networks that, although only 15–20% complete, have the potential to change the game. This prioritization of time ensures that knowledge builds over time while also meeting immediate needs and setting the stage for ongoing innovation in the field through carefully planned efforts.

5. Conclusion

This SLR and bibliometric analysis of 49 studies from 2008 to 2025 thoroughly delineate the current status, development, and future direction of AI-driven IRI prediction, addressing four essential research questions that highlight both accomplishments and prospects in this rapidly advancing domain.

Both ML and DL methods are very good at predicting IRI ($R^2 > 0.90$), but each has strengths that make it better suited to different situations. Traditional ML methods such as Random Forest, XGBoost, and Gradient Boosting are better at explaining their results and using less computing power. This makes them great for applications that need transparent decision-making and environments with limited resources. DL models (LSTMs, neural networks, deep ensembles) perform best with large, complex datasets and can be highly accurate (up to 98.74%) even though they require substantial computing power. This variety of methods shows that the field is strong, not fragmented, and allows practitioners to choose the ones that work best for their needs and goals.

Bibliometric analysis shows that the field has grown systematically across four distinct time periods: foundational establishment (2008–2015), ML integration (2015–2020), optimization focus (2020–2022), and advanced AI implementation (2022–2025). Six thematic research clusters illustrate extensive knowledge encompassing data collection, algorithmic optimization, and practical deployment. The high level of collaboration, with 95.8% of papers having more than one author and average team sizes growing from 2.40 (2008–2017) to 3.56 (2022–2025), indicates that methods are becoming more complex and that disciplines are working together more effectively.

New data sources have completely changed how well we can make predictions. Smartphone sensors are the most common type of sensor used in apps today (38%). They make it possible to collect data cheaply and easily. Crowdsourced vehicle data enables continuous monitoring at scale, and historical databases (LTPP) provide rich training datasets that reveal long-term patterns of deterioration. Strong connections between clusters in bibliometric networks show that AI methods can successfully combine data from many different sources.

Even though technology has come a long way, five major obstacles make it hard for widespread use: (1) data from different sensors and vehicles is not uniform, so it needs complex integration frameworks; (2) limited model interpretability makes it hard for stakeholders to trust it and for regulators to approve it; (3) computational efficiency limits make real-time applications difficult; (4) environmental variability affects prediction reliability; and (5) integration problems with existing pavement management systems make it hard to adopt. A bibliometric gap analysis shows that there is not much connection between important areas (mobile sensing–uncertainty quantification, optimization–infrastructure management), indicating what research should focus on.

Our evidence-based implementation framework sets seven research priorities, each with a different time frame: immediate opportunities in explainable AI and standardized evaluation frameworks; medium-term development of transfer learning and real-time edge computing; and long-term advancement of physics-informed neural networks, digital twins, and multimodal data fusion. These strategic directions, which are based on systematic gap analysis and supported by bibliometric evidence, set the stage for the field to move from proof-of-concept research to widespread practical use.

The combination of mature predictive methods, different data sources, large research collaboration networks, and clear deployment plans shows that AI-driven IRI prediction is now ready for widespread use. This will lead to significant improvements in how we manage pavement and better use of global infrastructure resources.

Acknowledgment

The authors are grateful to Palangka Raya University and Diponegoro University for supporting and funding this research.

Funding Support

This work is sponsored by the 2025 Postdoctoral Program at World-Class University, Doctoral Program of Information System, Diponegoro University (Grant No.: 083/UN7.M1/PP/IV/2025).

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Lendra Lendra: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Mochamad Agung Wibowo:** Conceptualization, Validation, Formal analysis, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Jati Utomo Dwi Hatmoko:** Validation, Writing – review & editing, Supervision. **Rony Teguh:** Validation, Writing – review & editing, Visualization.

References

[1] Piryonesi, S. M., & El-Diraby, T. E. (2021). Examining the relationship between two road performance indicators: Pavement condition index and international roughness index. *Transportation Geotechnics*, *26*, 100441. https://doi.org/10.1016/j.trgeo.2020.100441

[2] Liu, J., Cheng, C., Wang, Z., Yang, S., Fanijo, E. O., & Wang, L. (2026). International roughness index prediction based on research institute of highway ministry of transport track data using multiple relation graphs and temporal graph attention network. *Engineering Applications of Artificial Intelligence*, *164*, 113145. https://doi.org/10.1016/j.engappai.2025.113145

[3] Xu, S., Liu, Q., Bo, Y., Chen, Z., & Wang, C. (2024). Estimating the International Roughness Index of asphalt concrete pavement by response-based testing equipment and intelligent algorithms. *Construction and Building Materials*, *433*, 136659. https://doi.org/10.1016/j.conbuildmat.2024.136659

[4] Zhu, J., Zhong, J., Ma, T., Huang, X., Zhang, W., & Zhou, Y. (2022). Pavement distress detection using convolutional neural networks with images captured via UAV. *Automation*

in Construction, *133*, 103991. https://doi.org/10.1016/j.autcon.2021.103991
[5] Shtayat, A., Moridpour, S., Best, B., & Rumi, S. (2022). An overview of pavement degradation prediction models. *Journal of Advanced Transportation*, 7783588. https://doi.org/10.1155/2022/7783588 *2022*(1)
[6] Ali, A. A., Esekbi, M. O. I., & Sreh, M. (2023). Development of International Roughness Index from pavement distress using artificial neural networks (ANNs). *International Journal of Advanced Engineering and Business Sciences*, *4*(1), 179–192. https://doi.org/10.21608/ijaebs.2023.165192.1054
[7] Georgiou, P., Plati, C., & Loizos, A. (2018). Soft computing models to predict pavement roughness: A comparative study. *Advances in Civil Engineering*, 5939806. https://doi.org/10.1155/2018/5939806 *2018*
[8] Bashar, M. Z., & Torres-Machi, C. (2021). Performance of machine learning algorithms in predicting the pavement international roughness index. *Transportation Research Record*, *2675*(5), 226–237. https://doi.org/10.1177/0361198120986171
[9] Hossain, M., Gopisetti, L. S. P., & Miah, M. S. (2020). Artificial neural network modelling to predict international roughness index of rigid pavements. *International Journal of Pavement Research and Technology*, *13*(3), 229–239. https://doi.org/10.1007/s42947-020-0178-x
[10] Plati, C., Armeni, A., Kyriakou, C., & Asoniti, D. (2025). AI for predicting pavement roughness in road monitoring and maintenance. *Infrastructures*, *10*(7), 157. https://doi.org/10.3390/infrastructures10070157
[11] Purbasari, I. Y., Bayuseno, A. P., Isnanto, R. R., & Winarni, T. I. (2025). Readmission risk prediction after total hip arthroplasty using machine learning and hyperparameter optimized with Bayesian optimization. *International Journal of Advanced Computer Science and Applications*, *16*(2), 887–898. https://doi.org/10.14569/IJACSA.2025.0160288
[12] Nguyen, H. L., Le, T. H., Pham, C. T., Le, T. T., Ho, L. S., Tran, V. Q., . . . , & Prakash, I. (2019). Adaptive network based fuzzy inference system with meta-heuristic optimizations for international roughness index prediction. *Applied Sciences*, *9*(21), 4715. https://doi.org/10.3390/app9214715
[13] Pasupunuri, S. K., Thom, N., & Li, L. (2024). Roughness prediction of jointed plain concrete pavement using physics informed neural networks. *Transportation Research Record*, *2678*(11), 1733–1746. https://doi.org/10.1177/03611981241245991
[14] Sandamal, K., Shashiprabha, S., Muttil, N., & Rathnayake, U. (2023). Pavement roughness prediction using explainable and supervised machine learning technique for long-term performance. *Sustainability*, *15*(12), 9617. https://doi.org/10.3390/su15129617
[15] Aranha, A. L., Bernucci, L. L. B., & Vasconcelos, K. L. (2023). Effects of different training datasets on machine learning models for pavement performance prediction. *Transportation Research Record*, *2677*(8), 196–206. https://doi.org/10.1177/03611981231155902
[16] Liu, X., & Al-Qadi, I. L. (2021). Development of a simulated three-dimensional truck model to predict excess fuel consumption resulting from pavement roughness. *Transportation Research Record*, *2675*(9), 1444–1456. https://doi.org/10.1177/03611981211007849
[17] Damirchilo, F., Hosseini, A., Mellat Parast, M., & Fini, E. H. (2021). Machine learning approach to predict International Roughness Index using long-term pavement performance data. *Journal of Transportation Engineering, Part B: Pavements*, *147*(4), 04021058. https://doi.org/10.1061/jpeodx.0000312
[18] Wang, C., Xu, S., & Yang, J. (2021). Adaboost algorithm in artificial intelligence for optimizing the IRI prediction accuracy of asphalt concrete pavement. *Sensors*, *21*(17), 5682. https://doi.org/10.3390/s21175682
[19] Abdelaziz, N., El-Hakim, Abd., T, R., El-Badawy, S. M., & Afify, H. A. (2020). International Roughness Index prediction model for flexible pavements. *International Journal of Pavement Engineering*, *21*(1), 88–99. https://doi.org/10.1080/10298436.2018.1441414
[20] Sun, X., Wang, H., & Mei, S. (2023). Highway performance prediction model of International Roughness Index based on panel data analysis in subtropical monsoon climate. *Construction and Building Materials*, *366*, 130232. https://doi.org/10.1016/j.conbuildmat.2022.130232
[21] Lv, B., Zhang, Z., Wang, J., Zhu, Q., Wang, X., & Hou, Y. (2025). An explainable XGBoost model for International Roughness Index prediction and key factor identification. *Applied Sciences*, *15*(4), 1893. https://doi.org/10.3390/app15041893
[22] Ali, A. A., Heneash, U., Hussein, A., & Khan, S. (2023). Application of artificial neural network technique for prediction of pavement roughness as a performance indicator. *Journal of King Saud University - Engineering Sciences*, *36*(2), 128–139. https://doi.org/10.1016/j.jksues.2023.01.001
[23] Lendra, L., Wibowo, M. A., Utomo, J., & Hatmoko, D. (2023). A systematic literature network analysis: Research mapping of International Roughness Index. *Instrumentation Mesure Métrologie*, *22*(3), 1–27. https://doi.org/10.18280/i2m.220301
[24] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . , & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71
[25] Hakim, D. K., Gernowo, R., & Nirwansyah, A. W. (2024). Flood prediction with time series data mining: Systematic review. *Natural Hazards Research*, *4*(2), 194–220. https://doi.org/10.1016/j.nhres.2023.10.001
[26] Wu, D., Wang, L., Zhang, Y., Chen, X., Li, H., & Liu, J. (2025). Digital twin technology in transportation infrastructure: A comprehensive survey of current applications, challenges, and future directions. *Applied Sciences*, *15*(4), 1911. https://doi.org/10.3390/app15041911
[27] Kretzschmar, M., Dammann, M. P., Schwoch, S., Braun, F., & Saske, B. (2024). Evaluating the current role of generative AI in engineering development and design: A systematic review. In *NORDDESIGN Conference Reykjavik*, 1–10. https://doi.org/10.35199/NORDDESIGN2024.3
[28] Cole, R. (2023). Inter-rater reliability methods in qualitative case study research. *Sociological Methods and Research*, *53*(4), 1944–1975. https://doi.org/10.1177/00491241231156971
[29] Delgado-Martos, E., Pesqueira-Calvo, C., García-Tejedor, J., & Nogales, A. (2025). A systematic review of artificial intelligence for capturing real-world structures into building information modelling. *Journal of Building Engineering*, *113*, 114093. https://doi.org/10.1016/j.jobe.2025.114093
[30] Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis:

An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

[31] Milad, A., Ali, A. A., Al-Sulaimi, Z. S., & Al-Kindi, K. M. (2025). Utilizing spatial artificial intelligence to develop pavement performance indices: A case study. *Scientific Reports*, *15*, 39603. https://doi.org/10.1038/s41598-025-23290-7

[32] Marzi, G., Balzano, M., & Pellegrini, M. M. (2025). Guidelines for bibliometric-systematic literature reviews: 10 steps to combine analysis, synthesis, and theory development. *International Journal of Management Reviews*, *27*(1), 81–103. https://doi.org/10.1111/ijmr.12381

[33] Duan, C. (2024). Analyses of scientific collaboration networks among authors, institutions, and countries in FinTech studies: A bibliometric review. *FinTech*, *3*(2), 249–273. https://doi.org/10.3390/fintech3020015

[34] Miguel, S., González, C. M., Chinchilla, Z., & Miguel, S. (2024). Towards a new approach to analyzing the geographical scope of national research. An exploratory analysis at the country level. *Scientometrics*, *129*(7), 3659–3679. https://doi.org/10.1007/s11192-024-05045-9

[35] van Eck, N. J., & Waltman, L. (2023). *VOSviewer manual (Version 1.6.19)*.Universiteit Leiden: http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.19.pdf

[36] Rizelioğlu, M. (2025). An extensive bibliometric analysis of pavement deterioration detection using sensors and machine learning: Trends, innovations, and future directions. *Alexandria Engineering Journal*, *112*, 349–366. https://doi.org/10.1016/j.aej.2024.09.097

[37] Tsilika, K. (2023). Exploring the contributions to mathematical economics: A bibliometric analysis using Bibliometrix and VOSviewer. *Mathematics*, *11*(22), 4703. https://doi.org/10.3390/math11224703

[38] Liu, Q., Ali, N. L., & Lee, H. Y. (2025). Applying VOSviewer in a bibliometric review on English language teacher education research: An analysis of narratives, networks and numbers. *Cogent Education*, *12*(1), 2449728. https://doi.org/10.1080/2331186X.2025.2449728

[39] Koutrouli, M., Karatzas, E., Papanikolopoulou, K., & Pavlopoulos, G. A. (2022). NORMA: The network makeup artist—A web tool for network annotation visualization. *Genomics, Proteomics & Bioinformatics*, *20*(3), 578–586. https://doi.org/10.1016/j.gpb.2021.02.005

[40] Baneshi, M. R., & Dobson, A. (2024). Choices of measures of association affect the visualisation and composition of the multimorbidity networks. *BMC Medical Research Methodology*, *24*, 1–9. https://doi.org/10.1186/s12874-024-02286-3

[41] Strayer, N., Cenzer, L., Covinsky, J., & others. (2023). Data and text mining interactive network-based clustering and investigation of multimorbidity association matrices with association Subgraphs. *Bioinformatics*, *39*(1), 1–6. https://doi.org/10.1093/bioinformatics/btac780

[42] Barker, T. H., Stone, J. C., Sears, K., Klugar, M., Tufanaru, C., Leonardi-Bee, J., . . . , & Munn, Z. (2023). Revising the JBI quantitative critical appraisal tools to improve their applicability: An overview of methods and the development process. *JBI Evidence Synthesis*, *21*(3), 478–493. https://doi.org/10.11124/JBIES-22-00125

[43] Szomszor, M., Pendlebury, D. A., Adams, J., Fry, R., Gebert, C., Potter, R. W. K., & Rogers, G. (2021). Interpreting bibliometric data. *Frontiers in Research Metrics and Analytics*, *5*, 628703. https://doi.org/10.3389/frma.2020.628703

[44] Lim, W. M., & Kumar, S. (2024). Guidelines for interpreting the results of bibliometric analysis: A sensemaking approach. *Global Business and Organizational Excellence*, *43*(2), 17–26. https://doi.org/10.1002/joe.22229

[45] Marc, W., Kumar, S., & Donthu, N. (2024). How to combine and clean bibliometric data and use bibliometric tools synergistically: Guidelines using metaverse research. *Journal of Business Research*, *182*, 114760. https://doi.org/10.1016/j.jbusres.2024.114760

[46] Mejia, C., Wu, M., Zhang, Y., & Kajikawa, Y. (2021). Exploring topics in bibliometric research through citation networks and semantic analysis. *Frontiers in Research Metrics and Analytics*, *6*, 742311. https://doi.org/10.3389/frma.2021.742311

[47] Peng, X. (2025). Digitalization of the construction industry driven by AI and BIM: A bibliometric study using CiteSpace and VOSviewer. *Engineering*, *17*(11), 582–599. https://doi.org/10.4236/eng.2025.1711032

[48] Ali, U., Frawley, R., Gardiner, T., Ginther, R., & Hull, K. (2023). A method for analyzing text using VOSviewer. *MethodsX*, *11*, 102339. https://doi.org/10.1016/j.mex.2023.102339

[49] Ji, Z., Zhou, M., Wang, Q., & Huang, J. (2024). Predicting the International Roughness Index of JPCP and CRCP rigid pavement: A random forest (RF) model hybridized with modified beetle antennae search (MBAS) for higher accuracy. *Computer Modeling in Engineering and Sciences*, *139*(2), 1557–1582. https://doi.org/10.32604/cmes.2023.046025

[50] Wang, Q., Zhou, M., Sabri, M. M. S., & Huang, J. (2022). A comparative study of AI-based International Roughness Index (IRI) prediction models for jointed plain concrete pavement (JPCP). *Materials*, *15*(16), 5605. https://doi.org/10.3390/ma15165605

[51] Luo, Z., Wang, H., & Li, S. (2022). Prediction of International Roughness Index based on stacking fusion model. *Sustainability*, *14*(12), 6949. https://doi.org/10.3390/su14126949

[52] Wu, Y. (2024). From ensemble learning to deep ensemble learning: A case study on multi-indicator prediction of pavement performance. *Applied Soft Computing*, *166*, 112188. https://doi.org/10.1016/j.asoc.2024.112188

[53] Zhang, T., Smith, A., Zhai, H., & Lu, Y. (2025). LSTM+MA: A time-series model for predicting pavement IRI. *Infrastructures*, *10*(1), 10. https://doi.org/10.3390/infrastructures10010010

[54] Bral, S., Kumar, P. P., & Chopra, T. (2022). Prediction of International Roughness Index using CatBooster and SHAP values. *International Journal of Pavement Research and Technology*, *17*, 518–533. https://doi.org/10.1007/s42947-022-00253-z

[55] Chen, K., Torbaghan, M. E., Thom, N., & Faramarzi, A. (2025). Physics-guided neural network for predicting international roughness index on flexible pavements considering accuracy, uncertainty and stability. *Engineering Applications of Artificial Intelligence*, *142*, 109922. https://doi.org/10.1016/j.engappai.2024.109922

[56] Hossain, M. I., Gopisetti, L. S. P., & Miah, M. S. (2019). International Roughness Index prediction of flexible pavements using neural networks. *Journal of Transportation Engineering, Part B: Pavements*, *145*(1), 04018057. https://doi.org/10.1061/JPEODX.0000088

[57] Kaloop, M. R., El-Badawy, S. M., Ahn, J., Sim, H. B., Hu, J. W., & Abd El-Hakim, R. T. (2020). A hybrid wavelet-optimally-pruned extreme learning machine model for the estimation of international roughness index of rigid pavements. *International Journal of Pavement Engineering*, *23*(14), 5105–5119. https://doi.org/10.1080/10298436.2020.1776281

[58] Douangphachanh, V., & Oneyama, H. (2014). A study on the use of smartphones under realistic settings to estimate road roughness condition. *EURASIP Journal on Wireless Communications and Networking*, 114. https://doi.org/10.1186/1687-1499-2014-114 *2014*(1)

[59] Janani, L., Doley, R., Sunitha, V., & Mathew, S. (2021). Precision enhancement of smartphone sensor-based pavement roughness estimation by standardizing host vehicle speed. *Canadian Journal of Civil Engineering*, *49*(5), 716–730. https://doi.org/10.1139/cjce-2021-0182

[60] Sang, Y., Yu, Q., Fang, Y., Vo, V., & Wix, R. (2024). Smartphone-based IRI estimation for pavement roughness monitoring: A data-driven approach. *IEEE Internet of Things Journal*, *11*(11), 19708–19720. https://doi.org/10.1109/JIOT.2024.3369109

[61] Yu, Q., Sang, Y., Fang, Y., Vo, V., & Wix, R. (2024). Estimate road roughness using smartphone response data: A semi-supervised learning approach. In *Proceedings of the European Conference on Computing in Construction*, 454–460. https://doi.org/10.35490/EC3.2024.176

[62] Baykal, T., Ergezer, F., Eriskin, E., & Terzi, S. (2024). Using ensemble machine learning to estimate International Roughness Index of asphalt pavements. *Iranian Journal of Science and Technology - Transactions of Civil Engineering*, *48*(4), 2773–2784. https://doi.org/10.1007/s40996-023-01320-6

[63] Li, J., Guo, J., Li, B., & Meng, L. (2024). Novel instance-based transfer learning for asphalt pavement performance prediction. *Buildings*, *14*(3), 852. https://doi.org/10.3390/buildings14030852

[64] Wang, C., Xiao, W., & Liu, J. (2023). Developing an improved extreme gradient boosting model for predicting the international roughness index of rigid pavement. *Construction and Building Materials*, *408*, 133523. https://doi.org/10.1016/j.conbuildmat.2023.133523

[65] Molinero-Pérez, N., Montalbán-Domingo, L., Sanz-Benlloch, A., & García-Segura, T. (2025). Dual model for International Roughness Index classification and prediction. *Infrastructures*, *10*(1), 23. https://doi.org/10.3390/infrastructures10010023

[66] Rosa, F. D., Liu, L., & Gharaibeh, N. G. (2017). IRI prediction model for use in network-level pavement management systems. *Journal of Transportation Engineering, Part B: Pavements*, *143*(1), 04017001. https://doi.org/10.1061/jpeodx.0000003

[67] Gharieb, M., & Nishikawa, T. (2021). Development of roughness prediction models for Laos national road network. *CivilEng*, *2*(1), 158–173. https://doi.org/10.3390/civileng2010009

[68] Alnaqbi, A., Zeiada, W., & Al-Khateeb, G. G. (2024). Machine learning modeling of pavement performance and IRI prediction in flexible pavement. *Innovative Infrastructure Solutions*, *9*(10), 385. https://doi.org/10.1007/s41062-024-01688-y

[69] Al-Samahi, S., Zeiada, W., Al-Khateeb, G. G., Hamad, K., & Alnaqbi, A. (2024). A comparative study of pavement roughness prediction models under different climatic conditions. *Infrastructures*, *9*(10), 167. https://doi.org/10.3390/infrastructures9100167

[70] Dubey, A., & Rasool, A. (2022). Recent advances and applications of deep learning technique. *Journal of Theoretical and Applied Information Technology*, *100*(13), 500–509.

[71] Patil, T., Pandey, S., & Visrani, K. (2021). A review on basic deep learning technologies and applications. In *Lecture Notes on Data Engineering and Communications Technologies* (pp. 565–573). https://doi.org/10.1007/978-981-15-4474-3_61

[72] Qin, L., & Yin, K. (2024). An improved genetic algorithm for predicting road roughness coefficient. In *2024 5th International Conference on Big Data and Artificial Intelligence and Software Engineering*, 920–924. https://doi.org/10.1109/ICBASE63199.2024.10761989

[73] Rasheed, A. A. (2021). Improving prediction efficiency by revolutionary machine learning models. *Materials Today: Proceedings*, *81*(2), 577–583. https://doi.org/10.1016/j.matpr.2021.04.014

[74] Bosnić, Z., & Kononenko, I. (2008). Empirical analysis of reliability estimates for individual regression predictions. In *Lecture Notes in Computer Science*, *5182*, 379–388. https://doi.org/10.1007/978-3-540-85836-2_36

[75] Bosnić, Z., & Kononenko, I. (2009). Influence of domain and model properties on the reliability estimates' performance. *International Journal of Data Warehousing and Mining*, *5*(4), 58–76. https://doi.org/10.4018/jdwm.2009080704

[76] Carannante, G., Dera, D., Aminul, O., Bouaynaya, N. C., & Rasool, G. (2022). Self-assessment and robust anomaly detection with Bayesian deep learning. In *2022 25th International Conference on Information Fusion*, 1–8. https://doi.org/10.23919/FUSION49751.2022.9841358

[77] Hama, K., Matsubara, T., Uehara, K., & Cai, J. (2021). Exploring uncertainty measures for image-caption embedding-and-retrieval task. *ACM Transactions on Multimedia Computing, Communications and Applications*, *17*(2), 56. https://doi.org/10.1145/3425663

[78] Huang, C. H., & Lin, Y. K. (2024). Manufacturing system evaluation in terms of system reliability via long short-term memory. *Reliability Engineering and System Safety*, *251*, 110365. https://doi.org/10.1016/j.ress.2024.110365

[79] Xin, J., Akiyama, M., Frangopol, D. M., & Zhang, M. (2022). Multi-objective optimisation of in-service asphalt pavement maintenance schedule considering system reliability estimated via LSTM neural networks. *Structure and Infrastructure Engineering*, *18*(7), 1002–1019. https://doi.org/10.1080/15732479.2022.2038641

[80] Jindal, A., Gupta, A., & Rahul. (2022). Comparative analysis of software reliability prediction using machine learning and deep learning. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Smart Energy*, 389–394. https://doi.org/10.1109/ICAIS53314.2022.9743129

[81] Deng, A., Xiong, M., & Hooi, B. (2023). *Great models think alike: Improving model reliability via inter-model latent agreement*. *arXiv Preprint:2305.01481*

[82] Samake, A., & Boulmane, L. (2021). Acceptance and rejection zones for a classifier's predictions in deep learning. In *2021 5th International Conference on Intelligent Computing in Data Sciences*, 1–5. https://doi.org/10.1109/ICDS53782.2021.9626739

[83] Vasquez-Monteros, J., Palacios-Ortega, J., Samaniego-Rojas, P., & Vera, C. M. G. (2024). Searching for precision in pavement evaluation: A comparative analysis of smartphones for measuring the International Roughness Index (IRI).

In *Proceedings of the World Congress on New Technologies*, 1–7. https://doi.org/10.11159/icceia24.143
[84] Jinxi, Z., Qianqian, C., & Yongjie, D. (2024). Pavement roughness detection method based on smartphone and PSO-BP neural network. In *International Conference on Road and Airfield Pavement Technology*, 883–893. https://doi.org/10.1061/9780784485255.074
[85] Xu, Y., Zhang, Z., Ph, D., & Asce, A. M. (2022). Review of applications of artificial intelligence algorithms in pavement management. *Journal of Transportation Engineering*, *148*(3), 04022001. https://doi.org/10.1061/JPEODX.0000373
[86] Alqaydi, S., Zeiada, W., El Wakil, A., Alnaqbi, A. J., & Azam, A. (2024). A comprehensive review of smartphone and other device-based techniques for road surface monitoring. *Eng*, *5*(4), 3397–3426. https://doi.org/10.3390/eng5040177
[87] Jalili, F., Ghavami, S. M., & Afsharnia, H. (2024). An artificial neural network approach to assess road roughness using smartphone-based crowdsourcing data. *Engineering Applications of Artificial Intelligence*, *138*, 109308. https://doi.org/10.1016/j.engappai.2024.109308
[88] Pipanmekaporn, L., Kamonsantiroj, S., Sutthaweekul, R., Lee, W., & Srisomboon, K. (2024). Smartphone-based Road condition sensing using machine learning techniques. In *2024 Proceedings International SoC Design Conference*, 246–247. https://doi.org/10.1109/ISOCC62682.2024.10762144
[89] Moghadam, A., & Sarlo, R. (2021). Application of smartphones in pavement deterioration identification using artificial neural network. In *Sensors and Instrumentation, Aircraft/Aerospace, Energy Harvesting & Dynamic Environments Testing*, *7*, 167–174. https://doi.org/10.1007/978-3-030-75988-9_12
[90] Llopis-Castelló, D., Camacho-Torregrosa, F. J., Romeral-Pérez, F., & Tomás-Martínez, P. (2024). Estimation of pavement condition based on data from connected and autonomous vehicles. *Infrastructures*, *9*(10), 188. https://doi.org/10.3390/infrastructures9100188
[91] Jeong, J. H., & Jo, H. (2024). Toward real-world implementation of deep learning for smartphone-crowdsourced pavement condition assessment. *IEEE Internet of Things Journal*, *11*(4), 6328–6337. https://doi.org/10.1109/JIOT.2023.3312353
[92] Kargah-Ostadi, N., Vasylevskyi, K., Ablets, A., & Drach, A. (2023). Reconciling pavement condition data from connected vehicles with the International Roughness Index from standard monitoring equipment using physics-integrated machine learning. *Transportation Research Record*, *2678*(2), 416–429. https://doi.org/10.1177/03611981231174406
[93] Alatoom, Y. I., & Al-Suleiman, T. I. (2021). Development of pavement roughness models using artificial neural network (ANN). *International Journal of Pavement Engineering*, *23*(3), 4622–4637. https://doi.org/10.1080/10298436.2021.1968396
[94] Laubis, K., Simko, V., & Schuller, A. (2016). Road condition measurement and assessment: A crowd based sensing approach. In *2016 International Conference on Information Systems*, (p. *20*).
[95] Liu, W., Li, J. Q., Wang, G., & Wang, K. (2024). Pavement safety characteristics evaluation utilizing crowdsourced vehicular and cellular sensor data. *Journal of Transportation Engineering, Part B: Pavements*, *150*(3), 04024036. https://doi.org/10.1061/jpeodx.pveng-1486
[96] Khatoon, A., Wang, W., Ullah, A., Ahmed, I., Hussain, A., & Roslee, M. B. (2024). Advancement in pavement condition assessment: An AI-based computer vision approach. In *2024 Multimedia University Engineering Conference*, 1–6. https://doi.org/10.1109/MECON62796.2024.10776158
[97] Dong, J., Meng, W., Liu, Y., & Ti, J. (2021). A framework of pavement management system based on IoT and big data. *Advanced Engineering Informatics*, *47*, 101226. https://doi.org/10.1016/j.aei.2020.101226
[98] Mienye, I. D., & Swart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information*, *15*(12), 755. https://doi.org/10.3390/info15120755
[99] Bahji, A., Acion, L., Laslett, A. M., & Adinoff, B. (2023). Exclusion of the non-English-speaking world from the scientific literature: Recommendations for change for addiction journals and publishers. *Nordic Studies on Alcohol and Drugs*, *40*(1), 6–13. https://doi.org/10.1177/14550725221102227

How to Cite: Lendra, L., Wibowo, M. A., Hatmoko, J. U. D., & Teguh, R. (2026). A Comprehensive Literature Review and Bibliometric Analysis of the Development and Difficulties of Artificial Intelligence Methods for International Roughness Index Prediction. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62025993

Nomenclature

Acronym	Definition
AADTT	Annual Average Daily Truck Traffic
AI	Artificial Intelligence
ANFIS	Adaptive Network-Based Fuzzy Inference System
ANN	Artificial Neural Network
BAS	Beetle Antennae Search
CESAL	Cumulative Equivalent Single-Axle Load
CNN	Convolutional Neural Network
CRCP	Continuously Reinforced Concrete Pavement
DL	Deep Learning
DNN	Deep Neural Network
ESALs	Equivalent Single Axle Loads
FWD	Falling Weight Deflectometer
GA	Genetic Algorithm
GBDT	Gradient Boosting Decision Tree
GPS	Global Positioning System
GPR	Gaussian Process Regression
IRI	International Roughness Index
IRI(I)	Initial Roughness expressed as IRI, m/km
JPCP	Jointed Plain Concrete Pavement
KESAL	Equivalent Single-Axle Loads in Thousands
LR	Logistic Regressor
LSTM	Long Short-Term Memory
LTPP	Long-Term Pavement Performance
LWD	Light Weight Deflectometer
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MDI	Mean Decrease in Impurity
ML	Machine Learning
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MSE	Mean Squared Error
OP-ELM	Optimally Pruned Extreme Learning Machine
PCI	Pavement Condition Index
PGNN	Physics-Guided Neural Network
PMS	Pavement Management System
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSO	Particle Swarm Optimization
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RQ	Research Question
SHAP	SHapley Additive exPlanations
SLR	Systematic Literature Review

(*Continued*)

(*Continued*)

Acronym	Definition
SSL	Semi-Supervised Learning
SVM	Support Vector Machine
SVR	Support Vector Regressor
TFAULT	Total Joint Faulting cumulated per km
TPE	Tree-structured Parzen Estimator
VOSviewer	Visualization of Similarities Viewer
WOPELM	Wavelet-Optimally Pruned Extreme Learning Machine
XGBoost	Extreme Gradient Boosting