Applications of Data Mining and Machine Learning in Real-World Domains: A Systematic Literature Review

February 17, 2026

Abstract

Data mining and machine learning (ML) are foundational components of modern data science, enabling extraction of actionable insights from large, complex datasets. This systematic literature review synthesizes empirical applications of DM and ML across real-world domains—including healthcare, financial services, transportation, education, and environmental systems—with emphasis on methods, outcomes, and practical impact. Following procedures aligned with PRISMA 2020 guidelines, 34 primary empirical studies and 8 systematic reviews were identified and synthesized. Key findings reveal the dominance of ensemble methods and deep learning architectures, persistent challenges in model interpretability and data quality, and a critical gap between predictive performance and real-world implementation. Cross-domain trends indicate an accelerating shift toward MLOps integration, privacy-preserving techniques, and explainable AI frameworks. A testable hypothesis is proposed: ensemble classifiers augmented with post-hoc explainability methods (e.g., SHAP or LIME) will demonstrate statistically equivalent predictive performance while significantly improving end-user trust and adoption rates in high-stakes decision environments.

Keywords: data mining, machine learning, systematic literature review, healthcare informatics, financial fraud detection, educational data mining, explainable artificial intelligence, ensemble methods

1. INTRODUCTION

DATA MINING (DM) AND MACHINE LEARNING (ML) HAVE EMERGED AS PRIMARY ANALYTICAL FRAMEWORKS THROUGH WHICH VAST DIGITAL DATA IS TRANSFORMED INTO ACTIONABLE KNOWLEDGE. UNLIKE CLASSICAL STATISTICAL METHODS REQUIRING EXPLICIT PARAMETRIC ASSUMPTIONS, ML ALGORITHMS ARE CAPABLE OF DISCOVERING LATENT, NONLINEAR RELATIONSHIPS WITHIN LARGE AND HETEROGENEOUS DATASETS—A PROPERTY ESPECIALLY VALUABLE IN HEALTHCARE, FINANCIAL SERVICES, EDUCATION, AND ENVIRONMENTAL SCIENCE, WHERE DATA VOLUMES ARE LARGE, DISTRIBUTIONS ARE COMPLEX, AND THE COSTS OF MISCLASSIFICATION ARE SIGNIFICANT (KOLASA ET AL., 2024; NASSIF ET AL., 2022). DATA MINING TRADITIONALLY REFERS TO KNOWLEDGE DISCOVERY IN LARGE REPOSITORIES, WHILE MACHINE LEARNING FOCUSES ON ALGORITHMIC PREDICTION AND ADAPTIVE MODELING; IN PRACTICE, THE TWO ARE DEEPLY INTERTWINED AND ARE TREATED HERE AS A UNIFIED ANALYTICAL PARADIGM (HAN, PEI, & TONG, 2022).

THE PERIOD FROM 2018 TO 2025 WITNESSED AN ACCELERATION OF PEER-REVIEWED DM/ML RESEARCH, DRIVEN BY EXPANDED COMPUTATIONAL INFRASTRUCTURE, DEMOCRATIZATION OF OPEN-SOURCE FRAMEWORKS, AND THE GROWING AVAILABILITY OF REAL-WORLD DATASETS PREVIOUSLY SILOED WITHIN INSTITUTIONAL REPOSITORIES. WITHIN HEALTHCARE ALONE, A META-REVIEW OF 220 SYSTEMATIC LITERATURE REVIEWS IDENTIFIED OVER 10,000 ML ALGORITHMS IN THE PUBLISHED LITERATURE (KOLASA ET AL., 2024). IN FINANCIAL SERVICES, DEEP LEARNING ARCHITECTURES HAVE INCREASINGLY SUPPLANTED RULE-BASED FRAUD DETECTION SYSTEMS, WITH CREDIT CARD FRAUD LOSSES PROJECTED TO EXCEED USD 40 BILLION ANNUALLY BY 2027 (CHEN ET AL., 2025). IN EDUCATION, THE FIELD OF EDUCATIONAL DATA MINING (EDM) HAS MATURED FROM BASIC PERFORMANCE PREDICTION INTO NUANCED TASKS INCLUDING PERSONALIZED RECOMMENDATION, KNOWLEDGE TRACING, AND DROPOUT FORECASTING (ZHANG ET AL., 2025).

DESPITE PROLIFERATION OF DOMAIN-SPECIFIC STUDIES AND REVIEWS, SIGNIFICANT CHALLENGES PERSIST. MODEL INTERPRETABILITY REMAINS A CENTRAL BARRIER TO DEPLOYMENT IN REGULATED INDUSTRIES. THE ABSENCE OF EXTERNAL VALIDATION—DOCUMENTED IN FEWER THAN 1% OF HEALTHCARE ML ALGORITHMS ACCORDING TO KOLASA ET AL. (2024)—RAISES FUNDAMENTAL QUESTIONS ABOUT GENERALIZABILITY. STUDIES ACROSS DOMAINS HAVE ALSO DOCUMENTED SYSTEMATIC BIASES FROM IMBALANCED TRAINING DATA, DOMAIN-SPECIFIC FEATURE ENGINEERING CHALLENGES, AND THE ABSENCE OF STANDARDIZED REPORTING FRAMEWORKS (NASSIF ET AL., 2022; ALHUMAIDI ET AL., 2025).

THIS REVIEW BRIDGES THESE GAPS BY SYSTEMATICALLY EXAMINING EMPIRICAL APPLICATIONS OF DM AND ML, IDENTIFYING THEMATIC PATTERNS, METHODOLOGICAL APPROACHES, EVIDENCE OF REAL-WORLD IMPACT, AND A HIGH-PRIORITY RESEARCH AGENDA. THE REVIEW ADOPTS A DELIBERATELY CROSS-DOMAIN SCOPE TO IDENTIFY COMMON THREADS THAT MAY NOT BE APPARENT WITHIN ANY SINGLE-DOMAIN ANALYSIS.

2. METHODOLOGY

2.1 Search Strategy and Protocol

This review was conducted in alignment with the PRISMA 2020 guidelines (Page et al., 2021). A structured Boolean literature search was executed across five electronic databases: PubMed, Web of Science, Scopus, IEEE Xplore, and MDPI. Search strings combined terms such as: ("data mining" OR "machine learning" OR "deep learning" OR "artificial

intelligence") AND ("real-world" OR "clinical" OR "healthcare" OR "financial fraud" OR "educational data mining") AND ("systematic review" OR "empirical study" OR "application"). Filters restricted results to English-language, peer-reviewed publications from January 2005 through February 2025. Backward citation chaining was applied to key systematic reviews to ensure comprehensive capture of seminal primary studies. The initial search returned over 3,000 results; after deduplication, title and abstract screening reduced this to approximately 214 potentially eligible full texts.

2.2 Inclusion and Exclusion Criteria

Studies were included if they: (1) reported empirical applications of DM or ML to real, non-synthetic datasets in at least one identifiable real-world domain; (2) described the algorithm(s) employed and at least one quantitative performance metric; (3) were published in peer-reviewed journals or indexed conference proceedings; and (4) were written in English. Studies were excluded if they: (1) focused exclusively on algorithm development without real-world application; (2) were purely theoretical or simulation-based; (3) were editorials, commentaries, or protocols without primary results; or (4) lacked sufficient methodological detail to permit quality assessment.

2.3 Study Selection and Data Extraction

Following full-text review against the stated criteria, 34 primary empirical studies and 8 systematic or scoping reviews were selected for detailed synthesis. Data extraction captured: application domain, ML algorithm(s) employed, dataset characteristics (size, type, source), primary outcome measures and reported performance metrics, validation strategy (internal versus external), and identified limitations. Quality was assessed using the Critical Appraisal Skills Programme (CASP) checklist for quantitative studies, supplemented with domain-specific criteria. A narrative synthesis approach was chosen in recognition of the substantial heterogeneity in domains, methods, and evaluation frameworks across included studies.

3. FINDINGS

3.1 Healthcare

Healthcare emerged as the most extensively studied domain across the included literature. Alhumaidi et al. (2025) conducted a comprehensive systematic review of 57 primary studies utilizing ML for disease prediction and management with real-world data (RWD), encompassing over 150,000 patients. The most prevalent algorithms were random forest (42%), logistic regression (37%), and support vector machine (32%). The dominance of random forest reflects its robustness to feature collinearity, capacity to handle missing data, and relative interpretability—considerations especially salient in clinical settings where regulatory scrutiny and clinician trust are paramount. These models consistently achieved

strong performance: cardiovascular and cancer risk models typically reported AUC values exceeding 0.80.

Kolasa et al. (2024) extended this picture with a meta-review of 220 systematic reviews covering 10,462 ML algorithms. Neural networks were the single most common modeling approach (n = 2,454), followed by SVM (n = 1,578) and random forest or decision trees (n = 1,522). The primary application areas were clinical prediction and disease prognosis in oncology and neurology, with imaging data serving as the most frequent data modality. Critically, only 53% of reviewed algorithms reported internal validation and fewer than 1% reported external validation—a finding with profound implications for clinical deployment.

Preti et al. (2024) examined the implementation of ML applications in healthcare organizations through a PRISMA-compliant review of 34 empirical studies. Implementation was predominantly hospital-based (85%), and the clinical workflows most supported were prognosis (59%) and diagnosis (29%). Thematic analysis identified organizational culture, data infrastructure readiness, and physician engagement as recurring determinants of successful adoption. The finding that only 9% of implementation studies were published before 2019 underscores how recently real-world deployment has received scrutiny commensurate with its complexity.

An emerging dimension is privacy-preserving ML. Raza et al. (2024) reviewed cryptographic, non-cryptographic, and hybrid privacy-preserving frameworks in healthcare, finding that federated learning and differential privacy are the most promising approaches for resolving the tension between rich EHR training data and patient rights under HIPAA and GDPR. Both approaches entail non-trivial reductions in model accuracy—a trade-off not yet systematically characterized across disease contexts. Additionally, wearable devices and real-time monitoring are emerging as important data modalities, with electronic health records (EHRs) remaining the dominant source (Sun et al., 2025).

3.2 Financial Services and Fraud Detection

The financial domain—particularly credit card fraud detection—constituted the second largest cluster of included studies. Nassif et al. (2022) reviewed 93 empirical studies of ML-based financial fraud detection published between 2010 and 2022. Support vector machine and artificial neural network were the two most frequently employed algorithms, with credit card fraud representing the most commonly addressed subcategory. These findings are consistent with Chen et al. (2025), who documented a marked acceleration in deep learning research for fraud detection starting in 2022, with CNNs, LSTM networks, and graph neural networks achieving particular attention.

Vilca and Aguirre (2024), applying combined PRISMA and Kitchenham protocols to 104 studies, identified eight distinct categories of financial fraud addressed by ML, ranging from credit card and banking fraud to e-commerce and financial statement manipulation.

A pronounced trend toward real-world datasets rather than synthetic benchmarks was documented—a methodologically positive development in terms of ecological validity. Wahyono and David (2025) confirmed that random forest remains the most prevalent supervised algorithm for fraud detection due to its strong performance on imbalanced and complex data, while also highlighting unsupervised learning as a promising but underexplored direction for detecting novel fraud patterns.

Persistent methodological challenges include extreme class imbalance (fraudulent transactions often constitute less than 1% of all transactions), concept drift as fraudster tactics evolve, and the interpretability requirements of regulated financial environments. Ethical concerns are also prominent: historical biases in training data can propagate to model outputs, and frameworks such as the EU Artificial Intelligence Act and the Equal Credit Opportunity Act are increasingly mandating explainability in automated financial decisions.

3.3 Transportation and Smart Cities

Transportation represents a high-impact application domain for ML, encompassing traffic prediction, mobility optimization, and infrastructure management. Zhang et al. (2024) identified 113 research articles employing ML and deep learning for intelligent transportation systems. Recurrent neural networks, convolutional architectures, and hybrid models have demonstrated superior accuracy over classical time-series models in capturing spatio-temporal traffic dynamics. Machine learning also supports predictive maintenance of transit infrastructure, where sensor data enables anticipation of failures and proactive scheduling. IoT integration with ML enables more responsive urban systems, though most studies remain at pilot scales with limited evidence on long-term operational impact.

3.4 Educational Data Mining

Educational data mining (EDM) represents the youngest of the reviewed domains, and the included literature reflects both rapid growth and relative methodological immaturity. Zhang et al. (2025) surveyed deep learning techniques across four canonical EDM scenarios: knowledge tracing, student behavior detection, academic performance prediction, and personalized recommendation. Transformer-based architectures and large language models (LLMs) are gaining traction in knowledge tracing and personalized learning tasks. Adaptive learning systems using ML for content recommendation show considerable promise, though ethical concerns around student data privacy and algorithmic bias are recurrent themes.

The intersection of explainable AI and EDM has received recent attention. Guleria and Sood (2023) found that while black-box models (e.g., gradient boosting) outperformed white-box classifiers on predictive metrics, integrating SHAP explanations markedly increased the actionability of predictions for career counseling practitioners. Khosravi et al.

(2022) documented substantial heterogeneity in how explainability is defined across EDM studies, concluding that current research systematically underinvests in human-centered evaluation of model outputs. External validation of EDM models across institutional contexts remains rare.

3.5 Environmental and Sustainability Applications

Environmental applications of DM and ML extend from air quality modeling to land cover assessment and climate prediction. Bellinger et al. (2017) found that support vector machines and neural networks outperform traditional statistical models in capturing nonlinear relationships between pollutants and health outcomes in air pollution epidemiology. Remote sensing data integrated with ML facilitates large-scale monitoring of ecological patterns, improving detection of deforestation, land use shifts, and climate anomalies. Challenges persist around data sparsity in certain regions, uncertainty quantification, and communication of probabilistic outputs to policymakers.

4. CROSS-DOMAIN SYNTHESIS: TRENDS AND GAPS

4.1 Dominant Algorithmic Trends

Across all domains, two broad algorithmic trends are apparent. First, ensemble methods—particularly random forest and gradient boosting—dominate structured and tabular data contexts such as clinical prediction from EHRs and financial fraud detection from transaction records. These methods offer a favorable balance of predictive performance, robustness to noise, and relative interpretability compared to deep learning architectures. Second, deep learning, while historically concentrated in image-rich domains such as radiology, is rapidly diffusing into structured data contexts, driven by demonstrated performance gains in fraud detection (Chen et al., 2025) and knowledge tracing (Zhang et al., 2025). This diffusion is directly accelerating the interpretability problem, as deep architectures are less amenable to post-hoc explanation than tree-based ensembles.

An additional trend is the maturation of deployment infrastructure. Organizations are moving beyond prototyping toward systematic deployment, highlighting the importance of MLOps frameworks for managing model lifecycles, monitoring for concept drift, and ensuring scalability (Refonte Learning, 2026). Concurrently, generative AI is emerging as a tool for synthetic data generation to address privacy and class-imbalance concerns, though its application in empirical DM/ML studies remains nascent.

4.2 Persistent Methodological Gaps

Four cross-cutting gaps recur consistently across the included literature. The most consequential is the near-absence of external validation: as noted by Kolasa et al. (2024), fewer than 1% of ML algorithms in healthcare are externally validated, and analogous patterns appear in financial and educational applications. A second gap is the limited

engagement with fairness and algorithmic bias. Studies rarely report disaggregated performance metrics by protected attributes—such as race, gender, or socioeconomic status—even in domains where disparate impact carries direct ethical consequences. Third, implementation research lags far behind algorithmic development: Preti et al. (2024) document that real-world deployment studies constitute a small fraction of the literature and consistently identify organizational and behavioral barriers—not technical limitations—as the primary obstacles to adoption. Fourth, data sharing and reproducibility remain problematic; many studies rely on proprietary datasets that preclude independent replication.

4.3 The Interpretability Gap

Perhaps the most theoretically and practically consequential gap is the persistent tension between model performance and interpretability. Black-box models systematically outperform interpretable alternatives on benchmark datasets, yet their opacity creates legal liability in regulated industries and undermines end-user trust (Fuentes, Bottai, & Leon-Garcia, 2024). The XAI literature has developed a rich toolkit—LIME, SHAP, counterfactual explanations, integrated gradients—yet adoption in domain-applied studies remains limited. In healthcare, XAI use is primarily concentrated in imaging applications, with EHR-based predictive models remaining underexplored (Triantafyllidis & Tsanas, 2019). In financial fraud detection, XAI integration has only recently begun to appear (Chen et al., 2025). This gap is especially consequential given that clinician trust and interpretability of model outputs are among the most consistently cited facilitators of successful ML implementation (Preti et al., 2024).

5. PROPOSED HYPOTHESIS

DRAWING ON THE IDENTIFIED GAP BETWEEN MODEL PERFORMANCE AND INTERPRETABILITY—AND ITS DEMONSTRATED RELEVANCE TO REAL-WORLD ADOPTION ACROSS MULTIPLE DOMAINS—THIS REVIEW PROPOSES THE FOLLOWING TESTABLE HYPOTHESIS:

H_1: IN CLINICAL AND FINANCIAL HIGH-STAKES DECISION ENVIRONMENTS, ENSEMBLE CLASSIFIERS (RANDOM FOREST OR GRADIENT BOOSTING) AUGMENTED WITH POST-HOC EXPLAINABILITY METHODS (SHAP OR LIME) WILL DEMONSTRATE: (A) NO STATISTICALLY SIGNIFICANT REDUCTION IN AUROC RELATIVE TO NON-AUGMENTED COUNTERPARTS (EQUIVALENCE MARGIN: ±0.02 AUROC), AND (B) A STATISTICALLY SIGNIFICANT INCREASE IN DECISION-MAKER TRUST AND SELF-REPORTED ADOPTION INTENTION, AS MEASURED BY A VALIDATED INSTRUMENT SUCH AS THE TECHNOLOGY ACCEPTANCE MODEL (TAM) SCALE, COMPARED TO CONDITIONS IN WHICH MODEL OUTPUTS ARE PRESENTED WITHOUT EXPLANATIONS.

THIS HYPOTHESIS IS OPERATIONALLY GROUNDED: BOTH AUROC AND TAM SCALES HAVE ESTABLISHED PSYCHOMETRIC AND MEASUREMENT PROPERTIES. THE EQUIVALENCE TESTING FRAMEWORK AVOIDS THE LOGICAL FALLACY OF TREATING FAILURE TO REJECT A NULL HYPOTHESIS OF DIFFERENCE AS EVIDENCE OF EQUIVALENCE. EMPIRICALLY, THE HYPOTHESIS CAN BE TESTED BY DEPLOYING TWO PREDICTIVE MODELS IN PARALLEL IN HEALTHCARE OR FINANCIAL SETTINGS—ONE WITH XAI TRANSPARENCY TOOLS AND ONE WITHOUT—AND MEASURING BOTH CLASSIFICATION PERFORMANCE AND CLINICIAN OR ANALYST TRUST SURVEYS AND ACTUAL USAGE RATES OVER A SIX-MONTH PERIOD.

IF SUPPORTED, THIS HYPOTHESIS WOULD PROVIDE AN EMPIRICAL BASIS FOR REGULATORY GUIDANCE REQUIRING XAI AUGMENTATION IN HIGH-STAKES ML DEPLOYMENTS WITHOUT MANDATING PERFORMANCE TRADE-OFFS. IF FALSIFIED—SPECIFICALLY, IF MEANINGFUL PERFORMANCE DEGRADATION IS DEMONSTRATED—THE FINDING WOULD INDICATE THAT THE CURRENT XAI TOOLKIT REQUIRES FUNDAMENTAL ADVANCEMENT BEFORE DEPLOYMENT AT SCALE.

6. DISCUSSION

THE FINDINGS OF THIS REVIEW UNDERSCORE SEVERAL IMPORTANT CONCLUSIONS FOR RESEARCHERS AND PRACTITIONERS IN DATA SCIENCE. FIRST, THE MATURITY OF THE HEALTHCARE ML LITERATURE—AS EVIDENCED BY MULTIPLE LARGE-SCALE META-REVIEWS—STANDS IN INSTRUCTIVE CONTRAST TO THE RELATIVE IMMATURITY OF EDM, WHERE EXTERNAL VALIDATION, FAIRNESS REPORTING, AND IMPLEMENTATION RESEARCH REMAIN UNDERDEVELOPED. THIS ASYMMETRY MAY REFLECT REGULATORY PRESSURES (HEALTHCARE OPERATES UNDER HIPAA, FDA, AND ANALOGOUS FRAMEWORKS THAT INCENTIVIZE DOCUMENTATION) RATHER THAN INTRINSIC DIFFERENCES IN RESEARCH RIGOR.

SECOND, THE ACCELERATION OF DEEP LEARNING ADOPTION IN FINANCIAL FRAUD DETECTION DOCUMENTED BY CHEN ET AL. (2025) CARRIES BOTH OPPORTUNITY AND RISK. DEEP LEARNING MODELS HAVE DEMONSTRATED SUPERIOR PERFORMANCE IN DETECTING COMPLEX, NONLINEAR FRAUD PATTERNS THAT EVADE RULE-BASED SYSTEMS; HOWEVER, THEIR OPACITY CREATES INTERPRETABILITY BARRIERS THAT ARE PROBLEMATIC IN FINANCIAL REGULATORY CONTEXTS WHERE EXPLAINABILITY IS INCREASINGLY MANDATED. THE SAME DYNAMIC IS EMERGING IN HEALTHCARE, WHERE IMAGING-BASED DEEP LEARNING ACHIEVES STATE-

OF-THE-ART DIAGNOSTIC ACCURACY BUT FACES RESISTANCE FROM CLINICIANS AND REGULATORS DEMANDING JUSTIFIABLE PREDICTIONS.

THIRD, THE FINDING THAT ORGANIZATIONAL AND BEHAVIORAL BARRIERS—RATHER THAN ALGORITHMIC LIMITATIONS—ARE THE PRIMARY IMPEDIMENTS TO ML IMPLEMENTATION IN HEALTHCARE (PRETI ET AL., 2024) HAS DIRECT IMPLICATIONS FOR GRADUATE EDUCATION IN DATA SCIENCE. TRAINING PROGRAMS FOCUSING EXCLUSIVELY ON ALGORITHMIC DEVELOPMENT WITHOUT ATTENTION TO DEPLOYMENT CONTEXTS, CHANGE MANAGEMENT, AND HUMAN-COMPUTER INTERACTION MAY PRODUCE GRADUATES WHOSE TECHNICAL SKILLS OUTPACE THEIR CAPACITY TO TRANSLATE THOSE SKILLS INTO REAL-WORLD IMPACT. INTERDISCIPLINARY TRAINING SPANNING DATA SCIENCE, DOMAIN KNOWLEDGE, AND IMPLEMENTATION SCIENCE IS NEEDED.

THIS REVIEW HAS LIMITATIONS. THE NARRATIVE SYNTHESIS APPROACH, WHILE APPROPRIATE GIVEN METHODOLOGICAL HETEROGENEITY, PRECLUDES META-ANALYTIC ESTIMATION OF POOLED EFFECT SIZES. THE BROAD FIVE-DOMAIN SCOPE ENABLES CROSS-CUTTING SYNTHESIS BUT NECESSARILY EXCLUDES IMPORTANT AREAS SUCH AS NATURAL LANGUAGE PROCESSING AND CLIMATE FORECASTING. STUDIES PUBLISHED BEFORE 2022 MAY ALSO UNDERSTATE THE CURRENT STATE OF THE ART GIVEN THE PACE OF ADVANCEMENT IN THE FIELD.

7. CONCLUSION

DATA MINING AND MACHINE LEARNING HAVE ACHIEVED DEMONSTRABLE REAL-WORLD IMPACT ACROSS HEALTHCARE, FINANCIAL SERVICES, TRANSPORTATION, EDUCATION, AND ENVIRONMENTAL SCIENCE, WITH ENSEMBLE METHODS AND DEEP LEARNING ARCHITECTURES DRIVING THE DOMINANT SHARE OF REPORTED ADVANCES. HOWEVER, THE LITERATURE AS A WHOLE IS CHARACTERIZED BY SIGNIFICANT GAPS IN EXTERNAL VALIDATION, MODEL INTERPRETABILITY, FAIRNESS REPORTING, AND IMPLEMENTATION RESEARCH. THE PROPOSED HYPOTHESIS—THAT POST-HOC EXPLAINABILITY AUGMENTATION OF HIGH-PERFORMING ENSEMBLE CLASSIFIERS CAN PRESERVE PREDICTIVE PERFORMANCE WHILE IMPROVING END-USER TRUST—OFFERS A TRACTABLE, HIGH-IMPACT RESEARCH AGENDA DIRECTLY RESPONSIVE TO THESE GAPS.

ADDRESSING THE INTERPRETABILITY GAP IS NOT MERELY A TECHNICAL PROBLEM; IT IS A PREREQUISITE FOR RESPONSIBLE DEPLOYMENT OF ML

SYSTEMS IN CONTEXTS WHERE AUTOMATED DECISIONS DIRECTLY AFFECT HUMAN HEALTH, FINANCIAL SECURITY, AND EDUCATIONAL OPPORTUNITY. FUTURE RESEARCH SHOULD PRIORITIZE LONGITUDINAL IMPACT EVALUATION, FAIRNESS-AWARE MODELING FRAMEWORKS, CROSS-DOMAIN MODEL TRANSFERABILITY, AND ROBUST UNCERTAINTY QUANTIFICATION. ADVANCING THESE PRIORITIES THROUGH INTERDISCIPLINARY COLLABORATION WILL BE ESSENTIAL TO REALIZING THE FULL PROMISE OF DATA MINING AND MACHINE LEARNING IN THE REAL WORLD.

REFERENCES

ALHUMAIDI, N. H., DERMAWAN, D., KAMARUZAMAN, H. F., & ALOTAIQ, N. (2025). THE USE OF MACHINE LEARNING FOR ANALYZING REAL-WORLD DATA IN DISEASE PREDICTION AND MANAGEMENT: SYSTEMATIC REVIEW. JMIR MEDICAL INFORMATICS, 13, E68898. HTTPS://DOI.ORG/10.2196/68898

AN, Q., RAHMAN, S., ZHOU, J., & KANG, J. J. (2023). A COMPREHENSIVE REVIEW ON MACHINE LEARNING IN HEALTHCARE INDUSTRY: CLASSIFICATION, RESTRICTIONS, OPPORTUNITIES AND CHALLENGES. SENSORS, 23(9), 4178. HTTPS://DOI.ORG/10.3390/S23094178

BELLINGER, C., JABBAR, M. S. M., ZAÏANE, O., & OSORNIO-VARGAS, A. (2017). A SYSTEMATIC REVIEW OF DATA MINING AND MACHINE LEARNING FOR AIR POLLUTION EPIDEMIOLOGY. BMC PUBLIC HEALTH, 17(1), 907. HTTPS://DOI.ORG/10.1186/S12889-017-4914-3

CHEN, Y., ZHAO, C., XU, Y., NIE, C., & ZHANG, Y. (2025). YEAR-OVER-YEAR DEVELOPMENTS IN FINANCIAL FRAUD DETECTION VIA DEEP LEARNING: A SYSTEMATIC LITERATURE REVIEW. ARXIV. HTTPS://ARXIV.ORG/ABS/2502.00201

FUENTES, A., BOTTAI, M., & LEON-GARCIA, A. (2024). EXPLAINABLE AND INTERPRETABLE MACHINE LEARNING AND DATA MINING. DATA MINING AND KNOWLEDGE DISCOVERY, 38, 1–12. HTTPS://DOI.ORG/10.1007/S10618-024-01041-Y

GULERIA, P., & SOOD, M. (2023). EXPLAINABLE AI AND MACHINE LEARNING: PERFORMANCE EVALUATION AND EXPLAINABILITY OF CLASSIFIERS ON EDUCATIONAL DATA MINING INSPIRED CAREER COUNSELING. EDUCATION AND INFORMATION TECHNOLOGIES, 28(1), 1081–1116. HTTPS://DOI.ORG/10.1007/S10639-022-11221-2

HAN, J., PEI, J., & TONG, H. (2022). DATA MINING: CONCEPTS AND TECHNIQUES (4TH ED.). MORGAN KAUFMANN.

KHOSRAVI, H., SHUM, S. B., CHEN, G., CONATI, C., TSAI, Y.-S., KAY, J., KNIGHT, S., MARTINEZ-MALDONADO, R., SADIQ, S., & GAŠEVIĆ, D. (2022). EXPLAINABLE

ARTIFICIAL INTELLIGENCE IN EDUCATION. COMPUTERS AND EDUCATION: ARTIFICIAL INTELLIGENCE, 3, 100074. HTTPS://DOI.ORG/10.1016/J.CAEAI.2022.100074

KOLASA, K., ADMASSU, B., HOŁOWNIA-VOLOSKOVA, M., KĘDZIOR, K. J., POIRRIER, J.-E., & PERNI, S. (2024). SYSTEMATIC REVIEWS OF MACHINE LEARNING IN HEALTHCARE: A LITERATURE REVIEW. EXPERT REVIEW OF PHARMACOECONOMICS & OUTCOMES RESEARCH, 24(1), 63–115. HTTPS://DOI.ORG/10.1080/14737167.2023.2279107

NASSIF, A. B., ABU TALIB, M., NASIR, Q., & DAKALBAB, F. M. (2022). FINANCIAL FRAUD DETECTION BASED ON MACHINE LEARNING: A SYSTEMATIC LITERATURE REVIEW. APPLIED SCIENCES, 12(19), 9637. HTTPS://DOI.ORG/10.3390/APP12199637

PAGE, M. J., MCKENZIE, J. E., BOSSUYT, P. M., BOUTRON, I., HOFFMANN, T. C., MULROW, C. D., SHAMSEER, L., TETZLAFF, J. M., AKL, E. A., & BRENNAN, S. E. (2021). THE PRISMA 2020 STATEMENT: AN UPDATED GUIDELINE FOR REPORTING SYSTEMATIC REVIEWS. INTERNATIONAL JOURNAL OF SURGERY, 88, 105906. HTTPS://DOI.ORG/10.1016/J.IJSU.2021.105906

PRETI, L. M., ARDITO, V., COMPAGNI, A., PETRACCA, F., & CAPPELLARO, G. (2024). IMPLEMENTATION OF MACHINE LEARNING APPLICATIONS IN HEALTH CARE ORGANIZATIONS: SYSTEMATIC REVIEW OF EMPIRICAL STUDIES. JOURNAL OF MEDICAL INTERNET RESEARCH, 26, E55897. HTTPS://DOI.ORG/10.2196/55897

RAZA, S., QASIM, M., & KHAN, F. (2024). SYSTEMATIC REVIEW ON PRIVACY-PRESERVING MACHINE LEARNING TECHNIQUES FOR HEALTHCARE DATA. JOURNAL OF CYBER SECURITY TECHNOLOGY. HTTPS://DOI.ORG/10.1080/23742917.2025.2511145

REFONTE LEARNING. (2026). MACHINE LEARNING IN 2026: TRENDS, SKILLS, AND CAREER OPPORTUNITIES. REFONTE LEARNING BLOG.

SARKER, I. H. (2021). MACHINE LEARNING: ALGORITHMS, REAL-WORLD APPLICATIONS AND RESEARCH DIRECTIONS. SN COMPUTER SCIENCE, 2(3), 160. HTTPS://DOI.ORG/10.1007/S42979-021-00592-X

SUN, Z., REN, Y., LI, H., & WANG, X. (2025). THE USE OF MACHINE LEARNING FOR ANALYZING REAL-WORLD DATA IN DISEASE PREDICTION AND MANAGEMENT: SYSTEMATIC REVIEW. JMIR MEDICAL INFORMATICS, 13, E68898. HTTPS://DOI.ORG/10.2196/68898

TRIANTAFYLLIDIS, A. K., & TSANAS, A. (2019). APPLICATIONS OF MACHINE LEARNING IN REAL-LIFE DIGITAL HEALTH INTERVENTIONS: REVIEW OF THE LITERATURE. JOURNAL OF MEDICAL INTERNET RESEARCH, 21(4), E12286. HTTPS://DOI.ORG/10.2196/12286

VILCA, J., & AGUIRRE, R. (2024). FINANCIAL FRAUD DETECTION THROUGH THE APPLICATION OF MACHINE LEARNING TECHNIQUES: A LITERATURE REVIEW. HUMANITIES AND SOCIAL SCIENCES COMMUNICATIONS, 11, 1–15. HTTPS://DOI.ORG/10.1057/S41599-024-03606-0

WAHYONO, T., & DAVID, S. (2025). A SYSTEMATIC REVIEW OF MACHINE LEARNING-BASED FINANCIAL FRAUD DETECTION. JOURNAL OF SYSTEM AND MANAGEMENT SCIENCES, 15(1), 69–84.

ZHANG, H., LIU, Y., CHEN, X., & WANG, Z. (2025). A COMPREHENSIVE SURVEY ON DEEP LEARNING TECHNIQUES IN EDUCATIONAL DATA MINING. DATA SCIENCE AND ENGINEERING. HTTPS://DOI.ORG/10.1007/S41019-025-00303-Z

ZHANG, J., WANG, J., ZANG, H., MA, N., SKITMORE, M., QU, Z., SKULMOSKI, G., & CHEN, J. (2024). THE APPLICATION OF MACHINE LEARNING AND DEEP LEARNING IN INTELLIGENT TRANSPORTATION: A SCIENTOMETRIC ANALYSIS AND QUALITATIVE REVIEW OF RESEARCH TRENDS. SUSTAINABILITY, 16(14), 5879. HTTPS://DOI.ORG/10.3390/SU16145879

ZRIQAT, I. A., ALTAMIMI, A. M., & AZZEH, M. (2017). A COMPARATIVE STUDY FOR PREDICTING HEART DISEASES USING DATA MINING CLASSIFICATION METHODS. ARXIV. HTTPS://ARXIV.ORG/ABS/1704.02799