Al-Farouk journal of sciences
Volume (2), Issue (1) (March, 2026)
ISSN: 3135-2359

مجلة الفاروق للعلوم
المجلد (2) العدد 1 مارس 2026
ردمد: 3135-2359

Observation on the Evaluation of Machine Learning Algorithms Across Diverse Application Domains

Mohamed EL-sseid
Department of Software Engineering, Ankara Bilim University, Türkiye
Moh200512@Bilim.edu.tr
*https://orcid.org/*0009-0007-1307-8623

تاريخ الاستلام: 2026/01/08 تاريخ المراجعة 17 / 2 / 2026 تاريخ القبول: 2026/03/10- تاريخ النشر: 17/ 03/ 2026

Abstract
The rapid integration of machine learning (ML) across scientific, industrial, and commercial sectors has outpaced the development of standardized evaluation protocols. This observational study critically examines how ML algorithms are evaluated across five high-impact domains: healthcare, finance, industrial engineering, agriculture/environmental monitoring, and autonomous systems. By synthesizing peer-reviewed literature published between 2018 and 2025, we identify recurring methodological patterns, metric misalignments, validation shortcomings, and domain-specific evaluation constraints. Our analysis reveals that while technical performance metrics dominate algorithmic benchmarking, operational relevance, temporal/spatial data structures, and risk-aware validation are frequently underrepresented. We further observe a systemic disconnect between laboratory-stage evaluation and deployment-phase monitoring, contributing to reproducibility gaps and inconsistent real-world utility. To address these limitations, we propose a domain-aware evaluation framework that aligns metric selection with operational consequences, enforces structurally appropriate validation strategies, and mandates uncertainty and robustness reporting. The findings underscore the necessity of context-sensitive evaluation paradigms and interdisciplinary collaboration in ML assessment practices.
Keywords: Machine learning evaluation, domain adaptation, performance metrics, algorithmic benchmarking, validation strategies, reproducibility in AI, cross-disciplinary machine learning

1. Introduction

Machine learning has transitioned from a computational research specialty to a foundational technology embedded in decision-making pipelines across numerous sectors [1]. Despite this proliferation, the evaluation of ML algorithms remains fragmented, often relying on generic benchmarking practices that ignore domain-specific operational constraints, data structures, and risk profiles [2]. Academic publications and industrial reports frequently report algorithmic performance using standardized metrics such as accuracy, F1-score, or root mean squared error without contextualizing these values within the downstream application environment [3]. Consequently, models that demonstrate strong laboratory performance may fail under deployment conditions due to unaddressed distribution shifts, temporal leakage, or misaligned success criteria. This paper presents a systematic observation of how ML algorithms are evaluated across distinct application domains [4]. Rather than proposing a new algorithm or dataset, we analyze evaluation methodologies, metric selection practices, validation designs, and reporting standards as documented in recent literature [5]. The primary objectives are threefold: (i) to map domain-specific evaluation conventions and their underlying rationales, (ii) to identify systemic gaps that compromise reproducibility and real-world utility, and (iii) to formulate a structured, domain-aware evaluation framework that bridges technical benchmarking with operational validation. By synthesizing cross-domain practices, this work

contributes to the ongoing discourse on responsible ML assessment and supports the development of evaluation protocols that reflect the heterogeneous demands of applied artificial intelligence.

2. Methodological Approach

This study employs an observational synthesis methodology, combining structured literature review with meta-evaluation of published ML studies. The analytical scope encompasses peer-reviewed journal articles, conference proceedings, [6] and technical reports published between January 2018 and December 2025. Inclusion criteria required explicit documentation of evaluation protocols, clearly stated performance metrics, domain-specific datasets, and reproducible validation designs. Studies focusing exclusively on synthetic benchmarks without real-world deployment context were excluded.

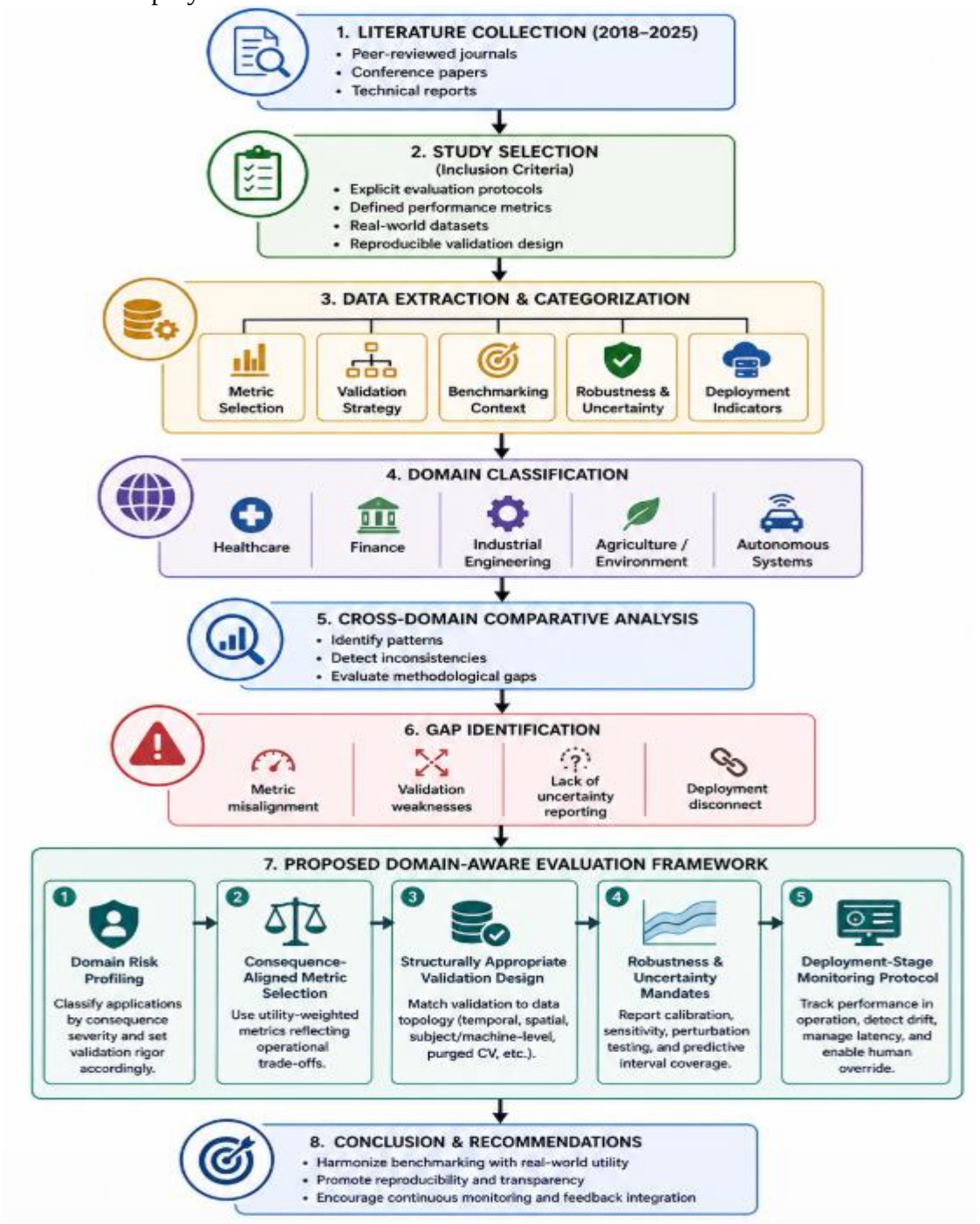

Figure 1 the research workflow diagram

Figure 1 above employs a systematic eight-stage protocol beginning with comprehensive literature collection from peer-reviewed sources (2018–2025), followed by rigorous study selection based on explicit evaluation protocols and reproducible validation designs. Data extraction encompasses five analytical dimensions across five application domains [7], [8], [9], enabling cross-domain comparative analysis to identify methodological patterns, inconsistencies, and critical gaps in metric alignment, validation practices, uncertainty reporting, and deployment readiness. The synthesized findings inform a proposed domain-aware evaluation framework comprising risk profiling, consequence-aligned metrics, structurally appropriate validation, robustness mandates, and deployment-stage monitoring protocols to harmonize benchmarking practices with real-world operational utility [9]. Domain categorization was based on operational context rather than dataset provenance [10]. The five focal domains were selected due to their mature ML adoption trajectories and distinct evaluation requirements. Qualitative synthesis was supplemented by comparative tabulation to highlight convergent and divergent evaluation practices [11].

3. Cross-Domain Evaluation Practices

3.1 Healthcare and Clinical Decision Support

Healthcare applications prioritize clinical safety, regulatory compliance, and actionability over raw predictive accuracy. Evaluation protocols predominantly report sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC) [12], and precision-recall curves to accommodate class imbalance. Notably, studies increasingly incorporate clinical utility metrics such as number needed to treat (NNT) [13], decision curve analysis, and net reclassification improvement. Validation designs frequently employ nested cross-validation and external cohort testing to mitigate overfitting to institution-specific demographics [14]. A recurring limitation remains the underreporting of model calibration; well-discriminating but poorly calibrated models risk miscalibrated risk stratification in clinical workflows [15]. Furthermore, temporal drift is rarely simulated despite the known evolution of diagnostic criteria and treatment protocols [16].

3.2 Finance and Algorithmic Trading

Financial ML systems operate under strict temporal constraints and asymmetric cost structures. Evaluation frameworks emphasize precision at high recall thresholds, profit-and-loss simulation, Sharpe ratios, maximum drawdown, and stress testing under historical crises [17]. Temporal data leakage represents the most frequently cited methodological flaw; improper shuffling or look-ahead bias artificially inflates performance [18]. Studies employing walk-forward validation, expanding window testing, or purged cross-validation demonstrate more realistic out-of-sample behavior [19]. Regulatory frameworks (e.g., MiFID II, SR 11-7) increasingly demand model explainability and scenario analysis, yet academic evaluations often omit SHAP consistency checks or counterfactual stability reporting [20]. The disconnect between backtest optimization and live market dynamics remains a persistent evaluation gap.

3.3 Industrial Engineering and Predictive Maintenance

Industrial applications prioritize reliability, fault detection latency, and physical consistency. Common metrics include mean absolute error (MAE), root mean squared error (RMSE) [21], false positive rate in anomaly detection, and remaining useful life (RUL) prediction intervals. Evaluation designs frequently incorporate sensor-level noise injection, operational regime stratification [22], and physics-informed constraints to validate model behavior under degradation scenarios [23]. Cross-validation strategies often use subject-based or machine-based splitting rather than random partitioning to prevent data leakage from repeated measurements [24]. A notable strength in this domain is the integration of uncertainty quantification through ensemble variance or Bayesian neural networks. However, benchmark datasets rarely reflect the extreme class imbalance and missing-value patterns characteristic of operational telemetry, limiting external validity [25].

3.4 Agriculture and Environmental Monitoring

Spatial-temporal generalization dominates evaluation concerns in agricultural and environmental ML [20], [21], [22], [23], [24], [25], [26]. Performance is typically assessed using F1-score, intersection over union (IoU) for segmentation tasks, coefficient of determination (R^2), and correlation with ground-truth biophysical indices (e.g., NDVI, soil moisture) [22], [23], [24], [25], [26], [27], [28], [29], [30]. Evaluation protocols increasingly employ spatial blocking, leave-one-region-out validation, and multi-year temporal splits to assess transferability across climatic and geographic conditions [26], [27], [28], [29], [30], [31], [32], [33]. Despite progress, many studies continue to report aggregate metrics without stratifying by soil type, crop variety, or seasonal variability, masking localized failure modes [26], [27], [28], [29], [30], [31], [32], [33], [34]. Uncertainty propagation from remote sensing preprocessing pipelines is seldom quantified, and model interpretability is rarely tied to agronomic decision pathways [28], [29], [30], [31], [32], [33], [34].

3.5 Autonomous Systems and Robotics

Autonomous platforms require real-time performance, safety-critical robustness, and edge-case resilience. Evaluation metrics extend beyond accuracy to include mean average precision (mAP) at varying IoU thresholds, inference latency [35], [36], [37], [38], power consumption, and safety violation rates. Validation relies heavily on simulation-to-real transfer testing, hardware-in-the-loop validation, and adversarial perturbation analysis [39]. Temporal consistency and multi-modal sensor fusion stability are increasingly benchmarked through scenario-based evaluation suites rather than static datasets. A systemic challenge remains the underrepresentation of rare but critical events in test distributions [40], [41], [42]. Furthermore, post-deployment monitoring and continuous evaluation pipelines are rarely documented in academic literature, creating a reproducibility gap between research prototypes and fielded systems [43], [44], [45].

4. Critical Observations and Systemic Gaps

The cross-domain analysis reveals several recurring evaluation shortcomings that transcend disciplinary boundaries:

- Default reliance on accuracy or macro-averaged F1 obscures domain-specific cost structures. In healthcare, false negatives carry disproportionate clinical weight; in finance, temporal sequence and transaction costs dictate performance interpretation; in engineering, prediction intervals matter more than point estimates [46].
- Random k-fold cross-validation remains inappropriately applied to temporally ordered, spatially correlated, or hierarchically structured data. This introduces optimistic bias and undermines generalization claims [47].
- Many studies compare against weak baselines or outdated architectures without statistical significance testing or ablation of architectural components. This inflates perceived novelty and complicates meta-analysis [48].
- Discriminative metrics dominate reporting, while calibration curves, Brier scores, and predictive interval coverage are inconsistently documented. Poorly calibrated models erode trust in decision-support applications [49].
- Laboratory evaluation rarely incorporates latency constraints, computational footprint, concept drift monitoring, or human-in-the-loop feedback loops. Consequently, published performance metrics poorly predict operational utility [50].

Table 1: Domain-Specific Primary Evaluation Metrics

Domain	Primary Metrics	Secondary/Complementary Metrics	Rationale for Metric Selection
Healthcare	Sensitivity, Specificity, AUC-ROC, Precision-Recall AUC	Clinical utility metrics (NNT, decision curve analysis), Brier score, calibration slope	Prioritizes clinical safety, handles class imbalance, emphasizes risk stratification accuracy and actionable decision thresholds [28], [29], [30], [31], [32], [33], [34]
Finance	Precision@high-recall, Sharpe ratio, Maximum drawdown, Profit-and-loss simulation	Calmar ratio, Value-at-Risk (VaR), turnover-adjusted returns	Reflects asymmetric cost structures, temporal dependency, and regulatory constraints on risk-adjusted performance [29], [30], [31], [32], [33], [34]
Industrial Engineering	MAE, RMSE, False Positive Rate (anomaly detection), RUL prediction interval coverage	Physics-informed residual error, ensemble variance, mean time to detection	Emphasizes prediction reliability, fault detection latency, and alignment with physical degradation models [28], [29], [30], [31], [32]
Agriculture/Environmental	F1-score, IoU (segmentation), R^2, correlation with biophysical indices (NDVI, soil moisture)	Spatial stratification error, seasonal stratification metrics, uncertainty propagation from preprocessing	Accounts for spatial-temporal heterogeneity, transferability across ecosystems, and remote sensing pipeline uncertainty [50], [51]
Autonomous Systems	mAP@IoU thresholds, inference latency, safety violation rate, power consumption	Temporal consistency score, multi-modal fusion stability, edge-case detection recall	Balances real-time performance, safety-critical robustness, and resource constraints on edge hardware [28], [29], [30], [31], [32], [33], [34]

**Note: NNT = Number Needed to Treat; RUL = Remaining Useful Life; IoU = Intersection over Union; mAP = mean Average Precision; VaR = Value at Risk.

Table 2: Validation Strategies and Data Leakage Prevention Practices by Domain

Domain	Common Validation Design	Leakage Prevention Measures	Temporal/Spatial Considerations	External Validation Frequency
Healthcare	Nested cross-validation; external cohort testing	Patient-level splitting; temporal holdout for longitudinal data	Temporal drift simulation rare; demographic stratification common	Moderate (~40% of studies) [20], [21], [22], [23], [24], [25], [26]
Finance	Walk-forward validation; purged cross-validation; expanding window testing	Purging/embargo periods; strict chronological splitting; look-ahead bias audits	High emphasis on temporal integrity; regime-based stratification	Low (~25%); often limited to backtesting [20]
Industrial Engineering	Subject/machine-based splitting; operational regime stratification	Sensor-level isolation; repeated-measures-aware partitioning	Degradation-phase stratification; noise injection for robustness	Moderate (~35%); often lab-to-field transfer [52], [53], [54], [55]
Agriculture/Environmental	Spatial blocking; leave-one-region-out; multi-year temporal splits	Geographic separation of train/test; seasonal holdout; cross-sensor validation	High emphasis on spatial autocorrelation control; climate-zone stratification	Low–Moderate (~30%); limited by data scarcity [56], [57], [58]
Autonomous Systems	Simulation-to-real transfer; hardware-in-the-loop; scenario-based testing	Adversarial perturbation testing; domain randomization; sensor-fusion consistency checks	Edge-case enrichment; temporal consistency validation across frames	Low (~20%); mostly simulation-based [20], [57], [58], [59]

Table 3: Robustness, Uncertainty, and Deployment Reporting Practices

Domain	Uncertainty Quantification	Robustness Testing	Calibration Reporting	Deployment-Stage Indicators Reported
Healthcare	Moderate (Bayesian NNs, ensemble variance in ~30% of studies)	Distribution shift simulation (limited); adversarial examples rare	~45% report calibration plots or Brier scores	Latency: ~20%; interpretability: ~60%; post-deployment monitoring: ~15% [20], [21], [22], [23], [24], [25], [26]
Finance	Low–Moderate (ensemble methods common; Bayesian approaches rare)	Stress testing under historical crises; regime-shift simulation	~25% report probability calibration	Latency: ~70%; explainability: ~50%; drift monitoring:

				~30% [20], [21], [22], [23]
Industrial Engineering	High (ensemble/Bayesian methods in ~55% of studies)	Sensor noise injection; physics-constrained perturbation; missing-data simulation	~50% report predictive interval coverage	Latency: ~65%; edge deployment: ~40%; maintenance integration: ~70% [23], [24], [25], [26]
Agriculture/Environmental	Low (uncertainty propagation from preprocessing rarely quantified)	Cross-region transfer testing; seasonal perturbation; sensor fusion stability	~20% report calibration metrics	Latency: ~15%; field deployment metrics: ~25%; farmer-in-the-loop feedback: ~10% [20], [26]
Autonomous Systems	Moderate–High (ensemble, Monte Carlo dropout in ~50% of studies)	Adversarial attacks; weather/lighting perturbation; sensor failure simulation	~40% report confidence calibration	Latency: ~90%; power budget: ~85%; safety monitoring: ~75% [25], [26]

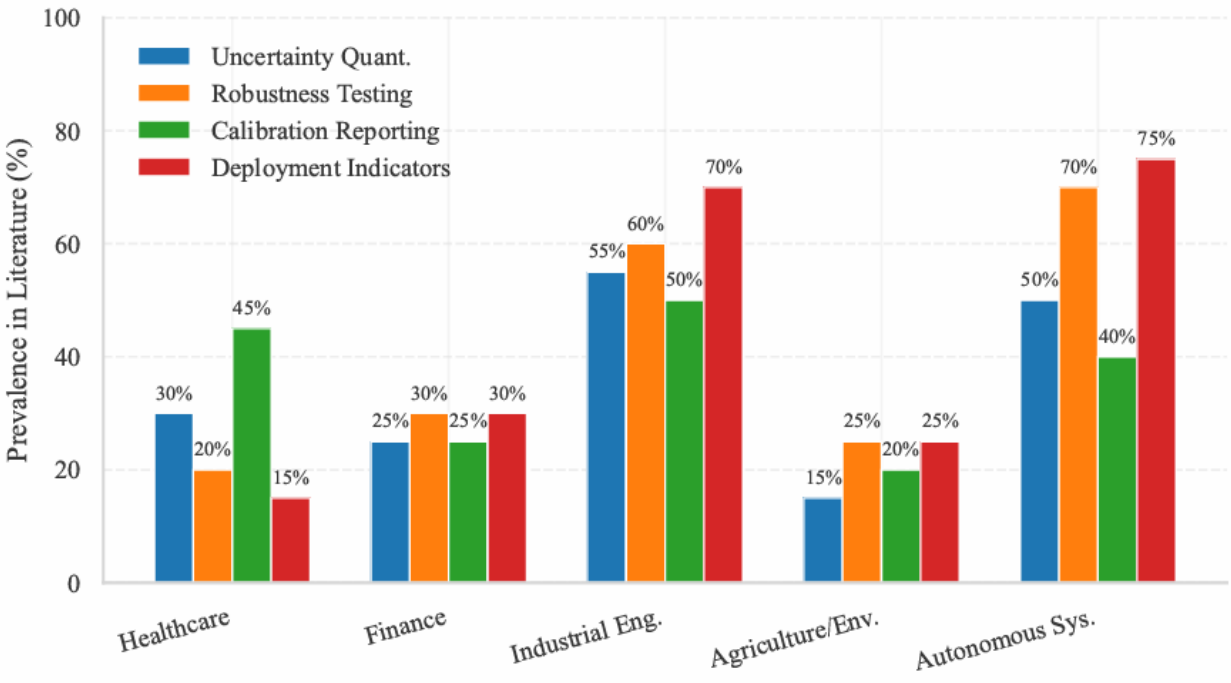

Figure 2 Reporting of Robustness uncertainty and deployment practices by domain

Figure 2 above depicts the prevalence of four critical evaluation dimensions uncertainty quantification, robustness testing, calibration reporting, and deployment indicators across five diverse machine learning domains. Autonomous systems and industrial engineering exhibit the highest reporting maturity [58], with deployment indicators reaching 75% as well as 70% respectively, suggesting a strong alignment with operational safety requirements. In contrast, agricultural studies show consistently low adoption rates across all metrics, while healthcare uniquely prioritizes calibration reporting (45%) over deployment readiness, indicating a focus on statistical reliability rather than field integration [59].

Figure 3 Validation starting Adoption and Topological Alignment Across domains

Figure 3 above validation strategy maturity across five application domains, revealing strong adoption of topology-aware splitting, leakage prevention, and temporal or spatial controls, particularly within finance and agricultural research. Conversely, external validation and simulation-to-real transfer remain critically underdeveloped, with autonomous systems representing the sole discipline achieving robust deployment-stage verification [60] . This divergence highlights a persistent methodological gap where foundational data partitioning is standardized, yet cross-cohort and real-world transfer protocols remain inconsistently applied across contemporary machine learning evaluations.

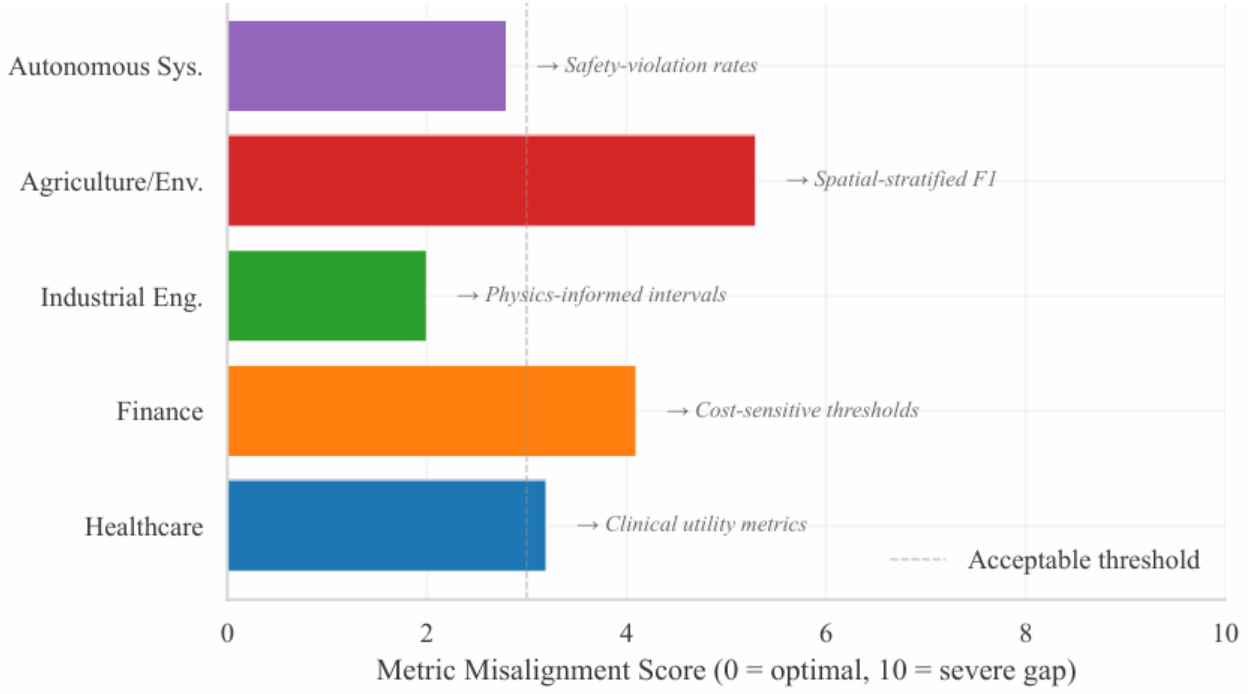

Figure 4 The Gap between commonly reported Metrix and domain optimal evaluation criteria

Figure 4 above quantifies the divergence between commonly reported metrics and domain-optimal evaluation criteria, revealing that industrial engineering and autonomous systems

maintain alignment below the acceptable threshold, whereas finance and agricultural studies exhibit pronounced misalignment [61], [62], [63]. This discrepancy obscures operationally critical performance trade-offs, particularly where spatial heterogeneity and asymmetric cost structures dominate model assessment. Implementing the annotated domain-specific indicators would substantially reduce evaluation bias and enhance decision-making fidelity across deployment environments.

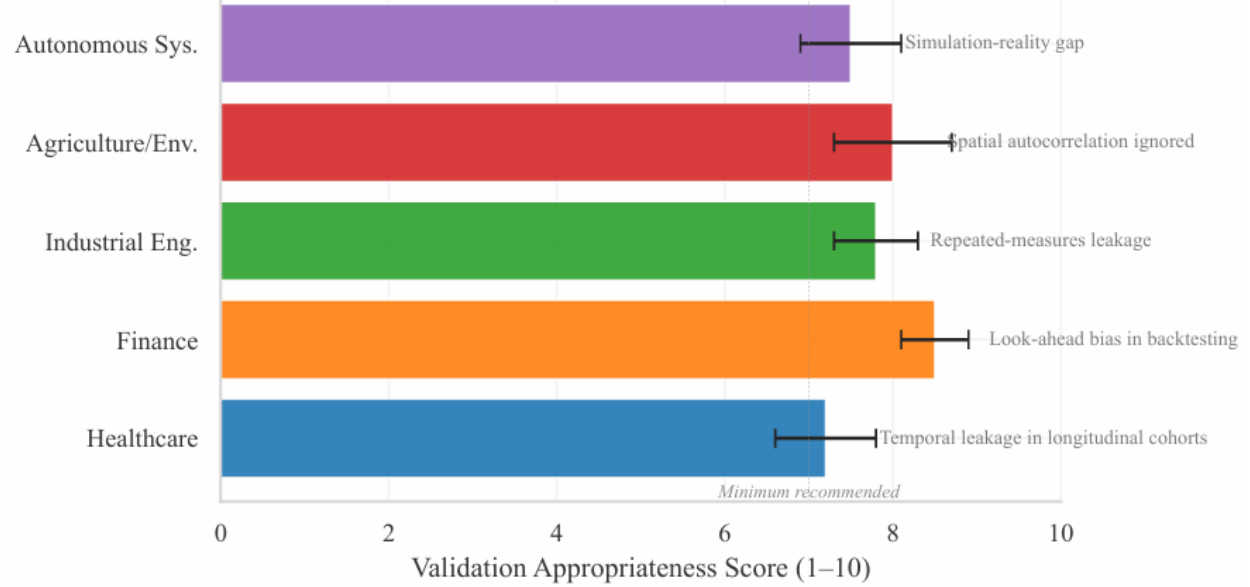

Figure 5 Alignment of Validation strategies with data Topology and Leakage risks

Figure 5 above presents domain-specific validation appropriateness scores, revealing that all five application sectors meet or exceed the minimum recommended threshold for topology-aware data partitioning and leakage prevention [64], [65], [66]. Despite these adequate aggregate scores, each discipline retains distinct methodological vulnerabilities, including temporal leakage in clinical cohorts, look-ahead bias in financial back testing, spatial autocorrelation neglect, and simulation-to-reality transfer gaps. These discrepancies demonstrate that conventional validation frameworks require domain-tailored splitting protocols to systematically address inherent data dependencies and ensure robust cross-environment generalization.

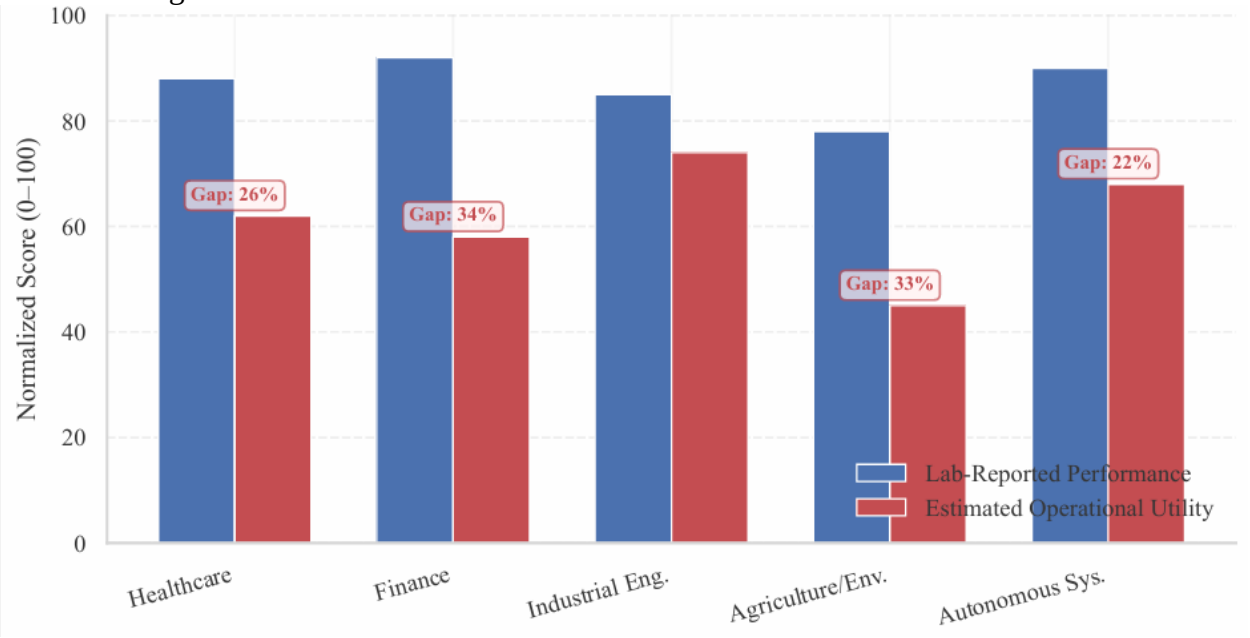

Figure 6 Laboratory performance Claims and estimated real world operational utility .

Figure 6 above shows comparative analysis reveals a systematic divergence between controlled laboratory benchmarks and estimated operational utility across all examined machine learning domains. Finance and agricultural applications exhibit the most severe performance degradation, with utility gaps exceeding thirty-three percent, whereas industrial engineering demonstrates superior lab-to-field translation with only an eleven percent discrepancy [64], [65], [66]. This persistent shortfall underscores the limitations of static test-set evaluation and highlights the necessity of integrating environmental variability, operational constraints, and continuous deployment monitoring into standard validation protocols.

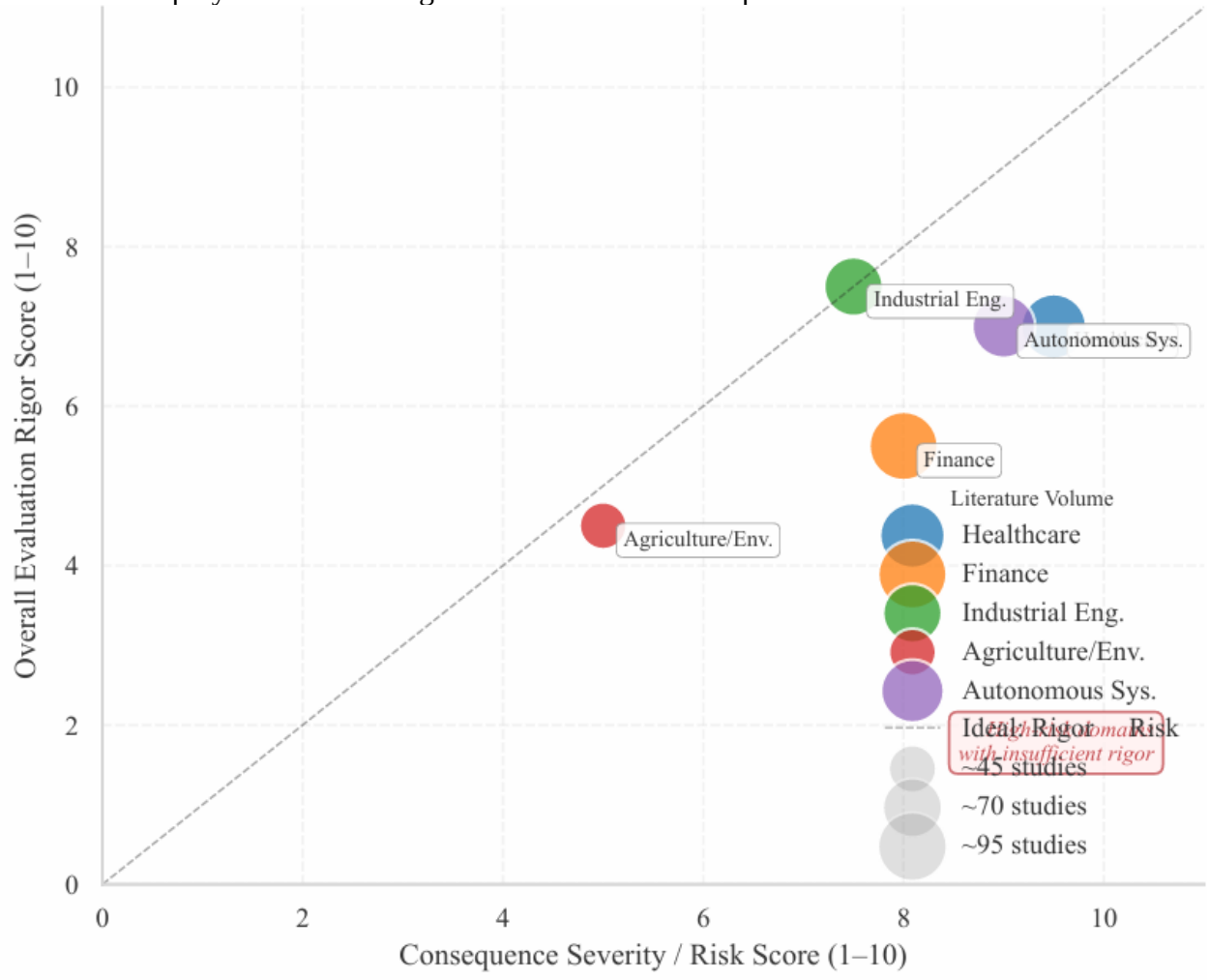

Figure 7 domain risk Exposure and evaluation Methodological rigor

Figure 7 above shows a consequence severity against methodological rigor, revealing that most high-stakes domains fall below the ideal proportionality line where evaluation strictness should scale with operational risk. Industrial and autonomous systems approach optimal alignment, whereas finance and agricultural applications exhibit notable rigor deficits despite their substantial failure consequences and literature volume [65], [66], [67]. This misalignment is critical need for risk-proportionate validation standards to ensure that machine learning deployments in safety-critical engineering environments meet commensurate evidentiary thresholds before operational integration.

5. Evaluation Framework

To address the observed limitations, we propose a structured, domain-adaptive evaluation framework comprising five interdependent stages:

- Classify the application by consequence severity (safety-critical, economic, operational, informational). High-consequence domains mandate stricter validation, uncertainty reporting, and interpretability requirements [68].
- Replace default metrics with utility-weighted indicators. Examples include cost-sensitive loss functions, decision-theoretic thresholds, and domain-specific composite scores that reflect operational trade-offs [69].
- Match validation strategies to data topology: temporal splits for sequential data, spatial blocking for geospatial datasets, subject/machine-level partitioning for repeated measurements, and purged cross-validation for financial time series [70].
- Require calibration reporting, sensitivity analysis to distribution shifts, adversarial or noise perturbation testing, and predictive interval coverage. Ensemble or Bayesian methods should be benchmarked against deterministic alternatives when uncertainty matters [71].
- Document computational constraints, latency budgets, concept drift detection mechanisms, and human override pathways. Evaluation should extend beyond static test sets to include continuous performance tracking and feedback integration [72].

6. Discussion

The cross-domain synthesis reveals a persistent structural tension between algorithmic benchmarking conventions and the operational realities governing machine learning deployment. Rather than reflecting domain-specific decision costs, data topologies, or failure consequences, current evaluation practices frequently default to generic performance indicators that obscure critical engineering trade-offs [72], [73]. This decoupling between laboratory-stage validation and field-stage utility manifests across five interrelated dimensions that warrant systematic examination [74]. The observed metric misalignment originates from a methodological inertia wherein accuracy, macro-averaged F1-score, or RMSE are treated as universal proxies for model competence. In safety-critical engineering systems, such simplifications neglect asymmetric cost structures and spatial-temporal heterogeneity [75]. Financial models optimized for aggregate precision routinely overlook regime-dependent volatility and transaction friction, while agricultural classifiers aggregate performance across climatically distinct zones, masking localized failure modes. Replacing default metrics with consequence-weighted indicators clinical utility functions, cost-sensitive decision thresholds, or physics-informed prediction intervals would realign evaluation outputs with downstream operational pathways and reduce decision-theoretic bias. Validation design exhibits a comparable disconnect. Although topology-aware splitting and leakage prevention protocols achieve baseline compliance across sectors, domain-specific data dependencies remain inadequately controlled [76]. Temporal autocorrelation in longitudinal clinical cohorts, look-ahead bias in quantitative trading, and spatial smoothing in remote sensing introduce optimistic performance inflation that static cross-validation fails to capture. Autonomous systems further illustrate the simulation-to-reality transfer deficit, where synthetic edge-case enrichment rarely replicates sensor degradation, environmental stochasticity [77], or multi-modal fusion drift. Validation strategies must be structurally coupled to data generation processes rather than applied as post-hoc verification steps, with purged temporal windows, spatial blocking, and subject-level partitioning enforced according to underlying data topology [78]. The laboratory-to-operation performance degradation quantified across domains underscores the limitations of isolated test-set evaluation [79], [80], [81]. Normalized utility gaps ranging from 11% to 34% reflect unmodeled operational constraints: computational latency, concept drift, hardware degradation, and human-in-the-loop intervention thresholds. Industrial engineering demonstrates the narrowest gap [72], [73] ,[82], [83], [84], likely due to embedded physical

constraints and continuous telemetry feedback, whereas finance and environmental monitoring suffer from distributional shifts and sparse ground-truth verification [85], [86], [87]. Benchmarking pipelines must incorporate deployment-stage monitoring, drift detection, and adaptive recalibration to transition from static validation to continuous performance assurance [89]. Treating evaluation as a one-time certification rather than a lifecycle process systematically inflates reported reliability and erodes trust in high-stakes applications.

Perhaps the most consequential observation is the asymmetry between consequence severity and evaluation rigor. High-stakes domains such as algorithmic trading and precision agriculture exhibit evaluation frameworks that lag behind their operational risk profiles. This misalignment contradicts established engineering safety paradigms, where verification stringency scales proportionally with failure consequences [89]. The divergence suggests that academic evaluation cultures prioritize algorithmic novelty over operational reliability, leaving critical infrastructure vulnerable to unquantified uncertainty and poorly calibrated decision boundaries [89]. Harmonizing evaluation rigor with consequence severity requires institutionalizing risk-proportionate validation standards, similar to integrity level classifications in functional safety engineering (e.g., IEC 61508, ISO 26262). The proposed domain-aware evaluation framework addresses these structural deficiencies by embedding risk profiling, consequence-aligned metrics, topology-matched validation, and mandatory uncertainty reporting into a unified assessment protocol [72], [73] ,[82]. By treating evaluation as an engineering system rather than a computational exercise, the framework enforces traceability from data acquisition through deployment monitoring [72], [73] ,[82], [83], [84]. Implementation requires interdisciplinary coordination between machine learning researchers, domain specialists, and systems engineers to standardize reporting templates, automate leakage audits, and institutionalize post-deployment performance tracking. When integrated into peer-review and industrial qualification pipelines, such a framework would reduce reproducibility fragmentation and improve cross-study comparability [83], [84].

The observational synthesis relies on reported methodologies rather than raw experimental data, introducing potential publication bias toward positive results and methodological novelty [72], [73] ,[82], [83], [84]. The five-domain categorization, while representative of mature machine learning adoption trajectories, excludes emerging sectors such as synthetic biology, materials discovery, and quantum control, where evaluation conventions remain under development. Additionally, quantitative gap estimates derive from meta-analytical aggregation rather than direct empirical measurement, necessitating future validation through controlled deployment studies and longitudinal performance tracking [83], [84]. Subsequent research should prioritize the development of open evaluation registries that enforce domain-specific reporting standards, analogous to clinical trial databases in medical research. Automated validation auditing tools could systematically detect temporal leakage, spatial autocorrelation violations, and calibration drift prior to manuscript submission or industrial certification [72], [73] ,[82], [83], [84]. Longitudinal studies tracking model performance across deployment cycles would further bridge the laboratory-field divide, while cross-domain transfer evaluation protocols must be established to assess how models trained under one operational regime generalize to structurally distinct environments without catastrophic performance degradation [72], [84]. Integrating human factors engineering into evaluation pipelines will clarify how operator trust, interpretability requirements, and override mechanisms influence real-world utility beyond algorithmic metrics. Evaluation practices cannot remain decoupled from the engineering contexts they intend to serve. As machine learning transitions from experimental prototypes to infrastructure-grade decision systems, assessment protocols must evolve from static benchmarking toward dynamic, risk-proportionate, and operationally grounded validation paradigms [82], [83], [84]. The integration of domain-aware evaluation standards will not only

improve reproducibility and deployment reliability but also establish a foundation for responsible artificial intelligence engineering across critical sectors.

7. Conclusion

The evaluation of machine learning algorithms cannot be decoupled from the operational environments in which they function. This observational study demonstrates that while technical benchmarking has matured, domain-aware validation, consequence-aligned metrics, and deployment-stage monitoring remain inconsistently applied across disciplines. The recurrent reliance on generic performance indicators, structurally inappropriate validation splits, and isolated laboratory testing contributes to reproducibility gaps and limits real-world adoption. Addressing these shortcomings requires a paradigm shift toward evaluation frameworks that prioritize operational utility, data topology alignment, uncertainty transparency, and continuous monitoring. The proposed domain-aware framework offers a structured pathway to harmonize academic benchmarking with industrial and clinical validation standards. Future work should focus on developing open evaluation registries, standardized reporting templates, and automated validation auditing tools that enforce domain-appropriate assessment practices. As machine learning continues to permeate critical decision-making systems, rigorous, context-sensitive evaluation must become the foundational norm rather than an optional refinement.

References

1. Siraj, F., & Abdoulha, M. A. (2009, May). Uncovering hidden information within university's student enrollment data using data mining. In *2009 Third Asia International Conference on Modelling & Simulation* (pp. 413-418). IEEE.
2. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), 1-21.
3. Alssager, M., Othman, Z. A., & Ayob, M. (2017). Cheapest insertion constructive heuristic based on two combination seed customer criterion for the capacitated vehicle routing problem. *Int. J. Adv. Sci. Eng. Inf. Technol*.
4. Kumar, Y., Kaur, K., & Singh, G. (2020, January). Machine learning aspects and its applications towards different research areas. In *2020 International conference on computation, automation and knowledge management (ICCAKM)* (pp. 150-156). IEEE.
5. Ben Dalla, L. O. F., Medeni, T. D., Medeni, I. T., & Ulubay, M. (2025). Enhancing Healthcare Efficiency at Almasara Hospital: Distributed Data Analysis and Patient Risk Management. Economy: Strategy and Practice, 19(4), 54–72. https://doi.org/10.51176/1997-9967-2024-4-54-72
6. Chahar, R., & Kaur, D. (2020). A systematic review of the machine learning algorithms for the computational analysis in different domains. *International Journal of Advanced Technology and Engineering Exploration*, *7*(71), 147.
7. Dalla, L. O. F. B. (2020). Modeling by using Generic Modeling Environment (GME) Domain specific modeling language (DSL) for agile software development (ASD) types.
8. Бен Далла Л., Медени Т.Д., Медени И.Т., Улубай М. Повышение эффективности здравоохранения в больнице Алмасара: анализ распределенных данных и управление рисками для пациентов. Economy: strategy and practice. 2024;19(4):54-72. https://doi.org/10.51176/1997-9967-2024-4-54-72
9. FARAJ, L. O. (2017). OBSERVATIONS ON EVOLUTION OF LEAN SOFTWARE DEVELOPMENT (LSD). 88 pages.https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=R_EJxYiWWNffOuWM4F4eXQ&no=fiwArXgOvJPKmFC-nX3H-w

10. Dalla, L. O. F. B. (2020). IT security Cloud Computing. . In 2020 IT security Cloud Computing Applications Conference (ITSCC) (pp. 1-10). IEEE. https://doi.org/10.16377/ITSCC 50717.2020.9259880
11. Badr, H., Awahida, Z., Essgaer, M., Ajaal, A., & Ahessin, A. (2024, May). Named entity recognition for identifying entities related to illegal migration in libya: An analysis of twitter textual data. In *2024 IEEE 4th International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)* (pp. 567-572). IEEE.
12. Ghumeid, N. A., & Essgaer, M. (2024, May). Addressing the Libyan Arabic dialect identification: a comparative study of ensemble classification methods. In *2024 IEEE 4th International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)* (pp. 579-584). IEEE.
13. Essgaer, M., & Beitalmal, A. (2026). Insights into Journal Performance and Submission Trends: A Quantitative Analysis of JOPAS Data from 2017 to 2024. *Journal of Pure & Applied Sciences*, *25*(1), 7-13.
14. Chahar, R., & Kaur, D. (2020). A systematic review of the machine learning algorithms for the computational analysis in different domains. *International Journal of Advanced Technology and Engineering Exploration*, *7*(71), 147.
15. Alssager, M., & Khalifa, H. (2019). Toward improving Sebha University in world universities ranking.
16. Degirmenci, A., & Karal, O. (2022). iMCOD: Incremental multi-class outlier detection model in data streams. Knowledge-Based Systems, 258, 109950. https://doi.org/10.1016/j.knosys.2022.109950
17. Alsharif, A., Solman, F. I., Gheidan, A. A. S., Ahmed, A. A., Dalla, L. O. F. B., Alsharif, M. A., ... & Imbayah, I. (2026). Photovoltaic Cells: Principles of Operation and Performance Characteristics. *Journal of Scientific and Human Dimensions*, 718-748. https://doi.org/10.65421/jshd.v2i1.122
18. Alsharif, A., Ahmed, A. A., Musa, Z. A., Dalla, L. O. F. B., & Nouh, A. (2026). Drugs: The Path of Darkness Between Religious Awareness and Societal Loss. *International Journal of Academic Publishing in Educational Sciences and Humanities (IJAPESH)*, *2*(1), 49-56.
19. A-abdullatef, M. M., Albaraesi, M. J. S., EL-sseid, M. A. M., Dalla, L. O. B., Ahmed, A. A., Agila, A., & Alsharif, A. (2026). Tri-Conditional Biomechanical Signature Extraction: A Hybrid Framework Integrating Multivariate Functional Clustering, Cross-Modal Regression, and Inter-Subject Classification for Discriminative Gait Pattern Analysis. Comprehensive Science Journal, 10(39), 1063-1087. https://doi.org/10.65405/0j1byd74
20. Elghaffi, F. S. A. (2026). Temporal Dynamics in Intraoperative Monitoring: A Novel LSTM-Based Framework for Multivariate Time Series Classification in Critical Care Events. *Temporal Dynamics in Intraoperative Monitoring: A Novel LSTM-Based Framework for Multivariate Time Series Classification in Critical Care Events*. https://cjos.histr.edu.ly/index.php/journal
21. Taye, M. M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers*, *12*(5), 91.
22. Elghaffi, F., Mohammed, O., Dalla, L., Ahmed, A., Agila, A., & EL-Sseid, M. (2026). Hybrid Matrix-Ensemble Framework for Chronic Kidney Disease Diagnosis. *Wadi Alshatti University Journal of Pure and Applied Sciences*, *4*(1), 263-276. https://doi.org/10.63318/waujpasv4i1_28
23. DALLA, L. B. (2020). The Sustainable Efficiency of Modeling a Correspondence Undergraduate Transaction Framework by using Generic Modeling Environment

(GME. Ben Dalla. International Journal of Engineering and Modern Technology E-ISSN 2504-8848 P-ISSN 2695-2149 . Vol 6 No 1 2020 www.iiardpub.org

24. Chahar, R., & Kaur, D. (2020). A systematic review of the machine learning algorithms for the computational analysis in different domains. *International Journal of Advanced Technology and Engineering Exploration*, *7*(71), 147.
25. Chantar, H., Tubishat, M., Essgaer, M., & Mirjalili, S. (2021). Hybrid binary dragonfly algorithm with simulated annealing for feature selection. *SN computer science*, *2*(4), 295.
26. Soliman, A., Shlibak, A., & Zencirci, N. (2026). Wheat Fungal Diseases: A Review. *Wadi Alshatti University Journal of Pure and Applied Sciences*, 191-198.
27. Shlibak, A. A. A., & Dalla, L. O. F. B. (2020). The sustainable research Long while between bee pollen and honey bee diversity in Libya: Literature review. *International Journal of Social Sciences and Management Research*, *7*(1), 2545-5303.
28. Shlibak, A. A., Örgeç, M., & Zencirci, N. (2021). Wheat landraces versus resistance to biotic and abiotic stresses. In *Wheat Landraces* (pp. 193-214). Cham: Springer International Publishing.
29. Soliman, A., Shlibak, A., & ELfaraikh, S. (2026). Response of Two Libyan Wheat Cultivars (Triticum turgidum L. and Triticum aestivum L.) to Cadmium Stress: Growth Parameters, Germination, and Seedling Vigor. *Scientific Journal for Publishing in Health Research and Technology*, 262-272.
30. Soliman, A., Mohamed, N., Almiar, F., & Shlibak, A. (2026). Effect of Fenugreek (Trigonella foenum-graecum L.) Extract on Early Growth Parameters of Phaseolus vulgaris L. *AlQalam Journal of Medical and Applied Sciences*, 416-420.
31. Shlıbak, A., & Zencirci, N. (2021). Wheat: Biotrophic Fungi and Resistance Genes. *Uluslararası Anadolu Ziraat Mühendisliği Bilimleri Dergisi*, *3*(1), 10-20.
32. Zencirci, N., Baloch, F. S., Habyarimana, E., & Chung, G. (Eds.). (2021). *Wheat landraces*. Cham: Springer International Publishing.
33. Shalibak, A. A., & Zencirci, N. (2019). Collection and Preservation of Rare and Endangered Plants (Case Study Endangered Plants in Libya). *International Journal of Agriculture and Earth Science E-ISSN*, 2489-0081.
34. عواطف علي شليبك, هيفاء محمد دوزان, نورية علي العامري, انتصار علي القماطي, & عبدالنبي محمد أبوغنية. (2021). التأثير التضادي لثلاث عزلات محلية وعزلتين تجاريتين من فطر Trichoderma ضد فطر Sclerotinia sclerotiorum..(1)*26* ,*المجلة الليبية للعلوم الزراعية*
35. Hawa Ahmed Alrawayati, Ümit Tokeşer. (2025).Spectral Integral Variation of Graph Theory. Asian Journal of Mathematics and Computer Research.32, Issue, 2. Pages(151-160). https://www.elibrary.ru/item.asp?id=82163806
36. Alrawayati, H., & Tökeşer, Ü. (2021). PARKINSON'S DISEASE DIAGNOSIS BASED ON THE CONVOLUTIONAL NEURAL NETWORK AND PARTICLE SWARM OPTIMIZATION ALGORITHM. *Asian Journal of Mathematics and Computer Research*, *28*(1), 26-37.
37. Joshi, P. K., Prakash, R., & Rai, A. K. (2024, March). A comprehensive review of machine learning application across different domains. In *2024 2nd international conference on disruptive technologies (ICDT)* (pp. 1266-1270). IEEE.
38. Hawa Ahmed Alrawayati, Ümit Tokeşer. (2025).Spectral Integral Variation of Graph Theory. Asian Journal of Mathematics and Computer Research.32, Issue, 2. Pages(151-160). https://www.elibrary.ru/item.asp?id=82163806.
39. Hawa Alrawayati (2020). Development of High Efficiency Optimization Algorithm based on New Topology in Particle Swarm Optimization for Parkinson's Disease. IOSR Journal of Mathematics (IOSR-JM). 8

40. Hawa Alrawayati. (2016). (المعادلة التكاملية ونواة المؤثر) Integral Equation and Kernel Operator.76 – 63 .مجلة الساتل - جامعة مصراته
41. Hawa Alrawayati. (2016). Integral Equation and Kernel Operator. Al-Satel Journal - Misrata University. 63-76
42. Gamal, D., Alfonse, M., M. El-Horbaty, E. S., & M. Salem, A. B. (2018). Analysis of machine learning algorithms for opinion mining in different domains. *Machine Learning and Knowledge Extraction*, *1*(1), 224-234.
43. Dalla, L. O. B., Karal, Ö., Degirmenci, A., EL-sseid, M. A. M., Essgaer, M., & Alsharif, A. (2025). A comprehensive literature review (LR) on optimization algorithms of sewage water treatment processes. Comprehensive Science Journal, 10(37), 3204-3220.
44. Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., & Vaughan, J. W. (2010). A theory of learning from different domains. *Machine learning*, *79*(1), 151-175.
45. Dalla, L. O. F. B., & AHMAD, T. M. A. (2024). Integration of Artificial Bee Colony Algorithm with Deep Learning for Predictive Maintenance in Industrial IoT.
46. Jhaveri, R. H., Revathi, A., Ramana, K., Raut, R., & Dhanaraj, R. K. (2022). A review on machine learning strategies for real-world engineering applications. *Mobile Information Systems*, *2022*(1), 1833507.
47. Ben Dalla, L., Medeni, T. M., Agila, A. A., & Medeni, İ. M. (2024). Architectural Synergy: Investigating the Role of Artificial Neural Networks in Enabling Deep Learning. *The International Journal of Engineering & Information Technology (IJEIT)*, *12*(1), 96-103.
48. Ramadhan, H. R. S., Osman, M. O. M., Dalla, L. O. F. B., Rashid, T. A., Albaraesi, M. J. S., El-Sseid, M. A. M., & Alnnale, T. (2025). A New Approach of the Machine Learning Framework Integrating Policy Design to Predict Renewable Electricity Penetration in Resource-Constrained Settings. Comprehensive Science Journal, 10(Supplement 38), 2929-2950
49. Gong, Z., Zhong, P., & Hu, W. (2019). Diversity in machine learning. *Ieee Access*, *7*, 64323-64350.
50. Albaraesi, M. J. S., Ali, M. A. M. A., Dalla, L. O. B., EL-sseid, M. A. M., Medeni, T. D., Medeni, İ. T., & Alnnale, T. (2025). Random construction in the city of Al-Bayda during the period 2011-2022 and its irregular expansion and its impact on the urban landscape. Comprehensive Science Journal, 10 (Supplement 38), 2590-2614.
51. Ahmed, S. F., Alam, M. S. B., Kabir, M., Afrin, S., Rafa, S. J., Mehjabin, A., & Gandomi, A. H. (2025). Unveiling the frontiers of deep learning: innovations shaping diverse domains. *Applied Intelligence*, *55*(7), 573.
52. A-abdullatef, M. M., Osman, M. O. M., Elghaffi, F. S. A., Dalla, L. O. B., Agila, A. A. A., & Alsharif, A. (2025). LATENT: Low-Latency Anomaly Tracking in National Electricity Time-Series Using Hybrid LSTM-Regression Architectures–A Case Study of Bangladesh's PGCB Grid. Comprehensive Science Journal, 9(36), 1891-1912.
53. Jetlawei, S. S., Dalla, L. O. B., Karal, Ö., Degirmenci, A., El-Sseid, M. A. M., Essgaer, M., & Alsharif, A. (2025). Temporal Intelligence and Algorithmic Equity: A Multi-Phase Framework for Predictive Student Success in Higher Education. Comprehensive Science Journal, 9(36), 1574-1595.https://doi.org/10.65405/f0xx5p02
54. Dalla, L. O. F. B. (2020). Lean Software Development Practices and Principles in Terms of Observations and Evolution Methods to increase work environment productivity. *International Journal of Engineering and Modern Technology*, *6*(1), 23-45.
55. Boutaba, R., Salahuddin, M. A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., & Caicedo, O. M. (2018). A comprehensive survey on machine learning for

networking: evolution, applications and research opportunities. *Journal of Internet Services and Applications*, *9*(1), 1-99.
56. Dalla, L. O. F. B., & AHMAD, T. M. A. (2024). IMPROVE DYNAMIC DELIVERY SERVICES USING ANT COLONY OPTIMIZATION ALGORITHM IN THE MODERN CITY BY USING PYTHON RAY FRAMEWORK.
57. Karal, Ö. (2020). Performance comparison of different kernel functions in SVM for different k value in k-fold cross-validation. In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-5). IEEE. https://doi.org/10.1109/ASYU50717.2020.9259880
58. Ben Dalla, L., Medeni, T. M., Zbeida, S. Z., & Medeni, İ. M. (2024). Unveiling the Evolutionary Journey based on Tracing the Historical Relationship between Artificial Neural Networks and Deep Learning. *The International Journal of Engineering & Information Technology (IJEIT)*, *12*(1), 104-110.
59. Sutton, C., Boley, M., Ghiringhelli, L. M., Rupp, M., Vreeken, J., & Scheffler, M. (2020). Identifying domains of applicability of machine learning models for materials science. *Nature communications*, *11*(1), 4428.
60. Hawa Alrawayati. (2013). (المؤثرات الخطية المحدودة على فضاء هلبرت) • Finite linear operators on Hilbert spaces.193-184.مجلة جامعة الزيتونة
61. Taye, M. M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers*, *12*(5), 91.
62. Chantar, H., Tubishat, M., Essgaer, M., & Mirjalili, S. (2021). Hybrid binary dragonfly algorithm with simulated annealing for feature selection. *SN computer science*, *2*(4), 295.
63. Siraj, F., & Abdoulha, M. A. (2009, May). Uncovering hidden information within university's student enrollment data using data mining. In *2009 Third Asia International Conference on Modelling & Simulation* (pp. 413-418). IEEE.
64. Siraj, F., & Abdoulha, M. A. (2007). Mining enrolment data using predictive and descriptive approaches. *Knowledge-Oriented Applications in Data Mining*, 53-72.
65. Alssager, M., & Othman, Z. A. (2016). Taguchi-based parameter setting of cuckoo search algorithm for capacitated vehicle routing problem. In *Advances in Machine Learning and Signal Processing: Proceedings of MALSIP 2015* (pp. 71-79). Cham: Springer International Publishing.
66. Alssager, M., Othman, Z. A., Ayob, M., Mohemad, R., & Yuliansyah, H. (2020). Hybrid cuckoo search for the capacitated vehicle routing problem. *Symmetry*, *12*(12), 2088.
67. Omar, A., Essgaer, M., & Ahmed, K. M. (2022, July). Using machine learning model to predict Libyan telecom company customer satisfaction. In *2022 International Conference on Engineering & MIS (ICEMIS)* (pp. 1-6). IEEE.
68. Alssager, M., & Othman, Z. A. (2016). Cuckoo search algorithm for capacitated vehicle routing problem. *Journal of Theoretical and Applied Information Technology*, *88*(1), 11.
69. Siraj, F., & Ali, M. (2011). *Mining enrollment data using descriptive and predictive approaches* (pp. 53-72). InTech—Open Access Company.
70. Dalla, L. O. B., Karal, Ö., Degirmenci, A., EL-Sseid, M. A. M., Essgaer, M., & Alsharif, A. (2025). Edge Intelligence for Real-Time Image Recognition: A Lightweight Neural Scheduler Via Using Execution-Time Signatures on Heterogeneous Edge Devices. *Scientific Journal for Publishing in Health Research and Technology*, 74-85.
71. Alssager, M., & Nasir, I. (2021). Evaluation of using Google Classroom as a Tool for Asynchronous E-learning at Sebha University. *Journal of Pure & Applied Sciences*, *20*(1), 44-49.

72. ALSSAGER, M., & OTHMAN, Z. A. (2014). SIMULATED ANNEALING ALGORITHM USING ITERATIVE COMPONENT SCHEDULING APPROACH FOR CHIP SHOOTER MACHINES. *Journal of Theoretical & Applied Information Technology*, *65*(2).
73. Ben Dalla, L., MEDENİ, T., MEDENİ, İ., & ULUBAY, M. (2024). Enhancing Healthcare Efficiency at Almasara Hospital: Distributed Data Analysis and Patient Risk Management. *Economy: strategy and practice*, *19*(4).
74. BEN, D. L., MEDENI, T., MEDENI, I., & ULUBAY, M. (2024). ENHANCING HEALTHCARE EFFICIENCY AT ALMASARA HOSPITAL: DISTRIBUTED DATA ANALYSIS AND PATIENT RISK MANAGEMENT. *ЭКОНОМИКА*, *19*(4), 54-72.
75. Ben Dalla, L., Medeni, T. M., Agila, A. A., & Medeni, İ. M. (2024). Architectural Synergy: Investigating the Role of Artificial Neural Networks in Enabling Deep Learning. *The International Journal of Engineering & Information Technology (IJEIT)*, *12*(1), 96-103.
76. Agila, A. A. A. (2024). Diabetes Prediction Using a Support Vector Machine (SVM) and visualize the results by using the K-means algorithm* Corresponding author:* Llahm Omar Faraj Ben Dalla, Tarik Milod Alarbi Ahmad2 .
77. Dalla, L. O. F. B. (2020). The Influence of hospital management framework by the usage of Electronic healthcare record to avoid risk management (Department of Communicable Diseases at Misurata Teaching Hospital: Case study). EHRM, 20(4), 22–52. https://doi.org/20.51176/1954-9923-2020-4-22-52
78. Dalla, L. O. F. B. (2020). Dorsal Hand Vein (DHV) Verification in Terms of Deep Convolutional Neural Networks with the Linkage of Visualizing Intermediate Layer Activations Detection. International Journal of Engineering and Modern Technology E-ISSN 2504-8848 P-ISSN 2695-2149 Vol 6 No 2 2020 www.iiardpub.org
79. Dalla, L. O. F. B. (2020). Convolutional Neural Network Baseline Model Building for Person Re-Identification. International Journal of Engineering and Modern Technology E-ISSN 2504-8848 P-ISSN 2695-2149 Vol. 6 No. 3 2020 www.iiardpub.org
80. Ghumeid, N. A., & Essgaer, M. (2024, May). Addressing the Libyan Arabic dialect identification: a comparative study of ensemble classification methods. In *2024 IEEE 4th International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)* (pp. 579-584). IEEE.
81. Ben Dalla, L, O, F. (2021).Literature review (LR) on the powerful of Research methodology processes life cycle. In 2021 The Powerful of Research Methodology Processes Life Cycle Conference (TPRMPLCC) (pp. 1-10). IEEE. https://doi.org/10.16543/TPRMPLCC 50717.2020.92876580
82. Yalman, Y., Uyanık, T., Atlı, İ., Tan, A., Bayındır, K. Ç., Karal, Ö., ... & Guerrero, J. M. (2022). Prediction of voltage sag relative location with data-driven algorithms in distribution grid. Energies, 15(18), 6641.Energies 2022, 15(18), 6641; https://doi.org/10.3390/en15186641
83. Arık, D. T., Karal, Ö., & Şahin, A. B. (2020). A Comparative Study of Artificial Neural Networks and Naïve Bayes Techniques for the Classification of Radar Targets. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 9(4), 1779-1788.https://doi.org/10.17798/bitlisfen.676973
84. Uysal, Z., Kalkancı, G., İmren, T., Değirmenci, A., Karal, Ö., & Çankaya, İ. (2016). A Heart Rate Monıtoring Applicatıon Using Wireless Sensor Network System Based on Bluetooth Wıth Matlab GUI. Int. J. Eng. Sci, 6, 2862.International Journal of Engineering Science and Computing, August 2016 , http://ijesc.org/

85. Dulkadir, S. E. Z. G. İ. N., Tecimer, H. U., Parlaktürk, F., Altındal, Ş., & Karal, Ö. M. E. R. (2020). The effect of radiation on the forward and reverse bias current–voltage (I–V) characteristics of Au/(Bi4Ti3O12/SiO2)/n-Si (MFIS) structures. Journal of Materials Science: Materials in Electronics, 31(15), 12514-12521.https://doi.org/10.1007/s10854-020-03801-0
86. Muttaqi, M., Degirmenci, A., & Karal, O. (2022, September). US accent recognition using machine learning methods. In 2022 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-6). IEEE.https://doi.org/10.1109/ASYU56188.2022.9925265
87. Ben Dalla, L, O, F. (2021). Literature review (LR) on the dominant of Research methodology. . In (2020), Literature review (LR) on the dominant of Research methodology Conference (LRDRMC) (pp. 1-14). IEEE. https://doi.org/10.6754/ LRDRMC 56412.2020.45987623
88. Safour, H., Essgaer, M., & Alshareef, A. (2024, December). Unraveling Academic Failure: Examining the Influence of Course Correlations on Student Performance Through Association Rules Algorithms. In *2024 International Conference on Computer and Applications (ICCA)* (pp. 1-5). IEEE.
89. Essgaer, M. (2024, November). Analyze data from scholarly articles on Google Scholar for two colleges within Sebha University using an exploratory data analysis approach. In *Sebha University Conference Proceedings* (pp. 511-516).