Optimizing Regularized Logistic Regression Pipelines on Cleaned Biomarker Datasets: A Leakage-Proof Benchmarking Protocol for Diagnostic and Prognostic Modeling

Stanley Nwakamma [1*], Serif Oyindamola Oyesiji [2], Harouna Wendpanga Yann Christian Sankara [3], Toussida Fatah Tanguy Minoungou [4]

[1] Independent Researcher, USA
[2] Harrisburg University of Science and Technology, Harrisburg, Pennsylvania, USA
[3] Independent Researcher, Dallas, TX, USA
[4] Independent Researcher, Dallas, TX, USA

* Corresponding Author: **Stanley Nwakamma**

Article Info

P-ISSN: 3051-3383
E-ISSN: 3051-3391
Volume: 07
Issue: 02
Received: 06-07-2026
Accepted: 04-08-2026
Published: 02-09-2026
Page No: 64-72

Abstract

Logistic regression remains the workhorse of biomarker-based diagnostic and prognostic modeling, and for principled reasons: it yields calibrated probabilities when fitted and validated correctly, its coefficients support biological and clinical interpretation, and its statistical behavior under limited events is thoroughly characterized. Yet the gap between the method's textbook properties and its published practice is wide and documented, with prediction research reviews finding pervasive deficiencies in validation, calibration assessment, sample size justification, and handling of missing data, deficiencies that persist even in high-stakes settings. The availability of cleaned, curated biomarker datasets removes the data-quality alibi and exposes the remaining variance for what it is: modeling protocol. This paper develops a research concept for a systematic optimization and benchmarking protocol for logistic regression pipelines on cleaned biomarker datasets. The protocol fixes a leakage-proof evaluation harness, nested resampling with all preprocessing inside the resampling loop and grouping that respects study and batch structure, and then treats the pipeline's design space as the experimental object: regularization family and strength, feature selection strategy including stability-based selection, treatment of continuous predictors, class imbalance handling, missing data strategy, rare-event corrections, and post-hoc calibration methods, each varied under prespecified contrasts. Performance is reported as a profile, discrimination with uncertainty, calibration curves and indices, and decision-analytic net benefit across clinically relevant thresholds, never as a single headline number, with optimism-corrected internal validation and temporal or cohort-external validation where data permit. Prespecified hypotheses target the conditions under which elastic net dominates unpenalized fitting, when stability selection changes selected panels, and how far calibration methods repair miscalibration without refitting. Reporting follows prediction-model standards throughout.

DOI: https://doi.org/10.54660/IJAIET.2026.7.2.64-72

Keywords: Logistic regression, Biomarker-based diagnostics, Prognostic modeling, Predictive modeling, Model calibration, Regularization

1. Introduction

The statistical case for logistic regression in biomarker modeling has never weakened. The model's foundations are classical (Cox, 1958; Hosmer *et al*., 2013) [39, 55], its regularized extensions address the correlated, wide feature sets that biomarker panels produce (Hoerl & Kennard, 1970; Tibshirani, 1996; Zou & Hastie, 2005) [54, 107, 118], efficient solvers make the whole regularization path routine (Friedman *et al*., 2010; Pedregosa *et al*., 2011) [48, 89], and the surrounding methodology, events-per-variable behavior, modern sample size criteria, rare-event bias correction, and validation strategy, is mature and prescriptive

(Peduzzi *et al*., 1996; van Smeden *et al*., 2019; Riley *et al*., 2020; Firth, 1993; Harrell, 2015; Steyerberg, 2019) [90, 112, 96, 47, 51, 103]. When more flexible learners are compared fairly on clinical tabular problems, regularized regression remains a strong and frequently undominated baseline (Hastie *et al*., 2009) [52]. Applied biomarker research continues to instantiate the setting directly, as in recent population-scale studies associating composite metabolic indices with estimated cardiovascular risk through regression-based analysis (Amadi *et al*., 2025) [15]. The same index literature extends that pattern across outcomes and cohorts, relating the triglyceride-glucose index to reduced kidney function along the cardio-renal continuum (Amadi *et al*., 2026a) [13], pairing it with echocardiographic measurement for stratification in Sub-Saharan African populations (Okwah, 2022) [79], and situating it against opportunistic screening estimates of hypertension burden (Amadi *et al*., 2026b) [14]. Adjacent work supplies the predictor classes such pipelines ingest: circulating inflammatory biomarkers and gene-expression readouts (Okwah *et al*., 2026) [80], nutritional exposures paired with DNA markers in diabetic cohorts (Ike *et al*., 2025) [56], and the modifiable-risk context that determines case mix and confounding structure (Abah *et al*., 2025a; Abah *et al*., 2025b; David *et al*., 2025) [1, 2, 40]. These are precisely the datasets on which the modeling choices this paper isolates are made.

The practice, by systematic account, does not match the method. Reviews of clinical prediction research document weak validation design, absent calibration assessment, unjustified sample sizes, and ad hoc missing data handling as the norm rather than the exception (Bouwmeester *et al*., 2012) [34], a pattern that reappeared at scale in the pandemic-era model literature, where a living systematic review appraised hundreds of models and found nearly all at high risk of bias (Wynants *et al*., 2020) [116]. The biomarker field carries additional documented pathologies of its own: selection bias and overfitting in discovery (Ransohoff, 2004; Ransohoff, 2005) [94, 95], inflated early effect sizes (Ioannidis & Panagiotou, 2011; Ioannidis, 2005) [61, 60], and evaluation by measures that overstate clinical usefulness (Pepe *et al*., 2004) [92], all against staged evidentiary frameworks that specify what validation should look like (Pepe *et al*., 2001) [91]. The same tension surfaces in evidence synthesis, where phenotype-stratified pooling is adopted precisely because averaged effects conceal the subgroup structure a prediction model is supposed to represent (Yusuf *et al*., 2026) [117], and in outcome-level reviews of complex interventions whose heterogeneity mirrors the case-mix variation that transportability analysis must handle (Obi *et al*., 2025) [71]. Methodological work has likewise shown that the most damaging errors are procedural: feature selection performed outside cross-validation biases error estimates catastrophically (Ambroise & McLachlan, 2002; Simon *et al*., 2003) [16, 101], model selection inside a single resampling loop biases them again (Varma & Simon, 2006) [113], and leakage in its many forms is now recognized as a general reproducibility threat in learning-based science (Kapoor & Narayanan, 2023; Whalen *et al*., 2022) [63, 115].

Cleaned data sharpens the question. As curated biomarker resources and standardized preprocessing become available, the residual variance in reported performance is increasingly attributable to pipeline choices, regularization, selection, imputation, calibration, and validation design, rather than to raw data quality, and those choices are rarely varied systematically or reported completely (Collins *et al*., 2015; Moons *et al*., 2015) [38, 68]. Curation also has an upstream that the term conceals: the cleanliness of a biomarker table is manufactured by laboratory design, pre-analytical control, and diagnostic workflow standardization long before an analyst sees it (Ogbete *et al*., 2018; Ayinde *et al*., 2026; Asiedu, 2023) [73, 32, 24], and by the governance and interoperability architecture that determines whether a variable means the same thing across contributing sources (Ilodigwe & Adesemoye, 2021; Aneke *et al*., 2023) [57, 21]. This paper develops the research concept for doing exactly that: a benchmarking protocol that treats the logistic regression pipeline as the experimental object on cleaned biomarker datasets, with a leakage-proof harness, prespecified contrasts, and profile-based reporting. Section 2 reviews background; Section 3 defines the problem; Sections 4 through 7 specify the protocol, the harness, the design space, and the calibration and decision-analytic layer; Section 8 sets out the evaluation and benchmarking plan; Sections 9 and 10 treat anticipated contributions and limitations; Section 11 concludes.

2. Background and Related Work

Prediction-model methodology supplies the protocol's spine. Modeling-strategy texts consolidate the discipline, prespecification, appropriate handling of continuous predictors, shrinkage, and internal validation by resampling (Harrell, 2015; Steyerberg, 2019) [51, 103], with fractional polynomial and related approaches giving continuous predictors principled functional forms rather than data-driven categorization (Royston & Altman, 1994; Sauerbrei *et al*., 2007) [97, 100]. Validation methodology distinguishes apparent, internally validated, and externally validated performance, quantifies optimism via bootstrap and cross-validation procedures (Efron & Tibshirani, 1993; Steyerberg *et al*., 2001) [43, 104], and warns precisely where naive designs fail: selection outside the loop (Ambroise & McLachlan, 2002; Simon *et al*., 2003) [16, 101] and tuning without nesting (Varma & Simon, 2006) [113]. Sample size methodology has moved from events-per-variable heuristics (Peduzzi *et al*., 1996) [90] to criteria targeting shrinkage and calibration stability directly (van Smeden *et al*., 2019; Riley *et al*., 2020) [112, 96], with penalized likelihood correcting small-sample and separation pathologies (Firth, 1993) [47]. Reporting standards close the loop, specifying what a trustworthy model paper discloses (Collins *et al*., 2015; Moons *et al*., 2015) [38, 68].

Performance assessment methodology defines the protocol's outcome space. Discrimination is measured by the area under the receiver operating characteristic curve with appropriate uncertainty and paired comparison machinery (Hanley & McNeil, 1982; DeLong *et al*., 1988) [50, 42], but discrimination alone misleads under imbalance and near decision thresholds, motivating precision-recall analysis (Saito & Rehmsmeier, 2015; Davis & Goadrich, 2006) [99, 41] and balanced summaries (Chicco & Jurman, 2020) [37]. Calibration is the clinically decisive property and the habitually neglected one (Van Calster *et al*., 2019) [111], assessed by calibration curves, slope and intercept, and proper scores (Brier, 1950; Steyerberg *et al*., 2010) [35, 105], and repaired post hoc, where refitting is impossible, by parametric and non-parametric mapping methods whose behavior is characterized in the calibration literature (Platt, 1999; Niculescu-Mizil & Caruana, 2005; Guo *et al*., 2017) [93, 69, 49]. Decision-analytic evaluation translates both into clinical terms, net benefit

across threshold probabilities (Vickers & Elkin, 2006) [114]. The auxiliary design space carries its own methodology: imbalance handling by

4. Protocol Overview

The protocol has four layers executed in fixed order. The readiness layer verifies that the cleaned dataset supports the intended experiment: outcome definition and event counts against modern sample size criteria [96, 112], missingness mapped by variable and mechanism plausibility [65], structural metadata sufficient for grouped partitioning, and a data dictionary frozen before any modeling, with variable definitions, units, assay provenance, and permissible ranges recorded under the data-integrity discipline that regulated life-science operations already impose [21, 57, 12]. The harness layer instantiates the leakage-proof evaluation machinery of Section 5 and registers the prespecified design space and hypotheses. The experimental layer executes the design space under the harness, producing per-configuration performance profiles with uncertainty. The reporting layer renders results under prediction-model standards, full specification of every configuration, optimism-corrected estimates, calibration graphics, and decision curves, with code and configuration versioned for exact reproduction [38, 68, 89], and with access to identifiable inputs handled under the privacy-preserving architectures such data demands [27, 8]. The protocol's products are twofold: for the dataset at hand, a defensible model choice with its evidence; for the field, an accumulating benchmark of which pipeline choices matter under which named conditions.

5. The Evaluation Harness

The harness is nested, grouped resampling with the pipeline inside. The outer loop estimates generalization performance; the inner loop performs all tuning, penalty strength and mixing, selection thresholds, imputation hyperparameters, so that no outer estimate is contaminated by its own selection, the nesting whose omission biases error estimates [113]. Grouping assigns whole studies, sites, or batches to folds, so outer estimates approximate transport across structure rather than interpolation within it [64, 62], the structural unit in practice being the laboratory or collection network whose design, catchment, and throughput differ systematically from its peers [18, 75]; where only one structural unit exists, temporal splitting substitutes as the honest alternative. Every data-dependent operation, imputation models, transformations, screening, selection, calibration mapping, is fitted on training partitions only and applied forward, the discipline whose violation in feature selection was shown to fabricate accuracy decades ago and still recurs [16, 101, 63].

Uncertainty and correction are built in rather than appended. Optimism-corrected estimates via bootstrap validation accompany cross-validated ones, following comparative evidence on internal validation efficiency [104, 43, 51], paired comparisons between configurations use resampled differences with appropriate tests for correlated performance measures [42], and multiple imputation is executed per training partition with pooled downstream estimation under the standard combining rules [98, 110, 102]. The harness emits, for every configuration, the full profile, discrimination with intervals, precision-recall behavior at prespecified operating points, calibration curve with slope, intercept, and proper score, and net benefit across the registered threshold range [50, 99, 105, 35, 114], plus stability outputs where selection is in play [67].

6. The Pipeline Design Space

The design space is prespecified as factors with registered levels. Estimation: unpenalized maximum likelihood as the reference, ridge, lasso, and elastic net across the mixing range [54, 107, 118, 48], and Firth-penalized likelihood where event counts or separation indicate it [47]. Selection: none, embedded selection via the lasso path, univariate screening with false discovery control as the deliberately included bad-practice comparator whose cost the benchmark quantifies [33, 106], and stability selection over subsamples with selection-frequency thresholds [67]. Continuous predictors: linear, fractional polynomial, and spline treatments, with categorization excluded on the methodology's settled advice except as a costed comparator [97, 100, 51]. Imbalance: unweighted likelihood, class weighting, and synthetic minority oversampling as a sensitivity arm given its known calibration distortions [53, 36], an arm that matters most in the rare-outcome regimes that dominate applied risk stratification and detection work [81, 26]. Missing data: complete-case as the costed comparator, single imputation, and chained-equations multiple imputation [110, 65]. Calibration: none beyond the fitted model, parametric mapping, and non-parametric mapping, each fitted within the harness [93, 69, 49].

Crossing all factors fully is neither feasible nor informative; the protocol registers a fractional design anchored on hypotheses. Contrasts of primary interest, stated as prespecified hypotheses in the registration: elastic net dominates unpenalized fitting on the profile whenever events per variable fall below modern adequacy thresholds, with the margin growing as dimensionality rises [112, 96, 118]; stability selection changes selected panels materially relative to single-pass lasso while costing little discrimination, and its selection frequencies predict cross-fold panel reproducibility [67]; univariate screening's optimism, the gap between its inner and outer estimates, exceeds that of embedded selection, reproducing the classic selection-bias result under modern conditions [16, 101]; post-hoc calibration repairs slope and intercept miscalibration from imbalance interventions but not the local distortions synthetic oversampling induces [36, 69, 111]; and multiple imputation's advantage over complete-case analysis on the profile grows with missingness rate and structure, per the imputation literature's predictions [102, 98].

7. Calibration and Decision-Analytic Layer

The protocol elevates calibration and clinical utility from afterthoughts to registered endpoints. Calibration assessment is graphical and quantitative per configuration, curves with confidence bands, slope and intercept against the ideal, and proper scoring [105, 35], on the position that a biomarker model's probabilities are its clinical interface and their trustworthiness is not implied by discrimination [111]. Post-hoc calibration methods are treated as pipeline components inside the harness, so their apparent repairs are validated out-of-fold like everything else [93, 69]. Decision-analytic evaluation registers the clinically relevant threshold range with justification, rule-in and rule-out uses named, and reports net benefit curves per configuration against treat-all and treat-none references [114], converting profile differences

into the currency that adoption decisions actually spend, and exposing the configurations whose discrimination advantages evaporate at the thresholds that matter [92]. Thresholds are registered with the deployment context named, since a probability consumed by an early-detection service, an operational triage dashboard, or a clinician-facing decision support tool carries different rule-in and rule-out costs and different tolerance for false positives [7, 5, 59, 58]. Where the dataset supports it, temporal or leave-one-study-out external validation is registered as the confirmatory tier above internal estimates, in the staged-evidence spirit of the biomarker frameworks [91, 103].

8. Evaluation and Benchmarking Plan

The plan proceeds in three phases. Phase one is harness verification: the machinery is validated on semi-synthetic data, real biomarker covariance with simulated outcomes of known coefficient structure, confirming that the harness recovers known signal, that its outer estimates are unbiased under the null, and that the registered leakage traps, selection outside the loop, unpooled imputation, ungrouped folds, are detected when deliberately introduced, the positive controls that give later results their credibility [63, 113].

Phase two executes the registered design space across a curated collection of cleaned biomarker datasets spanning the named conditions, events per variable, dimensionality, imbalance, missingness, and structural depth, reporting per-dataset profiles and cross-dataset patterns for each prespecified contrast, with disconfirmations reported as findings. The collection is assembled to vary provenance deliberately, because datasets drawn from different laboratory networks and diagnostic programs differ in measurement protocol and referral pattern as much as in case mix [19, 76, 25]. Phase three consolidates the benchmark: configuration-condition performance maps, the empirical answer to which choices matter when; stability atlases for selection methods across datasets [67]; and a reporting package per dataset that satisfies the prediction-model checklist end to end [38, 68], released with code, configurations, and seeds for exact reproduction [89], with the benchmark functioning as a release gate rather than a leaderboard [70, 3]. Acceptance for the protocol itself is defined operationally: results reproducible from the release, hypotheses adjudicated with registered analyses, and no reported estimate whose provenance the harness cannot certify leakage-free.

9. Anticipated Contributions

Three contributions are anticipated. For practice, the configuration-condition maps give biomarker modelers evidence-based defaults, when penalization is obligatory, when stability selection earns its cost, which imbalance interventions to avoid for probability-consuming uses, replacing the folklore that reviews show currently governs these choices [34, 116]. For methodology, the protocol operationalizes the literature's scattered prescriptions, nesting, grouping, imputation-within-validation, profile reporting, into one executable harness whose positive controls make violations visible, a transferable instrument for any tabular clinical modeling, not only biomarkers [51, 103, 115]. Because the harness is agnostic to the substantive domain, the same instrument answers the analogous question wherever tabular classification under imbalance and drift is deployed at scale, from anomaly detection and fraud screening to demand and segmentation modeling [28, 108, 66, 31, 78]. For the evidence base, the accumulating benchmark across cleaned datasets creates the comparative record that individual model papers cannot: how much performance variation is pipeline rather than biology, and where the honest ceilings lie for regression-based biomarker prediction under stated conditions [61, 94].

10. Limitations

The concept's boundaries are explicit. Scope: the protocol characterizes logistic regression pipelines; flexible learners enter only as registered challengers on identical folds, and claims about their relative merit are limited to the datasets and conditions studied [52]. Data dependence: cleaned does not mean representative, and curation itself can distort case mix and missingness in ways that qualify transportability; the structural grouping mitigates within-collection optimism but cannot manufacture external validity, which only the registered external tiers provide [91, 95].

Where a collection is dominated by a small number of laboratory networks, its case mix inherits their catchments and referral patterns [18, 72], and equity in who is measured at all remains a property of the diagnostic system rather than of the model [19, 74]. Design coverage: the fractional design answers its registered contrasts, not all interactions, and configuration-condition maps interpolate at their peril outside studied ranges.

Computational cost is real, nested grouped resampling with multiple imputation multiplies fits, and the released harness's efficiency work is part of the deliverable rather than an afterthought [48].

Two further boundaries are governance rather than statistics. The protocol assumes lawful access to individual-level data under prevailing privacy regimes, and the harness's appetite for structural metadata can conflict with the de-identification and minimization constraints those regimes impose [27, 8, 45, 85, 86]. And the protocol governs estimation and evaluation, not clinical adoption: net benefit curves inform but do not constitute impact evidence, the staged frameworks' downstream tiers, impact studies in deployment, remain beyond its remit [114, 103], and the organizational questions that decide whether a validated model is used at all, institutional readiness, oversight and accountability structures, and the regulatory route by which a model-derived tool reaches practice, sit outside the harness entirely [7, 9, 10, 22, 46, 20].

11. Conclusion

This paper has developed a research concept for the systematic optimization of logistic regression pipelines on cleaned biomarker datasets: a leakage-proof, structure-respecting evaluation harness with positive controls; a registered design space spanning regularization, selection, continuous-predictor treatment, imbalance handling, missing data strategy, and calibration; profile-based outcomes joining discrimination, calibration, and decision-analytic net benefit; and prespecified hypotheses that convert the field's methodological prescriptions into testable contrasts on real data. The concept's premise is that with curated data in hand, the pipeline is the experiment, and that the biomarker literature's documented gap between method and practice closes not through exhortation but through instruments, harnesses that make correct evaluation the path of least

resistance and benchmarks that make the cost of bad practice a measured quantity. Executed, the protocol yields defensible models per dataset, transferable machinery for the field, and an evidence base on which the oldest tool in clinical prediction can be used at the standard its own methodology has long specified.

References

1. Abah MA, Okwah MN, Oladosu MA, Yohanna NR, Blessing CT. Advances in the management of hypertension: A review of current guidelines, pharmacological therapies, and lifestyle interventions. Drug Pharm Sci Arch. 2025;5(4):35-41. doi:10.47587/DAP.2025.5401.
2. Abah MA, Okwah MN, Oladosu MA, Yohanna NR, Blessing CT, Vanessa OI. The impact of lifestyle modifications on cardiovascular health: A review of diet, exercise, and stress reduction intervention. Drug Pharm Sci Arch. 2025;5(3):26-34. doi:10.47587/DAP.2025.5301.
3. Adelanwa A, Basnet A, Anene UN. Data driven digital transformation models for lifecycle performance management in infrastructure delivery. Int J Adv Multidiscip Res Stud. 2023;3(6):2646-2662. doi:10.62225/2583049X.2023.3.6.5968.
4. Adelanwa A, Basnet A, Anene UN. Predictive analytics models for financial risk detection and fraud prevention in public systems. Int J Adv Multidiscip Res Stud. 2023;3(6). doi:10.62225/2583049X.2023.3.6.5969.
5. Adelanwa A, Basnet A, Anene UN. Advanced AI based decision support systems for healthcare operations and resource planning. Gyanshauryam Int Sci Refereed Res J. 2024;7(1):244-276. doi:10.32628/GISRRJ.
6. Afrihyia E, Akinse SG, Ojukwu PU. Development and implementation of an AI-driven early detection system for non-communicable diseases. Int J Adv Multidiscip Res Stud. 2023;3(6):2680-2691.
7. Afrihyia E, Akinse SG, Ojukwu PU. Comparative governance of AI-driven healthcare management: Executive oversight, regulatory structures, and accountability in the United States and developing countries. Int J Adv Multidiscip Res Stud. 2024;4(6):3087-3102. doi:10.62225/2583049X.2024.4.6.5920.
8. Afrihyia E, Ojukwu PU, Akinse SG. Privacy-preserving health data governance models: A comparative review of blockchain and cryptographic strategies in U.S. and developing healthcare systems. Int J Adv Multidiscip Res Stud. 2024;4(6):3071-3086. doi:10.62225/2583049X.2024.4.6.5919.
9. Afrihyia E, Akinse SG, Ojukwu PU. Organizational readiness for generative AI integration in healthcare operations: Comparative management capabilities between the U.S. and low- and middle-income countries. Iconic Res Eng J. 2025;8(10):1673-1697. doi:10.64388/IREV8I10-1714670.
10. Akintola AS, Fawehinmi YO, Aiyenitaju O, Chinemerem B, Emmanuella O. Digital transformation in telehealth: A systematic review of user trust, privacy, and regulatory governance in AI-powered remote monitoring systems. J Sci Res Rep. 2025;31(11):163-182. doi:10.9734/jsrr/2025/v31i113658.
11. Aliliele C, Mbonu IS, Iwuanyanwu U. Advances in predictive analytics models for student retention and institutional risk management systems. Int J Adv Multidiscip Res Stud. 2023;3(6):2692-2711. doi:10.62225/2583049X.2023.3.6.5990.
12. Aliliele C, Mbonu IS, Iwuanyanwu U. Advances in HIPAA compliant data architecture and secure analytics frameworks for community healthcare organizations. Shodhshauryam Int Sci Refereed Res J. 2024;7(2):277-324. doi:10.32628/SHISRRJ2472163.
13. Amadi CE, Akanbi PO, Okwah MN, Mbakwem AC, Ajuluchukwu JN. Triglyceride-glucose index and reduced kidney function in hypertensive Nigerian adults: Independent association within the cardio-renal continuum. BMC Cardiovasc Disord. 2026. Advance online publication. doi:10.1186/s12872-026-05985-5.
14. Amadi CE, Ale OK, Okorafor UC, Okwah MN, Akanbi PO, Mbakwem AC, *et al*. Unmasking the contemporary burden of hypertension in Lagos, Nigeria: Insights from an opportunistic screening. BMC Public Health. 2026. Advance online publication. doi:10.1186/s12889-026-26310-x.
15. Amadi CE, Okorafor UC, Okorafor CI, Okwah MN, Akanbi PO, Okam OV, *et al*. Association between triglyceride-glucose index and estimated 10-year risk of cardiovascular disease in Nigerian non-diabetic hypertensives: A cross-sectional study. PLOS Glob Public Health. 2025;5(7):e0004760. doi:10.1371/journal.pgph.0004760.
16. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A. 2002;99(10):6562-6566.
17. Aminu-Ibrahim AY, Ogbete JC. Healthcare infrastructure as a public health intervention using evidence from large laboratory networks. Shodhshauryam Int Sci Refereed Res J. 2023;6(1):256-286. doi:10.32628/SHISRRJ23678.
18. Aminu-Ibrahim AY, Ogbete JC, Ambali KB. Infrastructure driven expansion of diagnostic access across underserved and rural healthcare regions. Int J Multidiscip Res Growth Eval. 2020;1(5):691-706. doi:10.54660/IJMRGE.2020.1.5.691-706.
19. Aminu-Ibrahim AY, Ogbete JC, Ambali KB. Design and development of high-performance medical laboratories supporting pandemic preparedness and response. Shodhshauryam Int Sci Refereed Res J. 2022;5(3):302-335.
20. Aneke OB, Adesemoye AC. A conceptual framework for risk-based decision-making in good manufacturing practice environments. Iconic Res Eng J. 2020;4(6):325-341.
21. Aneke OB, Sydney E, Adesemoye AC. Data integrity and governance in regulated life sciences operations: A review of failures, controls, and remediation strategies. Int J Sci Res Comput Sci Eng Inf Technol. 2023;9(10):538-568.
22. Anioke SC, Atima ME. Public health governance models using process optimization and performance metrics for regulatory oversight. Int J Adv Multidiscip Res Stud. 2023;3(6):2534-2548. doi:10.62225/2583049X.2023.3.6.5491.
23. Anioke SC, Atima ME. Public health informatics

frameworks for protecting vulnerable populations through data-driven policy enforcement. Int J Adv Multidiscip Res Stud. 2023;3(6):2564-2579.
24. Asiedu CS. Diagnostic integration in infectious disease management: A review of microbiology, virology, and serology workflows. Int J Med Eval Phys Rep. 2023;7(4):173-195. doi:10.56201/ijmepr.v7.no4.2023.pg173.195.
25. Asiedu CS. Droplet digital PCR for HIV viral reservoir quantification: A review of diagnostic accuracy and clinical utility. Int J Med Eval Phys Rep. 2024;8(6):253-273. doi:10.56201/ijmepr.v8.no6.2024.pg253.273.
26. Atakpa MI, Abetoh NF. A systematic review of machine learning advances in financial fraud detection for banking systems. Int J Sci Res Comput Sci Eng Inf Technol. 2022;8(2):771-801. doi:10.32628/CSEIT23906220.
27. Atakpa MI, Abolaji TO. A privacy-preserving data architecture model for regulated industry analytics under GDPR and HIPAA compliance. Int J Sci Res Comput Sci Eng Inf Technol. 2022;8(3):781-808. doi:10.32628/CSEIT22558.
28. Atakpa MI, Fobellah AN. Anomaly detection in financial time-series data: A conceptual model for healthcare and banking applications. Int J Sci Res Comput Sci Eng Inf Technol. 2023;9(2):967-997. doi:10.32628/CSEIT2342441.
29. Atakpa MI, Abolaji TO, Akeju B. A review of natural language processing for social media-based public health surveillance and analytics. Int J Sci Res Comput Sci Eng Inf Technol. 2023;9(6):1030-1060. doi:10.32628/CSEIT23906783.
30. Atakpa MI, Abolaji TO, Abetoh NF. Statistical time-series models for long-range public health trend forecasting: A review and conceptual framework. Int J Sci Res Comput Sci Eng Inf Technol. 2024;10(6):2748-2780. doi:10.32628/CSEIT2410792.
31. Atima ME, Sanni JO, Attah A. Predictive audience segmentation models resolving targeting inefficiencies in regulated professional service enterprises. Shodhshauryam Int Sci Refereed Res J. 2022;5(1):271-303. doi:10.32628/SHISRRJ247134.
32. Ayinde M, Ohunyon G, Ojukwu PU. Optimization of pre-analytical and analytical processes in CLIA-certified clinical laboratories: Recent innovations and future directions. Int J Adv Multidiscip Res Stud. 2026;6(2):585-600. doi:10.62225/2583049X.2026.6.2.6006.
33. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B. 1995;57(1):289-300.
34. Bouwmeester W, Zuithoff NPA, Mallett S, *et al*. Reporting and methods in clinical prediction research: A systematic review. PLoS Med. 2012;9(5):e1001221.
35. Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev. 1950;78(1):1-3.
36. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321-357.
37. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21:6.
38. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. BMJ. 2015;350:g7594.
39. Cox DR. The regression analysis of binary sequences. J R Stat Soc Series B. 1958;20(2):215-242.
40. David FT, Abah MA, Oladosu MA, Agida OD, Clive NO, Adebukola AM, *et al*. The impact of plant-based diets on cardiovascular health: A comprehensive review. Plant Sci Arch. 2025;10(4):51-59. doi:10.51470/PSA.2025.10.4.51.
41. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: Proceedings of the International Conference on Machine Learning. 2006. p.233-240.
42. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics. 1988;44(3):837-845.
43. Efron B, Tibshirani RJ. An introduction to the bootstrap. Chapman & Hall; 1993.
44. Ejofodomi OA, Gideon EN, Oladipo GO, Oshomah ER. Automated detection of architectural distortion in mammograms using template matching. Int J Biomed Sci Eng. 2014;2(1):1-6. doi:10.11648/j.ijbse.20140201.11.
45. Ekechi NV, Ozowara DE, Anunagba CO. Conceptual framework for AI governance, data privacy compliance, and financial sustainability in digital health. Comput Sci IT Res J. 2026;7(4):275-299.
46. Eze FI, Akinleye OK, Anene UN. Development of a risk-based accelerated approval decision model balancing speed, safety, and evidence quality in pharmaceutical regulatory science. Res J Pure Sci Technol. 2024;7(6):120-161. doi:10.56201/rjpst.v7.no6.2024.pg120.161.
47. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27-38.
48. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1-22.
49. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: Proceedings of the International Conference on Machine Learning. 2017. p.1321-1330.
50. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
51. Harrell FE. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed. Springer; 2015.
52. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction. 2nd ed. Springer; 2009.
53. He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng. 2009;21(9):1263-1284.
54. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970;12(1):55-67.

55. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Wiley; 2013.
56. Ike PC, Enyinnaya AC, Ujam CG, Ayinde MK, Golden FI, Olatokun TP, *et al*. Investigating the relationship between nutrition and DNA markers associated with heart health in middle-aged adults with type 2 diabetes. Int J Converg Inform Sci Res. 2025;8(9). doi:10.70382/hijcisr.v08i9.035.
57. Ilodigwe L, Adesemoye AC. The data backbone of health system transformation: A governance and architecture framework for interoperability, data quality, and advanced analytics at national scale. Int J Health Pharm Res. 2021;6(2):52-76. doi:10.56201/ijhpr.vol.6.no2.2021.pg52.76.
58. Ilodigwe L, Adesemoye AC. Advances in real-world evidence, predictive analytics, and data-driven decision making for improving health system performance and patient outcomes. Int J Health Pharm Res. 2024;9(5):174-197. doi:10.56201/ijhpr.v9.no5.2024.pg174.197.
59. Ilodigwe L, Adesemoye AC. Advances, risks, and implementation challenges of artificial intelligence as a force multiplier for clinical decision making and health system efficiency. Int J Med Eval Phys Rep. 2025;9(7):165-185. doi:10.56201/ijmepr.v9.no7.2025.pg165.185.
60. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2(8):e124.
61. Ioannidis JPA, Panagiotou OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. JAMA. 2011;305(21):2200-2210.
62. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118-127.
63. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns. 2023;4(9):100804.
64. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, *et al*. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010;11(10):733-739.
65. Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. Wiley; 2002.
66. Mayo W, Ogbole JI, Okoruwa PO, Babatope OM. Designing an AI-predictive maintenance model for e-commerce systems using machine learning and cloud analytics. Int J Sci Res Comput Sci Eng Inf Technol. 2021;7(5):416-440. doi:10.32628/IJSRCSEIT.
67. Meinshausen N, Bühlmann P. Stability selection. J R Stat Soc Series B. 2010;72(4):417-473.
68. Moons KGM, Altman DG, Reitsma JB, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med. 2015;162(1):W1-W73.
69. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. In: Proceedings of the International Conference on Machine Learning. 2005. p.625-632.
70. Nwakamma S, Ojukwu J, Oyesiji SO. LLM-as-a-judge for automated model governance: A review of methods, biases, and release-gating practices. World J Innov Mod Technol. 2024;8(6):185-216. doi:10.56201/wjimt.v8.no6.2024.pg185.216.
71. Obi LI, Dogbanya G, Awah LC, Tawose OM, Ubani C, Ayo-ige AB, *et al*. A systematic analysis of effectiveness of nurse-led dementia care interventions on health outcomes among community-dwelling older adults. J Pharma Insights Res. 2025;3(5):24-33. doi:10.69613/3f5cq717.
72. Ogbete JC, Aminu-Ibrahim AY. Translating healthcare infrastructure investment into measurable population health and diagnostic outcomes. Int J Sci Res Humanit Soc Sci. 2024;1(2):955-985. doi:10.32628/IJSRSSH242775.
73. Ogbete JC, Aminu-Ibrahim AY, Ambali KB. Optimizing laboratory spatial planning strategies to improve diagnostic accuracy, safety, and clinical throughput. Iconic Res Eng J. 2018;2(1):87-113. doi:10.64388/IREV1I7-1713587.
74. Ogbete JC, Aminu-Ibrahim AY, Ambali KB. Regulatory compliant design systems for molecular and pathology laboratories in highly controlled environments. Iconic Res Eng J. 2019;3(4):607-631. doi:10.64388/IREV3I4-1713589.
75. Ogbete JC, Aminu-Ibrahim AY, Ambali KB. Lifecycle performance evaluation of purpose built diagnostic laboratories supporting long term healthcare delivery. Int J Adv Multidiscip Res Stud. 2023;3(6):2605-2621. doi:10.62225/2583049X.2023.3.6.5698.
76. Ogbete JC, Aminu-Ibrahim AY, Iwuanyanwu OC. Scaling molecular diagnostic facilities through standardized infrastructure and operational design models. World Sci News. 2026;212:220-258. doi:10.65770/ZUIY6924.
77. Ogunboye I, Adebayo IPS, Anioke SC, Egwuatu EC, Ajala CF, Awuah SB. Enhancing Nigeria's health surveillance system: A data-driven approach to epidemic preparedness and response. World J Adv Res Rev. 2023.
78. Ogundairo KM, Ayivi-Donkor SS. Machine learning approaches to customer acquisition and retention: A framework for predictive lifecycle management. IIARD Int J Econ Bus Manag. 2023;9(10):210-246. doi:10.56201/ijebm.v9.no10.2023.pg210.246.
79. Okwah MN. Integrating triglyceride-glucose index and echocardiographic parameters for improved cardiovascular risk stratification in Sub-Saharan Africa. Int J Cardiol. 2022;1(2):1-16. doi:10.34218/IJC_01_02_001.
80. Okwah MN, Abah MA, Oladosu MA, Blessing TB, Anku AI, Agida OD. Anti-inflammatory effects of omega-3 fatty acids: Evidence from circulating biomarkers and inflammatory gene-expression studies. Afr Res Rep. 2026;2(4):365-382. doi:10.65221/0167.
81. Oparah OS, Ezeh FE, Olatunji GI, Ajayi OO. AI-based risk stratification framework for large-scale public health emergency preparedness and response planning. Int J Sci Res Comput Sci Eng Inf Technol. 2021;7(1):332-366.
82. Oparah OS, Ezeh FE, Olatunji GI, Ajayi OO. Big data-enabled predictive models for anticipating infectious disease outbreaks at population and regional levels. Int J Adv Multidiscip Res Stud. 2022.
83. Oparah OS, Ezeh FE, Olatunji GI, Ajayi OO. Framework

for designing national real-time disease surveillance dashboards for public health stakeholders. Shodhshauryam Int Sci Refereed Res J. 2023;6(1):208-227.
84. Oparah OS, Ezeh FE, Olatunji GI, Ajayi OO. Framework for integrating climate data and health outcomes to improve mortality risk prediction systems. Int J Sci Res Comput Sci Eng Inf Technol. 2024;10(2):1128-1150.
85. Orise C, Niniola S. Privacy, safeguarding, and ethical data practices in healthcare EdTech: A conceptual model for compliance and trust. Shodhshauryam Int Sci Refereed Res J. 2023;6(1):507-529.
86. Orise C, Niniola S. Data governance in digital health learning systems: A systematic review of implementation frameworks and best practices. Int J Sci Res Comput Sci Eng Inf Technol. 2025;11(5):520-538.
87. Oyesiji SO, Nwakamma S, Ojukwu J. Parameter-efficient multilingual adaptation of on-device language models: A survey. Int J Eng Mod Technol. 2023;9(3):287-320. doi:10.56201/ijcsmt.v9.no3.2023.pg287.320.
88. Oyesiji SO, Nwakamma S, Ojukwu J. A conceptual framework for context-aware data curation in fine-tuning pipelines. Int J Eng Mod Technol. 2024;10(11):197-228. doi:10.56201/ijemt.v10.no11.2024.pg197.228.
89. Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.
90. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-1379.
91. Pepe MS, Etzioni R, Feng Z, *et al*. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst. 2001;93(14):1054-1061.
92. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004;159(9):882-890.
93. Platt JC. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Smola AJ, Bartlett P, Schölkopf B, Schuurmans D, editors. Advances in large margin classifiers. MIT Press; 1999. p.61-74.
94. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer. 2004;4(4):309-314.
95. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer. 2005;5(2):142-149.
96. Riley RD, Ensor J, Snell KIE, *et al*. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441.
97. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. J R Stat Soc Series C. 1994;43(3):429-467.
98. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581-592.
99. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015;10(3):e0118432.
100. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Stat Med. 2007;26(30):5512-5528.
101. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst. 2003;95(1):14-18.
102. Sterne JAC, White IR, Carlin JB, *et al*. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. BMJ. 2009;338:b2393.
103. Steyerberg EW. Clinical prediction models: A practical approach to development, validation, and updating. 2nd ed. Springer; 2019.
104. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-781.
105. Steyerberg EW, Vickers AJ, Cook NR, *et al*. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology. 2010;21(1):128-138.
106. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440-9445.
107. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B. 1996;58(1):267-288.
108. Tonoyan A, Dada O, Ayivi-Donkor SS. Advances in demand forecasting: Machine learning algorithms for revenue projection and financial planning accuracy. Gyanshauryam Int Sci Refereed Res J. 2021;4(1):282-317.
109. Tonoyan A, Dada O, Ayivi-Donkor SS. Conceptual model for predictive cost optimization: Machine learning applications in business transformation analysis. Int J Econ Bus Manag. 2022;8(6):68-102. doi:10.56201/ijebm.v8.no6.2022.pg68.102.
110. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1-67.
111. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17:230.
112. van Smeden M, Moons KGM, de Groot JAH, *et al*. Sample size for binary logistic prediction models: Beyond events per variable criteria. Stat Methods Med Res. 2019;28(8):2455-2474.
113. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics. 2006;7:91.
114. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-574.
115. Whalen S, Schreiber J, Noble WS, Pollard KS. Navigating the pitfalls of applying machine learning in genomics. Nat Rev Genet. 2022;23(3):169-181.

116. Wynants L, Van Calster B, Collins GS, *et al*. Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. BMJ. 2020;369:m1328.
117. Yusuf IA, Badero OJ, Ihesiulo A, Okwah MN, Oshodin A, Asikong E. Aldosterone synthase inhibitors in uncontrolled and resistant hypertension: A phenotype-stratified systematic review and network meta-analysis of randomized trials. PLoS One. 2026;21(6):e0349932. doi:10.1371/journal.pone.0349932.
118. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B. 2005;67(2):301-320.

How to Cite This Article
Nwakamma S, Oyesiji SO, Sankara HWYC, Minoungou TFT. Optimizing regularized logistic regression pipelines on cleaned biomarker datasets: a leakage-proof benchmarking protocol for diagnostic and prognostic modeling. Int J Artif Intell Eng Transform. 2026;7(2):64-72. doi:10.54660/IJAIET.2026.7.2.64-72.