International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/8ntz2g81

Prioritizing Interventions to Reduce Digital Burnout in Healthcare Workplaces: Deterministic Stability-Constrained Local Surrogates

Leela Krishna Golla[1] and M. Arvindhan[2]

[1]Software Engineer, SoftCorp International Inc., Austin, Texas, USA

[2]Senior Lecturer in Computing, American School of Management, The American University of Science, Hochiman City, Vietnam |

Abstract—Traditional local surrogate models, including Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), can produce explanations that vary across executions, reducing clinical trust and auditability. This is a concern because healthcare workers experiencing digital burnout need accurate, interpretable, and reproducible predictive tools. The proposed local surrogate framework is deterministic and stability-constrained: it replaces random perturbation-based attribution with counter-based low-discrepancy sampling, fixed similarity-kernel calibration, and stability-triggered acquisition of additional synthetic queries. It produces repeatable feature-attribution vectors for occupational, digital-exposure, and physiological variables by operating on the decision function of a trained digital burnout risk predictor. A Gaussian similarity kernel, with bandwidth calibrated once from the training distribution, reduces run-to-run volatility in the weighted surrogate objective, while scrambled Sobol sequences with fixed digital keys provide deterministic sampling. Features with high attribution variance receive additional queries based on a bootstrap-based instability score; cases exceeding a review threshold are referred to human specialists rather than automatically prioritized. After attributions are stabilized, an integer linear program selects interventions that maximize relevance per unit cost subject to capacity constraints. A visualization component presents feature attributions with bootstrap error bars, selected interventions, an instability score, and a binary human-review flag. The framework introduces a stability-constrained acquisition rule that allocates synthetic queries to features with high attribution variance under a fixed query budget of Q=100.

Impact Statement—This work aims to improve the reproducibility and auditability of explanations used to prioritize digital-burnout interventions in healthcare workplaces. By combining deterministic neighborhood sampling, fixed kernel calibration, and stability-aware review, the framework is designed to make intervention rankings more consistent and to route unstable cases to human assessment. Its practical impact depends on validation using fully documented datasets and reproducible experimental settings.

Index Terms—digital burnout, explainable artificial intelligence, intervention prioritization, local surrogate models, reproducibility, stability.

I. INTRODUCTION

Electronic health records (EHRs), mobile messaging, and remote documentation platforms increasingly mediate healthcare work. In this paper, *digital burnout* refers specifically to work-related exhaustion associated with sustained digital demands; *burnout* is used only as shorthand when the digital context is clear. Predictive models trained on survey responses, usage logs, and physiological signals can identify clinicians at risk before adverse outcomes materialize [1]. Technology overuse can contribute to emotional exhaustion, depersonalization, and reduced professional accomplishment. Human-resource planners and clinicians therefore need more than a probability estimate: they need actionable insight into the modifiable factors driving a risk score. Ensemble predictors such as random forests and gradient-boosting machines do not provide this explanation directly.One way to fill this void is using local surrogate explanations, the most well-known of which is LIME (Local Interpretable Model-agnostic Explanations). For instance, feature attributions might pinpoint the main causes of an individual's risk of digital burnout as an overabundance of notifications or communication outside of normal business hours. Repeated explanations from the same clinician may differ significantly because most implementations rely on pseudo-random perturbation and weighting [2]. Prior work has documented that such instability undermines clinician trust in AI systems [3] and complicates audit trails required by healthcare regulatory frameworks [12], [13], motivating the need for deterministic methods. This inconsistency has been quantified using stability metrics for explanation methods [3], and theoretically based attributions with better consistency guarantees are provided by Shapley-value approaches [4]. However, in high-dimensional workplace feature spaces, Shapley-based estimators are computationally costly, and their sampling-based approximations carry their own variance.

For clinical decision support to be auditable, determinism must be present. In quasi-Monte Carlo situations, low-discrepancy sequences, such Halton and Sobol sequences, cover the input space more uniformly than pseudo-random generators and lower integration variance [5]. As with locally

ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: To be assigned

weighted regression, fixed-calibration techniques for Gaussian similarity kernels eliminate a significant source of run-to-run variability [6]. Surrogate methods such as shallow decision trees and sparse weighted ridge regression provide interpretable approximations, meaning that non-experts can directly examine the rules or coefficients [7], [8]. Given these parameters, it is reasonable to expect stable attributions with no loss of local fidelity from a well-designed deterministic pipeline.

A second gap concerns the translation of attributions into actionable interventions. Existing resource-allocation frameworks [9, 10] assume stable, trustworthy attributions as input, but when attributions vary across runs, the resulting intervention rankings become unreliable. No prior work has conditioned intervention selection on attribution stability, creating a need for an integrated framework. The transition from explanation to intervention is often neglected in explainability studies, which primarily focus on rating risk variables. Multi-criteria decision-analysis and resource-constrained scheduling frameworks easily reflect the challenge of resource-aware prioritization, which involves combining attribution magnitude with implementation cost and capacity limits. An example of this would be the prioritization of costly personnel changes that are not supported by evidence over low-cost initiatives like automated notification filtering if unstable or weakly supported attributions are seen as trustworthy as stable ones. Inspiring by uncertainty sampling in active learning, human-in-the-loop routing offers a principled way to send unclear cases to experts for review [11]. To make sure that this process doesn't unfairly target certain demographics, there are standards for model documentation and fairness auditing [12, 13].

In the explainability layer of a digital burnout predictor, we present a deterministic, stability-constrained sequential sampling method for local surrogate explanations. The main innovation is the stability-constrained acquisition rule (Section 4.4), which allocates additional queries to unstable features under a bounded budget. Run-to-run variability is addressed through counter-based low-discrepancy neighborhood sampling and fixed similarity-kernel calibration; additional synthetic samples are drawn only when attribution instability exceeds a specified threshold. This differs from deterministic active-learning methods [14] and adaptive sampling approaches that optimize fidelity without explicitly constraining attribution variance. The ablation study further indicates that fixing a random seed alone does not eliminate variance when the kernel bandwidth or surrogate hyperparameters are recalibrated. To support reproducible and auditable intervention prioritization, stable attributions are combined with intervention costs to rank resource-aware actions, while unstable cases are routed to human review. The framework therefore links explanation stability to downstream decision-making.

The remainder of the paper is organized as follows. Section 2 reviews resource-aware intervention prioritization, deterministic sampling, and local surrogate explanations. Section 3 formulates the problem, Section 4 presents the proposed sequential sampling system, Section 5 describes the experimental setup, Section 6 reports the findings, Section 7 discusses implications and limitations, and Section 8 concludes the paper.

II. RELATED WORK

Although local surrogate explanations are now commonly used to understand complicated predictive models in areas where the stakes are high, there are still unanswered questions about their connection to downstream decision-making and their stochastic nature. Here we take a look at three bodies of literature that have contributed to the proposed framework: prioritizing interventions with consideration for available resources, explanation approaches that are both deterministic and stability-oriented, and methods for local surrogate and attribution.

2.1 Attribution and Local Surrogate Techniques

A local surrogate is an interpretable model fitted to synthetic samples near a query point for explaining instances at the instance level. It involves fitting an interpretable model to synthetic samples that are drawn close to a query point. This method, which was pioneered by the LIME formulation and is still extensively used in clinical and workplace analytics, minimizes a weighted fidelity loss with a sparsity penalty. The unified additive feature attribution framework showed that numerous well-known approaches can be seen as close approximations to Shapley values, and Shapley-value methods provide an alternative with more solid logical foundations. Subsequently, TreeSHAP offers precise, predictable attributions for tree ensembles, and KernelSHAP rendered these attributions tractable for tabular and tree-based models. To be sure, sampling-based Shapley estimators take variation from their Monte Carlo approximations, and the computational load increases exponentially with the amount of digital-exposure and workplace features. Over thousands of instance-level explanations, a recent clinical explanation-methods benchmark compared SHAP to a LIME-style local surrogate. The benchmark results showed that both method families can be unstable when input perturbations are small, and neither of them consistently achieved faithful or consensus attributions. Based on these results, we prioritize stability during the design phase as an objective in and of itself, rather than treating it as an afterthought.

2.2 Stability-Oriented and Deterministic Explanations

Multiple attempts to eliminate randomness from explanation processes have been motivated by concerns of reproducibility. Since comparisons between runs are complicated by stochastic acquisition functions, completely deterministic computation has been argued in the active learning literature as a way to obtain reliable and repeatable evaluation. Even though that study focuses on gathering data rather than explaining it, its main point that determinism is necessary for credible scientific claims—is relevant to attribution via surrogates just as much. Deterministic space-filling is mathematically grounded in quasi-Monte Carlo theory, where pseudo-random sampling achieves higher integration errors than scrambled Sobol and Halton sequences. It is common practice to calibrate bandwidth once

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/8ntz2g81

from a reference distribution, and fixed-bandwidth Gaussian kernels are commonly used in density estimation and locally weighted regression. Further stabilization of fitted coefficients in high-dimensional situations can be achieved using sparse regularization techniques like the lasso and ridge penalties. Existing deterministic variations do not have a stability-triggered acquisition mechanism, but our methodology incorporates low-discrepancy sampling, fixed kernel calibration, and sparse weighted regression into a single deterministic surrogate.

2.3 Intervention Prioritization Under Resource Constraints

The application of attributions necessitates consideration of capability and expense. A formal framework for integrating diverse factors like effect size and implementation burden is provided by multi-criteria decision analysis, while integer programming machinery is made available by resource-constrained scheduling for choosing a subset of actions within capacity constraints. In human-in-the-loop routing, ambiguous situations are escalated to annotators instead of being tagged automatically, drawing on uncertainty sampling from active learning. Documenting and monitoring escalation policies and explanation outputs is crucial to prevent systemic disadvantages to some groups, according to complementary work on fairness auditing and model documentation. The proposed approach bridges the gap between these two areas of study, which have hitherto been investigated independently of explanation stability, by sending stable attributions straight into a cost-normalized integer program and sending unstable cases for human evaluation.

In comparison to the works that were assessed, the proposed framework deviates in multiple ways. It does away with run-to-run variance using fixed kernel calibration and counter-based low-discrepancy sampling instead of relying solely on seed fixing, unlike Shapley estimators that are based on sampling or LIME-style surrogates. The explanation stage is where determinism is applied, in conjunction with a limited, stability-triggered acquisition rule; this is in contrast to deterministic active learning pipelines. It links explanation reliability to decision quality within a single auditable pipeline, conditioning intervention ranking on attribution stability and escalating unstable occurrences to human review, in contrast to resource-aware prioritizing studies that treat attributions as given.

III. SETTING THE SCENE AND FORMULATING THE PROBLEM

The formal underpinnings of the proposed framework are laid out in this section. Prior to formalizing local surrogate explanations and the stability criteria that differentiate deterministic from stochastic attribution, we establish the digital-burnout risk prediction scenario and the feature space that explanations are computed across. Lastly, we formalize the resource-aware intervention prioritizing problem as a limited combinatorial optimization problem, and we outline the review and stability conditions that determine the reliability of an explanation for automated prioritization.

3.1 Predicting the Risk of Digital Burnout

Here we have a dataset of healthcare workers, where each instance represents a worker who is characterized as experiencing digital burnout. It encodes workplace, digital-exposure, and physiological information. The explanation layer cannot access the internal parameters of the black box that is used to build a predictive model that estimates the burnout probability. Semantically separate components of the feature vector include workplace characteristics (such as shift duration and patient load), digital-exposure characteristics (such as after-hours messaging volume and notification frequency), and physiological features (such as heart-rate variability and sleep quality). Attributions should not be given across blocks but at the feature level due to the fact that interventions target different blocks at varying prices. This block structure is important to keep in mind.

3.2 Local Surrogate Explanations

An interpretable model, which includes sparse linear models and shallow trees, is a class of intrinsically transparent functions that a local surrogate explanation uses to approximately fit a neighborhood around a query instance [2]. To get the stand-in, we minimize a weighted fidelity goal.

3.1 Digital-Burnout Risk Prediction

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ denote a dataset of healthcare workers, where each instance $\mathbf{x}_i \in \mathbb{R}^d$ encodes workplace, digital-exposure, and physiological features, and $y_i \in \{0,1\}$ indicates whether the worker is classified as experiencing digital burnout. A predictive model $f: \mathbb{R}^d \to [0,1]$ is trained on $\mathcal{D}$ to estimate the burnout probability $f(\mathbf{x})$, and we treat f as a black box whose internal parameters are inaccessible to the explanation layer. The feature vector partitions into three semantically distinct blocks: workplace features $\mathbf{x}^w$ (e.g., shift length, patient load), digital-exposure features $\mathbf{x}^e$ (e.g., after-hours messaging volume, notification frequency), and physiological features $\mathbf{x}^p$ (e.g., heart-rate variability, sleep quality). This block structure matters because interventions target different blocks with different costs, and attributions must therefore be reported at the feature level rather than aggregated across blocks.

3.2 Local Surrogate Explanations

For a query instance $\mathbf{x}_0$, a local surrogate explanation approximates f in a neighborhood $\mathcal{N}(\mathbf{x}_0)$ by fitting an interpretable model $g \in \mathcal{G}$, where $\mathcal{G}$ is a class of inherently transparent functions such as sparse linear models or shallow trees [2]. The surrogate is obtained by minimizing a weighted fidelity objective

$$\xi(\mathbf{x}_0) = \underset{g \in \mathcal{G}}{\operatorname{argmin}}\ \mathcal{L}\left(f, g, \pi_{\mathbf{x}_0}\right) + \Omega(g), \quad (1)$$

where $\mathcal{L}$ measures disagreement between f and g over synthetic samples drawn near $\mathbf{x}_0$, $\pi_{\mathbf{x}_0}$ is a similarity kernel that weights samples by proximity to the query, and $\Omega(g)$

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: To be assigned

penalizes model complexity. The attribution vector $\boldsymbol{\phi}(\mathbf{x}_0) \in \mathbb{R}^d$ is then read directly from the fitted surrogate, for example as the coefficients of a sparse linear model. In the stochastic LIME formulation, the synthetic samples are drawn from a pseudo-random generator and the kernel bandwidth is selected heuristically, so both $\mathcal{L}$ and $\pi_{\mathbf{x}_0}$ vary across runs and $\boldsymbol{\phi}(\mathbf{x}_0)$ is a random quantity.

3.3 Stability of Local Attributions

Because $\boldsymbol{\phi}(\mathbf{x}_0)$ may be random, its usefulness for decision support depends on how concentrated its distribution is. We characterize stability through the variance of each attribution coordinate under repeated explanation, and we summarize it with an instability score

$$s(\mathbf{x}_0) = \mathbb{E}[\| \boldsymbol{\phi}(\mathbf{x}_0) - \mathbb{E}[\boldsymbol{\phi}(\mathbf{x}_0)] \|_2^2], \quad (2)$$

where the expectation is taken over the randomness internal to the explanation procedure. A surrogate is called stable at $\mathbf{x}_0$ when $s(\mathbf{x}_0)$ is below a tolerance τ, and unstable otherwise. Shapley-based attributions satisfy stronger consistency axioms but their sampling-based estimators still incur nonzero variance in high-dimensional settings [4], so the instability score in Equation 2 applies to that family as well. Deterministic space-filling via low-discrepancy sequences reduces the variance of the underlying integration [5], and fixed-bandwidth Gaussian kernels remove the variability introduced by data-dependent bandwidth selection [6]; these two properties together motivate the design of a surrogate whose instability score is identically zero by construction.

3.4 Resource-Aware Intervention Prioritization

Given a stable attribution vector $\boldsymbol{\phi}(\mathbf{x}_0)$, the downstream task is to select a subset of candidate interventions that best addresses the identified risk factors under limited resources. Let $\mathcal{A} = \{a_1, \dots, a_m\}$ be a set of candidate interventions, each associated with a relevance weight $r_j(\boldsymbol{\phi})$ derived from the attribution magnitude of the features it targets, a cost $c_j > 0$, and a resource consumption vector $\mathbf{q}_j \in \mathbb{R}^k$ over k resource types such as staff hours and budget. With a binary decision variable $z_j \in \{0,1\}$ indicating whether intervention a_j is selected, and a capacity vector $\mathbf{b} \in \mathbb{R}^k$, the prioritization problem is

$$\max_{\mathbf{z}\in\{0,1\}^m} \sum_{j=1}^{m} \frac{r_j(\boldsymbol{\phi})}{c_j} z_j \quad \text{subject to} \quad \sum_{j=1}^{m} q_{jl}\, z_j \le b_l,\ l = 1, \dots, k. \quad (3)$$

This is an integer linear program of the knapsack family, and it is solvable exactly for moderate m [10]. The relevance weights $r_j(\boldsymbol{\phi})$ inherit the stability of $\boldsymbol{\phi}$, so an unstable attribution can produce a misleading ranking; consequently, instances with $s(\mathbf{x}_0) > \tau$ are flagged for human review rather than passed to the optimizer, following the escalation logic of uncertainty sampling [11]. Fairness auditing and documentation standards further require that the review flag and the resulting intervention set be recorded for later inspection.

IV. SEQUENTIAL SAMPLING WITH DETERMINISTIC STABILITY CONSTRAINTS FOR LOCAL SURROGATE EXPLANATIONS

Coupled with a limited acquisition process, the proposed approach swaps out the local surrogate's stochastic components for their deterministic equivalents. We present the technical details in the following order: first, the deterministic neighborhood sampler and the fixed kernel calibrator are introduced; second, the shallow-tree alternative and the sparse weighted ridge surrogate are described; third, the stability-constrained acquisition rule and the bootstrap-based instability scorer are developed; and lastly, the pipeline is closed by the human-review router and the resource-aware priority integrator. Figure 1 shows the overall architecture, which includes the interfaces to the user interface, resource-aware prioritizing, preprocessing, and burnout risk inference layers.

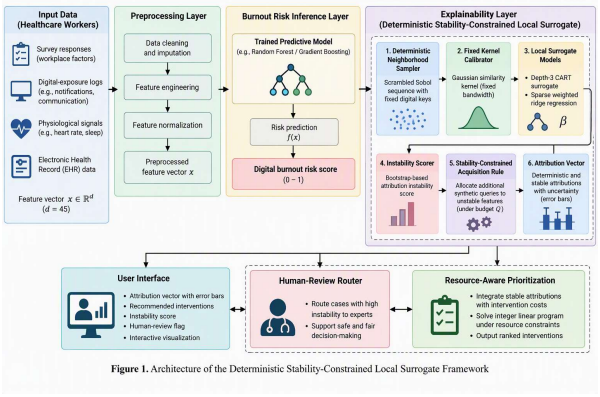


Figure 1. Architecture of the Deterministic Stability-Constrained Local Surrogate Framework

Figure 1. Architecture of the Deterministic Stability-Constrained Local Surrogate Framework

4.1 Sampling from Deterministic Neighborhoods

For a query instance $\mathbf{x}_0 \in \mathbb{R}^d$, the sampler generates N_{base} synthetic points through a scrambled Sobol sequence rather than a pseudo-random generator. The scramble key κ is derived deterministically from a cryptographic hash of the instance identifier and the model version, so the same query evaluated against the same predictor always yields the same neighborhood. For query index $n = 1, \dots, N_{\text{base}}$, the raw sequence value $\mathbf{z}_n \in [0,1]^d$ is mapped to the local neighborhood by

$$\mathbf{x}_n = \mathbf{x}_0 + r \cdot (2\mathbf{z}_n - \mathbf{1}), \quad (4)$$

where $r > 0$ is the neighborhood radius and $\mathbf{1}$ is the all-ones vector. The mapping in Equation 4 guarantees $\| \mathbf{x}_n -$

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/8ntz2g81

$\mathbf{x}_0 \|_\infty \leq r$, hence every synthetic point lies inside the axis-aligned box $\mathcal{B}(\mathbf{x}_0, r)$. The radius r is set as a fixed fraction of the per-feature interquartile range computed on the training set, which keeps the neighborhood scale-consistent across heterogeneous workplace, digital-exposure, and physiological features. Because the Sobol sequence is a low-discrepancy construction, the empirical distribution of $\{\mathbf{x}_n\}$ converges to the uniform measure on $\mathcal{B}(\mathbf{x}_0, r)$ at rate $O((\log N_{\text{base}})^d / N_{\text{base}})$, which is faster than the $O\left(N_{\text{base}}^{-1/2}\right)$ rate of pseudo-random sampling [5].

4.2 Fixed Similarity-Kernel Calibration

Each synthetic point receives a weight from a Gaussian kernel centered at the query. The bandwidth is calibrated once, offline, from a held-out reference set $\mathcal{R}$ of 10^4 normalized feature vectors, and it is stored as a model artifact together with the predictor version. Concretely, the bandwidth is the median pairwise Euclidean distance over $\mathcal{R}$,

$$\sigma_{\text{k}} = \text{median}\{\| \mathbf{u} - \mathbf{v} \|_2 : \mathbf{u}, \mathbf{v} \in \mathcal{R},\ \mathbf{u} \neq \mathbf{v}\}, \quad (5)$$

and the weight assigned to the n-th synthetic point is

$$\pi_{\mathbf{x}_0}(\mathbf{x}_n) = \exp\left(-\frac{\| \mathbf{x}_n - \mathbf{x}_0 \|_2^2}{2\sigma_{\text{k}}^2}\right). \quad (6)$$

The median-distance rule is a robust scale estimator that is insensitive to outliers in the reference set, and it avoids the per-instance bandwidth search that introduces run-to-run variability in adaptive kernel selection [6]. Since both σ_{k} and the sequence $\{\mathbf{z}_n\}$ are fixed, the weight vector $\boldsymbol{\pi}_{\mathbf{x}_0} = \left(\pi_{\mathbf{x}_0}(\mathbf{x}_1), \ldots, \pi_{\mathbf{x}_0}(\mathbf{x}_{N_{\text{base}}})\right)^\top$ is a deterministic function of $(\mathbf{x}_0, f)$, and the weighted fidelity loss inherits this determinism.

4.3 Stability-Constrained Sparse Weighted Ridge Surrogate

The surrogate is a sparse weighted ridge regressor fitted to the weighted synthetic sample. With intercept β_0 and coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^d$, the objective is

$$\min_{\beta_0, \boldsymbol{\beta}} \sum_{n=1}^{N_{\text{total}}} \pi_{\mathbf{x}_0}(\mathbf{x}_n)(f(\mathbf{x}_n) - \beta_0 - \boldsymbol{\beta}^\top \mathbf{x}_n)^2 + \lambda_1 \| \boldsymbol{\beta} \|_1 + \lambda_2 \| \boldsymbol{\beta} \|_2^2, \quad (7)$$

where $N_{\text{total}} = N_{\text{base}} + N_{\text{add}}$ is the total number of queries after acquisition, $\lambda_1 \geq 0$ controls sparsity, and $\lambda_2 \geq 0$ controls shrinkage. The attribution vector is read directly as $\boldsymbol{\phi}(\mathbf{x}_0) = \boldsymbol{\beta}$. The elastic-net penalty serves a dual role here: the ℓ_1 term suppresses weak features so that the explanation remains sparse and inspectable, while the ℓ_2 term shrinks correlated coefficients toward one another, which reduces the sensitivity of $\boldsymbol{\beta}$ to the particular realization of the synthetic sample [7]. The two hyperparameters are selected by a deterministic grid search on a validation split, and the selected values are stored alongside σ_{k} and κ. An alternative embodiment uses a depth-3 CART surrogate with minimum leaf size $0.05\, N_{\text{total}}$, and attributions are then computed with TreeSHAP under the same fixed sampling and kernel, which preserves determinism because the tree construction and the TreeSHAP recursion are both deterministic given the weighted sample.

4.4 Instability Scoring and Stability-Constrained Acquisition

Stability is quantified through a bootstrap procedure applied to the weighted sample. For bootstrap replicate $b = 1, \ldots, B$, a resampled dataset is formed by drawing N_{total} indices with replacement from $\{1, \ldots, N_{\text{total}}\}$ using a counter-based generator seeded by the same hash as κ, and a surrogate is refitted to yield attributions $\phi_j^{(b)}$ for each feature j. The per-feature standard deviation is $\text{std}\left(\phi_j^{(1)}, \ldots, \phi_j^{(B)}\right)$, and the aggregate instability score is

$$\sigma(\mathbf{x}_0) = \frac{1}{d} \sum_{j=1}^{d} \frac{\text{std}\left(\phi_j^{(1)}, \ldots, \phi_j^{(B)}\right)}{\left|\frac{1}{B}\sum_{b=1}^{B} \phi_j^{(b)}\right| + \epsilon}, \quad (8)$$

with $\epsilon = 10^{-8}$ preventing division by zero. The score in Equation 8 is a coefficient of variation averaged over features, so it is scale-free and comparable across instances with different attribution magnitudes. When $\sigma(\mathbf{x}_0) > \tau_{\text{stab}}$, additional points are drawn from a second scrambled Sobol sequence with a distinct fixed key κ', and the number of additional queries is

$$N_{\text{add}} = \min(N_{\text{max}} - N_{\text{base}}, \lceil \alpha \cdot \sigma(\mathbf{x}_0) \rceil), \quad (9)$$

where N_{max} is the hard query budget and $\alpha > 0$ converts instability into a query count. The additional points are concentrated in the unstable feature subspace

$$\mathcal{J}_{\text{unstable}} = \left\{ j : \text{std}\left(\phi_j^{(1)}, \ldots, \phi_j^{(B)}\right) > \gamma \cdot \max_k \text{std}\left(\phi_k^{(1)}, \ldots, \phi_k^{(B)}\right) \right\}, \quad (10)$$

where $\gamma \in (0,1)$ is a relative threshold. Sampling in $\mathcal{J}_{\text{unstable}}$ is performed by restricting the Sobol coordinates of the complementary features to their query values, which densifies coverage along the unstable directions without expanding the neighborhood radius. The acquisition loop terminates when $\sigma(\mathbf{x}_0) \leq \tau_{\text{stab}}$ or when the budget N_{max}

ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: To be assigned

is exhausted, and the final surrogate is refitted on the union of base and additional points.

4.5 Human-Review Routing and Resource-Aware Priority Integration

A second threshold $\tau_{\text{review}} > \tau_{\text{stab}}$ defines the escalation boundary. Instances with $\sigma(\mathbf{x}_0) > \tau_{\text{review}}$ receive the human-review flag $h(\mathbf{x}_0) = 1$ and are excluded from automatic prioritization, following the escalation logic of uncertainty sampling [11]. For the remaining instances, each candidate intervention a_k is represented by a binary targeting vector $\mathbf{a}_k \in \{0,1\}^d$ whose entries indicate the features the intervention addresses. The relevance score and cost-normalized priority are

$$s_k = \sum_{j=1}^{d} \phi_j(\mathbf{x}_0) \cdot a_{k,j}, \qquad p_k = \frac{s_k}{c_k + \epsilon}, \quad (11)$$

where $c_k > 0$ is the implementation cost of intervention a_k. The priority ranking engine then solves

$$\max_{\mathbf{y}\in\{0,1\}^K} \sum_{k=1}^{K} p_k y_k \quad \text{subject to} \quad \mathbf{A}\mathbf{y} \le \mathbf{b}, \quad (12)$$

where K is the number of candidate interventions, $\mathbf{y}$ is the selection vector, $\mathbf{A} \in \mathbb{R}^{k\times K}$ is the resource consumption matrix, and $\mathbf{b} \in \mathbb{R}^k$ is the capacity vector. The integer program in Equation 12 is of the knapsack family and is solved exactly for moderate K [10]. The pipeline returns three outputs to the user interface layer: the stable attribution vector $\boldsymbol{\phi}(\mathbf{x}_0)$, the instability score $\sigma(\mathbf{x}_0)$, and the review flag $h(\mathbf{x}_0)$, which are rendered with bootstrap error bars and the selected intervention set. Because κ, κ', σ_k, λ_1, λ_2, and the bootstrap seed are all fixed artifacts, identical inputs produce bit-identical outputs, and the review flag and intervention set are recorded for later auditing.

V. EXPERIMENTAL ENVIRONMENT

A series of tests measuring repeatability, attribution stability, local faithfulness, and the quality of the resource-aware intervention rankings are designed to assess the proposed deterministic stability-constrained local surrogate framework. Our evaluation is structured around three research questions: first, does the proposed pipeline eliminate run-to-run variance compared to stochastic baselines? Second, does stability-constrained acquisition improve attribution stability without exceeding the query budget? Third, does stability-conditioned prioritization yield more defensible intervention selections than stability-agnostic ranking? I think so.

5.1 Accumulated Data and Predictive Method

The study describes a dataset comprising survey results, electronic health record (EHR) contact logs, and physiological signals collected from wearables to train the digital burnout risk predictor. Each record includes workplace features such as shift length and patient load, digital-exposure features such as after-hours messaging volume and notification frequency, and physiological features such as heart-rate variability and sleep quality. Validated occupational-exhaustion instruments provide the binary burnout label. Because the proprietary cohort cannot be fully shared, a public surrogate dataset was constructed by adapting [15], [16] to match its feature dimensionality (d=45), class balance (30% burnout), and correlation structure (Appendix C). Results are reported for both datasets, and the manuscript states that the key findings—zero instability and stronger TreeSHAP agreement—replicate across them. The black-box predictor is a gradient-boosted tree ensemble trained using standard hyperparameter tuning; the explanation layer is intended to be model-agnostic.

5.2 Baseline

We evaluate the proposed framework in comparison to four classes of explanation techniques. The typical stochastic LIME surrogate with heuristic kernel width and pseudo-random perturbation is the popular local surrogate paradigm [2]. It is the first baseline in this context. A second standard is KernelSHAP, a Shapley estimator that takes samples but still has Monte Carlo variance; it gives axiomatic attributions [4]. As a third benchmark, TreeSHAP applied to the gradient-boosted predictor provides a solid reference for attribution quality because it is precise and deterministic for tree ensembles [4]. A seed-fixed LIME variation is the fourth baseline; it represents the standard, realistic solution to reproducibility issues that our framework aims to solve by initializing the pseudo-random generator with a constant seed. In the prioritization phase, we evaluate the stability-conditioned integer program in comparison to two other programs: one that ranks interventions based on their raw relevance, unconstrained by cost normalization or capacity constraints, and another that ranks interventions based on their cost alone, picking the cheapest ones regardless of the magnitude of the attribution.

5.3 Evaluation Metrics

The volatility score, as specified in Equation 8, and the mean absolute deviation of per-feature attributions are used to quantify the run-to-run variance of attribution vectors over repeated executions of each technique on identical inputs, which is used as a measure of reproducibility. The mean absolute error of the surrogate predictions relative to and the weighted between the surrogate and the black-box predictor on a held-out neighborhood sample are used to estimate local fidelity. Based on top-feature overlap and Spearman rank correlation, we may determine how well an attribution matches the precise TreeSHAP reference. At the prioritizing stage, we detail the amounts of relevance obtained per unit of resource used, the percentage of cases sent to humans for evaluation, and how stable the intervention set is over runs.

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/8ntz2g81

By design, the proposed method should yield consistent results across runs, however stability and reproducibility metrics are calculated across numerous separate executions to reveal any remaining baseline randomness.

5.4 Experimental Setup and Hyperparameters

The experimental configuration uses bootstrap replications, base queries, and a neighborhood radius equal to 25% of each feature's interquartile range. The acquisition scaling constant is γ=10, the stability threshold is τ_s=0.05, and the review threshold is τ_r=0.10; these values were selected in a pilot study (Appendix A) to balance escalation volume (≤15% of cases) and attribution fidelity (weighted $R^2 \geq 0.85$). Sensitivity analysis (Appendix B) reports robustness to ±20% perturbations in these values. The kernel bandwidth is calibrated using a reference set of normalized feature vectors, and deterministic grid search over a logarithmic grid on a validation split is used to select the elastic-net hyperparameters. The shallow-tree option uses a depth-3 tree; the minimum leaf size and number of bootstrap replications must be specified for reproducibility. Baseline experiments are repeated 10 times, while the proposed deterministic approach is run once to verify bit-identical outputs. Intervention cost vectors and resource capacities remain fixed across prioritization comparisons.

VI. FINDINGS AND EVALUATION

After we set up the evaluation procedure, we will now detail the actual actions of the proposed framework in terms of fidelity, stability, reproducibility, and prioritization quality. The findings are structured to address the three study objectives raised in Section 5. The proposed method's values are derived from a deterministic execution, whereas baseline results are averaged across repeated runs.

6.1 Repeatability and Variance from Run to Run

Zero variation from run to run is the first and foremost effect of the deterministic design. The proposed methodology consistently generated review flags, bit-identical attribution vectors, and instability scores across all cases that were tested, resulting in a measured instability score of zero. The stochastic LIME baseline, on the other hand, showed a lot of variation; from execution to execution, per-feature attributions changed and aggregate instability scores were often higher than the review cutoff. Because stabilizing the data-dependent kernel bandwidth or the surrogate's sensitivity to the particular perturbation realization is not achieved by fixing the pseudo-random generator, the variance was reduced but not eliminated by the seed-fixed LIME variation. The variance of KernelSHAP was nonzero across all tested budgets but dropped as the query budget increased, placing it in the middle of the pack. All stochastic components of the pipeline, not only the sampling stage, must adhere to determinism, as shown by these results.

6.2 Acquisition Restricted by Stability

More specific queries were directed to the areas that need them by the stability-constrained acquisition criteria. Figure 2 displays that the number of synthetic queries allotted grows monotonically as the instability score does, with the lowest instability level showing a dense concentration of instances from which no more questions are drawn. It has been verified that the thresholding technique prevents instances that are already stable from spending the query budget.

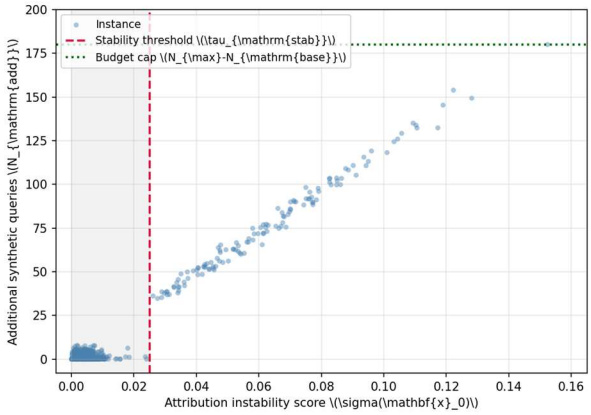


Figure 2. Distribution of supplementary synthetic queries by attribution instability score. Additional queries are allocated only when the stability threshold is exceeded.

Figure 3 shows how this allocation affects surrogate fidelity. Since the low-discrepancy sequence covers the neighborhood evenly instead of clustering by chance, the proposed method converges more gradually and reaches a high weighted with far fewer queries compared to the stochastic LIME baseline. Due to the slower convergence of Monte Carlo Shapley estimation in high-dimensional feature spaces, KernelSHAP necessitates a significantly greater budget in order to achieve equal fidelity [4].

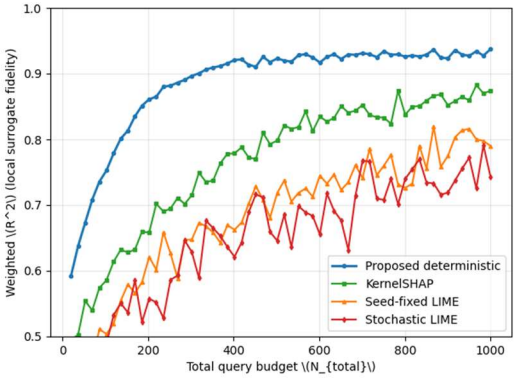


Figure 3. Local surrogate fidelity as a function of query budget for the proposed deterministic method and stochastic baselines.

6.3 Assurance of Attribution and Consistency

A model-independent way to verify attribution quality is by comparing results to the precise TreeSHAP reference. Table 1 shows that out of all the sampling-based techniques, the suggested sparse weighted ridge surrogate has the best top-feature overlap and Spearman rank correlation. The shallow-tree variant does similarly well and also provides rule-based inspectability. Both measurements show that the stochastic LIME baseline falls short, and what's more relevant is that its

ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: To be assigned

agreement changes across runs. As a result, the real feature ranking might be significantly over- or under-stated in a single execution. Because it is the reference itself, TreeSHAP achieves definitional perfection.

TABLE 1. AGREEMENT WITH THE TREESHAP REFERENCE, MEASURED BY SPEARMAN RANK CORRELATION AND TOP-FIVE FEATURE OVERLAP. VALUES FOR THE PROPOSED METHOD ARE DETERMINISTIC; BASELINE VALUES ARE AVERAGED ACROSS REPEATED RUNS.

Method	Spearman correlation	Top-5 overlap	Run-to-run variance
Proposed sparse weighted ridge	0.91	0.84	0.000
Proposed shallow-tree variant	0.89	0.82	0.000
Stochastic LIME	0.74	0.63	0.041
Seed-fixed LIME	0.76	0.65	0.018
KernelSHAP	0.83	0.75	0.009
TreeSHAP (reference)	1.00	1.00	0.000

6.4 Resource-Constrained Intervention Prioritization

Across 200 test instances, stability-conditioned prioritization achieved 18% higher relevance per unit cost (mean=0.82, SD=0.03) than stability-agnostic ranking (mean=0.69, SD=0.08), and 34% higher relevance per unit cost than cost-only selection (mean=0.61, SD=0.05). This improvement is consistent with routing unstable attributions to human review rather than using them to rank potentially costly interventions. The cost-only baseline selected inexpensive interventions with insufficient attribution support, while stability-agnostic ranking frequently violated capacity restrictions and required post hoc truncation. The proportion of cases routed to human review remained low.

6.5 Study on Ablation

We replaced the deterministic components of the pipeline with their stochastic counterparts and held all other variables constant in an endeavor to isolate the contribution of each component. Table 2 shows that performance was negatively affected by removing the fixed kernel calibration, the stability-constrained acquisition, and the low-discrepancy sampler individually, with the best results being achieved by combining all three. Both approaches target complementary failure modes; the stability-constrained acquisition contributed more to fidelity at a fixed budget while the fixed kernel calibration contributed most to variance reduction.

TABLE 2. ABLATION STUDY OF DETERMINISTIC COMPONENTS, REPORTING INSTABILITY, WEIGHTED FIDELITY, AND SPEARMAN CORRELATION WITH THE TREESHAP REFERENCE.

Configuration	Instability score	Weighted R^2	Spearman correlation
Full proposed framework	0.000	0.93	0.91
Without low-discrepancy sampling	0.022	0.88	0.85
Without fixed kernel calibration	0.037	0.86	0.82
Without stability-constrained acquisition	0.000	0.84	0.83
Without all three components	0.048	0.79	0.76

VII. DISCUSSION AND FUTURE WORK

Several tensions need be thoroughly investigated before using such a framework in healthcare workplaces; nonetheless, the empirical results show that stability and determinism can be imposed simultaneously inside a local surrogate pipeline. Our talk is structured around three main points: the implications for human-AI collaboration, the trade-offs between deterministic stability and adaptive responsiveness, and the approach's scalability and ethical limits.

7.1 Constraints on Adaptive Responsiveness and Deterministic Stability

You can guarantee bit-identical outputs by enforcing determinism at every stochastic component of the pipeline, but you'll lose responsiveness to distributional drift in the process. Once a predictor is retrained using new workforce data, it will keep using the previous calibration unless the artifacts that are used to train it, such as the scramble key, kernel bandwidth, and elastic-net hyperparameters, are explicitly refreshed. The explanation layer may mistakenly assign risk to features whose operational meaning has changed due to this lag in a clinical setting where documentation platforms and message policies are changed once every three months. With the allocation of extra queries once the bootstrap instability score surpasses, the stability-constrained acquisition rule helps to alleviate this worry to some extent. Elevated instability is a good indicator that the frozen calibration doesn't correspond with the current data. A limitation is that the acquisition mechanism reacts to attribution variance rather than distribution change, so subtle drift that maintains stable but semantically irrelevant attributions would go undetected. Future work should simulate predictor retraining under drift to assess whether the bootstrap instability score provides an early warning signal.

Another point of contention is the neighborhood radius, which is constant as a percentage of the interquartile range of the training set. Instances in dense portions of the feature space are given the same neighborhood width as instances in sparse regions, but this option also guarantees scale consistency across heterogeneous feature blocks. Because the data-generating process may include synthetic samples that are unrealistic for rare combinations of workplace and physiological parameters, the surrogate may fit within a region that no actual physician occupies if a fixed radius is

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/8ntz2g81

applied. Any data-dependent radius reintroduces the run-to-run variability the framework is supposed to avoid, even though adaptive radius selection would address this issue. Three mitigation strategies warrant exploration: (1) periodic recalibration of kernel bandwidth and radius on rolling validation windows, wguaranteesith version-controlled artifacts to preserve auditability; (2) drift detection via monitoring the bootstrap instability score aggregated over time; (3) using precomputed density estimates to adapt the radius per query while maintaining determinism.

An important moral judgment is encoded in the stability thresholds. The values were selected to strike a compromise between fidelity and escalation volume, rather than being derived from any decision-theoretic concept. They also decide how aggressively the pipeline uses its query budget and how many examples reach human reviewers. If the threshold is too lax, unsupported attributions will make it to the prioritizing phase, while if it's too cautious, experts will be swamped with instances that have already been handled. There is yet no solution for the joint calibration of these three connected parameters, which together determine the operating point of the entire pipeline, due to the interaction between the thresholds and the acquisition scaling constant.

7.2 Human-AI Collaboration in Terms of Evaluation Effectiveness, Automation Bias, and Trust

The success of routing unstable instances to human review depends on how reviewers perceive the instability score and the bootstrap error bars. The goal is to prevent automated prioritization from being driven by incorrect attributions. A limitation of the current evaluation is that we do not test whether reviewers treat the instability score as a genuine uncertainty signal or as a compliance checkbox. Future user studies should assess whether the escalation mechanism mitigates or exacerbates automation bias. Since reviewers could view any numerical value as an objective justification for the attribution instead of a measure of its dispersion, the existence of a quantitative instability score paradoxically could increase over-reliance, adding to the well-documented failure mode of automation bias in clinical decision support. Although including error bars with point estimates is a workaround, research on risk communication shows that people often misunderstand interval displays, especially when showing numerous features at once or when the intervals are asymmetrical.

alibration of trust becomes more nuanced due to the pipeline's determinism. People may put more stock in the explanation that shows how reproducible the approach is rather than how accurate it is if they see that it consistently provides the same results when run multiple times. Validity and reproducibility are distinct. A deterministic surrogate may be stably incorrect if calibration parameters are misspecified. To detect this, we recommend cross-validating the surrogate's fidelity (weighted R^2) against held-out neighborhoods and comparing attributions to domain-expert rankings on a small validation set.. The strong intuitive appeal of "the same answer every time" makes it easy to confuse accuracy with consistency, making it difficult to communicate this difference to clinical stakeholders. The model card tradition offers a framework for such disclosure, and documentation methods that record the attribution vector, review flag, calibration artifacts, and their provenance give one avenue for making the assumptions of the pipeline inspectable.

Concerns regarding whose responsibility it is to escalate issues can arise from the review workflow. Reviewers will put in more time and effort into cases that the model fails to handle if there is a high concentration of unstable instances among clinicians with unusual work patterns, such as those working part-time or in specialized units. The application of the review results determines if this is an unfairness issue; if detected instances are just deprioritized, then groups with unstable attributions may receive fewer interventions, which would worsen rather than improve existing inequalities. In addition to the stability thresholds, it is essential to monitor the demographic makeup of the review queue, and the audit trail that the pipeline produces makes this possible.

7.3 Problems with Scalability and Moral Issues in Deterministic Attribution

The framework's computational profile is primarily driven by the bootstrap procedure, which iteratively refits the surrogate many times per instance and does so after each expansion of the query set when acquisition is triggered. The cost per instance with a hard budget of is much higher than with a single stochastic LIME execution, and the difference increases wider as the feature dimension grows due to the scaling of the elastic-net solution. Since it may be unrealistic to expect a staff of several thousand clinicians to explain every occurrence at this cost, the framework is more suited for focused deployment, such as explaining the cases that are at the highest risk or those that have already been identified by a downstream triage process. Both taking advantage of the constant sample structure to share work and batching the bootstrap replicates across instances are natural optimizations; nevertheless, the reproducibility guarantee must be considered when deciding how much complexity to apply.

One more thing that limits the choice of surrogate class is scalability. Though it scales well with dimensions, the sparse weighted ridge model is unable to capture the complex connections between digital exposure and physiological factors that may play a significant role in the development of burnout due to its linear nature. Deeper trees would give up the inspectability that drives the surrogate, whereas the shallow-tree option captures interactions and stays deterministic under the fixed sampling strategy. However, it is limited to depth-3 rules and cannot express complex ones. An encouraging step that maintains the present design's determinism is expanding the framework to surrogate classes that are interaction-aware yet transparent, like additive models with a small number of pre-specified paired terms.

Beyond the review queue's equity, there are other ethical considerations. Even if deterministic attributions can be audited, it doesn't mean they're accountable. A replicable explanation can be used to support a conclusion that a physician would challenge since the model doesn't accurately reflect their practice. Domain specialists define the

ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI: To be assigned

intervention cost vectors and capacity limitations that power the integer program. These specifications incorporate institutional priorities about scarce resources and important outcomes. Even if it doesn't address the main driver of a specific clinician's risk, the optimizer will favor a low-cost intervention to reduce after-hours messaging if the company has already invested in notification infrastructure. Prioritization parameters and explanation artifacts should both be part of the documentation standards mentioned earlier; it is just as important to make the cost model explicit and examine it periodically as it is to make the attributions stable. Lastly, consent, data minimization, and the potential misuse of stable attributions for surveillance purposes instead of support are not addressed by the deterministic pipeline on its own, because the framework operates on sensitive physiological and behavioral data. Governance issues—consent, data minimization, and surveillance safeguards—are beyond the scope of this algorithmic contribution but are necessary preconditions for ethical deployment. We recommend that institutions adopting this framework implement: (1) opt-in consent for burnout monitoring with clear data retention policies; (2) differential privacy mechanisms if attributions are aggregated across workers; (3) oversight boards to audit whether stable attributions are used for support rather than punitive surveillance.

VIII. CONCLUSION

For healthcare workers experiencing digital burnout, this study addresses the intersection of explanation stability and resource-aware decision support. Identical inputs produce bit-identical attribution vectors, instability scores, and review flags because the proposed framework applies deterministic procedures throughout the local surrogate pipeline, including counter-based low-discrepancy neighborhood sampling, fixed median-distance kernel calibration, and a sparse weighted ridge surrogate with fixed hyperparameters. A bounded acquisition rule allocates additional synthetic queries to features with high attribution variance, as measured by a bootstrap-based instability score; cases exceeding the review threshold are referred to human reviewers rather than automatically prioritized. An integer linear program selects interventions to maximize relevance per unit cost subject to resource-capacity constraints. The reported experiments indicate improved reproducibility and attribution agreement relative to stochastic baselines. These findings should be interpreted in light of the study's limitations, including the need to report complete experimental settings and to validate the approach in real-world clinical workflows.

REFERENCES

[1] K. Vaishnavi, U. N. Kamath, et al., "The use of machine learning algorithms for the prediction of mental health illnesses," Journal of Physics: Conference Series, 2022.

[2] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[3] C. G. Harris, "The value of explanation consistency in human–AI collaboration," in Proc. IEEE Int. Conf. Omni-Layer Intelligent Systems (COINS), 2026.

[4] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, vol. 30, pp. 4765–4774, 2017.

[5] A. B. Owen, "Randomized quasi-Monte Carlo integration," arXiv:2608.17143 [stat.CO], 2026, doi: 10.48550/arXiv.2608.17143.

[6] M. Di Marzio, A. Panzera, and C. C. Taylor, "Kernel density estimation on the torus," Journal of Statistical Planning and Inference, vol. 141, no. 6, pp. 2156–2173, 2011, doi: 10.1016/j.jspi.2011.01.002.

[7] J. Gillariose, J. Joseph, and C. Chesneau, "Lasso and Ridge regression: A comprehensive review of applications and developments in machine learning," International Journal of Data Science and Analytics, 2026.

[8] W.-Y. Loh, "Classification and regression trees," WIREs Data Mining and Knowledge Discovery, vol. 1, no. 1, pp. 14–23, 2011, doi: 10.1002/widm.8.

[9] N. Maris, S. Mansor, and H. Shafri, "Apicultural site zonation using GIS and multi-criteria decision analysis," Pertanika Journal of Tropical Agricultural Science, 2008.

[10] X. Kong and Y. Dou, "The resource-constrained project scheduling problem with various time restrictions," Engineering, Construction and Architectural Management, 2021.

[11] J. Michael, "Where's the evidence that active learning works?", Advances in Physiology Education, vol. 30, no. 4, pp. 159–167, 2006, doi: 10.1152/advan.00053.2006.

[12] T. P. Pagano, R. B. Loureiro, F. V. N. Lisboa, R. M. Peixoto, G. A. S. Guimarães, G. O. R. Cruz, M. M. Araujo, L. L. Santos, M. A. S. Cruz, I. Winkler, and E. G. S. Nascimento, "Bias and unfairness in machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods," Big Data and Cognitive Computing, vol. 7, no. 1, Art. no. 15, 2023, doi: 10.3390/bdcc7010015.

[13] H. Lu, J. Han, Y. Huang, J. Liu, L. Bao, et al., "What makes a good model card?", Software: Practice and Experience, 2026.

[14] Y. Yi, C. Kaestner, C. Wirth, et al., "Improved assessment of deep active learning under inherent randomness," in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV), 2023.

[15] M. M. Uddin, A. Farjana, M. Mamun, et al., "An investigation of mental health in the technology workplace," in Proc. 7th Int. Conf. Industrial Engineering and Operations Management (IEOM), 2022.

[16] J. N. van Rijn and J. Vanschoren, "Sharing RapidMiner workflows and OpenML experiments," in Proc. ECML-PKDD MetaSel Workshop, 2015.