

Foundations for an Enhanced XAI System: A Systematic Literature Review on Optimising LIME for Distance Learning Environments

Dr. M. O. Oyediran¹, Dr. A. Ojo², Abdulsalaam Aminat Obakhume³
^{1,2}Department of Computer Science Faculty of Computing Ajayi Crowther University, Oyo
³Department of Computer Science, Federal College of Education (Special), Oyo

ARTICLE INFO	ABSTRACT
Published Online: 27 January 2026 Corresponding Author: Abdulsalaam Aminat Obakhume	Explainable Artificial Intelligence (XAI) has become increasingly essential in distance learning systems, where predictive models are used to support assessment, personalised feedback, learner monitoring, and early-warning interventions. Local Interpretable Model-Agnostic Explanations (LIME) is a widely adopted XAI techniques due to its model flexibility and ease of implementation. However, LIME suffers from methodological limitations and limited global interpretability, that undermine its suitability for high-stakes educational applications. This systematic literature review analysed research published on LIME optimisation, categorising existing enhancements into six major strategy groups: deterministic and uncertainty-based sampling, data-aware perturbations, nonlinear surrogate models, global and aggregated explainability methods, visual and relational interpretation enhancements, and hybrid or optimisation-driven frameworks. Findings show that while numerous variants significantly improve stability, fidelity, or interpretability within their respective domains, most remain domain-specific, computationally demanding, and insufficiently evaluated through human-centred or educational lenses. Only a few studies explore LIME optimisation within learning analytics, leaving a substantial gap in applying enhanced XAI methods to distance learning environments. The review identifies unresolved challenges like the need for multimodal and sequential data handling, unified optimisation across multiple objectives, scalable real-time implementation, and empirically validated human interpretability. Insights from literature inform a set of design requirements for an improved LIME-based XAI system designed for distance learning.
KEYWORDS: Explainable Artificial Intelligence (XAI), Local Interpretable Model-Agnostic Explanations (LIME), Distance Learning Analytics, Model Optimisation Techniques, Educational Machine Learning	

BACKGROUND TO THE STUDY

Advances in artificial intelligence (AI) have led to increasing adoption of predictive analytics, personalised tutoring systems, automated assessment, and learning support tools in distance education. These systems often rely on complex machine learning models whose internal workings are not transparent to educators or learners, raising concerns about accountability, fairness, and pedagogical trust (Raza et al., 2024). Explainable Artificial Intelligence (XAI) has therefore become an important requirement in digital learning environments to enable stakeholders understand why a prediction was generated and how student behaviours or learning patterns influence model outputs (Nnadi et al., 2024). Among existing XAI techniques, Local Interpretable

Model-Agnostic Explanations (LIME) has gained widespread use because it provides human-readable explanations for any classifier or regressor, making it particularly appealing for heterogeneous educational datasets (Hassija et al., 2024). Despite its popularity, LIME exhibits several methodological and practical limitations that constrain its usefulness in high-stakes educational settings (Bramhall, 2020). Studies consistently report that LIME’s explanations are unstable across repeated runs due to its reliance on random perturbations, making it difficult to ensure consistency in feedback and intervention decisions (Bacevicius et al., 2025). Its perturbation strategies often fail to respect the structure of educational data, such as temporal dependencies in learning sequences or correlated student behaviours, resulting in

unrealistic synthetic samples and reduced fidelity of explanations. Furthermore, LIME’s linear surrogate model struggles to approximate nonlinear relationships commonly found in learning analytics, such as interaction effects between time-on-task, resource difficulty, and cognitive engagement.

These challenges have significant implications for distance learning, where interpretability must be technically sound and pedagogically meaningful. Inconsistent explanations can erode trust, distort teacher judgments, and undermine the fairness and effectiveness of AI-driven educational support. Despite growing interest in improving LIME, existing optimisation efforts are fragmented, domain-specific, and rarely evaluated within educational contexts. The lack of a consolidated evidence base creates a notable research gap, limiting the development of enhanced XAI systems designed for distance learning environment. This motivates a systematic review to synthesise optimisation strategies for LIME and establish a foundation for designing more robust, reliable, and education-oriented explainability frameworks.

Aim of the Study

The aim of this study is to conduct a systematic literature review on existing optimisation approaches for the Local Interpretable Model-Agnostic Explanations (LIME) technique. The specific objectives are to:

1. Review and classify existing optimisation techniques applied to LIME across various domains.
2. Examine how LIME has been used in educational and distance learning contexts.
3. Evaluate the strengths, limitations, and performance outcomes of identified LIME optimisation methods.
4. Identify gaps in current optimisation approaches that limit their applicability in distance learning.

Research Questions

The research questions formulated for this study are:

1. What optimisation techniques have been applied to improve the performance and reliability of LIME?
2. How has LIME been used in distance learning or broader educational AI systems?
3. What are the reported strengths, weaknesses, and performance outcomes of existing LIME optimisation approaches?
4. What gaps exist in current research on LIME optimisation?

RESEARCH METHODOLOGY

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines to ensure a transparent, reproducible, and rigorous review of optimisation approaches for Local Interpretable Model-Agnostic Explanations (LIME) and their relevance to distance learning environments.

Research Design

A systematic literature review was conducted to consolidate and analyse research on LIME optimisation from 2016 to 2025. The PRISMA framework structured the identification, screening, eligibility assessment, and synthesis of relevant studies. This approach enabled a focused examination of methodological innovations addressing LIME’s main limitations, such as instability, low fidelity, and limited interpretability, within the context of distance learning.

Search Strategy

A comprehensive search was conducted across major scientific databases, including Scopus, IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, and Google Scholar (for supplementary retrieval). Search queries combined keywords related to explainable AI, LIME, optimisation techniques, and educational applications using Boolean operators to increase coverage while maintaining conceptual specificity. A representative search string was: (“LIME” OR “Local Interpretable Model-Agnostic Explanations”) AND (“optimisation” OR “optimization” OR “meta-heuristic” OR “Bayesian optimisation” OR “hybrid optimisation”) AND (“explainable artificial intelligence” OR “XAI”) AND (“education” OR “distance learning” OR “e-learning”).

Inclusion and Exclusion Criteria

Studies were included if they:

- i. focused on LIME or a LIME-derived variant;
- ii. proposed, implemented, or evaluated an optimisation technique;
- iii. were published between 2016–2025;
- iv. were written in English; and
- v. provided findings relevant to AI model interpretability, with preference for educational contexts.

Excluded studies were those that:

- i. addressed explainability techniques other than LIME without comparative analysis;
- ii. did not incorporate optimisation methods;
- iii. were non-peer-reviewed (e.g., opinion papers, theses, blog posts);
- iv. duplicated earlier publications; or
- v. focused solely on predictive modelling without interpretability components.

Study Selection Procedure

The review followed the four PRISMA stages: identification, screening, eligibility, and inclusion. All retrieved records were exported to a reference-management tool where duplicates were removed. Titles and abstracts were screened against inclusion criteria, after which full texts of potentially relevant studies were reviewed to confirm eligibility. The final set of studies was compiled after resolving any ambiguities through closer textual examination. A PRISMA flow diagram summarises the selection process and the number of studies retained at each stage.

Data Extraction

A structured extraction protocol was employed to ensure consistency across studies. Extracted data included:

- i. author(s) and publication year;
- ii. purpose of study;
- iii. machine-learning models used;
- iv. description of LIME optimisation technique;
- v. dataset characteristics;

- vi. evaluation metrics;
- vii. key findings; and
- viii. strengths and limitations.

This schema facilitated direct comparison of methodological features, optimisation strategies, and outcomes across diverse LIME variants.

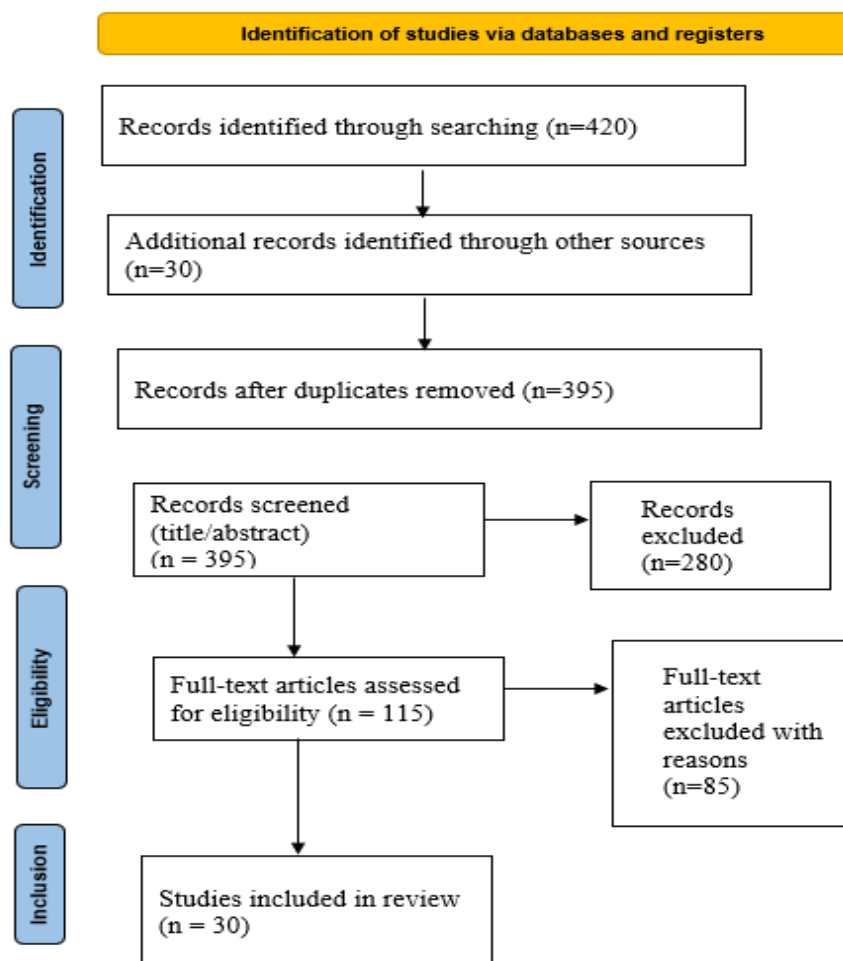


Figure 1: PRISMA 2020 flow diagram illustrating the study selection process

Data Analysis

A structured analytical framework was employed to extract, organise, and synthesise data from each eligible study. The framework captured analytic information, such as study context, machine learning model characteristics, optimisation techniques applied to LIME, dataset properties, evaluation metrics, and reported performance outcomes. Each study was further categorised by publication year, domain of application, optimisation strategy, and data modality (tabular, image, time series, multimodal). This categorisation enabled the visualisation of methodological and thematic patterns across the LIME optimisation landscape.

Data extraction and coding were conducted using Microsoft Excel to ensure consistency and traceability throughout the review process. An inductive–deductive analytical strategy

was adopted: deductively, the coding structure followed predefined domains aligned with the review questions (e.g., stability enhancements, perturbation control, surrogate model extensions); inductively, emerging patterns and methodological innovations were identified during full-text analysis. This dual approach ensured both theoretical alignment and flexibility to capture novel optimisation contributions.

Braun and Clarke’s (2006) six-phase thematic analysis guided the synthesis and interpretation of extracted data. The phases familiarisation, initial coding, theme development, theme review, theme refinement, and narrative reporting, were systematically applied to derive overarching thematic categories. Five major themes emerged: (1) perturbation and sampling optimisation, (2) fidelity and stability enhancement

strategies, (3) surrogate model advancements, (4) domain- and modality-specific adaptations, and (5) hybrid and automated optimisation frameworks. The thematic synthesis enabled the identification of recurrent methodological trends, unresolved challenges, and opportunities for advancing LIME within distance learning environments.

RESULTS

This systematic review identified and analysed 30 studies that proposed or evaluated optimisation techniques for the Local Interpretable Model-Agnostic Explanations

(LIME) method. The results are presented below, beginning with a descriptive summary of the literature landscape, followed by a thematic synthesis of the optimisation strategies.

Descriptive Overview of the Included Studies

The reviewed studies were published between 2018 and 2025, with a noticeable increase in publications from 2020 onward, indicating growing research interest in XAI and LIME optimisation. The applications spanned diverse domains, with a predominant focus on healthcare and computer vision.

Table 1: Distribution of Studies by Primary Domain and Data Modality

Primary Application Domain	Number of Studies	Primary Data Modality
Healthcare/Biomedical	8	Signals (ECG), Images, Tabular
Computer Vision	7	Images
General Tabular Data	6	Tabular
Signal & Time Series Processing	4	Time-series, Signals
Multimodal Data	2	Multimodal (Image+Text)
Other (Staffing, Gait, DOE)	3	Tabular

Table 1 reveals that research on LIME optimisation is concentrated in healthcare/biomedical (8 studies) and computer vision (7 studies), using signals, images, and tabular data. A smaller number of studies address time-series or multimodal data. This distribution shows a domain-specific focus, with no studies applying these optimised

methods to educational or distance learning datasets. This illustrates a major gap, as the techniques are being refined in fields like medicine and image analysis but remain untested for the behavioural, sequential, and multimodal data characteristic of learning environments.

Table 2: Categorisation of LIME Optimisation Techniques

Optimisation Strategy Category	Example Variants (from review)	Primary Goal	# of Studies
Sampling & Perturbation Enhancement	DLIME, GMM-LIME, ALIME, Sig-LIME, US-LIME, Guided-LIME	Improve stability & generate realistic data samples	11
Surrogate Model Advancement	QLIME, BMB-LIME, LIME-SUP, NormLIME, ILIME	Capture non-linearity & provide global explanations	8
Domain/Modality-Specific Adaptation	DSEG-LIME, LIMESegment, B-LIME, SurvLIME-Inf, DIME	Tailor explanations to specific data types (Image, TS, Survival)	7

Optimisation Strategy Category	Example Variants (from review)	Primary Goal	# of Studies
Theoretical/Modular Reformulation	S-LIME, bLIMEy, GLIME	Address foundational instability or offer flexibility	4
Parameter & Configuration Optimisation	CBR-LIME	Automate hyperparameter tuning	1

Table 2 classified the principal methodological approaches used to optimise LIME, highlighting their distinct objectives and prevalence in the literature. Sampling and Perturbation Enhancement is the most common strategy (11 studies), aiming to boost explanation stability and realism by improving how data samples are generated. Surrogate Model Advancement (8 studies) seeks to overcome LIME's linear limitations by using more complex models to capture non-linear patterns and provide global insights. A significant portion of work (7 studies) focuses on Domain/Modality-Specific Adaptation, customising LIME for particular data

types like images or time-series. In contrast, fewer studies pursue Theoretical/Modular Reformulation (4 studies) to fix core algorithmic issues, and only one study addresses Parameter Optimisation, underscoring a major gap in automated configuration for practical deployment.

Thematic Synthesis of Optimisation Outcomes

The analysis of the extracted data led to the identification of five major thematic categories, summarising the methodological innovations and their reported outcomes.

Table 3: Analysis of main findings by optimisation theme

Theme	Reported Improvements (vs. Standard LIME)	Limitations & Trade-offs
Perturbation & Sampling Optimisation	- Increased explanation stability (e.g., GMM-LIME). - Higher fidelity via realistic samples (e.g., ALIME, Sig-LIME). - More efficient sampling (e.g., US-LIME).	- Often domain-specific. - Can increase computational cost. - Potential loss of local detail for stability.
Surrogate Model Advancements	- Better approximation of non-linear boundaries (e.g., QLIME, BMB-LIME). - Provision of uncertainty estimates (BMB-LIME). - Coherent global/class-level explanations (e.g., NormLIME, ILIME).	- Increased model complexity. - Can reduce intuitive interpretability. - Higher computational demand.
Domain & Modality-Specific Adaptations	- Superior performance on target data type (e.g., DSEG-LIME for images, LIMESegment for time series). - Explanations that respect data structure (e.g., temporal dependencies in Sig-LIME).	- Lack of generalisability to other domains. - Require domain expertise to implement.

Theme	Reported Improvements (vs. Standard LIME)	Limitations & Trade-offs
Hybrid & Automated Optimisation Frameworks	- Automated hyperparameter selection (CBR-LIME). - Structured sampling improving global coherence (Guided-LIME).	- Frameworks are few and often conceptual. - Can add implementation overhead.
Enhanced Interpretability & Visualisation	- Improved human alignment of explanations (e.g., DSEG-LIME). - Tools for deeper analysis of LIME's behaviour (ExpLIMEable).	- Often an auxiliary contribution, not a core optimisation of LIME's algorithm. - User studies are rare.

Table 3 presents the benefits and inherent trade-offs of five major LIME optimisation themes, revealing a pattern of targeted gains coupled with new compromises. Perturbation and Sampling Optimisation improves stability and realism but often at the cost of computational efficiency and generalizability. Surrogate Model Advancements deliver more accurate and globally coherent explanations, yet they introduce complexity that can undermine the very interpretability LIME aims to provide. While Domain-Specific Adaptations yield superior performance for their target data type, they sacrifice broader applicability. The emerging themes of Hybrid Frameworks and Enhanced

Visualisation address automation and usability but remain underrepresented and insufficiently validated with end-users. The table collectively shows that current optimizations tend to solve one problem while introducing another, highlighting the absence of a holistic approach that balances all critical objectives for real-world deployment.

Performance Metrics and Evaluation Trends

The evaluation of optimised LIME variants relied on a mix of quantitative metrics and qualitative analysis. A clear trend was the focus on algorithmic performance rather than human-centric utility.

Table 4: Commonly used evaluation metrics across studies

Evaluation Dimension	Specific Metrics Reported	Frequency of Use
Explanation Stability	Variance across runs, Jaccard similarity of top features	High
Explanation Fidelity	Local accuracy of surrogate, Euclidean distance to reference explanation, MSE	High
Computational Efficiency	Execution time, sample count required	Medium
Interpretability Quality	Qualitative analysis, case studies, (rarely) user comprehension tests	Low
Model Performance	Accuracy, F1-score of the underlying black-box model	Low (often assumed stable)

Research Gap

A critical finding of this review is the disparity between the technical advancements in LIME optimisation and the requirements for their application in educational settings.

Table 5: Summary of research gaps

Gap Category	Specific Gap Identified	Evidence from Review
Domain Application Gap	No study applied or evaluated optimised LIME in distance learning or educational AI.	All studies used biomedical, vision, or general benchmarking datasets.
Human-Centred Evaluation Gap	Severe lack of empirical studies with educators or students to assess pedagogical utility.	Evaluation was primarily algorithmic; only 2 studies mentioned user comprehension indirectly.
Multimodal Data Gap	Limited frameworks for explaining models trained on integrated multimodal learning data.	Only DIME addressed multimodal explanations holistically.
Multi-Objective Optimisation Gap	Absence of frameworks balancing stability, fidelity, interpretability, fairness, and speed.	Studies typically optimized for one or two objectives in isolation.
Scalability & Real-time Gap	High computational cost of advanced variants (e.g., ALIME, BMB-LIME) precludes large-scale, real-time use.	Many methods introduced significant overhead from autoencoders, Bayesian inference, or generative models.

Table 5 shows that while significant methodological progress has been made in stabilising LIME, improving its fidelity, and adapting it to various data types, this research trajectory has not yet converged with the practical needs of AI-driven distance learning systems.

DISCUSSION

The systematic review of LIME optimisation approaches shows a trend towards addressing the three core limitations of the original LIME algorithm: instability due to random perturbation sampling, poor fidelity in complex and nonlinear decision boundaries, and limited interpretability at both local and global scales. Literature shows that researchers have applied series of optimisation techniques to improve reliability, accuracy, and human interpretability of explanations, thus, reflecting both methodological innovation and the practical needs of applied machine learning. The review categorised LIME optimisation methods into six approaches and they are deterministic sampling, data-aware perturbations, surrogate model enhancement, active and uncertainty-driven sampling, theoretical reformulations, and automated hyperparameter optimisation. The breadth of these techniques reflects an increasing awareness of LIME’s methodological limitations, and a concerted effort by researchers to address them in domain-specific contexts.

The review show that a good number of studies focus on deterministic sampling and neighbourhood construction as a

means of reducing LIME’s instability. DLIME, introduced by Zafar and Khan (2019), replaced LIME’s random perturbations with clustering-based neighbourhood selection to generate reproducible explanations. This approach improved run-to-run stability, though sometimes at the expense of local fidelity due to rigid neighbourhood boundaries. AL-DLIME (Holm & Macedo, 2023) extended this principle by integrating active learning into deterministic sampling, demonstrating that stability and fidelity can be partially reconciled when informative samples are prioritised. While these methods provide valuable insights into neighbourhood design, most retain fixed sampling parameters chosen heuristically, indicating a need for adaptive, optimised configuration in broader application domains such as education.

Another major optimisation approach involves data-aware perturbation strategies, which seek to generate realistic samples that lie closer to the true data manifold. Autoencoder-based approaches such as ALIME (Shankaranarayana & Runje, 2019) and generative-model-based MeLIME (Botari et al., 2020) show significant improvements in stability and fidelity by constraining perturbations to plausible data variations. Domain-specific extensions, such as Sig-LIME and B-LIME for biomedical signals (Abdullah et al., 2023; Abdullah et al., 2024), incorporate temporal segmentation and structured noise injection to preserve sequence properties. Together, these studies highlight the importance of

modality-sensitive perturbations, but they also introduce considerable computational overhead, raising concerns for real-time or large-scale environments like learning analytics systems.

A third category of improvements targets surrogate model expressivity and uncertainty quantification. BMB-LIME (Hung & Lee, 2024) replaced the conventional linear surrogate with a Bayesian mixture model to better capture local non-linearities and provide uncertainty estimates alongside feature attributions. Such probabilistic surrogates expand the explanatory value of LIME but risk reducing interpretability and increasing computational cost. These findings reinforce the general observation that surrogate choice involves trade-offs between fidelity, stability, interpretability, and efficiency, trade-offs that have not been adequately explored in educational settings.

Studies adopting active or uncertainty-driven sampling show that LIME’s performance depends critically on the informativeness of its perturbations. US-LIME (Saadatfar et al., 2024) demonstrated that targeting samples near the decision boundary yields higher-fidelity surrogates with smaller sample budgets. AL-DLIME similarly integrates active sample selection into deterministic neighbourhoods. These results suggest that sample selection strategies can substantially improve LIME’s efficiency and reliability, though few implementations explicitly optimise sampling policies across datasets or application contexts.

A smaller but noteworthy body of literature presents theoretically motivated reformulations of LIME. GLIME (Tan et al., 2023) corrects LIME’s sampling, weighting mismatch by embedding kernel weights within the perturbation distribution, yielding faster convergence and more stable explanations. Structured global surrogates, such as LIMETree (Sokol & Flach, 2025), reduce variance by replacing per-instance surrogates with deterministic decision-tree structures that generate consistent multi-class explanations. These theoretically grounded approaches contribute important insights into LIME’s statistical properties, but typically involve hyperparameters that must still be manually configured, highlighting the need for automated and data-driven optimisation.

A final, emerging stream of research concerns automated configuration and hyperparameter optimisation. OptiLIME (Visani, 2020) demonstrates that tuning parameters such as sample size, kernel width, and perturbation strategy can dramatically improve explanation fidelity and stability. Broader optimisation literature shows that Bayesian optimisation and hybrid metaheuristics (e.g., BO combined with evolutionary algorithms) are well-suited to black-box, mixed-type optimisation problems (Snoek et al., 2012; Salgotra et al., 2024). These findings strongly support the claim of LIME’s configuration as an optimisation problem, especially in domains requiring multiple competing

objectives such as interpretability, fidelity, fairness, computational efficiency, and human usability.

Across these optimisation strategies, several cross-cutting gaps emerge. Many methods were designed for specific domains (e.g., biomedical signals, images), with limited transferability to tabular or behavioural data typical of distance learning environments. Most approaches optimise a single objective, usually fidelity or stability, while neglecting multi-objective trade-offs crucial for educational applications where transparency, interpretability, fairness, and usability must be balanced. Moreover, few studies incorporate human-centred evaluation involving real users (such as teachers or students), leaving open questions about practical usefulness and cognitive interpretability. These gaps highlight the need for an improved LIME framework that is optimised for the needs and constraints of distance learning systems.

In answering the third research question, the studies reviewed showed that LIME optimisation methods result in improvements in stability, local fidelity, and interpretability. Deterministic sampling techniques such as DLIME reduce run-to-run variability by replacing random perturbations with structured neighbourhoods, resulting in more reproducible explanations (Zafar & Khan, 2019). Data-aware and domain-aligned perturbation strategies, such as autoencoder-based latent perturbations in ALIME, produce more realistic samples that better preserve underlying data structures, thereby enhancing surrogate fidelity (Shankaranarayana & Runje, 2019). Other optimisation efforts extend LIME’s capacity to handle complex or nonlinear decision boundaries by introducing more expressive surrogate models or uncertainty-aware explanations, as seen in Bayesian or mixture-based approaches. Additionally, global and aggregated variants (e.g., GLIME, LIMETree) improve interpretability at group or class level, providing more holistic insights than traditional instance-level explanations (Tan, Tian, & Li, 2023; Sokol & Flach, 2020). These enhancements prove that targeted optimisation can significantly improve LIME’s reliability, plausibility of perturbations, and explanatory value across diverse data types.

Despite these advancements, several limitations remain evident across the reviewed optimisation strategies. Many enhanced variants introduce higher computational cost, additional training requirements, or greater implementation complexity, particularly when incorporating autoencoders, generative models, or probabilistic surrogates. Domain-specific methods, such as those designed for signal or temporal data, lack generalisability and require careful tuning or extensive domain knowledge to be effectively deployed in new contexts. Global or aggregated explanation approaches, while useful for interpretability, tend to sacrifice fine-grained local fidelity. Moreover, theoretical reformulations such as GLIME still leave practitioners with hyperparameters that require expert adjustment, limiting ease of adoption. Across

methods, there is a persistent absence of multi-objective optimisation frameworks that jointly consider fidelity, stability, interpretability, robustness, and computational efficiency. These weaknesses emphasise the need for more unified, domain-adaptive, and systematically optimised LIME frameworks for sensitive contexts such as educational decision-making.

A prominent gap in the literature is the limited focus on educational and distance-learning data, despite the growing dependence on AI-driven predictive models in digital education. Most LIME optimisation studies are conducted on biomedical signals, computer vision, tabular diagnostics, or general benchmarking datasets (e.g., UCI repositories), with almost no work validating enhanced LIME variants on learning analytics or student performance prediction. This limits transferability to distance-learning environments, where behavioural, cognitive, and contextual variables differ significantly from conventional ML datasets and require domain-aware interpretability approaches.

Another unresolved challenge concerns multimodal and sequential data, which are increasingly common in modern online learning systems (e.g., clickstream logs, forum text, instructional videos, temporal engagement traces). Although techniques such as Sig-LIME and B-LIME address temporal or signal-based perturbations, there is little work extending LIME to handle multimodal educational contexts holistically. Current methods typically optimise perturbations or surrogate behaviour for a single modality at a time, making them insufficient for integrated, multimodal learning environments like MOOCs or LMS platforms.

A third gap involves the persistent trade-off between global interpretability and local fidelity. While aggregated or global techniques (e.g., ILIME, LIMETree) improve high-level interpretability and reduce explanation variance, they commonly sacrifice fine-grained, instance-level accuracy. Conversely, methods focused on local fidelity, such as uncertainty-driven sampling or nonlinear surrogates, rarely provide coherent global explanations. This trade-off remains unresolved, yet it is particularly crucial in education where both student-specific (local) and class-wide (global) insights are necessary for effective intervention.

In terms of practical deployment, scalability and computational cost remain major barriers. Approaches that rely on generative models, autoencoders, or Bayesian surrogates (e.g., ALIME, MeLIME, BMB-LIME) offer measurable improvements but substantially increase computational overhead, making them unsuitable for real-time feedback systems used in large-scale distance learning. This bottleneck is amplified in environments requiring continuous monitoring of thousands of learners.

Another notable gap is the absence of human-centric evaluation, which significantly limits the ecological validity of LIME optimisation research. A few studies empirically test

whether enhanced explanations improve teacher comprehension, support decision-making, or reduce misinterpretation. This is problematic because the primary purpose of XAI in education is not merely algorithmic improvement but human interpretability and pedagogical utility.

CONCLUSION

This systematic review synthesised and critically evaluated the diverse optimisation strategies developed to address the known limitations of the Local Interpretable Model-Agnostic Explanations (LIME) technique. The findings demonstrate that while LIME remains one of the most widely used and versatile XAI methods, its native formulation exhibits substantial weaknesses, such as instability, low fidelity in nonlinear regions, unrealistic perturbations, limited global coherence, and sensitivity to hyperparameter choices. In response, an extensive body of research has proposed methodological extensions to improve reliability, fidelity, interpretability, and domain alignment. These optimisation strategies span deterministic sampling, data-aware perturbations, surrogate model enhancements, uncertainty-driven sampling, theoretical reformulations, and automated hyperparameter optimisation. Collectively, these developments highlight meaningful progress towards a more robust and expressive LIME framework.

Despite these advances, several critical gaps persist. Most optimisation methods remain highly domain-specific, with limited evidence of generalisability to the behavioural, multimodal, and sequential data common in distance learning environments. Existing techniques often optimise a single objective of fidelity or stability, without addressing the multi-objective trade-offs required for educational deployment, such as balancing transparency, interpretability, fairness, computational cost, and real-time usability. Furthermore, there is limited research evaluating the human-centred impact of optimised LIME explanations, leaving open questions about how teachers, students, and instructional designers interpret and act upon model explanations in real educational settings. These limitations underscore the need for XAI solutions that are both technically rigorous and pedagogically grounded.

The insights from this review provide a clear foundation for the development of an enhanced, education-oriented XAI system. The evidence strongly supports the use of optimisation-driven approaches, that combine principled sampling design, non-linear surrogate flexibility, and automated configuration through Bayesian or hybrid metaheuristics. Such an approach is well positioned to overcome the instability and low fidelity associated with standard LIME, while addressing the unique challenges of distance learning data. Future systems can provide explanations that are not only more reliable and accurate but

also more meaningful to educational stakeholders, by integrating optimisation techniques with human-centred design.

REFERENCES

1. Abdullah, T. A. A., Zahid, M. S. M., Turki, A. F., Ali, W., Jiman, A. A., Abdulaal, M. J., ... & Attar, E. T. (2024). Sig-lime: A signal-based enhancement of lime explanation technique. *IEEE access*, 12, 52641-52658.
2. Abdullah, T. A., Zahid, M. S. M., Ali, W., & Hassan, S. U. (2023). B-LIME: An improvement of LIME for interpretable deep learning classification of cardiac arrhythmia from ECG signals. *Processes*, 11(2), 595.
3. Ahern, I., Noack, A., Guzman-Nateras, L., Dou, D., Li, B., & Huan, J. (2019). NormLime: A new feature importance metric for explaining deep neural networks. *arXiv preprint arXiv:1909.04200*.
4. An, J., Zhang, Y., & Joe, I. (2023). Specific-input lime explanations for tabular data based on deep learning models. *Applied Sciences*, 13(15), 8782.
5. Angiulli, F., De Luca, F., Fassetti, F., & Nisticò, S. (2025). LLiMe: enhancing text classifier explanations with large language models. *Machine Learning*, 114(12), 271.
6. Bacevicius, M., Paulauskaite-Taraseviciene, A., Zokaityte, G., Kersys, L., & Moleikaityte, A. (2025). Comparative analysis of perturbation techniques in LIME for intrusion detection enhancement. *Machine Learning and Knowledge Extraction*, 7(1), 21.
7. Bora, R. P., Terhörst, P., Veldhuis, R., Ramachandra, R., & Raja, K. (2024). Slice: Stabilized lime for consistent explanations for image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10988-10996).
8. Botari, T., Hvilshøj, F., Izbicki, R., & de Carvalho, A. C. (2020). MeLIME: Meaningful local explanation for machine learning models. *arXiv preprint arXiv:2009.05818*.
9. Bramhall, S., Horn, H., Tieu, M., & Lohia, N. (2020). Qlime-a quadratic local interpretable model-agnostic explanation approach. *SMU Data Science Review*, 3(1), 4.
10. Dehghani, Z., Aslansefat, K., Khan, A., & Akram, M. N. (2025). Explainability of Large Language Models using SMILE: Statistical Model-agnostic Interpretability with Local Explanations. *arXiv preprint arXiv:2505.21657*.
11. Dikopoulou, Z., Moustakidis, S., & Karlsson, P. (2021). GLIME: A new graphical methodology for interpretable model-agnostic explanations. *arXiv preprint arXiv:2107.09927*.
12. Duong, H., Hoang, L., & Le, B. (2023, October). Controlling lime kernel width to achieve comprehensible explanations on tabular data. In *International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making* (pp. 153-164). Cham: Springer Nature Switzerland.
13. ElShawi, R., Sherif, Y., Al-Mallah, M., & Sakr, S. (2019, August). Ilime: Local and global interpretable model-agnostic explainer of black-box decision. In *European Conference on Advances in Databases and Information Systems* (pp. 53-68). Cham: Springer International Publishing.
14. Garcia, W., Chen, P. Y., Clouse, H. S., Jha, S., & Butler, K. R. (2023, February). Less is more: Dimension reduction finds on-manifold adversarial examples in hard-label attacks. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)* (pp. 254-270). IEEE.
15. Gaudel, R., Galárraga, L., Delaunay, J., Rozé, L., & Bhargava, V. (2022, April). s-LIME: Reconciling locality and fidelity in linear explanations. In *International Symposium on Intelligent Data Analysis* (pp. 102-114). Cham: Springer International Publishing.
16. Gonzalez, A. V., & Søgaard, A. (2020). The reverse turing test for evaluating interpretability methods on unknown tasks. In *NeurIPS 2020 Workshop on Human And Model in the Loop Evaluation and Training Strategies*.
17. Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., ... & Hussain, A. (2024). Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, 16(1), 45-74.
18. Holm, S., & Macedo, L. (2023, July). The accuracy and faithfulness of AL-DLIME-active learning-based deterministic local interpretable model-agnostic explanations: A comparison with LIME and DLIME in medicine. In *World Conference on Explainable Artificial Intelligence* (pp. 582-605). Cham: Springer Nature Switzerland.
19. Hryniewska, W., Grudzień, A., & Biecek, P. (2024). LIMEcraft: handcrafted superpixel selection and inspection for Visual eXplanations. *Machine Learning*, 113(5), 3143-3160.
20. Hu, L., Chen, J., Nair, V. N., & Sudjianto, A. (2018). Locally interpretable models and effects based on supervised partitioning (LIME-SUP). *arXiv preprint arXiv:1806.00663*.

21. Hung, Y. H., & Lee, C. Y. (2024). BMB-LIME: LIME with modeling local nonlinearity and uncertainty in explainability. *Knowledge-Based Systems*, 294, 111732.
22. Knab, P., Marton, S., & Bartelt, C. (2024). Dseg-lime—improving image explanation by hierarchical data-driven segmentation. *arXiv preprint arXiv:2403.07733*.
23. Knab, P., Marton, S., Schlegel, U., & Bartelt, C. (2025, July). Which lime should i trust? concepts, challenges, and solutions. In *World Conference on Explainable Artificial Intelligence* (pp. 28-52). Cham: Springer Nature Switzerland.
24. Kovalev, M. S., Utkin, L. V., & Kasimov, E. M. (2020). SurvLIME: A method for explaining machine learning survival models. *Knowledge-Based Systems*, 203, 106164.
25. Laguna, S., Heidenreich, J. N., Sun, J., Cetin, N., Al-Hazwani, I., Schlegel, U., ... & El-Assady, M. (2023, October). ExpLIMEable: A visual analytics approach for exploring LIME. In *2023 Workshop on Visual Analytics in Healthcare (VAHC)* (pp. 27-33). IEEE.
26. Li, X., Xiong, H., Li, X., Zhang, X., Liu, J., Jiang, H., ... & Dou, D. (2024, February). G-LIME: statistical learning for local interpretations of deep neural networks using global priors (abstract reprint). In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence* (pp. 22705-22705).
27. Lyu, Y., Liang, P. P., Deng, Z., Salakhutdinov, R., & Morency, L. P. (2022, July). Dime: Fine-grained interpretations of multimodal models via disentangled local explanations. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 455-467).
28. Mulwa, M. M., Mwangi, R. W., & Mindila, A. (2024). GMM-LIME explainable machine learning model for interpreting sensor-based human gait. *Engineering Reports*, 6(10), e12864.
29. Nnadi, L. C., Watanobe, Y., Rahman, M. M., & John-Otumu, A. M. (2024). Prediction of students' adaptability using explainable AI in educational machine learning models. *Applied Sciences*, 14(12), 5141.
30. Rabold, J., Deininger, H., Siebers, M., & Schmid, U. (2019, September). Enriching visual with verbal explanations for relational concepts—combining LIME with Aleph. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 180-192). Cham: Springer International Publishing.
31. Rahimiaghdam, S., & Alemdar, H. (2025). MindfulLIME: A Stable Solution for Explanations of Machine Learning Models with Enhanced Localization Precision--A Medical Image Case Study. *arXiv preprint arXiv:2503.20758*.
32. Raza, S., Fatima, I., Arif, S., Sharif, M., Jalal, M. S., & Muhammad, Z. (2024). The future of learning: building trust and transparency in AI education. *Journal of management practices, humanities and social sciences.*, 8(3), 62-74.
33. Recio-García, J. A., Díaz-Agudo, B., & Pino-Castilla, V. (2020, June). CBR-LIME: a case-based reasoning approach to provide specific local interpretable model-agnostic explanations. In *International Conference on Case-Based Reasoning* (pp. 179-194). Cham: Springer International Publishing.
34. Saadatfar, H., Kiani-Zadegan, Z., & Ghahremani-Nezhad, B. (2024). US-LIME: Increasing fidelity in LIME using uncertainty sampling on tabular data. *Neurocomputing*, 597, 127969.
35. Saadatfar, H., Kiani-Zadegan, Z., & Ghahremani-Nezhad, B. (2024). US-LIME: Increasing fidelity in LIME using uncertainty sampling on tabular data. *Neurocomputing*, 597, 127969.
36. Salgotra, R., & Gandomi, A. H. (2024). A novel multi-hybrid differential evolution algorithm for optimization of frame structures. *Scientific Reports*, 14(1), 4877.
37. Salgotra, R., Lamba, A. K., Talwar, D., Gulati, D., & Gandomi, A. H. (2024). A hybrid swarm intelligence algorithm for region-based image fusion. *Scientific Reports*, 14(1), 13723.
38. Sangroya, A., Rastogi, M., Anantaram, C., & Vig, L. (2020, October). Guided-LIME: Structured Sampling based Hybrid Approach towards Explaining Blackbox Machine Learning Models. In *CIKM (Workshops)* (Vol. 17).
39. Shankaranarayana, S. M., & Runje, D. (2019, October). ALIME: Autoencoder based approach for local interpretability. In *International conference on intelligent data engineering and automated learning* (pp. 454-463). Cham: Springer International Publishing.
40. Sivill, T., & Flach, P. (2022, May). Limesegment: Meaningful, realistic time series explanations. In *International Conference on Artificial Intelligence and Statistics* (pp. 3418-3433). PMLR.
41. Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning

- algorithms. *Advances in neural information processing systems*, 25.
42. Sokol, K., & Flach, P. (2025). LIMETree: Consistent and faithful surrogate explanations of multiple classes. *Electronics*, 14(5), 929.
 43. Sokol, K., Hepburn, A., Santos-Rodriguez, R., & Flach, P. (2019). bLIMEy: surrogate prediction explanations beyond LIME. *arXiv preprint arXiv:1910.13016*.
 44. Stadler, A., Müller, W. G., & Harman, R. (2025). Green LIME: Improving AI Explainability through Design of Experiments. *arXiv preprint arXiv:2502.12753*.
 45. Tan, L., Huang, C., & Yao, X. (2025). A Faithful and Stable Explanation Method for Image Classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*.
 46. Tan, Z., Tian, Y., & Li, J. (2023). GLIME: General, stable and local LIME explanation. *Advances in Neural Information Processing Systems*, 36, 36250-36277.
 47. Utkin, L. V., Kovalev, M. S., & Kasimov, E. M. (2020, September). An explanation method for black-box machine learning survival models using the Chebyshev distance. In *Conference on Artificial Intelligence and Natural Language* (pp. 62-74). Cham: Springer International Publishing.
 48. Visani, G., Bagli, E., & Chesani, F. (2020). Optilime: Optimized lime explanations for diagnostic computer algorithms. *arXiv preprint arXiv:2006.05714*.
 49. Wang, G., Wang, H., Lv, S., Kang, X., & Cui, Q. A. (2025). Explainable artificial intelligence for manufacturing process modelling and optimisation: an integrated ANN and LIME approach. *International Journal of Production Research*, 1-22.
 50. Yasui, D., & Sato, H. (2025). Improving local fidelity and interpretability of LIME by replacing only the sampling process with CVAE. *IEEE Access*.
 51. Zafar, M. R., & Khan, N. M. (2019). DLIME: A deterministic local interpretable model-agnostic explanations approach for computer-aided diagnosis systems. *arXiv preprint arXiv:1906.10263*.

“Foundations for an Enhanced XAI System: A Systematic Literature Review on Optimising LIME for Distance Learning Environments”

Author(s) & Year	Purpose of Study	ML Model Used	LIME Optimisation Technique	Dataset Characteristics	Evaluation Metrics	Key Findings	Strengths	Limitations
TAA Abdullah et al., 2023 – B-LIME	Improve LIME explanations for ECG signals	Hybrid Deep Learning Model	Bootstrap-LIME: heartbeat segmentation + bootstrapping to improve temporal dependencies	ECG signal data	Comparison with standard LIME	More meaningful and credible explanations; stable and locally faithful	Addresses temporal dependencies; improved stability	Limited to ECG signals; not tested across other domains
TAA Abdullah et al., 2024 – Sig-LIME	Address LIME instability and local fidelity for signal data	Random Forest + heatmaps	Signal-based data generation preserving temporal dependence	ECG signals	ANOVA, Euclidean distance for fidelity	Improved stability and local fidelity over LIME	Tailored for signal data; strong empirical results	Domain-specific (ECG); may not generalize to other modalities
Guodong Wang et al., 2025	Integrate XAI into DOE to improve model interpretability	ANN	LIME to interpret ANN for local explanations	DOE experimental datasets	Model accuracy; local interpretability	LIME enhances understanding of important experimental factors	Combines high precision modeling with local explainability	Focused on DOE data; limited generalizability
I Ahern et al., 2019 – NormLIME	Aggregate local LIME explanations into global/class-specific interpretations	Deep Neural Networks	Normalized aggregation of local explanations	Tabular and image data	Human user study, numerical experiments	NormLIME improves class-specific interpretation and feature recognition	Bridges local and global interpretability	Focused on aggregation; may not improve local fidelity
J An et al., 2023 – Specific-input LIME	Interpret deep learning models for tabular data	Deep Learning	Uses feature importance + PDPs for stability; balances local and global interpretations	Tabular data	Interpretation stability, PDP analysis	Improves stability and extends beyond local interpretations	Balances global/local explanations	Limited to tabular data
RP Bora et al., 2024 – SLICE	Stabilize LIME for image	Image models	Feature selection + adaptive perturbation	Image datasets	Explanation consistency, fidelity	Significantly improved consistency	Directly addresses instability issue	Limited to image classification

“Foundations for an Enhanced XAI System: A Systematic Literature Review on Optimising LIME for Distance Learning Environments”

	classification		n for consistent explanations			y and fidelity		
S Bramhall et al., 2020 – QLIME	Handle non-linear local decision boundaries	Black-box model for staffing data	Quadratic relationships instead of linear for surrogate model	Tabular staffing data	MSE improvement	Improves feature interpretation for linear assumptions	Handles non-linear local boundaries	Tested on single domain; small-scale validation
T Botari et al., 2020 – MeLIME	Improve local explanations across different data types	Various (tabular, image, text)	Flexible perturbation sampling; modified local interpretable models; counterfactual examples	Tabular, image, text	Qualitative & comparative analysis	Produces more meaningful explanations	Generalizable across data types; supports counterfactuals	Less quantitative evaluation; stability metrics limited
Z Dikopoulou et al., 2021 – GLIME	Graphical interpretability for feature relationships	Classification models	LIME + GLASSO; graphical representation of direct/indirect impacts	BIOPSY & OAI datasets	Feature importance, robustness, consistency	Improved stability (76%-96%) over LIME	Reveals conditional feature relationships; robust	Limited domains; more complex workflow
R ElShawi et al., 2019 – ILIME	Local and global explanations of black-box models	Supervised ML	Uses most influential instances for explanation; aggregates local explanations globally	Various datasets	Explanation quality, mimic accuracy	Outperforms LIME in mimicking black-box models	Global + local explanations; improves trust	Complexity; implementation heavier than LIME
R Gaudel et al., 2022 – S-LIME	Reconcile locality and fidelity in LIME	Not specified	Adjust surrogate model to avoid degeneration in overly local neighborhoods	Not specified	Conceptual & illustrative	Highlights pitfalls of LIME locality assumption	Theoretical insight into LIME limitations	Limited empirical evaluation
L Hu et al., 2018 – LIME-SUP	Interpret supervised ML models with trees	Gradient Boosting, RF, NN	Fits trees to fitted response and	Simulated and real datasets	Comparative evaluation	Provides interpretable local explanation	Clear methodology; works	Complexity; may require careful tree tuning

“Foundations for an Enhanced XAI System: A Systematic Literature Review on Optimising LIME for Distance Learning Environments”

			derivatives ; local-global explainability		with KLIME	ns using trees	with large datasets	
YH Hung & CY Lee, 2024 – BMB-LIME	Model local nonlinearity & uncertainty in LIME	Not specified	Bootstrap aggregating MARS + Bayesian framework to model uncertainty	Simulated & real datasets	Local fidelity, stability, Jaccard similarity	High consistency (89.1%+) & improved local fidelity	Handles non-linear boundaries ; quantifies uncertainty	Computationally heavier; complex implementation
P Knab et al., 2024 – DSEG-LIME	Improve image explanation via hierarchical segmentation	Pre-trained image models	Data-driven + hierarchical segmentation	ImageNet class images	XAI metrics, human alignment	Better alignment with human-recognized concepts	Enhances interpretability for images	Focused on images; hierarchical step adds complexity
S Laguna et al., 2023 – ExpLIMEable	Visual analytics for LIME	Medical image models	Visual analytics app to explore LIME perturbation effects	Medical image datasets	User comprehension metrics	Helps understand LIME robustness issues	Enhances human understanding	Tool-focused; not direct model optimisation
LV Utkin et al., 2020 – SurvLIME-Inf	Explain survival ML models	Cox proportional hazards	-norm for distance; linear programming approach	Synthetic & real datasets	Feature importance	Outperforms SurvLIME on small datasets	Better suited for small datasets; simple LP	Limited to survival models
Y Lyu et al., 2022 – DIME	Fine-grained interpretations for multimodal models	Multimodal models	Disentangles unimodal & multimodal contributions	Synthetic & real multimodal datasets	Qualitative & quantitative analysis	Accurate disentangled explanations; supports debugging	Multimodal applicability; fine-grained	Complex; multiple modalities may complicate deployment
MM Mulwa et al., 2024 – GMM-LIME	Stabilize LIME for human gait data	NN (GaitRec), RF (HugaDB)	Gaussian Mixture Model sampling to avoid random perturbation	Gait datasets	Accuracy, stability, Jaccard similarity	Produces more accurate, reliable, stable explanations	High stability & fidelity	Limited to gait datasets; domain-specific
A Sangroya et al., 2020 –	Hybrid approach for	Various ML models	Combines LIME + Formal	IRIS, Heart Disease,	Prominent features, proxy	Hybrid method improves	Efficient for	Complexity; requires

“Foundations for an Enhanced XAI System: A Systematic Literature Review on Optimising LIME for Distance Learning Environments”

Guided-LIME	structured sampling		Concept Analysis (FCA) for structured sampling	Adult Earnings	model closeness, calibration error	global explanation and interpretability	structured datasets	FCA knowledge
H Saadatfar et al., 2024 – US-LIME	Improve fidelity via uncertainty sampling	Not specified	Sample near decision boundary; linear surrogate	Tabular datasets	Fidelity, stability, execution time	Improves fidelity and explanation constancy; faster than variants	Effective for tabular data; kernel-width independent	Limited to tabular; not tested for images/signals
JA Recio-García et al., 2020 – CBR-LIME	Use case-based reasoning for specific LIME parameter tuning	Image classifier	Case-based approach to reuse parameter configurations for similar inputs	Image datasets	Qualitative evaluation	Produces tailored explanations; reflects human perception	Parameter reuse; adaptable	Domain-specific; relies on case base
J Rabold et al., 2019 – LIME + Aleph	Enrich visual explanations with relational/logic rules	Image classifiers	LIME + Inductive Logic Programming (Aleph)	Blocksworld images	Qualitative analysis	Captures relational concepts; richer explanations	Combines visual and logical explanations	Domain-specific; not generalizable
SM Shankaranarayana & D Runje, 2019 – ALIME	Autoencoder-based weighting for local interpretability	Various ML models	Autoencoder for weighting local model	Multiple datasets	Stability, local fidelity	Improved stability & fidelity	Autoencoder weighting reduces perturbation instability	More complex than LIME; requires autoencoder training
T Sivill & P Flach, 2022 – LIMESegment	Time series adaptations of LIME	Time series classifiers	Meaningful representation + realistic perturbation; local neighbourhood definition	Time series datasets	Classification performance	Outperforms existing LIME adaptations	Handles time series; realistic perturbations	Limited to time series data
K Sokol et al., 2019 – bLIMEy	Modular framework for custom surrogate explainers	Any black-box model	Modular decomposition of LIME components	Tabular, text, image	Qualitative assessment	Supports flexible design of surrogate explainers; LIME as a special case	Highly generalizable; modular; conceptual clarity	Largely conceptual; limited quantitative validation