

An Osprey-Optimised LIME Framework for Explainable Student Performance Prediction in Distance Learning

Aminat Obakhume Abdulsalaam¹; Oyedepo Mayowa Oyediran²; Olufemi S. Ojo³ & Funmi W. Ipeyeda⁴

Department of Computer Science, Ajayi Crowther University, Oyo, Nigeria.

*Corresponding Author: Aminat Obakhume Abdulsalaam

DOI: <https://doi.org/10.5281/zenodo.23633725>

Article History	Abstract
Original Research Article	<i>Explainable Artificial Intelligence (XAI) is increasingly applied to student-performance prediction to improve the transparency and interpretability of machine-learning decisions. Local Interpretable Model-Agnostic Explanations (LIME) provides model-agnostic local explanations; however, its effectiveness depends on the configuration of parameters governing neighbourhood construction, perturbation sampling and feature selection. Conventional LIME relies largely on default or manually selected configurations, which may produce explanations with inconsistent fidelity, interpretability and computational efficiency when applied to educational data. This study developed an Osprey-Optimised LIME (OOA-LIME) framework that employs the Osprey Optimisation Algorithm (OOA) to automatically search for effective LIME configurations. The framework was evaluated using 10,000 anonymised student records from a distance-learning programme. Standard LIME and OOA-LIME were evaluated across dataset sizes of 2,000–10,000 records using fidelity, interpretability, comprehensibility, trustworthiness, coverage and explanation-generation time. OOA-LIME consistently outperformed Standard LIME across all five explanation-quality measures. Fidelity increased from 0.850–0.890 for OOA-LIME compared with 0.780–0.812 for Standard LIME, while interpretability increased from 0.800–0.836 compared with 0.710–0.738. Comprehensibility, trustworthiness and coverage similarly improved. Explanation-generation time decreased from 12.01–12.98 seconds for Standard LIME to 7.58–8.09 seconds for OOA-LIME. The findings indicate that metaheuristic optimisation can improve both the explanatory quality and computational efficiency of LIME for educational prediction. OOA-LIME provides an optimisation-based approach for improving LIME configurations and offers a practical framework for more effective explainable student-performance prediction in distance-learning environments.</i> Keywords: Explainable Artificial Intelligence (XAI), Student Performance Prediction, Local Interpretable Model-Agnostic Explanations (LIME), Osprey Optimisation Algorithm (OOA), Model Interpretability, Educational Data Mining.
Received: 26-08-2026	
Accepted: 30-09-2026	
Published: 25-10-2026	
Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited. Citation: Aminat Obakhume Abdulsalaam, Oyedepo Mayowa Oyediran, Olufemi S. Ojo, & Funmi W. Ipeyeda. (2026). An osprey-optimised LIME framework for explainable student performance prediction in distance learning. UKR Journal of Engineering, Innovation and Technology, 1(1), 76-87.	

Introduction

The application of explainable artificial intelligence (XAI) in educational data mining and learning analytics has attracted attention because machine-learning models can provide accurate predictions without explaining the factors underlying those predictions. LIME has emerged as an important post-hoc, model-agnostic technique for interpreting complex black-box models. Recent research has extended LIME through counterfactual reasoning,

hierarchical explanations, fairness-aware feature selection, explanation clustering, and integration with other XAI methods. At the same time, optimisation techniques have been applied to student-performance prediction models, however, literature has identified concerns regarding the reliability of perturbation-based explanations when applied to contextual, temporal, and interdependent educational data.

Cohausz (2022) examined the interpretability of student-success prediction by proposing Minimal Counterfactual-LIME, an extension of LIME designed to identify the smallest number of feature changes required to alter a student's predicted outcome. The study proposed a three-step interpretability model progressing from feature importance, through counterfactual understanding, to causal relationships among similar students. While conventional LIME provides information about which features influence a prediction, the proposed approach extends interpretation toward actionable counterfactual information.

Igoche *et al.* (2024) also enhanced LIME in educational decision-making by integrating a Structural Causal Model ontology with LIME to develop a Fair-LIME framework for educational admissions. Their approach used causal knowledge to identify potentially biased features for ablation before generating explanations, resulting in reported improvements in prediction accuracy from 89% to 91% and F1-score from 94% to 95%, together with more stable and less fairness-biased explanations. Compared with Cohausz (2022), which focused primarily on counterfactual interpretation, Igoche *et al.* introduced causal knowledge and fairness considerations into the explanation process. However, neither study optimised the numerical parameters governing LIME's explanation-generation process, indicating that improvements to the type and fairness of explanations do not necessarily address the underlying parameter optimisation problem.

Nnadi and Watanobe (2026) extended LIME through Hierarchical LIME (H-LIME) to provide explanations at multiple educational levels, including institution type, location, and educational level. Their approach moved beyond isolated individual explanations by aggregating local explanations to reveal subgroup and global patterns in student adaptability prediction. The study showed that hierarchical aggregation could identify both general predictors and subgroup-specific factors, although its effectiveness depended on the validity of the predefined hierarchy and the quality of the underlying LIME explanations. Carvalho *et al.* (2025) further advanced LIME-based interpretability through LIME-EC, which clusters local LIME explanations to identify recurring explanation patterns and provide a quantitative measure of interpretability. Applied to approximately 17,000 undergraduate students for early dropout-risk prediction, the approach connected explanation patterns with model performance and fairness. Whereas H-LIME expanded the level and scope of explanations, LIME-EC focused on quantifying and evaluating explanation patterns across predictions. Both studies therefore improve the organisation of LIME explanations, but neither directly

addresses the optimisation of the LIME mechanism that generates those explanations.

Tirumanadham *et al.* (2025) combined a multi-tier ensemble model with Bayesian Optimisation, Particle Swarm Optimisation and Random Search for student-performance classification and subsequently employed LIME to explain the resulting predictions. The study showed the value of combining optimisation with XAI, with LIME identifying influential e-learning attributes such as VisITedResources and AnnouncementsView. However, the optimisation was primarily directed toward the predictive model, rather than toward the LIME explanation-generation process. Nnadi and Watanabe (2026) developed a multi-level XAI framework integrating H-LIME, SHAP and DiCE for predicting student depression risk and generating global, subgroup, and individual-level insights using a Random Forest model. In contrast to Tirumanadham *et al.*, which emphasised predictive optimisation alongside LIME-based explanation, Nnadi *et al.* focused on combining complementary XAI methods to provide broader explanatory coverage. These studies demonstrate increasing integration between optimisation, LIME and other XAI techniques; however, they continue to leave unresolved the question of whether the LIME parameters themselves can be optimised to improve the quality and efficiency of local explanations.

Gunasekara and Saarela (2025), through a systematic review of explainable learning analytics and educational data mining, identified LIME as an important post-hoc model-agnostic XAI technique but highlighted concerns about its application to educational data. In particular, educational variables are often contextual, temporal and interdependent, which may make conventional perturbation-based explanations less reliable when generated synthetic observations do not adequately represent realistic student behaviour. This finding is important because it shows that simply applying standard LIME settings to educational datasets may not produce consistently reliable explanations. Fridkin and Bendersky (2026) provided a broader interpretable-machine-learning perspective, identifying LIME as one of the influential post-hoc model-agnostic explanation frameworks and recognising the refinement of established methods such as LIME as an important direction for methodological development. In comparison with the empirical studies, these reviews provide broader evidence that LIME is both influential and in need of continued methodological refinement. However, they do not provide a specific optimisation mechanism for determining LIME parameters that are appropriate for distance-learning student-performance data.

Across the reviewed studies, a clear progression is observed in the development of LIME-based explainability for educational prediction. Early work focused on improving feature-importance explanations toward counterfactual and causal understanding, while subsequent research introduced hierarchical explanations, fairness-aware

explanations, explanation clustering, and multi-method XAI frameworks. Parallel research has shown the usefulness of optimisation techniques for improving predictive models before LIME is applied. More studies showed that LIME presents particular challenges when applied to educational data because student-related variables may be contextual, temporal and interdependent. Importantly, the existing studies modify what LIME explains, how explanations are aggregated, how explanations are evaluated, or how the underlying predictive model is optimised. Limited attention has been given to metaheuristic optimisation of LIME's explanation-generation parameters.

The literature also establishes the importance of optimisation in educational prediction and recognises LIME as an influential model-agnostic explanation framework. However, existing approaches have largely focused on improving the scope, structure or interpretation of LIME explanations, or on optimising the predictive models to which LIME is subsequently applied. This research aims to develop an Osprey-Optimised LIME Framework for Explainable Student Performance Prediction in Distance Learning.

Explainable Artificial Intelligence

Explainable Artificial Intelligence (XAI) has emerged as an interdisciplinary area of research concerned with making the behaviour, decisions, and outputs of artificial intelligence systems understandable to humans. The increasing adoption of complex machine-learning and deep-learning models has created a tension between predictive capability and human understanding, particularly because sophisticated models may achieve high predictive performance while providing limited insight into how individual decisions are generated. Angelov *et al.* (2021) identify this challenge as a central concern of XAI, emphasising the need to develop approaches that enable humans to obtain meaningful information about the reasoning underlying AI decisions.

Similarly, Minh *et al.* (2022) observe that the growing complexity of machine-learning and deep-learning models has intensified the need for methods that can provide explanations without necessarily replacing high-performing predictive models with inherently simpler ones. More recent reviews further indicate that XAI has evolved from a primarily technical concern into a broader research area involving interpretability, transparency, trustworthiness, accountability, human understanding, and responsible deployment of AI systems (Ortigossa *et al.*, 2024; Kalasampath *et al.*, 2025; Kaya *et al.*, 2026). Thus, XAI can be understood not merely as a mechanism for revealing how an AI model operates, but as a means of

facilitating meaningful human understanding and appropriate interaction with AI-driven decisions.

In educational settings, XAI has special significance because AI outputs often inform pedagogical decisions that affect student learning and evaluation. Research in explainable learning analytics highlights that teachers, students, and administrators require different explanation modalities (Khosravi *et al.*, 2022; Ifenthaler and Yau, 2020).

Local Interpretable Model-Agnostic Explanations (LIME)

Local Interpretable Model-Agnostic Explanations (LIME) is a post-hoc, model-agnostic XAI technique proposed by Ribeiro *et al.* (2016). LIME explains an individual prediction by approximating the behaviour of a complex black-box model within the neighbourhood of the instance being explained. LIME constructs a simpler and interpretable surrogate model that provides an approximation of the black-box model locally, rather than interpreting the complete predictive model.

For an instance x , LIME first generates perturbed samples around x . The black-box model f is then used to obtain predictions for these samples, while a locality function assigns greater weights to samples that are closer to the original instance. An interpretable surrogate model g is subsequently fitted to the weighted samples. The resulting surrogate model is used to identify the features that contribute most strongly to the prediction.

The general LIME objective can be expressed as:

$$\varepsilon(x) = \arg \min_{g \in G} \{L(f, g, \pi_x) + \Omega(g)\} \quad (001)$$

where f represents the original black-box predictive model, g represents the interpretable surrogate model, G is the class of interpretable models, $L(f, g, \pi_x)$ represents the local fidelity loss between the black-box model and the surrogate model under the locality weighting function π_x and $\Omega(g)$ represents a complexity penalty imposed on the surrogate model.

The effectiveness of LIME depends on the choice of surrogate model and on how the local neighbourhood is constructed and sampled. Important configuration factors include the number of perturbation samples, kernel width, sampling strategy, feature-selection settings, and other parameters that determine the local approximation. These factors can influence the quality of the resulting explanation.

Despite its advantages, LIME has several recognised limitations. Its explanations may vary across repeated runs because of the stochastic nature of the perturbation and sampling process (Alvarez-Melis and Jaakkola, 2018). The

method is also sensitive to parameter selection, especially parameters controlling locality and the generation of neighbourhood samples (Molnar, 2020). In addition, generating and evaluating large numbers of perturbations can introduce considerable computational overhead, when explanations must be generated repeatedly for large datasets or time-sensitive applications (Mosqueira-Rey et al., 2022). LIME also has difficulty representing complex feature interactions when the local surrogate is overly simplified (Slack et al., 2020). The limitations of conventional LIME consequently create an optimisation problem.

Metaheuristic Optimisation

Metaheuristic optimisation refers to general-purpose search strategies designed to obtain high-quality solutions to complex optimisation problems where conventional mathematical optimisation may be impractical. Metaheuristics are useful when the search space is large, nonlinear, multimodal, stochastic, or does not readily provide gradient information (Salgotra et al., 2024). Metaheuristic algorithms can be classified into single-solution and population-based approaches. Single-solution approaches iteratively improve one candidate solution, whereas population-based approaches maintain multiple candidate solutions and update them through interactions among members of the population. Population-based methods can provide greater diversity during the search and reduce the likelihood of becoming trapped prematurely in a poor local solution.

Osprey Optimisation Algorithm

The Osprey Optimisation Algorithm (OOA) is a population-based metaheuristic proposed by Dehghani and Trojovský (2023). The algorithm is inspired by the hunting behaviour of ospreys, particularly their ability to identify prey, dive towards a selected target, capture it, and subsequently move the prey to a suitable location. These behaviours are translated into computational mechanisms for exploring and exploiting a search space. In the developed OOA-LIME framework, each osprey represents a candidate configuration of the LIME parameters being optimised. If the optimisation problem contains m decision variables, the position of the i -th osprey can be represented as:

$$X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,m}] \quad (002)$$

The initial population is generated within the predefined lower and upper bounds of each optimisation variable:

$$x_{i,j} = lb_j + r_{i,j}(ub_j - lb_j) \quad (003)$$

where lb_j and ub_j denote the lower and upper bounds of the j th variable, respectively, and $r_{i,j}$ is a uniformly distributed random number in the interval $[0,1]$.

Phase 1: Position Identification and Hunting

During the first phase, each osprey identifies a promising prey based on the objective-function values of candidate solutions. Better-performing solutions are considered potential prey, while the best solution identified by the population provides an additional reference point. The prey-selection set can be represented as:

$$FP_i = \{X_k | k \in \{1, 2, \dots, N\}, F_k < F_i\} \cup \{X_{best}\} \quad (004)$$

where F_i represents the objective value of osprey i , N is the population size, and X_{best} represents the best solution currently identified.

An osprey then selects a promising prey and updates its position towards the selected solution. The general position update during the exploration phase can be represented as:

$$x_{i,j}^{new,PI} = x_{i,j} + r_{i,j}(SF_{i,j} - I_{i,j}x_{i,j}) \quad (005)$$

where $SF_{i,j}$ represents the position of the selected prey and $I_{i,j}$ is an algorithmic factor controlling the direction and magnitude of movement. This phase promotes exploration of different regions of the solution space and helps identify potentially promising LIME configurations.

Phase 2: Carrying the Prey to a Suitable Position

The second phase represents the osprey carrying the captured prey towards a safer or more suitable position. Computationally, this phase performs a more focused search around promising regions and therefore contributes to exploitation of the search space. A general position update can be expressed as:

$$x_{i,j}^{new,P2} = x_{i,j} + t[Ib_j + r_{i,j}(ub_j - lb_j)] \quad (006)$$

where t represents the current iteration-related factor and $r_{i,j}$ is a random value within $[0,1]$.

The alternating exploration and exploitation mechanisms allow OOA to search broadly for promising solutions while progressively concentrating the search around high-quality regions. This characteristic makes OOA potentially suitable for LIME parameter optimisation, where the objective landscape may be nonlinear and different parameter combinations may produce substantially different explanation quality.

In this study, OOA is therefore used as an optimisation engine rather than as the predictive model itself. The underlying machine-learning model remains responsible for predicting student performance, while OOA searches for a LIME configuration that produces improved local explanations. The optimisation process is consequently intended to shift LIME from a predominantly manually configured explanation method towards a systematically optimised explanation framework.

Author(s) / Year	Study Focus	XAI Approach	Dataset	ML Algorithm(s)	Findings	Limitations
Cohausz (2022)	Improves interpretability of student-success prediction	Minimal Counterfactual-LIME (MC-LIME); counterfactual reasoning and causal analysis	Student-success prediction data	Student-success predictive model + LIME	counterfactual explanations can identify minimum changes capable of altering predictions	Does not optimise LIME's parameters or explanation-generation process; counterfactual feasibility and causal interpretation remain challenges
Igoche et al. (2024)	Fair and causally informed explanations for educational admissions	Structural Causal Model ontology integrated with LIME; causal feature ablation	Polytechnic admission database	Predictive ML model + LIME	SCM-guided feature ablation improved fairness and stability; reported improvements in accuracy and F1-score	Depends on the correctness of domain-specific causal knowledge; does not perform metaheuristic optimisation of LIME parameters
Gunasekara & Saarela (2025)	Systematic review of explainable Learning Analytics and EDM	PRISMA-based systematic literature review	1,656 initially identified; 212 studies included	Multiple ML and XAI approaches	identifies limitations in explainability evaluation and the need for better optimisation of educational XAI systems	Does not develop an algorithm for optimising LIME explanation parameters
Nnadi & Watanobe (2026)	Hierarchical explanation of student adaptability	Hierarchical LIME (H-LIME); aggregates local explanations across institution, location and educational level	Student Adaptability in Online Education dataset; 1,205 instances	Decision Tree and Random Forest	Revealed individual, subgroup and hierarchical patterns in factors affecting student adaptability; Random Forest explanations were more stable	User-defined hierarchy may not represent all underlying relationships; aggregation can obscure individual variation; does not optimise LIME parameters
Carvalho et al. (2025)	Quantifying interpretability through explanation patterns	LIME-EC; clustering of local LIME explanations	Approximately 17,000 undergraduate students at risk of early dropout	Model-agnostic predictive model + LIME	Explanation clustering identified recurring explanatory patterns and provided a quantitative approach to interpretability	Dependent on the quality and stability of the initial LIME explanations; clustering does not itself optimise explanation generation
Tirumanadham et al. (2025)	Accurate and explainable student-performance prediction in e-learning	Multi-tier ensemble model, Bayesian Optimisation, PSO, Random Search and LIME	E-learning/student-performance data	Ensemble classification models	Demonstrated that optimisation can improve predictive performance while LIME provides feature-level explanations	Optimisation is primarily directed toward the predictive model, not the LIME explanation-generation parameters
Nnadi et al. (2026)	Multi-level explanation of student depression risk	H-LIME combined with SHAP and DiCE	Student depression-risk data	Random Forest	Provided global, subgroup and individual explanations by combining complementary XAI methods	Adds multiple explanatory techniques but does not systematically optimise LIME's locality, sampling or perturbation parameters
Fridkin & Bendersky (2026)	Foundations and future directions of interpretable ML	Comprehensive review of interpretable and post-hoc ML methods	Broad IML literature	Multiple ML/XAI methods	Identifies LIME as an influential post-hoc model-agnostic approach and recognises refinement of established interpretability methods as an important research direction	Broad review rather than an empirical educational study; does not develop an educational LIME optimisation framework

Methodology

The dataset consisted of 10,000 anonymised student records obtained from the Distance Learning Programme of Ladoko Akintola University of Technology (LAUTECH), covering the period from January 2020 to December 2024. Each record represents a student learning profile constructed from academic, behavioural, and learning-management-system information. The variables were grouped into three principal categories. Academic variables included continuous assessment performance and assignment submission behaviour. Learning-engagement variables included login frequency, session duration, resource utilisation, forum participation, and quiz attempts. Background variables included selected demographic and programme-related characteristics where available. Additional derived indicators, such as engagement and consistency measures, were calculated from the underlying behavioural records.

The dataset was divided into training, validation, and test subsets using a 70:15:15 stratified split. Stratification was applied to preserve the Pass/Fail class distribution across the three subsets.

All operations involving information learned from the data, including imputation parameters, scaling parameters, feature-selection decisions, and class-balancing operations, were derived exclusively from the training set. The resulting transformations were subsequently applied unchanged to the validation and test sets. The validation set was used for model development and parameter selection, whereas the test set remained completely untouched until final evaluation. This procedure prevents information from the validation or test data from influencing model development or explanation optimisation.

Data preprocessing was performed to improve data quality, ensure numerical comparability, and prepare the variables for predictive modelling. Missing numerical observations were replaced using the median calculated from the training data, while missing categorical observations were replaced using the mode calculated from the training data. The corresponding training-derived imputation parameters were subsequently applied to the validation and test sets. Median imputation was selected for numerical variables because it is less sensitive to extreme observations than mean imputation. Mode imputation was used for categorical variables.

Feature Selection and Class Imbalance

Feature selection was performed exclusively on the training data to reduce redundancy and retain variables that contribute useful information to student-performance prediction. Feature-selection decisions were not based on the validation or test sets. Because the Pass/Fail distribution was not perfectly balanced, Synthetic Minority Over-Sampling Technique (SMOTE) was applied only to the training data after feature selection. SMOTE generates synthetic observations for the minority class by interpolating between existing minority-class observations. No synthetic observations were introduced into the validation or test sets. Consequently, these sets retained their original class distributions and provided a more realistic assessment of model performance.

Osprey-Optimised LIME Framework

The principal contribution of the study is the integration of the Osprey Optimisation Algorithm (OOA) with LIME to automatically search for improved LIME configurations.

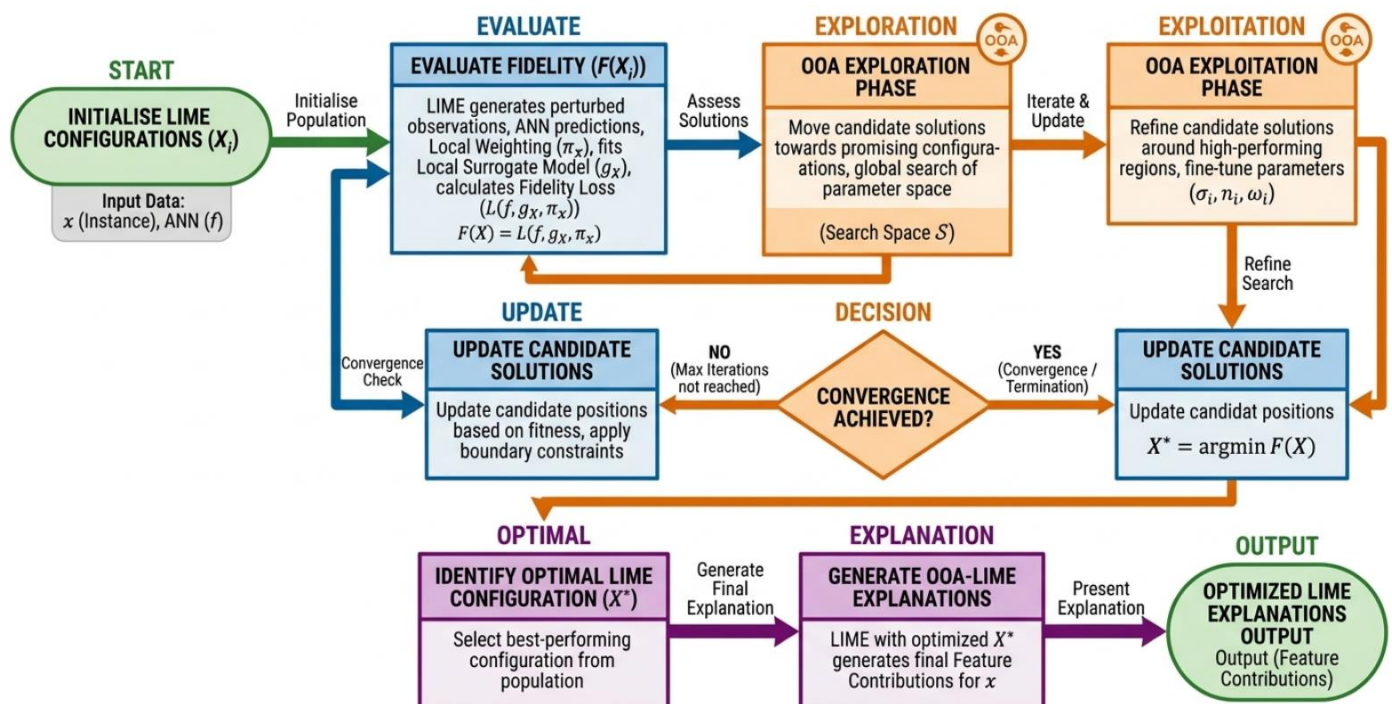


Figure 1: Flow diagram of the proposed OOA-LIME framework for optimising LIME configurations and generating local feature-based explanations

Figure 1 presents the operational flow of the proposed Osprey Optimisation Algorithm–LIME (OOA-LIME) framework, which integrates the Osprey Optimisation Algorithm (OOA) with Local Interpretable Model-agnostic Explanations (LIME) to automatically identify an effective configuration of LIME for explaining the predictions of an artificial neural network (ANN). The framework is organised into five principal stages: initialisation, evaluation, exploration, exploitation, and decision/update, followed by identification of the optimal LIME configuration and generation of the final explanation.

The process begins with the initialisation of the LIME configuration population. Given an input instance x and the trained ANN model f , a population of candidate LIME configurations X is generated. Each candidate represents a particular combination of LIME parameters, such as the kernel-width or locality parameter σ , number of perturbation samples n , and feature-selection or weighting parameter ω as shown in eqn 007.

$$X_i = [\sigma_i, n_i, \omega_i] \quad (007)$$

These candidate configurations constitute the initial search population from which the OOA searches for an improved LIME configuration.

In the evaluation stage, each candidate configuration is used to generate perturbed samples around the target instance. LIME applies the candidate configuration to the perturbed instances, obtains predictions from the ANN, and constructs a local surrogate model g . The quality of each candidate is then quantified using a predefined fitness function. In the proposed framework, the fitness function represents the discrepancy between the predictions of the original ANN and those of the LIME surrogate within the local neighbourhood. Thus, a candidate configuration that produces a surrogate model that closely approximates the ANN is assigned a better fitness value. This evaluation provides the information required by OOA to guide the subsequent search.

The exploration phase implements the global-search behaviour of OOA. Ospreys explore the search space by moving candidate solutions towards promising positions, thereby encouraging the algorithm to investigate different regions of the LIME parameter space. This stage is important because LIME performance can vary substantially according to its

parameter configuration. Rather than relying on a manually selected configuration, OOA systematically searches for configurations that may provide improved local fidelity. The candidate solutions generated through exploration are subsequently evaluated using the same fitness function.

The framework then enters the exploitation phase, in which the search becomes increasingly focused on promising regions identified during exploration. Candidate solutions are refined around high-performing configurations, with the objective of improving their fitness through local search. The resulting positions are again evaluated against the ANN–LIME fitness function. This alternating emphasis on exploration and exploitation enables the optimisation process to balance the discovery of new regions of the parameter space with the refinement of already promising configurations.

Following the exploration and exploitation operations, the framework performs a convergence check. The decision stage determines whether the optimisation process has achieved the specified convergence or termination criterion, such as reaching the maximum number of iterations or satisfying a predefined improvement threshold. If convergence has not been achieved, the candidate solutions are updated and constrained to remain within the permissible parameter boundaries. The updated population is then returned to the evaluation stage, creating an iterative optimisation loop. This cycle continues until the termination condition is satisfied.

Once convergence is achieved, the framework performs the identification of the optimal LIME configuration. The candidate with the best fitness value is selected as the optimal configuration, denoted as X^* , such that

$$X^* = \underset{X \in S}{\operatorname{argmin}} F(X) \quad (008)$$

where $F(X)$ represents the fitness function associated with a candidate LIME configuration. The selected configuration therefore represents the parameter combination that provides the lowest measured discrepancy between the ANN predictions and the corresponding local surrogate explanation, subject to the defined search constraints.

The optimal configuration is subsequently passed to the explanation-generation stage. LIME is executed

again using X^* to produce the final local explanation for the target instance. The explanation identifies the features that contribute to the ANN prediction and quantifies their respective contributions. The final output is therefore an optimised LIME explanation, consisting of the feature contributions generated using the OOA-selected configuration.

System Implementation and User Interface

The optimised XAI system was implemented using the MATLAB integrated development environment, using its comprehensive toolboxes for machine learning, optimisation, and visualisation. Figure 2 shows the graphical user interface (GUI) of the developed system, which provides an intuitive platform for users to interact with the optimised LIME explainer.

The interface, as shown in Figure 2, is organised into four main functional panels, which are the dataset display panel, the control panel, evaluation results panel and the output controls. The dataset display panel displays the loaded raw student data in a tabular format. This panel allows users to preview the dataset, verify data integrity, and confirm that the correct records have been loaded for analysis. The control panel allow users to load raw data, initiate the optimisation process, and generate explanations for individual student predictions. The evaluation result panel displays the performance evaluation metrics after the system has been executed. While the output controls provide functionality for saving the generated explanations to a file resultOOALIME and clearing the current results for new analysis.

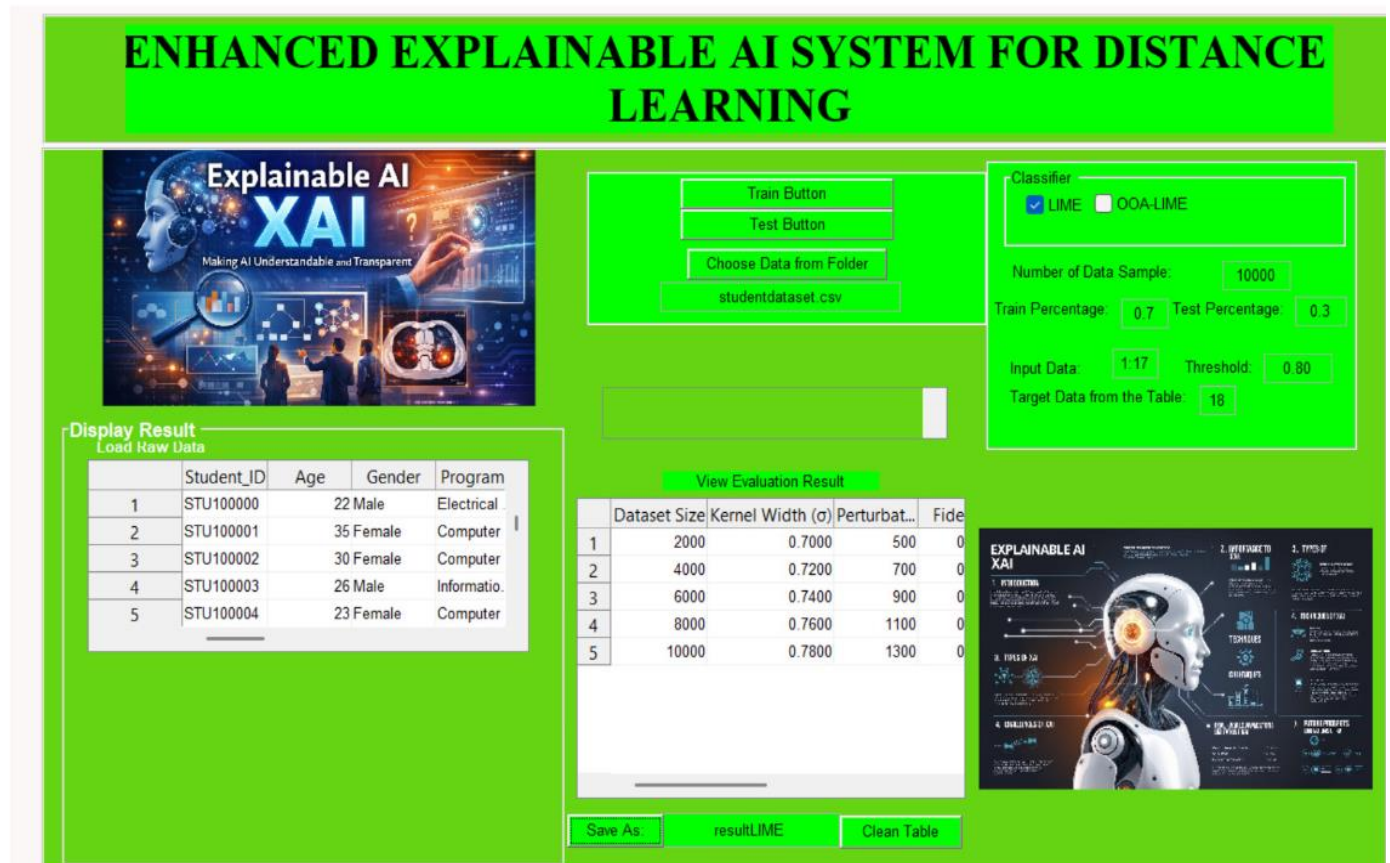


Figure 2: The Graphical User Interface (GUI) of the Optimised LIME

Performance Evaluation

Table 1 presents the performance evaluation of Standard LIME and the developed OOA-LIME across the evaluated dataset sizes. The results show that OOA-LIME consistently achieved higher values than Standard LIME across fidelity, interpretability, comprehensibility, trustworthiness, and coverage, while requiring less time to generate explanations. The

consistent difference between the two approaches indicates that optimising LIME configurations using the Osprey Optimisation Algorithm improved both the quality of the generated explanations and the efficiency of the explanation process.

The results further show that the performance advantage of OOA-LIME was maintained as the dataset size increased from 2,000 to 10,000 records.

For Standard LIME, fidelity increased from 0.780 to 0.812, whereas OOA-LIME increased from 0.850 to 0.890. Similarly, interpretability increased from 0.710 to 0.738 for Standard LIME and from 0.800 to 0.836 for OOA-LIME. This consistent pattern across the evaluated dataset sizes suggests that the effect of Osprey-based optimisation was not restricted to a particular dataset size.

The fidelity results show that OOA-LIME provided a closer approximation of the underlying predictive model than Standard LIME. Fidelity increased from 0.850 to 0.890 for OOA-LIME compared with 0.780 to 0.812 for Standard LIME across the evaluated dataset sizes. This represents an improvement of approximately 9.0–9.6% relative to Standard LIME. The result indicates that the optimised LIME configurations were better able to reproduce the local behaviour of the ANN around the instances being explained. The progressive increase in fidelity with dataset size also indicates that the optimised configuration maintained its effectiveness as the volume of data increased.

The interpretability results similarly demonstrate an improvement following Osprey-based optimisation. OOA-LIME recorded values ranging from 0.800 to 0.836, compared with 0.710 to 0.738 for Standard LIME. The corresponding relative improvement ranged from approximately 12.7% to 13.3%. This indicates that the optimised configurations produced explanations that were more readily understandable while preserving the local relationship between the predictive model and the explanation.

For comprehensibility, OOA-LIME increased from 0.770 at 2,000 records to 0.802 at 10,000 records, while Standard LIME increased from 0.680 to 0.704. The improvement of approximately 13.2–13.9% relative to Standard LIME indicates that the

optimisation process improved the clarity and conciseness of the generated explanations. This finding is important because an explanation that accurately represents the model but contains difficult-to-understand or excessive information may have limited practical value to end users.

The trustworthiness results show a similar pattern. OOA-LIME recorded scores between 0.830 and 0.862, compared with 0.740 and 0.768 for Standard LIME. This corresponds to an improvement of 12.2–12.3% relative to Standard LIME. The result suggests that the improved local fidelity and explanation characteristics associated with the optimised configurations were accompanied by greater consistency between the model's prediction and the generated explanation.

The coverage results provide further evidence of the effect of optimisation. OOA-LIME achieved coverage values ranging from 0.810 to 0.838, compared with 0.700 to 0.720 for Standard LIME. The relative improvement ranged from approximately 15.7% to 16.4%. The higher coverage indicates that the optimised configurations were able to provide meaningful explanations across a broader proportion of the evaluated instances.

In terms of timeliness, OOA-LIME required less time to generate explanations at every dataset size. Standard LIME required between 12.01 and 12.98 seconds, whereas OOA-LIME required between 7.58 and 8.09 seconds. This represents a reduction in explanation-generation time of approximately 32.6–41.1% relative to Standard LIME. The reduction is particularly relevant for distance-learning applications, where explanation systems may need to process predictions for a large number of students within a reasonable period.

Table 1: Comprehensive comparative performance of LIME and OOA-LIME across dataset sizes

Dataset size	Fidelity		Interpretability		Comprehensibility		Trustworthiness		Coverage		Timeliness (s)	
	Standard LIME	OOA-LIME	Standard LIME	OOA-LIME	Standard LIME	OOA-LIME	Standard LIME	OOA-LIME	Standard LIME	OOA-LIME	Standard LIME	OOA-LIME
2,000	.780	.850	.710	.800	.680	.770	.740	.830	.700	.810	12.01	8.09
4,000	.788	.860	.717	.809	.686	.778	.747	.838	.705	.817	12.04	7.82
6,000	.796	.870	.724	.818	.692	.786	.754	.846	.710	.824	12.98	7.74
8,000	.804	.880	.731	.827	.698	.794	.761	.854	.715	.831	12.92	7.66
10,000	.812	.890	.738	.836	.704	.802	.768	.862	.720	.838	12.86	7.58

Conclusion

This study developed and evaluated an Osprey-Optimised LIME (OOA-LIME) framework for explainable student-performance prediction in distance learning. The framework addresses a limitation of conventional LIME by using the Osprey Optimisation Algorithm to systematically search for LIME configurations that improve the quality and efficiency of local explanations. The results across dataset sizes of 2,000 to 10,000 records showed consistent improvements in fidelity, interpretability, comprehensibility, trustworthiness, and coverage compared with Standard LIME. OOA-LIME also reduced explanation-generation time, indicating improved computational efficiency.

The findings show that optimising the configuration of the explanation mechanism, rather than only the underlying predictive model, can influence the quality of model explanations. The consistent results across the evaluated dataset sizes further indicate that the benefits of OOA-based configuration search were maintained as the volume of student data increased. Thus, the developed framework provides an alternative to relying on default LIME settings for explaining student-performance predictions.

The study contributes to explainable learning analytics by introducing a metaheuristic optimisation approach targeted at the LIME explanation-generation process. The framework

can support teachers and other educational stakeholders in obtaining more informative and efficient explanations of machine-learning predictions. Future studies could examine the framework across different educational datasets, predictive models, LIME parameter spaces, and explanation-quality measures, as well as investigate its performance against other LIME optimisation and XAI approaches.

Funding

This research received no external funding.

Data Availability Statement

There are no data associated with this manuscript.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was not supported by any grant.

1. Alvarez-Melis, D. and Jaakkola, T.S. (2018) 'On the robustness of interpretability methods', *Proceedings of the 35th International Conference on Machine Learning*, pp. 109–118.
2. Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(5), e1424. <https://doi.org/10.1002/widm.1424>
3. Carvalho, C. S., de Mattos, J. C. B., & de Aguiar, M. S. (2025, November). LIME-EC: an Agnostic Approach to Quantifying Model Interpretability using Explanation Clustering. In *2025 12th International Conference on Soft Computing & Machine Intelligence (ISCMI)* (pp. 549-555). IEEE.
4. Cohausz, L. (2022). When probabilities are not enough-A framework for causal explanations of student success models. *Journal of Educational Data Mining*, 14(3), 52-75.
5. Dehghani, M., & Trojovský, P. (2023). Osprey optimization algorithm: A new bio-inspired metaheuristic algorithm for solving engineering optimization problems. *Frontiers in Mechanical Engineering*, 8, 1126450.
6. Fridkin, S., & Bendersky, M. (2026). Interpretable machine learning: a comprehensive review of foundations, methods, and the path forward. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 16(1), e70075.
7. Gunasekara, S. and Saarela, M. (2025) 'Systematic literature review on explainable Learning Analytics and Educational Data Mining', *IEEE Access*, 13, pp. 214387–214408. doi: 10.1109/ACCESS.2025.3643645.
8. Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational technology research and development*, 68(4), 1961-1990.
9. Igoche, B.I., Matthew, O., Bednar, P. and Gegov, A. (2024) 'Integrating structural causal model ontologies with LIME for fair machine learning explanations in educational admissions', *Journal of Computing Theories and Applications*, 2(1), pp. 65–85. doi: 10.62411/jcta.10501.
10. Kalasampath, K., Spoorthi, K.N., Sajeev, S.P., Kuppa, S.S., Ajay, K., & Maruthamuthu, A. (2025). A Literature Review on Applications of Explainable Artificial Intelligence (XAI). *IEEE Access*, 13, 41111-41140.
11. Kaya, O., Shah, A. S., Karabulut, M. A., Karahan, S. N., Osmanca, M. S., & Acir, N. (2026). Explainable artificial intelligence (XAI): Concepts, applications, challenges, and future perspectives. *IEEE Access*.
12. Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., ... & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and education: artificial intelligence*, 3, 100074.
13. Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, 55(5), 3503-3568.
14. Molnar, C. (2022) *Interpretable Machine Learning*. 2nd edn. Available at: <https://christophm.github.io/interpretable-ml-book/>.
15. Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: a state of the art: E. Mosqueira-Rey et al. *Artificial Intelligence Review*, 56(4), 3005-3054.
16. Nnadi, L. C., Isiwu, C. P., Ding, D., Muepu, D. M., & Watanobe, Y. (2026). Multi-level explainable ai for predicting student depression risk: Global, subgroup, and individual insights. *IEEE Access*, 14, 6271-6286.
17. Ortigossa, E.S., Gonçalves, T., & Nonato, L.G. (2024). EXplainable Artificial Intelligence (XAI): From Theory to Methods and Applications. *IEEE Access*, 12, 80799-80846.
18. Ribeiro, M.T., Singh, S. and Guestrin, C. (2016) "“Why should I trust you?”: Explaining the predictions of any classifier", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144. doi:10.1145/2939672.2939778.
19. 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.

20. Salgotra, R., Sharma, P., Raju, S., & Gandomi, A. H. (2024). A contemporary systematic review on meta-heuristic optimisation algorithms with their matlab and python code reference. *Archives of Computational Methods in Engineering*, 31(3), 1749-1822.
21. Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (pp. 180-186). <https://doi.org/10.1145/3375627.3375830>
22. Tirumanadham, N. K. M. K., Thaiyalnayaki, S., & Ganesan, V. (2025, February). Enhancing student performance prediction using E-learning through multimodal data integration and machine learning techniques. In *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)* (pp. 933-940). IEEE.