International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 31, No. 5, May, 2026, pp. 1655-1666
Journal homepage: http://iieta.org/journals/isi

Interpretable Learner's Dropout Feature Ranking Using Local Binary Patterns and Kullback-Leibler Divergence Across Educational Datasets

Houssam Ahmed Amin Bahi[1*], Karima Boussaha[2], Zakaria Laboudi[3], Christos Troussas[4]

[1] Department of Mathematics and Computer Science, Research Laboratory on Computer Science's Complex Systems (ReLa(CS)2), University of Oum El Bouaghi, Oum El Bouaghi 04000, Algeria
[2] Department of Mathematics and Computer Science, Artificial Intelligence and Autonomous Things Laboratory (LIAOA), University of Oum El Bouaghi, Oum El Bouaghi 04000, Algeria
[3] Department of Networks and Telecommunications, Artificial Intelligence and Autonomous Things Laboratory (LIAOA), University of Oum El Bouaghi, Oum El Bouaghi 04000, Algeria
[4] Department of Informatics and Computer Engineering, University of West Attica, Athens 12243, Greece

Corresponding Author Email: houssam.bahi@univ-oeb.dz

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/isi.310521

Received: 4 March 2026
Revised: 25 April 2026
Accepted: 5 May 2026
Available online: 31 May 2026

Keywords:
learner dropout prediction, feature ranking, Local Binary Patterns, Kullback–Leibler divergence, educational data mining, learning analytics, interpretable machine learning

ABSTRACT

Educational dropout prediction requires feature-ranking methods that are both informative and interpretable across heterogeneous learner data. This study introduces Local Binary Pattern – Kullback-Leibler (LBP-KL), a one-dimensional local binary pattern framework that ranks educational features by comparing class-specific local-pattern histograms with symmetric KL divergence. For sequential data, Local Binary Pattern (LBP) codes are extracted from timestamp-ordered learner-event streams; for static tabular data, feature values are ordered by magnitude before local-pattern encoding. LBP-KL was evaluated on the Capacitação da Administração Pública (SATDAP), Metaverse_Edu_Rich, and Stanford-ACT-MOOC datasets using 10 repeated stratified 80:20 train-test splits. Feature selection was fitted only on training data, and SMOTE, when required, was applied only after splitting. The method was compared with ANOVA F, chi-square, mutual information, random-forest importance, and recursive feature elimination for the event-stream dataset. Performance was assessed with accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC, with neighborhood sizes P = 4, 6, 8, and 10 examined. LBP-KL identified class-dependent local-pattern differences and produced interpretable rankings across static and sequential educational settings; however, it did not outperform the strongest conventional selectors on SATDAP or Stanford-ACT-MOOC. On Metaverse_Edu_Rich, all methods showed limited discrimination, with ROC-AUC values close to 0.50. The findings position LBP-KL as a complementary exploratory ranking method for datasets with meaningful ordered variation, rather than as a universal feature-selection solution. This boundary is important for responsible learning-analytics deployment.

1. INTRODUCTION

Massive Open Online Courses (MOOCs) have emerged as a transformative mode of delivering education to a vast and diverse number of learners all over the world. Despite their promise of accessible and scalable learning opportunities, MOOCs are known to exhibit high dropout rates, which is above 80% for several courses. This challenge has motivated extensive research into understanding dropout factors [1, 2], with the aim of enabling timely prediction and interventions to enhance retention and learner outcomes [3-5]. While numerous approaches ranging from statistical models [6] to deep learning architectures [7] have been investigated, the majority rely on conventional feature engineering and fail to capture the fine-grained behavioral nuances occurring in learner activity sequences [8-10].

In this study, Local Binary Pattern (LBP) is extended from its use in texture analysis of images to educational dropout data as a one-dimensional local-pattern feature. Instead of stating that the use of LBP renders other feature selection methods unnecessary, the primary objective here is to investigate if local relative variation in educational features provides an interpretable structure associated with dropouts. In this study, we regard the educational features as sequences. If the learner interaction log data is accessible, temporal sequences are applied; otherwise, we make use of value-based sequences in tabulated features. The resulting LBP histograms are then compared between the learner classes.

The proposed framework uses histograms and symmetric Kullback-Leibler (KL) divergence to compute Local Binary Pattern - Kullback-Leibler (LBP-KL) scores that measure the extent to which the local pattern distribution of a feature differs between dropout and non-dropout learners. These LBP-KL scores are then used to rank the features. Unlike mean difference or scalar-based univariate correlations, LBP-KL considers the organization of value patterns at a local level.

This makes the approach more sensitive to detecting differences between classes due to local pattern repetitions.

Another component of the proposed methodology is the direct comparison of LBP patterns. By evaluating the distributional differences between pattern occurrences across learner groups, behavioral contrasts that may not be captured by conventional statistical summaries or simple frequency counts are revealed. Such a comparative perspective improves both the interpretability and the actionable potential of predictive models.

Although the proposed representation may inspire future investigations in other educational analytics settings, the present study focuses exclusively on evaluating LBP-KL as a feature-ranking method for educational dropout prediction.

The key contributions of this study include:

- A one-dimensional LBP adaptation for dropout data in education, allowing static value ordering of features and sequence interactions between features.
- A feature ranking approach based on symmetric KL divergence, which involves comparing LBP histograms to discover locally discriminative features for education.
- A repeated and leakage-free experimental evaluation comparing LBP-KL with ANOVA F, Chi-Square, Mutual Information, and Random Forest feature importance measures on three educational datasets.
- A statistical validation of results through mean ± standard deviation, and pairwise significance test over repeated experiments.
- A neighborhood size sensitivity study analyzing different local neighborhood sizes of $P = 4, 6, 8$, and 10.
- A critical discussion on interpretability and limitations regarding the types of features and datasets LBP-KL works on and vice versa.

By combining a one-dimensional LBP adaptation, KL-divergence-based ranking, and cross-dataset evaluation, the proposed method investigates the usefulness of local pattern representations for interpretable educational dropout analysis.

The remainder of this paper is structured as follows: Section 2 reviews related work on MOOCs dropout analysis and LBP applications. Section 3 details our proposed methodology. Section 4 presents the results of our experiments. Section 5 discusses the findings, and Section 6 concludes the paper with a summary and outlines potential future research directions.

2. LITERATURE REVIEW

2.1 Dropout prediction and conventional feature engineering

Predictive modeling for learner's dropout has emerged as an important theme in education data mining and learning analytics, especially concerning MOOCs, which are often characterized by learner disengagement and non-completion issues. Prior research mainly made use of demographic, academic, socioeconomic, and engagement-related features to detect learners who may drop out of their courses. The features considered include age, gender, educational experience, registration details, login behavior, test scores, forum activity, and study time. Standard approaches like K-means clustering, MANOVA, chi-square tests, Pearson correlation, principal component analysis, and decision trees were used to investigate these features and predict dropout rates [8-12].

While such approaches give us valuable information, they normally work on the basis of a global representation of learning behavior. This can be in terms of how often a user accessed material, how many times assignments were submitted, or how often the user used the application itself. While this kind of information is very valuable, it can obscure local behavior in terms of sporadic activity, sudden inactivity, erratic behavior, or frequent fluctuation in use of the system. In this sense, traditional feature engineering can miss out on information that is not easily identifiable in terms of its own processes.

2.2 Sequential and temporal modeling in dropout prediction

To overcome some of the weaknesses of static features, several other studies have considered some more sophisticated data mining and time series modeling techniques. The utilization of techniques like Mutual Information, Random Forest importance measure, Recursive Feature Elimination, and dynamic feature selection has helped enhance the accuracy of predictors for learners' dropouts [13, 14]. Some other studies employed deep learning approaches such as self-attention techniques and recurrent neural networks such as LSTM to learn from hidden temporal patterns of learners' behavior sequences [7, 15].

Such techniques have proven effective in various scenarios, particularly when learner behavior is modeled through sequential behavior. Nonetheless, such techniques present another problem in terms of interpretability. While deep learning and sophisticated ensembles will be able to discern a pattern that allows for prediction, they may not necessarily explain which behavioral structures at a local level set apart a learner who is going to drop out from one that is not. This is a critical point since simply being able to predict a behavior is not enough. Understanding why predictions occur is equally important.

2.3 Interpretability gap in educational dropout analysis

Consequently, the present state-of-the-art in dropout prediction research suffers from an apparent methodological dilemma. Simple models and traditional approaches to feature selection tend to be more interpretable, but they may prove insufficiently expressive and miss some local or temporal behavior peculiarities. At the same time, sophisticated models can capture complex interactions and temporal dependencies but often remain black boxes, offering little insight into the patterns leading to prediction. Consequently, there is a need for a methodology that preserves local structures while remaining interpretable.

Against this background, feature representation itself gains equal importance with modeling. An appropriate representation should not only be able to enhance classification accuracy, but also offer a possibility to differentiate between dropouts and non-dropouts using behavioral patterns. This is particularly relevant in educational applications, as pattern-based analysis of learners' learning activities can help build an early warning system, construct learner profiles and implement appropriate interventions. Still, such claims to interpretability should not be easily justified - demonstrating only differences between two distributions will not suffice. The representation used for comparison needs to provide insight on how to interpret the patterns detected and what makes the representation valid.

2.4 Local Binary Patterns as local structure descriptors

The LBP technique was originally proposed as a texture descriptor in image processing. The traditional implementation of LBPs defines how pixels around the center pixel compare to the center pixel through an encoding scheme, whereby the value of each neighbor being greater than or equal to the center pixel results in a binary representation of such local texture. LBP has found wide applications in various fields, including but not limited to image analysis, face recognition, biometrics, medical imaging, and other pattern recognition domains [16-23].

The usefulness of LBP for educational data arises not from any images, but from its capability to represent local relative variation. As such, since learner behaviors can be considered as sequences, LBP can be modified from a method of extracting information from two-dimensional images to a method for extracting local patterns in one-dimensional sequences. In this adaptation, a comparison is made between the center value and neighboring values in the sequence, where the binary code indicates the local arrangement around the particular value. In the case of temporal data, this is done based on a timeline. On the other hand, with static data, this is achieved by sorting the values.

2.5 Positioning Local Binary Pattern against other sequence representations

LBP is different from sliding window statistic, Markov transition, and n-gram features with regard to the information that it captures. Sliding window statistics computes statistical measures for a given locality using the average, variance, slope, or frequency of the neighborhood, although the shape of the neighborhood may be lost in the process. On the other hand, while Markov transition and n-gram features concentrate on transitions between two states or events, they need a meaningful set of states or events and tend to be sparse as the number of states becomes large. Conversely, LBP captures the relative ordering of neighboring values to a given central value using binary codes.

LBP should not be viewed as a universal approach to supersede all other feature selection approaches or modeling of learner behaviors. The relevance of LBP would depend on the characteristics of the data and the choice of order. It will only be effective in case there is enough variation within features to distinguish between their values and there is some correlation between chosen ordering and learner behaviors. Furthermore, LBP is expected to perform poorly in case there are binary features, high sparsity in feature values, short sequences, or the data is mainly categorical and lacks any meaningful local order. Therefore, LBP will serve as an alternative way to rank the features and find out patterns in the current research.

2.6 Research gap and position of the present study

Three gaps have been detected after reviewing the literature. First, many previous attempts to predict dropouts used aggregate features without considering any local structure in learner behaviors. Second, although some methods of temporal and deep learning could detect more complex patterns, they lack interpretability. Third, there are not enough studies examining how LBP can be used in an educational setting for finding patterns and ranking features.

In order to fill these gaps, the current research attempts to design and evaluate an LBP-KL method of feature ranking related to learners' dropout status. Specifically, this method applies LBP to one-dimensional sequences of educational features, creates histograms of class-specific LBP codes, and computes feature rankings based on the symmetric KL divergence measure. This research considers only the actual implementation aspects, including adaptations of LBP, ranking algorithm based on KL divergence, comparisons with standard methods of selecting features, repeated tests, statistical analysis, and influence of different neighborhood sizes.

3. METHODOLOGY

In this section, the LBP-KL approach for interpretability of dropout features will be outlined. The proposed methodology consists five major steps: adaptation of LBP for one-dimensional education-related features, generation of class-based histograms based on LBP, ranking features according to symmetric KL divergence, predictive evaluation through repeatable leakage-aware experiments, and exploration of the effect of neighborhood size on different datasets.

3.1 Local Binary Patterns explanation

LBP represents a method designed for local texture description initially introduced for images. The algorithm works by comparing the central pixel with all its neighboring pixels and creating a binary number based on the comparison results. [17, 24]. The basic idea is to label each pixel by thresholding its neighborhood against the center pixel value (see Figure 1).

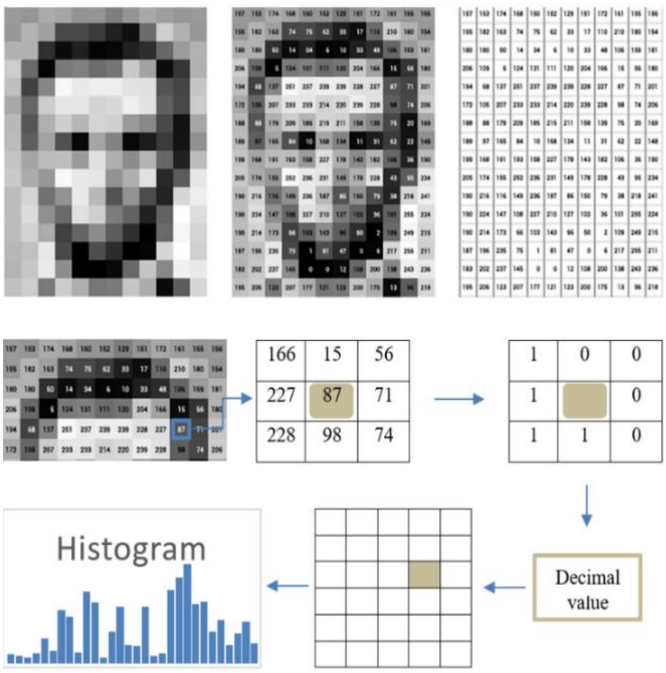

Figure 1. Local Binary Patterns (LBP) concept for image data

3.2 Adapting Local Binary Pattern to non-image data

The first adaptation that needs to be considered in relation to applying LBP in education is the shift from a descriptor of two-dimensional images to a one-dimensional descriptor of a local pattern. The algorithm is adjusted to analyze neighboring values, instead of neighboring pixels in the case of educational data. The order of the features depends on the nature of the data set and either represent a time sequence of interactions between the learner and the system or can reflect the sorting order of the feature values in tabular form.

3.2.1 Data preprocessing

Some educational data features may have high levels of sparseness and/or noisiness. To increase the consistency of patterns and avoid excessive fragmentation of LBP codes, the process of quantization can be conducted if necessary. Values are rounded to the previously determined level of precision (one decimal point, for example). In this way, the process of quantization reduces the effects of small deviations in feature values while keeping the larger pattern of local structure intact.

3.2.2 Neighborhood definition

For a one-dimensional sequence, a symmetric window of size P is considered for each center point value P_c. In this window, there will be $P/2$ neighbors on both sides of the window. Each neighbor value is compared with its corresponding center value and generates an output value of 1 if $P_i \geq P_c$; otherwise, it will generate an output value of 0. The sequence of output values is transformed into a decimal LBP code for the local structure of each center point.

For the early set of experiments, $P = 8$ is assumed, implying four neighbors on either side and $2^8 = 256$ possible patterns. But the modified experiment also considers the sensitivity to the size of the window, including P=4,6,8, and 10. Mathematically, the LBP value for a pixel (P_c) with neighbors (P_i) can be represented as:

$$LBP(x_i) = \sum_{P=0}^{P-1} s(x_{i+p} - x_i), 2^p \quad (1)$$

$$s(z) = \begin{cases} 1 \; if \; z \geq 0 \\ 0 if \; z < 0 \end{cases} \quad (2)$$

3.2.3 Data ordering

Handling Static and Sequential Data An essential part of tailoring LBP for use with educational datasets lies in specifying the ordering approach used for the sliding neighborhood window. The choice of the ordering procedure is determined by the structural features of the dataset as illustrated in Figure 2.

(1) Case 1: Sequential Event-Stream Features

In sequential datasets, like Stanford-ACT-MOOC, events are sequenced per learner based on their timestamps. The LBP sliding window is then applied without any changes to the existing sequence. For these kinds of datasets, the output LBP codes refer to local behavioral transitions, including sudden bursts of activity, a drop in participation levels, as well as varying intensity of interactions.

(2) Case 2: Static Tabular Features

Static datasets, such as Capacitação da Administração Pública (SATDAP) and Metaverse_Edu_Rich, have no temporal features available. Thus, feature values are sorted by increasing order of magnitude prior to running LBP window. For these datasets, LBP codes denote the local structure of the feature values distribution rather than transitions over time since there were no temporal relationships assumed in the data in the first place.

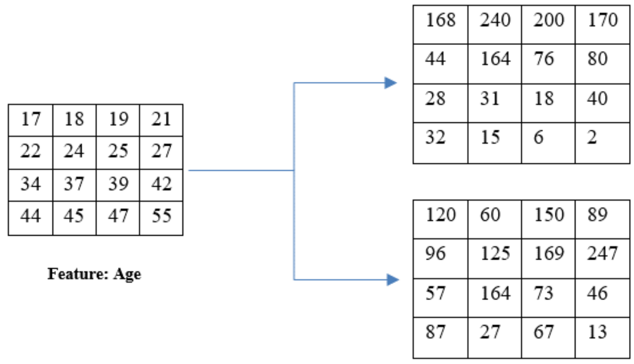

Figure 2. Value ordering and local neighborhood generation for static educational features

3.2.4 Pattern generation and histogram construction

Once LBP has been applied on each feature series in the given order, histograms corresponding to the frequency of occurrences of local patterns are computed from the extracted LBP codes. Two separate histograms are created for dropout learners and non-dropout learners. The histograms serve as local signatures of behavior/distribution in the given feature space.

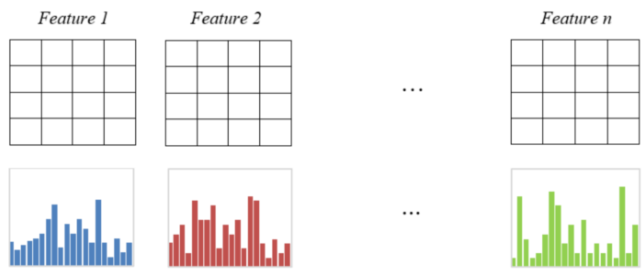

Figure 3. Example of class-wise Local Binary Pattern (LBP) histogram generation for a selected educational feature

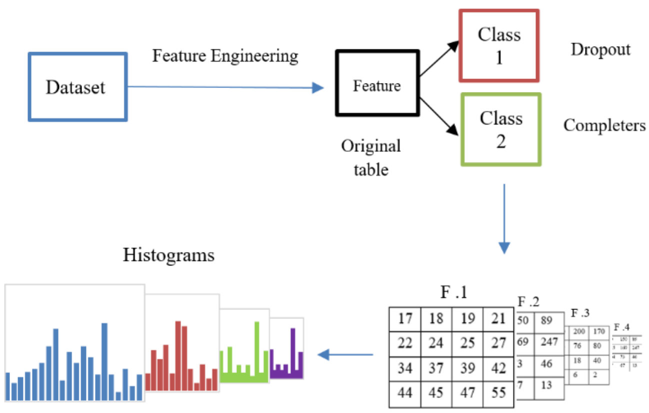

Figure 4. Overall Local Binary Pattern - Kullback-Leibler (LBP-KL) ranking and predictive evaluation pipeline

Those features that show considerably varying LBP histograms between learner types are deemed relevant for predicting dropouts. Those features with binary or little variation cannot be ranked using the proposed method due to insufficient variation to define any neighborhoods (see Figure 3 and Figure 4).

3.3 Kullback-Leibler-based feature ranking and predictive evaluation

Given the histograms computed based on the ordered LBP pattern occurrences, the difference can be quantified between dropout and non-dropout learners in terms of the underlying structural characteristics. To this end, symmetric KL divergence is calculated between class-wise histograms associated with a given feature.

These LBP-KL scores are then used for sorting the features based on their local pattern distinguishability. As opposed to classical statistical approaches, where feature selection is mostly based on global correlation or variance measures, the proposed algorithm focuses on local structural behavior, which is captured by the ordered feature sequence.

In order to assess the value of the selected features for classification, the experiments are conducted using repeated train-test splits with stratification. The feature selection is done only on the training dataset, since otherwise there might be a risk of information leakage. The classifiers being tested are Logistic Regression, Random Forest and Gradient Boosting. Metrics used include Accuracy, Recall, Precision, F1-score, ROC-AUC and PR-AUC.

In the case of class imbalance in the dataset, SMOTE is used only on the training subset and only after the train-test split. Importantly, LBP histograms are computed before the oversampling stage in order not to incorporate artificial local sequential patterns.

3.4 Cross-dataset evaluation and sensitivity analysis

To explore the performance characteristics of the proposed approach, including its robustness and limitations, a cross-dataset analysis is conducted using LBP-KL over three educational datasets with various structural properties, namely SATDAP, Metaverse_Edu_Rich and Stanford-ACT-MOOC, details shown in Table 1.

The assessment also involves a neighborhood size sensitivity test based on P = 4,6,8, and 10. As P grows larger, the size of the neighborhood also grows larger. This results in a higher number of combinations in the LBP. However, larger neighborhoods can result in sparser histograms especially for unbalanced or small datasets. Hence, a sensitivity test was performed to analyze the impact of neighborhood sizes in educational contexts.

Table 1. Summary of datasets used in this study

Dataset	Source	Samples	Dropout / Positive Class	Non-dropout / Negative Class	Feature Setting	Data Type
SATDAP	SATDAP Program / Realinho et al. (2021)	3,630	1,421 (39.1%)	2,209 (60.9%)	34 original features	Static institutional tabular data
Metaverse_Edu_Rich	Kaggle	3,000	1,521 (50.7%)	1,479 (49.3%)	11 original features	Static virtual-learning tabular data
Stanford-ACT-MOOC	Hugging Face / Stanford Project	411,749	4,066 (0.99%)	407,683 (99.01%)	6 predictive features after excluding USERID and ACTIONID	Sequential MOOC event-stream data

The new evaluation procedure also entails repeated experiments, reporting of mean ± SD, and paired statistical significance testing. This is due to the desire to enhance the quality of the experimental findings by making them more rigorous, reliable, and reproducible.

3.5 Complexity and applicability boundaries

The computational complexity of LBP extraction is approximately $O(NP)$, where N is the sequence length and P is the neighborhood size. Histogram construction requires $O(N)$, while the histogram dimensionality grows as 2^P. Thus, even though the framework maintains computational tractability, excessive values of P can lead to sparsely populated and statistically unstable histograms.

In turn, the applicability of the proposed LBP-KL framework also depends on the properties of the dataset itself. In particular, the framework can provide useful results if the features include enough non-binary variation and have a certain behavioral order that can be extracted from the local neighborhoods. At the same time, LBP-KL is not well-suited to work with binary variables, sparse features, extremely short sequences, and other datasets with insufficient local order.

4. RESULTS

This set of experiments seeks to test the novel LBP-KL approach using three educational datasets with various structures: a static institutional dropouts dataset (SATDAP), a static virtual-learning dataset (Metaverse_Edu_Rich), and a large sequence-based MOOC events stream dataset (Stanford-ACT-MOOC). Unlike the proof-of-concept evaluation described in the previous section, in this set of experiments we adopt repeated stratified train/test splits, show results as mean±standard deviation, conduct feature selection only on training sets, apply SMOTE only on training sets if required, and perform neighborhood size sensitivity analysis and statistical tests.

4.1 Datasets description

Dataset 1: In this study, an artificially created dataset from Portugal SATDAP program (Grant POCI-05-5762-FSE-000191) [25] has been used to reduce academic attrition via early academic risks detection. This original dataset consists of 4,427 learner cases and 34 attributes that have identified academic pathways for pre-enrollment and semesters. It should be noted that 797 learners enrolled currently were excluded from the analysis to get only the academic results of the participants since this research is based on educational data mining techniques. Therefore, the sample of this study consists of 3,630 learners, including 2,209 graduates (60.9%) and 1,421 dropouts (39.1%). There are four major attribute categories in the analyzed dataset: academic performance (semester-wise completion, grades, enrollment); demographics (age at enrollment, gender, nationality); socioeconomic (scholarship, tuition fees payment); and macroeconomic (regional unemployment, regional inflation).

Dataset 2: The Metaverse_Edu_Rich dataset has been retrieved from Kaggle, which has 3,000 learner observations with the objective of studying the dropout behaviors in immersive virtual learning environments. Out of the learners, 1,521 learners (51%) have been found as dropouts, whereas 1,479 learners (49%) remained active. It comprises of 11 variables that include the engagement, learning, and platform interaction behavior of the learners, such as UserID serving as the unique ID of the learners; ResourceName being the educational resource in use in the metaverse; Duration_Minutes showing the time spent on the resource; Quiz_Attempts and Quiz_Score_Percent for the quiz performance; Forum_Posts counting the number of times participants. participate in forums; Assignment CompletionRate being the proportion of coursework completion rate; Engagement_Level classified into low, medium, or high; Learning_Style as visual or auditory; Feedback_Score rating the learner's satisfaction level between 1 and 5; IsReturningUser, which determines whether the learner is a returning user that has been engaging with the learning environment before or a fresh user interacting with the system for the first time; and the Dropout_Likelihood, which signifies the likelihood or risk of leaving the course.

Dataset 3: Stanford-ACT-MOOC Dataset [26] was obtained from Hugging Face, consisting of about 411,000 records of user actions that were collected from a MOOC platform. The dataset captures the user actions sequence with timestamp information. Each record is labelled to check if it is the last action performed by the learner before they drop out. This dataset is distributed in three CSV files and then combined into a single unified dataset. It consists of the following attributes: ACTIONID: an identifier of every action taken; USERID: an identifier of the learner; TARGETID: the identifier of the activity interacted with; FEATURE0-FEATURE3: a four-dimensional behavioral feature vector; TIMESTAMP: timestamp in seconds after the session starts; and LABEL: an attribute with value 1 for dropout actions and 0 otherwise.

The data set exhibits an extremely imbalanced distribution. Within the data set, 407,683 observations are associated with class 0, corresponding to continuation events, and 4,066 observations are associated with class 1, which corresponds to dropout events. Therefore, 0.99% of the observations pertain to the rare dropout class. Consequently, the data set is appropriate for testing the behavior of feature selection techniques under highly imbalanced educational event streams.

4.2 Experimental protocol

For each dataset, LBP-KL was compared against baseline approaches of feature selection, which included ANOVA F, Chi-Square, Mutual Information, and Random Forest importance. Regarding Stanford-ACT-MOOC, the new experiment with ACT-MOOC dataset additionally includes the comparison with RFE since the experiment on ACT-MOOC dataset is based on the expanded set of baseline approaches. Top k features were selected by each approach and evaluated on three classifiers: Logistic Regression, Random Forest, and Gradient Boosting.

Feature selection approaches were fitted using only training data splits to prevent information leakage. 10 repeats of stratified 80/20 train/test splits were conducted. Stratified SMOTE was conducted for imbalanced datasets, but only using training data after conducting splits. LBP histogram calculation was performed on raw training data not on the SMOTE generated one. The presented scores are means and standard deviations obtained across multiple runs and classifiers. Used metrics included Accuracy, Recall, Precision, F1 Score, ROC-AUC and PR-AUC. Also, for statistical analysis paired Wilcoxon signed rank test and paired t-test were conducted as shown in Table 2.

Table 2. Experimental protocol

Component	Setting
Train-test split	10 repeated stratified 80/20 splits
Feature selection	Fitted on training data only
LBP neighborhood sizes	P = 4, 6, 8, 10
Main classifiers	Logistic Regression, Random Forest, Gradient Boosting
Baseline selectors	ANOVA F, Chi-Square, Mutual Information, Random Forest importance
Additional baseline	RFE included in the ACT-MOOC run
Imbalance handling	SMOTE applied only on training data when needed
Evaluation metrics	Accuracy, Recall, Precision, F1-score, ROC-AUC, PR-AUC
Statistical tests	Wilcoxon signed-rank test and paired t-test

4.3 Results on SATDAP

SATDAP dataset represents structured institutional scenario with class imbalance. Since the dataset lacks temporal learner events, LBP was implemented by sorting the features based on their values. The binary and low-variance features were omitted in LBP scoring. The value of top-k was chosen as 15, being equal to the greater number of institutional features.

From the findings from the SATDAP experiments and Tables 3 and 4, we note that LBP-KL exhibits robust results; however, the method is not better than the best of the classic methods. MI, ANOVA-F, and RF importance give the best F1 and ROC-AUC scores. On the other hand, LBP-KL is able to extract interpretable non-binary attributes where local distribution differs for dropouts and graduates.

Table 3. SATDAP performance at P = 8, mean ± standard deviation

Feature Selection	Accuracy	Recall	Precision	F1	ROC-AUC	PR-AUC
ANOVA F	0.903 ± 0.013	0.871 ± 0.031	0.880 ± 0.025	0.875 ± 0.018	0.950 ± 0.012	0.948 ± 0.012
Chi-Square	0.885 ± 0.012	0.848 ± 0.027	0.858 ± 0.022	0.852 ± 0.015	0.935 ± 0.013	0.926 ± 0.015
LBP-KL	0.827 ± 0.026	0.820 ± 0.032	0.762 ± 0.051	0.788 ± 0.024	0.903 ± 0.020	0.888 ± 0.026
Mutual Information	0.904 ± 0.011	0.874 ± 0.031	0.882 ± 0.024	0.877 ± 0.015	0.952 ± 0.012	0.950 ± 0.012
Random Forest	0.901 ± 0.013	0.868 ± 0.031	0.878 ± 0.029	0.873 ± 0.016	0.951 ± 0.011	0.950 ± 0.011

Table 4. SATDAP neighborhood-size sensitivity for Local Binary Pattern - Kullback-Leibler (LBP-KL)

P	Accuracy	Recall	Precision	F1	ROC-AUC	PR-AUC
4	0.822 ± 0.027	0.818 ± 0.032	0.755 ± 0.053	0.784 ± 0.023	0.903 ± 0.019	0.888 ± 0.025
6	0.827 ± 0.023	0.813 ± 0.032	0.765 ± 0.047	0.787 ± 0.022	0.903 ± 0.020	0.888 ± 0.026
8	0.827 ± 0.026	0.820 ± 0.032	0.762 ± 0.051	0.788 ± 0.024	0.903 ± 0.020	0.888 ± 0.026
10	0.822 ± 0.026	0.812 ± 0.027	0.757 ± 0.049	0.782 ± 0.025	0.902 ± 0.019	0.887 ± 0.025

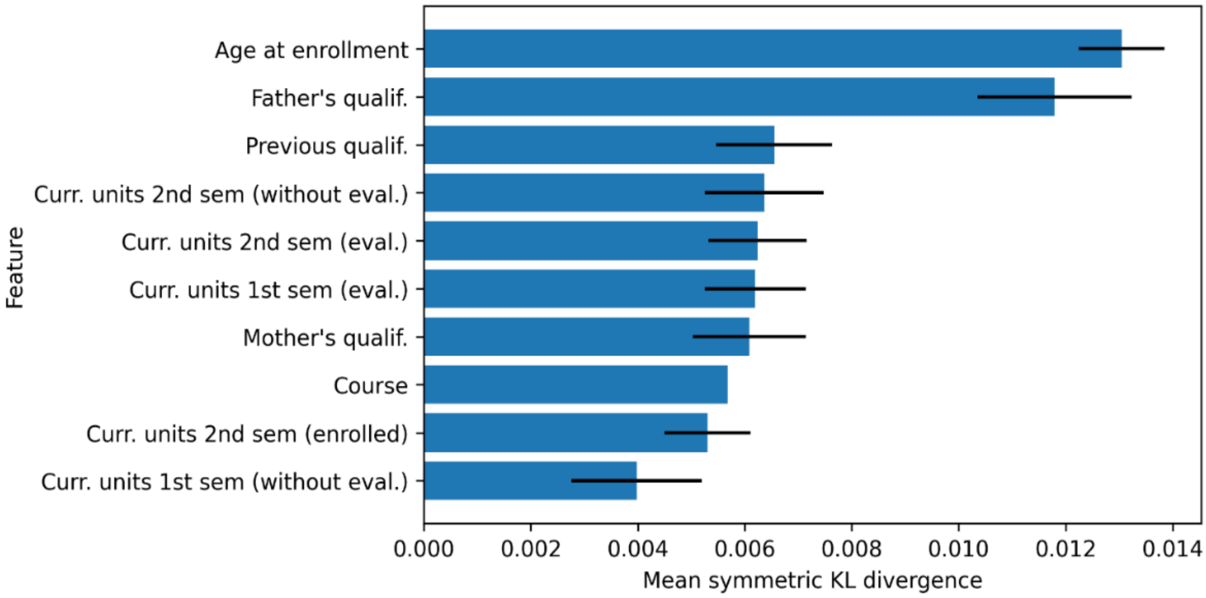

Figure 5. Top Local Binary Pattern - Kullback-Leibler (LBP-KL)-ranked SATDAP features at P = 8

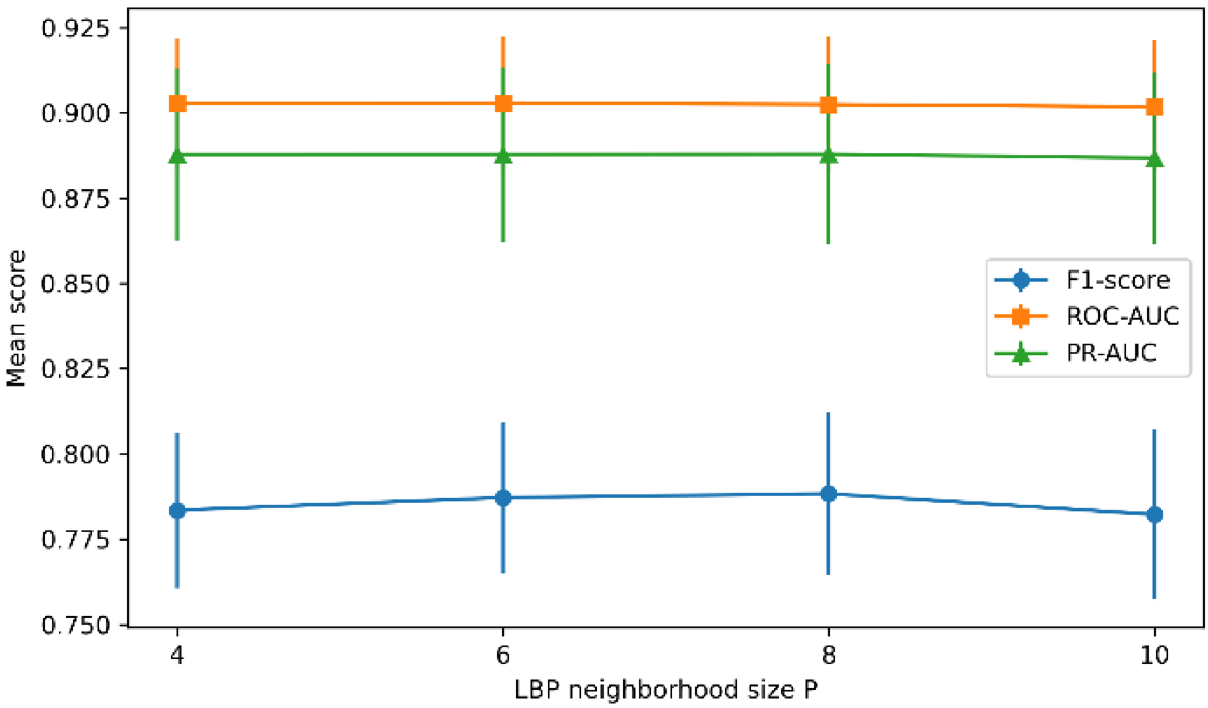

Figure 6. SATDAP P sensitivity plot

The top LBP-KL features identified when P = 8 as illustrated in Figure 5 include Age at enrolment, Father's Qualification, Previous Qualification, Curricular Units 2nd semester (No Evaluations), Curricular Units 2nd semester (Evaluations), Curricular Units 1st Semester (Evaluations), and Mother's Qualification. The results of sensitivity analysis in Figure 6 show that the performance of SATDAP is relatively stable with respect to P values. It yields the maximum F1 score with LBP-KL for P = 8, but P = 6 and P = 4 still yield comparable results. Hence, one can conclude that this approach is not particularly sensitive to the neighborhood sizes for structured static data.

4.4 Results on Stanford-ACT-MOOC

Stanford-ACT-MOOC denotes an imbalanced event-stream scenario. In the new test design, USERID is utilized for organizing temporal streams, but is not considered as a predictor. ACTIONID is not taken into account since it is just an event ID, but the TARGETID is used because it corresponds to the interacted resource.

The reason for choosing k = 4 in top-k is the lack of predictive features.

The proposed ACT-MOOC leakage proof evaluation affects the way the results of the initial experiment are interpreted. LBP-KL is comparable but does not perform better than the best classical selectors in terms of F1-score, ROC-AUC, and PR-AUC. ANOVA F yields the best F1-score and ROC-AUC results whereas Mutual Information yields the best recall result, which, however, is accompanied by very poor precision. It becomes clear that under severe class imbalance, the high recall value cannot serve as an indicator of high predictive power.

Notably, the ranking obtained using the LBP-KL method is still valuable from an interpretational perspective. The first 6 strongest features at P = 8 include: FEATURE1, TIMESTAMP, FEATURE2, FEATURE0, TARGETID, and FEATURE3. It means that even though the behavioral feature vectors' local structure and timestamps contain class-related information, it is not used in making accurate predictions.

As it is seen in the sensitivity analysis and performance in Tables 5 and 6 as well as in Figures 7 and 8, when P = 10, the performance of ACT-MOOC significantly outperforms that of P = 4, 6, and 8. This indicates that small neighborhoods might not be able to capture the beneficial local structures in the data since sparsity is a concern, while a larger neighborhood may capture more information regarding interaction. However, even when using P = 10, the LBP-KL still achieves similar, but not superior, results compared to the other classical selectors.

Table 5. Stanford-ACT-MOOC performance, mean ± standard deviation

Feature Selection	Accuracy	Recall	Precision	F1	ROC-AUC	PR-AUC
ANOVA F	0.935 ± 0.004	0.246 ± 0.021	0.041 ± 0.003	0.070 ± 0.005	0.710 ± 0.021	0.026 ± 0.002
Chi-Square	0.952 ± 0.013	0.168 ± 0.037	0.042 ± 0.010	0.066 ± 0.012	0.687 ± 0.043	0.026 ± 0.006
LBP-KL	0.898 ± 0.033	0.222 ± 0.082	0.023 ± 0.004	0.041 ± 0.006	0.628 ± 0.034	0.017 ± 0.002
Mutual Information	0.798 ± 0.001	0.454 ± 0.017	0.022 ± 0.001	0.043 ± 0.002	0.642 ± 0.009	0.016 ± 0.001
RFE	0.953 ± 0.015	0.163 ± 0.039	0.043 ± 0.010	0.066 ± 0.012	0.686 ± 0.041	0.026 ± 0.005
Random Forest	0.954 ± 0.015	0.159 ± 0.040	0.043 ± 0.010	0.066 ± 0.012	0.686 ± 0.041	0.026 ± 0.005

Table 6. Stanford-ACT-MOOC neighborhood-size sensitivity for Local Binary Pattern - Kullback-Leibler (LBP-KL)

P	Accuracy	Recall	Precision	F1	ROC-AUC	PR-AUC
4	0.898 ± 0.033	0.221 ± 0.082	0.023 ± 0.004	0.041 ± 0.006	0.628 ± 0.034	0.017 ± 0.002
6	0.898 ± 0.033	0.220 ± 0.082	0.023 ± 0.004	0.041 ± 0.006	0.628 ± 0.034	0.017 ± 0.002
8	0.898 ± 0.033	0.222 ± 0.082	0.023 ± 0.004	0.041 ± 0.006	0.628 ± 0.034	0.017 ± 0.002
10	0.952 ± 0.015	0.164 ± 0.038	0.043 ± 0.010	0.066 ± 0.012	0.686 ± 0.041	0.026 ± 0.005

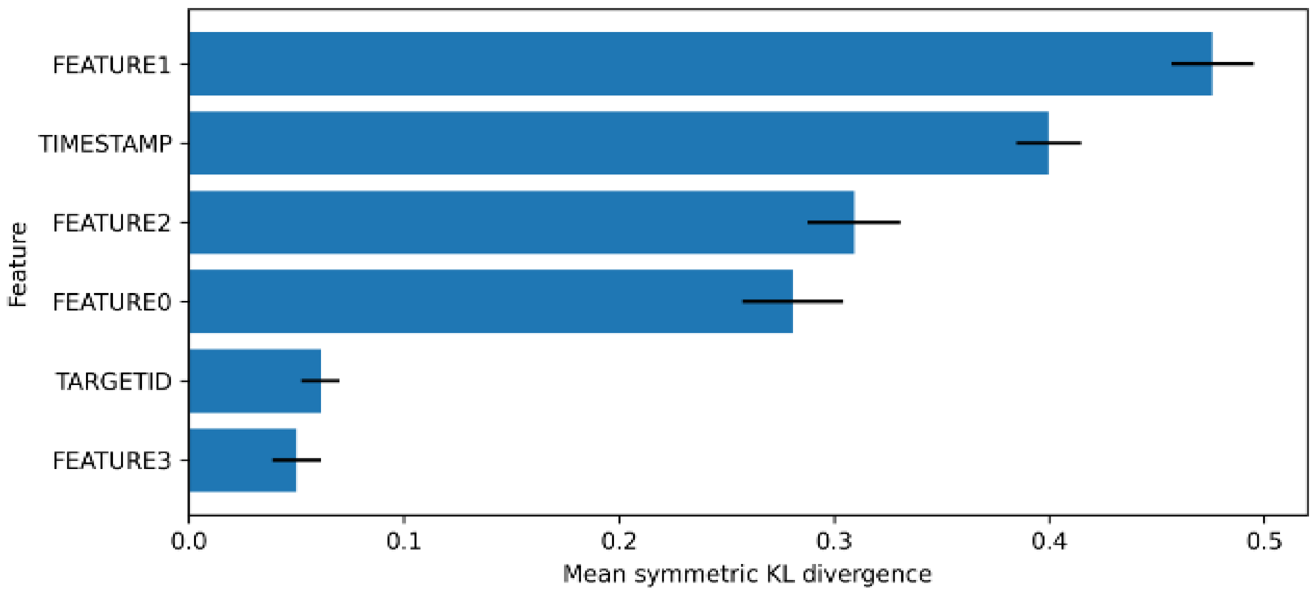

Figure 7. Top Local Binary Pattern - Kullback-Leibler (LBP-KL)-ranked ACT-MOOC features

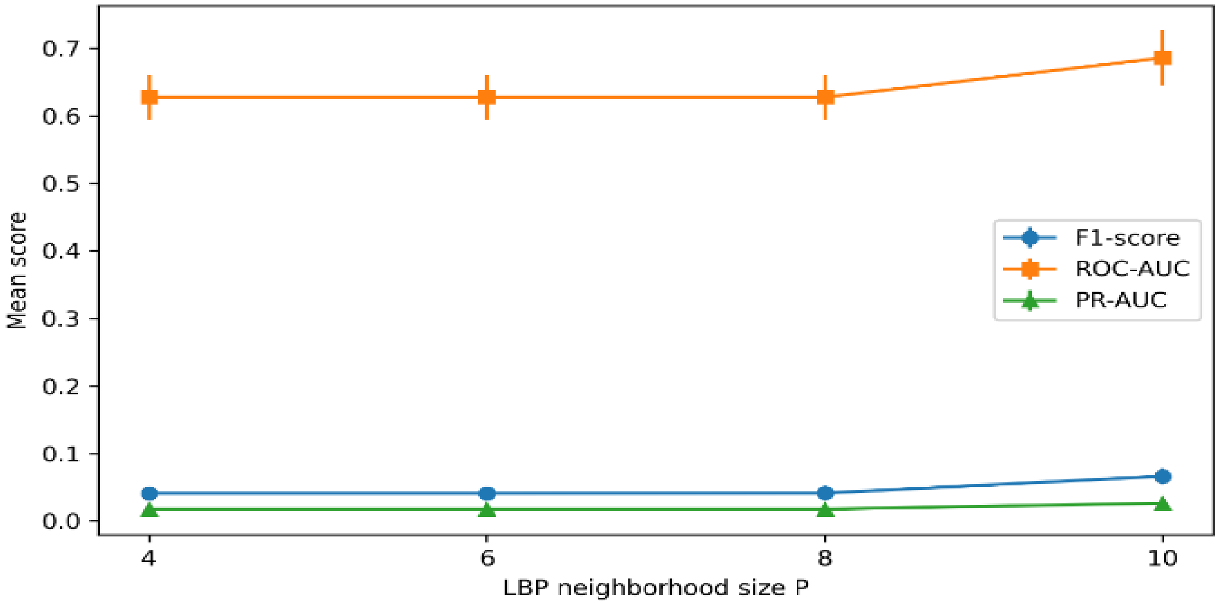

Figure 8. ACT-MOOC P sensitivity plot

4.5 Results on Metaverse_Edu_Rich

The Metaverse_Edu_Rich dataset is a static virtual learning dataset which includes engagement measures, assessment measures, resources utilization, and learner profile variables. Since this dataset lacks temporal sequences of learner events, LBP-KL was used based on sorted value. Binary and low-variance variables were excluded during LBP ranking process. Top-k value is set to 5.

From the Metaverse results in Tables 7 and 8, there is minimal discrimination across all techniques. Accuracy and ROC-AUC metrics are nearly at random level performance whereas F1-score metrics have a strong influence of high recall and balanced class distribution.

LBP-KL is equally effective as Random Forest Importance and the other ranking approaches applied to this data set; yet, it fails to provide any predictive superiority. The top features selected by LBP-KL at P = 8, depicted in Figure 9, are Duration_Minutes, Quiz_Score_ Percent, Assignment Completion_ Rate, Resource Name, and Forum_Posts. However, despite their high rankings, the associated KL-divergences are very low, suggesting insufficient pattern discrimination locally for each pair of classes. In order to understand the impact of the neighborhood parameter P, Figure 10 displays the sensitivity analysis of P on the Metaverse data set. From the figure, we conclude that P = 8 produces the optimal ranking of features.

Based on the sensitivity analysis, Metaverse performs nearly constant over P = 4, 6, 8 and slightly decreases at P = 10. Hence, growing neighborhood sizes cannot detect further improvements in terms of local structural differences in this dataset. Thus, these results can serve as a useful boundary condition regarding the proposed approach - LBP-KL yields little information when features do not exhibit significant local class-dependent patterns.

4.6 Statistical significance and cross-dataset interpretation

In this regard, the significance test results suggest that LBP-KL is comparable and in some cases even better but does not represent a better feature selection tool by default. The negative mean values obtained on SATDAP and ACT-MOOC indicate that the best baseline models are better than LBP-KL regarding the F1-score measure following the improved approach to detecting leakage. The Metaverse_Edu_Rich dataset demonstrates very small values with no statistically significant differences for most cases due to the low discriminative nature of data. Therefore, LBP-KL should be viewed as an additional exploratory tool for pattern analysis rather than the main one.

Overall, a distinct dataset-dependent characteristic has emerged in the course of the experimentation on the three datasets. LBP-KL is most effective when features exhibit sufficient variability of values besides being binary and where class-based histogram profiles vary stably in relation to the class-wise histograms. It would not be efficient where there are too many categorical, weakly ordered, binary, sparse, and imbalanced structures in the dataset.

Average results for each feature selection method have been given in the main document for the sake of comparing feature selection methods across different trials using different classifiers.

Plots for top-ranked LBP-KL features, selected class-wise histograms, and neighbor-size variations have been included in Table 9 for increased interpretation capability and parameter analysis purposes.

Table 7. Metaverse_Edu_Rich performance at P = 8, mean ± standard deviation

Feature Selection	Accuracy	Recall	Precision	F1	ROC-AUC	PR-AUC
ANOVA F	0.508 ± 0.015	0.827 ± 0.126	0.509 ± 0.010	0.627 ± 0.035	0.499 ± 0.021	0.511 ± 0.019
Chi-Square	0.507 ± 0.016	0.825 ± 0.132	0.509 ± 0.011	0.626 ± 0.038	0.495 ± 0.018	0.507 ± 0.016
LBP-KL	0.506 ± 0.012	0.835 ± 0.134	0.507 ± 0.008	0.628 ± 0.040	0.499 ± 0.016	0.509 ± 0.016
Mutual Information	0.509 ± 0.010	0.853 ± 0.112	0.510 ± 0.007	0.635 ± 0.029	0.500 ± 0.024	0.512 ± 0.023
Random Forest	0.506 ± 0.012	0.835 ± 0.134	0.507 ± 0.008	0.628 ± 0.040	0.499 ± 0.016	0.509 ± 0.016

Table 8. Metaverse_Edu_Rich neighborhood-size sensitivity for Local Binary Pattern - Kullback-Leibler (LBP-KL)

P	Accuracy	Recall	Precision	F1	ROC-AUC	PR-AUC
4	0.506 ± 0.012	0.835 ± 0.134	0.507 ± 0.008	0.628 ± 0.040	0.499 ± 0.016	0.509 ± 0.016
6	0.506 ± 0.012	0.835 ± 0.134	0.507 ± 0.008	0.628 ± 0.040	0.499 ± 0.016	0.509 ± 0.016
8	0.506 ± 0.012	0.835 ± 0.134	0.507 ± 0.008	0.628 ± 0.040	0.499 ± 0.016	0.509 ± 0.016
10	0.505 ± 0.014	0.812 ± 0.142	0.507 ± 0.010	0.620 ± 0.042	0.497 ± 0.022	0.507 ± 0.021

Table 9. Paired significance tests comparing Local Binary Pattern - Kullback-Leibler (LBP-KL) against baselines at P = 8

Dataset	Metric	Baseline	Mean Difference: LBP-KL − Baseline	Wilcoxon *p*-Value	Paired *t*-Test *p*-Value
SATDAP	F1	ANOVA F	-0.086	0.00195	<0.001
SATDAP	F1	Chi-Square	-0.064	0.00195	<0.001
SATDAP	F1	Mutual Information	-0.089	0.00195	<0.001
SATDAP	F1	Random Forest	-0.084	0.00195	<0.001
ACT-MOOC	F1	ANOVA F	-0.029	0.00195	<0.001
ACT-MOOC	F1	Chi-Square	-0.025	0.00195	<0.001
ACT-MOOC	F1	Mutual Information	-0.001	0.01367	0.009
ACT-MOOC	F1	RFE	-0.025	0.00195	<0.001
ACT-MOOC	F1	Random Forest	-0.025	0.00195	<0.001
Metaverse	F1	ANOVA F	+0.001	0.69531	0.874
Metaverse	F1	Chi-Square	+0.002	0.62500	0.673
Metaverse	F1	Mutual Information	-0.008	0.06445	0.056
Metaverse	F1	Random Forest	0.000	1.00000	N/A

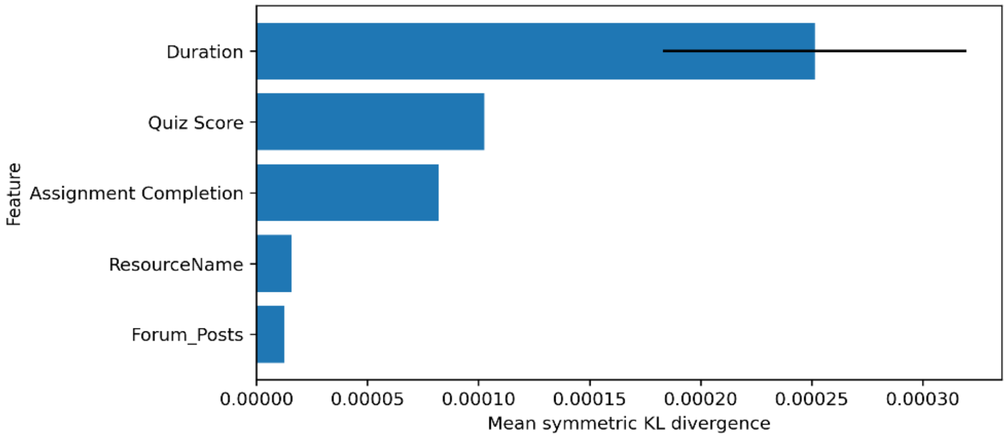

Figure 9. Top Local Binary Pattern - Kullback-Leibler (LBP-KL)-ranked Metaverse features

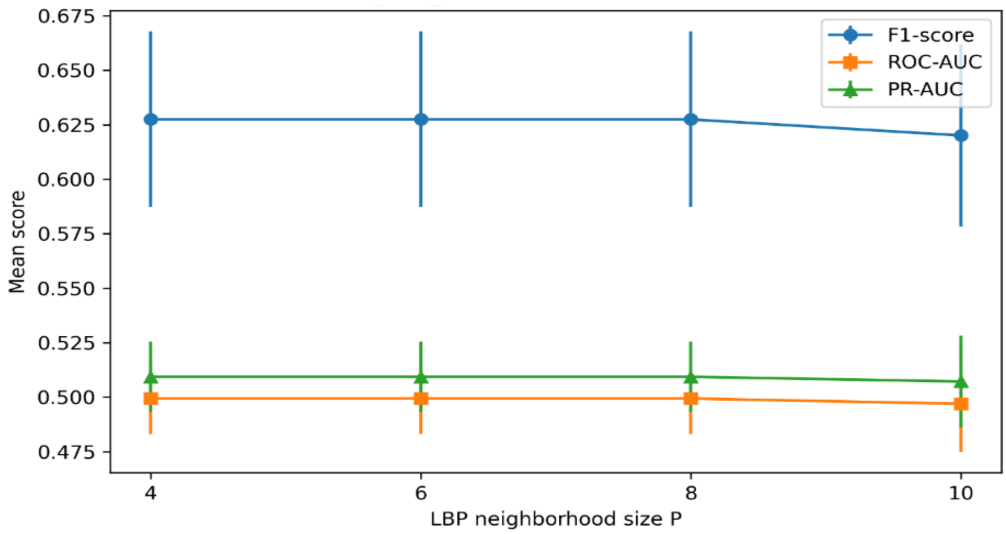

Figure 10. Metaverse P sensitivity plot

5. DISCUSSION

In this study, an adaptation of LBP is introduced as a means to extract significant temporal or structural patterns from sequential or pseudo-sequential educational data. Unlike traditional statistical descriptors or classical feature selection methods, which often depend on aggregate or scalar representations (For instance: mean, standard deviation), LBP captures micro-patterns in the data sequences of increases or decreases which are particularly useful in modeling complex learner behaviors. Below, several key applications and benefits of the proposed LBP framework in the context of dropout prediction and educational data analysis are presented.

5.1 What Local Binary Pattern - Kullback-Leibler adds

LBP-KL provides an interpretable perspective on the local structure of the feature space. In contrast to the other univariate statistical selector, it does not simply investigate whether there is a global difference in the values themselves. Rather, it investigates whether there is a global difference in the local structure of the values as reflected by pattern histograms.

5.2 Why performance differs across datasets

From the cross-dataset performance results, LBP-KL is dependent on the structure of the dataset. For structured static data like SATDAP, there are several academic and demographic variables that include sufficiently structured order within their values for LBP histograms to capture significant differences. For ACT-MOOC, the highly unbalanced nature of the event stream as well as its high sparsity make predicting dropout a challenge, but larger P values compensate for this with broader contextual interaction.

5.3 Interpretability and educational use

The usefulness of LBP-KL comes from providing support for exploratory learning analytics. High LBP-KL values mean that a particular feature displays distinctive local structures between dropouts and non-dropouts. Analysts can examine selected histograms and determine if the difference relates to spikes, drops, concentration around certain values, or even distribution irregularities. It should be noted that it cannot be considered as a guideline for action but serves as an informative base for educational analysis.

5.4 Failure cases and limitations

Several limitations of the approach have been discovered during experiments. First, the LBP method is currently not applicable for binomial features since they do not create valid local neighborhoods. Secondly, when the size of a minority class is small or unbalanced, histogram will have too few bins. Third, increasing the number of points P leads to an exponential growth of codes in the LBP space. Finally, in the case of a tabular dataset, value ordering is just an assumption of analysis and is not time-related.

Synthetic and not well-ordered features may lack the necessary local structure, rendering LBP-KL useless in that case.

5.5 Future work

Future research should explore hybrid LBP methods, which use a combination of binary LBP operators along with local pattern feature vectors, while using adaptation of neighborhood and considering sequence-wise imbalance. Future work may also investigate whether local-pattern representations can support transferability across educational datasets. Another area worth exploring could be mapping high-performing local patterns into interpretable early-warning signals, though this goes beyond the scope of current research.

6. CONCLUSION

This study proposed a one-dimensional LBP-KL framework for interpretable feature ranking for dropout prediction in educational datasets. The approach employs a modified version of Local Binary Pattern to generate patterns from ordered educational feature values and measures the divergence between local histogram distributions of classes using symmetric KL divergence. Compared with the original conceptualization, the study proposes a comprehensive evaluation based on the use of repeated train-test splits, reporting of mean values ± SD, significance testing of results, leakage-free selection of features, training-specific application of SMOTE, and sensitivity analysis to the size of the neighborhood. From the experiments conducted, LBP-KL achieves performance comparable to that of established feature-selection methods but must not be viewed as a better option per se. Instead, the effectiveness of the method depends largely on the dataset at hand, variability in the features, imbalance in the classes present, and relevance of the order chosen. In some instances, the method generates results that match those from traditional selectors, whereas in others it does not.

REFERENCES

[1] Mendoza, J.A.V., Ramírez, C.F., Aranda, C.M. (2024). Analysis and discovery of procrastination patterns in a language learning MOOC. Computers & Education, 223: 105154. https://doi.org/10.1016/j.compedu.2024.105154

[2] Wei, X., Chen, Y., Shen, J., Zhou, L. (2024). Fail or pass? Investigating learning experiences and interactive roles in MOOC discussion board. Computers & Education, 217: 105073. https://doi.org/10.1016/j.compedu.2024.105073

[3] Seo, E.Y., Yang, J., Lee, J.E., So, G. (2025). Prescriptive analytics for student success in an online university: Drawing learning profiles from trace observations for tailored support. Computers & Education, 237: 105384. https://doi.org/10.1016/j.compedu.2025.105384

[4] Wang, Y., Zhao, H., Peng, H. (2025). From 'feel-good' to 'fits-me': Task relevance as the key driver in older adults' video-based learning. Computers & Education, 105420. https://doi.org/10.1016/j.compedu.2025.105420

[5] Yan, L., Martinez-Maldonado, R., Jin, Y., Echeverria, V., Milesi, M., Fan, J., Gašević, D. (2025). The effects of generative AI agents and scaffolding on enhancing students' comprehension of visual learning analytics. Computers & Education, 105322. https://doi.org/10.1016/j.compedu.2025.105322

[6] Padilla Rodriguez, B.C., Armellini, A., Rodriguez Nieto, M.C. (2020). Learner engagement, retention and success: Why size matters in massive open online courses (MOOCs). Open Learning: The Journal of Open, Distance and e-Learning, 35(1): 46-62. https://doi.org/10.1080/02680513.2019.1665503

[7] Liu, H., Zhu, Y., Zang, T., Xu, Y., Yu, J., Tang, F. (2021).

Jointly modeling heterogeneous student behaviors and interactions among multiple prediction tasks. ACM Transactions on Knowledge Discovery from Data (TKDD), 16(1): 1-24. https://doi.org/10.1145/3458023
[8] Al-Shabandar, R., Hussain, A.J., Liatsis, P., Keight, R. (2019). Detecting at-risk students with early interventions using machine learning techniques. IEEE Access, 7: 149464-149478. https://doi.org/10.1109/ACCESS.2019.2943351
[9] Xing, W. (2019). Exploring the influences of MOOC design features on student performance and persistence. Distance Education, 40(1): 98-113. https://doi.org/10.1080/01587919.2018.1553560
[10] Mourdi, Y., Sadgal, M., El Kabtane, H., Berrada Fathi, W. (2019). A machine learning-based methodology to predict learners' dropout, success or failure in MOOCs. International Journal of Web Information Systems, 15(5): 489-509. https://doi.org/10.1108/IJWIS-11-2018-0080
[11] Mourdı, Y., Sadgal, M., Fathı, W.B., El Kabtane, H. (2020). A machine learning based approach to enhance MOOC users' classification. Turkish Online Journal of Distance Education, 21(2): 47-68. https://doi.org/10.17718/tojde.727976
[12] Gupta, S., Sabitha, A.S. (2019). Deciphering the attributes of student retention in massive open online courses using data mining techniques. Education and Information Technologies, 24(3): 1973-1994. https://doi.org/10.1007/s10639-018-9829-9
[13] Qiu, L., Liu, Y., Liu, Y. (2018). An integrated framework with feature selection for dropout prediction in massive open online courses. IEEE Access, 6: 71474-71484. https://doi.org/10.1109/ACCESS.2018.2881275
[14] Gagaoua, I., Labba, C., Brun, A. (2025). Class-imbalanced dynamic feature selection for dropout prediction in virtual learning environments. Procedia Computer Science, 270: 4645-4654. https://doi.org/10.1016/j.procs.2025.09.590
[15] Gao, Y., Sun, X., Wang, X., Guo, S., Feng, J. (2020). A parallel neural network structure for sentiment classification of MOOCs discussion forums. Journal of Intelligent & Fuzzy Systems, 38(4): 4915-4927. https://doi.org/10.3233/JIFS-191572
[16] Rahim, M.A., Hossain, M.N., Wahid, T., Azam, M.S. (2013). Face recognition using local binary patterns (LBP). Global Journal of Computer Science and Technology, 13(4): 1-8.
[17] Yang, B., Chen, S. (2013). A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image. Neurocomputing, 120: 365-379. https://doi.org/10.1016/j.neucom.2012.10.032
[18] Datta Rakshit, R., Nath, S.C., Kisku, D.R. (2017). An improved local pattern descriptor for biometrics face encoding: A LC-LBP approach toward face identification. Journal of the Chinese Institute of Engineers, 40(1): 82-92. https://doi.org/10.1080/02533839.2016.1259020
[19] Hallur, S., Gavade, A. (2025). Image feature extraction techniques: A comprehensive review. Franklin Open, 100366. https://doi.org/10.1016/j.fraope.2025.100366
[20] Sereewatanapong, K., Suebsang, S., Champathes, A., Laosunthara, A., Jattawa, D., Kantavat, P., Kijsirikul, B. (2025). Few-shot cattle muzzle biometric identification using a two-branch prototype network with adaptive-color local binary pattern and enhanced margin prototype loss. Smart Agricultural Technology, 12: 101199. https://doi.org/10.1016/j.atech.2025.101199
[21] Serhatlioglu, I., Kilic, I., Yaman, O., Kacar, E., Oz, Z.D., Ozdede, M.R., Kelestimur, H. (2025). A new method based on local binary Gaussian pattern for classification of rat estrous cycle stages using smear images. Biomedical Signal Processing and Control, 103: 107390. https://doi.org/10.1016/j.bspc.2024.107390
[22] Kumar, J., Pandey, V., Tiwari, R.K. (2025). Optimizing oral cancer detection: A hybrid feature fusion using local binary pattern and CNN. Procedia Computer Science, 258: 476-486. https://doi.org/10.62762/bish.2025.993395
[23] Khan, M.A., Arif, H., Arshad, H.A., Ijaz, H.M., Meeran, M.T. (2025). Unified computational intelligent framework for medical image smart retrieval using hybrid feature modeling. Kashf Journal of Multidisciplinary Research, 2(11): 34-61. https://doi.org/10.71146/kjmr755
[24] Fadaei, S., Azadimotlagh, M., Rashno, A., Beheshti, A. (2025). A new texture descriptor based on hexagonal local binary pattern for content-based image retrieval. Digital Signal Processing, 161: 105138. https://doi.org/10.1016/j.dsp.2025.105138
[25] Kumar, S., Zhang, X., Leskovec, J. (2019). Predicting dynamic embedding trajectory in temporal interaction networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, pp. 1269-1278. https://doi.org/10.1145/3292500.3330895
[26] Realinho, V., Martins, M. V., Machado, J., Baptista, L. (2021). Predict students' dropout and academic success. UCI Machine Learning Repository, 10: C5MC89. https://doi.org/10.24432/C5MC89