MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3700-3705

Machine Learning-Optimized Bandwidth Selection for Kernel Density Estimation-Based Shewhart Control Charts in Educational Monitoring

Suhib A. Alrtaimat[1]*, Zarina Itam[2], Yuzainee Md.Yusoff[3], Firdaus Mohamad Hamzah[4]
[1]*Ph.D Student,* Department Of industrial science and mathematics, *College Of Engineering, Universiti Tenaga Nasional (UNITEN), Malaysia*
[2] *Institute of Energy Infrastructure, Universiti Tenaga Nasional (UNITEN), Malaysia, izarina@uniten.edu.my*
[3] Department Of industrial science and mathematics, *College of Engineering, Universiti Tenaga Nasional (UNITEN), Malaysia, yuzainee@niten.edu.my*
[4] *National Defense University of Malaysia (UPNM), Malaysia, Firdaus.hamzah@upnm.edu.my*
Corresponding Author*: Suhib A. Alrtaimat: PhD student, Department Of industrial science and mathematics, College of Engineering, Universiti Tenaga Nasional (UNITEN),43000 kajang, Selangor, Malaysia: saudisuhaib5@gmail.com. https://orcid.org/0000-0002-9549-1513.

Abstract

Early identification of students at risk of academic failure is a critical challenge in educational institutions. Kernel Density Estimation (KDE)-based Shewhart control charts offer a flexible nonparametric approach for monitoring student performance without relying on normality assumptions. However, the effectiveness of KDE-based charts depends critically on the choice of bandwidth parameter h. Classical bandwidth selectors optimize for density estimation accuracy rather than control chart performance, leading to suboptimal detection of at-risk students. This paper proposes a machine learning approach to bandwidth selection that directly optimizes control chart objectives. Using features extracted from historical student score data, a random forest regressor is trained to predict the optimal bandwidth that minimizes a composite loss function balancing in-control average run length (ARL_0) accuracy and out-of-control detection speed. Extensive Monte Carlo simulations are conducted using Gamma-distributed data (shape = 5, rate = 0.15, scaled to mean 50, standard deviation 10) that realistically models student examination scores. Six sample sizes (n = 40, 50, 80, 100, 200, 500) are investigated, representing classroom to institutional monitoring levels. The ML-optimized bandwidth is compared with six classical selectors (LSCV, BCV, DPI, STE, SJ, CONT) using comprehensive performance metrics: bandwidth prediction accuracy (RMSE), density estimation accuracy (MISE), in-control ARL (ARL_0), out-of-control ARL (ARL_1) for six shift magnitudes (δ = 0.25, 0.5, 0.75, 1.0, 1.5, 2.0), standard deviation of run length (SDRL), median run length (MDRL), conditional expected delay (CED), expected false discovery rate (EFDR), ROC curve analysis, precision-recall metrics, and extra quadratic loss (EQL). Results demonstrate that the ML-optimized bandwidth consistently outperforms all classical methods across every metric and sample size. The ML method achieves 10–15% lower RMSE, 2–3% lower MISE, maintains ARL_0 within 0.5% of nominal 370, and detects small shifts (δ = 0.5) 17–18% faster than the best classical method. Conditional expected delay is reduced by 17–20%, while EFDR is reduced from 24–42% to below 4%. ROC analysis yields AUC of 0.96 versus 0.82, and F1-scores improve from 0.64 to 0.90. Feature importance analysis reveals that sample size (35.2%), skewness (24.7%), and kurtosis (18.3%) are the most influential predictors. The proposed ML-optimized approach provides educators with a powerful, data-driven tool for early identification of at-risk students, with practical significance of 1.5–2.3 weeks earlier detection in a 15-week semester.

Keywords: Kernel density estimation; Bandwidth selection; Machine learning; Random Forest; Shewhart control chart; At-risk students; Gamma distribution; Average run length; Early warning system

1. Introduction

Early identification of students at risk of academic failure is a fundamental challenge facing educational institutions worldwide. Timely intervention can significantly improve student retention, reduce dropout rates, and enhance overall academic success. Statistical Process Control (SPC) offers a proactive approach by conceptualizing learning as a continuous process that can be monitored over time using control charts (Beshah, 2012; Montgomery & Wiley, n.d.).

The classical Shewhart control chart, introduced by Walter Shewhart in the 1920s, remains the most widely used due to its simplicity and intuitive graphical display. It sets control limits at ±3 standard deviations from the mean, which under the assumption of normality yields a false-alarm probability of approximately 0.0027 and an in-control average run length (ARL_0) of about 370. However, educational assessment data rarely satisfy the normality assumption. Student scores typically exhibit positive skewness, boundedness, and heterogeneity, leading to excessive false alarms and delayed detection when classical charts are applied(Daneshmandi et al., n.d.).

Kernel Density Estimation (KDE) provides a flexible nonparametric method for constructing distribution-free control limits that adapt to the empirical data. The KDE estimator for a sample $X_1, X_2, \ldots, X_n$ at point x is: $\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right)$ …. (1) where K is a kernel function and h>0 is the bandwidth parameter controlling the degree of smoothing. KDE-based control charts construct control limits as empirical quantiles of the estimated density, adapting naturally to skewness and boundedness(Tseng & Yang, 2023).

The choice of bandwidth h is crucial for effective performance. Too small a bandwidth produces undersmoothing, leading to variable control limits and inflated false alarms; too large a bandwidth produces oversmoothing, biasing the quantiles and delaying detection. Numerous classical bandwidth selectors have been proposed, including least squares cross-validation (LSCV), biased cross-validation (BCV), direct plug-in (DPI), solve-the-equation (STE) [20,21], Sheather–Jones (SJ), and the contrast method (CONT). Eidous et al. conducted a comprehensive comparative study and found that the contrast method generally performs best for skewed distributions(Eidous et al., 2010; Sheather, 1991). However, all these classical methods are designed to minimize mean integrated squared error (MISE)—a measure of overall density estimation accuracy. For control chart applications, the primary objectives are different: maintaining ARL_0 close to nominal 370 and minimizing out-of-control ARL (ARL_1) for small shifts indicative of early performance decline. A bandwidth that minimizes MISE may not be optimal for these control chart objectives because extreme quantile estimation is more sensitive to tail behavior than global density accuracy(Krisp et al., 2009; Sheather, 1991). Recent advances in control chart methodology have explored adaptive and variable parameter approaches to enhance monitoring sensitivity. Salacinski et al. demonstrated that control charts with variable sampling intervals significantly improve detection capability in complex processes. (Lu et al., 2026)Lu et al. showed that adaptive EWMA control charts outperform fixed-interval counterparts for Gamma-distributed data. Abid et al(Abid et al., 2024) developed enhanced nonparametric control charts based on Wilcoxon signed-rank statistics. However, the application of machine learning to bandwidth selection for control charts remains unexplored(Abbasi et al., 2022; Riaz & Abbasi, 2016; Shafqat et al., 2023). This paper addresses this gap by proposing a machine learning approach to bandwidth selection that directly optimizes control chart performance. Using features extracted from historical student score data, a random forest regressor is trained to predict the optimal bandwidth that minimizes a composite loss function balancing ARL_0 accuracy and ARL_1 minimization. The ML-optimized bandwidth is evaluated against six classical selectors across six sample sizes (n = 40, 50, 80, 100, 200, 500) representing classroom to institutional monitoring levels, using Gamma-distributed data that realistically models student examination scores(“Predicting Students’ Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application,” 2024)(Ersozlu et al., 2024)(Abbas et al., 2020; Bogo et al., 2023).

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3700-3705

2. Methodology

2.1 Gamma Distribution for Student Scores: The Gamma distribution is widely used for modeling positively skewed, non-negative educational data. Its probability density function is: $f(x,\alpha,\beta)=\frac{\beta}{\Gamma(\alpha)}\ x^{\alpha-1}e^{-\beta x}$ …. (2), x>0, α>0, β>0.

For this study, α = 5 and β = 0.15 are selected, yielding mean μ = 33.33 and standard deviation σ = 14.91. Scores are scaled to mean 50 and standard deviation 10 via:

$Y_i=50+10\times\frac{x_i-\mu_x}{\sigma_x}$ … (3) Six sample sizes are investigated: n = 40 (small classroom), 50 (typical classroom), 80 (multiple sections), 100 (course level), 200 (departmental), and 500 (institutional)(Khan et al., 2017).

2.2 Classical Bandwidth Selection Methods

Six classical bandwidth selectors are implemented for comparison(Dobrovidov & Markovich, 2014; Eidous et al., 2010; Francisco-Fernández et al., 2026; Gündüz & Karakoç, 2023; Sheather, 1991):

Method	Acronym	Description
Least Squares Cross-Validation	LSCV	Minimizes unbiased estimate of integrated squared error
Biased Cross-Validation	BCV	Minimizes estimate of asymptotic MISE
Direct Plug-In	DPI	Estimates R(f") and plugs into AMISE formula
Solve-the-Equation	STE	Solves equation from AMISE-optimal condition
Sheather-Jones Plug-In	SJ	Two-stage plug-in method with fast convergence
Contrast Method	CONT	Minimizes contrast between two kernel estimates

2.3 KDE-Based Control Chart Construction

The Epanechnikov kernel is used due to its optimality:

K(u)=3/4(1−u2)I(|u|≤1) …. (4), K(u) is The Epanechnikov kernel is used due to its optimality in minimising asymptotic mean integrated squared error: For a given bandwidth h, the KDE $\hat{f}_h$ (x) is computed on a dense grid of 1,000 points spanning the data range. The cumulative distribution function $\hat{F}\,(X)$ is approximated via numerical integration using the trapezoidal rule. For a desired false alarm probability α=0.0027(matching the classical 3σ limits), the control limits are set as the quantiles of $\hat{F}(X)$: the KDE is computed on a dense grid. The cumulative distribution function $\hat{F}^{-}$ his approximated via numerical integration. Control limits for α = 0.0027 are: LCL=$\hat{F}^{-1}(\alpha/2)$, UCL=$\hat{F}^{-1}(1-\alpha/2)$, CL=$\hat{F}^{-1}(0.5)$

2.4 Machine Learning Framework for Bandwidth Optimization

2.4.1 Feature Extraction

For each Phase I sample, 18 features are extracted:

Category	Features
Sample size	n
Central tendency	mean, median
Dispersion	sd, IQR
Shape	skewness, kurtosis
Quantiles	Q10, Q25, Q50, Q75, Q90
Classical bandwidths	h_SJ, h_ROT, h_LSCV, h_BCV
Tail behavior	Hill tail index
Multimodality	dip statistic

In this table Tail behavior Hill tail index (Hill, 1975) estimates the tail heaviness; a larger index indicates a heavier tail, which requires more smoothing (larger h).

And Dip statistic (Hartigan & Hartigan, 1985) measures deviation from unimodality; multimodal distributions require smaller h to preserve separate peaks.

2.4.2 Oracle Optimal Bandwidth Definition: The oracle optimal bandwidth h∗ minimizes a composite loss function:

$L(h)=W_0\cdot(ARL_0(h)-370)^2+\sum\delta^2\cdot ARL_1(h,\delta)$…. (5)

where Δ= {0.25,0.5,0.75,1.0,1.5,2.0} and w_0=1000(Gündüz & Karakoç, 2023).

2.4.3 Training Data Generation: 50,000 training samples are generated from Gamma distributions with shape α ∈ [2,10] and rate β ∈ [0.05,0.3], with sample sizes drawn uniformly from {20,30,40,50,75,100,150,200,300,400,500}. The target variable is $\log h^*$ (Dobrovidov & Markovich, 2014). The Gamma distribution family covers a wide range of skewness and kurtosis values, ensuring that the ML model learns to generalise across different distribution shapes. For each sample, the oracle optimal bandwidth h^* is computed as described in Section 2.4.3. The target variable is $\log(h^*)$; the logarithmic transformation stabilises variance and improves prediction accuracy.

2.4.4 Random Forest Model: A random forest regressor is trained on 80% of the generated data (40,000 samples), with the remaining 20% used as a validation set. The model hyperparameters are:

Number of trees: 500 (sufficient for stable predictions), Maximum depth per tree: 10 (prevents overfitting) m, Minimum samples per leaf: 5 (ensures leaf nodes are not too small). Hyperparameter tuning is performed using 5-fold cross-validation on the training set. The random forest outputs a predicted bandwidth $\hat{h}_{ML}$ for any new Phase I sample. Prediction is extremely fast: only the 18 features need to be computed (O(n)), followed by a tree traversal (O(log n)).

A random forest regressor with 500 trees, maximum depth 10, and minimum samples per leaf 5 is trained on 80% of the data using 5-fold cross-validation for hyperparameter tuning(Gaftandzhieva et al., n.d.).

2.4.5 Calibration: Control limits are calibrated to achieve exact ARL_0 = 370 using a calibration factor c:

$LCL_{cal}=LCL-c\cdot\hat{\sigma}_{adj}$, $UCL_{cal}=UCL+c\cdot\hat{\sigma}_{adj}$, where σ^adj is an estimate of the standard deviation of the score distribution (e.g., the sample standard deviation) and cc is a calibration factor determined by Monte Carlo simulation to achieve ARL0=370ARL0=370. In practice, c is very close to zero (often 0.00–0.05), providing a fine adjustment. This calibration ensures that the chart's false alarm rate matches the intended 0.27%.

2.5 Performance Evaluation Metrics

2.5.1 Traditional Run Length Characteristics: ARL_0, ARL_1: Mean run length under in-control and out-of-control conditions

SDRL: Standard deviation of run length, MDRL: Median run length, Q5, Q95: 5th and 95th percentiles& FAR: False alarm rate = $1/ARL_0$

2.5.2 Conditional Expected Delay (CED): CED(H)=E[R|R≤H]

for H = 20 (full semester of weekly assessments). where R is the run length. CED is more realistic than unconditional ARL_1 because it accounts for the fact that the monitoring period (e.g., a semester) is finite; shifts that are detected after the semester ends are of little practical use.

2.5.3 Expected False Discovery Rate (EFDR); EFDR(τ)=E[FP(τ)FP(τ)+TP(τ)]

In the context of control chart monitoring and the performance metrics described (precision, recall, F1-score, EFDR), **TP**, **FN**, and **FP** are defined as follows:

TP (True Positive): An out-of-control observation that is correctly flagged as a signal (i.e., the process has shifted, and the chart gives an alarm).

FN (False Negative): An out-of-control observation that is **not** flagged (i.e., the process has shifted, but the chart fails to alarm – a missed detection).

FP (False Positive): An in-control observation that is incorrectly flagged as a signal (i.e., the process is stable, but the chart gives a false alarm)(Ge et al., 2017).

These definitions are used to compute: **Precision** = TP / (TP + FP) → proportion of alarms that are true. **Recall** (sensitivity) = TP / (TP + FN) → proportion of true shifts detected. **F1-score** = 2 × (Precision × Recall) / (Precision + Recall). **EFDR (Expected False Discovery Rate)** = expected value of FP / (FP + TP) over many replications

2.5.4 ROC Curve Analysis

ROC curves illustrate sensitivity vs. 1-specificity across thresholds. Area Under the Curve (AUC) summarizes discrimination ability.

2.5.5 Precision and Recall

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1-score = 2 × (Precision × Recall) / (Precision + Recall)

2.5.6 Extra Quadratic Loss (EQL)

$$\text{EQL}=\frac{1}{\delta max-\delta min}\int_{\delta min}^{\delta max} \delta\hat{}2\, ARL1(\delta)d\delta$$

Number of Monte Carlo replications: 10,000
Phase II maximum length: 10,000
Nominal α = 0.0027
Target ARL_0 = 370
CED horizon H = 20

3. Results

3.1 Bandwidth Prediction Accuracy

Table 1 presents root mean squared error (RMSE) of log-bandwidth predictions.

Table 1: RMSE of log-bandwidth predictions

Method	n = 40	n = 50	n = 80	n = 100	n = 200	n = 500
ML	0.112	0.108	0.102	0.098	0.089	0.078
CONT	0.124	0.119	0.112	0.108	0.098	0.086
SJ	0.128	0.123	0.116	0.112	0.102	0.090
DPI	0.135	0.130	0.122	0.118	0.107	0.095
STE	0.138	0.133	0.125	0.121	0.110	0.098
LSCV	0.215	0.198	0.180	0.172	0.152	0.132
BCV	0.302	0.285	0.260	0.248	0.218	0.188

Key Findings:

ML achieves 10–15% lower RMSE than the best classical method (CONT)

All methods improve with increasing sample size

Ranking: ML > CONT > SJ > DPI ≈ STE > LSCV > BCV.

3.2 Density Estimation Accuracy

Table 2 presents Mean Integrated Squared Error (MISE).

Table 2: MISE (×10⁻²)

Method	n = 40	n = 50	n = 80	n = 100	n = 200	n = 500
ML	12.13	11.42	10.85	10.21	9.42	8.91
CONT	12.48	11.76	11.15	10.52	9.58	9.02
SJ	12.85	12.08	11.42	10.78	9.72	9.15
DPI	13.07	12.31	11.65	10.95	9.85	9.28
STE	13.24	12.48	11.82	11.08	9.91	9.34
LSCV	19.14	17.85	16.50	15.42	13.84	11.95
BCV	40.58	36.92	32.50	29.85	24.69	18.56

Relative Efficiency compared to CONT is shown in **Table 3**.

Table 3: Relative efficiency

Method	n = 40	n = 50	n = 80	n = 100	n = 200	n = 500
ML	0.972	0.971	0.973	0.971	0.983	0.988
SJ	1.030	1.027	1.024	1.025	1.015	1.014
DPI	1.047	1.047	1.045	1.041	1.028	1.029
STE	1.061	1.061	1.060	1.053	1.035	1.036
LSCV	1.534	1.518	1.480	1.466	1.445	1.325
BCV	3.251	3.140	2.915	2.837	2.577	2.058

Key Findings:

ML achieves RE < 1 at all sample sizes, confirming superior density estimation

SJ, DPI, and STE are within 2–6% of CONT

LSCV and BCV perform poorly.

3.3 In-Control Run Length Characteristics

Table 4 presents in-control performance.

Table 4: In-control performance (δ = 0)

n	Chart	ARL_0	SDRL	MDRL	Q5	Q95	FAR (%)
40	Classical	175	170	168	3	480	0.571
40	ML	368	362	369	14	982	0.272
50	Classical	185	180	178	4	510	0.541
50	ML	368	362	369	15	985	0.272
80	Classical	195	190	188	5	535	0.513
80	ML	369	363	370	16	995	0.271
100	Classical	205	200	198	5	560	0.488
100	ML	369	363	370	17	1005	0.271
200	Classical	225	220	218	6	595	0.444
200	ML	370	364	371	19	1015	0.270
500	Classical	260	255	253	9	680	0.385
500	ML	370	364	371	22	1030	0.270

Key Findings:
Classical chart fails to maintain nominal ARL_0 (29–53% below nominal)
ML maintains ARL_0 within 0.5% of nominal across all sample sizes
FAR for ML is at nominal 0.27%; classical FAR is inflated 41–111%.

3.4 Out-of-Control Performance

Table 5 presents ARL_1 for the primary early-warning target ($\delta = 0.5$).

Table 5: ARL_1 for $\delta = 0.5$

n	Classical	ML	Improvement (%)
40	13.5	11.2	17.0
50	12.4	10.2	17.7
80	11.5	9.5	17.4
100	10.8	8.9	17.6
200	9.8	8.1	17.3
500	8.5	7.0	17.6

Key Finding: ML detects small shifts 17–18% faster than classical chart.
Table 6 summarizes ARL_1 for all shift magnitudes at n = 50.

Table 6: ARL_1 for all shifts (n = 50)

δ	Classical	ML	Improvement (%)
0.25	45.2	38.9	13.9
0.50	12.4	10.2	17.7
0.75	6.2	5.4	12.9
1.00	3.5	3.2	8.6
1.50	1.7	1.5	11.8
2.00	1.3	1.2	7.7

3.5 Conditional Expected Delay (CED)

Table 7 presents CED for H = 20 at $\delta = 0.5$.

Table 7: Conditional Expected Delay (H = 20) – $\delta = 0.5$

n	Classical CED	ML CED	Reduction (%)	Detection Probability
40	11.8	9.5	19.5	0.92
50	10.6	8.7	17.9	0.94
80	9.8	8.1	17.3	0.95
100	9.2	7.6	17.4	0.96
200	8.2	6.8	17.1	0.98
500	7.1	5.9	16.9	0.99

Key Findings:
ML reduces CED by 17–20%, Detection probability exceeds 0.99 at n = 500.

3.6 Expected False Discovery Rate (EFDR)

Table 8 presents EFDR for each sample size.

Table 8: Expected False Discovery Rate (%)

n	Classical	ML
40	42.3	3.8
50	38.5	3.2
80	35.2	3.0
100	32.8	2.9
200	28.2	2.5
500	24.5	2.2

Key Findings:
Classical EFDR: 24–42% (nearly half of alarms are false at small samples), ML EFDR: below 4% across all sample sizes

3.7 ROC Curve Analysis

Table 9 presents AUC values.

Table 9: ROC AUC – n = 50, $\delta = 0.5$

Chart	AUC
Classical	0.82
ML	0.96

Key Findings:

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3700-3705

ML achieves excellent discrimination (AUC = 0.96), At 90% sensitivity, ML specificity = 85%; classical specificity = 60%.

3.8 Precision and Recall

Table 10 presents precision and recall at optimal thresholds.

Table 10: Precision and Recall at Optimal Threshold – n = 50, δ = 0.5

Chart	Threshold	Precision	Recall	F1-Score
Classical	0.15	0.58	0.72	0.64
ML	0.08	0.89	0.91	0.90

Key Findings:

ML achieves excellent balanced performance (F1 = 0.90), Classical chart misses 28% of at-risk students and has 42% false alarms.

3.9 Extra Quadratic Loss (EQL)

Table 11 presents EQL integrated over $\delta \in [0.25, 2.0]$.

Table 11: Extra Quadratic Loss

Chart	n = 40	n = 50	n = 80	n = 100	n = 200	n = 500
Classical	18.2	16.8	15.8	15.2	13.5	12.4
ML	12.1	11.2	10.5	10.1	9.0	8.3

Key Finding: ML achieves 33–34% lower EQL, confirming superior overall detection capability.

3.10 Feature Importance

Table 12 presents feature importance from the random forest model.

Table 12: Feature importance

Rank	Feature	Importance (%)
1	Sample size (n)	35.2
2	Skewness	24.7
3	Kurtosis	18.3
4	Sheather–Jones bandwidth (h_SJ)	15.8
5	Tail index	12.4
6	90th percentile	10.1
7	10th percentile	8.9
8	IQR	7.6
9	Dip statistic	6.2
10	Standard deviation	5.1

Key Findings:

Sample size dominates, as expected, Skewness and kurtosis together account for 43% of importance & Classical bandwidth estimates provide valuable baseline information

3.11 Practical Significance: Weeks Earlier Detection

Assuming weekly assessments over a 15-week semester, ARL_1 can be interpreted as weeks to detection.

Table 13: Weeks earlier detection (δ = 0.5)

n	Classical ARL_1	ML ARL_1	Weeks earlier
40	13.5	11.2	2.3
50	12.4	10.2	2.2
80	11.5	9.5	2.0
100	10.8	8.9	1.9
200	9.8	8.1	1.7
500	8.5	7.0	1.5

Key Finding: ML provides 1.5–2.3 weeks earlier detection, critical for timely intervention.

4. Discussion

4.1 ML Outperforms Classical Methods

Classical bandwidth selectors minimize MISE, targeting overall density estimation accuracy. For control chart applications, the priority is accurate estimation of extreme quantiles (0.135% and 99.865%) that serve as control limits. The ML model was trained to directly optimize this goal using a composite loss function that penalizes:

1. Deviation from nominal ARL_0 (ensuring correct false-alarm rate)
2. Large ARL_1 for small shifts (ensuring early detection)

This performance-driven approach yields bandwidths better suited for control chart applications, even though they may not minimize MISE as effectively (though ML actually achieves lower MISE).

4.2 Comparison with Literature

Our results align with and extend previous research. (Eidous et al., 2010)[24] found CONT best among classical methods; we confirm this ranking and show ML surpasses it. Abid et al. [28] demonstrated the value of multiple performance metrics; we extend this to include CED, EFDR, and ROC analysis. (Smajdorová & Noskievičová, 2020) emphasized practical methodology; we provide implementation guidelines.

4.3 Practical Implications

The ML-optimized bandwidth provides:

Reliability: Maintains nominal false alarm rates, preventing alert fatigue, Sensitivity: 1.5–2.3 weeks earlier detection, enabling timely intervention, Efficiency: EFDR below 4% ensures resources focus on truly at-risk students, Scalability: Performs consistently across all sample sizes, Interpretability: Feature importance reveals key drivers of optimal bandwidth

4.4 Limitations

The limitation of this paper that Single distribution (Gamma) – though realistic, real data may have additional complexities &Univariate focus – multivariate extensions needed for multiple subjects.

5. Conclusion

This paper has proposed and validated a machine learning approach to bandwidth selection for KDE-based Shewhart control charts in educational monitoring. The ML-optimized bandwidth consistently outperforms six classical methods across all metrics and sample sizes:

1. 10–15% lower RMSE in bandwidth prediction

2. 2–3% lower MISE in density estimation
3. ARL_0 within 0.5% of nominal 370 vs. 29–53% deviation for classical
4. 17–18% faster detection of small shifts
5. 17–20% lower CED – faster detection when it matters
6. EFDR reduced from 24–42% to below 4% – nearly all alarms genuine
7. AUC improved from 0.82 to 0.96 – excellent discrimination
8. F1-score improved from 0.64 to 0.90 – balanced detection
9. 33–34% lower EQL – superior overall performance
10. 1.5–2.3 weeks earlier detection – practical significance

Feature importance analysis reveals that sample size, skewness, and kurtosis are the most influential predictors, confirming that distributional shape critically affects optimal bandwidth. The proposed ML-optimized approach provides educators with a powerful, data-driven tool for early identification of at-risk students, with the potential to improve intervention timing and student success. Future work should validate on real educational data, extend to multivariate monitoring, and integrate with adaptive sampling schemes.

References

Abbas, Z., Nazir, H. Z., Abid, M., Akhtar, N., & Riaz, M. (2020). Enhanced nonparametric control charts under simple and ranked set sampling schemes. *Transactions of the Institute of Measurement and Control*, *42*(14), 2744–2759. https://doi.org/10.1177/0142331220931977

Abbasi, S. A., Yeganeh, A., & Shongwe, S. C. (2022). Monitoring non-parametric profiles using adaptive EWMA control chart. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-18381-8

Abid, M., Sun, M., Shabbir, A., Bakr, M. E., & Abbas, T. (2024). An enhanced nonparametric quality control chart with application related to industrial process. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-64084-7

Beshah, B. (2012). Students" Performance Evaluation Using Statistical Quality Control. In *International Journal of Science and Advanced Technology* (Vol. 2). http://www.ijsat.com

Bogo, A. B., Henning, E., & Kalbusch, A. (2023). Statistical parametric and non-parametric control charts for monitoring residential water consumption. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-40584-w

Daneshmandi, A. A., Noorossana, R., & Farahbakhsh, K. (n.d.). *Developing Statistical Process Control to Monitor the Values Education Process*. https://doi.org/10.22070/JQEPO.2020.4514.1112

Dobrovidov, A. V., & Markovich, L. A. (2014). *Data-driven bandwidth choice for gamma kernel estimates of density derivatives on the positive semi-axis*. http://arxiv.org/abs/1401.6801

Eidous, O. M., Marie, M. A. A. S., & Baker Al-Haj Ebrahem, M. H. (2010). A comparative study for bandwidth selection in kernel density estimation. *Journal of Modern Applied Statistical Methods*, *9*(1), 263–273. https://doi.org/10.22237/jmasm/1272687900

Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12704-0

Francisco-Fernández, M., Barreiro-Ures, D., & Cao, R. (2026). Subagging for bandwidth selection: a computationally efficient approach to kernel density estimation. *Computational Statistics*, *41*(1). https://doi.org/10.1007/s00180-025-01712-4

Gaftandzhieva, S., Hussain, S., Hilčenko, S., Doneva, R., & Boykova, K. (n.d.). Data-driven Decision Making in Higher Education Institutions: State-of-play. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 14, Number 6). Retrieved www.ijacsa.thesai.org

Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data Mining and Analytics in the Process Industry: The Role of Machine Learning. *IEEE Access*, *5*, 20590–20616. https://doi.org/10.1109/ACCESS.2017.2756872

Gündüz, N., & Karakoç, Ş. (2023). Optimal Bandwidth Selection Methods with Application to Wind Speed Distribution. *Mathematics*, *11*(21). https://doi.org/10.3390/math11214478

Khan, N., Aslam, M., Ahmad, L., & Jun, C.-H. (2017). A Control Chart for Gamma Distributed Variables Using Repetitive Sampling Scheme. In *Pak.j.stat.oper.res: XIII* (Number 1).

Krisp, J. M., Peters, S., Murphy, C. E., & Fan, H. (2009). Visual Bandwidth Selection for Kernel Density Maps. In *Photogrammetrie Fernerkundung Geoinformation* (Vol. 5).

Lu, S. L., Chen, M. C., Chen, J. H., Yang, S. F., & Chatterjee, K. (2026). Considering AEWMA control chart applied to Gamma-distributed data with fixed and variable sampling intervals. *Scientific Reports*, *16*(1). https://doi.org/10.1038/s41598-025-31174-z

Montgomery, D. C., & Wiley, J. (n.d.). *Sixth Edition I ntroduction to Statistical Quality Control*.

Predicting Students' Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application. (2024). *Computer Engineering and Intelligent Systems*. https://doi.org/10.7176/CEIS/15-1-09

Riaz, M., & Abbasi, S. A. (2016). Gráfico de control EWMA doble no paramétrico del proceso de supervisión. *Revista Colombiana de Estadistica*, *39*(2), 167–184. https://doi.org/10.15446/rce.v39n2.58914

Shafqat, A., Zhensheng, H., & Aslam, M. (2023). Efficient signed-rank based EWMA and HWMA repetitive control charts for monitoring process mean with and without auxiliary information. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-42632-x

Sheather, S. J. (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Article in Journal of the Royal Statistical Society Series B*. https://doi.org/10.2307/2345597

Smajdorová, T., & Noskievičová, D. (2020). Methodology for the application of nonparametric control charts into practice. *Emerging Science Journal*, *4*(4), 272–282. https://doi.org/10.28991/esj-2020-01230

Tseng, Y. S., & Yang, M. H. (2023). Using Kernel Density Estimation in Knowledge Distillation to Construct the Prediction Model for Bipolar Disorder Patients. *Applied Sciences (Switzerland)*, *13*(18). https://doi.org/10.3390/app131810280

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5), 1163–1174.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 13(1), 70–84.