DOI: 10.5281/zenodo.22996295 **Volume 2 | Issue 1 | ISSN: 3141-643X**

Ktrend - African Journal of Mathematics, Statistics and Computer Science

A Machine-Learning Approach to Predictive Quality Assurance and Accreditation Monitoring in Nigerian Universities

Osemengbe Oyaimare Uddin[1,*], Susan Konyeha[2], Glory Nosawaru Edegbe[1]

[1]Department of Computer Science, Edo State University Iyamho, Uzairue, Edo State, Nigeria
[2]Department of Data Science, University of Benin, Benin City, Edo State, Nigeria
Corresponding author : uddin.osemengbe@edouniversity.edu.ng

Abstract

Programme accreditation by the National Universities Commission (NUC) is the main external quality assurance mechanism in Nigerian universities, but it is periodic and retrospective: weaknesses are usually discovered during the visit rather than before it. This study aimed to develop and evaluate a machine-learning (ML) framework that predicts accreditation outcomes from routinely collected pre-visit indicators so that institutions can remediate deficiencies early. Because programme-level accreditation records are not publicly available, we built a reproducible synthetic panel of 4,321 accreditation visits to 1,515 programmes in 149 simulated universities (2014–2024), with outcomes (full, interim, denied) generated by NUC-style scoring rules. Fifteen indicators covering staffing, curriculum, facilities, library, funding, research and employer rating, together with ownership, discipline and prior status, were used as predictors. Logistic regression (LR), support vector machine, multilayer perceptron, random forest and extreme gradient boosting (XGBoost) were trained on 2014–2021 visits with university-grouped cross-validation and tested on 2022–2024 visits. LR performed best on the temporal test set (accuracy 0.773; macro-F1 0.708, 95% CI 0.669–0.744; area under the curve [AUC] 0.908), exceeding a prior-status rule (macro-F1 0.537). For the binary task of identifying programmes at risk of not receiving full accreditation, LR achieved an AUC of 0.900 and good calibration; the 20% of programmes with the highest predicted risk included 52 of the 53 denied programmes. Proportion of PhD-holding staff and laboratory provision were the most influential predictors. ML-based risk scoring could support continuous, pre-emptive quality monitoring, but validation on real NUC data is required before operational use.

Keywords: Accreditation; Quality assurance; Machine learning; Higher education; Early warning; Nigeria

1. Introduction

The National Universities Commission (NUC) regulates university education in Nigeria. It became a statutory body in 1974 and now oversees more than 270 federal, state and private universities [1]. Its core quality assurance instrument is programme accreditation. Panels of peer assessors visit each academic programme and score it against minimum academic standards covering academic content, staffing, physical facilities, library, funding and employers' rating of graduates [2,3]. A programme receives full accreditation, valid for five academic sessions, when its aggregate score is at least 70% and it reaches at least 70% in each core area. It receives interim accreditation, valid for two sessions, when the aggregate lies between 60% and 69% or any core area falls below 70%. Accreditation is denied when the aggregate falls below 60%, and the programme must then stop admitting students [4].

Accreditation matters to many stakeholders. Students and parents use it as a signal of quality. Employers and professional bodies rely on it when recognising degrees. University managers are rewarded or penalised through admission quotas and reputation [3,5]. Quality, however, is multidimensional and depends on the stakeholder [6], and several studies have questioned whether episodic accreditation delivers lasting improvement. Reported problems include inadequate funding and staffing, last-minute ”window dressing” before visits, borrowed staff and equipment, and institutional politics around the exercise [2,7,8]. The underlying structural difficulty is timing: deficiencies usually come to light at the visit, when there is little time to correct them. Between visits, universities have limited tools for tracking whether a programme is drifting towards interim or denied status.

A precondition for continuous monitoring is that accreditation data are captured and exchanged reliably. In our earlier comparative study, we showed that electronic data interchange (EDI) between university departments and the NUC can replace manual compilation of accreditation returns with automated, faster and more accurate data exchange. That study also showed that EDI reduces errors, enables real-time monitoring, strengthens data security, and improves transparency and accountability in the accreditation process [9]. The present study builds directly on that work. Once accreditation indicators are available in structured, machine-readable form through EDI, the next question is what can be learned from them before a visit takes place.

Educational data mining and learning analytics have shown that institutional data can be turned into early-warning signals, mostly at the level of individual students [10]. Their use for programme-level quality assurance is more recent. Torrents et al. examined how artificial intelligence could predict quality risk in degree programmes for a quality agency and discussed both its potential and its limitations [11]. Kawesha and Phiri built data-mining models to support accreditation decisions at Zambia's Higher Education Authority [12], and Aldosari et al. proposed an explainable deep-learning framework that maps criterion-level evidence to accreditation grades [13]. We found no published study that applies ML to predictive accreditation monitoring in Nigerian universities.

We also found no study that evaluates such models under the conditions a regulator or university would actually face: forecasting future visits from past ones, with programmes clustered within institutions.

This study addresses that gap. Its objectives were to (i) design a predictive quality assurance framework that uses indicators available before an NUC visit; (ii) compare five supervised learning algorithms for predicting three-class accreditation outcomes and a binary "at-risk" outcome under temporal and institution-grouped validation; (iii) identify the indicators that drive predicted risk; and (iv) assess whether predicted risk can be turned into actionable monitoring tiers. Real programme-level accreditation records are not publicly available. We therefore conducted the study on a transparent, reproducible synthetic dataset modelled on the NUC scoring framework and report it as a methodological proof of concept.

2. Materials and methods

2.1. Study design and framework

The study followed a simulation-based predictive modelling design. We reported model development and validation in line with the relevant items of the TRIPOD statement for multivariable prediction models [14]. Fig. 1 shows the proposed framework, which has five layers: data, pre-processing, modelling, evaluation and monitoring. The data layer assumes that pre-visit indicators are collected through the EDI infrastructure we proposed previously [9], so that departmental returns reach the modelling pipeline without manual re-entry. The monitoring output is fed back to departmental and institutional quality assurance units so that they can act before the next accreditation visit.

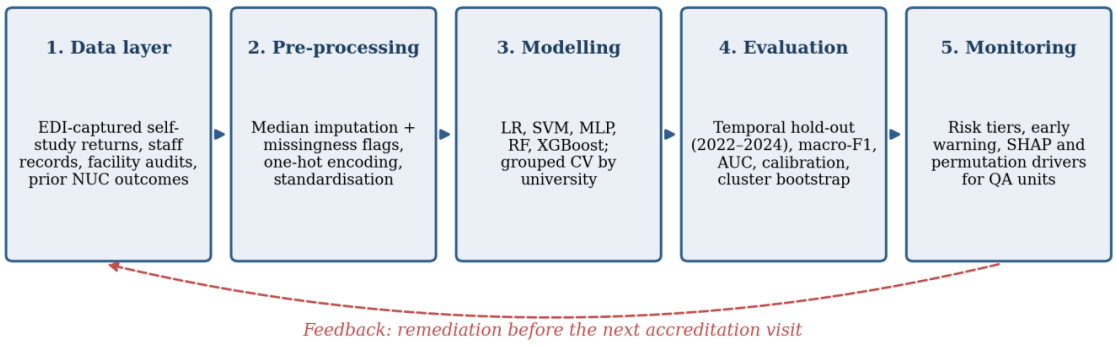

Fig. 1. Proposed predictive quality assurance and accreditation monitoring framework. EDI, electronic data interchange; LR, logistic regression; SVM, support vector machine; MLP, multilayer perceptron; RF, random forest; CV, cross-validation; QA, quality assurance.

2.2. Synthetic data generation

We wrote a simulator in Python 3.11 (NumPy, pandas) with a fixed random seed (2026) so that the dataset can be regenerated exactly. The simulator created 150 universities (40

federal, 45 state and 65 private), each with 6–14 programmes drawn from seven discipline clusters. Each university had a latent institutional capacity, shifted by ownership type. Each programme had a latent quality that was correlated with its university's capacity. Programmes received their first visit between 2014 and 2020, depending on the university's founding year, and were revisited according to the validity periods described above: five years after full accreditation and two years after interim or denied status. After each visit, latent quality changed by a random amount plus a remediation increment that was larger after interim or denied outcomes. This reflects the corrective investment that adverse outcomes usually trigger. One simulated university had no visits before the 2024 cut-off, so the final panel contained 149 universities.

At each visit, the simulator generated 15 observable pre-visit indicators from latent programme quality, institutional capacity, ownership, calendar year and random noise (Table 1). Separately, it generated hidden assessor scores (0–100) for six areas. These scores were combined with fixed simulator weights (staffing 0.32, physical facilities 0.25, academic content 0.23, library 0.12, funding 0.05, employer rating 0.03) and classified with the NUC rules [4]. The weights are simulation assumptions chosen to reflect the emphasis of the NUC instrument; they are not official values. The assessor scores were excluded from the predictors. This prevents target leakage and mirrors the real use case, in which the panel's scores are unknown before the visit. About 5% of values in four self-reported indicators were set to missing, with a slightly higher probability for weaker programmes, to mimic incomplete institutional returns.

Table 1. Predictor variables available before an accreditation visit

Variable	Description	Area
Staff with PhD (%)	Share of full-time academic staff holding a doctorate	Staffing
Student–staff ratio / benchmark	Observed ratio divided by discipline benchmark (10–30)	Staffing
Staff-mix deviation	Absolute deviation from the recommended rank mix	Staffing
Full-time academic staff (n)	Head count of full-time academic staff	Staffing
Curriculum compliance (%)	Coverage of the approved minimum curriculum	Academic content
Exam moderation rate (%)	Share of courses with external moderation	Academic content
Self-study completeness (%)	Completeness of the self-study form	Academic content
Laboratory/facility index	Audit score of laboratories, studios and space (0–100)	Facilities
ICT index	Audit score of ICT infrastructure (0–100)	Facilities
Library index	Composite of holdings, e-resources and seating (0–100)	Library
Funding per student	Recurrent programme spend per student (simulated units)	Funding

Variable	Description	Area
Research output per staff	Peer-reviewed outputs per staff member per year	Research
Employer rating	Mean employer rating of graduates (1–5)	Employer
Functional QA unit	Departmental QA committee active (yes/no)	Governance
Institution age	Years since the university was founded	Context
Ownership	Federal, state or private	Context
Discipline	Seven discipline clusters	Context
Prior status	Previous outcome (none, full, interim, denied)	History

QA, quality assurance; ICT, information and communication technology. All values are simulated.

2.3. Pre-processing

Numerical predictors were imputed with the training-set median, and a binary indicator was added for each variable that had missing values. Categorical predictors were one-hot encoded. For the distance- and gradient-based models (LR, SVM, MLP), numerical predictors were standardised to zero mean and unit variance. All pre-processing steps were fitted inside the modelling pipeline on training folds only, so that no information passed from validation or test data into training.

2.4. Learning algorithms

We compared five algorithms, all implemented in scikit-learn 1.8 [15] or the XGBoost 3.2 library: (i) multinomial LR (L2 penalty, $C = 1$); (ii) a support vector machine (SVM) with a radial basis function kernel ($C = 2$) and Platt-scaled probabilities [16]; (iii) a multilayer perceptron (MLP) with two hidden layers (64 and 32 units), L2 penalty 10^{-3} and early stopping; (iv) a random forest (RF) of 500 trees with a minimum leaf size of 3 [17]; and (v) extreme gradient boosting (XGBoost) with 400 trees, maximum depth 4, learning rate 0.05 and 80% row and column subsampling [18]. The denied class was rare, so class imbalance was handled through balanced class weights (LR, SVM, RF) or balanced sample weights (MLP, XGBoost) rather than synthetic oversampling. Hyperparameters were fixed a priori at conventional values to limit optimistic bias.

2.5. Validation strategy

We used two complementary validation schemes. First, within the training period (visits in 2014–2021; n = 3,071), we ran five-fold stratified group cross-validation in which all programmes of a university fell in the same fold. This prevents the model from learning institution-specific patterns and then being evaluated on the same institution. Second, and as the primary analysis, we trained each model on all 2014–2021 visits and evaluated it on a temporally separate test set of 2022–2024 visits (n = 1,250). This setting reproduces

prospective use, in which a model built on past cycles forecasts the next cycle. Two reference baselines were included: a majority-class rule and a ”prior-status” rule that predicts a programme's previous outcome (full for first visits).

2.6. Performance measures

For the three-class task, we report accuracy, balanced accuracy, macro-averaged F1 score (macro-F1, the primary metric because the classes are imbalanced), quadratically weighted Cohen's kappa (which respects the ordinal nature of the outcome) and macro one-versus-rest area under the receiver operating characteristic curve (AUC). For monitoring, the outcome was also collapsed into a binary ”at-risk” label (interim or denied versus full). The predicted risk was one minus the predicted probability of full accreditation. We summarised it by AUC, average precision (AP), which is informative under imbalance [19], and the Brier score, and we assessed calibration with reliability curves [20]. Ninety-five percent confidence intervals (CIs) were obtained by a cluster bootstrap over universities (1,000 resamples).

2.7. Interpretability and monitoring tiers

To explain predictions, we used Shapley additive explanations (SHAP) computed with the TreeExplainer on the XGBoost model [21] and model-agnostic permutation importance (20 repeats, decrease in macro-F1) on the best-performing model [17]. Importances of one-hot encoded levels were summed back to their parent variable. For operational use, predicted risk from the best model was grouped into three a priori tiers: low (< 0.30), medium (0.30–0.59) and high (0.60). We also report the yield of reviewing only the 20% of programmes with the highest predicted risk. All analyses were run in Python 3.11; the code and synthetic data are available from the corresponding author.

2.8. Ethical considerations

The study used only computer-generated data. No human participants, identifiable institutional records or animals were involved, so ethical approval and informed consent were not required.

3. Results and discussion

3.1. Characteristics of the synthetic panel

The panel contained 4,321 visits to 1,515 programmes in 149 universities (40 federal, 45 state and 64 private). Overall, 2,169 visits (50.2%) resulted in full accreditation, 1,833 (42.4%) in interim accreditation and 319 (7.4%) in denial. Fig. 2 shows the distribution by year. The temporal test set comprised 684 full, 513 interim and 53 denied outcomes. Median indicator values rose steadily from denied to interim to full programmes (Table 2). The largest contrasts were in the share of PhD-holding staff (24.1% vs 47.8% vs 72.2%) and the laboratory/facility index (18.3 vs 41.6 vs 68.2). A functional departmental QA unit was reported in 35.4%, 50.1% and 69.1% of denied, interim and full visits, respectively.

These gradients are consistent with the input-quality deficits that Nigerian studies have repeatedly linked to adverse accreditation outcomes [5,7,8].

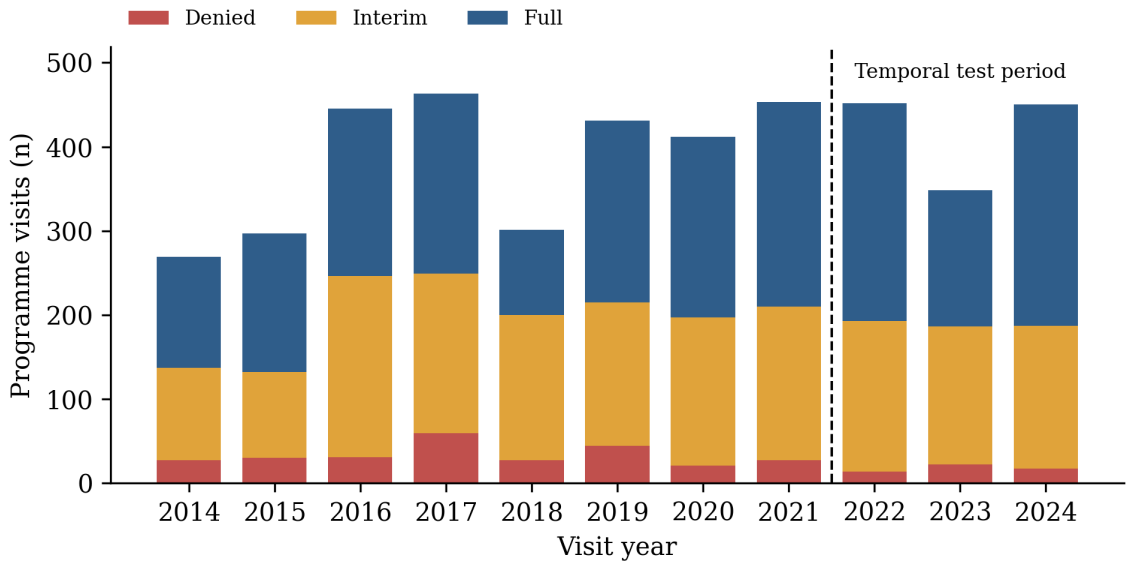

Fig. 2. Accreditation outcomes by visit year in the synthetic panel. The dashed line separates the training period (2014–2021) from the temporal test period (2022–2024).

Table 2. Median values of pre-visit indicators by accreditation outcome (all visits, n = 4,321)

Indicator	**Denied (n = 319)**	**Interim (n = 1,833)**	**Full (n = 2,169)**
Staff with PhD (%)	24.1	47.8	72.2
Student–staff ratio / benchmark	1.69	1.29	0.93
Staff-mix deviation	42.6	30.8	18.2
Full-time academic staff (n)	9	13	19
Curriculum compliance (%)	47.8	68.1	83.8
Exam moderation rate (%)	54.0	70.6	83.4
Self-study completeness (%)	64.3	76.0	86.3
Laboratory/facility index	18.3	41.6	68.2
ICT index	31.7	52.0	70.4
Library index	28.5	47.7	67.0
Funding per student (units)	86.1	138.6	218.3
Research output per staff	0.00	0.60	1.21
Employer rating (1–5)	2.40	2.97	3.53
Functional QA unit (%)[a]	35.4	50.1	69.1

[a] Percentage of visits rather than median. ICT, information and communication technology; QA, quality assurance.

3.2. Model performance

In university-grouped cross-validation on the training period, all five models reached a mean macro-F1 between 0.722 and 0.758 and a macro AUC between 0.911 and 0.923. RF had the highest mean macro-F1 (0.758 ± 0.026), and LR and XGBoost were close behind (both 0.748). Table 3 and Fig. 3 give performance on the 2022–2024 temporal test set. LR achieved the highest macro-F1 (0.708; 95% CI 0.669–0.744), balanced accuracy (0.751), weighted kappa (0.681) and macro AUC (0.908). RF and MLP followed, while SVM and XGBoost scored lowest. The confidence intervals overlapped substantially, so the differences among the five learners were small compared with sampling uncertainty. Every model clearly outperformed both baselines. The prior-status rule reached a macro-F1 of 0.537 and a weighted kappa of 0.414, and the majority-class rule reached a macro-F1 of 0.236.

Table 3. Performance on the temporal test set (visits in 2022–2024, n = 1,250)

Model	Acc.	Bal. acc.	Macro-F1 (95% CI)	$_w$	Macro AUC	Risk AUC	Risk AP	Brier
Logistic regression	0.773	0.751	0.708 (0.669–0.744)	0.681	0.908	0.900	0.888	0.130
Random forest	0.773	0.685	0.701 (0.650–0.744)	0.652	0.907	0.896	0.880	0.132
Multilayer perceptron	0.768	0.725	0.691 (0.646–0.729)	0.672	0.900	0.895	0.882	0.134
XGBoost	0.758	0.679	0.679 (0.630–0.718)	0.640	0.900	0.892	0.877	0.139
Support vector machine	0.764	0.650	0.675 (0.619–0.727)	0.631	0.896	0.886	0.876	0.138
Prior-status rule	0.633	0.545	0.537	0.414	–	–	–	–
Majority class	0.547	0.333	0.236	–	–	–	–	–

Acc., accuracy; Bal. acc., balanced accuracy; CI, confidence interval (cluster bootstrap over universities); $_w$, quadratically weighted Cohen's kappa; AUC, area under the receiver operating characteristic curve; AP, average precision. Risk AUC and risk AP refer to the binary at-risk task (interim or denied versus full accreditation).

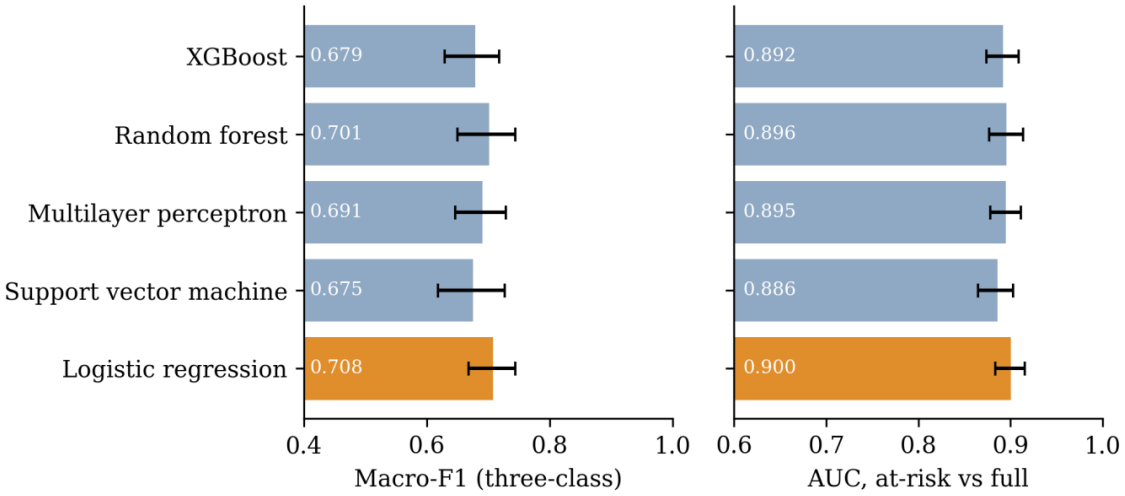

Fig. 3. Temporal test performance of the five learners: three-class macro-F1 (left) and AUC for identifying at-risk programmes (right). Error bars show 95% cluster-bootstrap confidence intervals; the best model is highlighted.

A simple linear model performing as well as the ensembles is a meaningful result. The simulated relationships between indicators and latent quality are mostly monotonic, and in the real accreditation instrument the outcome is also a weighted sum of area scores followed by thresholds. Models that can express additive structure are therefore well suited to the task. This agrees with reports that LR often matches more complex learners on tabular institutional data [10,11]. In practice, a well-specified LR model is also easier for quality assurance officers and regulators to audit and explain.

All models performed worse on the temporal test set than in cross-validation; for LR, macro-F1 fell from 0.748 to 0.708. This decline reflects temporal drift that was built into the simulator. Indicators such as the share of PhD-holding staff, ICT provision and curriculum compliance improve with calendar year, and programmes on repeat visits have already undergone remediation. Random cross-validation therefore overstates the accuracy a regulator would achieve in forecasting the next cycle, and future studies should use temporal validation as the primary estimate.

3.3. Class-level errors

Fig. 4 shows the confusion matrix of the LR model. Full accreditation was recognised well (recall 0.873; precision 0.808), and interim accreditation moderately well (recall 0.643; precision 0.766). The model identified 39 of the 53 denied programmes (recall 0.736) but at a precision of 0.487, because 41 interim programmes were predicted as denied. Almost all errors fell between adjacent categories. No denied programme was predicted as full, and no full programme was predicted as denied. For an early-warning system this pattern is acceptable: it is better to flag some interim programmes for closer attention than to miss a programme heading for denial.

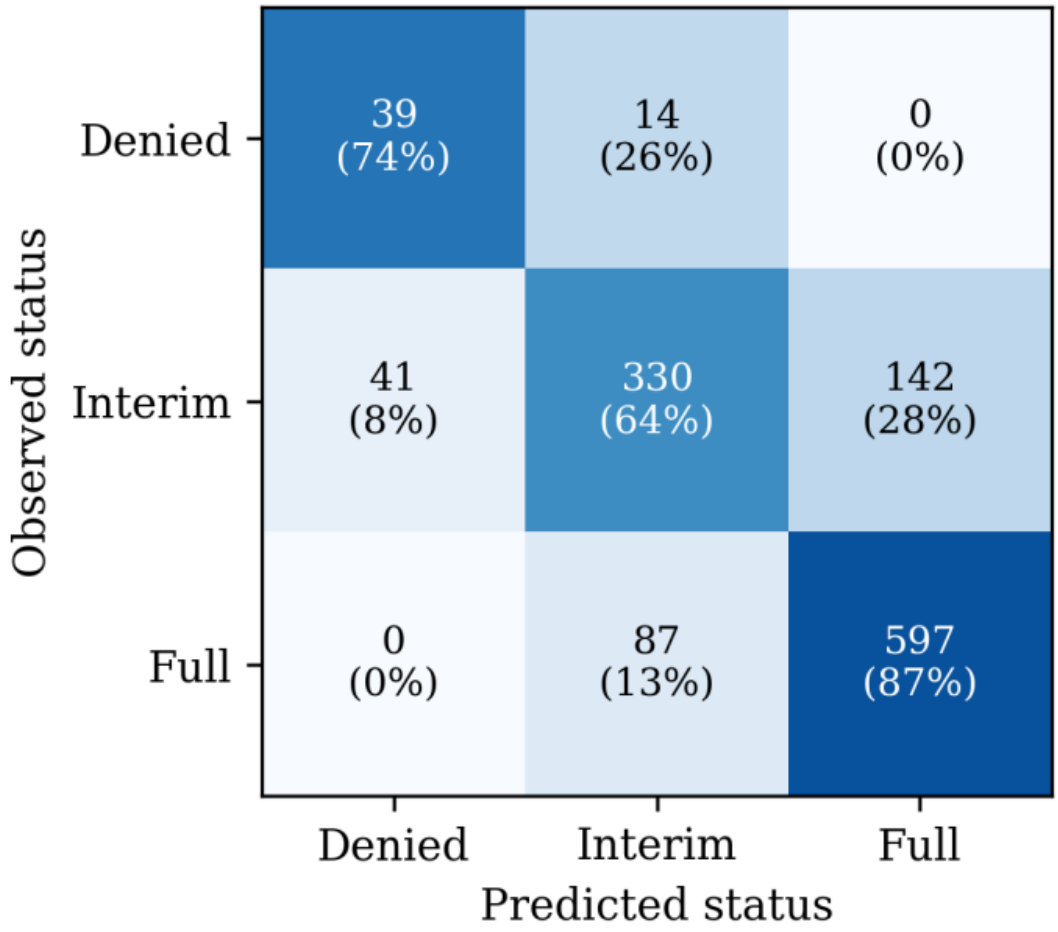

Fig. 4. Confusion matrix of the logistic regression model on the temporal test set. Cells show counts and row percentages (recall).

3.4. Drivers of predicted risk

Fig. 5 ranks the predictors by mean absolute SHAP value from the XGBoost model. The share of staff with a PhD dominated, followed by the laboratory/facility index, staff-mix deviation, ICT index, student–staff ratio relative to benchmark, library index and research output per staff. Contextual variables (ownership, institution age, discipline) contributed little once programme-level indicators were known. Permutation importance for the LR model gave a similar ranking. Removing the PhD share reduced macro-F1 by 0.035, and removing the laboratory index or funding per student reduced it by 0.019 and 0.018, respectively. Because the simulator placed most weight on staffing and facilities, this agreement is partly a check that the explanation methods recovered the data-generating structure. The substantive point is that the resulting explanations are expressed in the units an accreditation panel uses. A dean can be told that a programme's risk is driven mainly by its staffing profile and laboratory provision, which matches the input deficits that Nigerian studies have documented [2,3,7].

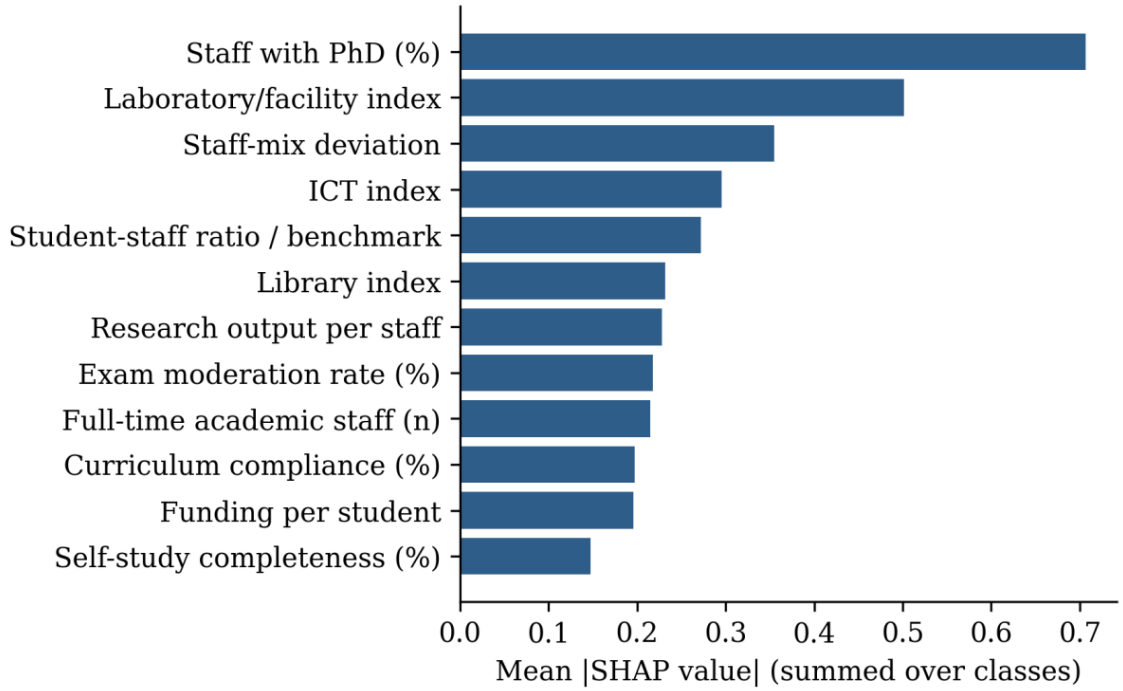

Fig. 5. Predictor importance from the XGBoost model (mean absolute SHAP value summed over the three classes; one-hot levels aggregated to their parent variable).

3.5. Calibration and monitoring tiers

For an early-warning system, predicted probabilities must be trustworthy as well as correctly ranked. The reliability curves for LR, RF and XGBoost lay close to the diagonal (Fig. 6, left), and LR, which discriminated at-risk programmes with an AUC of 0.900 (95% CI 0.884–0.916), had the lowest Brier score (0.130). When LR risk scores were grouped into tiers, the observed proportion of programmes not receiving full accreditation was 14.1% in the low tier (n = 611), 51.4% in the medium tier (n = 183) and 84.6% in the high tier (n = 456) (Fig. 6, right). All 53 denied programmes fell into the high tier. If a QA unit or the regulator reviewed only the 20% of programmes with the highest predicted risk (250 programmes), 94.4% of them would be truly at risk, and 52 of the 53 eventual denials (98.1%) would be covered. Mock accreditation, targeted staff recruitment or facility investment could then be concentrated on a small, well-defined set of programmes instead of being spread evenly or begun only when a visit is announced.

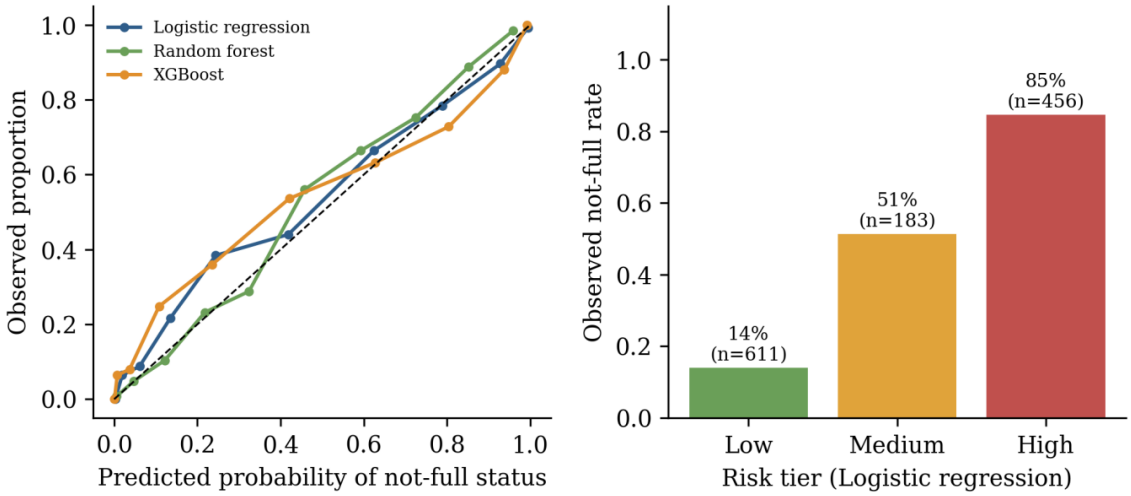

Fig. 6. Left: reliability curves for the predicted probability of not receiving full accreditation (deciles of predicted risk). Right: observed not-full rate by risk tier for the logistic regression model.

3.6. Implications for quality assurance practice

The results suggest three practical uses. First, universities could use the framework internally as a continuous self-assessment tool, updated each session from data they already compile for NUC self-study forms, staff audits and budgets. Combined with the EDI-based data exchange we described earlier [9], risk scores could be refreshed automatically whenever departmental records change, rather than once per accreditation cycle. This would move quality assurance from episodic compliance towards continuous monitoring, as critics of current practice have called for [7,8]. Second, the NUC could use risk scores to prioritise monitoring visits and technical support, directing limited assessor capacity to programmes most likely to fail. Third, explainable outputs could inform resource allocation, for example by linking a programme's risk to its staffing or laboratory gaps. Such tools must support peer judgement, not replace it. Predictions should trigger further inquiry rather than sanctions, and institutions could be tempted to game indicators, a known risk of metric-driven quality regimes [6,11].

3.7. Limitations

The main limitation is that the data are synthetic. The simulator encodes our assumptions about how indicators relate to accreditation outcomes, including the area weights, noise levels, remediation dynamics and missing-data mechanism. The absolute performance reported here therefore shows what is achievable under these assumptions and should not be read as the accuracy expected on real NUC data. Real records are likely to be noisier, to contain misreported values (including the ”window dressing” described in the literature [7,8]) and to be subject to policy changes, such as curriculum reforms, that the simulator does not represent. Second, hyperparameters were not tuned. Tuning might change the ranking slightly but is unlikely to alter the main conclusions, given the overlapping confidence intervals. Third, SHAP values were computed for XGBoost rather than for the best model, although permutation importance for LR gave a consistent rank-

ing. Finally, the denied class was small in the test period (n = 53), so estimates for that class are imprecise. The next step is external validation on anonymised programme-level accreditation records obtained in partnership with the NUC and individual universities.

4. Conclusion

This study presents a predictive quality assurance framework that uses indicators available before an NUC accreditation visit to forecast programme outcomes, and evaluates it under temporal and institution-grouped validation. On a reproducible synthetic panel modelled on the NUC scoring rules, all five learners clearly outperformed the prior-status and majority baselines. Logistic regression performed as well as the more complex ensembles and was well calibrated. A simple three-tier risk score concentrated almost all denials in a small, reviewable group of programmes, and staffing and physical facilities emerged as the main levers for remediation. ML-based monitoring could help Nigerian universities and the regulator move from retrospective, visit-driven compliance towards continuous, pre-emptive quality management. Its value on real accreditation data remains to be established through external validation.

Acknowledgements

The authors thank their respective departments at Edo State University Iyamho and the University of Benin for institutional support. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Use of AI technologies

An AI assistant (Claude, Anthropic) was used to help draft and edit the manuscript text, write the data-simulation and analysis code and produce the figures. The authors designed the study, reviewed and verified all code, results, references and text, and take full responsibility for the content of this article. The AI tool is not listed as an author.

Disclosure of conflict of interest

Osemengbe Oyaimare Uddin: no financial or non-financial conflict of interest to declare.

Susan Konyeha: no financial or non-financial conflict of interest to declare.

Glory Nosawaru Edegbe: no financial or non-financial conflict of interest to declare.

Statement of ethical approval

The present research work does not contain any studies performed on animals or human subjects by any of the authors.

Statement of informed consent

Not applicable. The study used only computer-generated (synthetic) data and involved no individual participants.

Data availability

The simulation code, the synthetic dataset (4,321 records) and the analysis scripts are available from the corresponding author on reasonable request.

References

[1] Council for Higher Education Accreditation. (n.d.). *CIQG member spotlight: Nigeria's National Universities Commission*. https://www.chea.org/ciqg-member-spotlight-nigerias-national-universities-commission

[2] Iguodala, W. A. (2023). An appraisal of the National Universities Commission quality assurance measures in the development of academic programmes in Nigerian universities. *British Journal of Education, 11*(7), 96–114. https://doi.org/10.37745/bje.2013/vol11n796114

[3] Ibijola, E. Y. (2014). Accreditation role of the National Universities Commission and the quality of the educational inputs into Nigerian university system. *Universal Journal of Educational Research, 2*(9), 648–654. https://doi.org/10.13189/ujer.2014.020907

[4] Akomolafe, C. O., & Adesua, V. O. (2020). Accreditation of academic programmes and university administration in public universities in South West Nigeria. *International Journal of Cross-Disciplinary Subjects in Education, 11*(1), 4221–4229.

[5] Osarenren-Osaghae, R. I. (2014). Influence of accreditation on quality assurance: An experiential report of the Nigerian public universities. *International Journal of Cross-Disciplinary Subjects in Education, 5*(2), 1626–1631. https://doi.org/10.20533/ijcdse.2042.6364.2014.0228

[6] Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education, 18*(1), 9–34.

[7] Ekpoh, U. I., & Edet, A. O. (2017). Politics of programme accreditation practices in Nigerian universities: Implications for quality assurance. *Journal of Educational and Social Research, 7*(2), 73–79. https://doi.org/10.5901/jesr.2017.v7n2p73

[8] Godgift, P. F. (2018). The paradox of quality assurance mechanism in the management and administration of universities for quality universities in Nigeria: The National Universities Commission in focus. *Wilberforce Journal of Social Sciences, 3*(1), 93–109. https://doi.org/10.36108/wjss/8102.30.0160

[9] Uddin, O. O., Konyeha, S., & Edegbe, G. N. (2024). Enhancing accreditation transparency and accountability through electronic data interchange: A compara-

tive study. *International Journal of Science and Research Archive, 12*(2), 2135–2140. https://doi.org/10.30574/ijsra.2024.12.2.1339

[10] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery, 10*(3), e1355.

[11] Torrents, D., Mancho, N., Mateos, J. L., & Prades, A. (2026). The role of artificial intelligence in higher education quality assessment: Opportunities and limitations in predicting the quality risk in degree programmes. *Quality in Higher Education, 32*(1), 41–57. https://doi.org/10.1080/13538322.2026.2642507

[12] Kawesha, F., & Phiri, J. (2024). Data mining and machine learning-based predictive model to support decision-making for the accreditation of learning programmes at the Higher Education Authority. In *Proceedings of Ninth International Congress on Information and Communication Technology* (pp. 351–361). Springer Nature. https://doi.org/10.1007/978-981-97-3289-0_29

[13] Aldosari, F. M., Badawood, D. Y., & Almotairi, K. H. (2026). AccreditNet: An explainable data-adaptive AI framework for accreditation quality analytics and decision support in the training sector. *Algorithms, 19*(9), 781.

[14] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine, 162*(1), 55–63.

[15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

[16] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

[17] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

[18] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.

[19] Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), e0118432.

[20] Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine, 17*(1), 230.

[21] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.

Copyright and License

Copyright © 2026 The Author(s).

This article is published by *African Journal of Mathematics, Statistics and Computer Science (AJMSCS)* and is distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal Website: https://ktrend.org/