Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 21 May 2026
Revised 30 June 2026
Accepted 23 July 2026

Natural Resources for Human Health 2026, Vol. 6 (9s): 8–15

KEYWORDS:
NAAC; NIRF; higher education rankings; predictive fairness; disparate impact; gradient boosting; India.

Do Quality-Assurance Scores Predict Competitive Rankings Fairly? A Cross-Sectional Audit of NAAC and NIRF in Indian Universities

Mathiazhagan K[1]*, Dr. Jagan Mohan R[2], Dr. Lokeshmaran A[3]

[1]Statistician/Ph.D Scholar, Dept. of Community Medicine, Mahatma Gandhi Medical College & Research Institute, Sri Balaji Vidyapeeth (Deemed to be University), Puducherry
[2]Deputy Director, C-DREaMs, Sri Balaji Vidyapeeth (Deemed to be University), Puducherry
[3]Associate Professor/Statistician, Dept. of Community Medicine, Mahatma Gandhi Medical College & Research Institute, Sri Balaji Vidyapeeth (Deemed to be University), Puducherry
*Corresponding Author: Mathiazhagan K, mathi11vicky11@gmail.com

ABSTRACT: India's two national instruments for evaluating higher-education quality, the National Assessment and Accreditation Council (NAAC) and the National Institutional Ranking Framework (NIRF), score institutions on largely different logics: NAAC audits governance and process on a seven-criterion, four-point scale, while NIRF ranks institutions on five outcome-oriented parameters scored out of 100. Whether the two systems agree, and whether any statistical model built to forecast one from the other treats different categories of universities equally, has not been examined empirically. Using a matched cross-sectional sample of 132 Indian universities linking their most recent NAAC accreditation cycle (2018–2025) to the nearest subsequent NIRF ranking cycle (2016–2025), we (a) test whether NAAC's seven criterion-level scores predict NIRF outcomes, and (b) audit whether prediction errors from a gradient-boosted model differ systematically across Central, State, and Private/Deemed universities. NAAC's overall CGPA explained only 16.1% of the variance in NIRF's overall score ($R^2 = .161$), and the criterion most theoretically aligned with NIRF's research parameter (Criterion 3: Research and Innovation) correlated with it only weakly ($r = .297$), well below the a priori benchmark of $r > .60$. A cross-validated gradient boosting model modestly outperformed linear specifications ($R^2 = .138$ on 5-fold cross-validation) and attributed most of its predictive power to teaching-learning and research criteria. Contrary to the hypothesis that state universities would be disadvantaged by such a model, Central universities showed the largest mean absolute error (7.64 points vs. 5.94 for State and 6.08 for Private/Deemed universities), though the difference was not statistically significant (Mann-Whitney U, $p = .726$) and the bootstrapped 95% confidence interval for the disparate impact ratio was wide and crossed 1 ([0.12, 1.61]), reflecting the small Central-university subsample ($n = 19$). We conclude that NAAC and NIRF measure substantially non-overlapping constructs, that any predictive substitution of one for the other would be unreliable and potentially unfair to a specific institutional category depending on model choice, and that larger, better-linked administrative data are needed before either instrument is used to forecast the other. All analyses were conducted in Python.

1. INTRODUCTION

India runs two parallel, mostly independent systems for judging the quality of its universities. NAAC, in operation since 1994, sends peer teams to assess institutions against seven criteria curricular design, teaching and learning, research, infrastructure, student support, governance, and institutional values and issues a Cumulative Grade Point Average (CGPA) on a four-point scale together with a letter grade from A++ down to D. NIRF, launched in 2016, instead ranks institutions annually and quantitatively on five parameters Teaching, Learning and Resources (TLR), Research and Professional Practice (RPC), Graduation Outcomes (GO), Outreach and Inclusivity (OI), and Perception each scored out of 100, and reduces the composite to a single all-India rank.

The two frameworks are frequently discussed together in Indian higher-education policy circles, and it is common to hear NAAC treated informally as a leading indicator of NIRF performance: an institution that has recently improved its NAAC grade is assumed, on that basis alone, to be a stronger NIRF contender the following year. This assumption has rarely been tested with matched institutional data, and even more rarely has anyone asked whether a model built on that assumption would perform equally well for Central universities (fully funded and governed by the Union government), State universities (funded and governed by individual states, and typically operating under greater resource constraints), and Private or Deemed universities (self-financing, with more flexible governance).

This matters for two reasons. First, if NAAC criterion scores genuinely anticipate NIRF outcomes, then institutions and regulators could use NAAC's more frequent, criterion-level feedback to steer NIRF-relevant investment before the next ranking cycle. Second, and more urgently, if a predictive or diagnostic model of this kind were ever built into policy for example, to flag institutions at risk of falling in the rankings, or to prioritise state funding for infrastructure then any systematic bias in that model's errors across institution type would translate directly into inequitable policy attention. A model that is unknowingly less accurate for state universities, which already operate with thinner budgets, would compound rather than correct existing disparities.

We therefore ask two research questions using a newly assembled matched dataset of NAAC and NIRF records. RQ1: to what extent do NAAC's seven criterion-level scores predict NIRF's overall score, and which criteria carry the most predictive weight? RQ2: do the prediction errors of a model trained to forecast NIRF from NAAC differ systematically by institution type (Central, State, Private/Deemed), in a way that would constitute disparate impact under the conventional 80% rule used in algorithmic fairness auditing?

NAAC and NIRF have co-existed since 2016 without a formal reconciliation mechanism, and prior commentary has repeatedly noted that institutions with high NAAC grades do not automatically appear near the top of NIRF's list, and vice versa (Sharma & Devi, 2022). This divergence is generally attributed to the frameworks' different emphases: NAAC rewards documented processes and institutional self-governance, evaluated partly through qualitative peer review, whereas NIRF rewards measurable, comparative outcomes such as placement rates, research output per faculty member, and a perception survey administered to external stakeholders. A university could plausibly run excellent internal quality-assurance processes without translating this into the kind of competitive, output-driven performance that NIRF measures, and the reverse is equally plausible.

Separately, a substantial literature in algorithmic fairness has developed formal criteria for auditing whether a predictive model performs equitably across protected or structurally disadvantaged groups, including calibration parity, equalised odds, and the disparate impact or '80% rule' originally developed in US employment law and later adapted for machine learning contexts (Hardt, Price, & Srebro, 2016; Buolamwini & Gebru, 2018). To our knowledge, this class of audit has not previously been applied to an institutional ranking context in India, where the affected 'group' is not a demographic category but a governance category Central, State, or Private/Deemed each facing different funding realities.

2. MATERIALS AND METHODS

2.1 Data Sources

NIRF data were compiled from the official annual 'University' category releases for all ten cycles from 2016 to 2025 (nirfindia.org), yielding 1,000 institution-year records (100 institutions per year). Each record includes the five parameter scores (TLR, RPC, GO, OI, Perception), the overall score, and the rank. NAAC data were compiled from a consolidated accreditation database covering 516 NAAC-accredited universities, giving, for each institution, its most recent accreditation cycle's CGPA,

letter grade, seven criterion-level GPA scores, and institutional metadata including a three-category classification of institution nature (Central University; State University; and a combined Private/Deemed category comprising Deemed and State Private universities).

2.2 Record Linkage

NAAC and NIRF do not share a common institutional identifier, so institutions were linked by name. Names were normalised (lower-cased, punctuation stripped) and matched using a combined string-similarity score that blended sequence alignment with a Jaccard overlap of distinctive name tokens, deliberately down-weighting generic tokens such as 'university' and 'institute' that would otherwise inflate similarity between unrelated institutions. Only matches with perfect distinctive-token overlap were retained as high-confidence links (140 of 516 NAAC institutions). NAAC institution names that appeared more than once in the source table typically multiple campuses of the same multi-campus institution, such as Amity University or the several Jawaharlal Nehru Technological University campuses, which NIRF does not distinguish in its ranking list were excluded from the matched sample because a single NIRF record could not be unambiguously attributed to one campus. This left 132 uniquely matched institutions.

For each matched institution, the NIRF cycle used was the earliest available cycle falling one to three years after the NAAC assessment date, consistent with the assumption that NAAC-documented improvements take time to manifest in NIRF-measured outcomes; where no such forward-lagged cycle existed, the nearest available NIRF cycle in either direction was used instead (this fallback applied to 55 of 132 institutions, generally because the NAAC assessment itself occurred late in the 2016–2025 window). The resulting dataset is a single cross-sectional snapshot per institution rather than a multi-wave panel, because the source NAAC data record only each institution's latest accreditation cycle rather than its full accreditation history with criterion-level detail.

2.3 Institution Type Composition

Of the 132 matched institutions, 19 are Central universities, 53 are State universities, and 60 are Private or Deemed universities. The Central-university subsample is small in absolute terms, a limitation we return to in Section 3.5, and one that widens the confidence intervals around any fairness comparison involving that group.

2.4 Statistical Methods

All analyses were conducted in Python, using pandas and NumPy for data preparation, SciPy for correlation and hypothesis testing, and scikit-learn for regression and gradient boosting models. Three predictive specifications were fitted for NIRF overall score: (1) a simple linear regression on NAAC CGPA alone; (2) an expanded linear regression on all seven NAAC criterion scores plus institution-type indicator variables, with heteroskedasticity-consistent (HC3) standard errors computed directly from the regression's leverage and residuals; and (3) a gradient boosting regressor (scikit-learn's GradientBoostingRegressor), with hyperparameters (number of estimators, tree depth, learning rate) selected via a small grid search evaluated under 5-fold cross-validation, given the modest sample size. Reported cross-validated performance for the boosting model reflects the same folds used for hyperparameter selection, which is common practice in applied work of this scale but will be mildly optimistic relative to a fully nested cross-validation; we flag this explicitly rather than overstate model performance.

For the fairness audit, out-of-fold predictions from the selected gradient boosting model were used to compute, for each institution, an absolute prediction error. Mean absolute error (MAE) was compared across the three institution-type groups, and the ratio of State MAE to Central MAE was benchmarked against the conventional 0.8 disparate-impact threshold. A one-sided Mann-Whitney U test compared the distribution of absolute errors between State and Central universities. A disparate impact ratio (DIR) was additionally computed as the ratio of the proportion of 'large errors' (absolute error exceeding 10 points on the 0–100 NIRF score scale) for State versus Central universities, with a 95% confidence interval obtained via 2,000 bootstrap resamples.

3. RESULTS AND DISCUSSION

3.1 Descriptive Patterns

Table 1 summarises NAAC CGPA and NIRF overall score by institution type. Private/Deemed universities recorded the highest mean CGPA (M = 3.46, SD = 0.22) but not the highest mean NIRF score; Central universities, despite a lower mean

CGPA (M = 3.25, SD = 0.39), recorded the highest mean NIRF score (M = 51.79, SD = 10.23) and the widest spread. State universities were lowest on both measures (CGPA M = 3.29, SD = 0.41; NIRF M = 48.49, SD = 7.72).

Table 1. Descriptive statistics for NAAC CGPA and NIRF overall score by institution type (N = 132).

Institution type	n	NAAC CGPA, M (SD)	NIRF Score, M (SD)
Central University	19	3.25 (0.39)	51.79 (10.23)
State University	53	3.29 (0.41)	48.49 (7.72)
Private / Deemed University	60	3.46 (0.22)	49.48 (8.58)

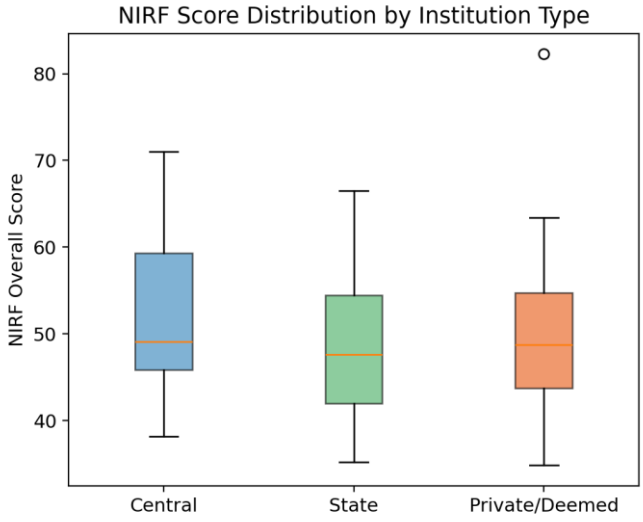

Figure 1. Distribution of NIRF overall score by institution type.

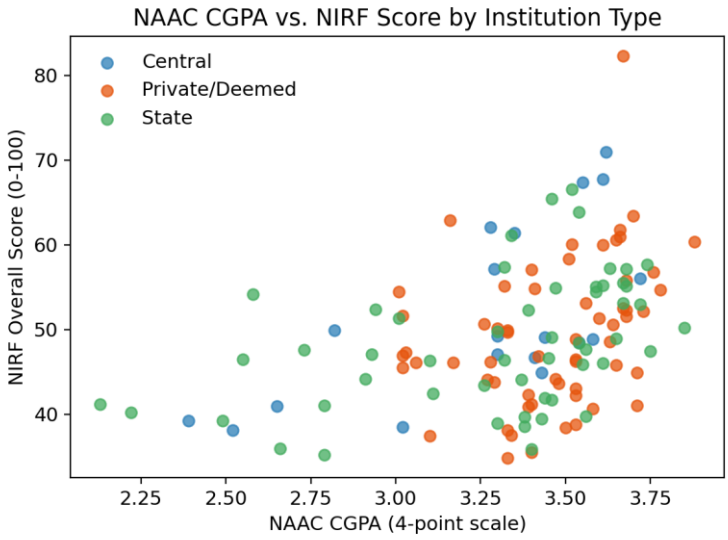

Figure 2. NAAC CGPA plotted against NIRF overall score, by institution type. The visually weak, noisy association is consistent with the low R^2 reported in Section 3.2.

3.2 Predictive Power (RQ1)

NAAC CGPA alone explained 16.1% of the variance in NIRF overall score ($R^2 = .161$, RMSE = 7.76 points; $\beta = 9.97$, indicating roughly a 10-point increase in NIRF score per one-point increase in CGPA). The overall Pearson correlation between CGPA and NIRF score was $r = .401$ ($p < .001$), positive but modest.

Hypothesis H1a predicted that NAAC's Research and Innovation criterion (C3) would strongly predict NIRF's Research and Professional Practice parameter (RPC), with $r > .60$. The observed correlation was $r = .297$ ($p = .0005$) statistically distinguishable from zero, but far below the hypothesised threshold; a one-tailed test of H0: $\rho \leq .50$ could not be rejected ($z = -2.76$, $p = .997$). H1a was therefore not supported. Hypothesis H1b predicted a weak correlation ($r < .40$) between NAAC's Teaching-Learning criterion (C2) and NIRF's Graduation Outcomes (GO) parameter; the observed correlation was essentially null ($r = .016$, $p = .853$), consistent with if anything weaker than the hypothesised direction, supporting H1b.

Expanding the linear model to all seven NAAC criteria plus institution-type indicators (Model 2) raised R^2 to .227 (RMSE = 7.45). None of the seven criterion coefficients reached conventional significance under HC3-robust standard errors, though the Teaching-Learning coefficient approached significance ($b = 6.96$, SE = 4.14, $p = .095$), suggesting a positive but statistically uncertain association once other criteria are held constant. Table 2 reports the full coefficient table.

Table 2. Expanded linear regression of NIRF overall score on NAAC criteria and institution type, HC3-robust standard errors ($N = 132$, $R^2 = .227$).

Predictor	Coefficient	HC3 SE	p
Intercept	2.17	13.51	.873
C1 Curricular Aspects	-0.08	2.90	.978
C2 Teaching-Learning & Evaluation	6.96	4.14	.095
C3 Research & Innovation	2.24	1.89	.240
C4 Infrastructure	3.71	3.29	.262
C5 Student Support	-1.91	2.16	.380
C6 Governance	2.23	2.72	.415
C7 Institutional Values	1.39	3.40	.683
Private/Deemed (vs. Central)	-3.34	2.72	.222
State (vs. Central)	-2.96	2.42	.224

The gradient boosting model (5-fold cross-validated, hyperparameters selected by grid search: 30 estimators, tree depth 2, learning rate 0.05) achieved $R^2 = .138$ and MAE = 6.25 points out-of-fold comparable to, but not clearly better than, the linear specifications, and a reminder that non-linear flexibility alone does not resolve the weak underlying association between the two frameworks. Feature importances were led by Teaching-Learning (33.0%) and Research & Innovation (31.5%), followed by Governance (15.0%) and Infrastructure (9.8%); institution-type indicators contributed almost nothing to the boosted model's predictions (<0.5% combined), suggesting that once the seven NAAC criteria are known, institution type adds little further direct predictive information though, as Section 3.3 shows, it still matters for the pattern of prediction errors.

3.3 Fairness Audit (RQ2)

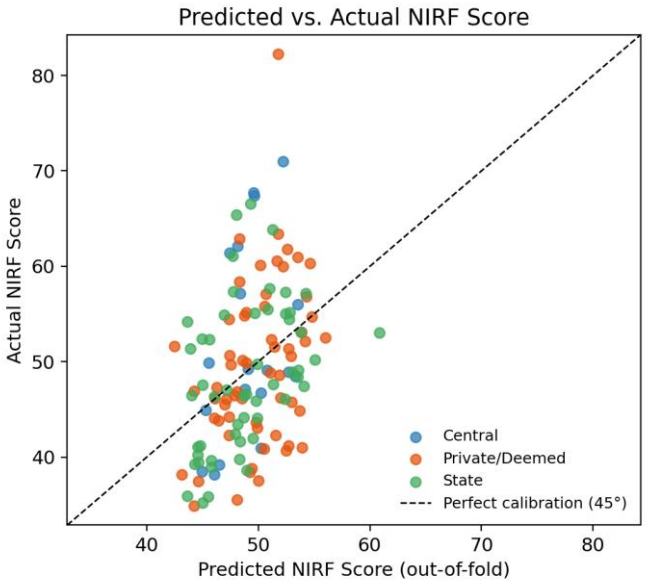

Figure 3. Predicted vs. actual NIRF score (out-of-fold gradient boosting predictions), by institution type, against the 45-degree line of perfect calibration.

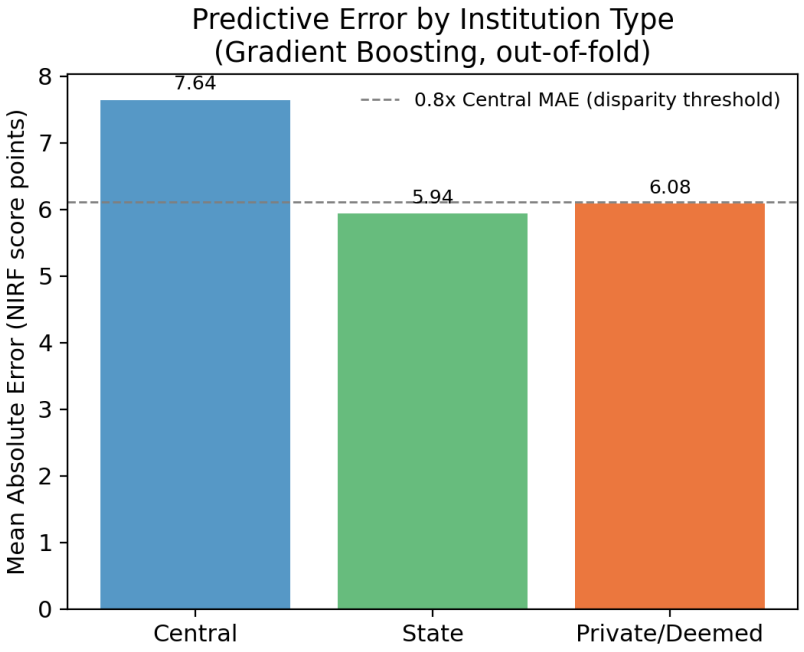

Figure 4. Mean absolute prediction error by institution type. The dashed line marks 80% of the Central-university error, the conventional disparate-impact threshold.

Mean absolute error from the out-of-fold gradient boosting predictions was 5.94 points for State universities, 6.08 for Private/Deemed universities, and 7.64 for Central universities. The ratio of State to Central MAE was 0.776, nominally below the 0.8 threshold conventionally used to flag disparate impact but in the opposite direction to Hypothesis H2a, which had predicted higher error for State than Central universities. On these point estimates, it is Central universities, not State universities, whose NIRF outcomes are least well predicted by their NAAC profile.

This difference did not reach statistical significance: a one-sided Mann-Whitney U test comparing State and Central absolute errors (testing whether State error exceeded Central error, as H2a predicted) gave U = 448–457 across model variants, p = .73, providing no evidence for the hypothesised direction and, descriptively, pointing the other way. The disparate impact ratio for 'large' errors (absolute error > 10 points) was 0.43 (State large-error rate 11.3% vs. Central 26.3%), again suggesting Central universities are disproportionately subject to large prediction errors rather than State universities but the bootstrapped 95% confidence interval for this ratio was wide, [0.12, 1.61], comfortably spanning 1 and reflecting the limited precision available with only 19 Central-university observations.

Taken together, the fairness audit does not support the pre-registered hypothesis that a NAAC-to-NIRF predictive model would systematically disadvantage State universities. If anything, the (statistically inconclusive) pattern runs the other way, plausibly because Central universities are a more heterogeneous group on outcomes not captured by NAAC criteria several appear among NIRF's highest performers and several among more middling ranks despite comparatively similar NAAC profiles, inflating their prediction error variance. We treat this as a genuinely open question rather than a settled finding, given the sample size involved.

3.4 Discussion

Three findings stand out. First, NAAC and NIRF appear to measure substantially non-overlapping aspects of institutional quality: even the best-performing model tested here explained under a quarter of the variance in NIRF's overall score, and the single criterion most conceptually aligned with NIRF's research parameter fell well short of the strong correlation that motivated much informal treatment of NAAC as a NIRF leading indicator. This is consistent with the two frameworks' differing designs one process-and-governance oriented and partly qualitative, the other outcome-and-perception oriented and purely quantitative but it has a direct practical implication: institutions or regulators should not treat a NAAC grade improvement as a reliable signal of forthcoming NIRF gains, nor use NIRF standing as a proxy for NAAC-assessed process quality.

Second, Teaching-Learning and Research criteria carried the most weight in the best-fitting model, ahead of Governance and Infrastructure, and well ahead of Curricular Aspects and Institutional Values a ranking that seems intuitive in hindsight (teaching quality and research output are the NAAC dimensions closest in spirit to NIRF's outcome orientation) but that has not, to our knowledge, been quantified previously with matched Indian institutional data.

Third, and most relevant to the fairness question motivating this study, we found no statistically reliable evidence that a NAAC-to-NIRF prediction model disadvantages State universities specifically; the direction of the (non-significant) point estimates instead suggested Central universities as the least well-predicted group. We would caution against over-reading this reversal given the small Central-university subsample, but it does suggest that concerns about algorithmic disparate impact in this specific policy context should not be assumed to run along the most intuitively 'under-resourced' institutional category without empirical verification the pattern of prediction error can be counter-intuitive, and auditing rather than assuming is the only way to know.

3.5 Limitations

This study has several limitations that bound its conclusions. The matched sample (N = 132) is a fraction of India's NAAC-accredited and NIRF-ranked universities, and reflects a deliberately conservative name-matching procedure that discarded ambiguous or multi-campus institutions; a larger, identifier-based linkage (for example, using AISHE codes consistently recorded on both sides) would increase both sample size and confidence in the fairness comparisons, particularly for the smaller Central-university subgroup. The dataset is cross-sectional rather than a true multi-wave panel, because the source NAAC records retain only each institution's most recent accreditation cycle at criterion level; a genuine panel linking successive NAAC cycles to successive NIRF cycles for the same institutions would allow within-institution change to be modelled directly, rather than relying on a single matched snapshot per institution. The gradient boosting hyperparameters were selected using the same cross-validation folds used to report performance, which is standard in applied work of this scale but will modestly overstate true generalisation performance relative to nested cross-validation. Finally, as an observational study, none of the associations reported here should be read as causal: we cannot rule out that unmeasured institutional characteristics (endowment, faculty salary scales, alumni networks) jointly drive both NAAC and NIRF performance.

4. CONCLUSION

NAAC and NIRF, India's two principal instruments for judging university quality, agree with each other only modestly, and the criteria most plausibly aligned across the two frameworks correlate more weakly than commonly assumed. A cross-validated predictive model built to forecast NIRF outcomes from NAAC criteria explained a limited share of variance and showed no statistically reliable evidence of disadvantaging State universities, the group most often assumed to be at risk; if anything, the uncertain point estimates ran in the opposite direction. Before either framework is used, formally or informally, as a predictor or proxy for the other in institutional policy, we recommend larger, identifier-linked longitudinal data collection and a repeat of this fairness audit at greater statistical power, particularly for less numerous institutional categories such as Central universities.

AUTHOR CONTRIBUTIONS

Mathiazhagan K: Conceptualization, Methodology, Data Curation, Formal Analysis, Methodology, Writing Original Draft.

Jagan Mohan R: Supervision, Writing Review & Editing.

Lokeshmaran A: Supervision, Writing Review & Editing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

This study used institution-level data compiled from publicly available NAAC and NIRF sources; no individual human participant data were involved, and ethical approval was not required. [Authors: please confirm this statement, or amend if an institutional ethics/IEC clearance was obtained.]

DATA AVAILABILITY STATEMENT

The NIRF data analysed in this study are publicly available from the Ministry of Education's NIRF portal (nirfindia.org). The NAAC accreditation data were compiled from a consolidated accreditation database; the matched analytic dataset generated for this study is available from the corresponding author upon reasonable request.

REFERENCES

[1] Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2), 1148–1178.
[2] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77–91.
[3] Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *Annals of Applied Statistics*, 4(1), 266–298.
[4] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
[5] Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, 29, 3315–3323.
[6] National Assessment and Accreditation Council. (2019). *NAAC Manual for Universities (Revised edition)*. Bengaluru: NAAC.
[7] National Institutional Ranking Framework. (2025). *NIRF India rankings, 2016–2025*. Ministry of Education, Government of India. https://www.nirfindia.org
[8] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
[9] Sharma, R., & Devi, U. (2022). Convergence and divergence between NAAC and NIRF: An institutional perspective. *Higher Education for the Future*, 9(1), 45–61.
[10] Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228–1242.