Utilizing Structured Equation Modeling and Machine Learning in Investigating Digital Competency in Public School Teachers in Bukidnon, Philippines

Nathalie Joy G. Casildo
Information Technology Dept.-College of Information Sciences and Computing,
Central Mindanao University, Bukidnon, Philippines

Abstract—**The rapid transition to digital education in the Philippines, accelerated by the COVID-19 pandemic, has highlighted significant integration challenges for public school teachers in rural provinces like Bukidnon. While digital proficiency is essential, existing studies often rely on either purely descriptive analytics or standalone machine learning models, which frequently fail to validate the complex, latent relationships between competency factors in terms of Digital Competency. To address these gaps, this research employs a two-phased hybrid analytical architecture. Phase I utilizes Structural Equation Modeling (SEM) to confirm the factor structure and establish the causal pathways of digital competency, ensuring that the theoretical framework is psychometrically sound. Phase II transitions these validated constructs into an optimized Machine Learning (ML) pipeline, incorporating SMOTE-ENN resampling to handle imbalanced regional data. Results from 1,275 participants demonstrate that "Professional Engagement" acts as the foundational engine of the digital competency system, while "Digital Pedagogy in Teaching" emerges as the most critical predictive determinant of teacher proficiency. The Random Forest algorithm achieved a high predictive accuracy of 89% and a Macro F1-Score of 85%, significantly outperforming traditional models. These findings indicate that the digital divide in this context is a pedagogical, rather than purely technical, bottleneck. The study provides a blueprint for the Department of Education to move from descriptive reporting toward Predictive Diagnostic Systems that can facilitate targeted, data-driven interventions.**

Keywords—Structural Equation Modeling (SEM); Machine Learning (ML); digital competence; DigCompEdu; teacher proficiency; SMOTE-ENN; predictive analytics; educational data mining

I. INTRODUCTION

Digital transformation in the Philippines was historically gradual, only accelerating rapidly out of necessity during the COVID-19 pandemic. This shift made remote work and digital education the primary modes of instruction. In response, national agencies like the Department of Education (DepEd), specifically in Region 10, mandated the full integration of Information and Communication Technology (ICT) in primary instruction. However, despite these mandates, public school teachers in inland provinces like Bukidnon continue to face significant integration challenges. Infrastructure remains a challenge; many schools still lack stable signals, and even the existence of satellite technologies is often insufficient. Beyond hardware, there is a persistent lack of awareness regarding data security and digital literacy, which suggests that infrastructure alone cannot solve the underlying threat to digital readiness.

While the Philippines has adopted various ICT frameworks, the actual digital proficiency of teachers in these rural settings remains under-explored through advanced analytics. In the Philippine educational context, most studies rely on either pure descriptive analytics or standalone Machine Learning (ML) and Educational Data Mining (EDM) [1]. While a few researchers have attempted to combine Structural Equation Modeling (SEM) and ML [2], these efforts are often hampered by limited datasets or a narrow focus on general digital literacy. Consequently, existing methods frequently fail to validate the complex, latent (hidden) relationships between competency factors, fail to accurately predict future teacher performance, and struggle to handle skewed data—where the majority of teachers might fall into a single proficiency category.

To address these gaps, this research employs a two-phased hybrid architecture. Phase I utilizes Structural Equation Modeling (SEM) to confirm the factor structure and establish the causal pathways of digital competency. Phase II then transitions these validated constructs into Machine Learning (ML) algorithms, such as Random Forest and Support Vector Machines, to build a predictive model optimized for regional educational data.

Unlike traditional studies that rely on either simple statistics or "black-box" machine learning, this study proposes a hybrid SEM-ML architecture. Using SEM as a rigorous validation layer for feature engineering, the research ensures that the inputs for the machine learning model are not just statistically significant but are theoretically grounded and psychometrically sound.

II. RELATED WORK

A. Computational Modeling of Digital Competence

Digital competence encompasses an integration of technical, cognitive, and socio-emotional proficiencies [3]. Assessing these skills involves using established frameworks [4] paired with computational modeling [5, 6]. Applying mathematical approaches, these models provide a scientific basis for interpreting the status of an individual's digital competence [7]. However, the focus of these frameworks must move beyond technical skills toward pedagogical intervention [8].

B. Digital Competence in DepEd Teachers

Existing research affirms the value of digital literacy (DL) and digital competence (DC) for Filipino teachers, enabling them to engage with new technologies critically within digital pedagogy [9]. Enhancing these competencies can drive higher work engagement, ultimately leading to improved job performance and satisfaction, particularly in remote work environments [10]. In terms of regional applications, [4] and [11] provide significant contributions. Specifically, [4] explores frameworks in Fig. 1 relevant at national, international, and regional scales, such as the framework developed by the UNESCO Asia and Pacific Regional Office. Furthermore, the work in [11] serves as a valuable benchmark for the Bukidnon context; it analyzes digital competence frameworks used in Ibero-America through 2022 to establish a conceptual foundation for assessment tools in regional educational settings.

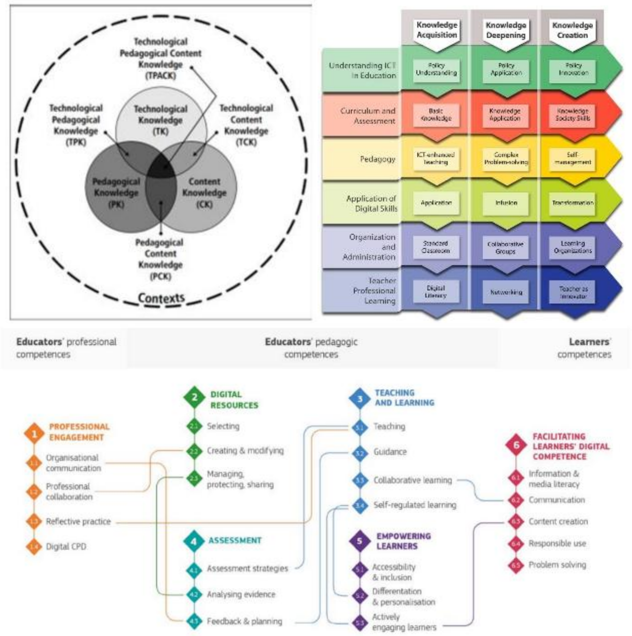


Fig. 1. ICT and competency frameworks [4].

C. Structural Equation Modeling (SEM) as a Validation Layer

Structural Equation Modeling (SEM) serves as a robust validation layer in contemporary research [12–14]. Since digital competence is a latent construct that eludes direct measurement, SEM utilizes Confirmatory Factor Analysis (CFA) to verify its underlying dimensions [15]. A practical application of this is seen in the validation of the TDC-S instrument [16]. In this study, Fig. 2 presents the CFA output for measurement model 3, reporting $X^2 = 1{,}480.693$, $(df = 584, p = 0.00)$ and an X^2/df ratio of 2.535. Further evaluation of the model's performance includes an RMSEA of 0.075 and a CFI of 0.897. Following the benchmarks established in recent literature, [17] provides an interpretation of these indices to determine the scientific status of the model's fit. A study establishes a robust theoretical framework for confirming theoretical constructs and quantifying latent relationships, which serves as the foundation for subsequent predictive modeling [18].

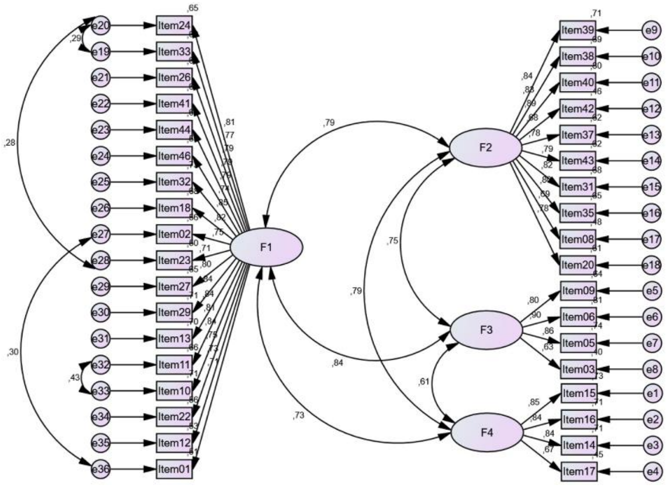


Fig. 2. CFA output of measurement model 3 [14].

D. Machine Learning and Educational Data Mining

Recent research has explored the application of machine learning (ML) to predict teachers' digital competency [19–21]. Within the field of Educational Data Mining (EDM), the most frequently utilized algorithms include Logistic Regression, Random Forest, and Decision Trees [19, 22, 23]. Furthermore, several studies have introduced hybrid methodologies that integrate Structural Equation Modeling (SEM) with Machine Learning [24, 25]. Specifically, [24] demonstrates that combining Bayesian SEM with ML achieves superior levels of predictive accuracy and robustness compared to traditional SEM or standalone ML techniques. Similarly, [25] bridges the gap between causal explanation and predictive accuracy by employing SEM to identify direct and indirect pathways, while using ML methods to reveal marginal contributions and nonlinear transition characteristics. These studies collectively indicate that the integration of SEM and ML offers a more effective and comprehensive analytical approach.

E. Hybrid SEM-ML Architectures in the Global Research Landscape

The integration of Machine Learning (ML) into Structural Equation Modeling (SEM) frameworks has emerged as a robust analytical paradigm in recent educational research. This hybrid approach allows for a fusion of explanation and prediction, addressing the limitations of theory-driven models in highly complex environments. Brandmaier and Jacobucci [26] emphasize that integrating ML techniques into SEM renders theory development more structured, efficient, and robust, particularly when dealing with high-dimensional data that traditional path analysis may struggle to parse.

Recent international studies have successfully operationalized this synergy across various educational domains. For example, Zhang et al. [27] applied a hybrid approach to predict student engagement in physical education, utilizing nine ML algorithms—with Random Forest yielding the highest predictive accuracy—to rank the influence of various digital tools. By concurrently using SEM to map causal pathways, the authors provided evidence on how technology usage indirectly impacts engagement through psychological mechanisms such as perceived usefulness and self-efficacy.

Furthermore, this hybrid architecture is instrumental in identifying the non-linear realities of digital transformation. In a large-scale study of 650 participants, a recent investigation [28] demonstrated that structural models can sometimes reveal statistically insignificant relationships (e.g., between technological infrastructure and actual digital adoption) that might be overlooked in simpler frameworks. This suggests that digital transformation is not a deterministic outcome of resource allocation or top-down mandates, but rather a complex, multi-faceted process. Positioning our study within this global methodological landscape, leveraging these advanced techniques to provide a nuanced, data-driven diagnostic of digital competency among public school teachers in Bukidnon, Philippines.

F. Resampling Techniques for Imbalanced Regional Data

Data imbalance is a prevalent challenge in educational datasets, often leading to biased models, untrustworthy outcomes, and degraded predictive performance [28–30]. To mitigate these issues, researchers frequently employ data-level interventions such as the Synthetic Minority Oversampling Technique (SMOTE). For instance, [29] implemented SMOTE-NC (for nominal and categorical data) to synthesize minority cases in high school dropout predictions, directly addressing class distribution gaps. Unlike Random Oversampling (ROS), which can lead to overfitting through sample duplication, SMOTE generates new synthetic instances by interpolating between a minority sample and its *k*-nearest neighbors [31]. Beyond oversampling, the literature suggests that various sampling strategies are essential for optimizing performance metrics [32–34]. Notably, [35] found that Random Undersampling (RUS) and hybrid techniques significantly improved Recall and F1-scores, demonstrating that prioritizing the detection of minority class instances is often more critical than maximizing overall accuracy.

III. METHODOLOGY

Fig. 3 shows the Knowledge Discovery and Predictive Analytics process. The activity started from research instrument design, where the conceptualization and schema design of the research instrument were involved. The questionnaire was adapted from the European Framework for Digital Competence in Education (DigCompEdu) to ensure construct validity. Once the framework is established, proceed to the Data Acquisition of Large-scale Survey Collection from the teachers of Bukidnon.

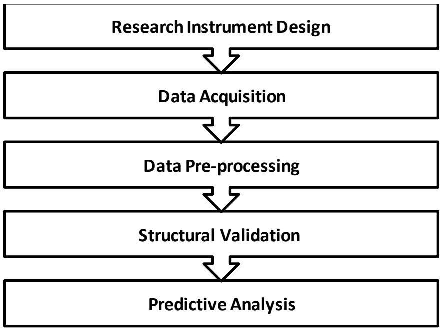


Fig. 3. Knowledge discovery and predictive analytics process.

The third phase is the Data Pre-Processing involving Data Cleaning & Feature Engineering. Proceeded with structural validation, in which Structural Equation Modeling was used using SPSS AMOS. The last step is the Predictive Analysis using logistic Regression, where accuracy calculation and Predictive classification of digital divide levels were conducted.

A. Dataset Acquisition and Characteristics

The dataset for this study was acquired through a large-scale survey involving public school teachers within the Division of Bukidnon, Philippines. Data collection was conducted between October 2022 and March 2023. A total of 1,275 valid responses were obtained, as shown in Table I: Summary of Dataset Characters. Responses represent a diverse cross-section of grade levels, school types, and geographic locations. The primary instrument was adapted from the European Framework for Digital Competence in Education (DigCompEdu), consisting of 22 items measured on a five-point Likert scale. In order to ensure the consistency of the multi-dimensional constructs, Cronbach's Alpha was calculated. The analysis yielded a reliability coefficient of 0.92, which significantly exceeds the standard threshold of 0.70.

TABLE I. SUMMARY OF DATASET CHARACTERISTICS (N=1,275)

Feature	**Category**	**Frequency**	**Percentage (%)**
Gender	Male	188	14.80%
	Female	1087	85.20%
School Type	Central Elementary School	847	66.22%
	Senior High School	123	9.69%
	Junior High School	305	24.02%
Professional Rank	Master Teacher	71	5.51%
	Teacher 1	909	71.34%
	Teacher 2	112	8.82%
	Teacher 3	170	13.31%
	Job Order	2	0.16%
	School Head/School School Principal	9	0.71%
	Volunteer Teacher	1	0.08%
	SpEd LGU	1	0.08%
Education	BS	841	65.83%
	MS	399	31.42%
	PhD	27	2.13%
	Post Doctorate	8	0.63%
Reliability	**Cronbach's Alpha**	**0.92**	**Excellent**

B. Data Pre-Processing

After data acquisition, the data is pre-processed. It was determined that there are 11 to 20 missing instances across several high-dimensional features. The Mode Imputation was applied, an imputation strategy that replaces missing values with the most frequent value, to address missing categorical and ordinal variables.

Handling missing data includes checking for missing values in the dataset and applying appropriate techniques such as imputation to fill in missing values. Data transformation was subsequently applied, which covers converting categorical variables to numerical format if necessary.

C. Theoretical Frameworks

- The theoretical foundation of this study is anchored in the European Framework for the Digital Competence of Educators (DigCompEdu) in Fig. 4 from [4].
- Digital competence [1] is defined not as a singular skill, but as a composite of six interrelated Latent Constructs, each representing a specific functional layer of the educational system:
- Professional Engagement: Focused on the organizational communication (PCO) and collaborative practices (PCC) of the educator.
- Digital Resources: The technical capacity to source, create (DRC), and manage digital assets.
- Digital Pedagogy: The core operational layer, encompassing the implementation of digital tools in teaching (DPT) and learner guidance (DPG).
- Assessment (Evaluation): The analytical layer where educators use digital technologies to enhance assessment strategies (EFA) and analyze evidence to provide feedback.
- Empowering Learners: The inclusive layer. It measures how technology is used to support learner diversity, accessibility (ESA), and personalized learning (ESD).

The final phase of the framework involved Predictive Analysis using a suite of supervised learning algorithms: Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest. Accuracy calculations and predictive classification of digital divide levels were conducted for all models, with comparative performance metrics utilized to select the optimal classifier for the regional dataset.

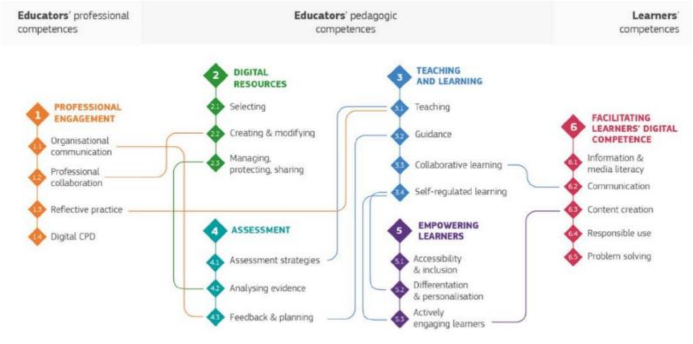
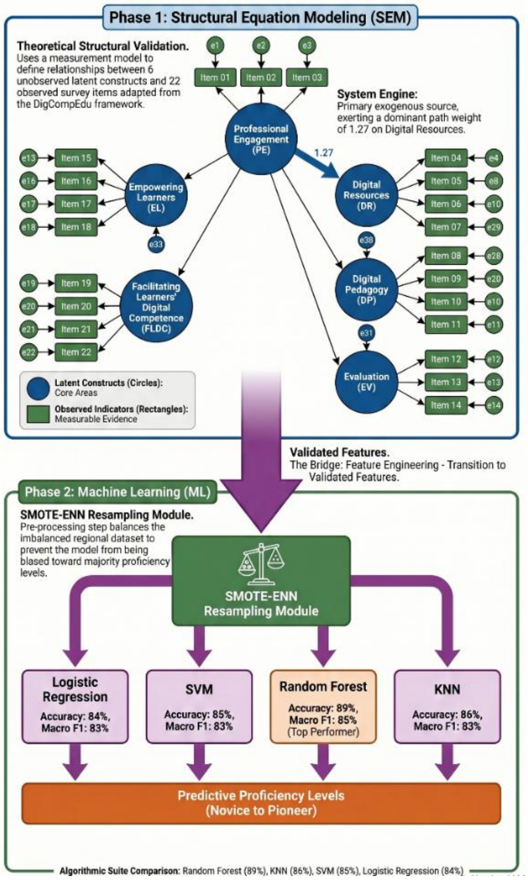


Fig. 4. Illustration of the digcompedu framework [4].

D. Structural Equation Modeling

The hybrid framework provides a two-stage approach to analyzing teacher digital competency in Fig. 5. Phase 1 uses Structural Equation Modeling (SEM) to validate theoretical constructs, identifying "Professional Engagement" as the system's foundational engine. These validated factors then bridge into Phase 2, which employs an SMOTE-ENN machine learning pipeline to handle imbalanced data and generate predictive proficiency classifications. The architecture demonstrates that an integrated approach—combining causal validation with high-performance algorithms like Random Forest (89% accuracy)—is more effective than isolated analytical methods for diagnosing and addressing pedagogical gaps in regional educational systems.



Theoretical Structural Validation. Uses a measurement model to define relationships between 6 unobserved latent constructs and 22 observed survey items adapted from the DigCompEdu framework.

System Engine: Primary exogenous source, exerting a dominant path weight of 1.27 on Digital Resources.

Validated Features. The Bridge: Feature Engineering - Transition to Validated Features.

SMOTE-ENN Resampling Module. Pre-processing step balances the imbalanced regional dataset to prevent the model from being biased toward majority proficiency levels.

Algorithmic Suite Comparison: Random Forest (89%), KNN (86%), SVM (85%), Logistic Regression (84%)

Fig. 5. Proposed model architecture.

The structural validation of the theoretical framework was conducted using Structural Equation Modeling (SEM) executed in SPSS AMOS. The model was specified as a full structural equation system comprising two distinct components: the measurement model and the structural model. The measurement model defined the relationship between six latent constructs (unobserved variables) and 22 observed indicators (survey items). Specifically, the constructs Professional Engagement, Digital Pedagogy, Digital Resources, Evaluation, Empowering Learners, and Facilitating Learners' Digital Competence were mapped to their respective observed variables. The Structural Model is the Directional paths that were established to represent hypothesized causal relationships. Professional Engagement was specified as the exogenous variable, while the remaining five constructs were specified as endogenous variables.

To account for measurement bias and unexplained variance, the model incorporated first Measurement Errors (e_1 through e_{22}), which are assigned to each observed indicator to capture variance specific to the individual survey items. Second Structural Disturbance Terms ($e_{29}, e_{30}, e_{31}, e_{33}, e_{34}$) are assigned to each endogenous latent construct to represent the residual variance not accounted for by the structural paths in the system.

The model was constrained to ensure mathematical identification and a stable metric scale. For each latent construct, one regression weight to an observed indicator was fixed to 1.0 (e.g., the paths to PCD, DPS, DRM, EFF, ESAS, and FSDP). This procedure transformed the latent constructs into a standard metric, allowing the system to estimate the remaining free parameters and path coefficients.

$$F_{\mathrm{ML}} = \ln|\Sigma(\theta)|+\mathrm{tr}(S\Sigma^{-1(\theta))-\ln|S|-p} \quad (1)$$

The Maximum Likelihood Estimation (MLE) algorithm in Eq. (1) was utilized to calculate the path coefficients and variance estimates.

The structural integrity of the model was evaluated against a suite of computational fit indices [36]. Parsimony Assessment for CMIN/DF ratio (also known as the Normed Chi-Square) is a measure used to evaluate the "parsimony" or the efficiency of the model [36]. Absolute Fit Testing for Root Mean Square Error of Approximation (RMSEA), Goodness of Fit Index (GFI), and Adjusted Goodness of Fit Index (AGFI) were employed, alongside Incremental Fit Comparison for Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Normed Fit Index (NFI) [36]. Information Criteria, such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), were also utilized for model comparison [36]. However, contemporary literature cautions against the use of rigid, fixed cutoff values for these indices, recommending instead that researchers consider model-specific characteristics and context [37][38].

$$\mathrm{P(Y=1|X)} = \frac{1}{1+e^{-(\beta_0+\sum_{i=1}^{n}\beta_i X_i}} \quad (2)$$

Lastly, in the final phase, the validated features were utilized to construct a predictive classifier using Logistic Regression in Eq. (2). To ensure the reliability of the predictive results and prevent model overfitting, the dataset (N=1,276) was partitioned using a train(80%)-test(20%) split. To validate the stability of the classifier across different subsets of the regional data, a 5-Fold Cross-Validation procedure was executed.

The predictive accuracy of the model was quantified using a comprehensive suite of classification metrics derived from the Confusion Matrix first is Accuracy - the overall ratio of correctly predicted instances to total instances. Precision and Recall are utilized to evaluate the model's ability to correctly identify positive instances of high ICT competency while minimizing false positives and false negatives. F1-Score - the harmonic mean of precision and recall, serving as a singular metric for balanced classification performance.

E. Class Imbalance

The original dataset exhibited a significant class imbalance, which could bias the predictive models toward majority classes. To mitigate this, a hybrid SMOTE-ENN (Synthetic Minority Oversampling Technique and Edited Nearest Neighbors) pipeline was applied to the training set (N=891). SMOTE was utilized to oversample minority classes by generating synthetic instances, while ENN was subsequently applied to remove ambiguous or 'noisy' samples from the feature space, ensuring well-defined decision boundaries. Table II illustrates the proficiency level distribution before and after the resampling process. The results confirm that this hybrid strategy successfully redistributed the instances to provide a more equitable analytical foundation for the machine learning models.

TABLE II. DISTRIBUTION OF PROFICIENCY LEVELS BEFORE AND AFTER SMOTE-ENN

Proficiency Level	**Count (Before SMOTE-ENN)**	**Count (After SMOTE-ENN)**
Pioneer	25	317
Novice	58	315
Leader	109	283
Explorer	328	201
Integrator	468	122
Expert	287	98
Total	**891**	**1,336**

IV. THE STRUCTURAL EQUATION MODEL

A. Goodness-of-Fit (GOF) and Model Verification

To determine the extent to which the theoretical architecture matches the empirical data collected from the 1,276 regional educators, the model was subjected to a rigorous Goodness-of-Fit (GOF) assessment. The observed covariance matrix was compared against the model-implied matrix using Maximum Likelihood Estimation in Eq. (1).

TABLE III. CALCULATED GOODNESS-OF-FIT FOR THE STRUCTURAL MODEL

Fit Index Category	**Metric**	**Calculated Value**	**Interpretation**
Parsimony Fit	CMIN/DF	4.749	Acceptable (Threshold < 5.0)
Absolute Fit	GFI	0.935	Good Fit
	AGFI	0.917	Good Fit
	RMSEA	0.054	Excellent (Threshold < 0.06)
Incremental Fit	NFI	0.936	Strong Fit
	CFI	0.948	Strong Fit
	TLI	0.940	Strong Fit

Table III shows that the model achieved convergence with high reliability. The RMSEA of 0.054 is a critical result, indicating that the structural model possesses a very low discrepancy per degree of freedom. This validates the model's generalizability across the diverse demographics of the Bukidnon teacher population. Furthermore, the GFI (0.935) and CFI (0.948) scores confirm that the proposed six-construct structure accounts for the majority of the variance in the observed indicators.

The modeling procedure followed a multi-stage mathematical specification to assess the inter-dependencies between latent constructs in Fig. 3.

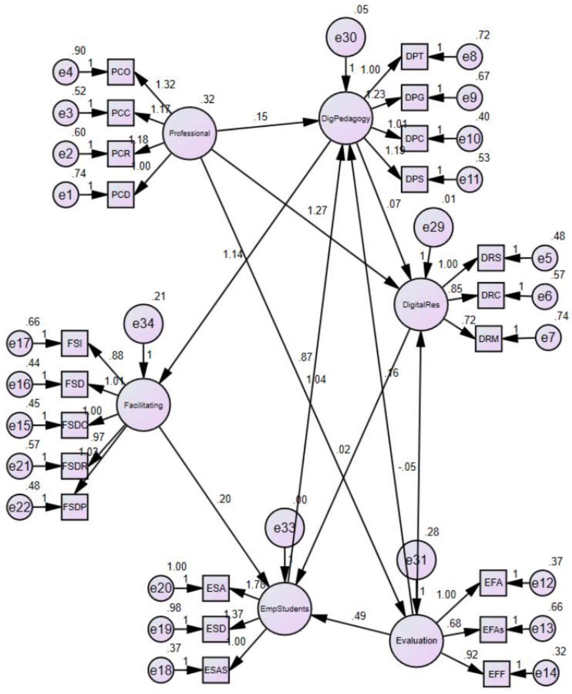


Fig. 6. SEM framework implemented in SPSS AMOS.

B. Path Analysis and Predictive Weights

Following the global validation of the model, the individual structural paths were analyzed to identify the primary drivers of teacher competency. Fig. 6 illustrates the final structural diagram with the calculated unstandardized regression weights.

The path analysis reveals a clear hierarchical influence within the digital competency system:

1) System driver: The path from Professional Engagement to Digital Resources yielded the highest weight in the system (1.27). This indicates that for every one-unit increase in professional commitment, the competency in sourcing and creating digital resources increases by 1.27 units, identifying "Professional Engagement" as the foundational engine of the system.

2) Pedagogical integration: A significant structural relationship was confirmed between *Evaluation and Empowering Students* (0.49). This confirms that data-driven assessment practices are essential precursors to achieving the ultimate inclusive goal of student digital empowerment.

3) Direct influence: The model shows that *Professional Commitment* also maintains a direct positive influence on Digital Pedagogy (0.15), suggesting that commitment to reflective practice directly modulates how tools are applied in the classroom.

C. Structural Validity Conclusion

The results of the SEM phase demonstrate that the digital divide is not a set of isolated skills but a highly integrated system. The high GOF scores prove that the DigCompEdu framework is an appropriate mathematical model for the Philippine regional context. Meeting all statistical thresholds, the latent constructs—specifically Digital Pedagogy and Professional Engagement—are validated as high-quality features. This structural certainty provides the necessary evidence base for the final Machine Learning Predictive Analysis presented in the following section.

V. MACHINE LEARNING PREDICTIVE RESULTS

Following the structural validation of the constructs, the predictive phase was executed using the resampled dataset. The application of the SMOTE-ENN (Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors) pipeline was utilized to ensure that the classifiers were not biased toward the majority proficiency levels, thereby providing an equitable analysis of the digital divide.

A. Comparative Model Performance

The predictive utility of the four supervised learning algorithms was evaluated based on Accuracy, Macro-averaged Precision, Recall, and the F1-Score. Table IV summarizes the performance of the models within the optimized SMOTE-ENN framework.

TABLE IV. PREDICTIVE PERFORMANCE METRICS WITH SMOTE-ENN RESAMPLING (N=1,276)

Model	**Accuracy**	**Macro Avg. Precision**	**Macro Avg. Recall**	**Macro Avg. F1-Score**
Logistic Regression	84%	84%	81%	83%
SVM	85%	84%	83%	83%
Random Forest	**89%**	86%	84%	85%
K-Nearest Neighbors	86%	84%	82%	83%

B. Algorithmic Analysis and Model Selection

As illustrated in Table IV, all four classifiers demonstrated high predictive performance, with F1-scores consistently exceeding 83%. The Random Forest (RF) algorithm emerged as the best performer, achieving an overall accuracy of 89% and a Macro F1-Score of 85%. The superior performance of RF is attributed to its ensemble nature, which effectively aggregates multiple decision trees to capture the non-linear relationships between the pedagogical and professional features identified in the SEM phase.

The K-Nearest Neighbors (KNN) and SVM models followed closely with accuracies of 86% and 85%, respectively. The high recall scores across these models (up to 84%) indicate that the SMOTE-ENN strategy successfully enabled the system to identify minority proficiency classes (such as "Novice" and "Leader"), which are often missed in traditional imbalanced models.

The Logistic Regression model, while having the lowest accuracy at 84%, still provided significant interpretive value through its 81% recall rate. The narrow margin of difference between the four models suggests a high degree of consensus regarding the predictive importance of the structural features used as inputs.

C. Confusion Matrix Evaluation

To visually verify the classification accuracy, a composite confusion matrix analysis was performed for each algorithm in Fig. 7.

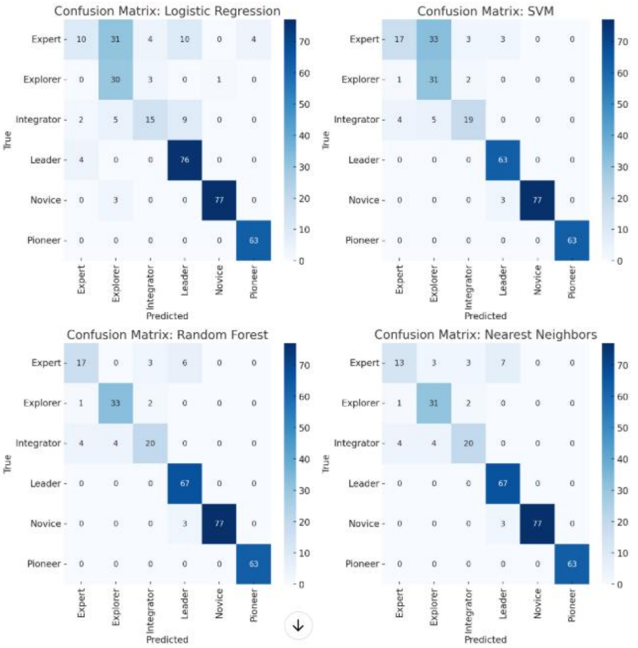


Fig. 7. Grid of Confusion Matrices for optimized models.

The diagonal values in Fig. 7 confirm a high rate of True Positives across all proficiency levels. The low off-diagonal counts demonstrate that the integrated SEM-ML framework is highly effective at minimizing misclassification errors, providing a reliable diagnostic tool for assessing the digital divide in regional education systems.

D. Feature Importance and Predictive Determinants

To enhance the transparency of the predictive framework and identify the primary drivers of the digital divide, a feature importance analysis was conducted. This stage translates the "black-box" predictions of the algorithms into actionable insights for educational policy.

1) Identifying the strongest predictors: Based on the coefficients derived from the Logistic Regression model and the feature importance rankings of the Random Forest classifier, Digital Pedagogy in Teaching (DPT) emerged as the most significant predictor of teacher ICT competency. The model yielded an Odds Ratio (OR) of 1.5176 for this feature, indicating that for every unit increase in a teacher's pedagogical integration score, the likelihood of belonging to a higher proficiency category increases by 51.76%.

2) The role of digital resource creation: The analysis further identified Digital Resource Creation and Modification (DRC) as a critical secondary predictor (OR = 1.3945). This aligns with the SEM results (Path weight = 1.27), confirming that the ability to customize digital content is a structural requirement for moving beyond the "Novice" stage of the digital divide.

3) Inverse predictors and system gaps: A compelling and critically revealing finding is the inverse relationship between Highest Educational Attainment (HighED) and ICT competency (OR = 0.2879). This negative coefficient indicates that, within this regional population, traditional advanced academic credentials do not serve as a proxy for digital proficiency. This "Academic-Digital Mismatch" highlights a systemic gap: postgraduate and teacher-education curricula in the region may be prioritizing traditional theoretical frameworks over the applied digital-pedagogical competencies defined by the DigCompEdu framework. This suggests that senior, highly-credentialed educators are not necessarily more digitally prepared; rather, they may be the most at risk of falling behind in the digital transition if their academic training is not synchronized with modern instructional technology requirements.

4) Strategic implications: The feature importance results demonstrate that the digital divide in Bukidnon is primarily pedagogical rather than technical. While "Professional Collaboration" and "Organizational Communication" are foundational (as proven in the SEM phase), the Machine Learning models prove that actual classroom implementation (Pedagogy) is the variable that ultimately decides a teacher's proficiency level.

VI. DISCUSSION

A. Characterizing the Digital Divide in Bukidnon

The integration of SEM path analysis and Machine Learning feature importance reveals that the digital divide among teachers in Bukidnon is not a binary gap of "haves vs. have-nots," but a multi-layered structural bottleneck.

1) The SEM result identifying professional engagement as the strongest driver (Path weight = 1.27) suggests that regional teachers possess the necessary organizational commitment and reflective drive to bridge the divide. However, the high variance in the Digital Resources construct indicates that this commitment does not always translate into resource access.

2) While SEM identifies commitment as the driver, the Machine Learning results identify Digital Pedagogy in Teaching (DPT) as the ultimate predictive determinant (OR = 1.5176). This synthesis suggests that while teachers are committed, the "Digital Divide" is actually a Pedagogical Divide. An educator's ability to cross into the "Expert" or "Leader" categories depends almost entirely on their capacity to integrate tools into instructional design rather than simply possessing the tools themselves.

3) The surprising inverse relationship between Highest Educational Attainment and competency (OR = 0.2879) highlights a significant regional gap. It suggests that traditional advanced academic training in the province has not yet integrated the digital competencies defined by the DigCompEdu framework, creating a scenario where senior, highly credentialed teachers may be the most at risk of falling behind in the digital transition.

B. Theoretical Implications for the DigCompEdu Framework

The study provides a successful large-scale validation (N=1,276) of the DigCompEdu framework in a predominantly agricultural, regional context.

The high global fit indices (CFI=0.948, GFI=0.935) confirm that the framework's six latent domains are cross-culturally applicable and mathematically stable even in resource-constrained environments. The structural analysis confirms the framework's theoretical assumption that *Professional Engagement* must precede *Digital Pedagogy*.

However, our findings suggest that in regional settings, the path between *Professional Commitment* and *Digital Resources* is more volatile than in urban settings, requiring the framework to perhaps place more weight on "Institutional Context" when applied to regional clusters.

C. Technical Contribution

A primary technical contribution of this research is the demonstration of the Hybrid SEM-ML Analytical Pipeline. In typical Machine Learning studies, features are often fed into models without conceptual validation. This study uses SEM as a pre-filtering layer, ensuring that only constructs with high structural validity and low measurement error are used for prediction. This reduces "noise" and increases the reliability of the classification.

From an Information Systems perspective, the success of the Random Forest model with 89% accuracy proves that SMOTE-ENN is an essential pre-processing layer for regional educational data. Balancing the sparse "Novice" and "Leader" classes, the system achieved a Macro F1-Score of 85% that far exceeds traditional linear methods.

This hybrid approach provides a blueprint for other regional government agencies to move from descriptive "snapshot" surveys to Predictive Diagnostic Systems. The model can be used to develop Decision Support Systems (DSS) that automatically identify teachers who require specific pedagogical interventions based on their structural profile.

VII. CONCLUSION AND RECOMMENDATIONS

A. Final Conclusion

This research successfully developed and validated a Hybrid Analytical Framework integrating Structural Equation Modeling (SEM) and Machine Learning (ML) to investigate the digital divide among 1,276 public school teachers in Bukidnon, Philippines. The study concludes that the digital divide is not merely a consequence of resource scarcity but a complex system of structural interdependencies.

The SEM phase established the mathematical validity of the DigCompEdu framework in a regional context, proving that Professional Engagement acts as the primary engine driving digital integration. The subsequent Machine Learning phase achieved a high predictive accuracy of 89% using the Random Forest algorithm, identifying Digital Pedagogy in Teaching (DPT) as the most critical predictive determinant of an educator's proficiency level.

Furthermore, the surprising inverse relationship between formal academic attainment and ICT competency highlights a critical mismatch between traditional higher education and the practical digital skills required in the modern classroom. Ultimately, this study demonstrates that a data-driven, hybrid approach provides the precision necessary to move beyond descriptive reports into predictive diagnostics for educational systems.

B. Study Limitations and Future Directions

While this research provides a robust hybrid SEM-ML framework for diagnosing digital competency, several limitations should be considered. First, the data were collected using a self-report Likert-scale instrument, which may be susceptible to social desirability and self-report bias; participants might overestimate their digital proficiencies. Second, the study's geographic scope is restricted to a single division (Bukidnon), which may limit the generalizability of the findings to broader or more urbanized regional contexts. Finally, the study employs a cross-sectional research design. While Structural Equation Modeling (SEM) identifies significant structural pathways, this design cannot definitively establish the causal directionality of these relationships. Future research should prioritize longitudinal studies to measure the evolution of teacher digital competency over time and incorporate objective performance data alongside self-reported metrics to further validate these findings.

C. Recommendations for Educational Policy

Based on the validated predictive weights of the model, the following interventions are recommended for the Department of Education (DepEd) and regional stakeholders:

1) Prioritize pedagogical integration over technical skills: Professional development programs should pivot from basic tool-use training to Digital Pedagogy. Training should focus on how to integrate ICT into instructional design, as this factor has the highest odds ratio for moving teachers out of the "Novice" and "Explorer" categories.

2) Bridge the 'academic-digital mismatch' for highly credentialed educators: The negative correlation between formal advanced degrees and digital competency points to a significant alignment gap between traditional Higher Education Institutions (HEIs) and the practical realities of the digital classroom. To resolve this, the Department of Education (DepEd) should initiate a collaborative curriculum audit with regional HEIs to integrate applied DigCompEdu modules into postgraduate programs. Furthermore, we recommend specialized, track-based upskilling programs for senior teachers and those with advanced degrees. These interventions should focus on moving these educators from theoretical academic mastery toward "Expert" and "Leader" levels in digital pedagogical implementation, ensuring that their tenure and qualifications are effectively leveraged to drive digital transformation.

3) Institutionalize collaborative learning communities: Given the foundational role of Professional Engagement in the SEM results, schools should establish formal digital mentorship programs that reward collaborative resource creation and reflective practice.

D. Technical Recommendations and Future Work

From a computational perspective, this study provides a blueprint for regional Data for Development initiatives.

Future research directions include system automation, algorithm expansion, and longitudinal modeling. System Automation involves integrating the validated SEM-ML pipeline into a web-based Decision Support System (DSS). This would allow educational administrators to input survey data and automatically receive real-time "Proficiency Risk" scores for their teaching staff.

For Algorithm Expansion, while Random Forest and Logistic Regression proved effective, future studies should explore Deep Learning architectures, such as Artificial Neural Networks (ANN), to see if predictive accuracy can be further enhanced for larger national-level datasets.

Longitudinal Modeling: Future work should utilize the proposed framework in a longitudinal study to measure the rate of competency growth following specific training interventions, allowing for the calculation of the "Return on Investment" (ROI) for digital transformation policies.

ACKNOWLEDGMENT

We sincerely thank the Department of Education, Bukidnon, for helping us to conduct this research. We would also like to thank Central Mindanao University for funding our research Project with a code, R-0355, which allowed us to conduct our study about ICT proficiency.

DECLARATION OF GENERATIVE AI IN SCHOLARLY WRITING

During the preparation of this manuscript, the authors utilized Google AI Studio to enhance the clarity, flow, and overall readability of the text. Following the use of this tool, the authors reviewed, validated, and edited the content to ensure accuracy. The authors take full responsibility for the integrity of the information presented. No AI tools were employed for the purposes of data analysis, hypothesis testing, or the derivation of research insights.

ETHICAL DECLARATIONS

This study was conducted in accordance with the ethical standards of Central Mindanao University and was formally approved by the Department of Education (DepEd) Division of Bukidnon. Research ethics clearance was secured from the Institutional Ethics Review Committee before the commencement of data collection. Informed consent was obtained from all 1,276 participants. To ensure participant protection, all data were anonymized and processed to maintain strict respondent confidentiality throughout the research process.

REFERENCES

[1] E. Onsay and J. Rabajante, “Measuring the Unmeasurable through Machine Learning Regressions and Classifications: Multidimensional Poverty Predictions in the Poorest Region of Luzon, Philippines,” 2024.

[2] C. S. R. Saflor, “Modeling Student Acceptance of AI Technologies in Higher Education: A Hybrid SEM–ANN Approach,” Future Internet, vol. 17, no. 12, p. 581, 2025.

[3] M. S. Asagar, “Digital competence in education: A comparative analysis of frameworks and conceptual foundations,” Synergy: Int. J. Multidiscip. Stud., vol. 2, no. 1, pp. 9–23, 2025. Online]. Available: https://sijmds.com/index.php/pub/article/view/34

[4] J. Mattar, C. C. Santos, and L. M. Cuque, “Analysis and comparison of international digital competence frameworks for education,” Education Sciences, vol. 12, no. 12, p. 932, 2022.

[5] V. Dolgopolovas and V. Dagiene, “ Competency ‐ based TPACK approaches to computational thinking and integrated STEM: A conceptual exploration,” Comput. Appl. Eng. Educ., vol. 32, no. 6, p. e22788, 2024.

[6] L. Sabyrkhanova et al., “Digital Literacy as a Tool for Developing Computational Thinking in Young Learners: A Literature Review,” Open Educ. Stud., vol. 8, no. 1, p. 20250121, 2026.

[7] C. Schreiner and C. Wiesner, “Modelling digital competence by combining computational thinking with general learning taxonomies,” Eur. Educ. Res., vol. 6, no. 1, pp. 21–42, 2023.

[8] S. K. Softić, “Teachers’ digital competences as a key factor for the digital transformation of education,” Adv. Online Educ., vol. 1, no. 1, pp. 75–86, 2022.

[9] J. Abella and E. D. Rosa, “Digital literacy and digital competence of selected Filipino teachers: Basis for a post-pandemic pedagogy,” 2023.

[10] M. R. Briones et al., “e-Work Self-Efficacy digital competence and work engagement of teachers in public secondary schools from two provinces in the Philippines,” Ho Chi Minh City Open Univ. J. Sci.-Soc. Sci., vol. 14, pp. 14–36, 2023.

[11] C. A. Velandia Rodríguez, A. F. Mena-Guacas, S. Tobón, and E. López-Meneses, “Digital teacher competence frameworks evolution and their use in Ibero-America up to the year the COVID-19 pandemic began: a systematic review,” Int. J. Environ. Res. Public Health, vol. 19, no. 24, p. 16828, 2022.

[12] R. S. Alotaibi and S. M. Alshahrani, “An extended DeLone and McLean’s model to determine the success factors of e-learning platform,” PeerJ Comput. Sci., vol. 8, p. e876, 2022.

[13] A. F. Ali, R. H. Abdullah, A. Abi Hassan, H. O. Abdullahi, and M. M. Mohamed, “COVID-19 pandemic impact on e-learning adoption and its utilization at higher education: A comparative analysis of institutions and students' perspectives,” Ingénierie des Systèmes d’Information, vol. 29, no. 2, p. 447, 2024.

[14] J. C. Vázquez-Parra, L. C. Henao-Rodriguez, J. P. Lis-Gutiérrez, I. M. Castillo-Martínez, and P. Suarez-Brito, “eComplexity: validation of a complex thinking instrument from a structural equation model,” Frontiers in Education, vol. 9, p. 1334834, Jun. 2024.

[15] X. Ju, “Unveiling the Structure of Teacher Competency: An Eight-Dimensional Model for Primary Education,” in Proc. 2025 21st Int. Conf. Comput. Intell. Security (CIS), Dec. 2025, pp. 1–5.

[16] M. K. Aydin, T. Yildirim, and M. Kus, “Teachers' digital competences: a scale construction and validation study,” Frontiers in Psychology, vol. 15, p. 1356573, 2024.

[17] F. Liaquat and M. S. Farooq, “Development and Validation of a Digital Competence Scale for Teachers at Higher Education level,” Acad. Educ. Soc. Sci. Rev., vol. 5, no. 3, pp. 348–358, 2025.

[18] M. Nowak and R. Zajkowski, “An Integrated Structural Equation Modelling and Machine Learning Framework for Measurement Scale Evaluation—Application to Voluntary Turnover Intentions,” AppliedMath, vol. 5, no. 3, p. 105, 2025.

[19] J. S. Moreira-Choez et al., “Machine Learning Algorithms to Predict Digital Competencies in University Faculty,” F1000Research, vol. 14, p. 573, 2025.

[20] L. Yang, Q. Wang, B. Zheng, X. Li, X. Ma, and T. Wang, “Assessing Digital Teaching Competence: An Approach for International Chinese Teachers Based on Deep Learning Algorithms,” Scalable Comput.: Pract. Exper., vol. 25, no. 1, pp. 495–509, 2024.

[21] R. Benzer, S. Benzer, and T. Doğan, “Assessing science teachers’ digital literacy with survey-based machine learning,” Inf. Dev., 2025, doi: 10.1177/02666669251350669.

[22] D. Yan, X. Yuan, and G. Li, “Key Factors and Predictive Models of Digital Collaborative Education Based on Machine Learning,” Annals of the New York Academy of Sciences, vol. 1555, no. 1, p. e70117, 2026.

[23] J. Cabero-Almenara, F. D. Guillen-Gamez, J. Ruiz-Palmero, and A. Palacios-Rodríguez, “Classification models in the digital competence of higher education teachers based on the DigCompEdu Framework: logistic regression and segment tree,” J. E-Learn. Knowl. Soc., no. 1, pp. 49–61, 2021.

[24] C. Magasi, “Evaluating Machine Learning Approaches in Structural Equation Modelling to Improve Predictive Accuracy in Marketing Research,” 2025.

[25] X. Li, Q. Qin, T. Zheng, and X. Wang, “Social Integration Among Rural-to-Urban Migrants: A Dynamic Analysis Using a Hybrid Approach Integrating Structural Equation Modeling and Machine Learning,” Evaluation Review, vol. 50, no. 3, pp. 423–461, 2026.

[26] . M. Brandmaier and R. C. Jacobucci, “Machine learning approaches to structural equation modeling,” in Handbook of Structural Equation Modeling, 2nd ed., R. H. Hoyle, Ed. New York, NY, USA: Guilford Press, 2023, pp. 722–739.

[27] L. Zhang, J. Gao, L. Zhao, Z. Liu, and A. Guan, “Predicting college student engagement in physical education classes using machine learning and structural equation modeling,” Appl. Sci., vol. 15, no. 7, p. 3884, Apr. 2025, doi: 10.3390/app15073884.

[28] J. Doe, "Challenges and Opportunities in Higher Education Digital Transformation: A Combined SEM and Machine Learning Perspective," Research Square, Preprint, 2026. [Online]. Available: https://assets-eu.researchsquare.com/files/rs-7708749/v1_covered_fca8c5fb-cb17-4b35-a92f-1dcadd928f1d.pdf

[29] T. Wongvorachan, O. Bulut, J. X. Liu, and E. Mazzullo, “A comparison of bias mitigation techniques for educational classification tasks using supervised machine learning,” Information, vol. 15, no. 6, p. 326, 2024.

[30] T. Wongvorachan, S. He, and O. Bulut, “A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining,” Information, vol. 14, no. 1, p. 54, 2023.

[31] Y. Chachoui, N. Azizi, R. Hotte, and T. Bensebaa, “Enhancing algorithmic assessment in education: Equi-fused-data-based SMOTE for balanced learning,” Comput. Educ.: Artif. Intell., vol. 6, p. 100222, 2024.

[32] G. Eom and H. Byeon, “Searching for optimal oversampling to process imbalanced data: Generative adversarial networks and synthetic minority over-sampling technique,” Mathematics, vol. 11, no. 16, p. 3605, 2023.

[33] M. Carvalho, A. J. Pinho, and S. Brás, “Resampling approaches to handle class imbalance: a review from a data perspective,” J. Big Data, vol. 12, no. 1, p. 71, 2025.

[34] R. Diallo, C. Edalo, and O. O. Awe, “Machine learning evaluation of imbalanced health data: a comparative analysis of balanced accuracy, MCC, and F1 score,” in Pract. Stat. Learn. Data Sci. Methods: Case Stud. LISA 2020 Global Netw., Cham, Switzerland: Springer, 2024, pp. 283–312.

[35] T. Wongvorachan, S. He, and O. Bulut, “A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining,” Information, vol. 14, no. 1, p. 54, 2023.

[36] S. Sathyanarayana and T. Mohanasundaram, “Fit indices in structural equation modeling and confirmatory factor analysis: Reporting guidelines,” Asian J. Econ. Bus. Account., vol. 24, no. 7, pp. 561–577, 2024.

[37] K. Groskurth, M. Bluemke, and C. M. Lechner, “Why we need to abandon fixed cutoffs for goodness-of-fit indices: An extensive simulation and possible solutions,” Behav. Res. Methods, vol. 56, no. 4, pp. 3891–3914, 2024.

[38] D. McNeish and M. G. Wolf, “Dynamic fit index cutoffs for confirmatory factor analysis models,” Psychol. Methods, vol. 28, no. 1, p. 61, 2023.