

Integrating Boolean Logic-Based Feature Representation and Artificial Neural Networks for Digital Divide Level Classification in Online Learning

Imalatul Hidayah^{1✉}, Ririen Kusumawati², Mochamad Imamudin³, Zainal Abidin⁴

^{1, 2, 3, 4} Departmen of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

✉ **Corresponding Author** : Imalatul Hidayah (e-mail: hidayahimalatul@gmail.com)

Article Information	ABSTRACT
Article History Received : May 25, 2026 Revised : June 14, 2026 Published : July 10, 2026 Keywords: <i>Artificial Neural Network, Classification, Digital Divide, Feature Engineering, Online Learning.</i>	Digital divide remains a major challenge in online learning, affecting students' ability to access and benefit from digital education. While previous studies have primarily focused on first-level digital divide factors or employed machine learning models using raw input variables, limited research has explored the integration of logical feature engineering and artificial intelligence to model the multidimensional nature of digital divide. This study proposes a hybrid classification framework that combines Boolean Logic-based feature engineering and Artificial Neural Networks (ANN) for digital divide classification in online learning. The proposed framework utilizes four variables, namely Primary Device, Internet Stability, Equity Score, and Accessibility Score. Boolean Logic operations (AND, OR, and XOR) were applied to generate additional logical representations of digital access conditions before the ANN classification process. Data were collected from 324 student respondents participating in online learning. The model was evaluated using accuracy, precision, recall, and F1-score metrics. Experimental results demonstrate that the optimized ANN model achieved 93% accuracy, 93% weighted precision, 93% weighted recall, and 93% weighted F1-score. Comparative analysis revealed that Boolean Logic features contributed additional logical representations of digital access patterns, although their impact on predictive performance was not consistently significant. The findings highlight that digital divide should be viewed as a multidimensional phenomenon and demonstrate the potential of ANN based predictive modeling to support data-driven digital inclusion strategies in education.

INTRODUCTION

The rapid digital transformation in higher education has accelerated the adoption of online learning as a primary mode of instruction. Although online learning provides flexibility and wider access to educational resources, its effectiveness is often constrained by digital inequality among students. This phenomenon, commonly referred to as the digital divide, reflects disparities in students' ability to access and utilize digital technologies for learning. Previous studies have shown that students with limited access to devices, unstable internet connectivity, or inadequate digital support experience greater challenges in participating effectively in online learning environments (Romero et al., 2025; Tamam & Aminudin, 2025; Tang et al., 2025; Werfhorst et al., 2022). These

inequalities ultimately influence learning quality, academic engagement, and educational outcomes (Deng & Hag, 2024; Kaur & Dutta, 2025).

Recent research suggests that the digital divide should no longer be viewed solely from the perspective of technology access. Instead, it has evolved into a multidimensional phenomenon involving connectivity quality, equity of access, institutional support, and accessibility of digital learning services (Celestin et al., 2025; Michikyan et al., 2025). For example, Werfhorst et al. (2022) demonstrated that disparities in digital readiness between students and educational institutions contribute significantly to unequal learning experiences. Similarly, Demır et al. (2023) and McIntyre, (2023) reported that limitations in devices, unstable internet connectivity, and restricted access to digital learning resources hinder students' participation in both synchronous and asynchronous learning activities. Furthermore, studies by Córdova-Esparza et al. (2025), Tamam & Aminudin (2025) and Wang & Yu (2025) emphasize that digital inequality is closely related to the broader quality of digital learning experiences rather than merely physical access to technology. Supporting this perspective, Hidayah & Kusumawati (2026), through a systematic literature review in Educational Data Mining (EDM), identified equity and accessibility as two critical dimensions that frequently influence online learning outcomes and should be considered when developing predictive and classification models in digital education environments..

Theoretical and empirical investigations on digital divide have predominantly focused on first-level digital divide factors, particularly access to devices and internet connectivity (Demır et al., 2023; Deng & Hag, 2024). While this perspective provides valuable insights into technological disparities, it is increasingly considered insufficient for explaining the complexity of online learning environments. Recent studies indicate that equity, accessibility, institutional support, digital engagement, and learning readiness also play significant roles in shaping students' digital learning experiences (Michikyan et al., 2025; Msomi & Behara, 2026). These findings suggest that digital divide classification requires a broader multidimensional perspective capable of capturing both technological and educational dimensions.

Parallel to these developments, Artificial Intelligence in Education (AIED) has emerged as a promising approach for analyzing educational data and supporting data-driven decision-making. Existing studies have demonstrated the effectiveness of artificial intelligence for educational classification, prediction, personalized learning, and academic intervention (Barrios-tao & Díaz-pérez, 2025; Estrada-molina et al., 2025; Liu & Zhong, 2025). In particular, machine learning methods have shown strong capabilities in identifying complex and nonlinear relationships among educational variables that are difficult to capture using conventional statistical approaches (Kamalov et al., 2023; Zhu et al., 2025). Within the context of digital inequality, AI-based approaches have been applied to identify inequitable access conditions and risks associated with digital services (Clara et al., 2026; Pinto et al., 2026). (Mcintyre (2023) further demonstrated the use of machine learning for analyzing online learning access from a social justice perspective, while studies by Liu & Zhong (2025), Murni (2026), and Rincón (2025) reported the effectiveness of Artificial Neural Networks (ANNs) in learning nonlinear educational patterns and predicting student outcomes.

Despite these advances, several limitations remain evident in the existing literature. First, most studies investigate digital divide factors using raw input variables without considering additional feature representations that may better capture relationships among variables. Second, previous AI based studies have primarily focused on predictive performance, while limited attention has been given to improving model interpretability in digital divide analysis. Third, the integration of feature engineering techniques and ANN models for digital divide classification in online learning remains relatively underexplored (Clara et al., 2026; Pinto et al., 2026). Given that relationships among digital divide variables are often nonlinear and condition dependent, additional feature representations may help machine learning models identify more meaningful patterns (Matjie et al., 2026; Szabó, 2024). Feature engineering has been recognized as a crucial stage in machine learning because it enhances data representation quality and facilitates the learning of complex relationships among variables (Zhu et al., 2025).

To address these research gaps, this study proposes a hybrid feature engineering Artificial Neural Network (ANN) framework for digital divide classification in online learning. The proposed framework incorporates four multidimensional variables Primary Device, Internet Stability, Equity Score, and Accessibility Score which collectively represent technology access, connectivity quality, equity, and digital learning support. In addition, Boolean Logic based feature engineering is applied through AND, OR, and XOR operations to generate additional logical representations of digital access conditions prior to the ANN classification process. ANN is selected as the primary classification model because of its proven ability to learn complex nonlinear relationships within educational data (Rincón, 2025; Zhu et al., 2025), while Boolean Logic is employed to provide supplementary logical representations that support model interpretability. Therefore, the objective of this study is to develop and evaluate a multidimensional ANN based classification model for digital divide detection in online learning. The novelty of this research lies in the integration of Boolean Logic-based feature engineering and ANN for digital divide classification. This study contributes to the literature by providing a more comprehensive representation of digital divide factors and demonstrating the potential of AI driven approaches to support data-driven digital inclusion strategies in higher education.

RESEARCH METHODS

This study uses a quantitative predictive modeling approach by integrating Boolean Logic feature engineering and Artificial Neural Networks (ANN) to classify the level of digital divide in online learning. This approach is designed to detect logical and nonlinear relationships between digital divide variables that are difficult to explain using conventional statistical methods. Previous research has shown that machine learning can model complex relationships in educational data more adaptively than linear approaches (Mcintyre, 2023; Rincón, 2025). This hybrid approach was chosen because Boolean Logic can improve the interpretability of relationships between variables, while ANN has the ability to effectively learn multidimensional nonlinear patterns (Szabó, 2024; Zhu et al., 2025).

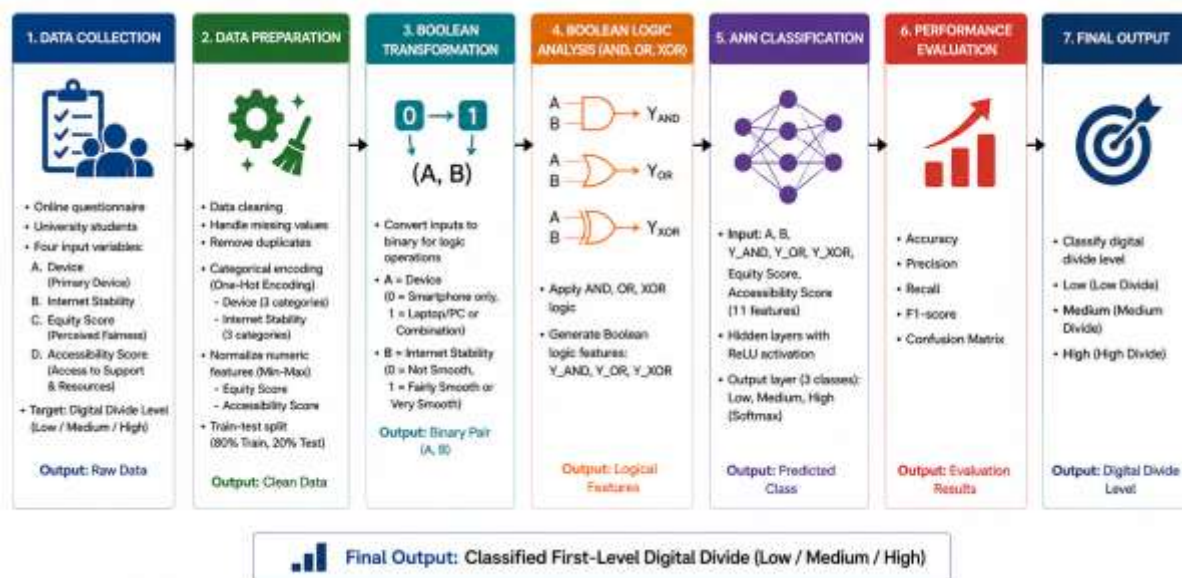


Figure 1. System Design of The Research Framework

Figure 1 shows the system design of the method proposed in this study to classify the level of digital divide in online learning using a hybrid Boolean Logic Artificial Neural Network (ANN) approach. The process begins with data collection and data preparation, followed by Boolean transformation and logic analysis (AND, OR, XOR) as the feature engineering stage, followed by classification using ANN, and concluded with model performance evaluation using a classification matrix.

Data Collection

At this stage, data was obtained through an online questionnaire distributed to 324 students at Raden Rahmat Islamic University Malang as research respondents. Data collection was conducted in December 2025. The sampling technique used in this research is Simple Random Sampling which is included in the Probability Sampling Technique. The variables used consisted of four input variables and one output variable. The Primary Device and Internet Stability variables were chosen to represent the basic technology access dimension, which is theoretically included in the first-level digital divide (Demir et al., 2023; Deng & Hag, 2024). Meanwhile, the Equity Score and Accessibility Score were used to represent the broader dimensions of fairness in access and support for digital learning (Michikyan et al., 2025; Msomi & Behara, 2026). We present the research variables in Table 1.

Table 1. Research Variables

Code	Variable	Type	Categories
X1	Primary Device	Categorical	Smartphone only, Laptop/PC, Combination
X2	Internet Stability	Categorical	Very Smooth, Fairly Smooth, Not Smooth
X3	Equity Score	Numerical	Composite score
X4	Accessibility Score	Numerical	Composite score
Y	Digital Divide Level	Output	Low, Medium, High

Table 1 presents the research variables used to identify the level of digital divide in online learning. The input variables consist of two categorical variables: Primary Device (X1), which describes the primary device used by students to access online learning (smartphone only, laptop/PC, or a combination of both), and Internet Stability (X2), which indicates the level of internet connection stability during the learning process (very smooth, fairly smooth, and not smooth). In addition, this study also uses two numeric variables: Equity Score (X3) and Accessibility Score (X4), which are composite scores resulting from measuring aspects of equal access and ease of access to digital learning technology. These four variables are used as predictors to determine the output variable, namely the Digital Divide Level (Y), which is classified into three categories: Low, Medium, and High. The selection of these variables is based on the consideration that the digital divide in online learning is influenced not only by technological infrastructure factors, but also by aspects of equity and accessibility, which together form the level of digital inequality among students.

Data Preparation

The data preparation stage is a crucial process in machine learning-based research because the quality of the input data significantly impacts the performance of the resulting classification model. In the context of educational data, variables derived from primary data generally have diverse characteristics, such as categorical data, numeric data, missing values, and uneven data distribution, requiring preprocessing before being used in modeling (Mcintyre, 2023; Rincón, 2025). Previous research has also shown that appropriate preprocessing, such as data cleaning, encoding, normalization, and data splitting, can improve the quality of data representation and help machine learning models learn patterns of relationships between variables more effectively (Kamalov et al., 2023; Zhu et al., 2025). Therefore, in this study, the data preparation stage was carried out through several main processes, namely data cleaning, one-hot encoding, normalization, and train-test split, to ensure that the data is ready for use in the feature engineering and classification stages using Artificial Neural Networks (ANN). The data preparation process in this study consisted of four sequential stages, which are described as follows:

1. Data Cleaning

This stage includes removing missing values, removing duplicate data, and validating the consistency of data categories. Previous research has shown that preprocessing is crucial

for improving the performance of machine learning and neural network models (Mcintyre, 2023; Rincón, 2025).

2. One-Hot Encoding

Because the Primary Device and Internet Stability variables are non ordinal categorical data, this study uses one-hot encoding to convert the categories into binary representations for processing by an Artificial Neural Network (ANN). This approach was chosen because it allows each category to be represented independently without assuming any numerical relationship between categories, making it more suitable for neural network based machine learning modeling (Kamalov et al., 2023; McIntyre, 2023; Zhu et al., 2025).

3. Normalization

Numeric variables (Equity Score and Accessibility Score) were normalized using Min-Max Scaling to ensure all values were within a range. Normalization was performed to improve training stability in neural network models (Kamalov et al., 2023) with equation 1.

$$x'_{i,k} = \frac{x_{i,k} - x_{k,\min}}{x_{k,\max} - x_{k,\min}} \quad (1)$$

4. Train Test Split

The dataset was divided into 80% training data and 20% testing data for model training and evaluation. The data was divided using a stratified split to maintain a proportional distribution of output classes in the training and testing data, allowing for a more representative model evaluation process (Mcintyre, 2023; Rincón, 2025).

Boolean Transformation

Two main digital access variables are transformed into Boolean representations as part of the feature engineering process to simplify digital access conditions into two logical states, namely conditions that are less supportive (0) and conditions that are supportive (1). Boolean transformations are used to build logical representations that are then used in the Boolean Logic analysis stage through AND, OR, and XOR operations (Kamalov et al., 2023; Szabó, 2024; Zhu et al., 2025). In this study, the Primary Device variable is transformed into a Device Boolean (A), while the Internet Stability variable is transformed into an Internet Boolean (B). Transformation rules are determined based on functional interpretations of digital access conditions, where categories considered less supportive of online learning are given a value of 0, while categories considered supportive are given a value of 1. The transformation rules are presented in Table 2.

Table 2. Device Boolean and Internet Boolean

Category	Value	Category	Value
Smartphone only	0	Not Smooth	0
Laptop/PC	1	Fairly Smooth	1
Combination	1	Very Smooth	1

Boolean Logic Analysis

This study applies three main logic operations AND, OR, and XOR-as part of feature engineering to generate additional logic features used as input to an Artificial Neural Network (ANN) model. These logic operations are not used as replacements for the primary variables, but as additional logical representations to help the model enrich the representation of digital access conditions in a more interpretable way (Kamalov et al., 2023; Szabó, 2024; Zhu et al., 2025). In this study, Boolean variable A represents the condition of the digital device, while Boolean variable B represents the internet stability condition. The AND operation is used to identify conditions

when both digital access factors are simultaneously in a state that supports online learning. This operation produces a value of 1 only when both variables are in good condition, which is mathematically formulated as:

$$Y \text{ AND} = A \wedge B$$

The OR operation is used to represent a condition where one of the digital access factors is sufficient to support online learning. This operation returns a 1 if at least one of the conditions is in a good state, which is mathematically formulated as:

$$Y \text{ OR} = A \vee B$$

Meanwhile, the XOR operation is used to represent the condition of digital access imbalance, namely when only one of the factors is in good condition, while the other factor is not supportive, which is mathematically formulated as:

$$Y \text{ XOR} = A \oplus B$$

The XOR operation is used to describe the digital divide that can arise due to imbalanced technology access, for example, students have adequate devices but poor internet connections, or have stable internet but only use devices with functional limitations. Such conditions indicate a relationship between digital access that is not always linear and can provide additional representation of the digital divide pattern in online learning. In this study, the Boolean features AND, OR, and XOR were used as logical feature representations that complement the main variables, namely Primary Device, Internet Stability, Equity Score, and Accessibility Score, before the classification process was carried out using ANN. This approach was chosen because feature engineering can help enrich input representation and improve the interpretability of relationships between variables in machine learning models (Kamalov et al., 2023; Zhu et al., 2025), although the performance of the main classification is still influenced by the overall combination of multidimensional variables.

ANN Classification

An Artificial Neural Network model was used to classify the digital divide into three classes: low, medium, and high. ANN was chosen because it has the ability to learn nonlinear relationships between educational variables more adaptively than linear statistical methods (Borgonovi et al., 2025; Estrada-molina et al., 2025; Murni, 2026; Rincón, 2025; Zhu et al., 2025). The model input consists of 11 features: Device (3 one-hot features), Internet Stability (3 one-hot features), Equity Score, Accessibility Score, AND, OR, XOR. The ANN architecture is presented in Table 3.

Table 3. ANN Architecture

Layer	Neurons	Activation
Input	11	-
Hidden Layer 1	32	ReLU
Hidden Layer 2	16	ReLU
Hidden Layer 3	8	ReLU
Output Layer	3	Softmax

Table 3 shows the Artificial Neural Network (ANN) architecture used for digital divide classification. The input layer consists of 11 neurons representing all preprocessing and feature engineering variables. The model uses three hidden layers with 32, 16, and 8 neurons, respectively, and a Rectified Linear Unit (ReLU) activation function to gradually learn the nonlinear relationships between input variables. The output layer uses three neurons with a Softmax activation function to generate classification probabilities into three digital divide categories: Low, Medium, and High. This architecture is designed to balance the model's ability to learn complex

data patterns with computational efficiency during the classification process. Neuron computation is performed using Equation 2.

$$z = \sum_{i=1}^n w_i x_i + b \quad (2)$$

The hidden layer uses the ReLU activation function with equation 3. The ReLU activation function is used in the hidden layer because it is able to reduce vanishing gradients and increase model training efficiency (Kamalov et al., 2023)

$$f(x) = \max(0, x) \quad (3)$$

Meanwhile, the output layer uses the Softmax function with equation 4 to produce multi-class classification probabilities.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (4)$$

In the training process, the model uses categorical cross-entropy with equation 5 as the loss function to measure classification error.

$$L = -\sum_{i=1}^N y_i \log(\hat{y}_i) \quad (5)$$

To address class imbalance, this study used class weighting in the training process, giving minority classes greater weight. This approach has been shown to improve classification performance on imbalanced educational data (Rincón, 2025).

Performance Evaluation

Model performance was evaluated using several classification metrics, namely accuracy, precision, recall, F1-score, and confusion matrix, to measure the model's ability to accurately classify the digital divide. The use of multiple evaluation metrics is necessary so that model performance is assessed not only by the overall level of prediction accuracy but also by the model's ability to more comprehensively recognize each output class (Mcintyre, 2023; Rincón, 2025; Zhu et al., 2025).

Accuracy was calculated using equation 6.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision is calculated using equation 7.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

Recall is calculated using equation 8.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

F1-Score is calculated using equation 9.

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

Through this evaluation, this study aims to produce a digital divide classification model that not only has high accuracy performance but is also more interpretable, nonlinear aware, and data-driven, so that it is able to provide more comprehensive insights in analyzing the digital divide in online learning.

Final Output

The final result of this study is the Level of Digital Divide, which serves as the target variable for the Artificial Neural Network (ANN) classification model. The target variable, namely the Level of Digital Gap, was adopted from the results of previous clustering studies conducted on the same dataset. The grouping process produces three categories, namely Low, Medium, and High Digital Gap which are then used as class labels for supervised classification using ANN. Therefore, this study focuses on predicting the level of previously identified digital divides rather than generating new clusters.

The use of clustering based labels was motivated by the multidimensional nature of digital divide, which encompasses not only technology access but also equity and accessibility dimensions. Previous studies have highlighted that digital inequality in online learning cannot be adequately represented using simple binary categories because students may experience different levels of digital disadvantage across multiple dimensions (Hidayah & Kusumawati, 2026; Michikyan et al., 2025). Therefore, clustering was employed to identify natural groupings within the data and to construct more representative digital divide categories.

The resulting clusters were interpreted according to their overall digital divide characteristics. Students assigned to the Low Digital Divide cluster generally exhibited favorable conditions across device access, internet stability, equity, and accessibility dimensions. The Medium Digital Divide cluster represented students experiencing moderate limitations in one or more dimensions, while the High Digital Divide cluster consisted of students facing substantial barriers that could negatively affect their participation in online learning.

RESULTS AND DISCUSSION

This section presents the research results based on the stages designed in the proposed framework, starting from the characteristics of respondent data, data preparation results, Boolean Logic feature engineering, classification using Artificial Neural Network (ANN), model performance evaluation, to comparative analysis of the contribution of Boolean features. The analysis is carried out in stages to show how each process contributes to the formation of a digital divide classification model in online learning. In addition, the results obtained are also interpreted to understand the effectiveness of the hybrid feature engineering ANN approach in identifying the level of digital divide more accurately, interpretably, and data driven.

Respondent Data Characteristics

Initial distribution of respondent data indicates that students' digital access is not homogeneous. Some respondents have adequate devices and connectivity, while others still experience limited access, potentially creating a digital divide.

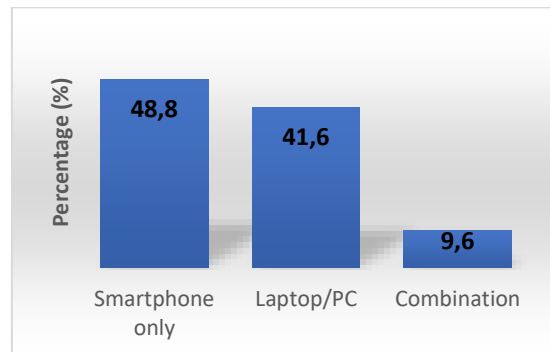


Figure 2. Primary Device Distribution

Figure 2 shows that smartphones alone are the most commonly used device by students in online learning, at 48.8%, followed by laptops/PCs at 41.6%, and a combination device at 9.6%. These results indicate that smartphones remain the primary device due to their greater accessibility and affordability. However, the dominance of smartphone-only use may also indicate limitations in supporting more complex learning activities compared to laptops or PCs. Meanwhile, the low use of a combination of devices indicates that only a small proportion of students have greater flexibility in accessing technology. These findings provide an initial overview of the conditions under which digital devices are used, potentially contributing to the student digital divide.

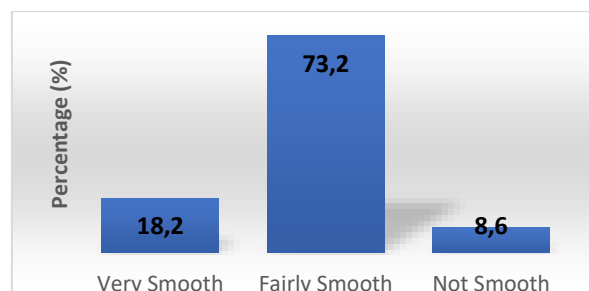


Figure 3. Internet Stability Distribution

Figure 3 shows that the majority of students are in the Fairly Smooth category with a proportion of 73.2%, followed by Very Smooth at 18.2%, and Not Smooth at 8.6%. These results indicate that most students have a relatively stable internet connection to support online learning, although it is not yet fully optimal. Meanwhile, there are still a small number of students who experience an unstable internet connection, which has the potential to hinder the learning process. This finding confirms that internet stability is a crucial factor that can influence the level of students' digital divide.

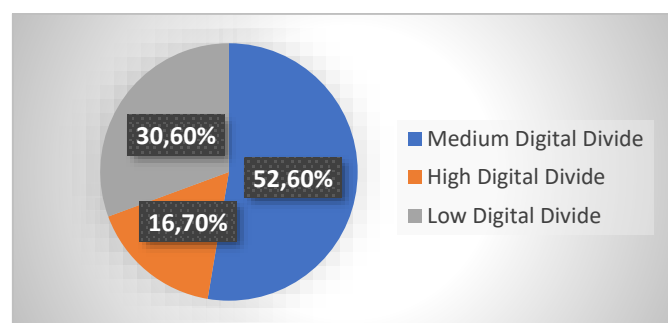


Figure 4. Digital Divide Class Distribution

Figure 4 shows that the Medium Digital Divide category is the most dominant group with a proportion of 52.6%, followed by Low Digital Divide at 30.6%, and High Digital Divide at 16.7%. These results indicate that the majority of students are at a medium level of digital divide, with adequate access to technology but still facing several limitations. Meanwhile, some students have good digital access, but there are still groups experiencing a high digital divide. This finding confirms that the condition of the student digital divide is not homogeneous, so more targeted interventions are needed according to the characteristics of each group.

Boolean Logic Feature Engineering Results

Unlike conventional classification approaches that rely solely on raw input variables, this study incorporates Boolean Logic based feature engineering to provide additional representations of digital access patterns. The generated AND, OR, and XOR features capture different logical combinations of device and internet conditions. Their distributions are shown in Figure 5.

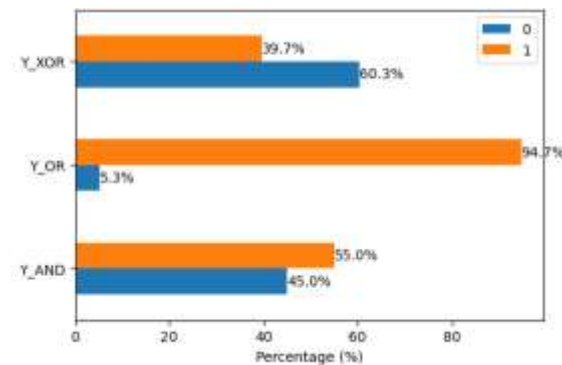


Figure 5. Boolean Logic Distribution

Figure 5 shows the distribution of the results of the Boolean logic operations AND, OR, and XOR formed from the variables A_bool (Device Boolean) and B_bool (Internet Boolean). In Y_AND, a value of 1 at 55% indicates that more than half of the respondents have adequate devices and a fairly stable internet connection. In Y_OR, the distribution is dominated by the value of 1 at 94.7%, indicating that almost all respondents have at least one factor supporting online learning. Meanwhile, in Y_XOR, the value of 0 at 60.3% is more dominant than the value of 1 at 39.7%, indicating that most respondents have relatively consistent device and internet conditions. Overall, this Boolean feature provides an additional representation that is able to capture the interaction between device and internet variables, thus enriching the input of the Artificial Neural Network (ANN) model in detecting digital divide patterns.

ANN Classification Results

After the feature engineering stage, the classification process was performed using an Artificial Neural Network (ANN). To evaluate the learning behavior of the model during training, the training and validation accuracy as well as the corresponding loss values were monitored across epochs. These metrics provide insights into the model's convergence, learning stability, and potential overfitting issues. The training results are presented in Figures 6.

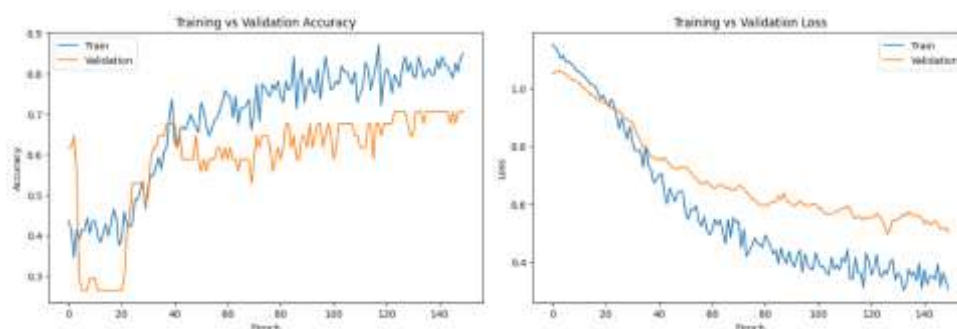


Figure 6. Training vs Validation Accuracy and Loss

Figure 6 presents the training and validation accuracy and loss curves of the ANN model across training epochs. The results show that training and validation accuracy generally increased, while both training and validation loss decreased throughout the learning process, indicating that the model successfully learned the underlying patterns of digital divide classification. Training accuracy reached approximately 88-90%, whereas validation accuracy stabilized at around 70-76%. Similarly, training loss decreased substantially and validation loss remained relatively stable toward the final epochs. Although a small gap between training and validation metrics suggests mild overfitting, the validation curves remained stable without significant degradation. Overall, these results indicate that the ANN model achieved satisfactory convergence and demonstrated adequate generalization capability for digital divide classification.

Digital Divide Level Classification Results

The proposed ANN model classified students into three digital divide categories, namely Low Digital Divide, Medium Digital Divide, and High Digital Divide. These categories were adopted from the labels generated in a previous clustering study and represent different levels of digital inequality experienced by students in online learning environments. The Low Digital Divide category represents students who generally have adequate access to digital devices, stable internet connectivity, equitable learning opportunities, and sufficient accessibility support. The Medium Digital Divide category includes students who experience moderate limitations in one or more dimensions of digital learning. Meanwhile, the High Digital Divide category consists of students facing substantial barriers related to technology access, connectivity, equity, and accessibility, potentially limiting their effective participation in online learning. Figure 7 presents the classification performance for each digital divide category.

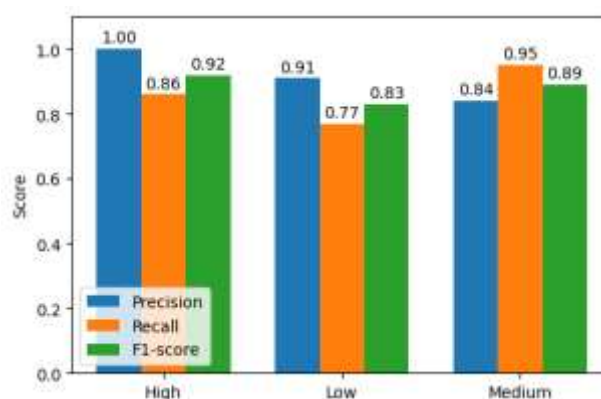


Figure 7. Classification Performance

Figure 7 indicate that the ANN model performed consistently across all classes. The High Digital Divide category achieved perfect precision, recall, and F1-score values, indicating that students experiencing severe digital inequality were identified without misclassification. This finding is particularly important because the accurate detection of highly disadvantaged students is essential for designing targeted digital inclusion interventions. For the Low Digital Divide category, the model achieved a precision of 0.91 and a recall of 0.77, indicating that most students with favorable digital learning conditions were correctly classified. Similarly, the Medium Digital Divide category achieved an F1-score of 0.89, demonstrating strong classification performance despite representing a more heterogeneous group. Overall, these results suggest that the proposed ANN model can effectively distinguish varying levels of digital divide among students.

Model Performance Evaluation

The classification performance of the ANN model was evaluated using a confusion matrix and several performance metrics. The confusion matrix was first analyzed to examine the distribution of correct and incorrect predictions across the three digital divide categories. The resulting confusion matrix is shown in Figure 8.

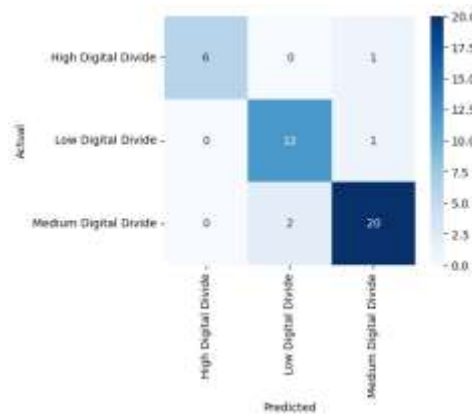


Figure 8. Confusion Matrix

Figure 8 shows that the Artificial Neural Network (ANN) model was able to classify the three digital divide categories quite well. In the High Digital Divide category, the model successfully classified 6 data points correctly and only 1 data point was incorrectly predicted as Medium. In the Low Digital Divide category, 12 data points were correctly predicted and 1 data point was incorrectly predicted as Medium. Meanwhile, in the Medium Digital Divide category, the model successfully classified 20 data points correctly, with 2 data points incorrectly predicted as Low. Overall, the main diagonal values in the confusion matrix (6, 12, and 20) indicate that most of the data points were correctly classified, while misclassification errors were relatively small and tended to occur in adjacent categories. These results indicate that the ANN model has good classification performance in detecting students' levels of digital divide.

Contribution of Boolean Logic Feature Engineering

A comparative analysis was conducted to evaluate the contribution of Boolean Logic based feature engineering. The results show that the inclusion of AND, OR, and XOR features did not consistently improve predictive performance compared with the baseline ANN model. While the OR representation produced a slight improvement in classification performance, the overall proposed model achieved a performance level comparable to the baseline ANN. Although the quantitative improvement was limited, Boolean Logic features provided additional logical representations of digital access conditions that cannot be directly observed from the original variables. For example, the XOR operation explicitly represents situations in which students possess adequate digital devices but experience unstable internet connectivity, or vice versa. Such conditions are difficult to identify through individual variables alone but are important for understanding the multidimensional nature of digital divide. Therefore, the primary contribution of Boolean Logic in this study lies not in significantly increasing prediction accuracy, but in enhancing the interpretability of digital access patterns. The generated logical features provide a complementary representation that helps explain how different combinations of access conditions contribute to digital inequality.

This finding has important implications for the novelty of the proposed framework. Rather than positioning Boolean Logic as a mechanism for maximizing predictive performance, this study demonstrates how logical feature representations can be integrated with ANN based classification to provide a more interpretable understanding of digital divide conditions. The novelty therefore lies in the combination of multidimensional digital divide variables, Boolean Logic based feature engineering, and ANN modeling within a unified framework. This approach contributes to the literature by extending digital divide classification beyond conventional prediction oriented models toward a framework that simultaneously supports prediction and interpretation of digital access patterns in online learning environments.

Comparative Analysis

A comparative analysis was performed to assess whether Boolean Logic based features contribute to classification performance beyond the original ANN model. By comparing the

accuracy and F1-score obtained from different logical representations, this analysis helps identify the practical role of Boolean features in digital divide classification. The comparison results are illustrated in Figure 9.

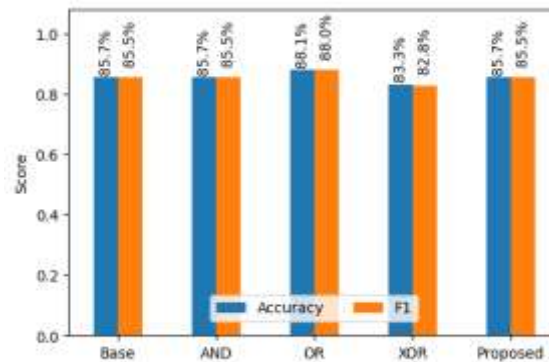


Figure 9. Performance Comparison of Boolean Logic Features

Figure 9 shows a comparison of the performance of the ANN model with the addition of the Boolean logic features AND, OR, XOR, and the Proposed model. The Base model produced an accuracy of 85.7% and an F1-score of 85.5% as a baseline. The addition of the AND feature did not produce significant changes, with the results remaining the same as the Base model. In contrast, the OR feature produced the best performance with an accuracy of 88.1% and an F1-score of 88.0%, indicating that this feature is the most effective in improving classification ability. Meanwhile, the XOR feature actually decreased performance to an accuracy of 83.3% and an F1-score of 82.8%, making it less effective as an additional feature. In the Proposed model, which combines all Boolean features, the performance was again the same as the Base model, namely an accuracy of 85.7% and an F1-score of 85.5%. Overall, these results indicate that the OR logic feature is the most effective Boolean feature in improving ANN classification performance, while adding all Boolean features simultaneously does not always produce optimal results.

The results of this study indicate that the digital divide in online learning is better understood as a multidimensional phenomenon, not solely influenced by basic technological access such as devices and internet connectivity. This is evident in the improved model performance after the Equity Score and Accessibility Score variables were added, resulting in the final model achieving 93% accuracy with a weighted F1-score of 93%. This finding suggests that equity in access and digital learning support play a significant role in determining the level of digital divide among students, consistent with previous research suggesting that the digital divide in education is influenced by a combination of technical and non-technical factors (Herman, 2022; Michikyan, 2025). Methodologically, the use of Artificial Neural Networks (ANNs) has proven effective in studying the nonlinear relationships between digital divide variables. The model performed very well in the High Digital Divide category, with precision, recall, and F1-score each reaching 100%, demonstrating the model's ability to detect student groups most at risk of experiencing the digital divide. These results support previous studies showing that ANN has high performance in classifying complex educational data (Guzman, 2025; Zhu, 2025).

Comparative analysis shows that the Boolean features AND, OR, and XOR do not provide significant improvements in classification performance compared to the Base Model, but they still provide additional logical representation of students' digital access conditions. Thus, the primary contribution of Boolean Logic in this study lies more in the aspects of feature representation and interpretability, while the primary classification performance is determined by the strength of the multidimensional variables learned through ANN. The findings also provide an important methodological insight regarding the role of feature engineering in digital divide classification. Although Boolean Logic features did not produce substantial performance gains, the results indicate that the original multidimensional variables Primary Device, Internet Stability, Equity Score, and Accessibility Score already contained strong predictive information. Consequently, the ANN model was capable of learning most relevant patterns directly from the input features.

Nevertheless, Boolean Logic representations remain valuable because they offer an interpretable view of digital access relationships and facilitate a better understanding of how different access conditions interact within online learning environments.

This finding suggests that the novelty of the proposed framework should be viewed from the perspective of integrating logical representations and predictive modeling rather than solely from performance improvement. The framework demonstrates how interpretable feature engineering can be combined with ANN-based classification to support a more comprehensive analysis of digital divide phenomena in education. Overall, this study demonstrates that the hybrid feature engineering–ANN approach is an effective framework for classifying the digital divide in online learning, as it is able to produce models that not only have high classification performance but are also more interpretable and data-driven, thus being able to be used as a basis for designing more targeted digital inclusion policies..

CONCLUSION

This study developed an Artificial Neural Network (ANN)-based framework for classifying digital divide levels in online learning using multidimensional factors, including Primary Device, Internet Stability, Equity Score, and Accessibility Score. The findings demonstrate that digital divide should not be viewed solely as a problem of technology access, but rather as a multidimensional phenomenon influenced by connectivity, equity, and accessibility conditions. The proposed model was able to effectively classify students into three digital divide categories Low, Medium, and High Digital Divide indicating its potential to support the identification of students with different levels of digital disadvantage in online learning environments.

The results further show that multidimensional variables contribute substantially to classification performance, highlighting the importance of considering both technical and non-technical aspects when analyzing digital inequality in education. Although Boolean Logic-based features (AND, OR, and XOR) did not produce substantial improvements in predictive performance, they provided additional logical representations of digital access conditions and demonstrated the feasibility of integrating feature engineering with ANN based classification. Therefore, the contribution of this study lies in combining multidimensional digital divide indicators, logical feature representation, and predictive modeling within a unified framework.

This study has several limitations. First, the classification model was developed using data collected from 324 respondents, which may limit its generalizability to broader educational contexts. Second, the input variables were restricted to four dimensions and did not include other potentially influential factors such as digital literacy, socioeconomic background, or academic engagement. Future studies are therefore encouraged to employ larger and more diverse datasets, incorporate additional digital divide dimensions, and compare ANN with other machine learning approaches to further evaluate the robustness and generalizability of digital divide classification models.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang, for the academic support provided during the completion of this study.

REFERENCES

- Barrios-tao, H., & Díaz-pérez, V. (2025). Impact of artificial intelligence on educational actors (2015-2023). *Revista Colombiana de Ciencias Sociales*, 16(1), 240–276. <https://doi.org/10.21501/22161201.4803>

- Borgonovi, F., Bastagli, F., Ochojska, M., & Piumatti, G. (2025). AI adoption in the education system: International insights and policy considerations for Italy. *OECD Artificial Intelligence Papers*, 52. <https://doi.org/10.1787/69bd0a4a-en>
- Celestin, M., Pal, T., Mungekar, P. R., Parmar, H., & Asamoah, P. J. (2025). The Future of Online Learning Innovations, Accessibility, and the Digital Divide. *International Journal of Advanced Trends in Engineering and Technology (IJATET)*, 10(1), 33–42. <https://doi.org/10.2139/ssrn.5212095>
- Clara, J., Jean-flavien, B., Grazia, C., Nessrine, O., & Paolo, P. (2026). Digital divide and artificial intelligence for health. *Technovation*, 151, 103392. <https://doi.org/10.1016/j.technovation.2025.103392>
- Córdova-Esparza, D.-M., Romero-González, J.-A., Terven, J., Córdova-Esparza, K.-E., López-Martínez, R.-E., García-ramírez, T., & Chaparro-SánchezRicardo. (2025). Predicting and Preventing School Dropout with Business Intelligence: Insights from a Systematic Review. *Information*, 16(326), 1–30. <https://doi.org/10.3390/info16040326>
- DemİR, İ., Olgun, C. K., & Özsöz, C. (2023). Confronting the First-Level Digital Divide in the Digital Age: A Comparison Between Public and Private University Students in Türkiye. *Journal of Economy Culture and Society*, 67, 135–153. <https://doi.org/10.26650/JECS2022-1213957>
- Deng, X. N., & Hag, S. El. (2024). Information Systems Education Digital Inequality and Two Levels of the Digital Divide in Online Learning: A Mixed Methods Study of Underserved College Students. *Journal of Information Systems Education*, 35(3), 377–389. <https://doi.org/10.62273/SSIF6302> Article
- Estrada-molina, O., Luis, J., Zapatero, R., Lacueva, P. R., Zambrano-acosta, J. M., Fabricio, L., & Chica, C. (2025). Artificial intelligence in primary education: systematic review (2000-2024). *Education in the Knowledge Society (EKS)*, 26, e32300. <https://doi.org/10.14201/eks.32300>
- Hidayah, I., & Kusumawati, R. (2026). Educational Data Mining in Online Learning: Data Mining Techniques and Algorithms, Factors, Equity and Accessibility Dimensions (A Systematic Literature Review). *G-Tech: Jurnal Teknologi Terapan*, 10(1), 190–205. <https://doi.org/10.70609/g-tech.v10i1.8628>
- Kamalov, F., Calonge, D. S., & Gurrib, I. (2023). New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution. *Sustainability*, 15, 12451. <https://doi.org/10.3390/su151612451>
- Kaur, R., & Dutta, S. (2025). Impact of the Digital Divide on Digital Learning: A Review Paper. *The Academic (International Journal of Multidisciplinary Research)*, 3(6), 728–742. <https://doi.org/10.5281/zenodo.15855529> ARTICLE
- Liu, X., & Zhong, B. (2025). Integrating generative Artificial Intelligence into student learning: A systematic review from a TPACK perspective. *Educational Research Review*, 49, 100741. <https://doi.org/10.1016/j.edurev.2025.100741>
- Matjie, M. A., Nethavhani, A., & Matlakala, M. (2026). AI and the digital divide in education. *Frontiers in Computer Science*, 8, 1759027. <https://doi.org/10.3389/fcomp.2026.1759027>
- Mcintyre, N. A. (2023). Access to online learning : Machine learning analysis from a social justice perspective. *Education and Information Technologies*, 28(4), 3787–3832. <https://doi.org/10.1007/s10639-022-11280-5>
- Michikyan, M., Subrahmanyam, K., Regan, P., Ángel, M., Linda, C., Harkness, A., & Schwartz, S. J. (2025). Differences in and relationship between digital access, use, connectivity, skill, engagement, and motivation: implications for digital inequality in online learning among first-generation and continuing-generation undergraduate students. *Discover Education*,

- 4(9). <https://doi.org/10.1007/s44217-024-00371-9>
- Msomi, A. M., & Behara, K. (2026). Investigating Digital Divide Barriers, Institutional Support, and Students' Perceptions of AI-Driven Mathematics Learning. *Education Sciences*, 16, 442. <https://doi.org/10.3390/educsci16030442>
- Murni, S. (2026). Artificial intelligence in higher education management a bibliometric analysis of challenges opportunities and future research directions. *Discover Education*, 5, 276. <https://doi.org/10.1007/s44217-026-01165-x>
- Pinto, A. S., Longhi, M. T., & Behar, P. A. (2026). Digital competencies of students for the use of artificial intelligence in distance higher education: a systematic literature review. *RDBCI: Revista Digital de Biblioteconomia e Ciência Da Informação*, 24, e026013. <https://doi.org/10.20396/rdbci.v24i00.8679434/en>
- Rincón, A. G. (2025). Dropout in rural higher education in virtual modality: a neural network approach n. *Cogent Education*, 12(1), 2493445. <https://doi.org/10.1080/2331186X.2025.2493445>
- Romero, C. S., Vacas, C. G., Munoz, E. M., Levterova, D., & Lazarov, A. (2025). Digital divide and accessibility in online education. *Edelweiss Applied Science and Technology*, 9(4), 2548–2560. <https://doi.org/10.55214/25768484.v9i4.6605>
- Szabó, R. Z. (2024). Overcoming the digital divide: A conceptual framework. *Journal of Infrastructure, Policy and Development*, 8(16), 1–15. <https://doi.org/10.24294/jipd10082>
- Tamam, B., & Aminudin. (2025). Bridging The Digital Divide In Education: A Systematic Review (SLR-Prisma) Of Digital Inclusion Strategies. *Proceeding of International Conference on Education, Society and Humanity*, 3(1), 138–153.
- Tang, Q., Kamarudin, S., Abdul Rahman, S. N., & Zhang, X. (2025). Bridging gaps in online learning: A systematic literature review on the digital divide. *Journal of Education and Learning*, 14(1), 161–176. <https://doi.org/10.5539/jel.v14n1p161>
- Wang, J., & Yu, Y. (2025). Machine learning approach to student performance prediction of online learning. *PLoS ONE*, 20(1). <https://doi.org/10.1371/journal.pone.0299018>
- Werforst, H. G. Van De, Kessenich, E., & Geven, S. (2022). The digital divide in online education: Inequality in digital readiness of students and schools. *Computers and Education Open*, 3, 100100. <https://doi.org/10.1016/j.caeo.2022.100100>
- Zhu, W., Wei, L., & Qin, Y. (2025). Artificial Intelligence in Education (AIEd): Publication Patterns, Keywords, and Research Focuses. *Information*, 16, 725. <https://doi.org/10.3390/info16090725>