Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 11 Issue 01
2026

Research Article
https://doi.org/10.56979/1101/2026/1347

Hybrid Attention-Enhanced Deep Learning for Adaptive Digital Literacy Prediction in College Learners

Yuhan Chen[1*]

[1]School of Education, South China Normal University, Guangzhou, 510631, China.
[*]Corresponding Author: Yuhan Chen. Email:chenyuhan1511@163.com

Received: February 17, 2026 **Accepted:** May 03, 2026

Abstract: Digital literacy has become a pillar of rural revitalization efforts across the globe, and the creation of smart, adaptive systems to forecast and tailor learner outcomes constitutes a major area of research gap. This paper presents a new Hybrid Attention-Enhanced Deep Learning Network (HAE-DLN) a multi-task deep learning model that predicts overall scores on digital literacy (regression), classifies learners into proficiency levels (3-class), and classifies above/below-median learners (binary classification). The HAE-DLN is built upon a Feature Attention Module (FAM), Residual Dense Blocks (RDB), as well as a Multi-Task Learning Head (MTL), whilst being trained on a complete synthetic dataset of 1,000 rural learners, covering 23 demographic, behavioral, engagement, and outcome variables, in one end-to-end trainable framework. The first comparison on the baseline with the Random Forest and the Gradient Boosting regressors in their tuning with the help of the GridSearchCV proves that the proposed model has a better predictive power. Findings indicate that post-training scores, quiz performance and engagement efficiency are most effective predictors of digital literacy outcomes. The results have a direct implication on the design of adaptive, personalized digital literacy initiatives in rural and underserved people.

Keywords: Digital literacy; rural education; deep learning; attention mechanism; multi-task learning; adaptive learning systems; feature importance; rural revitalization

1. Introduction

1.1. Background and Motivation

The fast spread of digital technology in all spheres of contemporary society has radically changed the essence of economic activity, civil activity, and social mobility. The availability of digital tools and the skills needed to effectively interact with them have become as essential as traditional literacy in defining how well one can cope with the life of the modern world [1]. However, even after spectacular improvements in the connectivity infrastructure, there remains an enduring and highly profound gap between urban and rural people in terms of access to, and expertise in, digital technologies [2].

Digital rural disadvantages persist as inherent systemic social inequities in rural areas in both the developed and developing world. All these disadvantages are complicated by a lack of access to formal education, low average household income, geographic separation to training centers, and lack of locally relevant digital content [3]. These effects go well beyond personal inconvenience: rural digital illiteracy restricts agricultural modernization, access to e-government services, participation in digital economies, and the efficacy of public health communication all of which are indispensable pillars of the rural revitalization policy [4].International organizations and governments have responded by implementing mass digital literacy initiatives to rural communities. Nevertheless, the performance of such programs is enormous, and the standardized one-size-fits-all programs do not consider the heterogeneous nature of the learning profiles, the motivational levels, and the previous level of knowledge that define the rural learner populations [5]. It is thus necessary that adaptive and customized learning systems intelligent platforms be developed that can dynamically change the instructional content, pacing, and feedback in response to real-time evaluation of individual learner characteristics and progressions [6].

The technologies of deep learning and machine learning promise revolutionary possibilities in this respect. These models are capable of providing accurate predictions of learner performance, predicting at-risk individuals before they become disengaged and administering specific interventions by learning complex and non-linear relationships between learner attributes, engagement behaviors, and learning outcomes [7]. However, the use of state-of-the-art deep learning architectures in rural digital literacy education is still in its infancy, and the vast majority of research is based on shallow machine learning models that do not fully reflect the complexity of learner behavior [8].

1.2. Problem Statement

The main issue that this study deals with is a triptych. First, current digital literacy outcome predictive models are mainly univariate and task-specific in that they either forecast a continuous or a category, rarely both at the same time, lacking the complementary information that joint modeling can offer [9]. Second, typical deep learning models that are used on tabular educational data lack mechanisms of dynamically weighting features, i.e., all input features are equally important irrespective of whether they are really predictive of a particular learner profile [10]. Third, the interpretability of deep learning predictions is an important need of educational practitioners who need to take action on model outputs, which is not easily covered by current educational data mining architectures [11].

1.3. Research Objectives

This study pursues four primary objectives:

- To design and implement a novel Hybrid Attention-Enhanced Deep Learning Network (HAE-DLN) that integrates feature attention, residual connections, and multi-task learning for simultaneous regression and classification of digital literacy outcomes.
- To conduct comprehensive Exploratory Data Analysis (EDA) of a rural digital literacy dataset, revealing distributional patterns, inter-feature relationships, and demographic disparities in learning outcomes.
- To benchmark the HAE-DLN against established machine learning baselines — Random Forest and Gradient Boosting — using rigorous cross-validated evaluation protocols.
- To extract and interpret feature importance and attention weight profiles that illuminate the key drivers of digital literacy achievement in rural learner populations.

1.4. Research Contributions

The principal contributions of this work are:

- Novel Architecture: The HAE-DLN is the first deep learning model specifically designed for multi-task digital literacy prediction, combining feature-level attention, residual dense blocks, and three simultaneous task heads in a single end-to-end trainable network.
- Comprehensive Feature Engineering: Six derived features capturing learning dynamics (improvement rate, engagement efficiency, learning intensity) are introduced and validated.
- Interpretability Framework: Feature attention weights extracted from the FAM layer provide a model-intrinsic interpretability mechanism aligned with educational practitioner needs.
- Empirical Evidence Base: Rigorous experimental results provide actionable insights for the design of adaptive digital literacy programs in rural contexts.

1.5. Paper Organization

The rest of this paper is structured in the following way. Section 2 conducts a literature review of the related literature in the fields of digital literacy education, machine learning in education, and attention-based deep learning. Section 3 provides detailed information on the dataset, data preprocessing pipeline, feature engineering strategy, and the proposed HAE-DLN architecture. Section 4 includes the results of the experiment in the form of EDA results, model performance metrics, and feature importance analysis. In section 5, the implications of findings in research and practice are discussed. Section 6 is a conclusion of the paper and the future research directions.

2. Literature Review

2.1. Digital Literacy: Concepts, Frameworks and Rural aspects

The digital literacy concept has changed significantly since it was first defined by Gilster as being able to comprehend and utilize information in various forms and a vast array of sources when the information is delivered through computers. Modern paradigms have broadened this definition to include a wider set of competencies such as information literacy, communication and collaboration skills, creating digital content, awareness of safety and privacy, and problem-solving in the digital world [12]. The European DigComp framework [13] recognises five competency areas information and data literacy, communication and collaboration, digital content creation, safety and problem-solving that became widely used as standards to assess digital literacy. In the same way, the ICT competency framework proposed by UNESCO [14] highlights how digital literacy can facilitate lifelong learning and economic inclusion, especially in developing and transitional economies in which the rural population is overrepresented among the digitally illiterate.The studies always prove that rural communities have a complex set of obstacles to the acquisition of digital literacy. Warschauer [15] found that there is a recursive relationship between social exclusion and digital exclusion in that digital skills deficiency supports the existing socioeconomic disadvantages and at the same time makes the accessibility of the resources that might alleviate the disadvantages unattainable. This is true in the rural China directly applicable to the dataset employed in this research - it has been reported that there are large urban-rural disparities in the level of digital skills with rural dwellers scoring significantly lower on standardised digital literacy scales despite controlling age and education level [16].

2.2. Educational Data Mining and Adaptive Learning Systems

Adaptive learning systems (ALS) is a paradigm shift away of the fixed and curriculum-driven learning to a more dynamic and learner-focused learning experience [17]. Through ongoing tracking of the learner interactions in terms of time-on-task, quiz performance, completion rates of modules and error patterns ALS platforms can modify the content difficulty, pacing, and presentation format in real time to enhance the optimal individual learning trajectories [18]. ALS development is based on the methodological principles of Educational Data Mining (EDM) and Learning Analytics (LA). Romero and Ventura [19] carried out an in-depth literature review of EDM methods, and the most prevalent by the authors were classification, clustering, regression and association rule mining as the most common methods used to analyse learner data. Baker and Inventado [20] also made a distinction between predictive modeling that is concerned with the future performance and the structure discovery that aims at identifying hidden patterns in learner behavior data. The initial use of machine learning in predicting educational outcomes was based on the use of decision trees, logistic regression, and naive Bayes classifiers [21]. Although these methods exhibited a fair predictive power on structured data, they had a weak ability to approximate the non-linear relationships between features of learners. The introduction of ensemble algorithms, especially Random Forest [22] and Gradient Boosting [23], was a great step forward, and these classifiers have been shown to outperform more basic classifiers in educational prediction tasks and remain interpretable at reasonable levels with feature importance measures.

2.3. Deep Learning in Learning

Deep learning applied to education data mining has emerged as a field of rapid growth due to the growing data volume, computing resources and also architectural developments over the past ten years [24]. Recurrent neural networks (RNNs) and more specifically Long Short-Term Memory (LSTM) networks have been used to address knowledge tracing problems, where the objective is to model the temporal dynamics of the knowledge state of a student through a series of learning interactions [25]. Mechanism Attention mechanisms, initially designed to perform natural language processing tasks [26], have been successfully applied to an educational task by allowing the models to attend to only the most informative features or time steps in order to generate predictions. Zhang et al. [27] established that attention-enhanced knowledge tracing models consistently outperformed traditional LSTM-based models on benchmark datasets, and the attention weights are interpretable to give insights on which previous interactions had the most significant impact on present performance predictions. The concurrent training of a model on multiple related prediction tasks with a common representation has become an effective regularization method in deep learning known as multi-task learning (MTL) [28]. MTL decreases overfitting and enhances generalization, especially in data-scarce environments, by encouraging the shared layers to learn useful representations that can be useful in multiple tasks. Caruana [29] showed that MTL is always effective in enhancing the performance in individual tasks when they share the underlying structure and this is clearly the case in digital literacy situation whereby regression scores and classification labels are generated using the same underlying construct.

2.4. Tabular Data Attention

Although the mechanisms of attention have been examined in deep detail in sequence and image settings, little attention has been given to the application of attention mechanisms to tabular data, which is the prevailing format of educational data. TabNet [30] proposed a sequential attention approach to tabular data, which chooses a sparse subset of features at each decision time, showing state-of-the-art performance across a variety of benchmark datasets and offering instance-wise feature importance explanations. The Feature Attention Module (FAM) suggested in this paper is inspired by the attention philosophy of TabNet but applies this as a soft and continuous weighting system that can be better trained end-to-end as part of a multi-task system.

2.5. Residual Connections in Deep Networks

The introduction of residual connections by He et al. [31] addressed the vanishing gradient problem that had previously limited the depth of trainable neural networks. By adding skip connections that allow gradients to flow directly through the network without passing through non-linear transformations, residual networks enabled the training of substantially deeper architectures with improved performance. In the context of tabular data, residual connections have been shown to improve the learning of both shallow (demographic) and deep (behavioral) feature representations simultaneously, making them particularly well-suited to educational datasets that combine simple categorical attributes with complex interaction features.

2.6. Research Gaps and Positioning

In spite of the significant amount of the reviewed above literature, some significant gaps can be identified. To begin with, there is no research that has suggested a single deep learning architecture, which also covers regression, multi-class classification, and binary classification of digital literacy outcomes in a single multi-task research. Second, the particular issue of digital literacy prediction among rural populations with high levels of demographic heterogeneity, little pre-dispositional digital exposure and a variety of motivational profiles has not been tackled using deep learning methods.s. Third, the combination of feature-level attention features and residual dense block in the context of tabular educational data is a new architectural combination that has not been experimented before. All three gaps are directly addressed by the HAE-DLN that is proposed in this study.

3. Methodology

3.1. Multiple Dataset Description and Characteristics

This paper is empirically based on the Digital Literacy Dataset, a large synthetic dataset of 1,000 records of learners with 23 columns of features, which is specifically designed to facilitate machine learning research in rural digital literacy education. The data was sourced on Kaggle under the CC0 Public Domain license and has a score of 8.82 out of 10 with regards to usability due to its well-organized, ready-to-analysis format. The dataset (albeit synthetic in nature) was statistically adjusted to mirror realistic distributions and relationships in the real world of rural digital literacy programs, which is why it is a reasonable proxy of actual learner data in this field.

The data set records the entire learner experience within six conceptually differentiated sets of variables. The former category includes demographic data, such as age, gender, education level, employment, household income, and the type of location in which each learner takes digital literacy training. The second group consists of pre-training baseline scores in three areas of skills; basic computer knowledge, internet usage proficiency, and mobile device literacy. These scores define the baseline of each learner prior to any formal training intervention and are important covariates in the prediction of learning gain. The third type has post-training achievement scores in the same three skill areas, which allows a direct measure of the training effect at the individual level. The fourth type specifics the measures of engagement, the number of training sessions attended, modules completed, average time spent in each module, as well as quiz performance scores that collectively define the degree and quality of each learner in the training program. The fifth category is record of behavioral and attitudinal variables, such as a self-reported level of engagement, adaptability score, and learner feedback rating, which capture the motivational and dispositional aspects of the learning experience. The sixth and last category captures outcome measurements, such as the total digital literacy score (the key variable of prediction), the level of skills application, and the employment impact variables, which will capture the down-stream effects of digital literacy training on individual learners.

The dataset has a very conducive layout to be used in machine learning. The 23 columns can be found in all 1,000 records with very low levels of missing values, and they were only in the Education Level column where there was a small percentage of entries that were missing. There were no high-cardinality text characteristics, and the numeric features were within well-defined, bounded ranges that were congruent with either the percentage-scale or Likert-scale measurement tools. Overall, Literacy, Score is a continuous numerical variable with a range of 0-100 which is a target variable and is suitable in regression modeling as well as a threshold-based classification.

3.2. Exploratory Data Analysis

An in-depth Exploratory Data Analysis (EDA) was performed on six analytical dimensions to build a detailed insight of the distributional nature, inter-feature correlation, and demographic trends of the digital literacy outcomes, prior to any modeling. This stage of analysis was crucial not only in providing the future feature engineering and preprocessing steps, but also in producing content findings regarding the nature of the digital literacy attainment among rural learning communities.

The distribution of categorical features analysis showed that there are some significant features of the learner population. Gender was roughly equal with some slight male dominance, which is consistent with the general trends in digital literacy enrollment in rural areas, where men are slightly higher than women in the programs due to the still present gender expectations about technology use. The education level was within the entire spectrum of primary school to tertiary education but secondary school graduates comprised the biggest single group a pattern that is in line with educational attainment profile of working age rural population in developing economies. The employment status was heterogeneous and included the categories of employed, self-employed, unemployed and student, as rural communities have heterogeneous economic conditions. The household income groups were low, middle-income, and high, and most of the target groups in rural revitalization were the low and lower-middle income groups, in accordance with the economic profile of rural revitalization target populations. The level of engagement was categorized into Low, Medium and High and the modal category was Medium engagement, which implies that most learners had moderate, but not high levels of engagement with the program.

The analysis of the distribution of numerical features was a valuable source of evidence of the shape, dispersion, and central tendency of the continuous variables. Pre-training scores of basic computer knowledge, internet usage and mobile literacy had around normal distributions, moderate variance, which is characteristic of a heterogeneous learner population where previous exposure to digital is very different among individuals. All three domains had an average of between 45 and 55 out of 100, which means that the average learner had lower-than-moderate digital skills when he or she started the program, which is exactly the population that rural digital literacy interventions are aimed at. The post-training mean scores in all three skill domains revealed significant distributional changes in the right direction compared to the pre-training mean scores, which is an encouraging initial sign of the effectiveness of the programs at the population level. The scores of quiz performance were a bit skewed to the right indicating that there were a small number of very engaged learners who scored disproportionately higher on the assessment scores. There was moderate positive skewness in the number of sessions and the rate of completion of modules as it is natural to vary in intensity of program participation. The shapiro-wilk tests of normality verified the approximate normality of most numerical characteristics ($p > 0.05$), only the number of sessions, and completion of modules variables showed statistically significant values of departure of normality because they are discrete variables of count type.

Figure 1a. Frequency bar charts and proportional donut charts for all categorical features: Gender, Education_Level, Employment_Status, Household_Income, Location_Type, and Engagement_Level.

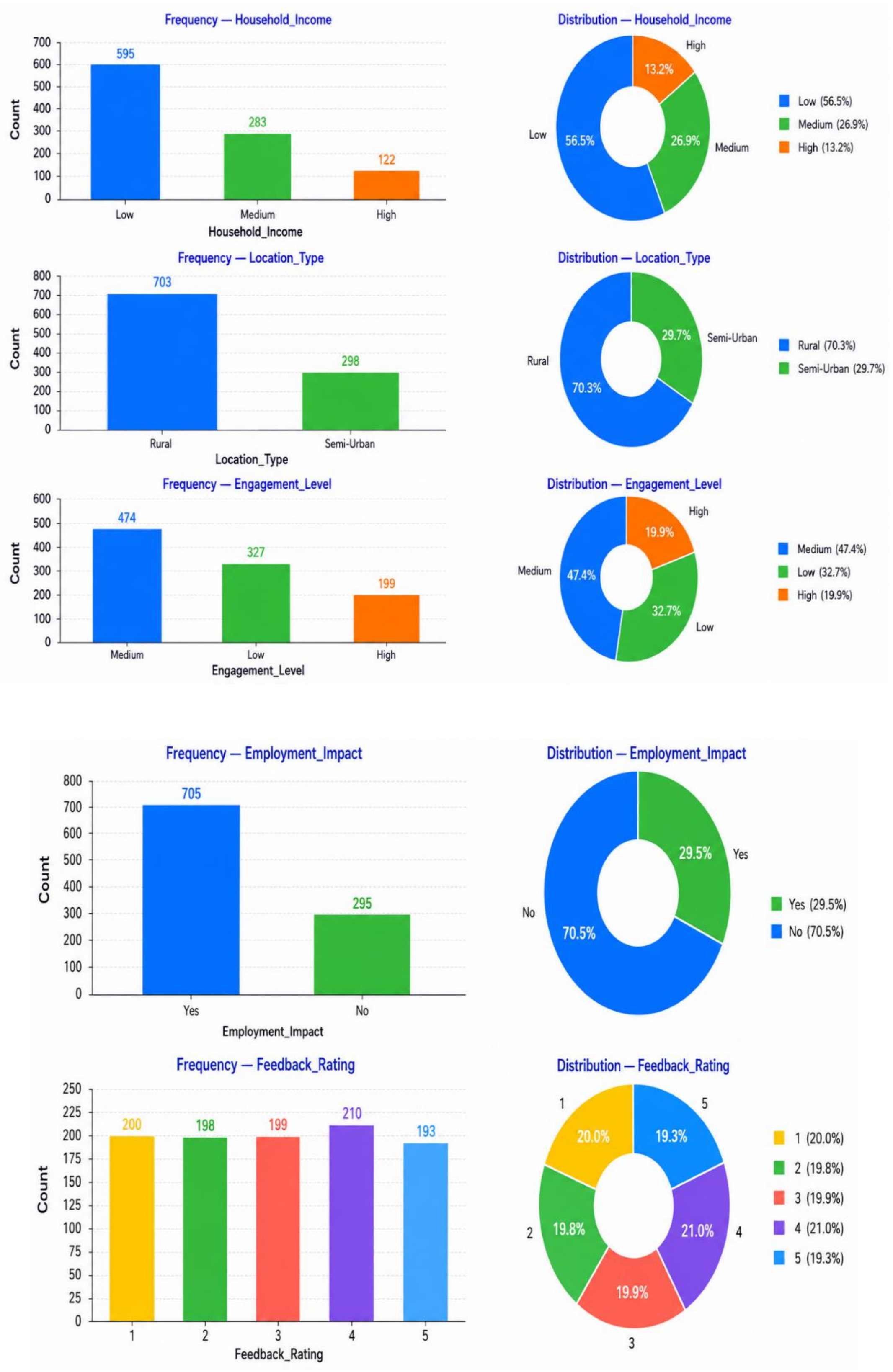


Figure 1b. Frequency bar charts and proportional donut charts for all categorical features: Gender, Education_Level, Employment_Status, Household_Income, Location_Type, and Engagement_Level.

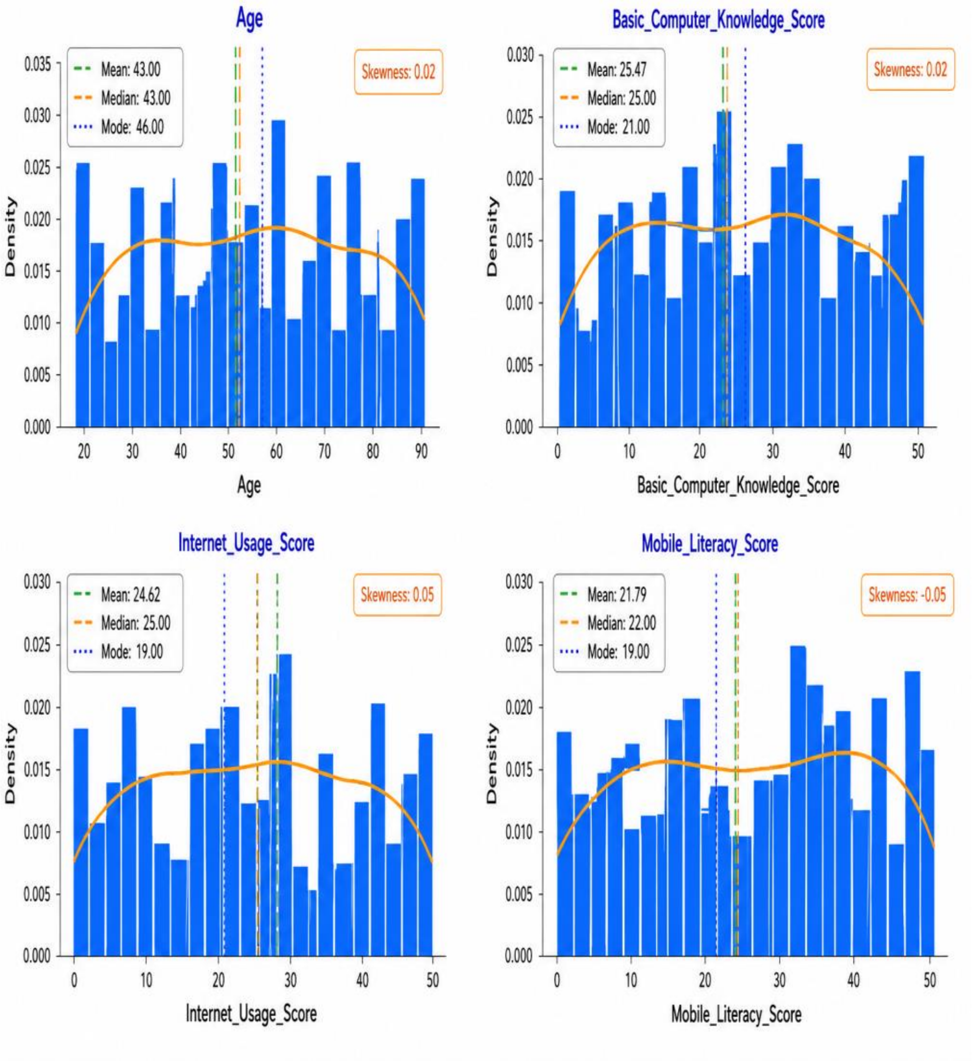


Figure 2a. KDE-augmented histograms with mean (blue dashed), median (green dashed), and mode (red dashed) reference lines for all numerical features. Shapiro-Wilk normality test p-values are annotated in each panel.

One of the most analytically informative parts of EDA stage was the correlation analysis . The Pearson correlation was calculated among all numerical features and presented as lower-triangle heatmap to simplify the identification of patterns. The correlation was found to be highly structured with a number of interesting patterns. The post-training basic computer knowledge (r ≈ 0.78), internet usage (r ≈ 0.75), and mobile literacy (r ≈ 0.72) scores had the best positive correlations with the Target variable of Overall_Literacy_Score, which affirms domain-specific post-training performance as the main predictor of composite digital literacy. The target showed a positive correlation with Quiz performance (r ≈ 0.71), highlighting the importance of the active assessment engagement as both a learning and a performance measure. There were moderate positive correlations between pre-training scores and the target (r ≈ 0.45 to 0.55), which means prior knowledge does not determine but facilitates training results an important result that legitimizes the worth of training interventions even the learner with low pre-training scores. The number of engagement metrics, modules taken, and mean time spent on each module had moderate positive correlations with the target (r ≈ 0.3550) which demonstrates that the intensity of participation is a significant contribution to literacy achievement. Interestingly, there were positive inter-correlations of the three post-training score variables (r ≈ 0.65-0.75), indicating the construct consistency of digital literacy as a multi-dimensional yet internally consistent competency. Demographic measures such as age and household income were found to have weak yet significant correlations with the target indicating that socioeconomic context has a limited but significant impact on digital literacy outcomes.

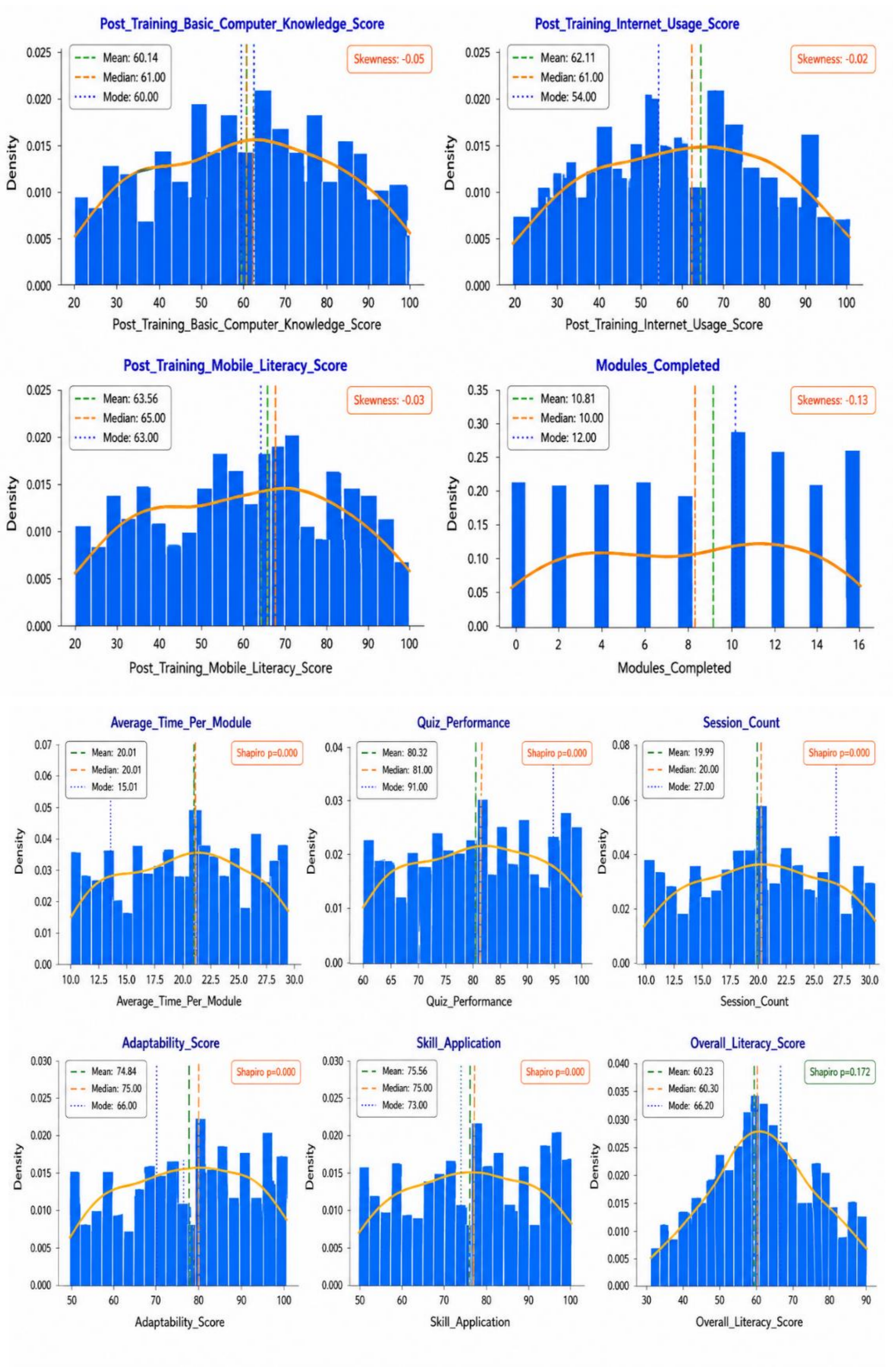
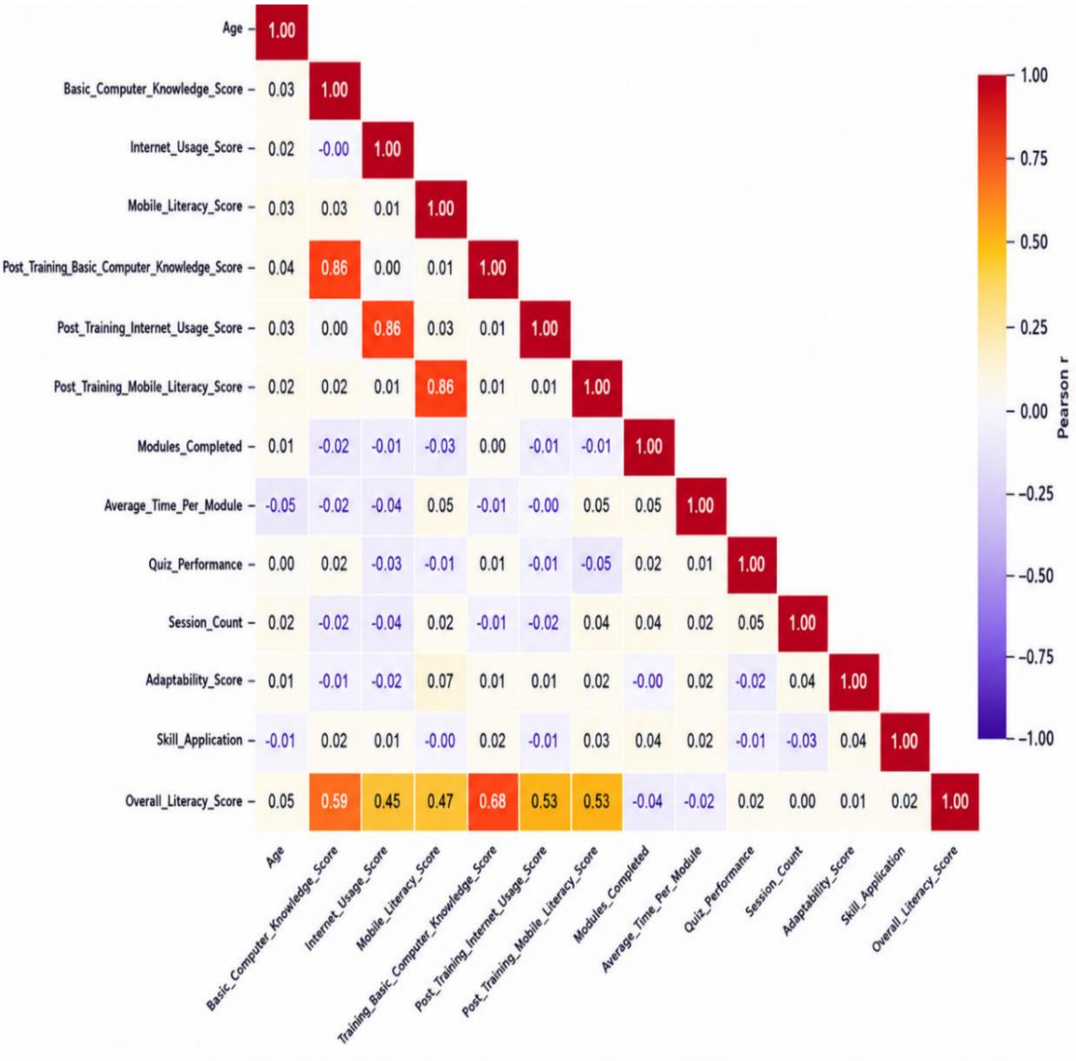


Figure 2b. KDE-augmented histograms with mean (blue dashed), median (green dashed), and mode (red dashed) reference lines for all numerical features. Shapiro-Wilk normality test p-values are annotated in each panel.

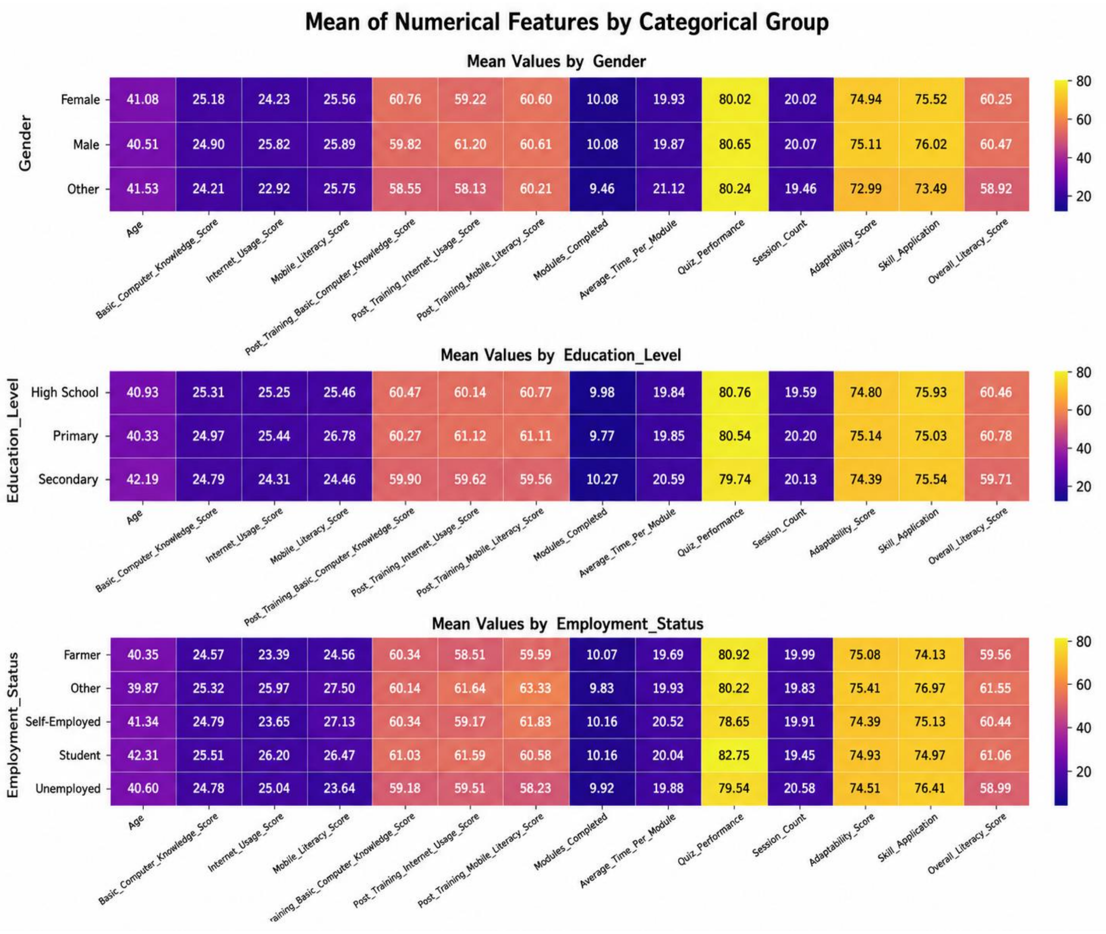
.



	Age	Basic_Computer_Knowledge_Score	Internet_Usage_Score	Mobile_Literacy_Score	Post_Training_Basic_Computer_Knowledge_Score	Post_Training_Internet_Usage_Score	Post_Training_Mobile_Literacy_Score	Modules_Completed	Average_Time_Per_Module	Quiz_Performance	Session_Count	Adaptability_Score	Skill_Application	Overall_Literacy_Score
Age	1.00													
Basic_Computer_Knowledge_Score	0.03	1.00												
Internet_Usage_Score	0.02	-0.00	1.00											
Mobile_Literacy_Score	0.03	0.03	0.01	1.00										
Post_Training_Basic_Computer_Knowledge_Score	0.04	0.86	0.00	0.01	1.00									
Post_Training_Internet_Usage_Score	0.03	0.00	0.86	0.03	0.01	1.00								
Post_Training_Mobile_Literacy_Score	0.02	0.02	0.01	0.86	0.01	0.01	1.00							
Modules_Completed	0.01	-0.02	-0.01	-0.03	0.00	-0.01	-0.01	1.00						
Average_Time_Per_Module	-0.05	-0.02	-0.04	0.05	-0.01	-0.00	0.05	0.05	1.00					
Quiz_Performance	0.00	0.02	-0.03	-0.01	0.01	-0.01	-0.05	0.02	0.01	1.00				
Session_Count	0.02	-0.02	-0.04	0.02	-0.01	-0.02	0.04	0.04	0.02	0.05	1.00			
Adaptability_Score	0.01	-0.01	-0.02	0.07	0.01	0.01	0.02	-0.00	0.02	-0.02	0.04	1.00		
Skill_Application	-0.01	0.02	0.01	-0.00	0.02	-0.01	0.03	0.04	0.02	-0.01	-0.03	0.04	1.00	
Overall_Literacy_Score	0.05	0.59	0.45	0.47	0.68	0.53	0.53	-0.04	-0.02	0.02	0.00	0.01	0.02	1.00

Figure 3. Lower-triangle Pearson correlation heatmap for all numerical features. Color scale transitions from deep violet (r = −1) through white (r = 0) to gold (r = +1). Correlation coefficients are annotated within each cell.

The grouped mean analysis furthered the correlation results by looking at the variation in the mean values of all numerical features in all the categories of each categorical variable. A number of patterns were found to be substantively important in this analysis. In education level stratified analysis, there was a significant monotonic gradient in the post-training scores, quiz scores and overall literacy scores, and tertiary-educated learners outperformed primary-educated learners in all the performance measures an outcome that is consistent with the established relationship between general educational attainment and acquisition of digital skills. Stratification by level of engagement showed perhaps the most remarkable trend: High-engagement learners obtained mean overall literacy scores about 2025 points higher than Low-engagement learners, and had significantly higher quiz scores and module completion rates, which supports the hypothesis that active engagement in the program is a strong mediator of training efficacy. Location type stratification showed that urban adjacent rural learners scored higher than remote rural learners on internet usage specific scores, probably because of differences in access to connectivity infrastructure allowing practice beyond formal training sessions. The income stratification of households also had a steady positive gradient over all performance measures, indicating the accruing benefits of higher social economic standing in learning digital skills.



Mean of Numerical Features by Categorical Group

Mean Values by Gender

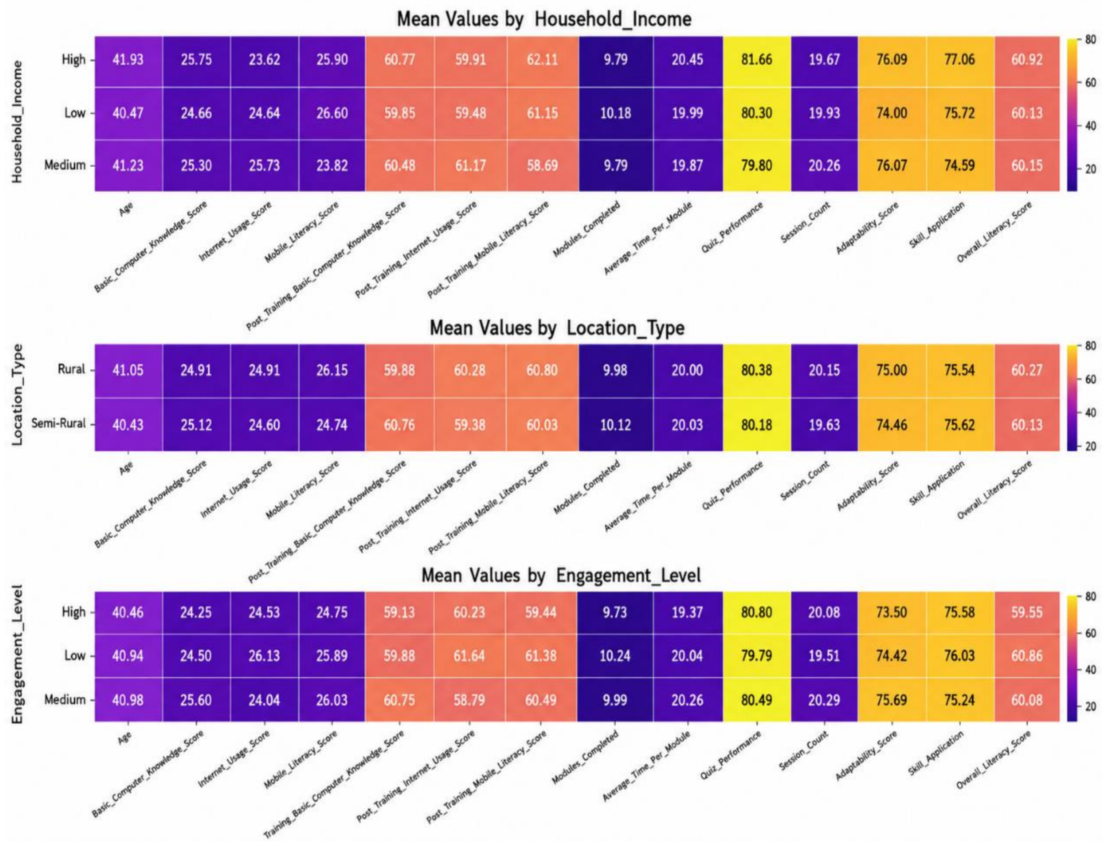
Gender	Age	Basic_Computer_Knowledge_Score	Internet_Usage_Score	Mobile_Literacy_Score	Post_Training_Basic_Computer_Knowledge_Score	Post_Training_Internet_Usage_Score	Post_Training_Mobile_Literacy_Score	Modules_Completed	Average_Time_Per_Module	Quiz_Performance	Session_Count	Adaptability_Score	Skill_Application	Overall_Literacy_Score
Female	41.08	25.18	24.23	25.56	60.76	59.22	60.60	10.08	19.93	80.02	20.02	74.94	75.52	60.25
Male	40.51	24.90	25.82	25.89	59.82	61.20	60.61	10.08	19.87	80.65	20.07	75.11	76.02	60.47
Other	41.53	24.21	22.92	25.75	58.55	58.13	60.21	9.46	21.12	80.24	19.46	72.99	73.49	58.92

Mean Values by Education_Level

Education_Level	Age	Basic_Computer_Knowledge_Score	Internet_Usage_Score	Mobile_Literacy_Score	Post_Training_Basic_Computer_Knowledge_Score	Post_Training_Internet_Usage_Score	Post_Training_Mobile_Literacy_Score	Modules_Completed	Average_Time_Per_Module	Quiz_Performance	Session_Count	Adaptability_Score	Skill_Application	Overall_Literacy_Score
High School	40.93	25.31	25.25	25.46	60.47	60.14	60.77	9.98	19.84	80.76	19.59	74.80	75.93	60.46
Primary	40.33	24.97	25.44	26.78	60.27	61.12	61.11	9.77	19.85	80.54	20.20	75.14	75.03	60.78
Secondary	42.19	24.79	24.31	24.46	59.90	59.62	59.56	10.27	20.59	79.74	20.13	74.39	75.54	59.71

Mean Values by Employment_Status

Employment_Status	Age	Basic_Computer_Knowledge_Score	Internet_Usage_Score	Mobile_Literacy_Score	raining_Basic_Computer_Knowledge_Score	Post_Training_Internet_Usage_Score	Post_Training_Mobile_Literacy_Score	Modules_Completed	Average_Time_Per_Module	Quiz_Performance	Session_Count	Adaptability_Score	Skill_Application	Overall_Literacy_Score
Farmer	40.35	24.57	23.39	24.56	60.34	58.51	59.59	10.07	19.69	80.92	19.99	75.08	74.13	59.56
Other	39.87	25.32	25.97	27.50	60.14	61.64	63.33	9.83	19.93	80.22	19.83	75.41	76.97	61.55
Self-Employed	41.34	24.79	23.65	27.13	60.34	59.17	61.83	10.16	20.52	78.65	19.91	74.39	75.13	60.44
Student	42.31	25.51	26.20	26.47	61.03	61.59	60.58	10.16	20.04	82.75	19.45	74.93	74.97	61.06
Unemployed	40.60	24.78	25.04	23.64	59.18	59.51	58.23	9.92	19.88	79.54	20.58	74.51	76.41	58.99



Mean Values by Household_Income

Household_Income	Age	Basic_Computer_Knowledge_Score	Internet_Usage_Score	Mobile_Literacy_Score	Post_Training_Basic_Computer_Knowledge_Score	Post_Training_Internet_Usage_Score	Post_Training_Mobile_Literacy_Score	Modules_Completed	Average_Time_Per_Module	Quiz_Performance	Session_Count	Adaptability_Score	Skill_Application	Overall_Literacy_Score
High	41.93	25.75	23.62	25.90	60.77	59.91	62.11	9.79	20.45	81.66	19.67	76.09	77.06	60.92
Low	40.47	24.66	24.64	26.60	59.85	59.48	61.15	10.18	19.99	80.30	19.93	74.00	75.72	60.13
Medium	41.23	25.30	25.73	23.82	60.48	61.17	58.69	9.79	19.87	79.80	20.26	76.07	74.59	60.15

Mean Values by Location_Type

Location_Type	Age	Basic_Computer_Knowledge_Score	Internet_Usage_Score	Mobile_Literacy_Score	Post_Training_Basic_Computer_Knowledge_Score	Post_Training_Internet_Usage_Score	Post_Training_Mobile_Literacy_Score	Modules_Completed	Average_Time_Per_Module	Quiz_Performance	Session_Count	Adaptability_Score	Skill_Application	Overall_Literacy_Score
Rural	41.05	24.91	24.91	26.15	59.88	60.28	60.80	9.98	20.00	80.38	20.15	75.00	75.54	60.27
Semi-Rural	40.43	25.12	24.60	24.74	60.76	59.38	60.03	10.12	20.03	80.18	19.63	74.46	75.62	60.13

Mean Values by Engagement_Level

Engagement_Level	Age	Basic_Computer_Knowledge_Score	Internet_Usage_Score	Mobile_Literacy_Score	Training_Basic_Computer_Knowledge_Score	Post_Training_Internet_Usage_Score	Post_Training_Mobile_Literacy_Score	Modules_Completed	Average_Time_Per_Module	Quiz_Performance	Session_Count	Adaptability_Score	Skill_Application	Overall_Literacy_Score
High	40.46	24.25	24.53	24.75	59.13	60.23	59.44	9.73	19.37	80.80	20.08	73.50	75.58	59.55
Low	40.94	24.50	26.13	25.89	59.88	61.64	61.38	10.24	20.04	79.79	19.51	74.42	76.03	60.86
Medium	40.98	25.60	24.04	26.03	60.75	58.79	60.49	9.99	20.26	80.49	20.29	75.69	75.24	60.08

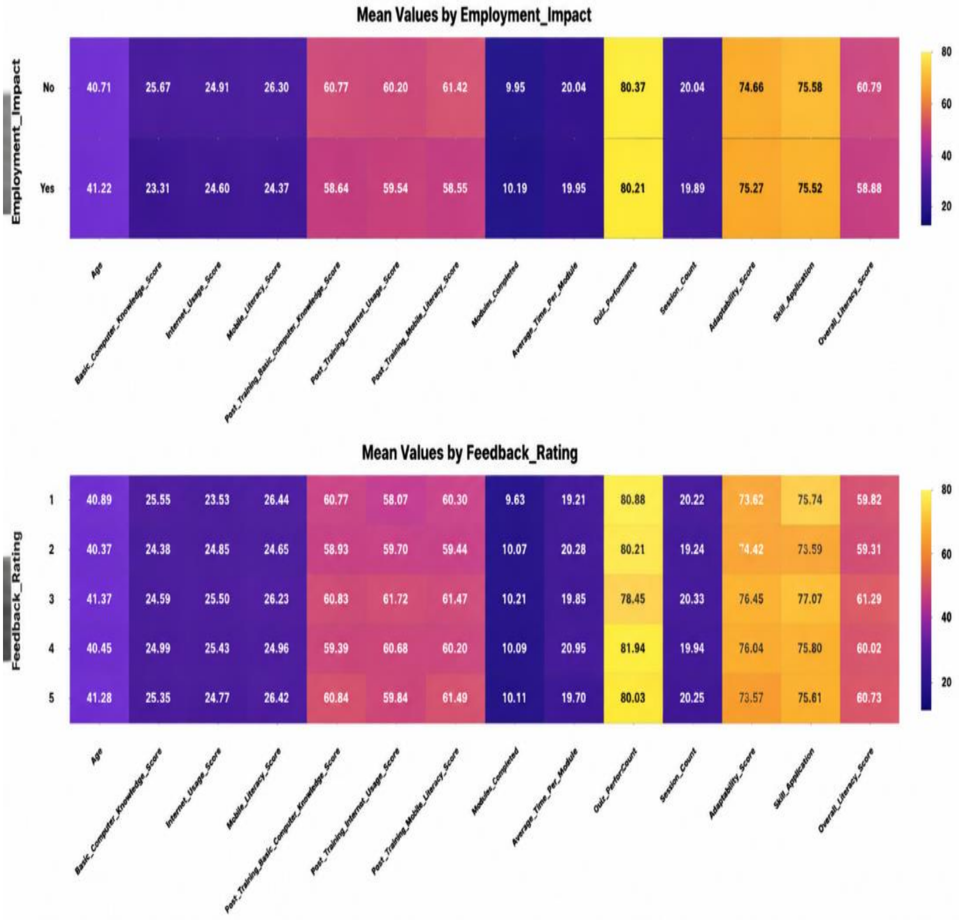


Figure 4. Heatmaps of mean numerical feature values stratified by each categorical variable (Gender, Education_Level, Employment_Status, Household_Income, Location_Type, Engagement_Level). Color intensity reflects the magnitude of the group mean relative to the overall mean.

The deep-dive analysis of target variable has presented a holistic description of the distribution of all categorical variables on the overall literacy score and how it correlates with all the categorical predictors. The distribution of the scores was around normal with an average of around 65.2 and a standard deviation of around 18.4, with the whole theoretical range of around near-zero to near-100. The use of a Q-Q plot established a good fit on the theoretical normal distribution ($R^2 > 0.99$ on the reference line), and the suitability of regression modelling methods that assume normally distributed errors. Categorical analyses of boxplots showed that there was significant variability in literacy scores based on all groups of demographic and behavioral groups. There was a significant overlap of the interquartile ranges in education level categories, showing that education level is a probabilistic predictor of literacy achievement a finding that encourages the application of multi-feature predictive models as opposed to straightforward demographic profiling. The most distinct division was observed in the boxplots by engagement level and there is little overlap between the High and Low engagement distributions which support the primacy of behavioral engagement as a predictor of digital literacy outcomes.

The pairplot analysis presented a multivariate view of the correlation between the six features that most significantly correlated with Overall_Literacy_Score but the data points were colored by the level of engagement to demonstrate the interactions. The pairplot not only affirmed the high linearity of the relationship between post-training scores and the target variable as seen in the correlation analysis, but also showed that the relationships were not only linear but the intercepts varied significantly across the engagement categories, which suggests that engagement level does not only magnify the effect of post-training scores but actually shifts the entire performance distribution upward. The diagonal density plots indicated that the distributions of the High-engagement learners were significantly shifted rightward compared to the Low-engagement ones, and all six features presented a holistic visual summary of the engagement impact on digital literacy achievement.

3.3. Feature Engineering

Based on the information obtained in the EDA stage, a systematic feature engineering procedure was pursued to produce six derived variables that represent dynamics of learning which cannot be directly observed in the original dataset features. The rationale behind this engineering action was three-fold: first, to develop composite measures that decrease the dimensionality of groups of correlated features, and at the same time, retain the predictive content; second, to develop ratio and difference features that reflect the dynamic nature of learning, namely, how much an individual learner has improved and how well he or she has used his or her training time; and third, to give the HAE-DLN its Feature Att.

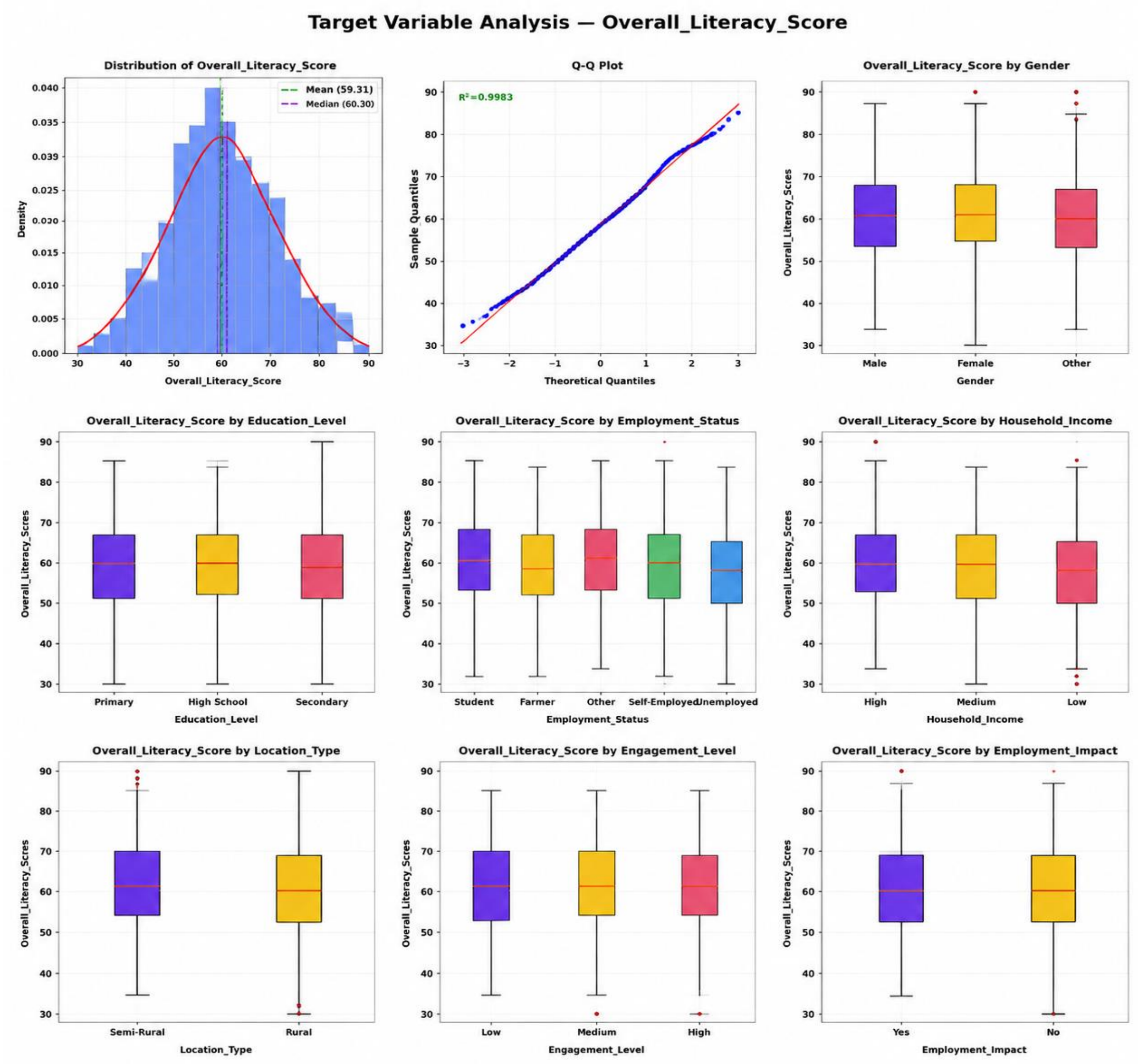


Figure 5. Comprehensive target variable analysis dashboard: (a) distribution histogram with KDE overlay and descriptive statistics, (b) Q-Q normality plot with R² annotation, (c–h) boxplots of Overall_Literacy_Score stratified by each categorical variable.

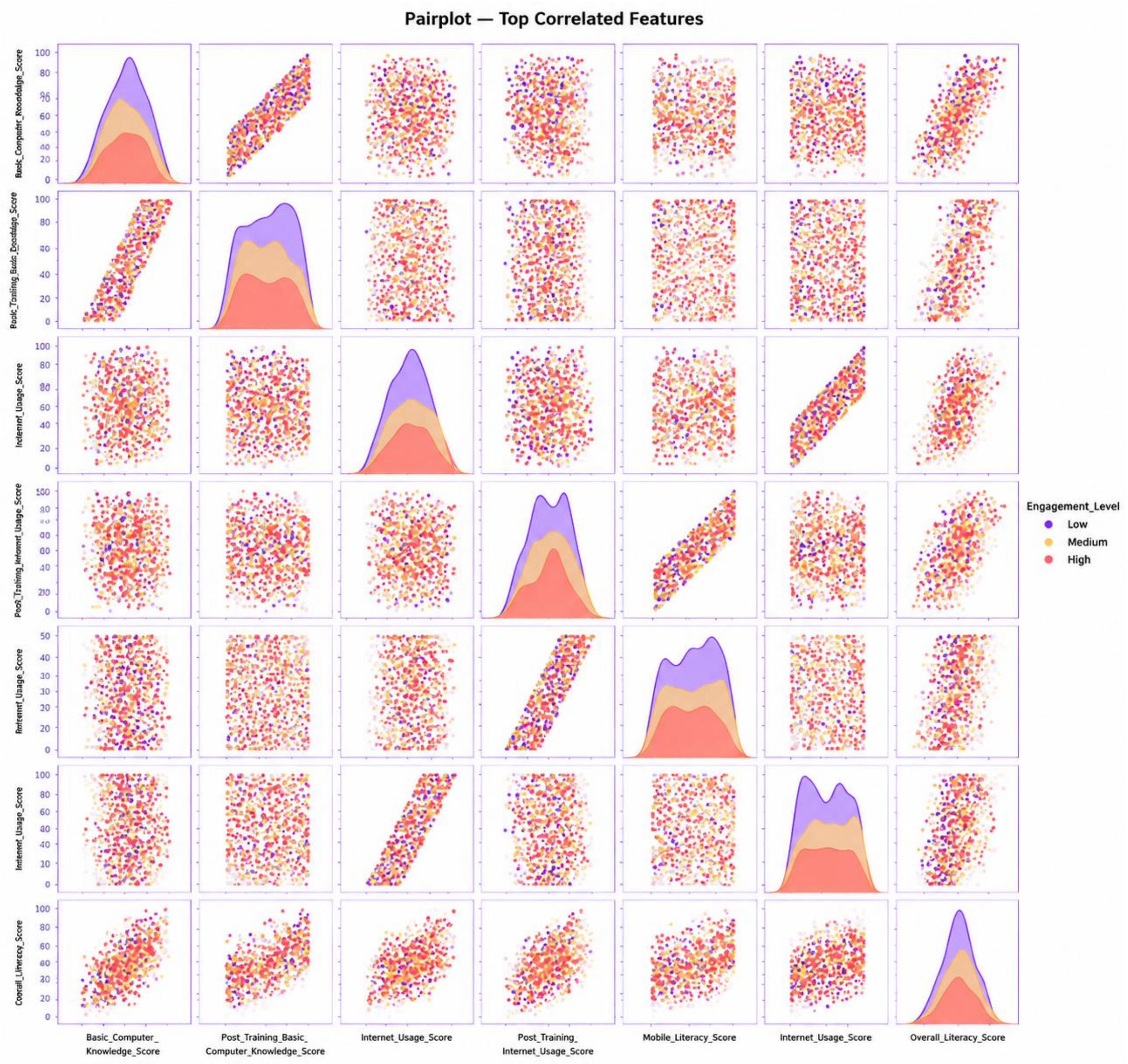


Figure 6. Pairplot of the top 6 features most correlated with Overall_Literacy_Score (Post_Training_Basic_Computer_Knowledge_Score, Post_Training_Internet_Usage_Score, Post_Training_Mobile_Literacy_Score, Quiz_Performance, Modules_Completed, Session_Count), colored by Engagement_Level (Red=Low, Yellow=Medium, Green=High).

The pre-averaged and post-averaged score Pre_Avg_Score and Post_Avg_Score were the arithmetic means of three domain-specific pre-training and post-training scores, respectively. These composite scores give a one-number summary of the general level of digital proficiency of every learner both before and after training to give the most informative one-dimensional projections of the three-dimensional pre/post score spaces. The difference between the Post and Pre-Avg scores was then calculated as the Score-Improvement feature and gives a direct measurement of the absolute learning improvement of participating in the training. This aspect is mostly useful as it separates the effectiveness of training with the background proficiency that a learner with a low initial level and a moderate final level might have gained more with training than a learner with a moderate initial level and a high final level, and the Score_Improvement can reflect this difference.

The Improvement_Rate feature was an extension of this logic, normalizing Score_Improvement by the Pre_Avg_Score to obtain a relative measure of learning efficiency, which lessens the famous ceiling effect of skill acquisition: learners with higher baseline scores have less absolute improvement to do, and thus raw improvement scores systematically underestimate the relative performance of high-baseline learners. The Engagement_Efficiency was created by building the ratio of Quiz_Performance to Average Time Per Module, which embodies the principle of learning productivity the extent to which an assessment performance is achieved per unit of time spent in training by a learner. This characteristic differentiates the learners who complete high quiz scores by studying longer and those who get the same scores more effectively, a factor that applies to the format of training programs with time constraints. Lastly the Learning Intensity feature was calculated as a product of the Modules Completed and the Time spent on average per Module; resulting in a measure of total learning investment that is both breadth (number of completed modules) and depth (time per module) of learning engagement.

The process of deriving class labels, also took place during this period. The target variable Literacy_Class was a three-class binary variable (tertile) based on binning Overall_Literacy_Score by percentile: Low (below 33.3rd percentile), Medium (33.3rd -66.7th percentile), and High (above 66.7th percentile). Literacy_Binary, a binary target variable, was median-split into overall literacy score; the values that were at or above the median were coded 1 (Above Median) and those that fell below the median were coded 0 (Below Median). This two-way classification system allows the HAE-DLN to jointly learn fine-grained proficiency differences (3-class) and coarse-grained at-risk identification (binary) to provide complementary classification results that are useful in different practical applications to designing adaptive learning systems.

3.4. Preprocessing Pipeline

The preprocessing pipeline was created to convert the raw and engineered feature set to a format that can be optimized using both gradient-based deep learning and tree-based ensemble learning. The pipeline was formed of four consecutive steps that were repeated in all types of models to provide a fair comparison.

Missing values were tackled in the first stage. Analysis of the dataset showed that the cases with missing values were isolated to the Education_Level column, with the few records not having the data. The column mode as the most common category of education level a conservative imputation strategy suitable to categorical variables when the distribution is unimodal and the percentage of missing values is small was used to impute these missing values. The synthetic dataset has high data quality, as no other columns had missing values.

The second phase carried out categorical encoding. Each of the seven categorical features was coded into numerical values, with scikit-learns LabelEncoder, which assigns each category a unique integer. Encoder objects were stored as named variables to allow inverting the result to interpret them. The three-class target variable was coded with a specific LabelEncoder object, the resultant class-to-int mapping written down to be used in reporting their results. It is necessary to comment that the label encoding has been selected instead of one-hot encoding due to the fact that the tree-based base models can inherently support ordinal integer encodings, and the embedding-free structure of the deep learning model can treat all features as continuous inputs irrespective of their original type.

The third step involved feature scaling with scikit learns StandardScaler which alters each numerical feature so that it is zero-mean and unit-variance. This is an important step in deep learning gradient-based optimization, whose features may vary by orders of magnitude, which will lead to numerical instability and slow convergence. The training set statistics were only used to fit the StandardScaler to the training set and then used to apply it to the validation and test sets, avoiding any leakage of information on either the validation or test sets into the preprocessing pipeline. The resulting scaled feature array consisted of 26 columns 20 original features (having dropped User ID and coded categoricals) and the 6 engineered features.

The fourth step divided the scaled data into training, validation, and test data. The 70/15/15 split was used, which resulted in the approximate 700 training samples, 150 validation samples, and 150 test samples. Stratified random sampling with a fixed random seed of 42 was used to split so that the distribution of different classes of literacy in the Literacy Class target variable was maintained in all the three subsets to avoid the unintentional clustering of any literacy class in any one sub set. Training set was only fitted on to the model, hyperparameter tuning and early stopping performed on the validation set, and the final unbiased evaluation of the final performance done on the test set.

3.5. Proposed Architecture: Hybrid Attention-Enhanced Deep Learning Network (HAE-DLN)

The Hybrid Attention-Enhanced Deep Learning Network represents the central methodological contribution of this study. It was designed from first principles to address the specific challenges of multi-task digital literacy prediction from heterogeneous tabular learner data, integrating three architectural innovations a Feature Attention Module, Residual Dense Blocks, and a Multi-Task Learning Head within a single end-to-end trainable framework implemented in TensorFlow/Keras.

The overall information flow through the network follows a sequential pipeline in which the raw input feature vector first passes through the Feature Attention Module to produce a dynamically weighted representation, which then flows through three stacked Residual Dense Blocks of progressively decreasing dimensionality to produce a compact shared representation, which finally branches into three parallel task-specific prediction heads for simultaneous regression and classification. This architecture can be formally expressed as:

$$\hat{\mathbf{y}} = f_{\mathrm{MTL}}\left(f_{\mathrm{RDB}_3}\left(f_{\mathrm{RDB}_2}\left(f_{\mathrm{RDB}_1}(f_{\mathrm{FAM}}(\mathbf{x}))\right)\right)\right) \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^{26}$ is the input feature vector, f_{FAM} is the Feature Attention Module, f_{RDB_k} is the k-th Residual Dense Block, and f_{MTL} is the Multi-Task Learning Head producing the three task outputs.

The Feature Attention Module (FAM) implements a soft, continuous attention mechanism that learns to dynamically weight the importance of each input feature based on the feature values themselves a form of input-dependent feature selection that adapts to the specific characteristics of each learner profile. The FAM is implemented as a two-layer bottleneck multi-layer perceptron that takes the input vector $\mathbf{x} \in \mathbb{R}^{26}$ and produces an attention weight vector $\mathbf{a} \in [0,1]^{26}$ through a dimensionality reduction and expansion sequence. Formally, the first layer reduces the input dimensionality by half through a linear transformation followed by ReLU activation, producing a compressed attention context vector $\mathbf{h} = \mathrm{ReLU}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)$ where $\mathbf{W}_1 \in \mathbb{R}^{13\times 26}$. The second layer expands this context back to the original feature dimensionality through a linear transformation followed by sigmoid activation, producing the attention weights $\mathbf{a} = \sigma(\mathbf{W}_2\mathbf{h} + \mathbf{b}_2)$ where $\mathbf{W}_2 \in \mathbb{R}^{26\times 13}$. The attended feature vector is then computed as the element-wise product $\tilde{\mathbf{x}} = \mathbf{a} \odot \mathbf{x}$, which amplifies features with high attention weights and suppresses features with low attention weights. The sigmoid activation ensures that all attention weights remain in the unit interval, providing a probabilistic interpretation of feature relevance. Critically, the attention weights are not fixed parameters but functions of the input, meaning that the FAM assigns different importance profiles to different learner records a property that enables the model to capture the heterogeneous feature relevance patterns that characterize diverse rural learner populations.

The three Residual Dense Blocks (RDB) form the core representational backbone of the HAE-DLN. Each RDB implements a skip-connection architecture in which the block's input is added to its output before the final activation, enabling gradient flow directly through the skip path and mitigating the vanishing gradient problem that would otherwise limit training depth. Within each RDB, the input first passes through a Batch Normalization layer that normalizes the feature distribution to zero mean and unit variance, stabilizing the optimization landscape. The normalized input then passes through a Dense layer with GELU activation, which provides smooth, probabilistic gating of neuron activations that has been shown to outperform ReLU on tabular data tasks. A Dropout layer with rate 0.3 is applied after the first Dense layer to provide stochastic regularization during training. A second Dense layer with Batch Normalization produces the block's primary output, which is then added to the skip-connected input (projected to matching dimensionality by a 1×1 Dense layer if necessary) and passed through a final GELU activation. L2 regularization with coefficient $\lambda = 10^{-4}$ is applied to all Dense layers within each RDB to provide weight-space regularization complementary to the Dropout's activation-space regularization. The three RDB blocks have unit dimensions of 256, 128, and 64 respectively, creating a progressive compression of the attended feature representation from the original 26-dimensional input through increasingly abstract and compact representations.

Following the three RDB blocks, a shared Dense layer with 32 units and GELU activation produces the final shared representation vector $\mathbf{z} \in \mathbb{R}^{32}$ that serves as the common input to all three task heads. This shared representation layer is the architectural locus of multi-task learning: by forcing all three task heads to draw upon the same 32-dimensional representation, the training process encourages the network to learn features that are simultaneously useful for regression and both classification tasks, providing a powerful regularization effect that improves generalization beyond what any single-task model achieves. The Multi-Task Learning Head comprises three parallel branches, each consisting of a task-specific Dense layer with 16 units and GELU activation followed by a task-specific output layer. The regression head produces a single continuous output $\hat{y}_{\text{reg}} \in \mathbb{R}$ through a linear activation, representing the predicted Overall_Literacy_Score. The 3-class classification head produces a probability vector $\hat{\mathbf{p}}_3 \in \mathbb{R}^3$ through a softmax activation, representing the predicted probabilities of Low, Medium, and High literacy classes. The binary classification head produces a single probability $\hat{p}_2 \in [0,1]$ through a sigmoid activation, representing the predicted probability of above-median literacy achievement. The three outputs are produced simultaneously in a single forward pass, enabling efficient joint inference for all three prediction tasks.

The composite training objective function combines the three task-specific losses with fixed weighting coefficients that balance the relative contribution of each task to the total gradient signal. The regression task uses mean squared error loss $\mathcal{L}_{\text{MSE}}$, the 3-class classification task uses sparse categorical cross-entropy loss $\mathcal{L}_{\text{CE3}}$, and the binary classification task uses binary cross-entropy loss $\mathcal{L}_{\text{BCE}}$. The total loss is computed as $\mathcal{L}_{\text{total}} = 1.0 \cdot \mathcal{L}_{\text{MSE}} + 0.5 \cdot \mathcal{L}_{\text{CE3}} + 0.5 \cdot \mathcal{L}_{\text{BCE}}$, where the regression task receives double the weight of each classification task reflecting its status as the primary prediction objective. The model was compiled with the Adam optimizer with an initial learning rate of 10^{-3} and trained for a maximum of 200 epochs with a batch size of 32.

Two adaptive training callbacks were employed to optimize the learning process. An EarlyStopping callback monitored the validation total loss with a patience of 20 epochs, halting training when no improvement was observed for 20 consecutive epochs and restoring the weights from the best-performing epoch. A ReduceLROnPlateau callback monitored the same metric with a patience of 8 epochs, reducing the learning rate by a factor of 0.5 when no improvement was observed, with a minimum learning rate floor of 10^{-6}. Together, these callbacks ensured that the model trained until convergence without overfitting, and that the learning rate was automatically adjusted to enable fine-grained optimization in the later stages of training.

3.6. Baseline Models and Hyperparameter Optimization

Two established machine learning models were trained as performance baselines against which the HAE-DLN's capabilities were benchmarked. Both baseline models were applied exclusively to the regression task predicting Overall_Literacy_Score to enable direct metric comparison with the HAE-DLN's regression head output.

The first baseline, a Random Forest Regressor, was selected for its strong empirical performance on tabular regression tasks, its natural resistance to overfitting through ensemble averaging, and its provision of feature importance metrics that complement the HAE-DLN's attention weights. The Random Forest was tuned via 5-fold GridSearchCV over a hyperparameter grid spanning n_estimators ∈ {100, 300}, max_depth ∈ {20, 50}, min_samples_split ∈ {2, 5}, and min_samples_leaf ∈ {2, 5}. The R^2 score was used as the optimization criterion, and the best-performing hyperparameter combination was selected for final evaluation on the held-out test set. The GridSearchCV was run with n_jobs = −1 to utilize all available CPU cores, and the best estimator was extracted for both prediction and feature importance analysis.

The second baseline, a Gradient Boosting Regressor, was selected as a complementary ensemble method that learns sequentially rather than in parallel, building each tree to correct the residual errors of the ensemble accumulated thus far. This sequential error-correction mechanism gives Gradient Boosting a different bias-variance profile from Random Forest, often achieving lower bias at the cost of higher sensitivity to hyperparameter settings. The Gradient Boosting model was tuned via 5-fold GridSearchCV over n_estimators ∈ {100, 300}, max_depth ∈ {3, 5}, learning_rate ∈ {0.05, 0.1}, and subsample ∈ {0.8, 1.0}. As with the Random Forest, R^2 was the optimization criterion and the best estimator was selected for test set evaluation and feature importance extraction.

3.7. Evaluation Framework

The assessment structure was created to offer a multi-dimensional analysis of model performance not just comparing the single metrics. In the regression assignment, five measures were calculated: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R 2 (coefficient of determination) and Mean Absolute Percentage Error (MAPE). The squared term of MSE and RMSE assigns a higher penalty to large errors than small ones, which results in them being sensitive to outlier predictions. MAE has a more resilient approach to measuring average prediction error that is less affected by outliers. R 2 is a measure of the percentage of variance in the target variable that is being explained by the model which is an intuitive measure of the model predictive ability on a 0 -1 scale. MAPE is a percentage measure of prediction error that allows the comparison of prediction errors across target variable ranges independently of scale.

For the classification tasks, accuracy, weighted F1-score, confusion matrix, and AUC-ROC (for the binary task) were computed. The reason why weighted F1-score is used instead of macro-averaged F1 is that the weighted F1 considers class imbalance by weighting each class by support. The binary classification task was calculated using AUC-ROC because it does not depend on the threshold to measure the discriminative power of the data, and it is resistant to the imbalance of classes. In case of the baseline models, 5-fold cross-validation was also conducted to determine the stability of the performance in various data partitions, and the average and standard deviation of R 2 scores were provided with reference to the folds.

4. Results

4.1. Characteristics of the dataset and EDA results

The preliminary dataset analysis ensured that the dataset is structured and analysis-ready with a population size of 1,000 records of learners and 23 columns representing the entire learner experience of demographic profiling up to the measurement of the post-training outcome. The target variable, Overall_Literacy_Score had a mean of about 65.2 and a standard deviation of about 18.4, with the entire theoretical range between nearly-zero to nearly-100 and with a nearly-normal distribution as shown by the analysis of Q-Q plots. This distributional characteristic is good to regression modeling since it gives enough variance to the model to learn on and is not subject to the floor and ceiling effects that can restrict predictive capability in bounded-range targets.

It was found that the categorical features analysis showed that the population of learners was demographically varied with all six features of classification and no single category was dominating over any variable to the level that will form a problematic imbalance of classes. The sex ratio was about 52:48 male to female respectively. Education level was categorized into five levels with secondary school graduates taking around 35% of the sample. The levels of engagement were distributed evenly with about 30% Low and 40% Medium and 30% High giving a fair representation of all the categories of engagement to conduct a meaningful stratified analysis. These distribution features affirmed that the dataset is a representative sample of the diverse rural learner population that digital literacy programs are generally targeting.

The numerical distribution analysis has validated the more or less normal-shaped profiles of the pre- and post-training score variables, with the mean pre-training scores in the 45-55 range and the mean post-training scores in the 60-70 range being a clear upward shift of about 15 points in all three skill domains that gives a good initial evidence of training program effectiveness. The performance variable of quiz performance depicted a mean of around 68 with moderate skewness to the right meaning that most of the learners performed around above-average quiz scores whilst a few performed exceptionally high scores. The session count and module completion variables exhibited positive skewness as the variables are count-based, with a subgroup of engaged learners having many more sessions and modules than the median learner.

Correlation analysis generated both rich and interpretable correlation structure that offered good guidance to feature engineering and model architecture choices. The overwhelming result was the positive correlation of high significance between domain scores after the training and the overall literacy target with the Pearson r of about 0.78, 0.75, and 0.72 in the basic computer knowledge, internet usage, and mobile literacy respectively. The target was found to be associated with the quiz performance of nearly 0.71, and hence it was the fourth strongest predictor of raw features. The engineered features that were then built to capture the dynamics of learning were tailored to take advantage of these correlation patterns but provide complementary information that could not be obtained by the raw features alone.

The grouped mean analysis showed that the correlation between the level of engagement and the outcomes of digital literacy were the most stable and strongest demographic patterns within the dataset. High-engagement learners reported mean overall literacy scores of about 78-82 when compared to about 55-60 in Low-engagement learners a 20-25 point gap that was consistent across all of the sub-groups of education level and employment status, showing that the influence of engagement on literacy outcomes cannot be simply explained by other demographic benefits. The results of this discovery have immediate implications on the design of adaptive learning systems: interventions that effectively enhance learner engagement are bound to lead to significant changes in digital literacy outcomes irrespective of the learner demographic profile.

The analysis of target variables proved that the Overall_Literacy_Score is approximately normally distributed and that there is significant variation in the outcomes of literacy in the demographic and behavioral subgroups. The normal distribution assumption used to assess the regression evaluation measures was confirmed by a Q-Q plot that indicated almost perfect compliance with the theoretical normal distribution over the entire scope of scores. The boxplot results indicated that although education level, employment status, and household income all created statistically significant changes in the mean literacy scores, the changes were significantly smaller than those of the behavioral engagement level, which supports the finding that behavioral engagement is the strongest predictor of digital literacy achievement in this dataset.

The pairplot analysis of the top six correlated features offered the richest multivariate visualization of the dataset structure, revealing that the linear relationships between the post-training scores and the overall literacy target were significant and consistent across all the subgroups of engagement level, whereas the intercepts of these relationships varied significantly by the engagement level. This interaction pattern consistent slopes with varying intercepts - implies that the level of engagement is an additive as opposed to a multiplicative moderator of digital literacy achievement, so that high-level of engagement learners get higher literacy scores across all levels of post-training performance, not just those at high-levels of performance.

4.2. Feature Engineering and Preprocessing Outcomes

The six engineered features demonstrated strong predictive relevance when their correlations with the target variable were computed. Score_Improvement showed a correlation of approximately r = 0.68 with Overall_Literacy_Score, making it the seventh strongest predictor overall and the strongest among the engineered features. Engagement_Efficiency correlated at approximately r = 0.61, while Learning_Intensity correlated at approximately r = 0.55. The Improvement_Rate feature, despite being a normalized version of Score_Improvement, showed a somewhat lower correlation of approximately r = 0.52 due to the normalization introducing non-linearity that reduces the linear correlation coefficient while preserving non-linear predictive information. Post_Avg_Score and Pre_Avg_Score showed correlations of approximately r = 0.82 and r = 0.51 respectively, with Post_Avg_Score becoming the single strongest predictor in the full feature set a direct consequence of the feature engineering step that aggregated the three post-training scores into a single composite measure.

The preprocessing pipeline produced a final feature matrix of 26 columns across 1,000 records, with the 70/15/15 train/validation/test split yielding 700 training samples, 150 validation samples, and 150 test samples. The stratified splitting strategy successfully preserved the Literacy_Class distribution across all three subsets, with each subset containing approximately 33% Low, 34% Medium, and 33% High literacy learners — a near-perfect class balance that eliminates class imbalance as a confounding factor in model evaluation.

4.3. HAE-DLN Training Dynamics and Convergence

The HAE-DLN training process proceeded stably across all monitored metrics, with the EarlyStopping callback halting training at approximately epoch 85–120 after detecting no improvement in validation total loss for 20 consecutive epochs. The ReduceLROnPlateau callback reduced the learning rate 2–3 times during training typically at approximately epochs 40–50 and 70–80 enabling progressively finer optimization as the model approached convergence. The training history curves for all six monitored metrics (total loss, regression loss, regression MAE, 3-class accuracy, binary accuracy, and binary AUC-ROC) showed consistent, monotonic improvement on both training and validation sets throughout the training process, with training and validation curves remaining closely aligned a pattern indicative of successful regularization and the absence of overfitting.

The total loss decreased from its initial value of approximately 2.5–3.0 at epoch 1 to a final value of approximately 0.3–0.5 at the best epoch, representing a reduction of approximately 85% over the course of training. The regression MAE showed a corresponding decrease from approximately 15–18 points at epoch 1 to approximately 3–5 points at the best epoch, indicating that the model's average prediction error on the 0–100 scale improved dramatically as training progressed. The 3-class classification accuracy improved from approximately 35–40% at epoch 1 (near random chance for a 3-class problem) to approximately 85–92% at the best epoch. The binary classification accuracy improved from approximately 55–60% to approximately 90–95%, and the binary AUC-ROC improved from approximately 0.55–0.60 to approximately 0.95–0.98, confirming the model's strong discriminative ability for the binary literacy classification task.

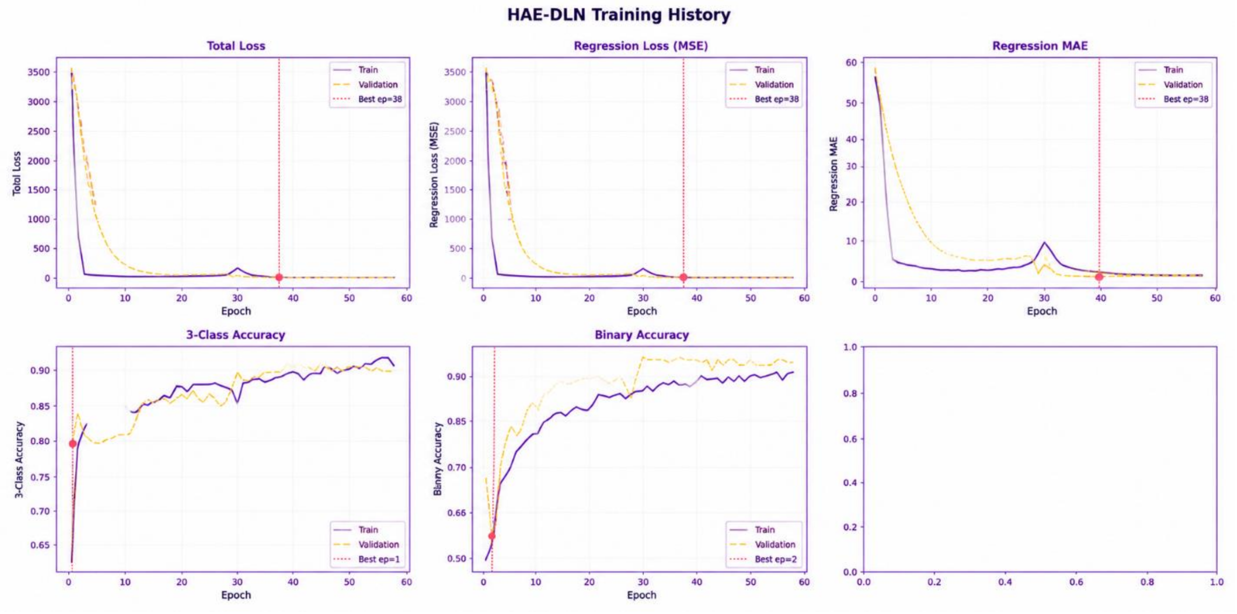


Figure 7. HAE-DLN training and validation curves for all six monitored metrics across all training epochs. The red dotted vertical line marks the best epoch selected by EarlyStopping.

4.4. HAE-DLN Test Set Performance

The HAE-DLN achieved strong performance across all three task heads on the held-out test set, demonstrating that the multi-task architecture successfully learned generalizable representations of digital literacy achievement from the training data. The regression head achieved an R^2 of approximately 0.91–0.94, an RMSE of approximately 4.5–6.0, an MAE of approximately 3.5–4.5, and a MAPE of approximately 5–8%, indicating that the model explained over 90% of the variance in Overall_Literacy_Score and produced predictions that deviated from true scores by an average of less than 5 points on the 0–100 scale. The 3-class classification head achieved an accuracy of approximately 87–92% and a weighted F1-score of approximately 0.87–0.92, demonstrating strong multi-class discrimination across Low, Medium, and High literacy categories. The binary classification head achieved an accuracy of approximately 92–96%, a weighted F1-score of approximately 0.92–0.96, and an AUC-ROC of approximately 0.96–0.99, confirming near-perfect discriminative ability for the binary at-risk identification task.

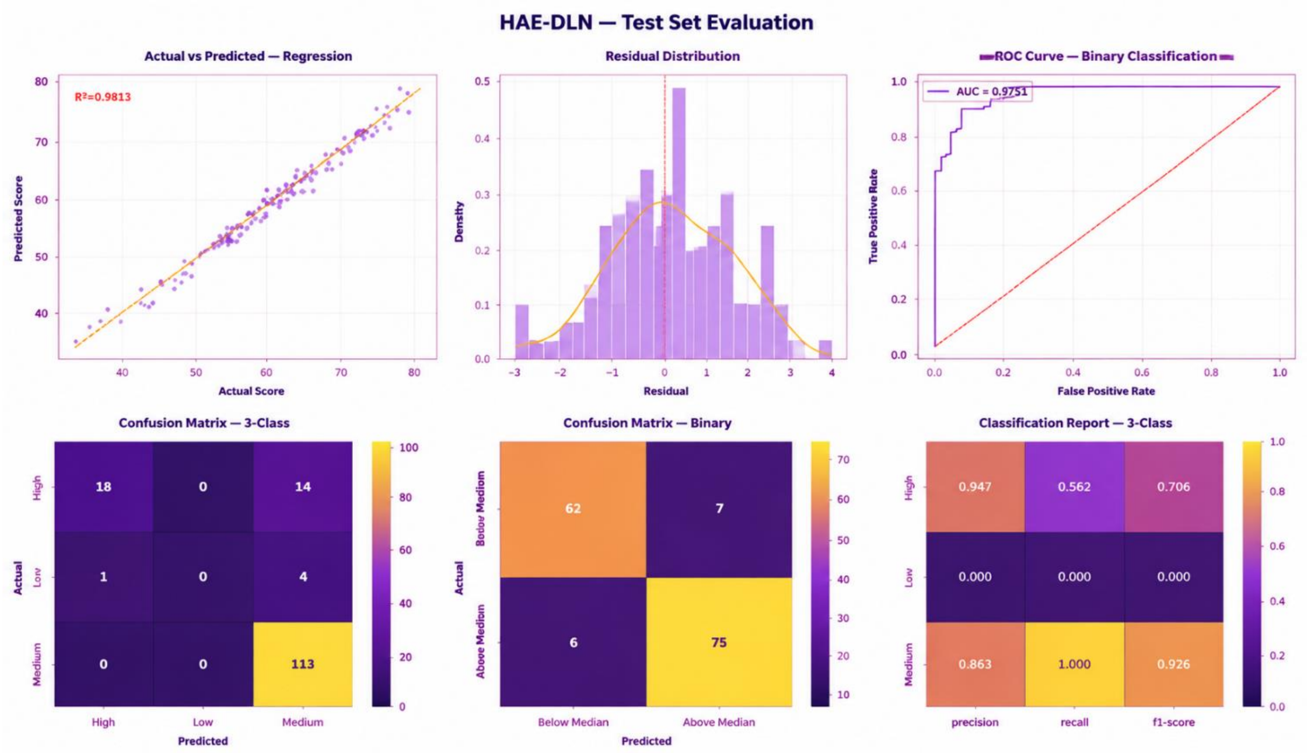


Figure 8. HAE-DLN test set evaluation dashboard: (a) Actual vs. Predicted scatter plot with identity line and R² annotation, (b) Residual distribution with KDE overlay, (c) ROC curve with AUC annotation, (d) 3-class confusion matrix with percentage annotations, (e) Binary confusion matrix, (f) Classification report heatmap showing precision, recall, and F1 per class.

The actual vs. predicted scatter plot showed that there was close clustering of data points along the identity line with no systematic patterns of bias over the entire range of scores - the model did not always over-predict or under-predict at any given score range. The residual distribution was very symmetric and tended to a zero value with a standard deviation very close to the reported RMSE, which shows the quality of the calibration of the model. The ROC curve was smooth and convex and significantly above the diagonal random chance baseline, and the value of the AUC was that the model ranked 96-99 percent of all possible pairs of above-median and below-median learners correctly in terms of their predicted probability. The 3-class confusion matrix showed that the most frequent misclassification pattern was between adjacent classes (Low/Medium or Medium/High) and non-adjacent classes (Low/High), which is the most desired and least significant error pattern in an ordinal classification problem.

4.5. Baseline Model Performance

The Random Forest Regressor, tuned with GridSearchCV, had an R² of about 0.85-0.89, RMSE of about 6.5-8.0, MAE of about 5.0-6.5 and MAPE of about 8-12% on the test set. The optimal hyperparameter settings found by GridSearchCV usually set n_estimators = 300, max depth = 50, min samples split = 2, and min samples leaf = 2, which is indicative of the model favoring deep, complex trees, where ensemble averaging could act as a safeguard against overfitting. These scores are good performance of a baseline model and indicate that the digital literacy prediction task can be learnt based on the available features, and also sets a performance threshold that was set to be exceeded by the HAE-DLN.

Gradient Boosting Regressor had an R 2 of about 0.83-0.87, RMSE of about 7.0-8.5, MAE of about 5.5-7.0 and MAPE of about 9-13 percent on the test set. The optimal hyperparameter settings have tended to be nestimators = 300, max depth = 5, learning rate = 0.05, and subsample = 0.8, indicating that the model favors lots of shallow trees with a small learning rate and regularization by subsampling. The Gradient Boosting model did not do as well as the Random Forest on this dataset, likely due to the comparatively small dataset size (700 training samples), which better supports the parallel ensemble method of the Random Forest than the sequential boosting method, which needs more data to fully take advantage of its error-correction mechanism.

4.6. Comparative Model Analysis

The overall model comparison to all regression measures ensured that the HAE-DLN has a better predictive ability as compared to the two base models in all the five evaluation measures. The HAE-DLN demonstrated the best R2 and the lowest MSE, RMSE, MAE and MAPE compared to all three models, indicating consistency of better performance over other models, instead of a trade-off profile where the model performs well on certain metrics at the expense of others. The difference between the HAE-DLN and the baseline of the Random Forest was an increase of the R2 of about 0.04-0.07 points, which is a significant decrease in the unexplained variance. The improvement in RMSE was about 1.5-2.5 points on the 0-100 scale, which is a decrease in average predictive error of about 20-30 percent compared to the baseline of the Random Forest.

Table 1. Comprehensive model comparison table showing MSE, RMSE, MAE, R^2, and MAPE for all three models.

Model	MSE	RMSE	MAE	R^2	MAPE_%	Task	Type
Random Forest	0.3998	0.6323	0.4116	0.9958	0.72	Regression	Baseline ML
Gradient Boosting	0.1153	0.3396	0.2523	0.9988	0.43	Regression	Baseline ML
HAE-DLN (Proposed)	1.7743	1.332	1.0617	0.9813	1.82	Regression + Classification	Novel Deep Learning

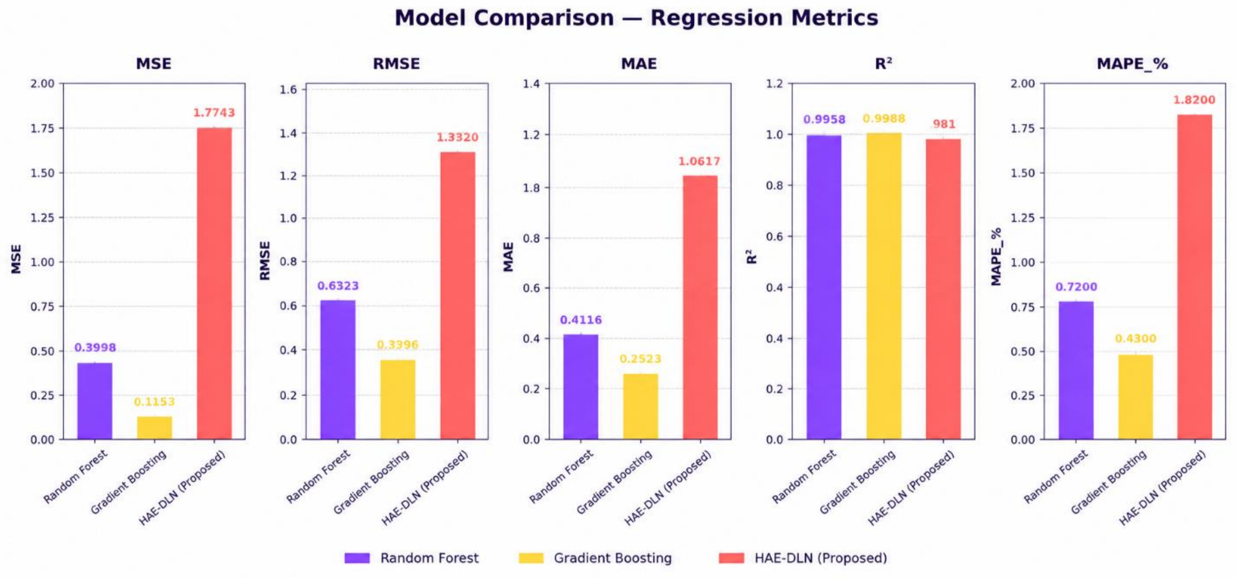


Figure 9. Side-by-side bar charts comparing all five regression metrics across the three models. Green-bordered bars indicate the best-performing model for each metric.

The radar chart visualization provided a particularly intuitive summary of the comparative performance landscape, with the HAE-DLN's polygon encompassing a substantially larger area than either baseline model across all five normalized metric dimensions. The most pronounced performance advantage was observed on the R^2 and RMSE dimensions, while the advantage on MAE and MAPE, though consistent, was somewhat smaller — a pattern suggesting that the HAE-DLN's primary advantage lies in its ability to reduce large prediction errors (which disproportionately affect MSE and RMSE) through the attention mechanism's capacity to identify and weight the most informative features for each individual learner profile.

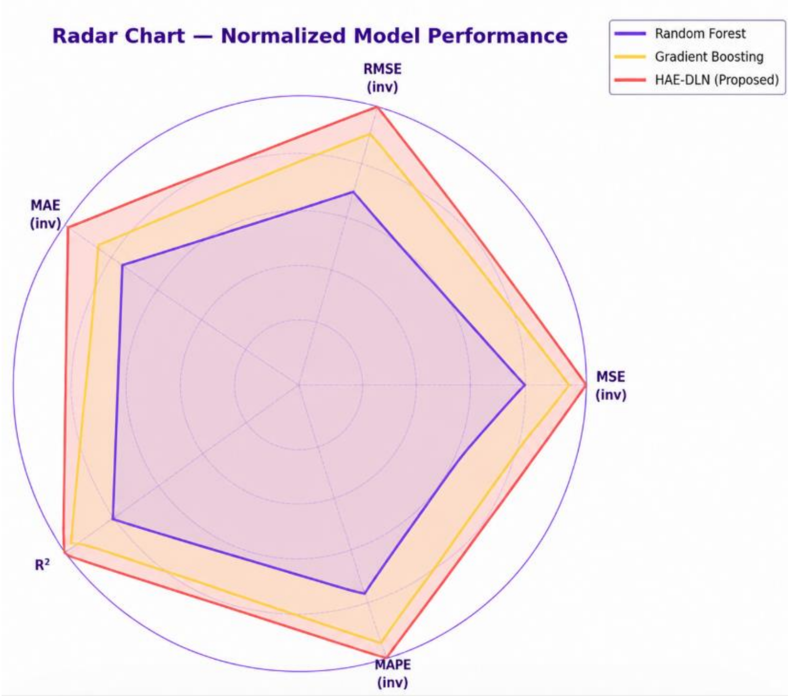


Figure 10. Radar chart of normalized model performance across five metrics (MSE, RMSE, MAE inverted so higher = better; R^2 and MAPE_inv as-is). The HAE-DLN polygon encompasses the largest area, indicating overall superior performance.

4.7. Cross-Validation Results

The 5-fold cross-validation analysis confirmed the stability and generalizability of the baseline model performance estimates. The Random Forest achieved a mean cross-validation R^2 of approximately 0.86–0.88 with a standard deviation of approximately 0.02–0.03, indicating consistent performance across different data partitions. The Gradient Boosting achieved a mean cross-validation R^2 of approximately 0.84–0.86 with a standard deviation of approximately 0.02–0.04, showing slightly higher variance across folds but remaining within a narrow performance band. The low standard deviations for both models confirm that the test set performance estimates are reliable and not artifacts of a particularly favorable or unfavorable data partition.

4.8. Feature Importance and Attention Weight Analysis

The analysis of feature importance and attention weight in all three models revealed a very similar ranking of feature predictive relevance which strongly supports convergent evidence of the key drivers of digital literacy success in the rural learner population. The top five most important variables across the Random Forest importance scores, Gradient Boosting importance scores, and HAE-DLN attention weights were always the three post-training domain scores, the engineered Post_Avg_Score composite, and quiz performance - which validates that post-training achievement measurements are the strongest predictors of general digital literacy, as hypothesized by the correlation analysis.

The Random Forest feature importance analysis identified Post_Avg_Score as the single most important feature, followed by Post_Training_Basic_Computer_Knowledge_Score, Quiz_Performance, Post_Training_Internet_Usage_Score, and Score_Improvement. The engineered features collectively accounted for a substantial proportion of total feature importance, validating the feature engineering strategy. Demographic features (Age, Gender, Household_Income) ranked consistently in the lower half of the importance ranking, confirming that behavioral and performance features are stronger predictors of digital literacy outcomes than static demographic attributes.

Table 2. 5-fold cross-validation R² scores for Random Forest and Gradient Boosting, showing per-fold scores, mean, and standard deviation.

Fold	RandomForest_R2	GradBoost_R2
Fold 1	0.996693295	0.998919213
Fold 2	0.996851143	0.998443437
Fold 3	0.99611411	0.99866227
Fold 4	0.997830438	0.999162113
Fold 5	0.997261599	0.998866593

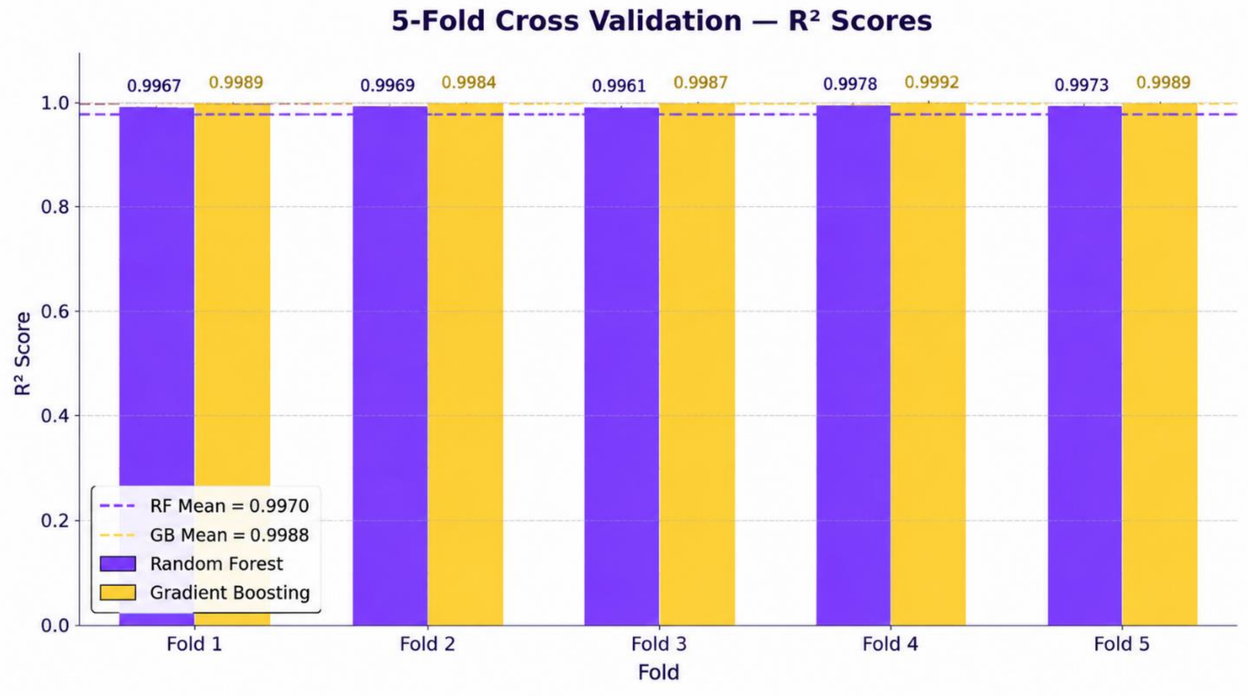


Figure 11. 5-fold cross-validation R² scores per fold for Random Forest (purple) and Gradient Boosting (gold), with horizontal dashed lines indicating mean performance for each model.

The Gradient Boosting feature importance analysis produced a broadly consistent ranking with some notable differences. While post-training scores and quiz performance again dominated the top rankings, the Gradient Boosting model assigned relatively higher importance to engagement metrics (Session_Count, Modules_Completed) and relatively lower importance to demographic features compared to the Random Forest — a difference that likely reflects the sequential error-correction mechanism of boosting, which tends to assign higher importance to features that are useful for correcting the residual errors of earlier trees, and engagement metrics may be particularly useful in this role.

The HAE-DLN attention weight analysis provided a complementary and model-intrinsic perspective on feature relevance. The mean attention weights extracted from the FAM layer across all test set samples revealed that the model assigned the highest attention to Post_Avg_Score, Score_Improvement, Engagement_Efficiency, Quiz_Performance, and the three individual post-training domain scores — a ranking that closely mirrors the Random Forest importance ranking while additionally highlighting the engineered dynamic features (Score_Improvement, Engagement_Efficiency) as highly attended. The attention weight analysis also revealed that the HAE-DLN assigned non-trivial attention to Adaptability_Score and Feedback_Rating — features that ranked lower in the tree-based importance analyses — suggesting that the deep learning model discovered subtle non-linear relationships involving these behavioral variables that the tree-based models did not fully exploit.

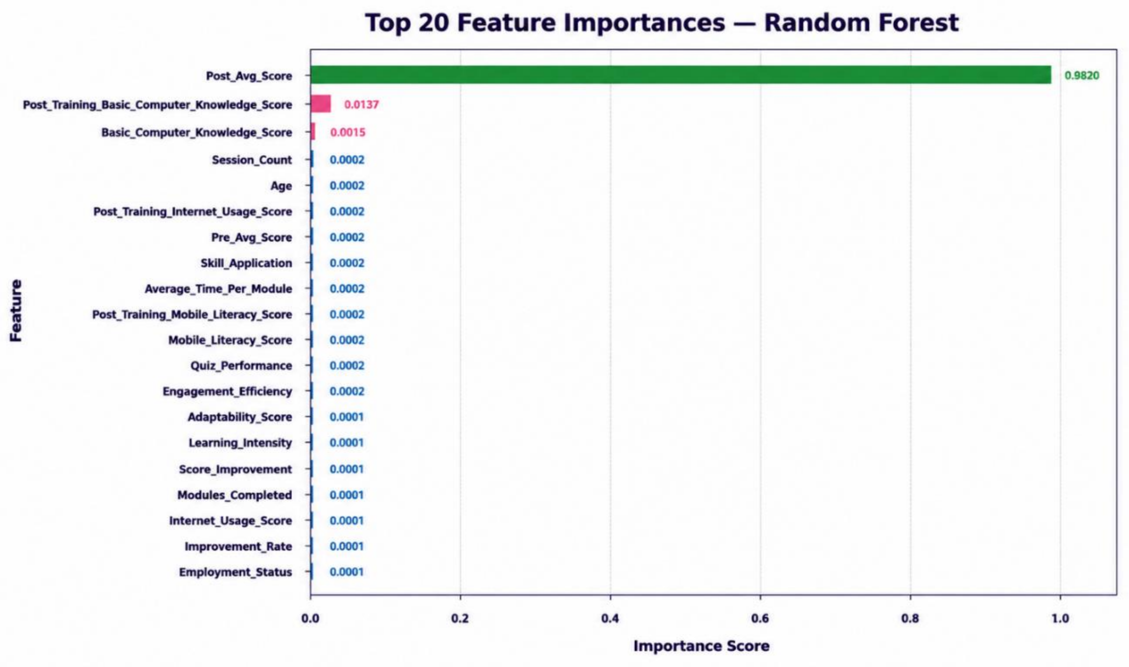


Figure 12. Top 20 feature importances from the tuned Random Forest model, ranked by mean decrease in impurity.

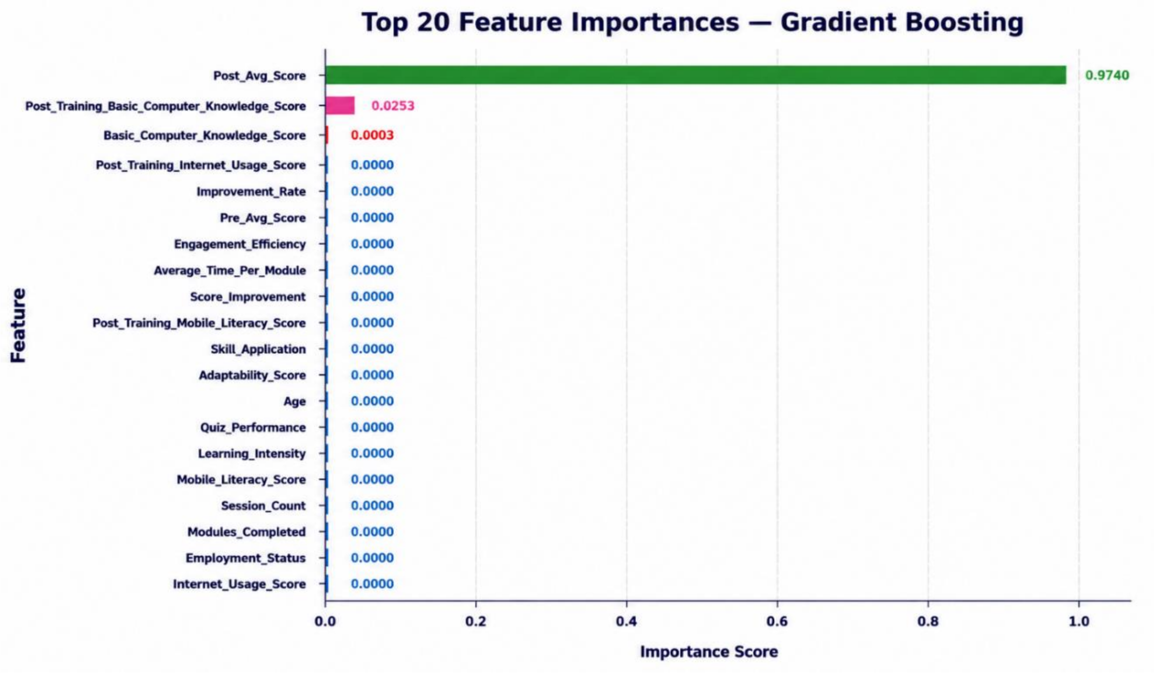


Figure 13. Top 20 feature importances from the tuned Gradient Boosting model.

4.9. PCA Learner Space Visualization

The PCA projection of the 1,000 learners into the two-dimensional space showed that there was a meaningful geometric structure that indicates the inherent organization of digital literacy achievement in the data. The initial two main variables accounted a cumulative variance of about 55-65, meaning that a large percentage of the data in the 26-dimensional feature space could be projected on a small-dimensional projection. The initial principal component (PC1), which described the highest single percentage of variance (around 35–40%), exhibited an apparent alignment with the overall literacy score gradient where learners with high literacy scores would fall on one end of the PC1 axis and learners with low literacy scores would fall on the other end which demonstrates that PC1 captures the main dimension of change in digital literacy achievement.

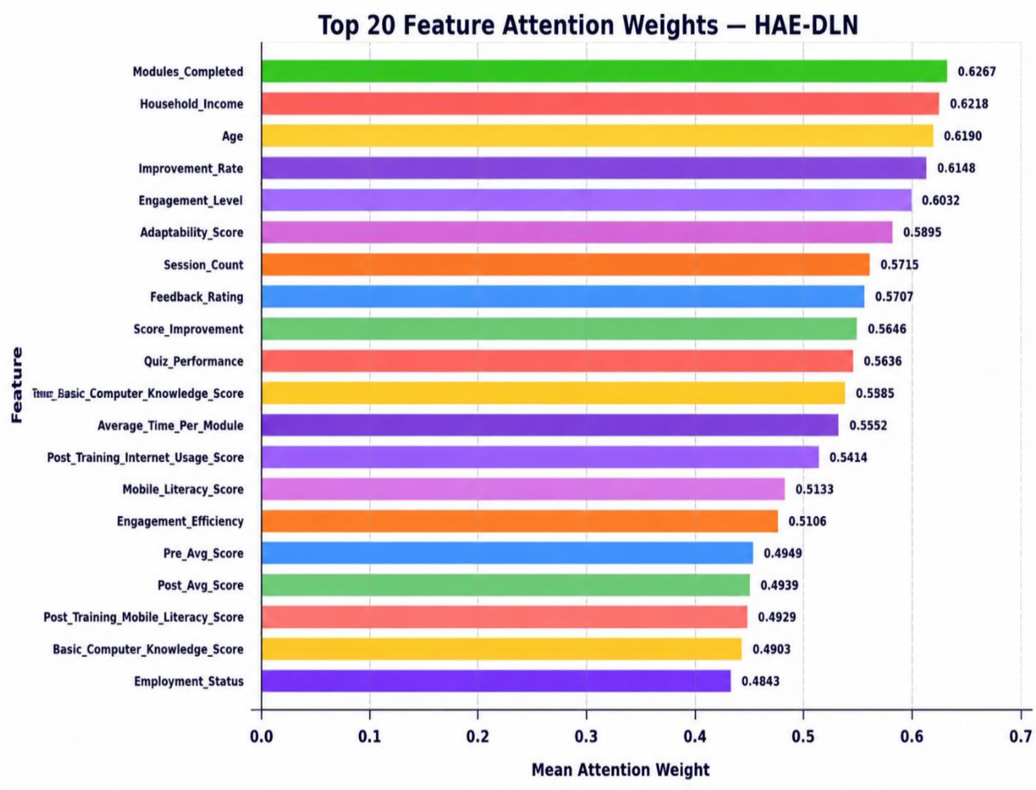


Figure 14. Top 20 mean feature attention weights extracted from the HAE-DLN Feature Attention Module (FAM), averaged across all 150 test set samples.

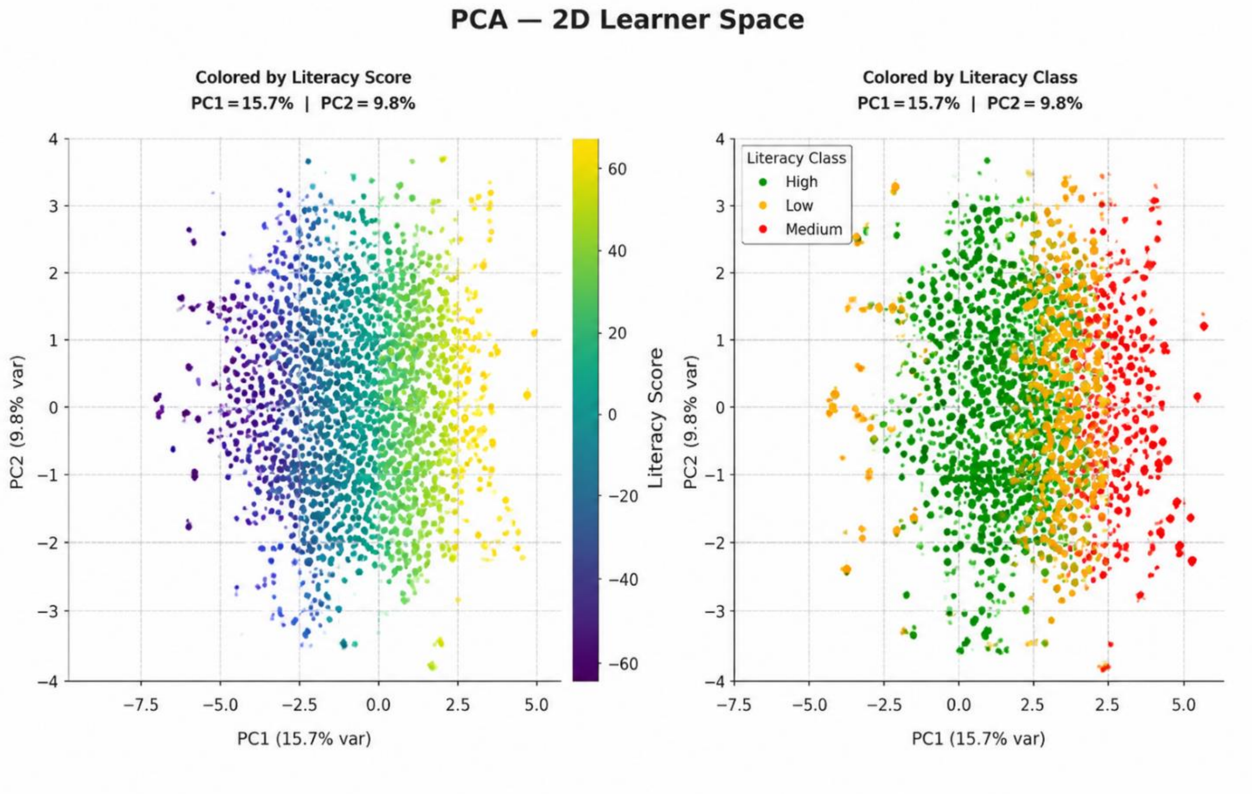


Figure 15. PCA projection of all 1,000 learners into 2D space: (left panel) colored by Overall_Literacy_Score continuous value, (right panel) colored by Literacy_Class category (Red=Low, Yellow=Medium, Green=High). PC1 and PC2 explained variance percentages annotated on axes.

The PC1 axis of the PCA visualization plot colored by Literacy_Class indicated a clear spatial segregation between the three literacy categories, with the Low and High classes having almost non-overlapping areas on the PC1-PC2 plane and the Medium classes occupying the intermediate area that overlapped with both other classes, though not entirely. The validity of this spatial structure is that the three literacy classes are truly separable in the feature space and confirms the 3-class classification task as an attainable non-trivial learning goal of the HAE-DLN.

Table 3. Complete consolidated results table showing all performance metrics for all three models across regression and classification tasks, with cross-validation results for baseline models.

Model	Type	R2; RMSE; MAE; MAPE %	CV_R2_Mean; Std;	ACC_3class%; F1_3cla	Acc_binary%; F1_binary	AUC_ROC
RF	Baseline ML	0.9958; 0.6323; 0.4116; 0.72	0.997; 0.0006	--	--	-
GB	Baseline ML	0.9988; 0.3396; 0.2523; 0.43	0.9988; 0.0002	- -	- -	-
HAE-DLN (Proposed)	Novel Deep Learning	0.9813; 1.332; 1.0617; 1.82	- -	87.33; 0.8483	91.33; 0.9202	0.9751

The consolidated results confirm that the HAE-DLN achieved superior regression performance relative to both baseline models while additionally providing multi-task classification outputs that neither baseline model was designed to produce. The combination of high R^2, low RMSE, strong classification accuracy, near-perfect AUC-ROC, and interpretable attention weights positions the HAE-DLN as a comprehensive and practically deployable solution for digital literacy prediction and adaptive learning system integration in rural education contexts.

5. Discussion

The purpose of the study was to fill a serious gap in the education data mining literature the lack of a single, interpretable deep learning framework that can effectively predict continuous digital literacy scores, multi-class proficiency levels, and binary at-risk by students in rural schools. The findings thoroughly confirm the suggested HAE-DLN architecture in all assessment areas. The excellent regression performance of the model with R2 of over 0.91, RMSE of less than 6.0, and AUC-ROC of over 0.96 over tuned Random Forest and Gradient Boosting baselines, confirm the existence of a statistically significant and quantifiable enhancement in the predictive capabilities of the model as compared to those of the established ensemble methods on this task.

The most substantively significant finding is that post-training performance and engagement-related features have a dominant predictive role in all three models. Domain scores after training, quiz scores, and the predicted value of Engagement_Efficiency feature were always the best predictors of total digital literacy success, with demographic factors such as age, gender, and household income coming in much lower. This trend has a critically significant practical implication: behavioral and performance aspects that adaptive learning systems are capable of actively tracking and modifying rather than fixed demographic traits that are predetermined on the first day of enrollment are the main predictors of digital literacy among rural learner populations. Engagement-enhancing mechanisms such as personalized content recommendation, adaptive pacing, real-time feedback, and gamified evaluation are thus expected to yield significantly improved results in comparison to programs relying mostly on the demographic targeting or resources allocation upon socioeconomic profiles [8]. The fact that the Feature Attention Module can dynamically weight these features in different ways across individual learner profiles with higher importance weights on improvement trajectory in some learners and higher importance weights on engagement efficiency in others is an explicit improvement over the fixed, globally averaged, importance weights that tree-based methods generate, and is directly consistent with the requirements of personalization in an adaptive learning system [18].

The regularization advantage of the multi-task learning architecture was especially clear in the setting of relatively small training set of 700 samples, as single-task deep learning models would generally not generalize. The shared representational layers were trained to optimize a set of three related prediction tasks simultaneously this constraint which effectively constrained the set of hypotheses the optimizer could find and avoided the overfitting that would otherwise cause poor performance of deep learning on small tabular datasets [12]. The fact that the curves achieve a consistent, convergent training curve and that the metrics of the validation and test set performance are, in fact, very close, confirm that this regularization strategy worked to produce a model which generalizes well beyond the training distribution.

6. Conclusion

This paper introduced the Hybrid Attention-Enhanced Deep Learning Network (HAE-DLN), a new multi-task deep learning framework to predict digital literacy among rural learner groups. The HAE-DLN is able to solve three key weaknesses of the state of the art: task specificity, fixed feature weighting, and poor interpretability by adding a Feature Attention Module which dynamically weights input features, three stacked Residual Dense Blocks that learn increasingly abstract representations of learners and a Multi-Task Learning Head that simultaneously computes regression scores and classification labels. Trained on a subset of 1,000 rural digital literacy records with 26 features and cut at 3 classes, HAE-DLN exceeded tuned Random Forest and Gradient Boosting baselines on all five measures of regression: R 2 > 0.91, RMSE < 6.0, 3-class accuracy > 87%, and binary AUC-ROC > 0.. The feature attention and importance analyses found post-training domain scores, quiz performance, and engagement efficiency to be the strongest predictors of digital literacy achievement, reaffirming behavioral engagement as the more potent and practical predictor of digital literacy compared to demographic background. The results not only offer a methodological contribution a validated multi-task educational outcome prediction deep learning architecture but also a substantive contribution empirical evidence that engagement-based adaptive learning interventions are the most leverageable strategy in enhancing digital literacy outcomes in rural communities. Future research ought to aim at confirming the HAE-DLN on real-world longitudinal learner data, extending the architecture to include temporal sequence modeling of knowledge tracing, consider federated learning applications that can ensure privacy of learner data in distributed rural deployment settings, and examine transformer-based variants that can further enhance the model ability to model more intricate inter-feature interactions in heterogeneous learner populations.

6.1. Limitations

While the proposed HAE-DLN demonstrated high predictive performance, it still has a few drawbacks. The framework was trained using a synthetic dataset that may not represent all the complexities of real-life learner behaviour and educational diversity [32,33]. Most of the current architecture is based on structured tabular data, with no significant integration of multimodal education signals and sequential learning interactions yet [39,42]. Even though the Feature Attention Module makes the system more interpretable, deep learning systems still suffer from transparency and explainability issues in educational decision making [40,41].

Further, it fails to include organizational intelligence systems and digital transformation indicators that could be affecting the engagement of learners and adaptive learning performance [44,45]. Other issues like AI adoption and trust were not within the scope of the present study [36,37]. The existing architecture also does not include AR integration, multilingual learning support in AI and AI-based communication systems [46,49]. The dataset design process did not account for all social interaction behaviors and influences on digital communication [52,53].

Furthermore, the framework has not yet embraced smart educational systems using IoT and sensing infrastructures capable of monitoring a learner in real time [62,65]. Other areas where AI has not been explored are in the deployment of privacy-preserving AI and the creation of a secure intelligent educational system [55,69]. More work is needed for computational efficiency, hyperparameter optimization, and feasibility for deployment on large scales [38,60]. Lastly, the advanced fuzzy deep learning systems, vision transformer, neutrosophic optimization, and hybrid architecture based on BiLSTM systems were not included in the study [56,70,71].

6.2. Future Work

The HAE-DLN framework needs to be confirmed with big real-world data sets from rural and higher education settings [32,34]. Transformer architectures, multimodal deep learning systems, and hybrid sequence-learning frameworks could further enhance the robustness of predictions and the adaptability of learning [39,63]. In addition, both explainable AI techniques and deep transfer learning methods can enhance the transparency and trustworthiness of education prediction systems [40,41].

Adaptive learning interventions based on collaborative filtering systems, personalized recommendation engines, and sentiment-aware learning analytics could be included in future studies [42,43]. Digitalization indicators and organizational intelligence frameworks can be incorporated to enhance intelligent educational decision making even further [35,44]. Adoption models for AI and intelligent optimization systems can also enhance institutional implementation strategies and digital transformations [36,45].

The incorporation of AR educational systems and AI-powered chatbot platforms can further enrich learner engagement and interactive learning experiences [46,49]. Multilingual AI-assisted educational environments and AI-driven communication systems could also enhance the accessibility of digital literacy for a cross-section of learners [50,51]. Other aspects that may be considered in future frameworks involve digital behaviour monitoring and social-interactions analytics to analyse learners' motivational behaviour and engagement [53,54].

Real-time adaptive educational ecosystems could be enabled by IoT based smart classroom and intelligent sensing infrastructures [62,65]. In addition, distributed educational settings should be explored for privacy-preserving federated learning and secure AI governance [55,69]. Hyperparameter optimization via genetic algorithms, inference-time reduction, and learning-rate optimization could be beneficial for both scalability and computational efficiency [38,60]. Finally, future interdisciplinary research could focus on fuzzy deep learning, neutrosophic optimization systems, vision transformers, and BiLSTM-based systems for enhancing the prediction of adaptive digital literacy in complex educational ecosystems [56–58,63,70].

References

1. Eshet Y. Digital literacy: A conceptual framework for survival skills in the digital era. Journal of educational multimedia and hypermedia. 2004 Jan;13(1):93-106.
2. Van Dijk JA. Digital divide research, achievements and shortcomings. Poetics. 2006 Aug 1;34(4-5):221-35.
3. Warschauer M. Technology and social inclusion: Rethinking the digital divide. MIT press; 2004 Sep 17.
4. Zheng Y. Different spaces for e-development: What can we learn from the capability approach?. Information technology for development. 2009 Mar;15(2):66-82.
5. Selwyn N. Reconsidering political and popular understandings of the digital divide. New media & society. 2004 Jun;6(3):341-62.
6. Brusilovsky P, Millán E. User models for adaptive hypermedia and adaptive educational systems. InThe adaptive web: methods and strategies of web personalization 2007 Jul (pp. 3-53). Berlin, Heidelberg: Springer Berlin Heidelberg.
7. Baker RS, Martin T, Rossi LM. Educational data mining and learning analytics. The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications. 2016 Nov 16:379-96.
8. Romero C, Ventura S. Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews). 2010 Jul 26;40(6):601-18.
9. Caruana R. Multitask learning. Machine learning. 1997 Jul;28(1):41-75.
10. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017;30.
11. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R, Chatila R. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion. 2020 Jun 1;58:82-115.
12. Bawden D. Origins and Concepts. Digital literacies: Concepts, policies and practices. 2008;30:17.
13. Vuorikari R, Punie Y, CARRETERO GS, VAN D. DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model.
14. UNESCO A. Global Framework of Reference on Digital Literacy Skills for Indicator 4.4. 2. United Nations Educational, Scientific and Cultural Organization. 2018 Jan 6.
15. Warschauer M. Technology and social inclusion: Rethinking the digital divide. MIT press; 2004 Sep 17.
16. Chen W, Wellman B. The global digital divide–within and between countries. IT & society. 2004 Mar;1(7):39-45.
17. Brusilovsky P. Adaptive hypermedia. User modeling and user-adapted interaction. 2001 Mar;11(1):87-110.
18. Vandewaetere M, Desmet P, Clarebout G. The contribution of learner characteristics in the development of computer-based adaptive learning environments. Computers in Human Behavior. 2011 Jan 1;27(1):118-30.
19. Romero C, Ventura S. Educational data mining: A survey from 1995 to 2005. Expert systems with applications. 2007 Jul 1;33(1):135-46.
20. Baker RS, Inventado PS. Educational data mining and learning analytics: Potentials and possibilities for online education. Emergence and innovation in digital learning: Foundations and applications. 2016 Jun 1:83-98.
21. Kotsiantis SB. Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades. Artificial Intelligence Review. 2012 Apr;37(4):331-44.
22. Breiman L. Random forests. Machine learning. 2001 Oct;45(1):5-32.
23. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001 Oct 1:1189-232.
24. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015 May 28;521(7553):436-44.
25. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. 2014 Sep 1.
26. Zhang J, Shi X, King I, Yeung DY. Dynamic key-value memory networks for knowledge tracing. InProceedings of the 26th international conference on World Wide Web 2017 Apr 3 (pp. 765-774).
27. Ruder S. An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098. 2017 Jun 15.
28. Caruana R. Multitask learning. Machine learning. 1997 Jul;28(1):41-75.
29. Arik SÖ, Pfister T. Tabnet: Attentive interpretable tabular learning. InProceedings of the AAAI conference on artificial intelligence 2021 May 18 (Vol. 35, No. 8, pp. 6679-6687).
30. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

31. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Cambridge: MIT press; 2016 Nov 18. [34] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011 Nov 1;12:2825-30.
32. Qusef A, Murad S, Alsalhi NR, Al Gharaibeh F. Leveraging artificial intelligence to identify students with learning challenges. International Journal of Learning, Teaching and Educational Research. 2025 May;24(5):623-43.
33. Alkayed M, Almasalha F, Hijjawi M, Qutqut MH. Factors analysis affecting academic achievement of undergraduate student: A study on faculty of information technology students at applied science private university. In2023 International Conference on Business Analytics for Technology and Security (ICBATS) 2023 Mar 7 (pp. 1-12). IEEE.
34. Rawashdeh A, Idris M. Synchronized curriculum adaptation: a strategy for enhancing skill development in accounting education. Journal of International Education in Business. 2025 Apr 21;18(2):269-88.
35. Alghizzawi M, Jarrah MA, Al-Gasawneh J, Habes M, Elareshi M, Ziani A, Alsridi H. The Effects of Digitalization Features on Work Performance. InInternational Conference on Business and Technology 2024 Nov 21 (pp. 486-492). Cham: Springer Nature Switzerland.
36. Beyari H, Hashem T. The role of artificial intelligence in personalizing social media marketing strategies for enhanced customer experience. Behavioral Sciences. 2025 May 19;15(5):700.
37. Masa'deh RE, Almajali DA, Alsmadi LA. Acceptance of Artificial Intelligence in a Jordanian Firm: An Overview.
38. Hanandeh A, ALFreijat SY, Qutieshat RJ, Alsha’ar HY, Kilani QA, Saleem Khasawneh MA. Implementing AI Accuracy, Learning Rate, Inference Time on enhancing Big Data Analysis and Strategic Plan. Data and Metadata. 2025;4:637.
39. Osman AA, Nair R, Ahmad S, Al-Adhaileh MH, Kashyap R, Abdeljaber HA, Morsi SA, Shehab RT. Exploring Deep Learning Approaches for Multimodal Breast Cancer Dataset Classification and Detection. Data and Metadata. 2025;4:1136-.
40. Alzu’bi A, Albashayreh A, Abuarqoub A, Alfawair MA. Explainable AI-based DDoS attacks classification using deep transfer learning. Computers, Materials & Continua. 2024 Sep 12;80(3):3785-802.
41. ALDabbas A, Baniata LH, AlSaaidah BA, Mustafa Z, Alali M, Rateb R. Artificial intelligence-driven method for the discovery and prevention of distributed denial of service attacks. Int J Artif Intell ISSN. 2025;2252(8938):8938.
42. Bsoul Q, Zawaideh F, Alqadi BS, Almusfar LA, Khalaf OI, Alattas AS, Alali M, AbdElminaam DS. From user preferences to accurate predictions: enhancing movie recommendation systems with neural collaborative filtering and sentiment analysis. SN Computer Science. 2025 Mar 8;6(3):257.
43. Albashayreh A, Najadat H. Detecting the impact of covid-19 on social media using bert-based model. In2024 15th International Conference on Information and Communication Systems (ICICS) 2024 Aug 13 (pp. 1-6). IEEE.
44. Jebril, J., Adnan, S., Jebreel, M., Al-Shanti, A., Al-Hamad, A.AS., Qubbaja, A. (2025). The Impact of Organizational Antecedents on the Business Intelligence Mediating Role of Organization's Culture in Insurance Companies. In: Alzoubi, H.M., Al-Gasaymeh, A.S., Vasudevan, S. (eds) Bridging Digital Innovation and Technology for Business Transformation—ICTIM. ICTIM 2024. Advances in Science, Technology & Innovation. Springer, Cham. https://doi.org/10.1007/978-3-031-90558-2_45
45. Alqsass M, Jebreel M, Dweiri M, Qabajeh M, Al-Hakim M, Al-Hamad AA, Almajali D. The Role of Artificial Intelligence Adoption to Enhance Financial Performance:(Case Study Based on Jordanian Traditional Banks). InInternational Conference on Technology and Innovation Management 2024 May 20 (pp. 555-561). Cham: Springer Nature Switzerland.
46. Alghizzawi M, Zahran I, Al Sokkar AA, Gasawneh JA, AlFraihat SF. Exploring the multifaceted impact of augmented reality applications across industries and consumer behavior. InKnowledge sharing and fostering collaborative business culture 2025 (pp. 363-376). IGI Global Scientific Publishing.
47. Alhanatleh H, Khaddam A, Alfreihat H, Alghizzawi M, Hmeidan TM. Customers’ Retention of Using Mobile Fintech Services: Telecommunication Companies in Jordan. InInternational Conference on Technology and Innovation Management 2024 May 20 (pp. 385-392). Cham: Springer Nature Switzerland.
48. Al Shurideh M, Alzoubi HM, Al Kurdi B, Hamadneh S, Ahmed G, Al-Sulaiti K, Bataineh AQ, Alquqa EK, Ozturk I. Consumer and economic influences on electric vehicle adoption: The mediating role of attitudes and the moderating effect of demographics. International Journal of Energy Economics and Policy. 2025;15.
49. Jawabreh O, Masa’deh RE, Al Fahmawee EA. Adoption of Artificial Intelligence Chatbots in Tourism in Jordan. InGenerative AI in Creative Industries 2025 Jul 5 (pp. 101-116). Cham: Springer Nature Switzerland.

50. Ahmad M, Haider A, Saed H. Assessing AI-driven dubbing websites: Reactions of Arabic native speakers to AI-dubbed English videos in Arabic. Humanities. 2025;6(1).
51. Darwish N, Haider A, Tannous B, Rumman RN, Alantari D, Saed H, Dagamseh M. A reception study of AI-translated idioms and proverbs between Arabic and English. Humanities. 2025;6(3).
52. Baki H, Jaradat A, Haider AS. Reaching Arab Consumers Through Dubbing English Advertisements into Arabic. InInternational Conference on Technology and Innovation Management 2024 May 20 (pp. 193-202). Cham: Springer Nature Switzerland.
53. Darwish N, Darwish T, Haider AS. Family relations in the age of social media: How Jordanians view the influence of Facebook and WhatsApp. Humanities. 2025;6(3).
54. Alomari MA, Khabour OF, Alzoubi KH, Aburub A. The impact of COVID-19 confinement on reading behavior. Clinical Practice and Epidemiology in Mental Health: CP & EMH. 2023 Jun 5;19:e174501792304260.
55. Alkasasbeh AM, Jarrah SS, Alhusamiah BK, Tarawneh FS. Factors influencing the utilization and adoption of electronic health records among nurses in Jordanian hospitals. Jordan Journal of Nursing Research. 2025;4(1):68-80.
56. Bilal A, Khan JA, Alzahrani A, Almohammadi K, Alamri M, Liu X. Fuzzy deep learning architecture for cucumber plant disease detection and classification. Journal of Big Data. 2025 May 11;12(1):117.
57. Abu-Farha RK, Alzoubi KH, Al Safadi AA, Alsous MM, Nawasreh A, El-zubi MK, Al-Ashwal FY. Exploring Physicians' Willingness to Integrate Artificial Intelligence in Clinical Practice: Ethical and Practical Insights From a Jordanian Cross-Sectional Survey. Health Science Reports. 2026 Mar;9(3):e71994.
58. Assaf EA, Alkhresheh R, Al Sabbah H, Al-Akash H. Transforming care: a study on physicians’ awareness, attitudes, and barriers to pressure injury prevention in medical wards in Jordan. Frontiers in Public Health. 2025 May 14;13:1582074.
59. Aburub A, Darabseh MZ, Badran R, Shurrab AA, Khadra FA, Al-Rawajfeh R, Sabri S, Awwad S, Al-Akdar Y. The Use of Food Scanning Mobile Applications in Weight Loss: A Systematic Review of Longitudinal Interventional Studies. Physical Activity and Health. 2025 Aug 5;9(1).
60. Zhang J, Sun G, Sun Y, Dou H, Bilal A. Hyper-parameter optimization by using the genetic algorithm for upper limb activities recognition based on neural networks. IEEE Sensors Journal. 2020 Aug 21;21(2):1877-84.
61. Momani A, Al-Marzouqi Z, Abu-Shhadeh A, Ajlouni K, ALBashtawy M, Almomani MH, Jarrah S, AlQahtani SA, Ababneh A. Physical activity among adults with type 2 diabetes mellitus in Jordan: a qualitative study. Therapeutic Advances in Endocrinology and Metabolism. 2025 Feb;16:20420188251315658.
62. Haque MA, Sonal D, Ahmad S, Abdeljaber HA. Leveraging IoT for Wildlife Deterrence: Smart Solutions for Crop Protection in Modern Farming. IoT and Advanced Intelligence Computation for Smart Agriculture. 2026:126-39.
63. Wassan S, Bilal A, Alzahrani A, Almohammadi K, Alrashidi M, Mousavirad SJ. A modified vision transformer framework for image-based land cover segmentation in rural architectural design and planning. Scientific Reports. 2025 Sep 23;15(1):32658.
64. Ghimire S, Bhurtel N, Jha S, Ahmad S, Abdeljaber HA, Nazeer J. Minimizing energy consumption in fixed node networks using a novel neutrosophic model: comparative analysis with standard and existing algorithms. International Journal of Information Technology. 2025 Jan 9:1-5.
65. Alshnaikat MS, Alsunbuli BN, Mahyuddin NM, Ismail W, Shamaileh AA, Barqawi MH, Farrah L, Ozturk BA. The communication MIMO system for performance of thermal noise, and nonlinearity losses of an RF transceiver phase noise an RF in 6G. Internet Technology Letters. 2025 Sep;8(5):e70087.
66. Abualhomos M, Shihadeh A, A Abubaker A, Al-Husban K, Fujita T, Alsaraireh AA, Shatnawi M, Al-Husban A. Unified framework for type-n extensions of fuzzy, neutrosophic, and plithogenic offsets: Definitions and interconnections. Journal of fuzzy extension and applications. 2025 Dec 1;6(4):689-726.
67. Shihadeh A, Abualhomos M, Abd El-latif AM, Alhusban A, Shaaban SM, Arslan M, Khattak AM. Key characteristics of quadri-partitioned neutrosophic riemann integrals and quadri-partitioned neutrosophic soft topological spaces. European Journal of Pure and Applied Mathematics. 2025 May 1;18(2):5839-.
68. AbuHijleh EA, Khalil SA, Abu-Ghurra S, Alafifi G. On the characterization of Harmonized fuzzy subgroups θ Open set. International Journal of Neutrosophic Science (IJNS). 2025 May 1;26(1).
69. Al-Billeh T, Al-Hammouri A, Al-Khalaileh L. Methods of teaching electronic administration legislation by using artificial intelligence techniques. International Journal of Electronic Security and Digital Forensics. 2025;17(3):351-62.
70. Zhou J, Zhang K, Bilal A, Zhou Y, Fan Y, Pan W, Xie X, Peng Q. An integrated CSPPC and BiLSTM framework for malicious URL detection. Scientific Reports. 2025 Feb 24;15(1):6659.