A Hybrid Deep Ensemble Framework for Predicting Learning Impacts across Diverse Student Cohorts in Electronic Learning Systems

A. Thilaka [1,*], B. Mahalakshmi [2], J. Devagnanam [1], S. Anitha Rajathi [3], T. Nathiya [2], and E. Sundaravalli [4]

[1] Department of Computer Applications, Faculty of Science and Humanities, SRM Institute of Science and Technologies, Kattankulathur, India
[2] Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technologies, Kattankulathur, India
[3] Department of Computer Science and Business School, R.M.D Engineering College, Kavaraipettai, India
[4] Department of Computer Science, Thavathiru Santhalinga Adigalar Arts Science and Tamil College, Coimbatore, India
Email: thilakaa@srmist.edu.in (A.T.); mahalakb1@srmist.edu.in (B.M.); devagnaj@srmist.edu.in (J.D.); anitharajathi91@gmail.com (S.A.R.); nathiyat@srmist.edu.in (T.N.); esundaravalli23@gmail.com (E.S.)
*Corresponding author

Manuscript received February 11, 2026; revised April 13, 2026; accepted May 22, 2026; published November 24, 2026

***Abstract*—Education is the foundation for building values, autonomy, strength, and meaningful lives for all human beings. In this era of technology, the education system has evolved through multiple advanced techniques. E-learning is a significant turning point where more effective and inclusive learning is achieved. The e-learning platforms enable the learning process to be flexible, accessible and personalized. The proposed system exhibits a novel approach by integrating machine learning and deep learning models to predict the impacts of the e-learning systems on a diverse set of students from various schools, universities and aspirants. A hybrid deep ensemble model was implemented by integrating Random Forest, XGBoost, LightGBM, CatBoost, MLP Regressor and ElasticNet Regression algorithms. The feature vectors are generated by the Embedding Neural Networks and combined with the original encoded data to serve as the input data to the advanced stacking model. This model achieved a Root Mean Squared Error of 0.8831, a Mean Absolute Error of 0.6840 and an R^2 Score of 0.8663. This evaluation result helps both students and developers to build strategic plans, identify risk, drive motivation, foster personalization features, improve participation and detect the struggles of the students.**

***Keywords*—e-learning, machine learning, deep learning, embedding neural networks and advanced stacking**

I. Introduction

E-learning, or electronic learning, is the process of transferring skills and knowledge via the internet [1]. Any individual can learn new things from anywhere and at any time through their smartphones, tablets or computers. It includes significant key features such as digital delivery, remote accessibility, self-paced/instructor-led, and its scalability satisfies the demands across the world. There are several types of e-learning available in the market.

- Synchronous E-learning: It is in the form of real-time learning where the learners can raise their doubts and the instructor solves their doubts in real-time at the same time. Live webinars or virtual classrooms are some of the examples here.
- Asynchronous E-learning: It is a self-paced form of learning in which pre-recorded videos and forums are streamed on e-learning platforms. The live interaction is not possible during the learning.
- Blended Learning: A Combination of online learning and face-to-face classroom interaction can occur in university courses. The learner can learn from the instructor with the online components in their physical classrooms.
- M-learning: Mobile learning is the most common one, especially in this era of smartphones. Most of the digital learners use their hand-held devices to utilize e-learning platforms.
- Micro-learning: The contents of micro-learning are keenly focused on a particular topic within a 10-minute duration. As it contains short content, it enables quick learning or revision on small topics.
- Gamified Learning: It is one of the innovative e-learning models that generate points, levels and badges for each activity of online learners. It follows a game-like environment to increase the engagement of learners.

The COVID-19 pandemic expedited the virtual transition of education and restructured the way of learning. Universities, institutions and schools were switched into e-learning to mitigate the spread of COVID-19 [2]. The virtual classrooms like Google Meet, MS Teams and Zoom Meetings were adapted by professors, teachers and students at the time of the worldwide shutdown [3]. These classrooms facilitate teaching as well as tracking attendance, assignments, and performances of the students. Technologies like Virtual Reality (VR) and Augmented Reality (AR) supported the virtual labs and simulations for science, medicine and engineering streams. It also gave opportunities to adopt e-learning platforms such as Coursera, edX, Udemy, YouTube, Khan Academy, and BYJU'S to upskill or to bridge learning gaps. AI tutors, self-paced learning and interactive learning are the key drivers to jump into the e-learning systems. The technology enables global access to students from remote locations and in developing countries. Even though digital divides exist, smartphones and the internet pave the path to world-class studies. The teachers also adapted these digital environments and new teaching modes to do their jobs. The new teaching modes, such as asynchronous learning, blended learning and flipped classrooms, enable technology-driven tutoring. This approach provides real-time analysis of student performance and tracks students' challenges to construct tailored strategies. The pandemic period had a profound impact on e-learning systems due to their deep, flexible, scalable, and personalized features. Still, a lot of institutions continue

hybrid learning as a core element of mainstream education.

E-learning has opened up the gateway of education, overcoming geographical, financial, social and cultural barriers by removing physical infrastructures globally. Some e-learning platforms offer free lessons to school students who can't afford to go to traditional schools. Some of the platforms offer top university courses as well as free auditing of courses and scholarships to afford certificate courses. Moreover, the e-learning applications have a wide range of subjects like academics, vocational, personality development and technical skills. It follows diversified styles of learning process such as videos, audio, quizzes and hands-on practices. The customization of content in e-learning satisfies every individual requirement. This ensures that no one is out. Especially, e-learning extends the learning process lifelong by encouraging continuous learning and development in each learner.

Though e-learning has revolutionized the education system, it has some limitations and challenges for a diverse set of learners. Typical e-learning models could not address their preferences and learning styles, which leads to a lack of understanding and engagement. The problems of the diverse e-learners should be addressed and rectified by the e-learning platforms. The learners from rural or unprivileged areas are affected by the digital divide. The stable internet connection and basic financial needs are the main challenges for these students. The e-learning providers should enhance their accessibility and also communicate with the Government to bridge this gap for the students from rural/ unprivileged locations. Information overload is another concern in e-learning engagements. The excessive and complex content makes them exhausted and annoyed, and leads to discontinuing the learning process. The proper design of the course content with motivational thoughts is required to avoid this problem. E-learners strive for self-paced learning content to fulfil their educational requirements rather than sitting on previously established or more advanced content. The e-learning platforms must adjust their content according to the user's performance and feedback. The disabled learner's problems should be addressed along with the content formats. The cultural and language barriers should be eliminated to grant efficient and equitable education to the students across the globe.

Most of the time, educational institutions apply the tools or insights of theories at the learner level. The beginner-level learner needs more time and detailed explanations to understand the concepts; at the same time, the topper wishes to explore advanced subjects and concepts. The adaptive learning of content is fundamental to reaching an effective education system. In a traditional learning environment, a lot of challenges and struggles exist to make it possible. The dynamically adjusted contents are not possible to implement in traditional schools/universities. It will create unavoidable pressure on diverse learners. Students' disengagements, drop-outs and poor performances are increasing without a dynamically adjusted curriculum. The adaptation of a dynamically adjusted curriculum in educational institutions is beyond reach, and this problem affects both students and teachers.

The e-learning system aids in overcoming all these issues. Still, the e-learning platforms have mitigations for content delivery. The learner expectations are unsatisfied due to its rigid deadlines, passive learning environment, abstract theories, generic responses, lack of personalization and real-time applications. The time gap among these diverse learners makes it even more complex to utilize these e-learning platforms. The diverse sets of students are:

(1) K-12 learners: The K-12 learners are the dependent learners who are pursuing their school education from kindergarten to 12th grade by default, in the age range of 5 to 18 years. They opted for synchronous and asynchronous e-learning platforms. These students are Enthusiastic about the interactive sessions and gamified content as they are at a beginner level in e-learning.
(2) Undergraduates: The undergraduate students utilize e-learning platforms for their academic purposes and personality development. Most of the students access these platforms through mobile applications. These applications act as a centralized repository to access their video lectures, notes, assignments and test materials.
(3) Post-Graduates and Research Scholars: The Post-Graduates and Research Scholars are self-directed and more independent students. These learners belong to an advanced level in e-learning. They engage with advanced and research-oriented content.
(4) Aspirants: Aspirants are individual learner who seeks learning resources from e-learning platforms. These resources offer learning materials to prepare for their competitive examinations, skill development and career up-gradation.

The e-learning platforms furnish a flexible learning environment, a wide range of resources, world-class instructors, personalized learning, free materials and certificates. Still, the platform needs to be reformed by considering the limitations and struggles of diverse learners. Digital divide, lack of personalization, cognitive overload, language barriers, financial barriers, assessments and technological illiteracy are some of the struggles faced by e-learners.

The e-learning systems often fail to fulfill the requirements of each learner globally. It needs to be designed based on different learners' experiences and feedback. There is a gap between the content and the learner's subject knowledge of the delivered content. It should consider their educational background, intention, age group, cognitive level, and engagement levels, etc. But it is difficult to assess as an instructor or course designer. The construction of adaptive learning paths leads to effective and inclusive e-learning systems. Through the targeted actions, the personalization of content can be achieved. The analysis of e-learning impacts on diverse learners reveals the huge interpretations and perspectives. The interpreted results ensure the maximum outcome in designing e-learning platforms.

The realization of real-time impacts of different levels of learners on e-learning is crucial to develop high-performing and result-driven platforms to achieve maximum engagement, accommodate a diverse set of students and provide an inclusive learning experience. The detection of different cognitive abilities, learning style preferences and motivations addresses the unique demands of beginners, intermediate and advanced levels of e-learners. The unabridged potential of learners, drop-out reduction, and an

impactful and more equitable learning environment will be uncovered with the use of enhanced e-learning systems.

Machine learning and deep learning models made huge transformative impacts on various fields, including education, media, health care, manufacturing, finance and more. Machine learning enables the machine to learn from the dataset. The machine learning systems recognize the patterns of the data and yield data-driven decisions. Minimal human intervention is required to build a machine learning model. It comprises multiple algorithms which execute step-by-step progress to derive the solution to a problem using the structured dataset. Whereas, the deep learning models apply neural networks to yield a solution to a problem, even with an unstructured dataset. The integration of ML and DL technologies solves many of the complex problems. They can handle a large amount of data, automation, enhanced accuracy in decision making, personalization, creating ground for innovations, boosting productivity, uplifting data-driven resolutions, and safety and security enhancements. These technologies unlocked many solutions which were impossible before. In the field of education, machine learning and deep learning play a vital role in many forms. It provides adaptive learning mechanisms, predictive analytics with students' performances, outcomes and dropouts, automated grading systems, virtual teaching assistance, speech and learning models for accent training and real-time translation, plagiarism detection and content optimizations.

This study examines the gap in the existing literature and implements a novel approach for evaluating e-learning impacts on different cohorts of students by considering the advantages of the machine learning deep learning integration mechanism. The objective of this study is to combine the machine learning algorithms and neural networks to identify the patterns that map the input data to the impact values of e-learning systems. This approach recommends the data-driven improvements on e-learning platforms to improve the personalized contents and eliminate the dropouts, disengagements and more. It also helps the different levels of students to find the impact of e-learning. This will help them build their academics more efficiently and powerfully. As a result, aids both educators and developers to model e-learning systems based on the requirements of students.

II. Literature Review

The indoor classroom environment imposes boundaries on the well-being of the students. It could not focus on their physical health and comfort. It also needs to tackle their psychological and social well-being. Studies show that the academic performance of students can be assessed by their cognitive health [4]. The traditional grading system needs to be improved or replaced by modern approaches. Even though it has been practiced for a long time, it needs to be changed and adapted to the current trend and lifestyle. The traditional grading system comes with various limitations which affect the integrity and efficiency of students and it is depicted in Fig. 1. It deals with four major issues—scalability, inconsistent assessments, lack of individual feedback and subjective perception of students [5].

With the Information and Communication Technology, significant changes have occurred in educational systems. It is predominantly carried through virtual learning platforms [6]. Electronic learning (e-learning) has turned into one of the most prominent trends across the world, especially after COVID-19 [7]. It is offering an enhanced and user-friendly learning environment to the users. It provides personalized educational content and assessments to individual participants, while the traditional learning system is equipped with uniform educational content for all learners. In today's scenario, an Artificial Intelligence (AI) based customized learning environment is necessary. It is beneficial even for beginners to tailor their content according to their purposes and requirements [8]. The study says that it received positive responses from most of the participants regarding the e-learning system. E-learning provided the opportunity to study from home. As a result, the advantage of their own pace in studies was achieved. Another convenience is cost savings, which removes the expenses of transportation and accommodation. It also contains options to access the study materials and record the online classes [7]. E-learning brought several improvements in accessibility, customization and interaction by leveraging AI techniques. Integration of AI and e-learning delivers exceptional potential in creating a personalized and persuasive learning environment. The sustained influence of this e-learning system endorses inclusive and proficient learning [6].

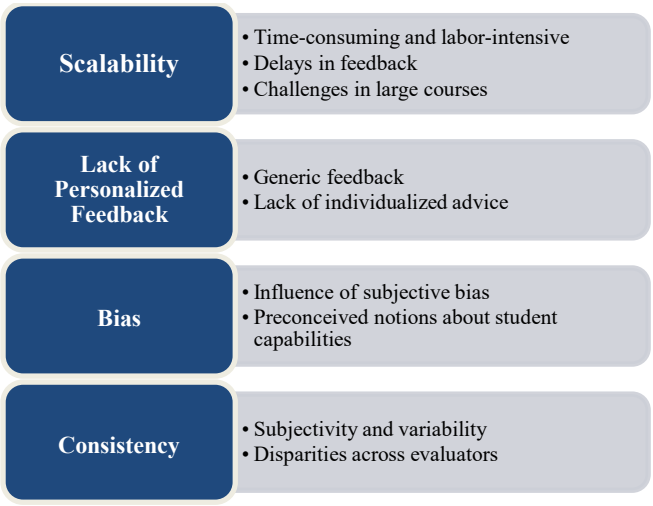

Fig. 1. Limitations of traditional grading systems.

A. Strength of E-learning

- E-learning allows broader accessibility regardless of the learner's geographical location and time.
- The flexibility achieved by adjusting the learner's own pace and Itinerary learning.
- Elimination of physical classrooms and transportation in e-learning makes the learning process cost-effective.

The variety of multimedia tools like videos, pictures, animations and simulations provides the interactive learning mechanism [9].

B. Weaknesses of E-learning

- The technical issue is that reliable internet and hardware tools may not be available for some learners.
- The e-learning process isolates the learners from social interactions.
- Subjects like science and engineering need a hands-on learning method, which is challenging to implement in

e-learning platforms.

- It creates fewer opportunities for collaboration and networking in terms of career development [9].

On top of everything, lack of self-discipline is the major concern in the student community due to digital distractions. This makes the e-learning methods less effective compared to the traditional classroom learning method [10].

C. Machine Learning and Deep Learning in E-learning Environment

Machine learning and deep learning leverage an innovative approach for data analysis over statistical methods. It enables the machine learning prediction systems working with the resampling techniques like cross-validation to prevent the educational sciences from the replication shortfall. Additionally, it enables the ML to handle multiple features and to design complex relationships [11].

Numerous studies on e-learning systems were developed through machine learning and deep learning techniques to enable the AI-driven approaches. The early identification of underperforming students and supporting them early is a challenging process in e-learning methods. The technological intervention is necessary to prevent and support the students who are falling behind. The clickstream data is used to monitor the student's online activities. It captures every click of the students who are involved in online learning platforms to deliver precise insights into their engagement with the program. It also detects the learning patterns which indicate the complexity level of the learning. The advanced machine learning approach was applied to clickstream data to predict the performance of the students. The traditional techniques and time-series models were used to construct a comprehensive analysis model. The traditional techniques—Logistic Regression, XGBoost, K-NN, Random Forest, and Naive Bayes and deep learning techniques—Long Short-Term Memory and Gated Recurrent Unit were implemented to design the novel approach. The Long Short-Term Memory neural network and Extreme Gradient Boosting algorithm were combined to build the hybrid model. The Gated Recurrent Unit (GRU) algorithm outperforms other models with above 90% of accuracy. This model acts as an early warning system to reduce the failures, dropouts and aids to improve the education quality [12]. The educational data mining approach was executed to analyze the student's outcomes on their test. The video clickstream data, such as forward, backwards and pause, exhibits the behaviour of the students during the online learning. The two consecutive experiments were performed with 22 and 16 university students. The first experiment was implemented with the fundamental statistical approaches, and the second experiment was performed with the advanced data mining techniques. The outcomes of these experiments show that the total number of clicks positively correlated with the student's test results [13]. In contrast, the Text Graph Convolutional Networks (Text-GCN) were used to study the learner's performances with video viewing behavior rather than click stream data. The participant's video viewing behavior is tested with the pedagogical video courses. It observes the clicks made for pedagogical sequences. The clicks and test scores were collected to analyze the performance of the participants (learners). The Text-GCN model achieved 67.23% accuracy [14].

Academic assessment's improvisation was demonstrated by the AI-driven models by addressing the current challenges in the process of attaining academic excellence. The developed AI tools can support dynamic and inclusive assessment frameworks. Still, it needs to focus on student motivation and behaviors to improve prediction metrics to add a more accessible and productive learning ecosystem [15]. The explainable machine learning model facilitates automatic and smart functional feedback and strategic suggestions for students during e-learning. The assignment and quiz level feedback and recommendations sent to the students enhance their academic outcomes. This model applies the student's data from the Learning Management System. The dashboard was developed in the real-time course environment to evaluate the results of the proposed system. This evaluation identifies the impacts and benefits of the dashboards to boost self-discipline and motivation towards the learning process. Furthermore, the collaboration of course tutors to deploy and estimate the dashboards can improve the efficiency of the model [16]. The computer systems can store a large amount of student information. This information was applied to machine learning algorithms to predict whether the student will succeed in the course or not. The purpose of this research is to implement the multi-classifier model, which extends the performance over other machine learning models. It took 25 characteristics of academic performance and applied them to Extra Tree (57.41%), Random Forest (61.96%), Decision Tree (91.44%), Ada Boost (59.65%) and XGBoost (83.3%) algorithms. The combination of these algorithms frames a multi-classifier model and produces 92.86% accuracy, which is higher than that of the other independent machine learning models [17].

The Educational Data Mining (EDM) technique captures the particular data types of the educational system to perceive the student's behavior as well as the system. This dataset was employed in XGBoost and Random Forest algorithms to detect the specific factors which steer the student achievement prediction [18]. The productivity of students in e-learning platforms is predicted through ensemble machine learning techniques. The keyboard and mouse events were observed to derive the results of this proposed system. As it deals with the binary and multi-class classification tasks, the Ensemble Learning-based XGBoost (EL-XGB) model was developed. This model is designed by an efficient modelling of the cross-validation algorithm. The EL-XGB model delivered higher accuracy than the existing systems [19]. The unique educational dataset was organized from a broad perspective to predict student achievements. The six traditional machine learning models, such as Random Forest, XGBoost, Lasso, Decision Tree, Elastic Net, and Support Vector Machine models, were implemented and compared in terms of their accuracy metrics. The XGBoost model dominates other models with a considerable growth in R2 by 92% [20].

The student performance prediction is required to eliminate the problem of dropouts in society. The prediction system tackles two issues regarding traditional machine learning analysis. It manages class imbalance and data complexity in the academic dataset. The imbalanced class is

handled with oversampling and undersampling techniques. The proposed system employed a decision tree ensemble model to classify the risk potential of the students. The classifier model's performance was examined by changing the ratios of oversampling and undersampling of data. The enhanced risk classification system was built using the decision tree ensemble method with a feature selection technique. This system achieved 0.849 as an F1-score [21]. It is also important to consider student's engagement in an online environment to confront its impact on efficiency, continuity and Perseverance on their online learning journey. It is essential in the higher education system. Still, the engagement level prediction is one of the most difficult tasks in the education field. The objective of this research is to develop an intelligent forecasting model for student engagement prediction using machine learning techniques. It provides feedback with three pre-defined classes—Not Engaged, Passively Engaged, and Actively Engaged to the students to boost their motivation and commitment towards e-learning. By comparing the results of different machine learning classifiers (Decision tree, SVM and ANN), the Artificial Neural Network classification model secured 85% accuracy. It alerts both students and teachers about the decreased engagement level of the students to identify the struggles in the middle of the learning and sends the motivation, course messages or schedule through e-mails [22].

The learning effectiveness can be indicated by multiple dimensions of the students. They are emotional, behavioral and cognitive intelligences. The student engagement on e-learning platforms explains their motivation, sustainability and outcomes. The goal is to design the predictive system to categorize the learners according to their engagement level with the help of student's academic results. It analyses the engagement levels in three categories of e-learning dataset: exams, quizzes and tutorials, which were derived from the Kaggle database. The exam and quiz scores were used to construct the predictive system for tutorial engagement. The decision tree approach attained 97% accuracy in the quiz and 98.49% accuracy in exam engagement. These produce the overall engagement level of the students in the e-learning process. This supports the execution of targeted intervention to refine the engagement of the students [23]. The integrated machine learning model was developed to predict the student's performances using the dataset, which was extracted from the electronic learning management systems. The ensemble methods bagging, boosting, stacking, and voting were applied to five conventional ML algorithms. The ensemble models performed better than the independent traditional models, Decision Tree, Random Forest, Naïve Bayes, KNN and Gradient Boosting. Still, the integration of five algorithms with the stacking technique accomplished the highest F1-score of 0.8195 [24].

This study explores the conduct of pupils in e-learning ecosystems. The Decision Tree, Support Vector Machine and Naive Bayes algorithms were implemented to find the best action plan to understand the student's communication in virtual learning. The implementation is done with the help of Jupyter Notebook using the Python language. The accuracy metrics, Relative Root Mean Squared Error, Root Mean Squared Error and Relative Absolute Error were analyzed to detect a well-performed machine learning system. The Support Vector Machine system produced RRSE of 0.02117, RMSE of 0.02714 and RAE of 0.00279, which is a greater performance when compared with other systems [25]. There is invisible collaboration in the student community to achieve the course certificate on MOOC platforms. The close submitters of the course are carrying out some unethical behavior for their certification [26]. The Large Language Models (LLMS) like ChatGPT offer both advantages and disadvantages in the field of higher education. The LLMs guide the students for their research, assessments, analysis and writing. However, it delivers the risk of unethical behaviors, inadequate or incorrect misinformation, leading to academic misconduct [27].

III. Machine Learning Algorithms

A. Random Forest Algorithm

The Random Forest algorithm is a supervised machine learning algorithm, and it implements a bagging approach to address the problems. The bagging is an ensemble method. It creates multiple models and trains them independently. Each independent trained model computes on different subsets of data simultaneously. The random forest algorithm generates several decision trees and integrates the output of each decision tree model with an averaging or voting technique. It reduces the dependency by random selection of features and subsets of data while securing better performance and mitigating the mistakes of overfitting. It handles both classification and regression tasks. A random forest regression model builds numerous decision trees, which produce individual predictions. The final output is calculated by the average of every decision tree.

$$\hat{Y}_{RF}(X) = \frac{1}{T}\sum_{k=1}^{T} \hat{y}_k(X)$$

where, $\hat{y}_k(X)$ is the k-th tree prediction for input X and T is the total number of decision trees. Number of decision trees, maximum depth of each tree, maximum features for each split and minimum samples to split a tree are considered as hyperparameters. It is an advanced ensemble learning technique to develop complex and non-linear relationships between input and output features [28].

B. XGBoost Algorithm

Extreme gradient boosting algorithm is one of the most influential and commonly used machine learning algorithms. It is a supervised machine learning algorithm. The XGBoost Algorithm is an enhanced boosting algorithm which builds several trees in sequential order. Each tree boosts the performance better than the previous one. This boosting process is achieved through the regularization and gradient descent techniques [29]. Gradient descent helps to estimate the errors of the previous decision tree. It also minimizes the loss function during the splitting of trees. It consumes a variety of hyperparameters. Such as the number of boosting, learning rate, tree depth, size of subsets for each tree, size of the features per tree, loss reduction function, lasso and ridge regularizations. The final prediction of the XGBoost

algorithm is derived from the sum of the new tree's predictions and the existing tree's predictions.

$$\hat{Y}_{XGB} = \hat{y}_{k-1} + \eta . f_k(X)$$

where, $\hat{y}_{k-1}$ is previous tree prediction, $f_k(X)$ is the k-th tree output and η is the learning rate to control the contribution of each tree. It supports early stopping to mitigate the training time, regularization reduces overfitting, and it also offers tree pruning to strengthen its objective. As a result of its parallel training, it achieves a fast and accurate model.

C. Light Gradient Boosting Machine

The light gradient boosting machine is an ensemble-driven supervised machine learning algorithm, and it works on a gradient boosting framework and exploits decision trees. It is a well-known algorithm for its efficiency, speed and outstanding result [30]. It can work with a large amount of data and attributes. It encourages leaf-wise growth. The highest loss leaf is selected to split, which grows deeper trees. The values of continuous attributes are aggregated into individual bins to speed up the training process while minimizing the memory consumption. LGBM regression algorithm trains the decision tree models in a parallel manner and also fixes the previous model's errors.

$$\hat{Y}_t(X) = \hat{y}_{t-1}(X) + \eta \sum_{l=1}^{L_t} \omega_{j,t} . I\left(X \in R_{j,t}\right)$$

where, $\hat{Y}_t(X)$ is prediction at iteration t, $\hat{y}_{t-1}(X)$ the previous iteration prediction η is the learning rate, $\sum_{l=1}^{L_t} \omega_{j,t} . I\left(X \in R_{j,t}\right)$ is the sum of leaf contributions. Since it is a gradient boosting framework, it consumes every hyperparameter of the XGBoost algorithm and adds a number of leaves as one more hyperparameter to control the complexity of the model.

D. CatBoost Algorithm

Categorical Boosting Algorithm is a type of gradient boosting framework and is specially designed for categorical datasets. It can address both classification and regression tasks. CatBoost Regression Algorithm analyses the categorical variables automatically by applying ordered target statistics. This technique works efficiently without data leakage. It can leverage the CPU and GPU for faster and more efficient predictions, particularly on high-volume categorical datasets. It supports the missing values naturally, and no need for imputation. It establishes the ordered boosting method, which processes the data with ordered permutations to protect the model's decision-making capability from future information. This key approach minimizes the overfitting to generalize efficiently on categorical variables. The CatBoost Regression Algorithm generates binary decision trees symmetrically and forms an ensemble of these decision trees sequentially. Each fresh tree fixes the errors of the previous tree.

$$\hat{Y}_{CAT} = \sum_{t=1}^{T} \alpha_t . f_t(X)$$

where, $f_t(X)$ is the prediction of decision tree t, T is the total number of decision trees and α_t is the weight of the tree t. It consumes every hyperparameter of other gradient boosting algorithms and also adds its own parameters. Categorical features to handle categorical datasets, bagging temperature to control the randomness of bagging, border count to regulate the continuous feature segmentation and feature border type to select the structure of continuous feature sorting [30].

E. MLP Neural Network

MLP Neural Network MLP is a deep learning model. It is one of the most commonly used neural networks to analyses regression-type problems. It absorbs the dependencies between input features and target attributes through multiple hidden layers [31]. During the training process, the MLP applies techniques like activation functions, forward propagation and back propagation to build a reliable and efficient model. The errors can be minimized by adjusting the weights of the links which connect two layers of the MLP. This weight adjustment improves the accuracy of the model [32].

$$\hat{Y} = \sum_{l=1}^{L} W_l \; h_l + b$$

where, W_l is hidden layers weight, h_l is hidden layers output and b is bias in output layer. It includes hyper parameters, hidden layer sizes to determine the number of layers and neurons in each layer, activation functions, solver to optimize the training, alpha for L2 regularization, learning rate to decide the learning strategy, initial learning rate, maximum iteration, training epochs, early stopping to stop the epochs once the validation drops.

F. Elastic Net (ELNET) Regression

The Elastic Net Regression technique combines the regularization techniques LASSO (L1) and Ridge regression (L2) to seek the advantages of both. In that case, the ELNET regression model can enhance attribute selection and multicollinearity successfully. This integration produces a cost-effective and interpretable model which keeps important predictors and increases the performance. The multicollinearity in L2 regularization retains only the crucial variables and diminishes the overfitting. This improves the generalization of the model's output. It is useful for high-dimensional datasets [33]. It encompasses the values for hyperparameters-regularization ratio (l1 ratio), overall regularization strength, intercept, normalization, maximum iterations, tolerance and feature updating strategy for the optimization process.

G. Regression Analysis

The Regression analysis method is familiar with the prediction of the continuous target variable. It often estimates the errors in the process of detecting the cause and effect relationships between the multivariate input features and the continuous output variable with the help of machine learning algorithms. This analysis is finalized by drawing a straight line on the instances of the data to minimize the differences between the actual and predicted variables. This difference measures the errors. It can be either a straight line/curve,

depending on the regression type. It is the most adoptable and impactful computational method for real-time applications. This deals with three performance metrics to detect the accuracy of the proposed system [34].

(1) Mean Absolute Error (MAE): MAE computes the average absolute differences between the original and predicted target values.

$$\frac{1}{n}\sum_{i=1}^{n}|a_i - p_i|$$

Lower the value, easy to interpret the performance.

(2) Mean Squared Error (MSE): MSE detects the average of squared differences between actual and predicted target values.

$$\frac{1}{n}\sum_{i=1}^{n}(a_i - p_i)^2$$

Though MSE gives larger errors more than MAE due to squaring, it is effective in detecting large prediction mistakes.

(3) R-squared / Coefficient of Determination (R^2): R^2 value indicates effectiveness of the system to identify the variability in the data.

$$1 - \frac{\sum_{i=1}^{n}(a_i - p_i)^2}{\sum_{i=1}^{n}(a_i - a_m)^2}$$

R^2 value 1 denotes better fit, 0 denotes model is not suitable and less than 0 means poorly fitted.

where,

- a_i = actual value
- p_i = predicted value
- a_m = mean of the actual values
- n = number of data points

H. Important Applications of Regression Analysis

(1) Forecasting the results of the observed data. (e.g., election results forecasting, sales outcome prediction, or health status prediction)

(2) Association analysis to detect the relationship between variables. (e.g., risk indicators, patterns, or social trends).

(3) Generalize the results from the samplings to the overall population for broader insights.

(4) Assess the impact of treatment by matching the outcomes of the treated and control groups. (e.g., performance of the new drugs or a new teaching strategy) [35].

IV. Proposed Methodology

A. Data Collection

The structured questionnaire was designed and shared via Google Forms to collect the data for this study. The main objective of this research is to observe the e-learning impacts on learners across various educational institutions and self-motivated learners. The response form includes three types of questions: closed-ended, multiple-choice questions, and check boxes. It was designed to obtain Demographic information and learning preferences to address the individual's learning preferences, Content engagement patterns to identify the interesting content patterns, Platform usage data for platform optimizations, motivational levels and perceived benefit to analyses the e-learning impacts of the learners. In aggregation of these data types, it assists the tutors, researchers and platform developers to make informed decisions assess the learner outcomes and elevate the educational standards. The data were obtained from 1141 participants from various universities, colleges, schools and aspirant learners. The demographic distribution of participants is shown in (Fig. 2).

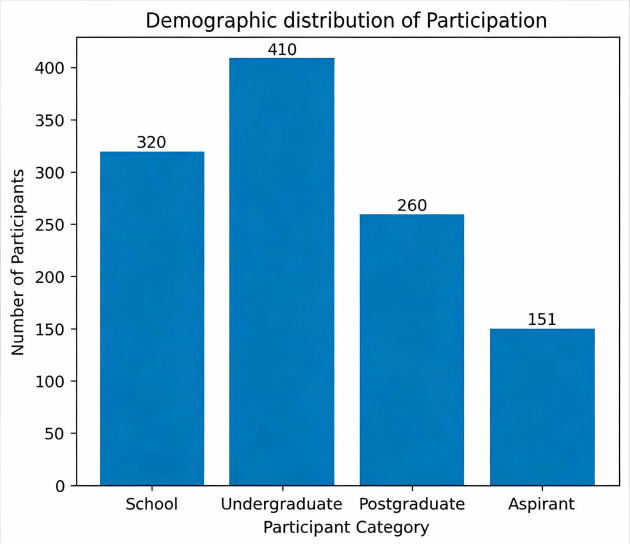

Fig. 2. Demographic distribution of participants.

B. Data Description

The data set includes various types of attributes such as demographic, behavioral, and usage patterns, learning preferences, platform usage, motivation and benefits. The Google form was distributed for three months to receive the student's responses, later anonymised to ensure privacy and confidentiality.

1) Demographic attributes

- Name: Open-text field.
- Age: Categorised into six sets (Below 18, 18–22, 23–27, 28–35, 36–45, and 46 and above).
- Educational Level: Participants selected from predefined categories (School, Undergraduate Level, Post Graduate Level and Aspirant).

2) Behavioural and usage patterns

- Types_of_content: Respondents could choose from options such as Video Lectures, Quizzes & Interactive Exercises, Assignments/Projects, Text-based Notes/PDFs, and Discussions & Forums.
- Hours_spent_weekly: Four time-range options were provided (Less than 2 h, 3–5 h, 5–10 h, and more than 10 h).
- Purpose_of_content: Participants indicated the main intent behind using E-learning platforms (Hobby, Academic, Professional, Self-development, and Aspirants).

3) Learning preferences

- Learning_preferences: Choices included preferred session lengths: Short modules (5–10 min), Medium sessions (30–45 min), and Long lectures (1+ h).

4) Platform usage

- E-learning_platforms: Respondents reported which e-learning platforms they use, including Coursera, edX,

Udemy, LinkedIn Learning, YouTube, Testbook, BYJU'S, and Other (with an option to specify).

5) Motivation level indicators

- Motivation_Level: It is to evaluate the motivational indicators of the learners. It indicates the levels of engagement and personal drive. These included factors like: Encourages me to explore more about the subject on my own; Helps me achieve my learning or career goals; Content keeps me engaged and interested; Excited to log in and Complete new lessons and The platform's structure and design make learning enjoyable.

6) Perceived benefit level

- Benefit_Level: This attribute is to analyse the perceived enhancements in knowledge and real-time applicability, such as: Alignment with personal/professional goals, Quality resource, Boost in confidence, Updated industry knowledge and Real-world applicability of acquired skills.

C. Data Preparation

The data preparation transforms the unprocessed or messy data into clean and consistent data. This phase involved loading the raw dataset, handling irrelevant attributes and missing values.

(1) Importing the Dataset: The dataset was ingested using the pandas library, which operates on large datasets and provides high performance in data manipulation.
(2) Quick Data Inspection: The shape and number of missing values of the dataset were retrieved to understand the overall size and its incomplete records.
(3) Dropping irrelevant attributes: The dataset encompasses 1141 rows and 10 columns. The column includes the names of the participants, which have been removed as they do not give any insight into the implementation.
(4) Handling Incomplete Data: It is found that 40 records have missing values. Those rows have been dropped to ensure the consistency of the dataset during training.

At this point of implementation, the dataset contains 1101 rows and 9 columns. It is guaranteed that the dataset will not show any variances or anomalies in the training phase.

D. Feature Engineering

Feature engineering is the process of creating and transforming the dataset into an understandable and meaningful format for modeling machine learning systems. The strong feature engineering highlights the significance of the features and their relationships. It enhances the performance of the model while speeding up the training process. This study focuses on three types of feature engineering techniques: label encoding, feature interactions, target variable creation and visualizing target variable distributions.

(1) Label Encoding: The categorical variables contain pre-defined classes. These pre-defined categories should be converted into numeric values to handle the non-numeric features effectively. This enables the machine learning algorithms to construct effective machine learning models. Categorical features such as age, occupation, content types, learning preferences, and e-learning platforms were transformed into numerical values using label encoding.
(2) Interaction Features: The complex relationships among the features should be captured and validated to improve the interpretability of the model. The interaction features were created by combining two categorical features. The interaction features are:

- Level_of_Education_x_Types_of_content
- Purpose_of_content_x_E-learning_platforms
- Types_of_content_x_E-learning_platforms
- Level_of_Education_x_E-learning_platforms
- Learning_preferences_x_E-learning_platform
- Hours_spent_weekly_x_E-learning_platform
- Age_x_E-learning_platforms,
- Age_x_Purpose_of_content and
- Purpose_of_content_x_Learning_preferences

(3) Target Variable Creation (E-learning Impact Score): The recent study demonstrated the conversion of multiple Likert scale responses (learners' perceptions and self-evaluations) into composite numerical measures [36]. The target variable construction is the main element of this feature engineering process. The target variable, E-learning Impact Score, is created by combining two attributes:

- Motivation_Level
- Benefit_Level

These two attributes give two different perceptions of the respondents, which comprise multiple-choice options. Each attribute contains five options, and each option holds 2 points. The average score of these selected options was assigned to the new target attribute: E-learning_Impact_Score. This procedure established a new type of dataset, which is ideal for supervised machine learning models such as regression. This process makes the dataset contain seven independent variables and one dependent variable.

(4) Target Variable Visualization: The target variable, E-learning_Impact_Score distribution, was plotted to analyse its shape and identify any asymmetry. This distribution chart (Fig. 3) shows the bell-curve, and it is in normal distribution.

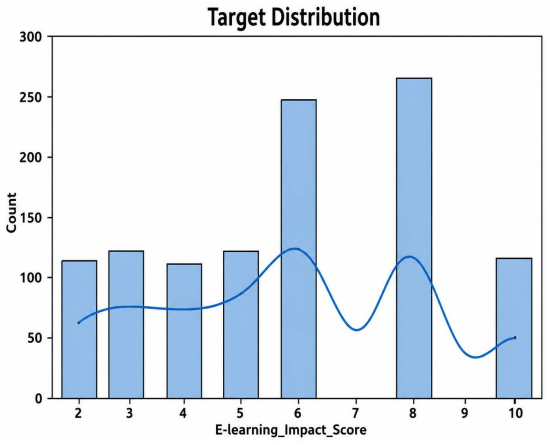

Fig. 3. Target variable distribution.

E. Model Training

The model training phase deals with the machine learning algorithms and the training dataset. It is also involved with the optimization and evaluation metrics. The machine learning model is developed with the help of machine learning algorithms by applying 80% of the dataset. These training datasets consume both independent and target

attributes. Various machine learning algorithms were considered to build the machine learning model, and the advanced ensemble model was discovered to provide higher accuracy.

V. Proposed System

The proposed system incorporates the applications of both deep learning and machine learning algorithms. A machine learning algorithm provides step-by-step processes to build a system. The machine learning system executes the data-driven decision-making approach by understanding the patterns in the dataset. The deep learning system comprises a variety of neural networks, often with multiple layers, to design deep learning models. Though Deep learning is a subfield of machine learning, it deals with a different approach while training the model. The deep learning model learns the patterns and features automatically, especially when it handles complex data like images, audio or text. This proposed system architecture (Fig. 4) consumes three different segments: Neural Network Structure, Hybrid Feature Engineering and Stacked Ensemble Model.

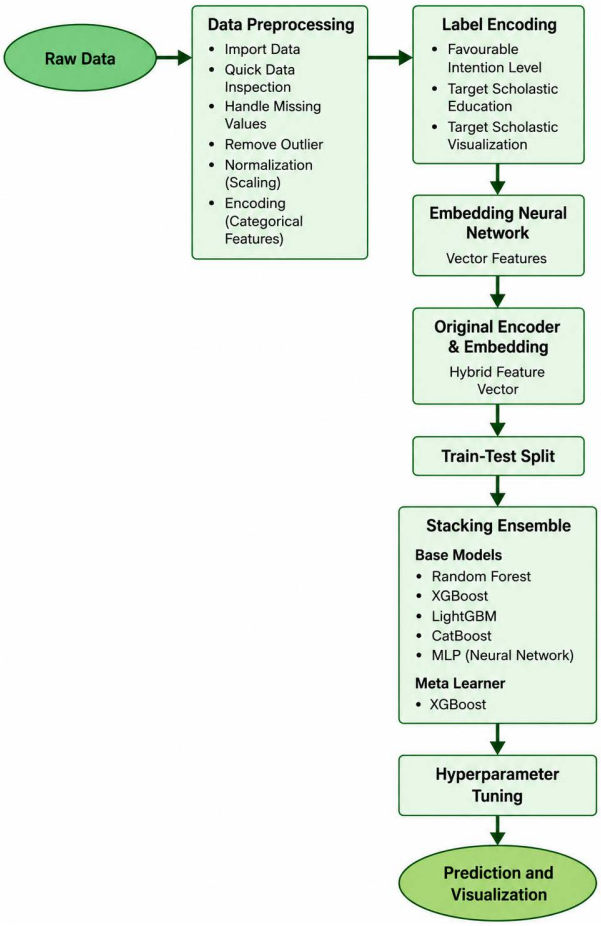

Fig. 4. Proposed system architecture.

A. Deep Embedding Neural Network

The deep embedding neural network is the first stage of implementation. In this stage, the categorical variables are converted into continuous vector representations. This is called dense embeddings, and it conveys the meaningful information. The closer vectors are allocated to the categories with similar semantics. It encompasses the understanding of relationships among the data better than the raw labels in categorical attributes. The neural network starts with the input layers. Each categorical variable holds the unique input layers. This enables the neural network to secure the individuality of these variables at the initial stage of development. Following that, the inputs are moved to the embedding layer, which holds the embedding dimension within 10. The embedding layer dynamically transforms the categorical input features i_c, into continuous vectors.

$$E(i_c) = W_c\,[i_c]$$

where, i_c is the integer encoded categorical variables and W_c is a trainable embedding matrix designated to the feature. This process allows training the neural network with compact representations. Depending on the categorical values involved in the prediction task, semantic equivalences are captured. The Ridge regularization is employed on the embedding weights, W_c to avoid the overfitting issues. Once the embedding process is completed, the operations are flattened and concatenations are executed to obtain the unified depiction of the feature Z .

$$Z = \cup\big(E(i_1), E(i_2), \ldots\ldots, E(i_k)\big)$$

The concatenated vector Z is sent to the deep feed-forward neural network which is meant to do complex and non-linear interactions among the features. The architecture of the Deep Embedding Neural Network consists of multiple dense layers, and every layer applies activation functions and regularization techniques. At the beginning of dense layer h_1, 128 units are activated through the ReLU function. Subsequently, batch normalization is applied to accelerate and normalize the training process. The overfitting is reduced by dropout the regularization. The next dense layer h_2 comprises 64 units and is enhanced with a skip network to boost the gradient flow and feature reusability. This hidden layer h_2 improves the training potential. The final dense layer with 32 units directs the result to the output layer y_p which holds a single neuron. The output y_p provides continuous vector representations of the categorical values. The forward propagation equations are illustrated here.

$$h_1 = ReLU\ (W_1\,Z + b_1)$$

$$h_2 = ReLU\ (W_2\ h_1 + b_2)$$

$$y_p = (W_1\ h_2 + b_3)$$

where, $ReLU$ is the activation function, W_1, W_2, W_3 are the weight matrices and b_1, b_2, b_3 are the biases. This neural network calculates the Mean Squared Error loss function L minimization by the differences between the actual target values y_a and predicted output values y_p.

$$L = \frac{1}{n}\sum_{i=1}^{n}\left(y_a - y_{p,a}\right)^2$$

The adaptive training potentials and efficient sparse gradient handling is achieved by the Adam optimizer with 0.001 learning rate. It also includes early stopping and

learning rate reduction techniques through callback functions. It allows the network to adjust the learning rate dynamically to enhance the integration.

B. Hybrid Feature Engineering

Hybrid feature engineering integrates the label-encoded categorical features with their corresponding neural embedding features to create hybrid features. It merges the traditional feature engineering with deep learned representations. This technique is effective for building detailed and predictive attributes. The embedding dense vectors E_{all} are created at the first stage of the proposed system using a deep embedding neural network.

$$E_{\text{all}} = \text{Embedding model.predict}(I)$$

The concatenation of these label-encoded categorical features I_c and their corresponding dense vectors E_{all} is ensued in this stage to acquire the hybrid feature sets X_{hybrid}.

$$X_{\text{hybrid}} = [I_c, E_{\text{all}}]$$

The hybrid feature matrix offers access to both the implicit and explicit structures of the dataset. It elevates the learning capability of the proposed model. The hybrid feature engineering fosters the trained representations while retaining the original variable values and builds multiple synergistic perspectives of the same feature. The augmented hybrid feature set is forwarded to the final stage of the proposed framework.

C. Splitting Training and Test Datasets

The training process enables the machine learning model to predict the outcomes for the unseen dataset. The testing process provides the accuracy of the trained model. The whole dataset was split into two parts: a training dataset and a testing dataset. The training data and testing data include 80% and 20% of the whole data, respectively.

D. Stacked Ensemble Model

The stacked ensemble model is designed with a 2-tiered architecture. The first tier includes five machine learning models to handle this regression problem. The Random Forest Regressor, XGB Regressor, LGBM Regressor, CatBoost Regressor and MLP Regressor act as the base learners. The second tier takes the ElasticNet regression model as the meta-learner. This stage utilizes the advantages of multiple machine learning algorithms to construct an effective e-learning impact prediction system. The base models, Random Forest Regression model creates 200 trees with a maximum depth of 15, XGB Regressor model forms 300 trees with a maximum depth of 6, LightGBM Regression model generates 300 trees with a maximum depth of 7, CatBoost Regressor model composes 300 trees with a depth of 6 and MLP is designed with two hidden layers with 64 and 32 units Individually. These base learners form the new feature set Z_{meta} ,

$$Z_{meta} = f_1(i), f_2(i), \dots\dots, f_m(i)$$

where, $f_1, f_2, \dots\dots, f_m$ denotes the base models. This new feature set Z_{meta} prediction is transferred to the second-tier meta-model, ElasticNet Regressor G. The final output $\hat{\boldsymbol{y}}$ is calculated as follows,

$$\hat{\boldsymbol{y}} = G(Z_{meta})$$

The ElasticNet regression model is preferred for its capability to stabilize Lasso and Ridge regularization. These regularizations support the high correlation and feature selection among the variables. The meta-model finds the vector of weight β to minimize the prediction error while blocking the overfitting through the integration of L1 and L2 regularizations. This process is ensured by the following equation:

$$min\,\beta \;\; \frac{1}{2n}\sum_{i=1}^{n}\left(y_i - Z_i^T\beta\right)^2 + \lambda(\alpha \parallel \beta \parallel_1 + (1-\alpha) \parallel \beta \parallel_2^2)$$

where, y_i is actual value Z_i is meta-model predictions and λ, β, α denoted as strength of regularization, model weights and the balance between Lasso and Ridge, respectively. This optimization process confirms that the proposed model trained with well-optimized base models by leveraging their distinct advantages and their combined range of strengths to predict the final output $\hat{\boldsymbol{y}}$.

E. Hyperparameter Tuning

Hyperparameter tuning is performed in the third stage of the proposed model. The optimization process is applied to the base models, and the meta-model is fine-tuned for its computations. The base learners—Light Gradient Boosting Machine Regressor, Extreme Gradient Boosting Regressor, Categorical Boosting Regressor, and Random Forest Regressor utilise GridSearchCV hyperparameter tuning and out-of-fold predictions processes. GridSearchCV finds all possible combinations of parameters and evaluates the performance of each to select the best one. This exhaustive search is accomplished automatically according to the performance metrics. The cross-validation reduces overfitting across the evaluation of multiple subsets of the data. The out-of-fold prediction is used to avoid data leakage and achieve consistent model performance. In this proposed model, the base models modify their pre-defined tree sizes, depth of the tree and learning rate to find the best combination that minimizes the Negative Mean Squared Error. The GridSearchCV estimate the performance metrics using 5-fold cross-validation for 5 equal parts of the dataset to acquire the best hyperparameters and a strong performance model. In the training process, four subsets of the dataset are utilized. The remaining subset is consumed to conduct the out-of-fold predictions. The results of out-of-fold predictions are kept in an array later; these predictions are provided to the meta-model as input. The meta-model, ElasticNet Regressor, is trained with the out-of-fold predictions. As it combines L1 and L2 regularization techniques, the hyperparameters alpha and L1-ratio are optimized. It enhances the regularization strength and balances the regularization L1 and L2. This optimization is fulfilled with the support of the net round GridSearchCV. Multiple values are applied to alpha and l1-ratio to reduce the Mean Squared Error and find the optimal parameters for accurate forecasts.

F. Computational Complexity

The proposed stacked ensemble model includes five base models and was tuned with 5 folds of GridSearchCV. The total training time complexity of this proposed system is:

$$O\left(\sum_{i=1}^{5} K \times P_i \times T_i\right)$$

where, $K = 5$, P_i represents number of hyperparameter combinations for the ith model and T_i represents training complexity of the model. The computation was dominated by the tree based models (Random Forest, XGBoost, LightGBM and CatBoost) with individual training complexity from $O(n \log n)$ to $O(n \times \text{max_depth})$. The meta-learner (Elastic net) adds linear complexity of $O(n \times d)$, where d represents number of input features (Here, $d = 8$).

The overall space complexity is defined as the sum of all base-model memory usage.

$$O(\text{Ensemble Trees})$$

The memory usage is dominated by the ensemble tree methods due to their multiple decision trees.

G. Model Testing

In this phase, the testing dataset is utilized to test the predictive accuracy of the trained machine learning model. The process of testing includes only independent variables to predict the corresponding target variables. The testing data includes 20% of the whole dataset. The evaluation metrics are defined based on the outcome of this testing process. It measures the generalization capability of the trained model.

H. Model Evaluation and Interpretation

Model Evaluation and Interpretation: The evaluation and interpretation of the E-Learning Impact Score Prediction Model were trained and tested to assess its performance. The prediction accuracy of this model is evaluated with the help of RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and R^2 (Coefficient of Determination) evaluation metrics. These performance metrics yield the interpretation of accuracy and its goodness of fit.

VI. Result and Discussion

The proposed system predicts the impact values of e-learning systems on a different set of students. The integration of machine learning and deep learning algorithms makes it possible to combine the strengths of both ML and DL techniques. The outcomes of this proposed system were evaluated with the accuracy metrics: Root Mean Squared Error, Mean Absolute Error, and Coefficient of Determination. The performance of the proposed system was better than the baseline models: Linear Regression, Random Forest, Gradient Boosting, XGBoost, LightGBM, CatBoost, and MLP. Fig. 5 shows the performance of the baseline models with the evaluation metrics - RMSE, MAE, and R^2. The negative R^2 values exhibit poor predictive performance. RMSE and MAE values are higher than the proposed system's values. This shows higher prediction errors.

The hybrid predictive model with a hyperparameter tuning process can predict the e-learning impact values with low error rates (RMSE and MAE) and a positive coefficient of determination (R^2 Score), which is higher than the model without hyperparameter tuning. The results were shown and compared in Figs. 6 and 7 and Table 1. It is shown that the predictive model with hyperparameter tuning gained 41.63% improvement in RMSE, 29.64% improvement in MAE and 43.08% improvement in Coefficient of Determination. The proposed system's R^2 score was nearly 1, which is adequate for a strong model.

Final Comparison Table

#	Model	RMSE	MAE	R2
0	Linear Regression	2.806443	2.397786	-0.022066
1	Random Forest	2.925145	2.517246	-0.112184
2	Gradient Boosting	2.895179	2.482877	-0.088731
3	XGBoost	3.384023	2.790510	-0.489747
4	LightGBM	3.069859	2.601407	-0.225376
5	CatBoost	3.166865	2.680945	-0.305208
6	MLP	2.923095	2.502284	-0.109139

Fig. 5. Baseline model result analysis.

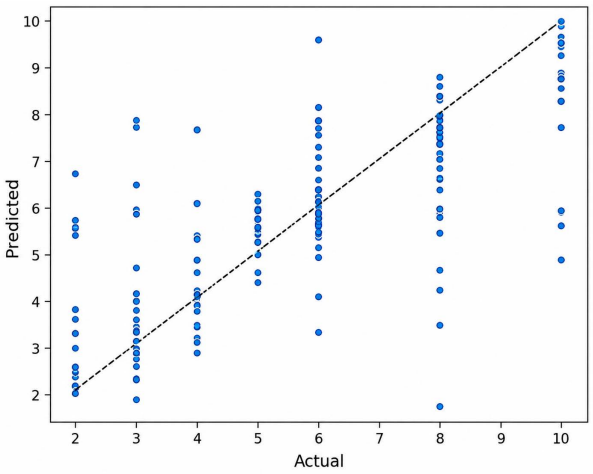

Fig. 6. The predictive model without hyperparameter tuning.

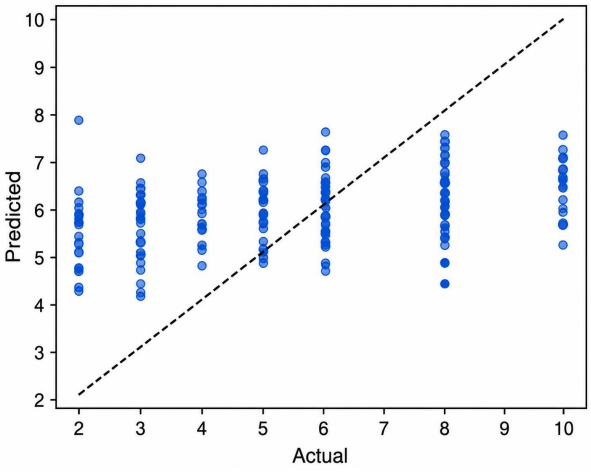

Fig. 7. The predictive model after hyperparameter tuning.

Table 1. Result comparison table

Accuracy/ variance metrics	Predictive model without hyper parameter tuning	Predictive model with hyper parameter tuning	Improvement
RMSE	1.5126	0.8831	**41.63%**
MAE	0.9719	0.6840	**29.64%**
R^2 Score	0.6056	0.8663	**43.08%**

As per the feature importance analysis report (Fig. 8), the feature E-learning_platforms contributed more than the other features, while the feature Learning_preferences contributed less to the prediction of the E-learning impact score. The interpretability of the proposed model was enhanced through the SHAP (SHapley Additive exPlanations) analysis, which calculates each feature's contribution to the final prediction, the E-learning_Impact_Score. The feature—e-learning platforms influence the outcomes highly. The other factors—age and type of content also show significant influence over the outcome. It indicates that both demographic and technical features play a crucial role in identifying the E-learning Impact Score. This demonstration provides a meaningful and practically interpretable model.

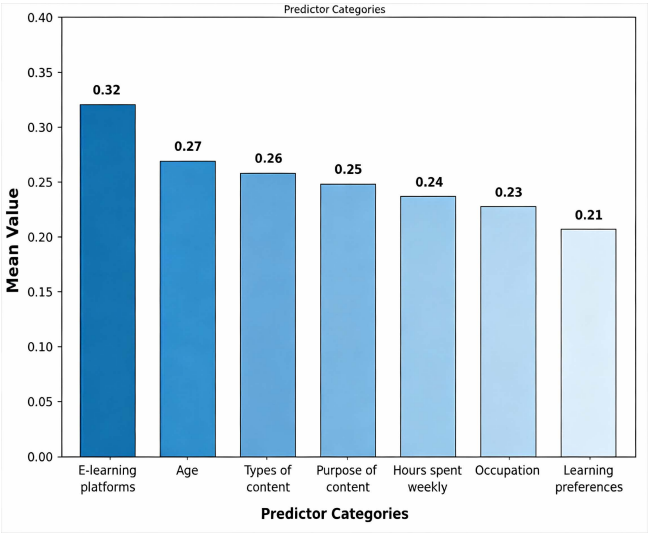

Fig. 8. Feature importance analysis.

The survey interpreted the analysis of learning motivation with several options (Fig. 9). These options-engagement level, structure of the course, excitement level, encouragement and career achievement- are used to detect the motivation level of the students. The analysis tells that the engagement and structure of the online courses need to be improved. Most of the students utilize these e-learning systems to achieve their career goals.

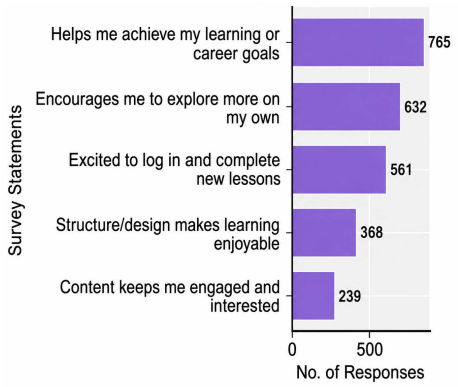

Fig. 9. Learning motivation analysis.

Fig. 10 shows the analysis of the benefits which were obtained by the students. These options describe the gained confidence in subjects, updating with the current trends, high-quality content, achieving personal/professional goals and skills for real-time applications. Most of the students obtained more knowledge on real-world applications and personal or professional goals. Even though the high-quality resources are useful, the e-learning platforms should be designed with motivation and up-to-date.

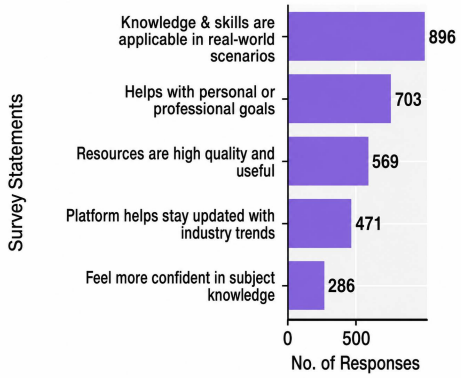

Fig. 10. Perceived benefit analysis.

The embedding-enhanced features from the embedded neural networks are a significant contribution to this system. The neural network with an embedding system extracted the complex relationships among the categorical features and produced vector input features about the categorical similarity. This enables the system to understand the patterns of the data easily, which boosts the training capability. The traditional encoding techniques were followed and combined with the vector input features as the training data. The stacking ensemble approach included deep learning and machine learning algorithms and was trained with the hybrid feature vector inputs. It also blends the tree-based model with neural networks and boosting techniques to reduce overfitting. The enhanced stacking model incorporates hyperparameter tuning and delivers a better and more robust prediction model.

VII. Conclusion

E-learning systems have become an important part of the student communities. The modern e-learning platforms provide the advantages of learning from anywhere and at any time. It helps the students to balance their lifestyle both financially and personally. It is cost-effective, time-saving, breaks spatial limitations, and is individual-focused. It is important to detect the challenges and behaviors of the students on these e-learning platforms. The e-learning platform developers, teachers, and students should address the impact of the values of e-learning. This study aims to analyze the impacts on the e-learning system among different groups of students. It leverages the advantages of both machine learning algorithms and deep neural networks to construct the proposed system. The dataset was collected from various schools, universities and individual aspirants. The survey questions were evaluated for the impact scores (scale of 10) based on their motivation and benefits. The other variables facilitate the training process of this model. The advanced hybrid model encompasses several techniques: data cleaning, encoding, vector input features generation, feature interaction creation, hybrid feature vector and stacking with ML and DL regression algorithms. This advanced hybrid model fuses the hyperparameter tuning techniques and predicts the impact scores with accuracy metrics, Root Mean Squared Error as 0.8831 and Mean Absolute Error as 0.6840.

The variance metric, Coefficient of Determination, is 0.8663. The R^2 value is nearly 1, and it is in an acceptable condition. The prediction of impact scores on e-learning systems helps the e-learning platform developers to improve their quality of content, interactive tools and the number of courses. The students can predict their outcomes, benefits and engagement levels to upgrade their knowledge and skills. The teachers can bridge their gaps with the student's community to find the difficulties, preferences, schedules and academic levels of their students. This study focused on the e-learning impact analysis of all groups of students with self-reported data. It may exhibit bias. It focuses only on e-learning behavior aspects rather than other influential factors such as academic outcomes or socio-economic levels. While providing the impacts of a diverse set of students, it would be better to focus on the specific groups with the same opportunities and awareness. It is also possible to convert the target variable into pre-defined classes to show the clear-cut impacts: high, moderate and low. The addition of a large amount of data will provide more accuracy without dealing with a biased target variable distribution.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Dr. A. Thilaka as a corresponding author conceptualized the article problem and she designed the hybrid deep ensemble framework first. And Dr. B. Mahalakshmi contributed to find out the methodology and the proposed model architecture and design. Dr. J. Devagnanam supported in this work to evaluate the proposed model. The important part of the study which is data collection and this task completed by S. Anitha Rajathi. She completed the preprocessing of the dataset also. Dr. T. Nathiya taken the relevant literature review for the study. E. Sundaravalli contributed a lot to the work, who helped us to create documentation and drafting of the manuscripts. All authors had approved the final version.

REFERENCES

[1] M. A. H. K. R. Bhai, "E-Learning: Types, tools, and techniques," *Knowledge Management in Higher Education Institutions*, vol. 14, 2022.

[2] S. K. M. Brika, K. Chergui, A. Algamdi, A. A. Musa, and R. Zouaghi, "E-learning research trends in higher education in light of COVID-19: a bibliometric analysis," *Frontiers in Psychology*, vol. 12, 762819, 2022.

[3] A. Vorina, N. Vrcelj, and V. Bevanda, "Usage of E-platforms Google Meet, Microsoft Teams and Zoom in Education," in *Proc. International Scientific Conference ITEMA*, 2022, vol. 6, pp. 63–69.

[4] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed., J. Peters, Ed. New York: McGraw-Hill, 1964, vol. 3, pp. 15–64.

[5] N. Makaremi, S. Yildirim, G. T. Morgan, M. F. Touchie, A. Jakubiec, and J. Robinson, "Impact of classroom environment on student wellbeing in higher education: Review and future directions," *Building and Environment*, 111958, 2024.

[6] A. R. Masalimova, M. R. Zheltukhina, O. V. Sergeeva, Z. M. Sizova, P. N. Novikov, and A. R. Sadykova, "Exploring higher education students' attitudes toward e-learning after COVID-19," *Contemporary Educational Technology*, vol. 16, no. 1, ep488, 2024.

[7] M. Imran, N. Almusharraf, S. Ahmed, and M. I. Mansoor, "Personalization of e-learning: Future trends, opportunities, and challenges," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 10, pp. 4–18, 2024.

[8] O. Fagbohun, N. P. Iduwe, M. Abdullahi, A. Ifaturoti, and O. M. Nwanna, "Beyond traditional assessment: Exploring the impact of large language models on grading practices," *Journal of Artificial Intelligence and Machine Learning & Data Science*, vol. 2, no. 1, pp. 1–8, 2024.

[9] A. Z. Sarnato, W. D. Sari, S. T. Rahmawati, R. Hidayat, and H. Patry, "The evolution of E-Learning platforms: From U-Learning to AI-Driven adaptive learning systems," *Journal of Social Science Utilizing Technology*, vol. 2, no. 2, pp. 289–300, 2024.

[10] M. Kamraju, J. Krishnaiah, G. Durgesham, N. Shaba, S. A. Begum, N. Fatima, and Y. Madhuri, "Exploring the impact of online education on higher education," *ASEAN Journal of Educational Research and Technology*, vol. 3, no. 1, pp. 27–36, 2024.

[11] S. Hilbert *et al.*, "Machine learning for the educational sciences," *Review of Education*, vol. 9, no. 3, e3310, 2021.

[12] Z. Khoudi, N. Hafidi, M. Nachaoui, and S. Lyaqini, "New approach to enhancing student performance prediction using machine learning techniques and clickstream data in virtual learning environments," *SN Computer Science*, vol. 6, no. 2, 139, 2025.

[13] O. R. Yürüm, T. Taşkaya-Temizel, and S. Yıldırım, "The use of video clickstream data to predict university students' test performance: A comprehensive educational data mining approach," *Education and Information Technologies*, vol. 28, no. 5, pp. 5209–5240, 2023.

[14] H. El Aouifi, M. El Hajji, Y. Es-Saady, and H. Douzi, "Predicting learner's performance through video viewing behavior analysis using graph convolutional networks," in *Proc. Fourth International Conference on Intelligent Computing in Data Sciences (ICDS)*, IEEE, 2020, pp. 1–6.

[15] M.-R. Borna, H. Saadat, A. T. Hojjati, and E. Akbari, "Analyzing click data with AI: Implications for student performance prediction and learning assessment," in *Frontiers in Education*, vol. 9, 1421479, Frontiers Media SA, 2024.

[16] M. Afzaal *et al.*, "Explainable AI for data-driven feedback and intelligent action recommendations to support students self-regulation," *Frontiers in Artificial Intelligence*, vol. 4, 723447, 2021.

[17] A. D. Vergaray, C. Guerra, N. Cervera, and E. Burgos, "Predicting academic performance using a multiclassification model: Case study," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, pp. 881–889, 2022.

[18] A. Khosravi and A. Azarnik, "Leveraging educational data mining: XGBoost and random forest for predicting student achievement," *International Journal of Data Science and Advanced Analytics*, vol. 6, no. 7, pp. 387–393, 2024.

[19] S. K. Siddaiah and M. S. Pm. (2022). Technique to Predict Student Performance Through Ensemble Learning Algorithm in E-Learning Environment. [Online]. Available: https://ssrn.com/abstract=4054489

[20] K. Yan, "Student performance prediction using XGBoost method from a macro perspective," in *Proc. 2nd International Conference on Computing and Data Science (CDS)*, IEEE, 2021, pp. 453–459.

[21] A. Ahmad, S. Ray, M. T. Khan, and A. Nawaz, "Student Performance Prediction with Decision Tree Ensembles and Feature Selection Techniques," *Journal of Information & Knowledge Management*, vol. 24, no. 2, 2550016, 2025.

[22] S. Ayouni, F. Hajjej, M. Maddeh, and S. Al-Otaibi, "A new ML-based approach to enhance student engagement in online environment," *PLOS ONE*, vol. 16, no. 11, e0258788, 2021.

[23] Y. Bellarhmouch, H. Majjate, A. Jeghal, H. Tairi, and N. Benjelloun, "Detecting Student Engagement in an Online Learning Environment Using a Machine Learning Algorithm," *Informatics*, vol. 12, no. 2, 44, 2025.

[24] F. Saleem, Z. Ullah, B. Fakieh, and F. Kateb, "Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning," *Mathematics*, vol. 9, no. 17, 2078, 2021.

[25] H. R. Ahamed and D. K. Hanirex, "Efficient Student Behaviour Analysis in E-Learning Using Data Mining Approaches," in *Explainable AI Applications for Human Behavior Analysis*, IGI Global, 2024, pp. 308–325.

[26] J. A. Ruipérez-Valiente, S. Joksimović, V. Kovanović, D. Gašević, P. J. Muñoz-Merino, and C. D. Kloos, "A data-driven method for the detection of close submitters in online learning environments," in *Proc. 26th International Conference on World Wide Web Companion*, 2017, pp. 361–368.

[27] T. Rasul *et al.*, "The role of ChatGPT in higher education: Benefits, challenges, and future research directions," *Journal of Applied Learning and Teaching*, vol. 6, no. 1, pp. 41–56, 2023.

[28] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest algorithm overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, 2024.

[29] J. Avanijaa, "Prediction of house price using xgboost regression algorithm," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 2, pp. 2151–2155, 2021.
[30] D. N. Muhammady, H. A. E. Nugraha, V. R. S. Nastiti, and C. S. K. Aditya, "Students final academic score prediction using boosting regression algorithms," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 1, pp. 154–165, 2024.
[31] J. Hwang, J. Lee, and K.-S. Lee, "A deep learning-based method for grip strength prediction: comparison of multilayer perceptron and polynomial regression approaches," *PLOS ONE*, vol. 16, no. 2, e0246870, 2021.
[32] T. Gerodimos *et al.*, "Artificial Intelligence Analysis of Macroscopic X-Ray Fluorescence Data: A Case Study of Nineteenth Century Icon," in *Proc. International Conference on the Theory of Machines and Mechanisms*, pp. 29–39, Springer, 2023.
[33] A. S. Al-Jawarneh, M. T. Ismail, and A. M. Awajan, "Elastic net regression and empirical mode decomposition for enhancing the accuracy of the model selection," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 6, no. 2, 564, 2021.
[34] K. Tyagi, C. Rane, and M. Manry, "Regression analysis," in *Artificial Intelligence and Machine Learning for EDGE Computing*, Academic Press, 2022, pp. 53–63.
[35] A. Gelman, J. Hill, and A. Vehtari, *Regression and Other Stories*, Cambridge University Press, 2021.
[36] O. Kayhan and Ö. Korkmaz, "Measuring 21st-century competencies: Development and validation of the information, media, and technology skills self-efficacy scale," *Information Development*, vol. 42, no. 4, pp. 2093–2110, 2026. doi: 10.1177/02666669241297661

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).