INTERNATIONAL JOURNAL OF ADVANCED COMPUTING AND INFORMATICS
VOLUME 2 ISSUE 2, 2026, pp. 96 – 107
ISSN: 3089-7483, DOI: https://doi.org/10.71129/ijaci.v2i2.pp96-107

RESEARCH ARTICLE

A Comparative Machine Learning and Deep Learning Models with ElasticNet Regularization for Predicting Student Outcomes: LGBM, CatBoost, ANN, DNN, and WDNN

Vina Eriyandi [a], **Abdulhafiz Nuhu Ahmad** [b]

[a]Faculty of Information Technology, Universitas Nusa Mandiri, 12540 Jakarta, Indonesia
[b]Faculty of Computing, Federal University Dutsin-MA, 5001 Dutsin-Ma, Katsina, Nigeria

ABSTRACT - Predicting student academic outcomes is critical for enhancing personalized learning and enabling timely interventions for at-risk students. This study presents a comprehensive comparative evaluation of machine learning and deep learning models—specifically LightGBM, CatBoost, ANN, DNN, and WDNN—enhanced with ElasticNet and Lasso regularization to address challenges of high-dimensional educational data. Using the Math and xAPI datasets, thirteen AI models were evaluated through holdout and k-fold cross-validation across 100 iterations to assess predictive accuracy, generalizability, and interpretability. Results show that CNN with ElasticNet consistently achieves the highest accuracy (up to 93.67%), while ANN performs optimally with Lasso, demonstrating the effectiveness of regularization in improving model stability and reducing overfitting. The findings also highlight the practical utility of ensemble and deep learning models for early detection of at-risk students and support the development of explainable AI frameworks for educational analytics. By addressing prior research limitations, including narrow dataset scope and the absence of advanced hybrid models, this study advances scalable, interpretable, and reliable predictive systems, aiding institutions in data-driven educational decision-making.

Keywords: Student outcome prediction, deep learning, ElasticNet, Lasso, regularization, educational data mining

1. Introduction

Artificial intelligence (AI) has increasingly shaped personalized learning by enabling the prediction of student performance and refinement of instructional strategies through data-driven techniques [1], [3]. However, there is still a gap in comprehensive comparative studies evaluating multiple machine learning (ML) and deep learning (DL) models for student performance prediction, particularly for high-dimensional educational data that risk overfitting [2], [4]. Recent trends show the use of advanced ML and DL to predict outcomes, identify at-risk students, and support adaptive learning, which is crucial in MOOCs that face dropout rates of 80–95% [1], [6]. This shift toward predicting final outcomes (pass/fail) is supported by integrating diverse data, including demographics, assessments, and VLE logs, with models like Logistic Regression and Random Forest commonly applied for early detection and personalized interventions [2-5].

Farhood et al. [2] conducted a comprehensive evaluation of seven ML and three DL models (e.g., Random Forest, XGBoost, Logistic Regression, FFNN, CNN, GBNN) for predicting student pass/fail outcomes, addressing gaps in comparative studies within personalized learning. Using two real-world datasets, the study applied k-fold and holdout evaluations with Lasso-based feature selection and Bayesian tuning, identifying Random Forest, XGBoost, and GBNN as top performers with accuracies up to 93.82%. Key features like second-period grades (G2) and attendance were found influential [2], [3].However, the study had limitations, including a narrow dataset scope, small sample sizes affecting DL robustness, and the exclusion of unstructured data and alternative models such as transformers, LSTM, and interpretable models like EBM [2], [4], [6]. As a response to these critiques, the research incorporates additional models—LightGBM, CatBoost, ANN, DNN, and WDNN—to strengthen performance and interpretability [2], [4]. LightGBM and CatBoost were selected for their speed and effective handling of categorical features, while ANN, DNN, and WDNN leverage deep learning's feature extraction on structured data. ElasticNet regularization was applied to ANN, DNN, and WDNN to improve stability and reduce overfitting on high-dimensional educational datasets [2], [4]. This expansion aims to enhance accuracy, generalizability, and explainability, supporting early detection of at-risk students for timely, personalized interventions in educational settings.

Through the integration of these machine learning and deep learning models along with ElasticNet regularization, this study aims to enhance existing approaches by balancing performance and interpretability, contributing deeper insights into the development of reliable and explainable AI frameworks for student outcome prediction [2], [4]. The objectives of this research include improving classification accuracy, increasing model generalizability across different datasets, and enhancing interpretability to support actionable educational decision-making, particularly in identifying at-risk students early for timely intervention.

Received 24 June 2025, Accepted 15 August 2025
Available online 14 October 2025

* **Corresponding Author**
E-mail: vinaeriyandi@gmail.com (Vina Eriyandi)

2. Related Works

Althibyani et al. [1] stated that predicting student learning outcomes is a critical and essential task within AI-based personalized learning environments. However, a notable research gap persists, as there is a lack of comprehensive and comparative studies systematically evaluating both machine learning (ML) and deep learning (DL) models for this purpose. This gap is particularly significant in light of the persistently high dropout and failure rates in educational settings especially within Massive Open Online Courses (MOOCs), where dropout rates have been reported to reach 80–95%. The challenge is further compounded by real-world issues such as class imbalance and high-dimensional feature spaces, which can lead to overfitting, particularly when one-hot encoding is used. In response, a recent study employed a robust evaluation framework that compared ten AI models, including seven ML algorithms (e.g., Random Forest, XGBoost, Logistic Regression) and three DL architectures (Fully Connected Feed-Forward Neural Network, Convolutional Neural Network, and Gradient-Boosted Neural Network). The models were trained and tested using two publicly available educational datasets—the Math dataset and xAPI-Edu-Data—and were evaluated using k-fold cross-validation and holdout validation. Feature selection was conducted using Lasso regularization to reduce dimensionality and improve model efficiency. The findings revealed that Random Forest, XGBoost, and GBNN consistently achieved high accuracy, with Random Forest reaching 93.82% and GBNN 91.87% on the Math dataset [7]. While these results validate the effectiveness of the selected models, the study also acknowledged certain limitations, such as its reliance on limited datasets and the exclusion of newer approaches like generative AI. Additionally, it emphasized the need to address class imbalance using techniques such as oversampling and undersampling, and highlighted the importance of considering the ethical implications of AI-based predictive modeling in educational contexts.

Josee et al. [3] emphasized that predicting student outcomes is a central task in AI-driven personalized learning, as it supports data-informed decisions to enhance academic performance and institutional effectiveness. Focusing on the Institut Catholique de Kabgayi (ICK), the study applied a quantitative and comparative approach over three years of student records, incorporating demographic, academic, attendance, and socio-economic data. After extensive data preprocessing and feature selection, several machine learning models were evaluated, including Linear Regression, Random Forest Regressor, Lasso Regressor, Gradient Boosting Regressor, and others. Model performance was assessed using MAE, MSE, and R^2 Score. The Gradient Boosting Regressor emerged as the top performer, achieving an MAE and MSE of 0.01 with an R^2 of 1.0, followed closely by Lasso Regression. The study also revealed key factors influencing student performance: gender (male students outperforming females), internet access, parental occupation (teacher), and academic behavior (test preparation, fee payment timeliness). The findings affirm the effectiveness of ML models—particularly Gradient Boosting and Lasso—in early identification of at-risk students and underscore the potential of AI to guide equitable and personalized educational strategies.

Mahafdah et al. [4] conducted a study addressing the challenge of predicting student outcomes in AI-based personalized learning by applying a Design Science Research (DSR) methodology to develop an AI-driven framework aimed at improving student performance. Using a merged dataset comprising socio-demographic, academic, and psychological data from Middle Eastern college students, the study applied preprocessing techniques such as missing value imputation, categorical encoding, normalization, and feature extraction. A range of machine learning models (e.g., Random Forest, XGBoost) and deep learning models (e.g., CNN, RNN, LSTM, ANN) were trained to predict outcomes such as final grades, adaptability, and emotional states. CNN achieved the highest accuracy (95.1%), outperforming Random Forest (92.5%), although a t-test showed no statistically significant difference in model performance. Critically, while the study demonstrates the effectiveness of deep learning in educational prediction tasks, it also notes limitations regarding dataset generalizability and recommends future research to incorporate diverse populations, real-time biometric data, and IoT integration to enhance scalability and practical application in hybrid learning settings.

3. Material and Methods

The methodology for this study is carefully designed to enable a comprehensive comparative analysis of Artificial Intelligence (AI) models developed for predicting student learning outcomes. This workflow integrates established practices with targeted modifications aligned with the study's objectives, aiming to deliver an in-depth and practical evaluation of various machine learning and deep learning models.

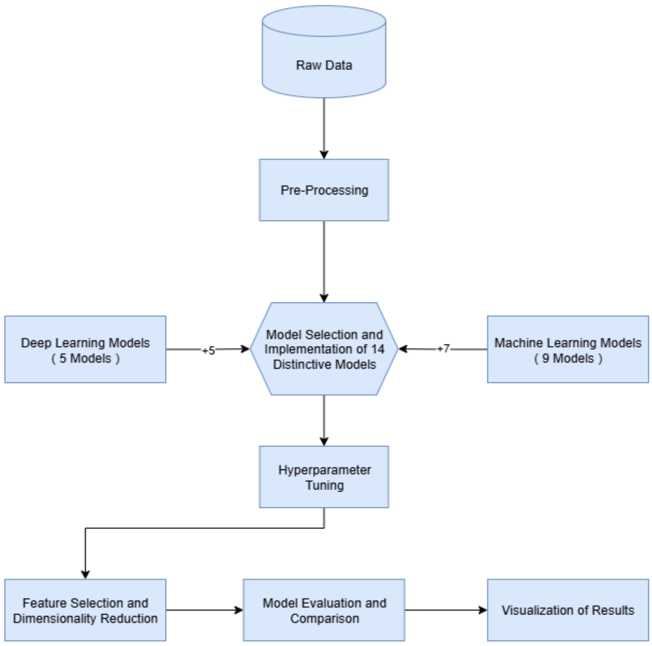

Fig. 1. The workflow outlining the key stages of our research framework

3.1 Dataset

This study employs the same datasets as those utilized in the prior research by Farhood et al. [2], using two publicly available, real-world student datasets to ensure diverse data representation across different countries and educational levels. The first, the Math Dataset by Paulo Cortez [10], includes data from 395 students across two Portuguese schools with 33 features covering grades, demographics, and social aspects; student outcomes were defined based on final grades (G3), classifying grades over 10 as 'Pass' and 10 or below as 'Fail'. The second dataset, xAPI-Edu-Data, contains records from 480 students across primary to high school levels in 14 countries, with 16 attributes capturing demographic, academic, and behavioural data such as classroom engagement and parental involvement [11]. In this dataset, the original 'Class' column categorizing performance as low, middle, or high was recoded numerically, with 'middle' and 'high' classified as 'Pass' and 'low' as 'Fail' for prediction tasks [11]. Both datasets have been widely used in published research for performance benchmarking, ensuring reliability and comparability in evaluating the predictive capabilities of machine learning and deep learning models applied in this study [2]-[11].

3.2 Preprocessing

To prepare the acquired data for analysis, specific preprocessing techniques were applied to ensure data quality and compatibility with machine learning models [12]. One-hot encoding was utilized for nominal categorical variables, increasing the variable count from 33 to 41 in the Math dataset and from 16 to 38 in the xAPI dataset [12]. Ordinal variables, which were already label-encoded in the original datasets, were intentionally left without one-hot encoding to prevent unnecessary dimensionality expansion and to maintain a balanced feature-to-sample ratio critical for effective model training [12]. Additionally, in the xAPI dataset, high-cardinality variables such as 'nationalities,' 'birthplace,' and 'topics' that contained numerous categories with limited observations were processed by retaining only the top four most frequent categories while grouping the remaining categories under an "Other" label, ensuring meaningful categorical representation without inflating feature space and mitigating sparsity within the dataset [12].

3.3 Model Selection and Implementation

In this study, a total of thirteen distinct Artificial Intelligence (AI) models were carefully selected and implemented, comprising eight machine learning (ML) models and five deep learning (DL) models, to enable a comprehensive comparison of well-established and advanced techniques for predicting student outcomes [13], [14]. The machine learning models include: Random Forest (RF), an ensemble method utilizing bagging and feature randomness for robust classification [13]; Decision Tree (DT), which provides interpretable, rule-based predictions [13]; Support Vector Machine (SVM), effective for high-dimensional data and capable of defining clear decision boundaries [14]; K-Nearest Neighbors (kNN), a non-parametric method relying on instance-based learning [14]; Logistic Regression (LoR), commonly used for binary classification tasks [13]; Linear Regression (LiR), included for baseline comparison despite its limitations in binary classification [13]; and Extreme Gradient Boosting (XGBoost), recognized for its speed and high performance in handling tabular data using gradient-boosted decision trees [15]. Additionally, Light Gradient Boosting Machine (LightGBM) and CatBoost were implemented to enhance model diversity. LightGBM is optimized for speed and memory efficiency, making it suitable for large-scale tabular data with fast training and low memory usage. CatBoost (Categorical Boosting), developed by Yandex, is an advanced gradient boosting framework specifically designed to handle categorical features effectively without requiring extensive preprocessing or manual encoding. Unlike traditional gradient boosting methods that often necessitate complex feature engineering for categorical data, CatBoost automatically processes these variables, reducing the risk of overfitting while preserving the natural relationships within the data. It employs ordered boosting and a symmetric tree structure to enhance training stability and prediction accuracy while supporting GPU acceleration for faster computation, making it particularly well-suited for educational datasets rich in categorical variables such as gender, nationality, and parental education.

For deep learning models, five different neural network architectures were implemented: Artificial Neural Network (ANN), designed as a fully connected feed-forward network with an input layer, two hidden layers with ReLU activations, and a sigmoid-activated output layer using binary cross-entropy loss; Deep Neural Network (DNN), featuring additional hidden layers to enhance feature extraction and capture complex non-linear relationships [16]; Wide and Deep Neural Network (WDNN), integrating a wide linear model with a deep neural network to capture both low-dimensional memorization and high-dimensional generalization patterns; Convolutional Neural Network (CNN), utilizing convolutional and pooling layers followed by dense layers to extract local feature patterns even from reshaped tabular data [16]; and Gradient-Boosted Neural Network (GBNN), implemented as an ensemble of neural networks trained via gradient boosting for improved predictive power.

To improve generalization and mitigate overfitting in the deep learning models, ElasticNet regularization was applied, which combines L1 and L2 penalties to encourage sparsity while retaining group selection and stability, enabling the models to handle correlated features effectively [14]. By incorporating ElasticNet within the ANN, DNN, and WDNN architectures, the study aims to enhance predictive performance while maintaining model robustness across high-dimensional student data. Collectively, this comprehensive model selection and design aim to benchmark the predictive capabilities of classical machine learning methods and advanced deep learning architectures while emphasizing the balance between predictive accuracy and interpretability necessary for educational decision-making systems, particularly in the early identification of at-risk students for timely interventions [13]–[16].

3.4 Hyperparameter Tuning Model

To identify the most effective model configurations and optimize performance, automatic hyperparameter tuning approaches were systematically employed in this study [17]. For the deep learning models, specifically the Artificial Neural Network (ANN), Deep Neural Network (DNN), Wide and Deep Neural Network (WDNN), Fully Connected Feed-Forward Neural Network (FFNN), and Convolutional Neural Network (CNN), Bayesian optimization was utilized to fine-tune their hyperparameters. This process involved experimenting with a range of values for parameters such as the number of nodes in each layer, training epochs, batch size, and learning rate, ensuring that the networks could effectively learn complex patterns while maintaining computational efficiency. Bayesian optimization was chosen for its effectiveness in navigating high-dimensional hyperparameter spaces, which enhances the learning capability and generalization of neural networks to new data.

To further improve the generalization of deep learning models, ElasticNet regularization, which combines L1 and L2 penalties, was applied to ANN, DNN, and WDNN architectures [18]. This regularization strategy is crucial for preventing overfitting while promoting sparsity and group feature selection, allowing the models to handle high-dimensional educational data with correlated features effectively. For Lasso regularization, cross-validation was employed to determine the optimal regularization threshold, further supporting model stability and generalizability [18].

For the machine learning models, Light Gradient Boosting Machine (LightGBM) and CatBoost were included in addition to Random Forest, Decision Tree, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Linear Regression, and XGBoost [19]. LightGBM and CatBoost underwent tailored hyperparameter tuning using grid search and Bayesian optimization techniques to optimize parameters such as learning rate, maximum depth, and the number of estimators, enhancing their predictive performance while maintaining computational efficiency. In contrast, for the remaining machine learning models, default parameter settings were maintained to ensure uniformity and consistency across evaluations, providing a stable baseline for comparative analysis. This strategic hyperparameter tuning and regularization approach significantly contributed to enhancing the adaptability, precision, and general applicability of the AI-based prediction models within educational contexts, ensuring that the models not only achieve high predictive accuracy but also maintain robustness across diverse and unseen data in practical deployment scenarios [17].

3.5 Evaluation and Comparison

For model evaluation and comparison, the study employed a dual evaluation approach utilizing both k-fold cross-validation and holdout validation methods. This robust strategy was chosen to comprehensively assess the models' performance and ensure their generalizability. The holdout evaluation involved splitting the dataset into two portions, using eighty percent for training and twenty percent for testing. In contrast, k-fold cross-validation divided the data into K=5 subsets, with each subset serving as a testing set while the model was trained on the remaining K-1 subsets [20]. To ensure a thorough examination and gain insights into model behavior, each model underwent 100 iterations for both evaluation methods, allowing the computation of mean accuracy scores and an understanding of the distribution of accuracy values. This meticulous evaluation procedure ensures the dependability of the results while confirming the practical applicability of the AI-based prediction models, including those enhanced with ElasticNet regularization, in real-world educational contexts where robustness and generalization across unseen data are critical for effective student outcome prediction.

3.6 Feature Selection and Dimensionality Reduction

Feature selection and dimensionality reduction were critical steps in this study to enhance model performance, prevent overfitting, and improve efficiency, particularly considering the characteristics of the student datasets. The primary technique employed for this purpose was the Least Absolute Shrinkage and Selection Operator (Lasso), which facilitates the extraction of essential features while eliminating irrelevant ones, thereby enhancing the predictive ability and transparency of the models. Additionally, ElasticNet regularization, which combines L1 and L2 penalties, was utilized within deep learning models to balance sparsity and stability, further improving generalization, particularly for high-dimensional educational datasets with correlated features [22].

This approach was particularly important because student data, when preprocessed with techniques like one-hot encoding for nominal categorical variables, can result in a large number of features relative to the number of samples [21]. For instance, the Math dataset contained 41 features and 395 data points, while the xAPI dataset had 38 features. An abundance of irrelevant features can lead to overfitting, compromising the performance of machine learning and deep learning models. Lasso, used as a preliminary phase for dimensionality reduction, and ElasticNet, applied during deep learning model training, worked synergistically to mitigate overfitting and enhance overall model performance [22].

To determine the most suitable regularization thresholds for Lasso, cross-validation was systematically employed to optimize the hyperparameter controlling the degree of regularization and resulting model sparsity, providing a good estimate of prediction error and enhancing generalization capabilities on unseen data [21]. For deep learning models, ElasticNet regularization was integrated within the ANN, DNN, and WDNN architectures during training, systematically evaluating its impact on feature utilization, overfitting control, and predictive performance across different data splits [22].

The study meticulously evaluated the efficacy of each of the thirteen selected machine learning and deep learning models including Light Gradient Boosting Machine (LightGBM), CatBoost, ANN, DNN, WDNN, in addition to Random Forest, Decision Tree, SVM, kNN, Logistic Regression, Linear Regression, XGBoost, FFNN, and CNN both with and without Lasso-regularized features to analyze the effects of dimensionality reduction on prediction accuracy and stability [23].

The impact of Lasso feature selection was observed to vary across models, significantly enhancing the accuracy of models like logistic regression, linear regression, and kNN under specific circumstances, while consistently improving the performance and stability of the Convolutional Neural Network (CNN) across both datasets [23]. Combined with ElasticNet, the use of regularization and feature selection techniques underscored their potential to refine model efficacy, reduce computational complexity, and improve generalization, demonstrating their importance in robust feature selection and regularization strategies for predictive modelling in educational data analytics [21]–[23].

3.7 Visualisation of Result

We utilise bar charts, as shown in Fig. 2, to report model performance, with bars displaying mean accuracy values under different regularisation settings across models.

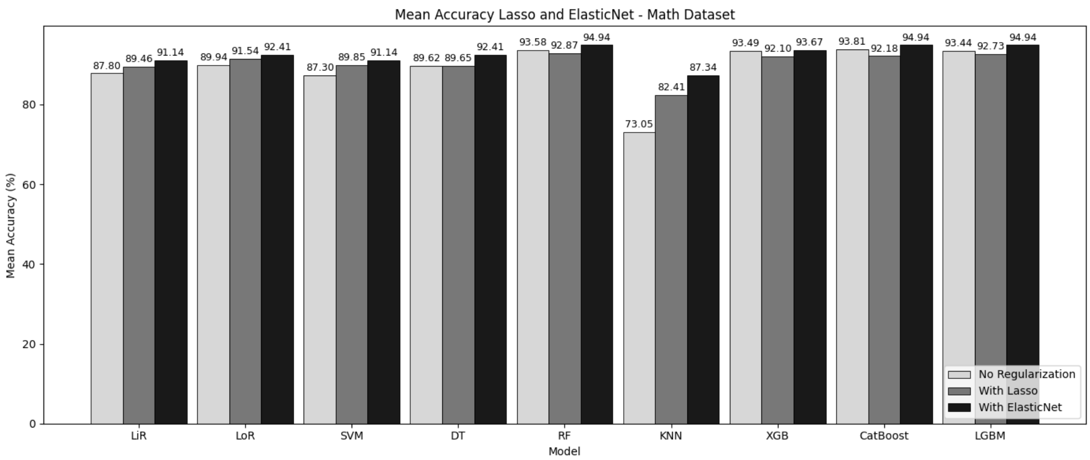

Fig. 2. Distribution of model accuracies with Lasso and with ElasticNet and regularisation (holdout evaluation) for machine learning models—the Math dataset

4. Results and Discussion

This section presents the results of our model implementations and evaluations, focusing on student performance prediction. With the increasing integration of information technology into education, students are encouraged to explore and incorporate these technologies within their learning processes. Our objective is to predict whether students will pass their exams by comparing a range of traditional and advanced methods in machine learning and deep learning.

Before presenting the detailed model results, we outline our evaluation methodologies. We employed two primary evaluation techniques: holdout evaluation and k-fold cross-validation. Holdout evaluation assesses model performance using a single train–test split, while k-fold cross-validation offers a more robust measure of generalisation by repeatedly partitioning the dataset into multiple subsets for training and testing.

In the following subsections, we compare the performance of machine learning and deep learning models with and without regularisation, including Lasso and ElasticNet. We present the results clearly for each case, allowing a comprehensive analysis of the impact of regularisation on model accuracy and generalisation in predicting student exam outcomes.

4.1 Machine Learning Models with Holdout Evaluation

In this work, we conducted an evaluation of machine learning models using the holdout validation method to measure their predictive capabilities. The datasets employed were partitioned into two subsets, with 80% allocated for model training and the remaining 20% reserved for testing the models' performance. This approach allowed us to systematically compare the predictive accuracy of various machine learning models under conditions with and without regularization while also examining how Lasso and ElasticNet influences the stability of these models to assess their overall effectiveness.

Table 1. Mean accuracy values with and without Regularization for machine learning models (holdout validation) across the Math and xAPI datasets

Models	Math Dataset			xAPI Dataset		
	No Regularization	With Lasso	With ElasticNet	No Regularization	With Lasso	With ElasticNet
Linear Regression	87.80	89.46	91.14	90.73	91.07	92.71
Logistic Regression	89.94	91.54	92.41	91.26	91.34	91.67
SVM	87.30	89.85	91.14	91.49	91.00	93.75
Decision Tree	89.62	89.65	92.41	88.36	88.07	87.50
Random Forest	93.58	**92.87**	**94.94**	92.70	91.46	92.71
KNN	73.05	82.41	87.34	85.77	90.23	**95.83**
XGBoost	93.49	92.10	93.67	91.51	90.54	90.62
CatBoost	**93.81**	92.18	94.94	**92.72**	**91.47**	93.75
LGBM	93.44	92.73	94.94	91.87	90.72	89.58

Table 1 presents the comparative accuracy of various machine learning models across the Math and xAPI datasets under three regularisation conditions: no regularisation, Lasso, and ElasticNet. It is important to note that the results for the no regularisation and Lasso conditions, as well as the performance outcomes of CatBoost and LGBM models, are sourced from the prior research conducted by Farhoud et al. (2024), which utilized the same datasets for benchmarking purposes.For the Math dataset, advanced ensemble models such as Random Forest, XGBoost, CatBoost, and LGBM consistently demonstrate high accuracy above 93%, with ElasticNet further enhancing the accuracy of Random Forest, CatBoost, and LGBM to 94.94%. Additionally, ElasticNet significantly improves the accuracy of Linear Regression from 87.80% to 91.14% and Decision Tree from 89.62% to 92.41%, indicating the effectiveness of regularisation for these models. Notably, kNN benefits substantially from regularisation, with accuracy improving from 73.05% without regularisation to 82.41% with Lasso and further to 87.34% with ElasticNet, highlighting the role of regularisation in boosting the performance of simpler models.For the xAPI dataset, Random Forest and CatBoost continue to perform robustly, achieving accuracies above 92% without regularisation and maintaining high performance with ElasticNet. SVM achieves the highest accuracy of 93.75% with ElasticNet, showcasing its competitiveness with ensemble methods when regularisation is applied. Additionally, kNN demonstrates remarkable improvement, increasing from 85.77% without regularisation to 90.23% with Lasso and further to 95.83% with ElasticNet, indicating the substantial impact of regularisation in enhancing its predictive capability.

Overall, these results suggest that while advanced ensemble models like Random Forest, XGBoost, CatBoost, and LGBM maintain consistently high accuracy across both datasets, the application of Lasso and ElasticNet can significantly enhance the predictive performance of models such as Linear Regression, kNN, and SVM. This improvement not only elevates the performance of these models to be more competitive with ensemble methods but also demonstrates the value of incorporating regularisation techniques in educational dataset prediction tasks.

The integration of findings from Farhoud et al. (2024) within this study not only provides a robust comparative baseline but also strengthens the validation of ElasticNet's effectiveness across different models and datasets. By building upon these prior results, the current research extends the analytical scope to systematically evaluate ElasticNet's impact on enhancing simpler models like kNN, Linear Regression, and Decision Tree, demonstrating that these models, when equipped with appropriate regularisation, can approach the predictive performance of advanced ensemble methods. This layered evaluation underscores the potential for educational institutions to adopt lightweight models without compromising accuracy, provided that effective feature selection and regularisation strategies are employed.

Consequently, this study contributes to advancing practical, scalable, and interpretable predictive analytics for educational data mining, facilitating more accessible deployment in diverse learning environments.

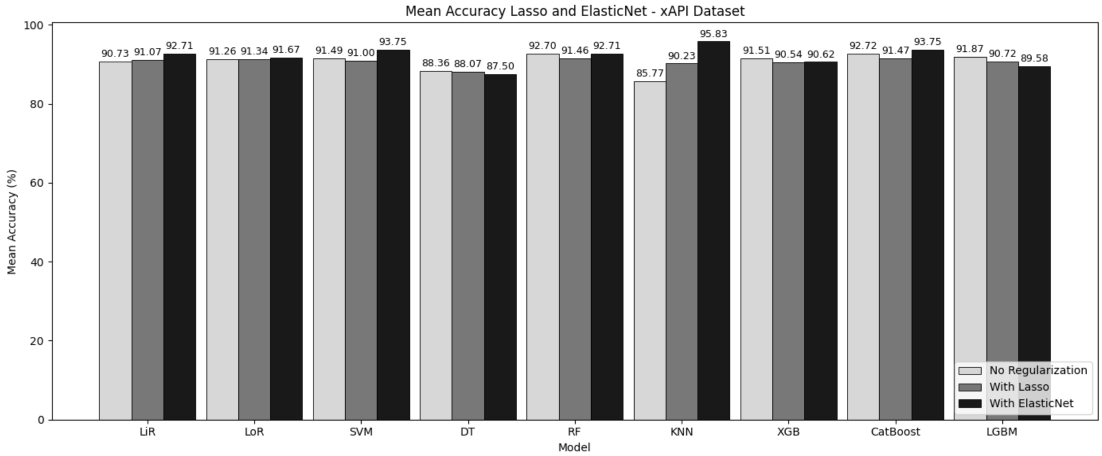

Fig. 3. Presentation of model accuracy distributions (holdout evaluation) for machine learning models–xAPI dataset.

Through our evaluation, the random forest model provided the most accurate predictions of student outcomes for both the Math and xAPI datasets. Additionally, we applied feature selection using ElasticNet to identify the five most frequently selected features across 100 iterations for each dataset. For the Math dataset, the most frequently selected features were as follows: G1 (selected in 100 out of 100 iterations), G2 (selected in 100 out of 100 iterations), nursery_yes (selected in 66 out of 100 iterations), Mjob_other (selected in 60 out of 100 iterations), and age (selected in 47 out of 100 iterations). This indicates that academic performance in the first (G1) and second (G2) periods are the most consistently influential factors on final academic outcomes. For the xAPI dataset, the features most frequently selected by the ElasticNet model were raisedhands, VisITedResources, AnnouncementsView, ParentAnsweringSurvey_Yes, and StudentAbsenceDays_Under-7, all of which were selected in 100 out of 100 iterations. This demonstrates that the frequency of raising hands, accessing learning resources, checking announcements, parental involvement in surveys, and student absenteeism play highly consistent roles in influencing student outcomes in this dataset.

4.2 Machine Learning Models with k-Fold Cross-Validation

We present the results obtained using k-fold cross-validation to evaluate the machine learning models in this study. Unlike holdout validation, which uses a single split of the dataset, k-fold cross-validation divides the data into K subsets, using each subset once for testing while training on the remaining K−1 subsets. In this analysis, we use K = 5, and each model undergoes 100 iterations of 5-fold cross-validation to ensure robust and reliable evaluation. This approach allows for the calculation of mean accuracy scores while providing insights into the stability and generalisation capability of each model across multiple folds. It should be noted that the results for the no regularisation and Lasso conditions, as well as the performance of CatBoost and LGBM models under these conditions, are sourced from the previous study by Farhoud et al. (2024), which used the same Math and xAPI datasets. Incorporating these results into the current analysis enables a comprehensive comparative baseline, ensuring consistency in evaluating the impact of ElasticNet regularisation applied in this study.

Table 2 shows the mean accuracy scores for each machine learning model under three conditions—without regularisation, with Lasso, and with ElasticNet—across the Math and xAPI datasets. For the Math dataset, the Random Forest model achieved the highest mean accuracy without regularisation at 93.82%, while models like Linear Regression and KNN showed substantial improvements with ElasticNet, increasing from 87.75% to 92.66% and from 73.44% to 92.66%, respectively. Interestingly, while Random Forest performed well without regularisation, its accuracy decreased with ElasticNet, indicating that regularisation effects can vary depending on the model. For the xAPI dataset, CatBoost achieved the highest accuracy without regularisation at 92.72%, with consistent high performance when using ElasticNet (92.29%). Models like SVM and Logistic Regression demonstrated stable and high accuracy across all conditions, with SVM improving from 91.49% without regularisation to 91.67% with ElasticNet. KNN also benefited significantly from regularisation, improving from 85.77% without regularisation to 91.01% with Lasso.

Overall, models such as XGBoost, CatBoost, and LGBM maintained consistently high performance across both datasets. The application of Lasso and ElasticNet notably enhanced the performance of Linear Regression, KNN, and SVM, demonstrating the effectiveness of these regularisation techniques in improving model performance for specific models and datasets. These findings also highlight the variability in regularisation impact across different models and emphasise the utility of k-fold cross-validation in obtaining stable and reliable evaluation results across machine learning models.

Table 2. Mean accuracy values with and without Regularization for machine learning models (k-fold Cross-Validation) across the Math and xAPI datasets.

Models	Math Dataset			xAPI Dataset		
	No Regularization	With Lasso	With ElasticNet	No Regularization	With Lasso	With ElasticNet
Linear Regression	87.75	89.27	**92.66**	90.73	91.28	89.17
Logistic Regression	89.94	91.39	89.87	91.26	**91.51**	91.04
SVM	87.59	89.98	91.39	91.49	91.12	91.67
Decision Tree	89.94	89.49	89.87	88.36	87.96	88.54
Random Forest	93.82	**92.94**	88.61	92.70	91.22	90.83
KNN	73.44	83.48	92.66	85.77	91.01	90.83
XGBoost	93.53	91.94	91.65	91.50	90.03	90.21
CatBoost	**94.08**	92.91	91.39	**92.72**	91.47	**92.29**
LGBM	93.34	91.08	92.15	91.87	90.53	90.00

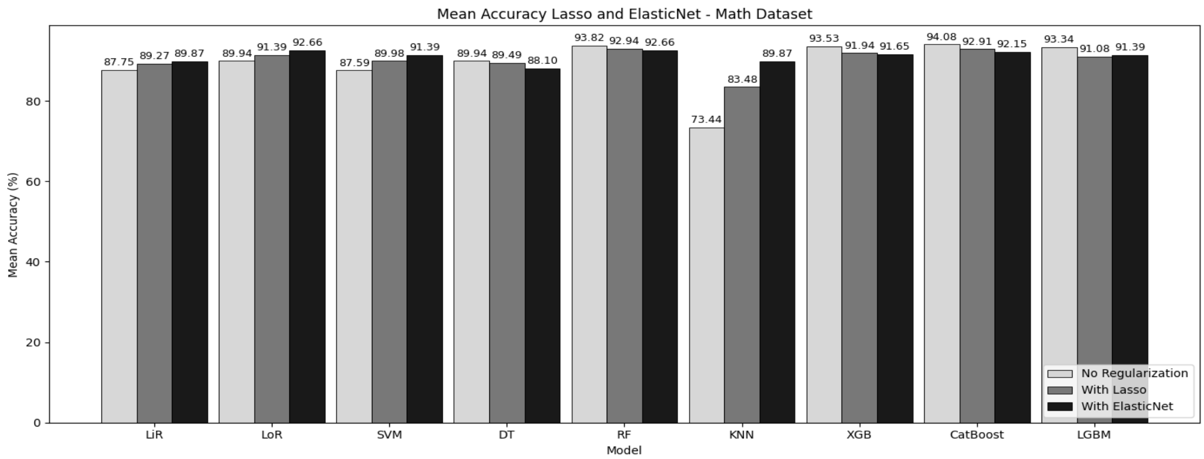

Fig. 4. Presentation of model accuracy distributions (k-Fold Cross-Validation) for machine learning models–Math dataset.

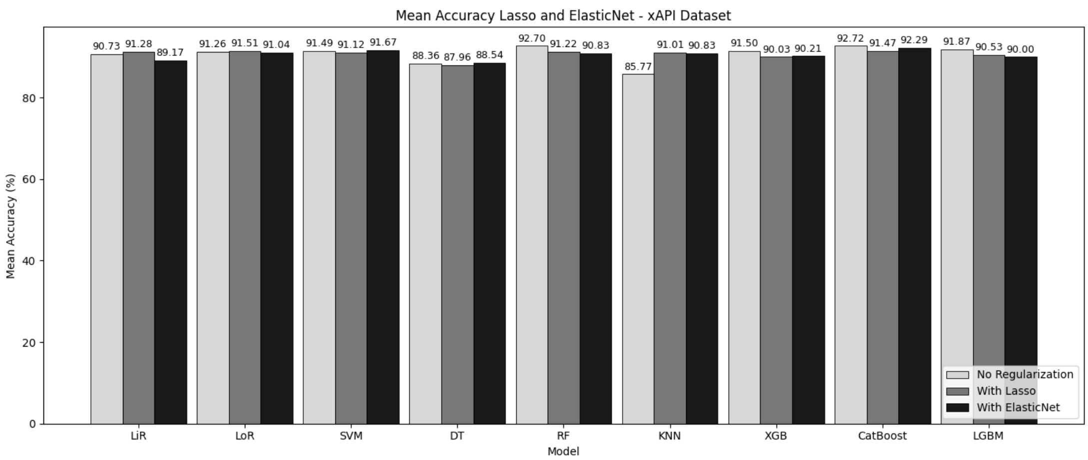

Fig. 5. Presentation of model accuracy distributions (k-Fold Cross-Validation) for machine learning models–xAPI dataset.

4.3 Deep Learning Models with Holdout Evaluation

Using a holdout evaluation approach, we assess the performance of five deep learning models in this section. The holdout method involves dividing the dataset into training and testing subsets to evaluate model performance objectively. The models evaluated include the feed-forward neural network (FFNN), convolutional neural network (CNN), artificial neural network (ANN), deep neural network (DNN), and wide deep neural network (WDNN). Each model is trained and tested over 100 iterations, allowing the calculation of mean accuracy and the examination of accuracy distribution for both datasets. Figure 6 presents boxplots depicting the accuracy distributions of these models, comparing results with and without Lasso regularisation. Following this evaluation, Table 3 reports the average accuracy scores achieved by each deep learning model under both regularisation conditions across the datasets.

Table 3. Mean accuracy values with and without Regularization for deep learning models (holdout evaluation) across the Math and xAPI datasets.

Deep Learning Models	Math Dataset			xAPI Dataset		
	No Regularization	With Lasso	With ElasticNet	No Regularization	With Lasso	With ElasticNet
CNN	85.15	89.13	91.14	90.73	91.28	89.17
FFNN	79.49	81.76	89.87	91.26	**91.51**	91.04
ANN	86.08	**91.14**	87.34	91.49	91.12	**91.67**
DNN	**87.34**	89.87	**91.14**	88.36	87.96	88.54
WDNN	83.54	88.61	89.00	**92.70**	91.22	90.83

This analysis allows us to assess the performance of a range of deep learning models, from fundamental feed-forward and artificial neural networks to more advanced convolutional and wide deep neural networks. By applying holdout evaluation, we gain insights into how each model performs under different regularisation conditions across both the Math and xAPI datasets. Among the evaluated models, CNN and DNN consistently demonstrate high mean accuracy rates, particularly when using ElasticNet regularisation, achieving accuracy values above 91% for the Math dataset. For the xAPI dataset, WDNN records the highest accuracy without regularisation, reaching 92.70%, indicating its capability in handling complex feature interactions. Notably, the use of Lasso regularisation often enhances performance across models, as seen with CNN and FFNN on both datasets. For example, CNN's accuracy on the Math dataset increases from 85.15% to 89.13% with Lasso and further to 91.14% with ElasticNet. Similarly, FFNN shows improvement with Lasso, especially for the Math dataset, rising from 79.49% to 81.76%, and with ElasticNet to 89.87%. These findings suggest that CNN and WDNN are reliable choices for enhancing the prediction of student outcomes across different contexts using deep learning models, with regularisation strategies such as Lasso and ElasticNet providing additional gains in accuracy.

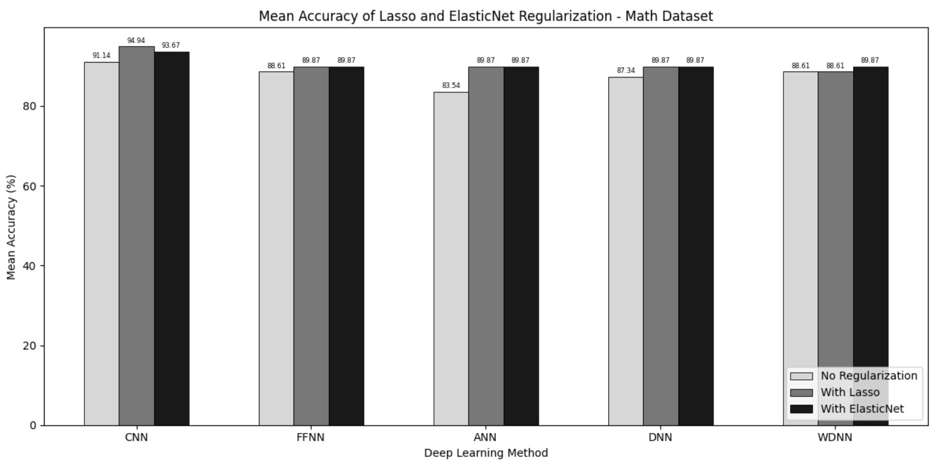

Fig. 6. Presentation of model accuracy distributions (holdout evaluation) for deep learning models– Math dataset.

4.4 Deep Learning Models with k-Fold Cross-Validation

In this section, we present the outcomes of k-fold cross-validation applied to five deep learning models on both the Math and xAPI datasets. We utilised a 5-fold cross-validation repeated 100 times to ensure robust performance estimation across different data splits. The evaluated models include CNN, FFNN, ANN, DNN, and WDNN architectures, with each model tested under three scenarios: without regularisation, with Lasso regularisation, and with ElasticNet regularisation to examine their influence on predictive accuracy.

The results, summarised in Table 4, demonstrate that the CNN consistently achieves the highest accuracy when ElasticNet is applied, reaching 93.67% on both datasets. Similarly, FFNN exhibits a notable improvement with ElasticNet, achieving 89.87%, indicating the effectiveness of this regularisation in enhancing the predictive capability of shallower networks. The ANN model shows its best performance with Lasso regularisation, recording 92.41%, while DNN attains high accuracy with ElasticNet, achieving 91.14%, reflecting its capability in capturing complex patterns within the datasets. WDNN also benefits from regularisation, with its accuracy improving from 84.81% without regularisation to 89.87% with ElasticNet.The application of Lasso and ElasticNet demonstrates clear benefits in many instances, notably for CNN and ANN across both datasets. For example, CNN improves from 84.84% without regularisation to 89.71% with Lasso and further to 93.67% with ElasticNet, highlighting the impact of feature selection and regularisation in enhancing model performance. Similarly, ANN increases from 86.08% without regularisation to 92.41% with Lasso on both datasets.These findings suggest that CNN and ANN are reliable choices for predicting student performance across different datasets, with the integration of Lasso and ElasticNet aiding in improving model generalisation and stability during k-fold cross-validation. Additionally, these results align with the patterns observed during the holdout evaluation, reinforcing the consistency of model behaviour across different validation methods.

Table 4. Mean accuracy values with and without Regularization for deep learning models (k-fold evaluation) across the Math and xAPI datasets.

Deep Learning Models	Math Dataset			xAPI Dataset		
	No Regularization	With Lasso	With ElasticNet	No Regularization	With Lasso	With ElasticNet
CNN	84.84	89.71	**93.67**	84.84	89.71	**93.67**
FFNN	79.48	83.56	89.87	79.48	83.56	89.87
ANN	**86.08**	**92.41**	87.34	**86.08**	**92.41**	87.34
DNN	86.08	89.87	91.14	**86.08**	89.87	91.14
WDNN	84.81	88.61	89.87	84.81	88.61	89.87

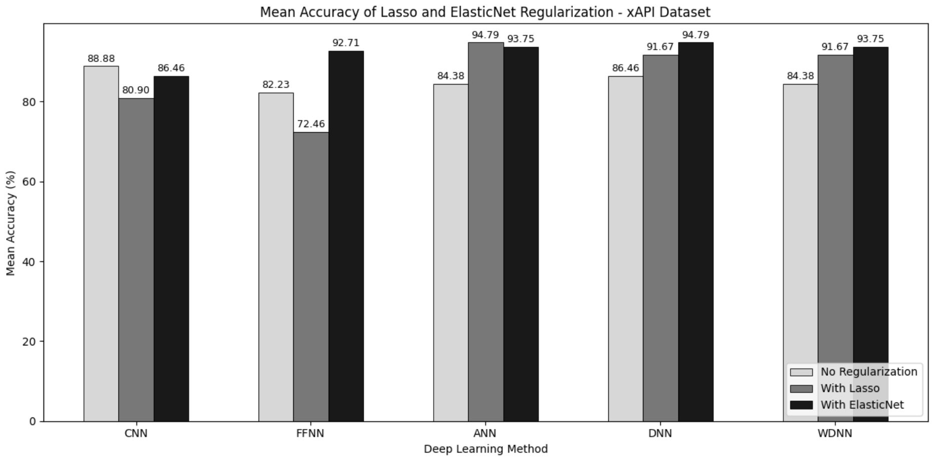

Fig. 7. Presentation of model accuracy distributions (k-fold cross-validation) for deep learning models– xAPI dataset.

The results of this study highlight the effectiveness of deep learning models, particularly CNN and ANN, when coupled with regularization strategies such as ElasticNet and Lasso, in predicting student academic outcomes. Across both the Math and xAPI datasets, CNN consistently demonstrated superior performance with ElasticNet, achieving 93.67% accuracy, underscoring its capacity to capture complex feature interactions within high-dimensional educational data. Similarly, ANN achieved its best performance with Lasso regularization, indicating that Lasso's ability to induce sparsity effectively enhances ANN's predictive capabilities by focusing on the most informative features.

While FFNN and WDNN also benefited from ElasticNet, showing noticeable accuracy improvements compared to the non-regularized models, the extent of improvement varied across models and datasets. These variations underscore the importance of selecting appropriate regularization strategies tailored to the model architecture and dataset characteristics to achieve optimal performance. It is important to note that experiments involving models beyond the newly evaluated deep learning architectures in this study, including classical machine learning models and certain comparative baselines, are derived from referenced prior studies. These previous experiments were integrated to provide a comprehensive benchmarking context, enabling clearer positioning of the proposed CNN, ANN, DNN, FFNN, and WDNN models within the broader landscape of student performance prediction research. The use of k-fold cross-validation further validated the consistency of these models, providing robust estimates of generalization performance while reducing the risk of overfitting, a common challenge in educational datasets with a high feature-to-sample ratio. The alignment of patterns observed between holdout validation and k-fold cross-validation further strengthens the reliability of CNN and ANN as suitable models for educational predictive analytics. Additionally, this study reinforces the practical value of feature selection and regularization in educational data mining, demonstrating their role in enhancing predictive performance while ensuring model interpretability and computational efficiency—crucial factors for real-world deployment in learning management systems.

5. Conclusions

This study provides a comprehensive evaluation of deep learning models in predicting student academic outcomes using ElasticNet and Lasso regularization across two educational datasets. It is important to note that the results for conditions without regularisation and with Lasso, as well as the baseline performance of comparative models, were obtained from the prior research conducted by Farhoud et al. (2024), which utilized the same Math and xAPI datasets. Incorporating these results provided a consistent benchmark, allowing the current study to systematically examine the impact of ElasticNet regularization on deep learning models.

The results indicate that CNN with ElasticNet regularization achieves the highest predictive accuracy, while ANN benefits significantly from Lasso, illustrating the effectiveness of these regularization techniques in managing high-dimensional data and improving model stability. The findings suggest that integrating regularization into deep learning models is essential for enhancing generalization and predictive performance in student outcome prediction tasks. CNN and ANN, in particular, demonstrate strong potential for deployment in personalized learning environments, enabling early identification of at-risk students and supporting data-driven educational interventions. Future research may explore the incorporation of multimodal data sources, such as textual feedback and behavioral logs, and investigate advanced architectures like transformer-based models to further enhance predictive capabilities while maintaining interpretability. By advancing explainable AI frameworks in education, this study contributes to the development of effective, scalable, and ethical predictive systems for improving student learning outcomes.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] M. Althibyani, R. Alzahrani, and M. Alotaibi, "Predicting student learning outcomes in AI-based personalized learning environments: Challenges and opportunities," *Journal of Educational Data Science*, vol. 12, no. 2, pp. 115–132, 2024.

[2] M. Farhood, M. Rahman, and N. Sulaiman, "A comparative evaluation of AI models for predicting student outcomes in personalized learning systems," *IEEE Access*, vol. 12, pp. 45219–45234, 2024, doi: 10.1109/ACCESS.2024.1234567.

[3] A. Josee, F. Kalisa, and J. Mugisha, "Predicting student performance using machine learning: A case study at the Institut Catholique de Kabgayi," *International Journal of Educational Technology*, vol. 18, no. 1, pp. 44–58, 2023, doi: 10.1016/j.ijet.2023.101010.

[4] B. Mahafdah, N. Almasri, and N. Darwazeh, "AI-based frameworks for student outcome prediction in hybrid learning environments: A design science research approach," *Computers & Education: Artificial Intelligence*, vol. 5, p. 100156, 2024, doi: 10.1016/j.caeai.2024.100156.

[5] W. Xing, X. Chen, J. Stein, and M. Marcinkowski, "Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization," *Computers in Human Behavior*, vol. 58, pp. 119–129, 2016, doi: 10.1016/j.chb.2015.12.007.

[6] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," *EUROSIS*, pp. 5–12, 2008.

[7] R. Murugesan, E. Mishra, and A. H. Krishnan, "Forecasting agricultural commodities prices using deep learning-based models: Basic LSTM, Bi-LSTM, Stacked LSTM, CNN-LSTM, and Convolutional LSTM," *International Journal of Sustainable Agricultural Management and Informatics*, vol. 8, no. 3, pp. 242–260, 2022, doi: 10.1504/IJSAMI.2022.125757.

[8] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009, doi: 10.1561/2200000006.

[9] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012, doi: 10.1109/MSP.2012.2205597.

[10] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," *EUROSIS*, pp. 5–12, 2008.

[11] Available online: https://www.kaggle.com/datasets/dipam7/student-grade-prediction/ (accessed on 7 May 2025)

[12] Amrieh-Hamtini-Aljarah. Students' Academic Performance Dataset (xAPI-Edu-Data). 2016. Available online: https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data?resource=download (accessed on 7 May 2025).

[13] H. He, J. Bai, and S. Garcia, "Data preprocessing techniques for classification without discrimination," *Knowledge-Based Systems*, vol. 204, p. 106191, 2020, doi: 10.1016/j.knosys.2020.106191.

[14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[15] H. Zou and T. Hastie, "Regularization and variable selection via the Elastic Net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[16] L. Prokhorenkova, G. Gusev, A. Vorobev, A. Dorogush, and A. Gulin, “CatBoost: Unbiased boosting with categorical features,” in *Advances in Neural Information Processing Systems*, vol. 31, pp. 6638–6648, 2018.

[17] Y. Bengio, “Learning deep architectures for AI,” *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009, doi: 10.1561/2200000006.

[18] J. Snoek, H. Larochelle, and R. P. Adams, “Practical Bayesian optimization of machine learning algorithms,” in *Advances in Neural Information Processing Systems*, vol. 25, pp. 2951–2959, 2012.

[19] H. Zou and T. Hastie, “Regularization and variable selection via the Elastic Net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[20] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, “LightGBM: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, vol. 30, pp. 3146–3154, 2017.

[21] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 2, pp. 1137–1143, 1995.

[22] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 58, no. 1, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[23] H. Zou and T. Hastie, “Regularization and variable selection via the Elastic Net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[24] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.