Academic performance prediction based on convolutional neural networks and IRT parameters as RGB images

Paulo Flavio Faia Nogueira[1], Cleber Gustavo Dias[2], Fabio Henrique Pereira[2], Francisco Elânio Bezerra[2], Vitor Augusto Dobrochinski[1], Douglas Senta Môr[1], and Flavio Ceban dos Anjos[1]

[1]PREMIA IA LTDA, São Paulo, SP, Brasil; {paulo, vitor, douglas, flavio}@premiagroup.org; Orcid ID: {0009-0006-8890-3037, 0009-0003-1525-656X, 0009-0002-5913-3644, 0009-0008-1013-3014}
[2]Acurai – Pesquisa, Desenvolvimento e Inovação Ltda; {cleberdias, fabiopereira}@acurai.com.br/elanio@hotmail.com; Orcid ID: {0000-0002-4232-2409, 0000-0002-6075-5566,0000-0003-2829-1148}

Abstract

In today's competitive educational environment, institutions face the crucial challenge of effectively assessing student performance, a problem of utmost importance to ensure quality education and develop strategies that improve academic performance and anticipate future demands. The literature explores various approaches to predict student performance using Item Response Theory (IRT) parameters and machine learning techniques. However, there needs to be more in computer vision to capture the behaviour of question assertiveness in image form. This work proposes transforming the IRT parameters into RGB matrices to generate images, which are used to train a convolutional neural network model. The results demonstrate the effectiveness of this method, showing that the images corresponding

https://doi.org/10.32388/K7EHEV

to the highest scores have a lighter tone, reflecting a more significant number of correct answers and, consequently, greater pixel intensity. Furthermore, the model successfully learned the students' scoring patterns, generating a Spearman Correlation for RGB Images of 0.86 for 20,000 images, showcasing its strong generalization capabilities.

Keywords: Predict student performance; Item Response Theory; Computer vision; Convolutional neural network

1 Introduction

Education is pivotal in promoting equal opportunities and reducing social and economic disparities. A well-educated society tends to be more innovative, productive, and resilient, better prepared to tackle challenges and seize future opportunities Zheng and Li [2024]. Thus, measuring students' success is paramount for educational institutions, being a primary criterion in evaluating educational quality Alyahyan and Düstegör [2020].

In today's highly competitive and intricate educational landscape, modern institutions face the critical task of assessing student performance effectively. Evaluating student outcomes is essential for delivering high-quality education, developing strategies to improve academic achievement, and anticipating future educational needs. This rigorous analysis of student performance is pivotal for universities to stay ahead in a competitive environment and to ensure they meet the evolving demands of both students and the workforce Albreiki et al. [2021].

In light of this, academic performance is extensively studied with several approaches, such as assessing the performance of first-level undergraduate students Khairy et al. [2024], enhancing educational assessment through predicting student performance Xuan Lam et al. [2024], student retention and graduation in education Okoye et al. [2024], grading classification to evaluate student performance Zheng and Li [2024], learning recommendation system for Test of English for International Communication (TOEIC) exams Loh et al. [2021], video analysis to forecast student performance Hasan et al. [2020], and academic procrastination due to fear of failure and self-regulation Abdi Zarrin and Gracia [2020], among other studies. These justifications motivate countries to enhance learning quality and student success. Thus, finding an optimal learning trajectory is crucial in educational systems. To achieve this, educational institutions must assess and measure students' performance, as early detection of at-risk students can lead to preventive measures to develop a strategic plan and consequently enhance academic success Zheng and Li [2024], Loh et al. [2021], Alyahyan and Düstegör [2020].

In response to this challenge, several works propose data-driven approaches and machine learning (ML) techniques to improve the learning system, ranging from video-based learning Hasan et al. [2020], analysis of academic procrastination, goal setting, memory strategy, self-evaluation, seeking help, responsibility, and organization Abdi Zarrin and Gracia [2020], use of an interactive learning management system Xuan Lam et al. [2024], knowledge tracking based on performance factor analysis Amal Asselman and Aammou [2023], recommendation system directly optimizing students' learning trajectory for standardized exams Loh et al. [2021], data mining to strategically analyze and manage students' performance Zheng and Li [2024], intelligent tutoring systems designed to provide personalized learning experiences, creating optimized learning paths for each student Lee et al. [2020], among others.

Moreover, in recent decades, artificial intelligence (AI) and machine learning (ML) have been employed in academic institutions to predict students' performance Mastour et al. [2023]. In works such as Alwarthan et al. [2022], Mengash [2020], Singh and Pal [2020], Xuan Lam et al. [2024], Hasan et al. [2020], Olaleye and Vincent [2020], Alyahyan and Düstegör [2020], Okoye et al. [2024], Khairy et al. [2024], authors have focused on predicting academic performance.

In Alwarthan et al. [2022], Singh and Pal [2020], Mengash [2020], Xuan Lam et al. [2024], authors explore different ML algorithms and techniques, such as Logistic Regression, Random Forest, Neural Networks, and Ensemble Learning, to forecast students' performance.

Conversely, Hasan et al. [2020], Namoun and Alshanqiti [2020], Olaleye and Vincent [2020], Alyahyan and Düstegör [2020], Okoye et al. [2024], Khairy et al. [2024] investigated factors influencing students' performance, including previous grades, socio-demographic data, interactions with the education system, student retention, candidate selection, and pedagogical guidance. Furthermore, works based on recommendation systems optimize the learning trajectory.

In Lee et al. [2020], authors developed a score prediction model based on realistic goals, enhancing confidence in the system and student engagement. Meanwhile, in Loh et al. [2021], authors optimized the learning trajectory of students preparing for standardized exams by recommending questions that maximize expected scores. Meanwhile, in Shin and Park [2021], authors proposed a pedagogical word recommendation system for second language learning to predict whether a specific student knows a particular word based on other known words.

1.1 Contribution of this work

Despite the literature showing various works applying machine learning techniques to evaluate student performance, some aspects have received little attention. For example, most studies have used limited data to train the machine learning methods. Another vital point involves the application of deep learning algorithms, as few works have investigated the potential of these algorithms in evaluating student performance Albreiki et al. [2021].

In addition, over the past decade, image classification, which organizes images into pre-established categories, has been the primary strategy for developing visual representations. However, studies evaluating students' grades using computer vision and deep learning have yet to be found in the literature (Wei, 2022) Wei et al. [2022].

Therefore, to fill these gaps, this paper presents the following contributions:

- Utilization of historical data from performance exam of over 4 million students in Brazil.
- Exploratory analysis to explore the characteristics, patterns, and relationships within the data.
- Transformation of IRT parameters of the exam questions into images.
- A Convolutional Neural Network (CNN) trained with the resulting images to predict student performance in the exam.
- Performance metrics to evaluate the model.
- Comparison the prediction model with models recently introduced in the literature.

1.2 Manuscript structure

The remainder of this manuscript organizes as follows: Section 2 describes the Literature Review, Section 3 outlines the proposed method for predicting student performance, Section 4 presents the results and discussions, and Section 5 concludes this article.

2 Machine learning for predicting student performance - Literature Review

Machine learning has revolutionized research in the educational sector by applying algorithms that analyze academic data to discover important patterns, aiding in decision-making and policy formulation Olaleye and Vincent [2020].

In Hasan et al. [2020], the authors utilized Random Forest with data transformation techniques and a genetic algorithm to predict academic performance based on video learning data. The RF approach achieved an accuracy of 88.3%, and multivariate analysis highlighted that selecting key variables resulted in more accurate predictions.

The study presented in Abdi Zarrin and Gracia [2020] investigated the prediction of academic procrastination based on factors such as fear of failure and self-regulation, using t-tests, Pearson correlations, and linear regression. They found that responsibility and fear of failure are significant predictors of academic procrastination.

In Lee et al. [2020], a score prediction model based on deep learning for a multipurpose English learning platform was presented, showing that the attentive model improves the completion rate of diagnostic tests and promotes greater engagement and profitability.

According to Singh and Pal [2020], Naive Bayesian achieved the highest accuracy, and the results suggest that multiple machine learning algorithms, including Decision Tree, Naive Bayesian, K-Nearest Neighbours, and Extra Tree, with Bagging and Boosting methods, can be used to improve the prediction of student performance and identify low-performing students.

In Olaleye and Vincent [2020], the authors identified relevant attributes from academic data of postgraduate teachers in Nigeria. Using algorithms such as C4.5, CART, and Naive Bayes, the study concluded that the C4.5 classifier achieved the highest accuracy (96.3%). Additionally, the authors highlighted that, despite the low academic scores associated with the teaching practice program, it is essential for the practical experience of trainee teachers.

The study presented in Loh et al. [2021] and Shin and Park [2021] focused on developing methods for TOEIC exams. While Loh et al. [2021] developed a learning recommendation system using bidirectional recurrent neural networks with Long Short-Term Memory (LSTM), where the Recommendation for Calibrated Expected Score (RCES) approach outperformed traditional methods and other baselines in terms of learning gain and showed a more uniform probability distribution of correctness in recommended questions, in contrast to the trend of Recommendation for Expected Score (RES) to suggest more straightforward questions, Shin and Park [2021] developed a pedagogical word recommendation system using Matrix Factorisation (MF) and Neural Collaborative Filtering (NCF). The study suggested that vocabulary prediction could be crucial for other Natural Language Processing (NLP) tasks, such as lexical simplification and improving computer-assisted learning.

In Alwarthan et al. [2022], the authors explored data mining techniques to predict the academic performance of university applicants in Saudi Arabia. By utilizing artificial neural networks (ANN), Decision Tree (DT), Support Vector Machine (SVM), and Naive Bayes, they found that student performance could be predicted before admission based on pre-admission criteria, highlighting the accuracy of the School Performance Admission Test.

As reported in more recent works, Zheng and Li [2024], Khairy et al. [2024], Xuan Lam et al. [2024], and Amal Asselman and Aammou [2023], the authors focused on the accuracy of student performance prediction models. In Zheng and Li [2024], the authors investigated grade classification and prediction using Naive Bayes with specific optimizations, such as Naive Bayes with Artificial Rabbits (NBAR). The NBAR outperformed Naive Bayes with Jellyfish Search (NBJS) and the Naive Bayes Classifier (NBC), excelling in both F1-score and recall predicting G1 and G3 grades.

In Khairy et al. [2024], the authors applied machine learning techniques such as Random Forest (RF), Decision Tree (DT), Naive Bayes (NB), Neural Network (NN), and K-Nearest Neighbours (KNN) to predict the academic performance of first-level Computer Science students. The study revealed that RF and DT achieved an accuracy of 98.7%, while NN, NB, and KNN obtained 96.4%, 94%, and 89.6%, respectively. Regarding accuracy, recall, and F1-score metrics, RF and DT also excelled, achieving 0.99 in all three metrics.

In Xuan Lam et al. [2024], the focus was on improving educational assessment through predictive modeling of student performance. Using algorithms such as RF, Logistic Regression (LR), Support Vector Machine (SVM), Naive Bayes (NB), k-Nearest Neighbour (kNN), and Stacking, the authors emphasized the importance of attendance monitoring and targeted interventions to improve student outcomes, offering valuable tools for supporting academic progress and intervention. In Amal Asselman and Aammou [2023], the authors compared different machine learning models, including Random Forest, AdaBoost, and XGBoost, to predict student performance. XGBoost proved superior, providing significant improvements in predicting student performance when handling large datasets from educational platforms.

Finally, Okoye et al. [2024] have focused on the retention of undergraduate students and used an ensemble algorithm based on Bagging to predict student retention. The results indicated a high efficacy of the model in predicting retention and graduation status, with exceptionally high accuracy and recall for retention (0.909 and 1.000, respectively) and slightly lower but still significant performance for graduation prediction.

3 Related Theory

3.1 Item Response Theory applied to the National High School Examination

The National High School Exam (ENEM) is a standardized assessment tool for evaluating students' performance in Brazil. Administered by the Brazilian Federal Government, it is used not only for university admissions in Brazil and Portugal but also for student funding and support applications. The National Institute for Educational Studies and Research (INEP) utilizes IRT to analyze the competencies and skills demonstrated by candidates across four key areas: Mathematics, Languages, Natural Sciences, and Humanities. This approach allows for a detailed understanding of students' proficiency levels in these subjects Brasil [2022], Costa et al. [2023].

The item response theory (IRT) resembles a statistical estimate theory. It utilizes latent characterizations of individuals and items as predictors of observed answers. One of the fundamental models in IRT is the 1-parameter response model, also known as the Rasch model. However, there are more complex models that incorporate the 2 parameters model (difficulty parameter) and the 3 parameter model (a setting parameter at random) Bock and Gibbons [2021], Baker and Kim [2004].

According to Bock and Gibbons [2021], the equation below describes the IRT of 3 parameters:

$$P(X_{ij} = 1) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-(a_i(\theta_j - b_i))}} \tag{1}$$

in which:

- $P(X_{ij} = 1)$ is the probability that individual j answer correctly the item i,
- θ_j is the ability parameter of individual j,
- a_i is the discrimination parameter of item i,
- b_i is the difficulty parameter of item i, and

- c_i is the parameter for the probability of a correct response by chance for item i

3.2 Image digitization process

A digital image is a visual representation composed of pixels, the most minor units of an image. Each pixel contains information about color and intensity, arranged in a grid of rows and columns, forming a two-dimensional matrix. In an RGB image, each pixel is represented by three color channels: R (red), G (green), and B (blue). These channels correspond to three two-dimensional matrices, one for each color. Each value in the matrix can range from 0 to 255, where 0 signifies the total absence of the color and 255 signifies maximum color intensity A. Baskar [2023], Umbaugh [2023].

Fig. 1 presents a simple example of generating a digital image by transforming matrices R, G, and B into a single RGB matrix. Each matrix R, G, and B is a two-dimensional matrix of size m×n, where m is the number of rows (height), and n is the number of columns (width) of the image. The elements of these matrices represent the intensity of the respective color channels (Red, Green, Blue) for each pixel A. Baskar [2023], Umbaugh [2023].

$$\underset{\text{Matrix R}}{\begin{bmatrix} 255 & 0 & 0 \\ 255 & 0 & 0 \\ 255 & 0 & 0 \\ 255 & 0 & 0 \end{bmatrix}} \quad \underset{\text{Matrix G}}{\begin{bmatrix} 0 & 255 & 0 \\ 0 & 255 & 0 \\ 0 & 255 & 0 \\ 0 & 255 & 0 \end{bmatrix}} \quad \underset{\text{Matrix B}}{\begin{bmatrix} 0 & 0 & 255 \\ 0 & 0 & 255 \\ 0 & 0 & 255 \\ 0 & 0 & 255 \end{bmatrix}} \quad \Rightarrow \quad \underset{\text{Matrix RGB}}{\begin{bmatrix} (255,0,0) & (0,255,0) & (0,0,255) \\ (255,0,0) & (0,255,0) & (0,0,255) \\ (255,0,0) & (0,255,0) & (0,0,255) \\ (255,0,0) & (0,255,0) & (0,0,255) \end{bmatrix}}$$

Figure 1: Transformation of R, G, and B matrices into an RGB image.

3.3 Images Classification by using Deep Learning

The convolucional neural network (CNN) recognises patterns in high-dimensionality data developed by LeCun et al. [1998]. This neural network operates through convolution, an operation where a filter (also known as kernel), which is sliding on the input, calculates the scalar product between the filter and the corresponding region of the input. This results in a map of characteristics that highlights different aspects of the image, such as edges, textures, and patterns LeCun et al. [1998], Rosebrock [2017].

The CNN network has several convolution layers, each applying filters to extract different input aspects. These layers are responsible for learning hierarchical representations of the data, starting with superficial characteristics in the initial layers and progressing to more complex characteristics in the posterior layers. After each convolution operation, Rectified Linear Unit (ReLU) activation function is applied to results, introducing non-linearity into the network and allowing it to learn more complex relationships in the data. Then, the pooling layers are used to reduce the size of the characteristics maps, making processing more efficient and helping to avoid overfitting LeCun et al. [1998], Rosebrock [2017].

The fully connected layers are responsible for combining the features learned from all areas of the input to make final predictions. The output layer consists of a fully connected layer with a number of neurons equal to the number of classes in the problem LeCun et al. [1998], Rosebrock [2017].

3.4 Metrics to evaluate the performance of machine learning models

The model performance evaluation is crucial to understanding how well it can generalize. For regression models, several metrics can be used to evaluate how well the model adjusts to the data. For example, the R-squared (R^2) metric measures the proportion of variance in the dependent variable that is predictable from the independent variable. The average quadratic error (MSE) calculates the average squares of errors between the expected and actual values. The absolute average error (MAE) calculates the average of the absolute error values between the expected and actual values. The root of the average quadratic error (RMSE) is the square root of the MSE and provides an interpretation on the same scale as the dependent variable. The average percentage error (MAPE) calculates the average of the absolute percentages of errors between the expected values and the actual values Rosebrock [2017].

Eq. 1 and 2 have mathematical MAE and MAPE errors.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{2}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

in which:

- n is the number of samples,
- y_i is the target,
- $\hat{y}_i$ is the model predicted value.

4 Materials and Methods

The proposed model was developed in the context of a computational platform aimed at enhancing the teaching-learning process for high school students. This educational platform (Aprova+™) is based on the concept of quizzes, which aims to provide students with an environment that fosters their performance and attention while respecting individual differences, as detailed in the study in Nogueira et al. [2024]. The platform establishes a set of steps for the evaluation process of students, including the automatic selection of questions based on the student's development history, diagnostic stages, content recommendations, and overall performance analysis to construct a learning path, identify gaps, and provide recommendations and reinforcement. Specifically for this work, the application of the model was considered in the analysis of questions in the National High School Exam (ENEM) in Brazil.

The process outlined in Fig. 2 was used to predict student performance in the ENEM exam. This process consists of the following steps: data collection, exploratory analysis, data preparation and normalization, matrix calculation and image generation, model construction and training and model validation.

4.1 Data collection and exploratory analysis

The data collection process involved gathering information on student grades, the number of correct answers, and the parameters of the exam itens. The analyses were conducted considering the areas of interest independently.

4.2 Data preparation and Normalization

Data preparation involved deleting null results and values, selecting items, and coordinating and normalizing data only from the humanities area. After that, we used the Scikit-Learn Minmaxscaller library to normalize the data, adjusting the values to vary between 0 and 1.

4.3 Matrix Calculation and Image Generation

To create an image using the parameters from a set of n questions answered by a student, the parameters are organized into vectors and then transformed into matrices. The matrices are structured as two-dimensional arrays with dimensions $m \times n$, where m is the number of rows, and n is the number of questions. Subsequently, these three matrices are combined to form a resulting RGB matrix that creates a digital image.

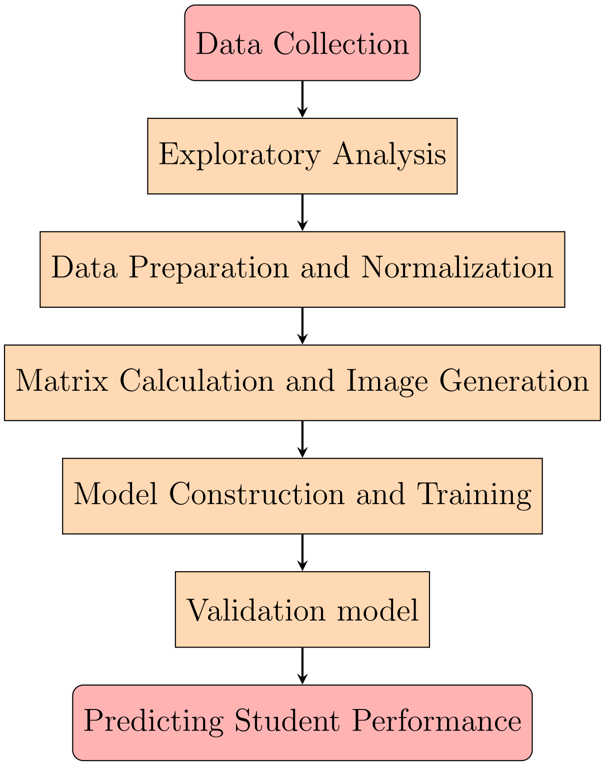

Figure 2: Flowchart of the student performance prediction process

The size of the RGB matrix depends on the number of correct answers to the questions. Different-sized images are created based on the number of correct answers.

4.4 Model Construction and Training

To build the model using a CNN neural network (Fig. 3), the TensorFlow/Keras Sequential library was used. The model consists of two convolutional layers followed by 2D pooling layers. Each layer was set up with two 2x2 convolutional filters and activated using the ReLU function. The 'same' padding parameter was used to maintain consistent output dimensions. A 2D MaxPooling layer was applied to reduce spatial dimensions. The process also involved flattening and adding four dense layers, each containing 64 neurons. Finally, a dense layer without activation was included for the regression. The model was compiled using the Adam optimizer, and the loss function was defined as MSE and MAE. With the model trained, we evaluated its performance using MAE and MAPE metrics.

The CNN neural network configuration shown in Fig. 3 was trained using

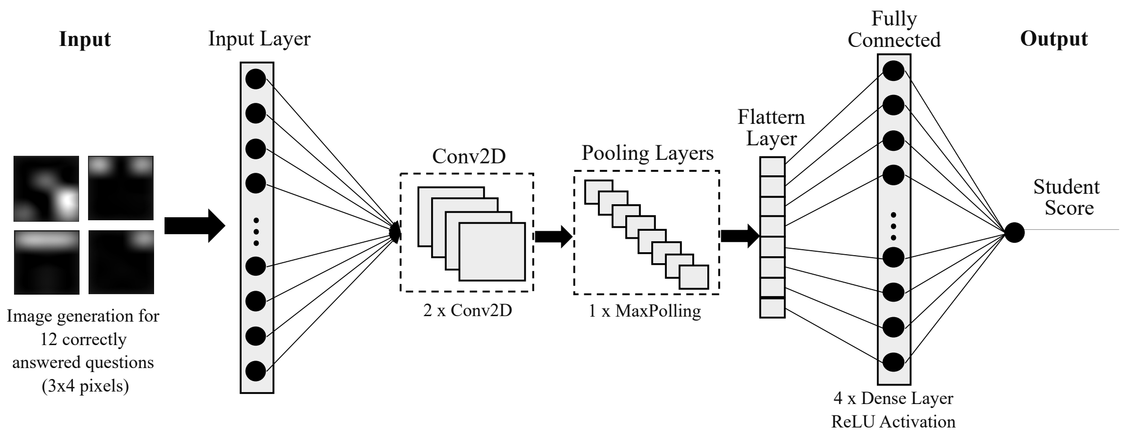

Figure 3: Student score prediction model
.

the following parameters:

- Layers 2 x Conv2D = (16, 32)
- Kernel size = 2 x (2,2)
- Activation of Conv2D = ReLU
- Layers Dense = (64, 64, 64, 64)
- Activation of layers Dense = ReLU and Linear
- Optimizer = Adam
- Epochs = 80
- Batch size = 64
- Train size = 80%
- Test size = 20%
- Loss = MSE, MAE

5 Results and Discussion

The research conducted in Alyahyan and Düstegör [2020] describes that the appropriate and efficient use of techniques for analyzing educational data can transform education, allowing a better understanding of student progress and facilitating more effective interventions to improve student success and retention.

The findings shown in Fig. 4 display the generated images and their corresponding notes for one of the areas of interest. Images with higher scores are more straightforward, as expected, containing more hits and resulting in more intense pixels. Furthermore, the higher the item parameter values, the greater the intensity of each pixel, as these values are multiplied when the answer is correct. This behavior aligns with Baker and Kim [2004] research, where item parameters, such as difficulty and discrimination, affect the likelihood of correct answers. This implies that the intensity of the pixels varies according to the highest scores in the visual representations of the test data.

According to the findings presented in Hellas et al. [2018], a review examines various predictive models used to identify students at risk for academic failure. They discuss how prediction models use different student data to increase prediction accuracy, which relates to how item parameters and correct answers are visually represented. In Fig. 4, the figures have different shapes even with the same score, indicating that students correctly answered questions with different parameters. This demonstrates how the model can capture the behaviour of IRT parameters and visually represent these variations.

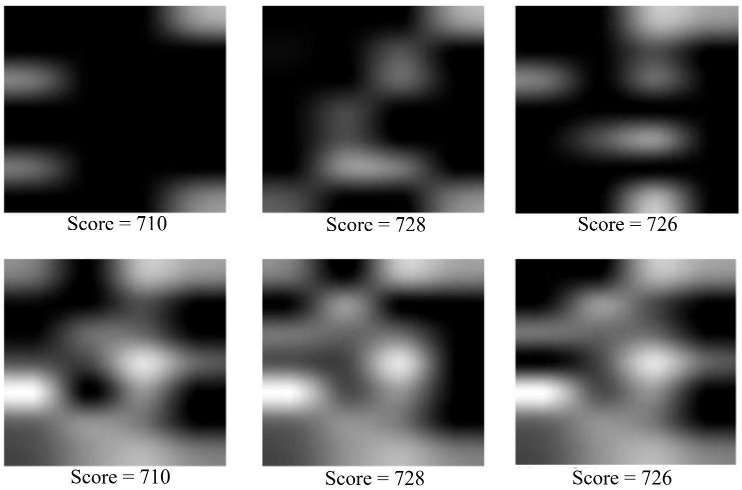

Figure 4: Examples of RGB images and their respective notes for the Human Science area

.

Tab. 1 shows the correlations between the images created for different questions and the final grades, based on a total of 20,000 images (corresponding to 20,000 students). The analysis revealed that the highest correlation was observed for around 16 questions, and for these questions, the correlations frequently exceeded 0.8. This information is important for determining the optimal number of images to be analyzed in order to ensure the desired accuracy in predicting students' final scores.

Table 1: Spearman Correlation for RGB Images.

RGB Images	**Spearman Correlation for 20,000 images**
Generated with 8 questions	0.648
Generated with 12 questions	0.751
Generated with 15 questions	0.755
Generated with 16 questions	0.860
Generated with 20 questions	0.837
Generated with 24 questions	0.825
Generated with 28 questions	0.847

Tab. 2 summarizes some works, including the variables, techniques, and accuracy applied in predicting student performance. Although the table presents results with greater accuracy, such as 1, 4, 6, and 16, this work presented a different approach based on IRT and images to predict students' grades.

Although the results obtained were based on questions from the National High School Exam, the model proposed in this work can be applied to estimate the score in any other test or exam, as long as the calculation of the student's final score in these exams is derived from Item Response Theory. Finally, it is important to highlight that the model proposed in this research was entirely developed by the authors of the study for integration into the Aprova+™Platform, and therefore does not use any third-party models for score prediction.

6 Conclusions

This study introduced an approach for predicting student scores on an exam in Brazil using images and a Convolutional Neural Network. In terms of accuracy, the results of this work are comparable to those of some studies presented in the literature. It successfully learned the students' scoring patterns from images representing 16 correctly answered questions, showcasing its strong generalization capabilities. Moreover, this study highlights the significant role of understanding the IRT parameters through image representations. The parameters are critical in IRT as they influence the probability of a correct response based on a student's latent ability. The model's successfully interpreting these parameters visually means it can effectively capture the nuances of test item characteristics, contributing to its predictive accuracy. This approach not only enhances the model's reliability but also holds the potential to provide valuable insights into the underlying factors affecting student performance, thus offering a powerful tool for educational assessments and interventions. Future work will explore additional neural networks, such as Generative Adversarial Networks, and incorporate more comprehensive datasets to further enhance model performance and accuracy.

Patents

The technological solution presented in this article is currently undergoing a patentability review process at the National Institute of Industrial Property in Brazil.

Author Contributions

Conceptualization and methodology, C.G.D. and F.E.B.; software, C.G.D.; validation, C.G.D. and F.H.P.; formal analysis, P.F.F.N., C.G.D and F.H.P; investigation, C.G.D. and F.H.P.; resources, P.F.F.N.; data curation, C.G.D.; writing—original draft preparation, F.E.B. and C.G.D.; writing—review and editing, F.E.B., C.G.D. and F.H.P.; visualization, C.G.D.; supervision, P.F.F.N., C.G.D and F.H.P; project administration, P.F.F.N.; funding acquisition, P.F.F.N. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

We have no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AdaBoost	Adaptive Boosting
AI	Artificial Intelligence
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DT	Decision Tree
ENEM	National High School Examination
ET	Extra Tree
INEP	National Institute of Educational Studies and Research
KNN	K-Nearest Neighbors
LR	Logistic Regression
LSTM	Long Short-Term Memory
MF	Matrix Factorization
ML	Machile Learning
NB	Naive Bayesian
NBAR	Naive Bayes com Artificial Rabbits
NBJS	Naive Bayes com Jellyfish Search
NCF	Collaborative Neural Filtering
NLP	Natural Language Process
NN	Neural Network
RCES	Recommendation for Calibrated Expected Score
ReLU	Rectified Linear Unit
RES	Recommendation for Expected Score
SVM	Support Vector Machine
IRT	Item Response Theory
XGBoost	Xtreme Gradient Boosting

References

Shriram K Vasudevan T.S. Murugesh A. Baskar, Muthaiah Rajappa. Digital image processing. *Chapman and Hall/CRC*, 1:208, 2023. doi: 10.1201/9781003217428.

Sohrab Abdi Zarrin and Esther Gracia. Prediction of academic procrastination by fear of failure and self-regulation. *Educational Sciences: Theory and Practice*, 20(3):34–43, 2020. doi: 10.12738/jestp.2020.3.003.

Balqis Albreiki, Nazar Zaki, and Hany Alashwal. A systematic literature review of student' performance prediction using machine learning techniques. *Education Sciences*, 11(9), 2021. ISSN 2227-7102. doi: 10.3390/educsci11090552.

Sarah A. Alwarthan, Nida Aslam, and Irfan Ullah Khan. Predicting student academic performance at higher education using data mining: a systematic review. *Applied Computational Intelligence and Soft Computing*, 2022, 2022. doi: 10.1155/2022/8924028.

Eyman Alyahyan and Dilek Düstegör. Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1):3, 2020. doi: 10.1186/s41239-020-0177-7.

Mohamed Khaldi Amal Asselman and Souhaib Aammou. Enhancing the prediction of student performance based on the machine learning xgboost algorithm. *Interactive Learning Environments*, 31(6):3360–3379, 2023. doi: 10.1080/10494820.2021.1928235.

Frank B. Baker and Seock-Ho Kim. *Item Response Theory: Parameter Estimation Techniques, Second Edition*. CRC Press, 2nd edition, 2004. doi: 10.1201/9781482276725.

R. Darrell Bock and Robert D. Gibbons. *Item Response Theory*. Wiley, Hoboken, first edition edition, 2021. doi: 10.1002/9781119716723. University of Chicago.

Ministério da Educação Brasil. Fazer o exame nacional do ensino medio. `https://gov.br/pt-br/servicos/fazer-o-exame-nacional-do-ensino-medio`, 2022.

H.G. Costa, L.A.d. Souza, and M.C. Roboredo. Applying electre tri to sort states according the performance of their alumni in brazilian national high school exam (enem). In A. Abraham, T.P. Hong, K. Kotecha, K. Ma, P. Manghirmalani Mishra, and N. Gandhi, editors, *Hybrid Intelligent Systems*, volume 647 of *Lecture Notes in Networks and Systems*. Springer, Cham, 2023. doi: 10.1007/978-3-031-27409-1_73.

Raza Hasan et al. Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences*, 10(11):3894, 2020. doi: 10.3390/app10113894.

A. Hellas, P. Ihantola, and A. Petersen. An overview of predictive models for early identification of at-risk students. *Journal of Educational Data Mining*, 10(3):1–20, 2018.

Dalia Khairy et al. Prediction of student exam performance using data mining classification algorithms. *Education and Information Technologies*, pages 1–25, 2024. doi: 10.1007/s10639-024-12619-w.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Youngnam Lee et al. Prescribing deep attentive score prediction attracts improved student engagement. *arXiv preprint arXiv:2005.05021*, 2020. doi: 10.48550/arXiv.2005.05021.

Hyunbin Loh et al. Recommendation for effective standardized exam preparation. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, pages 397–404, 2021. doi: 10.1145/3448139.3448177.

Haniye Mastour, Toktam Dehghani, Ehsan Moradi, and Saeid Eslami. Early prediction of medical students' performance in high-stakes examinations using machine learning approaches. *Heliyon*, 9(7):e18248, 2023. ISSN 2405-8440. doi: 10.1016/j.heliyon.2023.e18248.

Hanan Abdullah Mengash. Using data mining techniques to predict student performance to support decision making in university admission systems. *IEEE Access*, 8:55462–55470, 2020. doi: 10.1109/ACCESS.2020.2981905.

Abdallah Namoun and Abdullah Alshanqiti. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1):237, 2020. doi: 10.3390/app11010237.

Paulo Flavio Faia Nogueira, Cleber Gustavo Dias, Fabio Henrique Pereira, and Alberto Malta Junior. The impact of using a computational platform on the enhancement of the teaching-learning process: a case study applied to high school students. *Revista Caderno Pedagógico*, 2024. doi: https://doi.org/10.54033/cadpedv21n6-144.

Kingsley Okoye, Julius T. Nganji, Jose Escamilla, and Samira Hosseini. Machine learning model (rg-dmml) and ensemble algorithm for prediction of students' retention and graduation in education. *Computers and Education: Artificial Intelligence*, 6:100205, 2024. ISSN 2666-920X. doi: 10.1016/j.caeai.2024.100205.

Taiwo Olapeju Olaleye and Olufunke Rebecca Vincent. A predictive model for students' performance and risk level indicators using machine learning. In *2020 International Conference in Mathematics, Computer Engineering* *and Computer Science (ICMCECS)*, pages 1–7. IEEE, 2020. doi: 10.1109/ICMCECS47690.2020.240897.

Adrian Rosebrock. *Deep Learning for Computer Vision with Python, Starter Bundle*. PyImageSearch, 1.1.0 edition, 2017.

Jamin Shin and Juneyoung Park. Pedagogical word recommendation: A novel task and dataset on personalized vocabulary acquisition for l2 learners. *arXiv preprint arXiv:2112.13808*, 2021. doi: 10.48550/arXiv.2112.13808.

Randhir Singh and Saurabh Pal. Machine learning algorithms and ensemble technique to improve prediction of students performance. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3):3970–3976, May 2020. ISSN 2278-3091. doi: 10.30534/ijatcse/2020/221932020.

Scott E. Umbaugh. *Digital Image Processing and Analysis: Computer Vision and Image Analysis*. CRC Press, 4th edition, 2023. doi: 10.1201/9781003221135.

Yixuan Wei, Yue Cao, Zheng Zhang, Zhuliang Yao, Zhenda Xie, Han Hu, and Baining Guo. icar: Bridging image classification and image-text alignment for visual recognition, 2022.

Pham Xuan Lam, Phan Quoc Hung Mai, Quang Hung Nguyen, Thao Pham, Thi Hong Hanh Nguyen, and Thi Huyen Nguyen. Enhancing educational evaluation through predictive student assessment modeling. *Computers and Education: Artificial Intelligence*, 6:100244, 2024. ISSN 2666-920X. doi: 10.1016/j.caeai.2024.100244.

Xin Zheng and Conghui Li. Predicting students' academic performance through machine learning classifiers: A study employing the naive bayes classifier (nbc). *International Journal of Advanced Computer Science & Applications*, 15(1), 2024. doi: 10.14569/IJACSA.2024.0150199.

Table 2: Summary of Techniques and Accuracy in prediction student performance.

Authors	**Variables**	**Types of Techniques**	**Accuracy**
Khairy et al. [2024]	Year, midterm, practical exam, writing exam, final total degree, and grade	Random Forest, Decision Tree, Naive Bayes, Neural Network, and K-Nearest Neighbours	Precision - 0.99, Recall - 0.96, F-Score - 0.94
Okoye et al. [2024]	Number of Retention and Graduation	Ensemble algorithm based on the Bagging method	Precision 0.909, Recall 1.000, Accuracy 0.909, F1-score 0.952, Error-rate 0.090
Zheng and Li [2024]	Students' school, gender, age, home address, family size, parental cohabitation status, education and occupations of both parents	Ellyfish Search Optimizer and Artificial Rabbits Optimization	F1-Score 1.03%, Recall 0.18%
Xuan Lam et al. [2024]	Enhancing educational evaluation through predictive student assessment modeling	Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), k-nearest Neighbor (k-NN) e Stacking	Accuracy 0.826, Precision 0.800, Recall 0.790, F1 0.794
Okoye et al. [2024]	Prediction of students' retention and graduation in education	Ensemble algorithm based on the Bagging method	Precision 0.909, Recall 1.000, Accuracy 0.909, F1-score 0.952, Error-rate 0.090
Olaleye and Vincent [2020]	Cumulative Grade Point Aggregates, including other demographic data and Grade Point Aggregates of courses taken by students in the analysis	C4.5, CART, and Naïve Bayes algorithms	96.3% accuracy for C4.5
This approach	Information regarding student grades, number of correct answers, and exam item parameters for the year 2022	Convolutional neural networks and IRT parameters as RGB images	0.86 of Spearman Correlation for 20,000 images