Designing Low-Resource Pedagogical Intervention Controllers: A Comparative Study of Compact Neural and Neuroevolutionary Architectures

Houssam Ahmed Amin Bahi, Karima Boussaha, and Zakaria Laboudi

***Abstract*—In the context of Massive Open Online Courses (MOOCs), data-driven models play a vital role in assisting students through personalized interventions. Earlier works have focused on the prediction of students' performance and dropouts. However, the design of efficient decision-making controllers has received less attention. In this paper, the selection of pedagogical intervention is formulated as a low-resource controller design. A comparative study of traditional machine learning models, compact multilayer perceptron, and sparse neural controllers obtained through neuroevolutionary search is proposed. In addition, a novel framework is proposed as a combination of NEAT and weight-agnostic neural networks. In the experiments, a large-scale educational dataset is used. It is observed that compact neural models achieve high performance with reduced model sizes. Moreover, sparse neuroevolutionary controllers achieve extreme compression with acceptable performance. Thus, the need for balancing efficiency in the design of scalable intelligent educational systems is highlighted.**

***Index Terms*—Educational Data Mining, MOOCs, Pedagogical Intervention, Model Compression, Low-Resource Learning, Intelligent Educational Systems.**

I. Introduction

ALL Massive Open Online Courses (MOOCs) have been seen to contribute to the expansion of access to education by promoting large-scale participation among diverse learner groups. Despite the advantages, MOOCs still experience high dropout rates and limited personalization, especially during early learning stages [1]. Intelligent educational systems have been proposed to mitigate the challenges by providing adaptive feedback and interventions [2, 3, 4]. Recent developments in artificial intelligence, such as large language models, have improved the efficiency of the systems, but at the expense of increased complexity, especially during deployment, especially in environments with limited resources.

Manuscript received April 7, 2026; revised April 27, 2026. Date of publication August 17, 2026. Date of current version August 17, 2026. The associate editor prof. Ana Sović Kržić has been coordinating the review of this manuscript and approved it for publication.

H. A. A. Bahi is with the Department of Mathematics and Computer Science, Research Laboratory on Computer Science's Complex Systems (ReLa(CS)2), University of Oum El Bouaghi, Algeria (e-mail: houssam.bahi@univ-oeb.dz).

K. Boussaha is with the Department of Mathematics and Computer Science, LIAOA Laboratory, University of Oum El Bouaghi, Algeria (e-mail: karima.boussaha@univ-oeb.dz).

Z. Laboudi is with the Department of Networks and Telecommunications, LIAOA Laboratory, University of Oum El Bouaghi, Algeria (e-mail: laboudi.zakaria@univ-oeb.dz).

Digital Object Identifier (DOI): 10.24138/jcomss-2026-0090

One of the most critical, yet not well explored, components of the intelligent educational systems is the decision layer, which guides the choice of interventions. Much work has been done on predicting the performance and dropout probability of the learners [5, 6], yet little attention has been given to the design of lightweight controllers for the decision layer. However, in real-world cases, especially in situations where there are resource constraints, the effectiveness of this controller may also be equally important to its prediction accuracy in real-world intelligent system implementations [7]. The current state of affairs in controller selection for intelligent educational systems generally employs rule-based heuristics and machine learning models. Although these models are effective to a certain degree, they may also be less flexible and/or may not provide a proper balance between model simplicity and decision accuracy. At the same time, more advanced models may also incur unnecessary computational costs if they are to be implemented for selecting discrete interventions [8]. Thus, it is also essential to explore which controller architectures can be most suitable for low-resource intelligent educational systems.

In this paper, we explore the problem of selecting pedagogical interventions within low-resource controllers for intelligent educational systems. We propose a comparative approach to controller selection that considers a variety of architectures, including classical machine learning models, compact multilayer perceptron models, and sparsely connected neural network controllers discovered via neuroevolutionary optimization. Specifically, we also explore a possible hybrid model that combines NeuroEvolution of Augmenting Topologies (NEAT) and Weight Agnostic Neural Networks (WANN) principles to discover more structurally efficient controllers. The aim of this research is to explore performance/efficiency trade-offs with these controllers. We carry out this task by evaluating models on a large-scale educational data set with varying parameter budgets and multiple random seeds. This allows us to comprehensively evaluate these controllers. The results demonstrate that compact neural models maintain a considerable percentage of performance with significantly lower parameter counts compared to larger models. Additionally, sparse neuroevolutionary controllers enable extreme compression with reasonable performance. This research demonstrates the need to consider deployment constraints and predictive accuracy when developing intelligent educational systems.

The main contributions of this work are as follows:

- Formulating pedagogical intervention selection as a low-resource controller design problem within intelligent educational systems, explicitly incorporating both predictive performance and computational efficiency.
- Providing a comprehensive comparative evaluation of diverse controller architectures, including classical machine learning models, compact multilayer perceptrons, and sparsely connected neuroevolutionary controllers, under consistent experimental conditions.
- Proposing a hybrid neuroevolutionary framework that combines principles from NEAT and weight-agnostic neural networks to discover structurally efficient and compact controller architectures.
- Demonstrating, through extensive experiments, the performance–efficiency trade-offs across models, highlighting the effectiveness of compact neural networks and the potential of sparse controllers in extreme resource-constrained scenarios.

The rest of the paper is structured as follows: Section II discusses the relevant background, Section III explains the methodology, Section IV presents the results, Section V discusses the findings, and Section VI concludes the paper.

II. Related Work

The research on the design of intelligent educational systems using learning analytics and EDM has been quite active, with a focus on student modeling and predictions such as student performance, engagement, and dropout rates. The initial research was centered around traditional machine learning algorithms such as logistic regression, decision trees [11, 12], and ensembles with interaction and demographic data [13, 14].

Recent research has also been conducted on the use of deep learning algorithms such as recurrent neural networks [15], convolutional neural networks [16], and attention models [17] to effectively model student behaviour. Although the results are encouraging, the algorithms do require significant computational power.

On the other hand, research on the design of interventions for MOOCs has been conducted through personalized feedback [2], recommendation systems [18, 19], and adaptive content delivery systems [20]. Although rule-based systems have been popular due to their interpretability [21], they are less flexible. Research on the use of reinforcement learning is also present [22]. With recent developments in intelligent assistant technology, especially large language models, interactive tutoring and explanation generation have become possible [23]. However, the need for efficient decision-making layers persists. This is an unexplored research challenge. For efficiency, compression techniques like pruning, quantization, and knowledge distillation have been used extensively [24, 25].

In the context of education, compact neural networks, especially multilayer perceptron, have been found to achieve high performance with low parameter counts. Neuroevolution provides a different perspective through the evolution of the network. NEAT allows the simultaneous evolution of the structure and the weights [9], whereas WANN focuses on the efficiency of the structure with minimal reliance on weights [10]. Even a combination of the two methods, along with gradient-based training, has been proposed [26], although it is still in its infancy.

In conclusion, although substantial progress has been achieved in the areas of prediction, intervention design, and efficient modelling, the problem of designing efficient controllers while balancing performance and efficiency in the context of pedagogical intervention selection remains understudied. This paper seeks to bridge the gap by presenting a comparative analysis of efficient controllers in a unified framework.

III. Methodology

In the following section, the methodology that is adopted to study the low-resource pedagogical intervention controller is explained. The methodology is centered around the proposed framework to evaluate different controller architectures under the same conditions with a focus on balancing the predictive accuracy of the controller with efficiency considerations. The methodology is divided into three different sections: problem formulation/data processing, controller architectures, and the neuroevolutionary search framework.

A. Problem Formulation and Data Processing

In the proposed methodology, the problem of selecting the pedagogical interventions is cast as a multi-class classification problem, where each data point is represented as the state of the learner at a given time step, with the target value being the appropriate action to be taken as the pedagogical intervention. Let the feature vector of the state of the learner be represented as $x \in \mathbb{R}^d$ with the target value represented as $y \in \{1, \dots, C\}$. A controller $\hat{y} = f_\theta(x)$ is learned such that:

$$\hat{y}_i = \arg\max_c \; p_\theta(y = c \mid x_i) \tag{1}$$

The model is trained using the cross-entropy loss:

$$\mathcal{L}(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} \mathbf{1}(y_i = c) \log p_\theta(y = c \mid x_i) \tag{2}$$

Due to the presence of deployment constraints, the controller design is formulated as a performance efficiency trade-off. This is represented as follows:

$$\max_\theta \; P(\theta) - \lambda\, \Omega(\theta) \tag{3}$$

where $P(\theta)$ represents the predictive performance, whereas $\Omega(\theta)$ represents the model complexity. The features used for the controller are derived from the interaction data. This includes behavioral, performance, and demographic characteristics. The features used represent the short-term and long-term learning behaviour. In the context of the proposed framework, the missing values are handled using the appropriate imputation strategy. In addition, the categorical values are represented using the one-hot encoding scheme. Finally, the numerical values are standardized.

One important aspect of the proposed formulation is the intervention label. In the context of real-world datasets, it is difficult to obtain the optimal intervention label. In this regard, the intervention label is obtained using the heuristic rule. This is obtained using the learner behaviour. This is a realistic intervention policy. In this regard, the intervention label is obtained using the motivation prompt, explanation, detailed instruction, quiz-based reinforcement, and resource recommendation [27].

The following provides an explicit representation of the heuristics defining the proxy labeling process for reproducibility, based on learner behavior. Proxy labels are assigned as follows:

- *MOTIVATE*: If the learner shows decreased activity levels with moderate performance.
- *EXPLAIN_SIMPLE*: If recent performance is poor while engagement remains moderate.
- *EXPLAIN_DETAILED*: If both engagement and performance have been low for more than one time step.
- *QUIZ_CHECK*: If moderate performance exhibits noticeable fluctuations.
- *RESOURCE_RECOMMEND*: If both engagement and performance are low and the learner is at increased risk.

Each label is applied based on conditions derived from aggregated learner behavior features, including recent scores, activity levels, and risk factors. While proxy labeling leverages collected data, it relies on aggregated behavioral indicators rather than exposing explicit labeling conditions to the model, reducing potential circularity.

In order to have a realistic evaluation protocol, the data is split using a group-aware splitting method, avoiding any information leak between learners. All the time observations of a specific learner are assigned to the same partition, making sure that learners are tested on unseen students. Multiple random seeds are employed, and the results are reported as the mean values.

B. Controller Architectures

We examine a variety of controller architectures. Conventional machine learning approaches, including logistic regression and tree-based ensembles, are used as interpretable and computationally efficient baselines. Simultaneously, a variety of feedforward neural networks are tested, ranging from larger multilayer perceptron to extremely compact ones. A larger neural network is utilized as a teacher model to assist the training of smaller controllers via knowledge distillation. This allows small controllers to mimic the larger model's behaviour while being much simpler [28].

C. Neuroevolutionary Search Framework

Additionally, we propose sparse neural controllers that were learned using neuroevolutionary search. Each controller is defined as a sparse connected feedforward network. In the sparse connected networks, the connectivity is defined using binary masks that specify the presence or absence of the connection between the different layers. This allows us to explore different

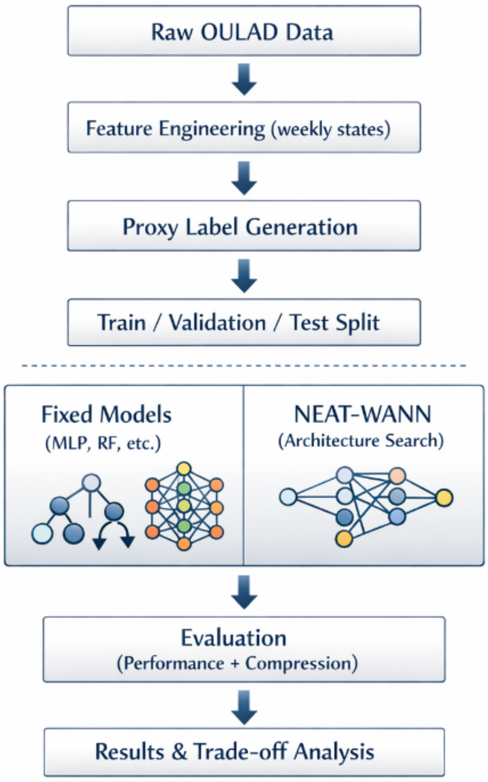

Fig. 1. Overview of the proposed framework for low-resource pedagogical intervention selection.

networks in a flexible manner, while still controlling the size of the models. The search for the architecture is done using a set of candidate architectures. The candidate architectures evolve through the iterations. In each iteration, the controllers are evaluated using a fitness function. In the hybrid architecture, the evaluation of the architectures is done using a set of shared weights:

$$F(a) = \frac{1}{K}\sum_{k=1}^{K} P(a, w_k) - \lambda\,\Omega(a) \tag{4}$$

This formulation encourages the selection of architectures that are resistant to weight initialization and efficient in terms of structure. An important feature of the framework is the distinction between architecture search and weight optimization. In the evolutionary stage, the architectures are assessed using simplified forward passes. However, the best architectures are fully optimized using a gradient-based optimization method. Multiple top-performing architectures are retained, fine-tuned separately, and the best model is selected based on the validation performance.

D. Evaluation Protocol

The evaluation process is not limited to traditional classification evaluation metrics but also includes the evaluation of the compactness and efficiency of the models. Besides the evaluation of the models in terms of their performance, i.e., Macro-F1 and balanced accuracy, we also evaluate them on the basis of the number of parameters, scores, and compression rates compared to larger neural baseline models. Models are also evaluated with different parameter budget constraints. This framework is used to evaluate the models comprehensively in terms of performance and efficiency, which is an important aspect of designing low-resource intelligent educational systems.

E. Implementation Details

All multilayer perceptrons (MLPs) employ ReLU activation functions and are optimized using the Adam algorithm with a learning rate of 0.001.

The architectures of the models are defined as follows:

- *MLP_micro*: one hidden layer with 16 units
- *MLP_tiny*: one hidden layer with 32 units
- *MLP_small*: two hidden layers with 64 and 32 units
- *MLP_medium*: two hidden layers with 128 and 64 units
- *MLP_large*: three hidden layers with 256, 128, and 64 units

To mitigate overfitting, a dropout rate of 0.2 is applied.

Knowledge distillation is performed using the *MLP_large* model as the teacher. Student networks are trained using a combined objective function, where cross-entropy loss and distillation loss are equally weighted, with a temperature parameter $T = 2$.

In the neuro-evolution approach, the population size and the number of iterations are set to 80 and 60, respectively. Three mutation operators are considered: connection addition, connection deletion, and weight adjustment. In Equations (3) and (4), the parameter λ is set to 0.01.

For all experiments, multiple random seeds are used, and the reported results correspond to the average performance.

IV. Results

In the following section, the results of the experimental analysis based on the comparative evaluation of various controller architectures are presented. The analysis of the results is focused on the predictive performance as well as the efficiency of the models, specifically regarding the trade-off between the compactness of the model and its effectiveness in the selection of pedagogical interventions.

We first discuss the overall performance of all models under various feature configurations. As presented in Table 1, the results are reported based on macro-F1, balanced accuracy, as well as weighted F1-score metrics, averaging over various random seeds during the experiments. As seen, multilayer perceptron (MLP) models are confirmed to perform better than classical machine learning models, capturing complex interactions in the data of learners. Specifically, it is clear that medium-sized neural networks achieve the highest performance, implying that beyond a certain model size, the performance gains are minimal.

Another important observation from Table 1 is that the compact neural models manage to retain a significant amount of performance from the larger models. For example, the performance of the smaller MLP models is comparable to larger models, yet they have significantly fewer parameters. The performance difference between the compact and larger neural models is relatively small, which indicates that a significant amount of compression can be obtained without a significant loss of performance.

In order to further evaluate the efficiency, Table 2 presents the parameter count, compression ratio with respect to a large neural model, and performance retention. The results clearly show that the compact neural models, particularly the MLP models, have a high compression ratio while retaining a significant amount of performance. On the other hand, classical approaches, such as logistic regression and ensemble methods, have comparable performance but lack the flexibility that can be obtained from neural models.

To better illustrate the comparison between the proposed approach and existing methods, we have analyzed predictive performance and efficiency together. Compact MLP models have higher Macro F1 scores than classical models like Logistic Regression and HistGradientBoosting, i.e., up to 0.470, while maintaining competitive computational cost.

Moreover, when comparing compact MLP models with other neural models, they have over 95% predictive performance while reducing parameters up to 12×–95×. In contrast, the proposed Hybrid NEAT-WANN controller outperforms existing sparse search approaches in terms of compression, achieving up to 219× reduction in parameters. Even though its predictive performance is low compared to compact MLP models, it still provides an advantage over traditional approaches in extreme low-resource conditions where they become impractical.

Thus, unlike existing approaches, it is seen that the proposed approach allows for trade-off analysis, enabling the choice of the best controller based on conditions.

Although the Macro-F1 score appears moderate, this is expected given the complexity of the task and the reliance on proxy labels. The discrepancy between balanced accuracy and Macro-F1 can be attributed to class imbalance: while the models achieve relatively consistent recall across classes, precision remains lower for minority classes.

Macro-F1 is selected as the primary evaluation metric because it provides a more comprehensive assessment than accuracy and weighted F1, which tend to be biased toward majority classes.

From a practical perspective, the evaluated classifiers should be considered as decision-support components, as they are intended to operate alongside a supervisory mechanism rather than function as fully autonomous systems.

The performance of sparse neuroevolutionary controllers is also assessed. As may be seen in Table 1, the models learned through the random sparse search method, the NEAT evolutionary algorithm, and the proposed hybrid NEAT-WANN method have a reduced performance in terms of prediction. However, the models have a reduced number of parameters, as may be seen in Table 2. This makes the models competitive in extreme compression scenarios where resource constraints are considered. Among the sparse methods, the proposed method has a competitive performance compared to other evolutionary methods, especially in feature sets where risk indicators are removed. Figure 2 illustrates the performance-parameter trade-offs for all architectures. The

TABLE I
MODEL PERFORMANCE ACROSS FEATURE SETS

Feature Set	*Model*	*Macro-F1*	*Balanced Accuracy*	*Weighted F1*
Full	RandomForest	0.467	0.467	0.923
	MLP_medium	0.470	0.460	0.915
	MLP_small	0.444	0.434	0.926
	MLP_large	0.465	0.442	0.902
	MLP_tiny	0.443	0.417	0.921
	MLP_micro	0.409	0.389	0.896
	LogisticRegression	0.378	0.630	0.764
	RuleBased	0.354	0.702	0.692
	HistGradientBoosting	0.348	0.349	0.851
	RandomSparseControllerSearch	0.294	0.328	0.807
	Hybrid_NEAT_WANN_sparseMLP	0.271	0.280	0.689
	NEAT_only_sparseMLP	0.289	0.279	0.702
	Majority	0.150	0.200	0.450
No Risk	MLP_large	0.435	0.468	0.878
	RandomForest	0.451	0.455	0.899
	MLP_medium	0.426	0.423	0.892
	MLP_small	0.421	0.413	0.904
	MLP_tiny	0.403	0.392	0.892
	MLP_micro	0.375	0.373	0.876
	LogisticRegression	0.369	0.596	0.752
	HistGradientBoosting	0.341	0.339	0.838
	RandomSparseControllerSearch	0.288	0.325	0.802
	Hybrid_NEAT_WANN_sparseMLP	0.293	0.316	0.789
	NEAT_only_sparseMLP	0.290	0.282	0.689
	RuleBased	0.139	0.463	0.200
	Majority	0.150	0.200	0.450

figure demonstrates a performance-parameter trade-off, with the compact MLPs operating in a desirable region. The sparse neuroevolutionary controllers operate in the low-parameter regime, achieving decent performance with significantly fewer parameters. The figure indicates that there exist a number of desirable design points, depending on the application-specific resource constraints.

To offer a more structured comparison, the models have been categorized according to the parameter budget tier, as shown in Table 3. This allows the models to be compared within a similar resource constraint. From the results, it is evident that the compact MLP models perform well in small to moderate budget regimes, whereas the sparse evolutionary models become more applicable in extremely budget-constrained regimes. This provides a clearer perspective on the models that may be more applicable in different deployment regimes.

In addition to these quantitative measures, we also examine the performance of the top-performing models by using a confusion matrix as well as qualitative analysis. Figure 3 shows the confusion matrix of the top-performing controller, indicating its effectiveness in distinguishing between various types of intervention. Although all the classes are well represented, some degree of confusion is noticed between various types of intervention, indicating the complexity of the task itself. Table 4 below shows qualitative analysis of the model's predictions during various states of learners. The results indicate that the model is able to predict the types of intervention required by learners based on the level of risk as well as recent performance. For example, learners who are at high risk but have low recent scores are correctly identified as requiring resource recommendation types of intervention. Although some degree of confusion is noticed between various types of intervention, it is evident that the model is able to handle the complexities of the task itself.

Thus, it can be concluded that even though compact neural models perform best in terms of predictive accuracy, the proposed hybrid approach based on NEAT and WANN possesses an added advantage in situations where extreme compression is a requirement.

Finally, the convergence property of the neuroevolutionary search process is depicted in Figure 4. The findings demonstrate the stable convergence property of the proposed hybrid NEAT-WANN model, where the fitness value improves over the generations. Though the final fitness value is not high, the model is able to evolve compact and functional neural architectures, thus proving the efficiency of the search process.

V. DISCUSSION

The results also provide insights for the design of pedagogical intervention controllers in intelligent educational systems. In particular, the results do not point to a single optimal architecture, but rather the need for alignment of model design with deployment constraints and system-level objectives.

In comparison to the existing techniques in the literature, which focus primarily on the accuracy of the predictive results, the current study has explicitly incorporated the concept of efficiency as an evaluation metric. While the existing techniques are based on the use of large neural structures or pre-defined models, the current study has demonstrated the feasibility of achieving comparable results while keeping the size of the model much smaller, as in the case of the hybrid NEAT-WANN technique, which extends the concept of compression to include efficient structures rather than the existing ones.

Compact feedforward neural networks are identified as a strong candidate architecture, demonstrating high prediction

TABLE II
MODEL EFFICIENCY AND COMPRESSION ANALYSIS

Feature Set	*Model*	*Params*	*Compression Ratio*	*Performance Retention*
Full	MLP_large	60,421	1.0×	1.000
	MLP_medium	18,053	3.35×	1.010
	MLP_small	5,061	11.94×	0.956
	MLP_tiny	1,269	47.61×	0.953
	MLP_micro	637	94.85×	0.880
	RandomForest	365	165.54×	1.000
	LogisticRegression	365	165.54×	0.814
	HistGradientBoosting	365	165.54×	0.748
	NEAT_only_sparseMLP	740	81.65×	0.706
	RandomSparseControllerSearch	1,256	48.11×	0.633
	Hybrid_NEAT_WANN_sparseMLP	275	219.71×	0.583
No Risk	MLP_large	60,165	1.0×	1.000
	MLP_medium	17,925	3.36×	0.979
	MLP_small	4,997	12.04×	0.967
	MLP_tiny	1,253	48.02×	0.925
	MLP_micro	629	95.65×	0.862
	RandomForest	360	167.13×	1.036
	LogisticRegression	360	167.13×	0.848
	HistGradientBoosting	360	167.13×	0.784
	NEAT_only_sparseMLP	325	185.12×	0.665
	RandomSparseControllerSearch	260	231.40×	0.661
	Hybrid_NEAT_WANN_sparseMLP	279	215.65×	0.674

TABLE III
PERFORMANCE BY PARAMETER BUDGET

Budget Level	*Models*	*Key Observation*
High ($>$ 10k params)	MLP_large, MLP_medium	Highest performance, diminishing returns beyond medium size
Medium (1k–10k)	MLP_small, MLP_tiny	Best trade-off between performance and efficiency
Low ($<$ 1k)	MLP_micro, NEAT-based models	Noticeable drop in performance, but strong compression
Extreme ($<$ 500)	Hybrid_NEAT_WANN, RandomSparse	Maximum compression with acceptable but limited performance

TABLE IV
QUALITATIVE EXAMPLES OF MODEL PREDICTIONS

Week	*Risk Score*	*Recent Activity*	*Recent Score*	*True Intervention*	*Predicted Intervention*
2	0.744	Moderate	Low	RESOURCE_RECOMMEND	RESOURCE_RECOMMEND
4	0.787	Moderate	Low	EXPLAIN_SIMPLE	RESOURCE_RECOMMEND
5	0.542	Moderate	Medium	EXPLAIN_SIMPLE	EXPLAIN_SIMPLE
7	0.611	Moderate	Low	RESOURCE_RECOMMEND	RESOURCE_RECOMMEND
9	0.690	Moderate	Medium	EXPLAIN_SIMPLE	EXPLAIN_SIMPLE
12	0.739	Moderate	Low	RESOURCE_RECOMMEND	RESOURCE_RECOMMEND

capability with low parameter count. This implies that the intervention selection problem is well-suited for moderate capacity models, potentially arguing against the need for overly complex architectures.

On the other hand, sparse neuroevolutionary controllers propose a complementary approach that focuses on extreme compression. While they do not compete with the prediction capability of the compact neural network baselines, they achieve significantly smaller architectures with non-trivial prediction capability. The proposed NEAT-WANN framework shows that constrained architecture search can identify efficient models that would be difficult to design manually.

This shows that controller development should be treated as a multi-objective optimization task. While focusing on predictive performance alone could miss better options under resource constraints, including parameters such as the number of parameters or the compression ratio offers a more complete evaluation, as seen in recent trends in efficiency-aware machine learning. Comparing the fixed and search-based methods also shows differences in stability between them. While the small neural network controllers show consistent performance, the evolutionary method shows more variation in results, reflecting the random nature of the optimization method. This shows the importance of multi-seed evaluations, suggesting that a combination of methods could offer better results. The use of knowledge distillation offers improvements over the small neural network method by using larger networks as teachers. This method, however, has to be carefully configured based on the capabilities of the student networks, especially in highly constrained settings. The use of proxy labels allows supervised learning without needing actual pedagogical action labels, while still offering interpretability of results. The results should be seen as modeling the proxy labeling scheme rather than actual instructional effectiveness.

One key insight derived from the confusion matrix (Figure 3) is the imbalanced distribution of classes, where cer-

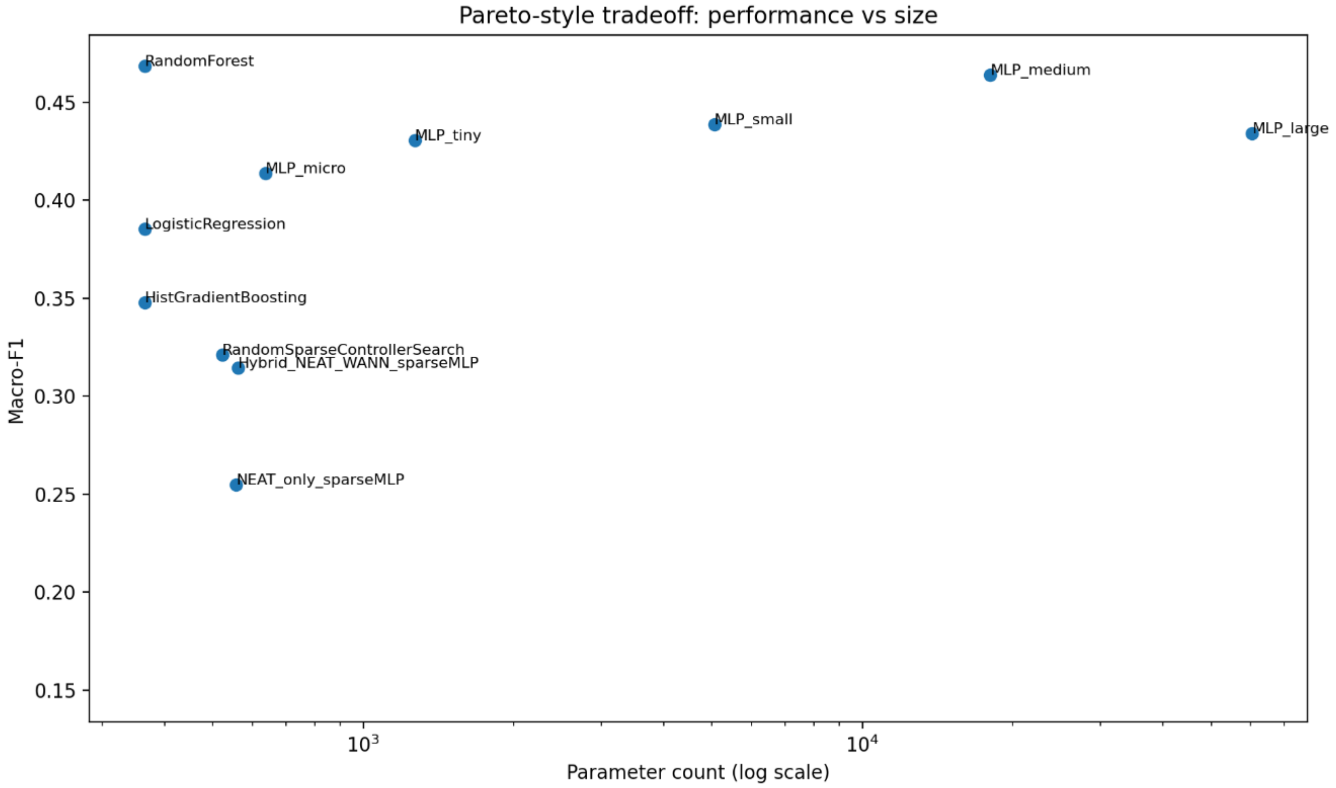

Fig. 2. The relationship between model performance (macro-F1) and parameter count on a logarithmic scale.

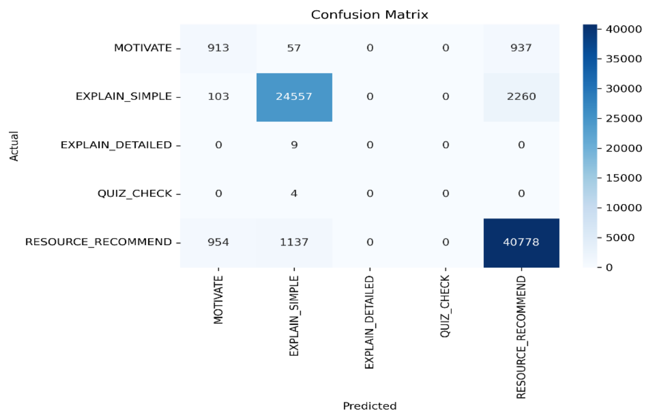

Fig. 3. Confusion matrix of the top-performing controller, illustrating its ability to distinguish between different pedagogical intervention classes.

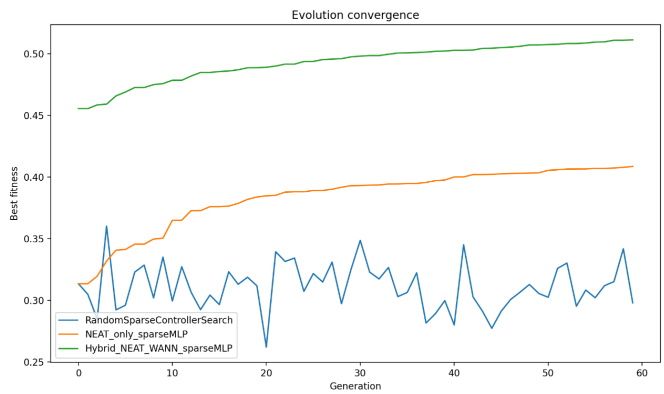

Fig. 4. The convergence behaviour of the evolutionary search process.

tain intervention types, such as EXPLAIN_DETAILED and QUIZ_CHECK, are underrepresented. This imbalance originates from the proxy label generation process, which inherently reflects skewed patterns in learner behavior. Therefore, the imbalanced nature of the distribution of proxy classes is the reason for certain prediction behaviors.

Class balancing and cost-sensitive learning techniques were not explicitly employed in this study, as the primary objective was to evaluate model performance under realistic conditions. However, this limitation introduces a potential bias, as the controller may favor dominant intervention types.

Future work should investigate strategies to address class imbalance, including the use of weighted loss functions and data sampling techniques, to improve the robustness and fairness of the model.

From a systems perspective, the controller is a light-weight decision component in a modular architecture. Optimizing this component helps to reduce the overall computational cost, especially if combined with more complex components, like a retrieval module or a language model. Overall, the experiments show that the proper optimization of a controller requires a balance between performance and efficiency. Neural models offer a powerful general solution, whereas sparse neuroevolutionary models offer a promising alternative in resource-constrained scenarios. This provides valuable insights for the construction of efficient intelligent educational systems.

VI. Conclusion

The focus of this paper was to explore the design of low-resource controllers for the selection of pedagogical interventions, specifically for intelligent educational systems, with a focus on predictive performance and computational efficiency. The focus was not to introduce a particular model, but rather to conduct a comparative analysis of different approaches, including classical machine learning, compact neural networks, and sparsely connected controllers through neuro-evolutionary search methods. The results show that compact neural network architectures, including small and medium-sized multilayer perceptrons, are a strong baseline since they are able to deliver high predictive performance while maintaining low parameter count. Additionally, sparse neuro-evolutionary search methods are shown to be a viable alternative for extreme compression, delivering high compression while maintaining predictive performance. The contribution of this paper is that it introduces the problem of low-resource controller design for the selection of pedagogical interventions and conducts a thorough analysis of performance and efficiency, specifically for intelligent educational systems.

Despite these contributions, some limitations are noted. Firstly, the intervention labels are based on heuristic proxy rules rather than ground-truth pedagogical decisions. This does not necessarily capture the true effectiveness of the intervention. Secondly, while the proposed method for hybrid neuroevolutionary optimization does result in significant compression, it does not quite live up to the state-of-the-art of the best performing small neural networks. Thirdly, the evaluation is performed on just one dataset. Future work will seek to address these limitations by investigating more sophisticated optimization techniques using hybrids of evolutionary optimization and gradient optimization. Also, the evaluation will be expanded to more datasets. More sophisticated learner models, incorporating temporal and sequential models, could also be used to further improve the intervention selection. Finally, the incorporation of deployment constraints such as latency and energy could be used to develop more practical and adaptable controllers.

References

[1] M. O. González-Morales, D. López-Aguilar, P. R. Álvarez-Pérez, and P. A. Toledo-Delgado, "Dropping out of higher education: Analysis of variables that characterise students who interrupt their studies," *Acta Psychologica*, vol. 252, p. 104669, 2025.

[2] A. U. Khan, K. U. Khan, F. Atlas, S. Akhtar, and F. Khan, "Critical factors influencing MOOCs retention: The mediating role of information technology," *Turkish Online Journal of Distance Education*, vol. 22, no. 4, pp. 82–101, 2021.

[3] S. Wang and S. K. Lee, "Exploring the role of learning goal orientation, instructor reputation, parasocial interaction, and tutor intervention in university students' MOOC retention: A TAM-TRA based examination," *PLoS One*, vol. 19, no. 9, p. e0299014, 2024.

[4] C. Troussas, A. Krouska, and C. Sgouropoulou, "Impact of social networking for advancing learners' knowledge in E-learning environments," *Education and Information Technologies*, vol. 26, no. 4, pp. 4285–4305, 2021.

[5] J. A. V. Mendoza, C. F. Ramírez, and C. M. Aranda, "Analysis and discovery of procrastination patterns in a language learning MOOC," *Computers & Education*, vol. 223, p. 105154, 2024.

[6] X. Wei, Y. Chen, J. Shen, and L. Zhou, "Fail or pass? Investigating learning experiences and interactive roles in MOOC discussion board," *Computers & Education*, vol. 217, p. 105073, 2024.

[7] Y. Hao, "Multi-teacher knowledge distillation for lightweight speech interaction in embedded educational robots," *Informatica*, vol. 50, no. 5, 2026.

[8] B. Tokoli, L. Jaimes, and A. S. Dina, "ALIGNAgent: Adaptive learner intelligence for gap identification and next-step guidance," *arXiv preprint arXiv:2601.15551*, 2026.

[9] E. Papavasileiou, J. Cornelis, and B. Jansen, "A systematic literature review of the successors of neuroevolution of augmenting topologies," *Evolutionary Computation*, vol. 29, no. 1, pp. 1–73, 2021.

[10] A. Gaier and D. Ha, "Weight agnostic neural networks," in *Proceedings of Advances in Neural Information Processing Systems*, vol. 32, 2019.

[11] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, "Predicting students' dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100066, 2022.

[12] K. Talebi, Z. Torabi, and N. Daneshpour, "Ensemble models based on CNN and LSTM for dropout prediction in MOOC," *Expert Systems with Applications*, vol. 235, p. 121187, 2024.

[13] M. Vitiello, S. Walk, D. Helic, V. Chang, and C. Guetl, "User behavioral patterns and early dropouts detection: Improved users profiling through analysis of successive offering of MOOC," *Journal of Universal Computer Science*, vol. 24, no. 8, pp. 1131–1150, 2018.

[14] R. Al-Shabandar, A. J. Hussain, P. Liatsis, and R. Keight, "Detecting at-risk students with early interventions using machine learning techniques," *IEEE Access*, vol. 7, pp. 149464–149478, 2019.

[15] M. Sultani and N. Daneshpour, "Extracting urgent questions from MOOC discussions: A BERT-based multi-output classification approach," *Arabian Journal for Science and Engineering*, vol. 50, no. 2, pp. 1169–1190, 2025.

[16] S. X. Guo, X. Sun, S. X. Wang, Y. Gao, and J. Feng, "Attention-based character-word hybrid neural networks with semantic and structural information for identifying urgent posts in MOOC discussion forums," *IEEE Access*, vol. 7, pp. 120522–120532, 2019.

[17] H. Liu, X. Chen, and F. Zhao, "Learning behavior feature fused deep learning network model for MOOC dropout prediction," *Education and Information Technologies*, vol. 29, no. 3, pp. 3257–3278, 2024.

[18] Y. Ouyang, H. Long, R. Gao, and J. Liu, "Course recommendation model based on layer dropout graph differential contrastive learning," *IEEE Access*, vol. 12, pp. 7762–7774, 2024.

[19] X. Yu et al., "CR-LCRP: Course recommendation based on learner–course relation prediction with data augmentation in a heterogeneous view," *Expert Systems with Applications*, vol. 249, p. 123777, 2024.

[20] S. Li, J. Du, and S. Yu, "Diversified resource access paths in MOOCs: Insights from network analysis," *Computers & Education*, vol. 204, p. 104869, 2023.

[21] C. Blundo, V. Loia, and F. Orciuoli, "A time-aware approach for MOOC dropout prediction based on rule induction and sequential three-way decisions," *IEEE Access*, vol. 11, pp. 113189–113198, 2023.

[22] M. Y. Boyajian, "Student intervention system using machine learning," Ph.D. dissertation, 2019.

[23] B. Khosrawi-Rad et al., "Conversational agents in education—A systematic literature review," 2022.

[24] H. Cheng, M. Zhang, and J. Q. Shi, "Mini-LLM: Memory-efficient structured pruning for large language models," *arXiv preprint arXiv:2407.11681*, 2024.

[25] X. Ma, G. Fang, and X. Wang, "LLM-Pruner: On the structural pruning of large language models," in *Proceedings of Advances in Neural Information Processing Systems*, vol. 36, pp. 21702–21720, 2023.

[26] K. O. Stanley, J. Clune, J. Lehman, and R. Miikkulainen, "Designing neural networks through neuroevolution," *Nature Machine Intelligence*, vol. 1, no. 1, pp. 24–35, 2019.

[27] S. Drissi, A. Chefrour, K. Boussaha, and H. Zarzour, "Exploring the effects of personalized recommendations on student's motivation and learning achievement in gamified mobile learning framework," 2024.

[28] C. Hu et al., "Teacher-student architecture for knowledge distillation: A survey," *arXiv preprint arXiv:2308.04268*, 2023.

Houssam Ahmed Amin Bahi is a Ph.D. researcher in Artificial Intelligence with the Department of Mathematics and Computer Science and the Research Laboratory on Computer Science's Complex Systems (ReLa(CS)^2) at the University of Oum El Bouaghi, Algeria. He received the B.S. and M.S. degrees in computer science from the same university in 2019 and 2021, respectively. His research focuses on time-series modeling, predictive analytics, and intelligent systems, with particular emphasis on developing novel feature representations and difficulty-aware learning frameworks for early prediction in complex data environments. He has authored multiple international conference papers and journal submissions, with contributions spanning feature engineering, model evaluation, and optimization methods. His interests also include the application of artificial intelligence to real-world systems involving temporal and multimodal data, particularly in the context of e-learning environments.

Karima Boussaha is an Associate Professor in the Department of Mathematics and Computer Science at Oum El Bouaghi University in Algeria. She is obtained her Ph.D. at Annaba University in Algeria since 2016. She is a Member in Artificial Intelligence and Autonomous Things Laboratory(LIAOA) at Oum El Bouaghi University. Her Ph.D. research is based on computer-assisted assessment of learners, application in the programming language field. She obtained her HU (Habilitation Universitaire) in Computer Science from Oum El Bouaghi university in Algeria since 2021. She has served as a member of the Editorial Board for numerous other international journals and has contributed as a Program Committee chair or member for various international conferences. She has published extensively in international journals, books and conference proceedings. Her main areas of interest include computer-based/assisted assessment, assessment of procedural knowledge, and assessment in collaborative learning, computer-based/assisted collaborative tutoring, computer-based/assisted collaborative learning, Ontologies, collaboration Learner's motivation, MOOC platforms, Game based learning, Artificial intelligence, maching learning, deep learning.

Zakaria Laboudi is a full Professor at University of Oum El Bouaghi, Algeria. He received his M.Sc degree in computer sciences from University of Constantine1, Algeria in 2009. He holds a Ph.D degree in Computer Sciences from University of Constantine2, Algeria since 2017. He also holds Habilitation degree (known also as Accreditation to Supervise Research) from University of Oum El Bouaghi, Algeria since 2019. Previously, he joined the MISC Laboratory at University of Constantine2, Algeria, and the ReLa (CS)2 Laboratory at University of Oum El Bouaghi, Algeria, for the period from 2010 to 2022, in which he was invited to collaborate in several research projects and tasks. His current research interests include artificial intelligence, machine learning, optimization, and data analysis, ontology-based computing, context-aware pervasive systems, ubiquitous learning, and multimedia adaptation.