Heliyon 9 (2023) e14506

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Preliminary evidence on machine learning approaches for clusterizing students' cognitive profile

Matteo Orsoni *, Sara Giovagnoli, Sara Garofalo, Sara Magri, Martina Benvenuti, Elvis Mazzoni, Mariagrazia Benassi

Department of Psychology, University of Bologna, Italy

ARTICLE INFO

Keywords:
Machine learning
Self-organizing maps
Cognitive profiling
Specific learning difficulties (SLD)
K-means

ABSTRACT

Assessing the cognitive abilities of students in academic contexts can provide valuable insights for teachers to identify their cognitive profile and create personalized teaching strategies. While numerous studies have demonstrated promising outcomes in clustering students based on their cognitive profiles, effective comparisons between various clustering methods are lacking in the current literature.

In this study, we aim to compare the effectiveness of two clustering techniques to group students based on their cognitive abilities including general intelligence, attention, visual perception, working memory, and phonological awareness. 292 students, aged 11–15 years, participated in the study.

A two-level approach based on the joint use of Kohonen's Self-Organizing Map (SOMs) and k-means clustering algorithm was compared with an approach based on the k-means clustering algorithm only. The resulting profiles were then predicted via AdaBoost and ANN supervised algorithms.

The results showed that the two-level approach provides the best solution for this problem while the ANN algorithm was the winner in the classification problem.

These results laying the foundations for developing a useful instrument for predicting the students' cognitive profile.

1. Introduction

Specific cognitive functions are typically assessed to explain learning heterogeneity in students, particularly for those with atypical development [1–3]. Indeed, even if students with atypical development are grouped in exact diagnostic group by means of specific diagnostic criteria, it is possible to find cognitive subgroups within each diagnosis aiming to personalize teaching interventions [4,5].

In research on typical developmental populations, cluster analysis techniques have been used. This is a type of multivariate analysis that helps to classify subjects into groups that are internally highly homogeneous and externally highly heterogeneous. Depending on the variables used to cluster the subjects, different solutions have been proposed to describe the cognitive profiles of students.

For instance, Yokota et al. [6] used a k-means clustering technique that considered four factors (verbal comprehension, perceptual organization, freedom from distractibility, and processing speed) of the Wechsler Intelligence Scale for Children (Third Edition). The authors discovered the presence of six cognitive subtypes that differed in verbal comprehension, perceptual organization, processing

* Corresponding author. Piazza Aldo Moro 90, 47521 Cesena, Italy.
E-mail address: matteo.orsoni2@unibo.it (M. Orsoni).

https://doi.org/10.1016/j.heliyon.2023.e14506
Received 12 May 2022; Received in revised form 17 February 2023; Accepted 9 March 2023
Available online 16 March 2023

speed, and distractibility. The interesting part is that the authors validated the clustering solution by confirming differences among clusters based on fMRI measures. The study results suggest that cognitive profiles can differentiate children with typical development, and these differences may be reflected in specific neural patterns.

Recently, Poletti et al. [7] conducted a cluster analysis study on ten core subtests of the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) and identified four subgroups of students with Specific Learning Difficulties (SLD). These subgroups differed in their performance on the WISC-IV subtests, particularly in the areas of verbal comprehension, coding, and executive functions. The authors also observed that while impairments in reading and mathematics were associated with low reasoning and executive functioning, difficulties in written expression were linked to low verbal and coding abilities.

Despite the relevant evidence obtained from different studies, results from cluster analysis could be sometime considered unsatisfactory in terms of rigor in methodology and replicability [8].

A major limitation of previous studies is that they lacked two crucial steps for clustering techniques, namely, comparison between different clustering solutions and validation of the selected cluster solution [9,10]. In Yokota et al. [6], only k-means clustering method was used, and the number of clusters was selected arbitrarily by progressively increasing them until a minimum of one cluster containing less than 10% of the sample appeared. Although this method has an advantage of including a parsimonious selection criterion, it may not result in the most meaningful clusters. In Poletti et al. [7], the authors relied on visual inspection of the agglomeration coefficients and dendrogram figure to identify the best cluster solution. Although they acknowledged the possibility of using multiple methods, their approach focused solely on supporting a single solution, rather than comparing various methods using statistical indices. Additionally, the authors evaluated agreement between clustering solutions using Cohen's kappa and Intraclass correlation coefficient but did not assess the accuracy of the proposed solution. While the method of comparing two solutions is intriguing, it fails to determine which of the two is superior, thus hindering a meaningful comparison. To date, as far as we know, no study has explicitly aimed to compare various clustering techniques and assess the feasibility of implementing efficient and replicable methods for the topic at hand. Therefore, further research is needed to investigate and compare multiple clustering approaches, while also evaluating their effectiveness and reproducibility.

In this study, we aim to cluster different cognitive profiles of secondary school students by using a two-level approach based on the joint use of an artificial neural network, the Kohonen's Self Organizing Maps [11], and the k-means clustering algorithm. The proposed approach would be beneficial in: 1. Allowing to compare different clustering approaches and select the best one on statistical index supporting the choice; 2. Enhancing the clustering solution accuracy.

Recent findings in non-psychological fields showed an improvement in the clustering solution by applying the Kohonen's Self Organizing Maps (SOMs) [11] before the k-means or hierarchical clustering implementation compared to the clustering methods only [12–14]. In this study, we evaluated the performance of k-means clustering with and without a SOM-based pre-processing step. We selected this clustering algorithm based on previous research [12–14], which showed superior performance compared to other algorithms such as hierarchical clustering. Despite the advantages of this approach, its application in the field of psychology, particularly in cognitive profiling, is still limited.

Our objective was to identify the optimal clustering solution by comparing two clustering methods, based on cognitive functions that previous research had identified as the most distinguishing between students with SLD and controls, while also supporting academic skills. Specifically, we focused on executive functions, language, and visual perception abilities (logical reasoning, visual attention, visual perception, verbal comprehension, and working memory), as reported in studies [15–23]. Next, we developed and tested the validity of a machine learning (ML) model to determine whether cognitive profiles could be accurately predicted by the model. This final step can serve to verify the replicability of the selected clustering solution.

2. Material and methods

2.1. Participants

To recruit participants for the study, several schools in the Emilia-Romagna Region of Italy were invited to participate. Of those invited, four consented to the use of an online digital game for cognitive assessment (see Cognitive Assessment section), and three of those also agreed to standardized battery tests. This resulted in a total sample of 292 secondary school students (104 females (36%), age range: 11–15 years) for cluster analysis. Of these, 99 (33.96%) were attending secondary school, while 193 (66.04%) were attending secondary high school. All participants were Italian, and 12 (4.11%) were bilingual.

Form the total, a subgroup of 105 students was selected for full clinical data collection, with 71 (29 females (41%), age range: 11–14 years) attending secondary school and 34 (13 females (38%), age range: 14–15 years) attending secondary high school. This subsample was assessed by four psychologists using standardized battery tests to evaluate the cluster solution's ability to differentiate between typically and atypically developing subjects. Of these, 30 (28.6%) met the criteria for a specific learning difficulties (SLD), with 7 having dyslexia, 7 having dyscalculia, and 16 exhibiting multiple disorders.

2.2. Cognitive assessment

All students cognitive abilities have been assessed by an online digital game called PROFFILO developed for the assessment of the student's cognitive profile [24] and with standardized tests for logical reasoning (Raven's Progressive Matrices) [25], working memory (WISC-IV Inverse SPAN) [26], and Visual Attention (NEPSY-II Visuospatial Attention subtest) [27].

PROFFILO was administered in class to the students and lasted 20/25 min. It is composed of five different sub-tests (games), each developed for the assessment of a specific cognitive function (logical reasoning, visuospatial attention, motion perception, phonological awareness, and working memory). A previous study showed a good correlation between these games and standardized tests for the evaluation of the same clinical functions, with the only exception of the phonological awareness game that, for this reason, was excluded from the subsequent analyses [24]. The tests used showed good convergent validity with standardized tests (see Supplementary Materials).

2.3. Reading, spelling and math assessment

Students aged 11–14 were assessed for reading and spelling abilities by standardised italian reading test, DDE-2 and for math abilities, by AC-MT [28]. Students aged above 14–15 years, were assessed by means of Advanced MT-3 battery test [28] both for reading and math abilities. The student was considered as having SLD when her/his standardized reading or spelling or math score was below 2SD, while the general intelligence evaluated by Raven's Matrices [29] was within the normative range (scoring above 25th centile). As documented by an interview with teachers and parents, all participants had no evidence of brain injury, socio-cultural detriment or relevant behavioral problems.

In Table 1, one-way ANOVAs have been carried out to inspect differences between the.

SLD group, and no-SLD in reading and arithmetic clinical tests.

2.4. From raw data to clustering: preprocessing, Self-Organizing Maps and K-means

To improve algorithm performance, we used Min-Max normalization as a preprocessing procedure on the sample. This procedure allowed for unifying the feature's orders of magnitude [30]. It performs a linear transformation on the original data, mapping a value in a range between [0,1], and it is not dependent on the distribution of the variable [31,32]. Additionally, before applying the k-means algorithm, we used Self-Organizing Maps (SOMs) for a subsequent clusterization. SOMs are competitive or unsupervised artificial neural networks that provide a topological representation of the input data [11]. A thorough explanation can be found in the supplementary materials.

The pseudo-code (Algorithm 1) explains the SOM implementation.

The k-means method [33] was then applied to find the best clustering solution. The optimal number of clusters was selected using the Elbow method [34]. We implemented two k-means cluster algorithms, with 1000 as the maximum number of iterations allowed and 100 as the number of random sets chosen. The cluster's quality was evaluated by qualitatively inspecting the cases gathered in each cluster, by reviewing all clusters by hand to evaluate the meaning of the membership of each data to a given cluster.

Following [14], the cluster accuracy index I_c was calculated. It refers to a single cluster and takes the form of Eq. (1), where a_v is the number of correctly assigned cases and n_c is the number of cases grouped in the cluster.

Table 1
z score in all the reading and math tests in SLD and no-SLD group (Mean and SE are reported). One-way ANOVAs for comparing SLD and no-SLD groups in reading, spelling, and math tests.

Age group	Test	SLD	no-SLD	$F_{(1,103)}$	p	η^2
11–14	Reading tests					
	Word reading Speed	−2.863(0.349)	−0.267(0.226)	38.978	<0.001	0.361
	Word reading Accuracy	−2.619(0.478)	0.040(0.310)	21.766	<0.001	0.240
	Word reading Speed	−2.573(0.376)	−0.093(0.241)	30.876	<0.001	0.306
	Word reading Accuracy	−0.524(0.252)	0.745(0.162)	17.956	<0.001	0.204
	Math tests					
	ACMT1a	−0.614(0.214)	0.333(0.137)	13.918	<0.001	0.166
	ACMT2a	−1.818(0.268)	−0.359(0.172)	21.048	<0.001	0.231
	ACMT3a	−1.327(0.275)	−0.328(0.176)	9.353	0.003	0.118
	ACMT4a	−1.402(0.229)	0.378(0.147)	42.850	<0.001	0.380
	ACMT1v	−1.657(0.271)	0.009(0.178)	26.394	<0.001	0.280
	ACMT2v	−2.245(0.405)	−1.140(0.262)	5.243	0.025	0.071
14–15	Reading tests					
	Word reading Speed	−3.576(0.375)	−0.531(0.225)	48.568	<0.001	0.603
	Word reading Accuracy	−1.016(0.318)	0.263(0.191)	11.911	0.002	0.271
	Word reading Speed	−1.924(0.315)	−0.044(0.189)	26.239	<0.001	0.451
	Word reading Accuracy	−0.847(0.300)	0.125(0.180)	7.711	0.009	0.194
	Math tests					
	MT3a	−0.198(0.292)	0.767(0.175)	8.042	0.008	0.201
	MT3t	−1.355(0.368)	−0.154(0.221)	7.831	0.009	0.197
	MT3af	−0.693(0.302)	0.120(0.181)	5.344	0.027	0.143

Note. ACMT1a: mental calculation accuracy; ACMT2a: written calculation accuracy; ACMT3a: numeracy test; ACMT4a: arithmetic facts retrieval; ACMT1v: mental calculation time; ACMT2v: written calculation time; MT3a: accuracy; MT3t: time; MT3fa: arithmetic facts.
The analyses showed that SLD and no-SLD groups were significantly different considering reading and math ability tests scores. The SLD group compared to the no-SLD performed the worst in all the tests administered.

$$I_c = \frac{a_v}{n_c} \tag{1}$$

We calculated the accuracy index by examining the number of variables that fell within the centroid's membership boundaries, as shown in Table 2. We graphically represented these boundaries for each cluster and variable, and considered a subject misclassified if it fell outside of them. If a subject was misclassified in more than two variables, the entire case was deemed misclassified by the algorithm.

2.5. From clustering to prediction: imbalance classification, AdaBoost and artificial neural networks

The cluster solution revealed the presence of various groups of different sizes, resulting in an imbalanced classification problem. Imbalance arises when one or more classes have significantly lower proportions in the training data than the other classes [35]. As a result, the impact of class imbalance on classification performance metrics is a significant concern [36].

To address the imbalance problem, we utilized the Synthetic Minority Over-Sampling Technique (SMOTE) proposed by Chawla et al. [37]. This approach combines up-sampling and down-sampling techniques that are determined by the class. Three parameters guided the SMOTE algorithm, including the amount of up-sampling, the amount of down-sampling, and the number of neighbors used to create new cases. During the up-sampling, SMOTE generated new cases by randomly selecting a data point from the minority class (es) and determining its K-nearest neighbors (KNNs). The new synthetic data point was a random combination of the selected data point predictors and its neighbors. Additionally, the SMOTE algorithm down-sampled cases from the majority class via random sampling to achieve balance in the training set [35]. The number of neighbors used in the algorithm implementation was set to 3. After the SMOTE implementation, we got a training sample of 540 subjects, 60 for each class. This sample characterized by real and synthetic data was used for the subsequent analyses where we compare the performances of two supervised ML algorithms in the prediction of our clusters both in the imbalanced dataset (only real data) and in the balanced (real and synthetic data) ones. The choice of using ML algorithms based on their best predictive ability than linear models [38,39].

After applying the Synthetic Minority Over-Sampling Technique (SMOTE) to balance the training set, we used the Adaptive Boosting algorithm (AdaBoost) [40] and a fully connected Artificial Neural Network (ANN) to predict the clusters emerged from the previous steps and test the replicability of the solution. We compared the performances of these two supervised ML algorithms, with AdaBoost being selected for its good performance in imbalance classification problems [41,42].

The Adaptive Boosting algorithm is an ensemble method that functions in a boosting network. Boosting is a technique that can significantly reduce the error of any weak learning algorithm to create classifiers that only need to be slightly better than random guessing [40]. The AdaBoost algorithm assigns weights to each sample based on its importance and places the most weight on those examples that are most frequently misclassified by the previous classifiers. This emphasis may cause the learner to produce an ensemble function that differs significantly from the single learning [42].

We implemented the algorithm on both balanced and imbalanced training samples. To prevent model overfitting and identify the optimal parameters, we performed 5-fold cross-validation and hyperparameter search on both samples. The tuning process involved two parameters: the number of estimators and the learning rate. For the number of estimators, the search range was set from 10 to 700 in increments of 10. For the learning rate, the search range was set from 0.0001 to 1 in increments of 0.1.

The second algorithm employed a fully connected artificial neural network (ANN) with two hidden layers. The first comprised 512 units, while the second contained 256 units. Rectified linear unit (ReLU) activation functions were used for the hidden layers, with the Softmax function employed for the output layer. The Adam optimizer was employed with the Sparse Categorical Crossentropy function utilized as the loss function. The Accuracy metric was implemented to evaluate the model. To prevent overfitting, the model was trained for 600 epochs with a dropout rate of 0.5. Furthermore, an early stopping callback was implemented. The algorithm was also tested on both the balanced and imbalanced training samples.

2.6. Evaluate the performances

To evaluate the performance of the model, the accuracy metric is commonly used. However, in cases of class imbalance, accuracy may not be a suitable measure as the minority class has little impact on the overall accuracy compared to the majority class [42].

Table 2
Centroids for each cluster and variable found after the implementation of k-means and SOM and clusters' numerosity for both the sample used in the k-means clustering.

Cluster- ID	Logical Reasoning	Visual Perception	Visuospatial Attention	Working Memory	N (%)
vhLR-aAll	0.17	0.28	0.53	0.55	52 (17.808%)
aALL-lWM	0.55	0.43	0.40	0.25	12 (4.110%)
vhLR-lWM	0.16	0.33	0.65	0.22	37 (12.671%)
vlVP-lVA	0.24	0.83	0.29	0.61	10 (3.424%)
vhLR-vlVA	0.18	0.30	0.13	0.46	23 (7.877%)
vhLRWM	0.16	0.39	0.54	0.93	22 (7.534%)
ahAll	0.21	0.29	0.76	0.52	75 (25.685%)
vhLR-lVP	0.18	0.67	0.68	0.54	19 (6.507%)
vhALL	0.10	0.18	0.79	0.81	42 (14.384%)

To address this issue, several other metrics can be derived from the confusion matrix to assess model performance. The confusion matrix compares the true classes with the predicted classes obtained from the model and can be used to calculate various error parameters based on the counts of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. These values form the basis for computing the Precision (PRE) and Recall (REC) error metrics. Precision (Eq. (2)) is the ratio of the number of correct predictions of an event (class) to the total number of times the model predicts it.

$$PRE = \frac{TP}{TP + FP} \tag{2}$$

The lower is the value of False Positive, higher is Precision.

Recall Eq (3) reflects the model's sensibility. It is the ratio of the correct predictions for a class of the total cases in which it occurs.

$$REC = \frac{TP}{TP + FN} \tag{3}$$

The lower is the value of False Negative, higher is Recall.

Usually, Precision and Recall are combined to obtain the F1 score. F1 Eq. (4) represents the harmonic mean between Precision and Recall:

$$F1 = \frac{2}{\frac{1}{R} + \frac{1}{P}} \tag{4}$$

In general, the harmonic mean of two numbers is closer to the smaller of the two. Therefore, having a high F1 score indicates that both the Recall and Precision are relatively high [42].

In this study, the performance of each class was evaluated based on metrics such as TP, TN, FP, FN, PRE, REC, and F1 scores. Moreover, the Balanced Accuracy Score and Weighted F1 Score were also calculated.

2.7. Software and packages

The analyses were carried on by using JASP statistical software [43], R v4.03 [44], and Python v3.8 [45]. On R, the 'kohonen' package was used for the Self-Organizing Maps implementation [45], and the k-means algorithm was carried out within the 'stats' base package. In Python, the 'scikit-learn' package [46] was used for the implementation of the Adaptive Boosting algorithm and the TensorFlow for the implementation of the Artificial Neural Network (ANN) [47].

3. Results

3.1. Self-organizing maps and K-means

To evaluate the reliability to use the SOM as pre-processing for a subsequent clusterization, we compared the variance explained between the k-means cluster after the SOM implementation and the clusterization held by the normalized data alone.

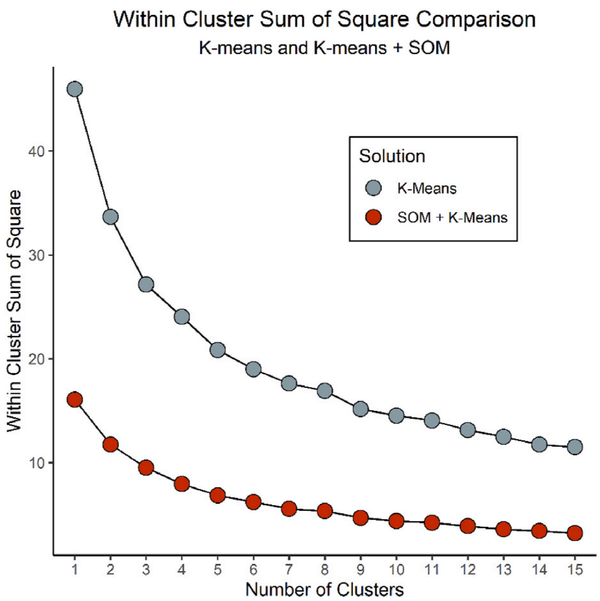

Fig. 1. Within Cluster Sum of Square (WCSS) solution for the k-means and k-means + SOM solutions.

The SOM algorithm was executed 200 times in order to minimize the mean distance between the codebook vector and the real vector (represented by the mean sum of square error, or SSE) to the closest unit on the map. The resulting mean SSE was 0.009. To visualize the distances in the original space, we employed the U-matrix method [48], and the resulting plot can be found in Supplementary Materials Fig.B1 (in color). This method calculates the average distances between the prototype vector of each cell and the prototype vectors of its neighboring cells, which are then represented by different color shades ranging from blue to red. Blue shades correspond to the smallest average distances, while red shades represent the largest ones.

To graphically display the properties of the variables, we created a property plot for each variable, which can be found in Supplementary Materials Fig.B.2 (in color). These plots allow for the visualization of the similarity of a particular object to all units on the map, as well as how these units are organized.

Before the k-means cluster implementation, the Elbow method was implemented both in the unit of the SOM and in the normalized data.

To determine the optimal number of clusters using the method described above, we conducted a k-means run for 45 steps, with a maximum of 1000 iterations allowed. As shown in Fig. 1 (in color), the results indicated that for the normalized data, the optimal number of clusters was 9, resulting in a within-cluster sum of squares of 15.15. However, when using the data that had been preprocessed by the SOM, we identified 9 clusters with a within-cluster sum of squares of 4.66. It is possible to evaluate how the cluster solution found implementing the SOM as pre-preprocessing improving the clusterization by reducing the WCSS compared to k-means alone.

The k-means implemented in normalized data and SOM showed the ratio between the sum of squares (BSS) and the total sum of squares (TSS) as equal to 0.673 (67.3%) and 0.712 (71.2%), respectively. This result outline how the hybrid approach (SOM + k-means) enhances the BSS/TSS ratio of 3.9%, highlighting how the clusterization preceded by the SOM is the one that best embodied the properties of internal cohesion and external separation explaining most of the variance.

In addition, both solutions were compared by using the BIC criterion. The results showed the hybrid approach as associated with the lowest BIC value (170.4), as compared to the single solution with k-means (219.4).

3.2. Clusters' description

Table 2 summarized the centroids' mean for each cluster and variable, along with their numerosity. To evaluate the performance, we proposed to consider different thresholds. In Visuospatial Attention, and Working Memory, subjects with values between 0 and 0.20 were classified as very low performance, 0.20-0.40 as low performance, 0.40-0.60 as average performance, 0.60-0.80 as high performance, and 0.80-1 as very high performance. In Logical Reasoning and Visual Perception, performance values are reversed. In other words, subjects with values between 0 and 0.20 were classified as very high performance, 0.20-0.40 as high performance, 0.40-0.60 as average performance, 0.60-0.80 as low performance, and 0.80-1 as very low performance. Moreover, these values allowed us to calculate the Cluster Accuracy Index as reported in Table 3.

By inspecting the centroids found after clusterization, we highlighted the characteristics of each cluster.

Cluster (1) Very high Logical Reasoning average All (vhLR-aAll) (n = 52) consists of subjects with very high performance in logical reasoning, high in visual perception, average in working memory, and visuospatial attention.

Cluster (2) Average All, Low Working Memory (aALL-lWM) (n = 12) consists of subjects with average performance in logical reasoning, visual perception, and visuospatial attention but low in working memory.

Cluster (3) Very High Logical Reasoning and Low Working Memory (vhLR-lWM) (n = 37) consists of subjects with very high performance in logical reasoning, high in visuospatial attention, and visual perception, and low performance in working memory.

Cluster (4) Very Low Visual Perception and Low Visuospatial Attention (vlVP-lVA) (n = 10) consists of subjects with very low performance in visual perception, low visuospatial attention, and high in working memory and logical reasoning.

Cluster (5) Very High Logical Reasoning and Very Low Visuospatial Attention (vhLR-vlVA) (n = 23) consist of subjects with very high performance in logical reasoning, average performance in working memory, high in visual perception, and very low in visuospatial attention.

Cluster (6) Very high Logical Reasoning and Working Memory (vhLRWM) (n = 22) consists of subjects with very high performance in logical reasoning and working memory, high in visual perception, and average in visuospatial attention.

Table 3
Clusters and Accuracy Index overall and divided for each cluster.

Cluster-ID	N	Accuracy %
vhLR-aAll	52	100
aALL-lWM	12	91.7
vhLR-lWM	37	91.9
vlVP-lVA	10	80
vhLR-vlVA	23	91.3
vhLRWM	22	90.9
ahAll	75	88
vhLR-lVP	19	94.7
vhALL	42	90.5
Average weighted Accuracy		91.8

Cluster (7) Average-High All (ahAll) ($n = 75$) It is the cluster with the highest representativeness and consists of subjects with high performance in all the tasks.

Cluster (8) Very high Logical Reasoning, Low Visual Perception (vhLR-lVP) ($n = 19$) consists of subjects with very high performance in logical reasoning, low in visual perception, and high in visuospatial attention and working memory.

Cluster (9) Very High All (vhALL) ($n = 42$) consists of subjects with a high or very high performance in all the tests. In particular, they present a very high performance in logical reasoning, high, visual perception, working memory, and high performance in visuospatial attention.

The Cluster Accuracy Index (Ic), weighted for the cluster numerosity showed a 91.8% overall accuracy, as illustrated in Table 3.

When inspecting the presence of SLD in 105 students, we found that 30 (28.6%) of them reached the criteria for an SLD (dyslexia or dyscalculia, or both).

Furthermore, in Table 3 the presence of SLD in the cluster solution found has been summarized.

Furthermore, according to the data presented in Table 4, we observed a higher frequency of students with SLD in certain clusters compared to those without SLD. Specifically, the clusters aALL-lWM and vhLR-lWM showed a prevalence of students with SLD that was 2.5 and 4.16 times greater, respectively, than those no-SLD. Additionally, both clusters exhibited poor performance in the working memory task, which supports previous research identifying working memory as a cognitive risk factor for students with SLD [49–51].

Table 2 displays the results of the cluster analysis, which grouped the original sample into nine different categories based on numerosity. This grouping was used to implement two classification algorithms: Adaptive Boosting and Artificial Neural Network (ANN) for both the original imbalanced sample and the balanced sample after the SMOTE process. A summary of the most relevant metrics can be found in Table 5.

3.3. *Adaptive Boosting*

The results of the algorithm trained on the original sample of 233 students (80%) showed a learning rate of 0.5001 and the number of estimators of 550 as the best parameters during the training, reflecting on an accuracy score of 80.68% at the training set and 84.75% on the test set on a sample of 59 students (20%).

The reliability of the classifier was compared with the imbalance of the training set. In our sample, Cluster (7) ahAll occupies 25.75% of the total frequency on the training set. Therefore, the algorithms' reliability was evaluated by considering that we can obtain the 25.75% of accuracy at the test set by predicting the majority class without the help of any supervised classifier. In our situation, the global accuracy of 84.75% and the balanced accuracy score of 77.23% highlight the ability of this classifier to learn most of the rules for predicting starting with the feature variables under consideration. However, not all the classes have been predicted correctly.

By inspecting the F1 score for each cluster in Table B1 (see Supplementary Materials) may be noted a good classification for clusters 1,5,6,7,8 and 9, whereas the other clusters showed a low F1 score meaning a worse classification rate. This resulted in a weighted F1 score of 0.846.

The results of the Adaptive Boosting on the balanced sample of 540 students showed a learning rate of 0.8001 and the number of estimators of 690 as the best parameters during the training reflecting an accuracy score of 69.81% at the training set and 57.63% on the test set on a sample of 59 students (20%). This shed light on a balanced accuracy score of 59.83%. By inspecting the F1 score for each cluster in Table B2 (see Supplementary Materials) clusters 1, and 5, showed a good classification. Compared to the remaining clusters, all exhibited a low F1 score. This resulted in a weighted F1 score of 0.595.

3.4. *Artificial neural network (ANN)*

After 140 epochs the model reached a global accuracy score of 94.42% at the training set and 89.83% at the test set. The balanced accuracy score of this model is 89.58%.

By inspecting the score F1 for each cluster in Table B3 (see Supplementary Materials) all the clusters exhibit scores over 0.8. This

Table 4
Distribution of SLD and no SLD students concerning the cluster solution found.

Cluster-ID	no SLD (%)	SLD (%)
vhLR-aAll	16 (21.3)	7 (23.3)
aALL-lWM	3 (4)	3 (10)
vhLR-lWM	6 (8)	10 (33.3)
vlVP-lVA	2 (2)	0 (0)
vhLR-vlVA	6 (8)	1 (3)
vhLRWM	1 (1.3)	0 (0)
ahAll	21 (28)	7 (23.3)
vhLR-lVP	9 (12)	1 (3)
vhALL	11 (14.6)	1 (3)

The Chi-Squared Test analysis carried out in our sample, showed a statistically significant difference between the distribution of the cluster in the students with SLD and no-SLD groups ($\chi2(8) = 16,587$, $p = 0.035$).

Table 5
Metrics of the algorithms implemented both in the imbalanced and balanced sample. It is possible to observe how the ANN on the balanced dataset performs better than the others in all the metrics under evaluation.

Principal Metrics	AdaBoost Imbalanced	AdaBoost Balanced	ANN Imbalanced	ANN Balanced
Global Training Accuracy (%)	80.7	69.8	94.4	94.8
Global Testing Accuracy (%)	84.7	57.6	89.8	91.5
Balanced Accuracy (%)	77.2	59.8	89.6	91.7
Weighted F1 Score	0.85	0.59	0.90	0.92

resulted in a weighted F1 score of 0.899.

The results after the implementation of the ANN on the imbalanced sample showed better results compared to the previous AdaBoost algorithm on the same sample.

The ANN implemented in the balanced showed even better results. After 82 epochs the model reached a global accuracy score of 94.81% at the training set and 91.53% at the test set, resulting in a balanced accuracy score of 91.66%.

By inspecting the score F1 for each cluster in Table B4 (see Supplementary Materials) no cluster exhibits scores below 0.7 while others over 0.8. This resulted in a weighted F1 score of 0.916.

These results displayed the ANN as the best algorithm for this type of problem both for the imbalanced and for the balanced sample.

4. Discussion

In this study, a novel clustering method was utilized for classifying cognitive abilities in secondary school students. This new approach involved preprocessing the k-means algorithm to achieve the most precise and reliable classification. The results revealed that the accuracy of classification and discrimination was at its peak when this method was applied.

The study demonstrated that a hybrid clustering approach, which combined Kohonen's Self-Organizing Maps (SOMs) and k-means, enhanced the replicability of clustering among students with typical development. The efficiency of this profiling technique was confirmed by an ANN algorithm, suggesting that it is highly effective in profiling new users.

Our findings confirm the results of prior research in various fields, as reported in Refs. [12–14]. The self-organizing map (SOM) technique groups similar cases into map units during the initial clustering stage. This reduces the amount of data to be classified in subsequent clustering procedures and diminishes the amount of noise [14]. As a result, applying the k-means algorithm to the map units divides the dataset into distinct partitions. Our study reveals that this approach provides a more accurate representation of the clustering space, explaining a higher degree of variance than the single cluster method alone. Furthermore, the overall cluster accuracy is excellent, achieving 91.8%.

The solution obtained indicated nine different groups having very high, high, average, low, or very low performance in the cognitive domains investigated. One group with difficulties in visual perception (vhLR-lVP), another group with impaired visuospatial attention (vhLR-vlVA), two groups with difficulties in working memory (aALL-lWM, vhLR-lWM), and one group with visuospatial and perceptual deficits (vlVP-lVA). This solution is partially in agreement with Yokota et al. (2014) study, indicating that perceptual organization and attention are important factors in clusterizing typically development children. Moreover, the proposed solution allowed us to distinguish between the distribution of the clusters in the SLD and no-SLD groups. In particular, by inspecting the aALL-lWM and vhLR-lWM clusters, were 2.5 and 4.16 times more for SLD students than no-SLD respectively. This corroborates previous findings where low working memory has been reported as a cognitive risk factor for developing dyslexia and dyscalculia [52–54].

The application of this cluster approach puts a novelty in the psychological field given that ML is not extensively used in the analysis of psychological experiments as compared to other fields (e.g., genetics) [55].

The results showed the ANN algorithm carried out in the balanced sample as the best one for this problem. The average F1 score of 0.92 indicates a very good ability of the algorithm to learn the rules which hold the cluster differences by considering a wide range of cognitive variables. These results, although preliminary, reveal that this approach could be an efficient tool for clustering cognitive profiles.

However, some limitations should be considered when interpreting the results of the present study. Above all, in the face of good results both for the cluster evaluation and prediction phases, the sample size consists of 292 students, thus reducing the generalizability of the results. Further studies will aim to increase the sample size, allowing a precise evaluation of the external validity of the clusters and, arguably, the cluster prediction from ML algorithms.

In addition, due to the nature of the tool used we have not been able to include the phonological awareness of the students inside the clustering procedure, but given its importance in learning, could be important its inclusion. This could point out more precisely students with educational special needs. Further studies would include a more reliable assessment of the phonological awareness inside PROFFILO and then include it in the clusterization procedure. Moreover, we would also focus on and compare other clustering algorithms (e.g. DBSCAN, spectral clustering, and gaussian mixture models) in the joint use with SOM.

5. Conclusion

The current study employs machine learning techniques to cluster the cognitive profiles of Italian secondary school students. The study measures cognitive abilities, including logical reasoning, visuospatial attention, motion perception, and working memory, using an online digital game called PROFFILO. The use of this clustering approach is a novelty in the psychological field, as machine learning is not widely used in psychological experiments as compared to other fields [55]. However, Orrù et al. [55] enumerated the benefits of using ML in psychological research, including improved generalization, replication of results, and personalized predictions at a single subject level. The present study adds to this literature, suggesting that ML can be especially useful for clustering heterogeneous populations, as it improves classification accuracy and allows testing the replicability of results. Psychologists often need to explain the heterogeneity of clinical populations and find a way to group patients in order to settle down successful interventions. The presented results evidenced that the use of ML within a cognitive profiling test such as PROFFILO may have important practical implications for clinical practice. Indeed, having a clustering model that is validated as the most accurate as possible, and could be replicated in other samples, allows the clinician to implement personalized based models of intervention. Moreover, the model is advantageous because it is expected to increase its validity and efficiency by adding cases and information. Furthermore, on the methodological point of view, this study is the first to compare the benefits and reliability of using both the Self-Organizing Maps algorithm and k-means for cognitive profiling, and to investigate the potential utility of supervised machine learning algorithms (specifically, AdaBoost and ANN) in predicting the cognitive profile of new users.

The findings of this study demonstrate that applying a hybrid clustering approach, which involves multiple steps using Self Organizing Maps and k-means, can enhance the reliability of clustering when analyzing diverse measures, such as cognitive profiling. This approach provides a better understanding of how clusters are distributed in groups with and without specific learning difficulties (SLD). Overall, these results suggest that hybrid clustering techniques can be useful in the field of psychology to improve the dependability of clustering and the accuracy of solutions.

Institutional permissions

All procedures complied with the ethical standards of national committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The study was approved by the University of Bologna Bioethics Committee. Parents and youths gave their written informed consent to participate in the study.

Author contribution statement

Matteo Orsoni: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Sara Giovagnoli: Conceived and designed the experiments.

Sara Garofalo: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Sara Magri; Martina Benvenuti; Elvis Mazzoni: Performed the experiments.

Mariagrazia Benassi: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data associated with this study has been deposited at OSF repository: https://osf.io/4qf8v/

Declaration of interest’s statement

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e14506.

References

[1] T.P. Alloway, M. Elsworth, An investigation of cognitive skills and behavior in high ability students, Learn. Indiv Differ 22 (6) (2012) 891–895, https://doi.org/10.1016/j.lindif.2012.02.001.

[2] M. Foley-Nicpon, S.G. Assouline, R.D. Stinson, Cognitive and academic distinctions between gifted students with autism and Asperger syndrome, Gift. Child. Q. 56 (2) (2012) 77–89, https://doi.org/10.1177/0016986211433199.

[3] D. Menghini, et al., Different underlying neurocognitive deficits in developmental dyslexia: a comparative study, Neuropsychologia 48 (4) (Mar. 2010) 863–872, https://doi.org/10.1016/j.neuropsychologia.2009.11.003.

[4] S. Heim, et al., Cognitive Subtypes Of Dyslexia Heim Et Al, 2008, pp. 73–82.

[5] H.W. Catts, D. Compton, J.B. Tomblin, M.S. Bridges, Prevalence and nature of late-emerging poor readers, J. Educ. Psychol. 104 (1) (Feb. 2012) 166–181, https://doi.org/10.1037/a0025323.
[6] S. Yokota, et al., Individual differences in cognitive performance and brain structure in typically developing children, Aug, Dev. Cogn. Neurosci. 14 (2015) 1–7, https://doi.org/10.1016/j.dcn.2015.05.003.
[7] M. Poletti, E. Carretta, L. Bonvicini, P. Giorgi-Rossi, Cognitive clusters in specific learning disorder, J. Learn. Disabil. 51 (1) (Jan. 2018) 32–42, https://doi.org/10.1177/0022219416678407.
[8] J. Clatworthy, D. Buick, M. Hankins, J. Weinman, R. Horne, The use and reporting of cluster analysis in health psychology: a review, Br. J. Health Psychol. 10 (3) (Sep. 2005) 329–358, https://doi.org/10.1348/135910705X25697.
[9] M. Benassi, et al., Using two-step cluster analysis and latent class cluster analysis to classify the cognitive heterogeneity of cross-diagnostic psychiatric inpatients, Front. Psychol. 11 (Jun) (2020), https://doi.org/10.3389/fpsyg.2020.01085.
[10] J.M. Kraus, C. Müssel, G. Palm, H.A. Kestler, Multi-objective selection for collecting cluster alternatives, Jun, Comput. Stat. 26 (2) (2011) 341–353, https://doi.org/10.1007/s00180-011-0244-6.
[11] T. Kohonen, The self-organizing map, Proc. IEEE 78 (9) (1990) 1464–1480.
[12] P. Juntunen, M. Liukkonen, M. Lehtola, Y. Hiltunen, Cluster analysis by self-organizing maps : an application to the modelling of water quality in a treatment process, Appl. Soft Comput. J. 13 (7) (2013) 3191–3196, https://doi.org/10.1016/j.asoc.2013.01.027.
[13] Z. Dong, D. Yang, T. Reindl, W.M. Walsh, A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance, Energy 82 (Mar. 2015) 570–577, https://doi.org/10.1016/j.energy.2015.01.066.
[14] F. Palamara, F. Piglione, N. Piccinini, Self-Organizing Map and clustering algorithms for the analysis of occupational accident databases, Saf. Sci. 49 (8–9) (2011) 1215–1230, https://doi.org/10.1016/j.ssci.2011.04.003.
[15] N.P. Allan, L.E. Hume, D.M. Allan, A.L. Farrington, C.J. Lonigan, Relations between inhibitory control and the development of academic skills in preschool and kindergarten: a meta-analysis, Dev. Psychol. 50 (10) (2014) 2368–2379, https://doi.org/10.1037/a0037493.
[16] A.G. Carlson, E. Rowe, T.W. Curby, Disentangling fine motor skills' relations to academic achievement: the relative contributions of visual-spatial integration and visual-motor coordination, J. Genet. Psychol. 174 (5) (Sep. 2013) 514–533, https://doi.org/10.1080/00221325.2012.717122.
[17] M.E. Fenwick, et al., Neuropsychological profiles of written expression learning disabilities determined by concordance-discordance model criteria, Apr, Appl. Neuropsychol. Child 5 (2) (2016) 83–96, https://doi.org/10.1080/21622965.2014.993396.
[18] E.S. Johnson, Understanding why a child is struggling to learn, Top. Lang. Disord. 34 (1) (Jan. 2014) 59–73, https://doi.org/10.1097/TLD.0000000000000007.
[19] M.F. Kudo, C.M. Lussier, H.L. Swanson, Reading disabilities in children: a selective meta-analysis of the cognitive literature, Res. Dev. Disabil. 40 (May 2015) 51–62, https://doi.org/10.1016/j.ridd.2015.01.002.
[20] Pradeep Kumar Gupta and Dr, Vibha sharma, "working memory and learning disabilities: a review, Jul, Int. J. Ind. Psychol. 4 (4) (2017), https://doi.org/10.25215/0404.013.
[21] C.R. Reynolds, S.E. Shaywitz, Response to Intervention: ready or not? Or, from wait-to-fail to watch-them-fail, Jun, Sch. Psychol. Q. 24 (2) (2009) 130–145, https://doi.org/10.1037/a0016158.
[22] C. Stevens, D. Bavelier, The role of selective attention on academic foundations: a cognitive neuroscience perspective, Feb, Dev. Cogn. Neurosci. 2 (2012) S30–S48, https://doi.org/10.1016/j.dcn.2011.11.001.
[23] M. Vock, F. Preckel, H. Holling, Mental abilities and school achievement: a test of a mediation hypothesis, Intelligence 39 (5) (Sep. 2011) 357–369, https://doi.org/10.1016/j.intell.2011.06.006.
[24] M. Orsoni, et al., PROFFILO: a New Digital Assessment Tool to Evaluate Learning Difficulties in Secondary School, 2021.
[25] J. Raven, The raven progressive Matrices: a review of national norming studies and ethnic and socioeconomic variation within the United States, J. Educ. Meas. 26 (1) (1989) 1–16, https://doi.org/10.1111/j.1745-3984.1989.tb00314.x.
[26] A. Orsini, L. Pezzuti, L. Picone, Wechsler Intelligence Scale for Children IV Edizione Italiana, 2012.
[27] B.L. Brooks, E.M.S. Sherman, E. Strauss, 1, in: second ed. 16NEPSY-II: A Developmental Neuropsychological Assessment, 2010, pp. 80–101, https://doi.org/10.1080/09297040903146966. Child Neuropsychol.
[28] C. Cornoldi, D. Giofrè, A.P. Baldi, PROVE MT AVANZATE - 3 - CLINICA, Giunti O.S, 2017.
[29] J. Raven, The raven progressive Matrices: a review of national norming studies and ethnic and socioeconomic variation within the United States, J. Educ. Meas. 26 (1) (Mar. 1989) 1–16, https://doi.org/10.1111/j.1745-3984.1989.tb00314.x.
[30] M. Walesiak, A. Dudek, The choice of variable normalization method in cluster analysis, Educ. Excell. Innov. Manag. (2020) 325–340. A 2025 Vis. to Sustain Econ. Dev. Dur. Glob. Challenges, no. June.
[31] M.M. Suarez-Alvarez, D.T. Pham, M.Y. Prostov, Y.I. Prostov, Statistical approach to normalization of feature vectors and clustering of mixed datasets, Proc. R. Soc. A Math. Phys. Eng. Sci. 468 (2145) (2012) 2630–2651, https://doi.org/10.1098/rspa.2011.0704.
[32] N.K. Visalakshi, K. Thangavel, Impact of normalization in distributed K-means clustering, Int. J. Soft Comput. 4 (4) (2009) 168–172.
[33] J. MacQueen, Some methods for classification and analysis of multivariate observations, Proc. 5th Berkeley Symp. Math. Stat. Probab. 1 (1967) 281–297.
[34] R.L. Thorndike, Who belongs in the family? Psychometrika 18 (4) (Dec. 1953) 267–276, https://doi.org/10.1007/BF02289263.
[35] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer, New York, 2013.
[36] A. Luque, A. Carrasco, A. Martín, A. De, The impact of class imbalance in classification performance metrics based on the binary confusion matrix, Pattern Recogn. 91 (2019) 216–231, https://doi.org/10.1016/j.patcog.2019.02.023.
[37] N V Chawla, K.W. Bowyer, L.O. Hall, P.W. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[38] G. Xu, M. Hu, C. Ma, Secure and smart autonomous multi-robot systems for opinion spammer detection, Inf. Sci. 576 (2021) 681–693, https://doi.org/10.1016/j.ins.2021.07.072.
[39] G. Xu, W. Li, J. Liu, A social emotion classification approach using multi-model fusion, Future Generat. Comput. Syst. 102 (2020) 347–356, https://doi.org/10.1016/j.future.2019.07.007.
[40] Y. Freund, R.E. Schapire, Experiments with a new boosting algorithm, Proc. 13th Int. Conf. Mach. Learn. (1996) 148–156.
[41] A. Luque, A. Carrasco, A. Martín, A. de las Heras, The impact of class imbalance in classification performance metrics based on the binary confusion matrix, Pattern Recogn. 91 (Jul. 2019) 216–231, https://doi.org/10.1016/j.patcog.2019.02.023.
[42] Y. Sun, M.S. Kamel, A.K.C. Wong, Y. Wang, Cost-sensitive boosting for classification of imbalanced data, Pattern Recogn. 40 (12) (2007) 3358–3378, Dec, https://doi.org/10.1016/j.patcog.2007.04.009.
[43] J. Team, JASP (Version 0.16.1), 2022.
[44] R. R Core Team, A Language and Environment for Statistical Computing, Vienna, Austria, 2020 [Online]. Available: https://www.r-project.org/.
[45] G. Van Rossun, F. Drake, Python 3 Reference Manual, Create Space, Scotts Valley, CA, 2009.
[46] F. Pedregosa, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2012) 2825–2830, Jan [Online]. Available: http://arxiv.org/abs/1201.0490.
[47] M. Abadi, et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, Mar. 2016 [Online]. Available: http://arxiv.org/abs/1603.04467.
[48] A. Ultsch, L. Herrmann, The architecture of emergent self-organizing maps to reduce projection errors, in: ESANN 2005 Proc. - 13th Eur. Symp. Artif. Neural Networks, 2007, pp. 1–6.
[49] K. Moll, S.M. Göbel, D. Gooch, K. Landerl, M.J. Snowling, Cognitive risk factors for specific learning disorder, J. Learn. Disabil. 49 (3) (May 2016) 272–281, https://doi.org/10.1177/0022219414547221.
[50] J.F. McLean, G.J. Hitch, Working memory impairments in children with specific arithmetic learning difficulties, J. Exp. Child Psychol. 74 (3) (Nov. 1999) 240–260, https://doi.org/10.1006/jecp.1999.2516.

[51] D.C. Geary, M.K. Hoard, Numerical and arithmetical deficits in learning-disabled children: relation to dyscalculia and dyslexia, Jul, Aphasiology 15 (7) (2001) 635–647, https://doi.org/10.1080/02687040143000113.

[52] K. Moll, S.M. Göbel, D. Gooch, K. Landerl, M.J. Snowling, Cognitive risk factors for specific learning disorder: processing speed, temporal processing, and working memory, J. Learn. Disabil. 49 (3) (2016) 272–281, https://doi.org/10.1177/0022219414547221.

[53] D.C. Geary, M.K. Hoard, Numerical and arithmetical deficits in learning-disabled children: relation to dyscalculia and dyslexia, Aphasiology 15 (7) (2001) 635–647, https://doi.org/10.1080/02687040143000113.

[54] J.F. McLean, G.J. Hitch, Working memory impairments in children with specific arithmetical difficulties, J. Exp. Child Psychol. 74 (1999) 240–260.

[55] G. Orrù, M. Monaro, C. Conversano, A. Gemignani, G. Sartori, Machine learning in psychometrics and psychological research, Front. Psychol. 10 (Jan) (2020), https://doi.org/10.3389/fpsyg.2019.02970.