CEIG - Spanish Computer Graphics Conference (2024)
J. Marco and G. Patow (Editors)

Serious Games and Artificial Intelligence for the detection of mathematical difficulties at school

J. Hornos-Arias[1,3], J.M. Serra-Grabulosa[2,3], J. Gómez-Berengueras[1,3] and S. Grau-Carrion[1,3]

[1]Digital Care research group. University of Vic-Central University of Catalonia (UVic-UCC), Vic (Barcelona). [2] Institut de Neurociències, Universitat de Barcelona (UB), Barcelona. [3] NeurekaLAB.

Abstract

Mathematical attainment at the beginning of primary school is the strongest predictor of later mathematical achievement. Mathematical difficulties are assessed objectively using screening tools based on cognitive assessment for numerical processing and calculation on an individual basis under professional supervision, which is why early detection/intervention in school is difficult to implement. Our main objective was to validate a tool that combines a serious game with machine learning (ML) algorithms to perform accurate prediction for cognitive assessment of numerical processing and calculation, facilitating early detection of unsupervised mathematical difficulties at school. Following an uncontrolled open trial with a small sample size (90 children in 2nd grade of primary school) we were able to train and compare different ML algorithms with the data generated with our serious game and traditional cognitive assessments. The best fitted models for each cognitive area offered promising results, showing accuracies between 65% and 96% that combined with other good performance metrics (high recall and F1 scores for some cases) appointed to a high fidelity on diagnose. Although the results are not totally conclusive, as this was an exploratory study and more research must be done, we were able to validate the system.

1 Introduction

Learning difficulties (LD) and their associated problems have been the subject of study in recent decades. By understanding how they manifest and evolve (especially at the neuropsychological level), not only early detection but also appropriate intervention may be possible. One of the best known and most studied specific learning difficulty is dyslexia, in which individuals have difficulties in the tasks of reading (fluent/accurate word recognition and spelling) and text comprehension. [PP15].

Another learning difficulty is developmental dyscalculia (DD). Although it has a similar prevalence, it is less known and understood than dyslexia. As a result, many teachers have limited knowledge about dyscalculia [KS20], and students with this difficulty do not get the support they need while learning mathematics in school. DD is a specific learning disorder with a neurobiological origin that affects numerical processing and calculation [Ame22] [Wor04]. People with low numeracy skills have problems understanding and mastering different mathematical concepts, such as number sense or memorizing arithmetic facts, not being able to perform calculations and basic math reasoning in an accurate and/or fluent manner. Different studies stated that about 3% to 6% of the children are affected by mathematical difficulties [KVA14] [Zer15].

For its inherent nature, DD is a very heterogeneous and complex learning disability [LGM13] [RH09], as far as in numeracy skills are included concepts such as counting, transcoding between spoken/read words, digits and/or representations, ordering, magnitude, simple arithmetic (addition and subtraction), subitizing (that involves memory and attention), spatial representation (mental line number) or problem solving [KVA14]. Apart from this, taking into account that great part of the numeracy skills enumerated involve knowledge in other areas (reading, speaking, working memory), DD commonly presents high comorbidity with other LD such as dyslexia or Attention-Deficit/Hyperactivity Disorder (ADHD) [WDP*06] [Rap16] [WMP*19] [MDLL*21], although it is possible to suffer low numeracy difficulties in isolation [MKN*14]. It is important to remark that having low or high intelligence quotient (IQ) has no direct relation with being diagnosed with dyscalculia [But18].

Mathematical level at the beginning of primary school is the strongest predictor for both later mathematical achievement and success in other academic domains [DDC*07]. Low numeracy skills in early stages not only affects the children in terms of confidence, but also have a great impact on their adulthood: from

DOI: 10.2312/ceig.20241141

problems solving daily tasks or school failure to unqualified employment/unemployment or even mental health issues [But18], apart from the derived cost for society in terms of investmentment to cover special education needs, treatments or the economical charge for unemployment benefits. For all these reasons it is necessary to detect DD in early stages of education to provide a proper intervention [RSW*09] to low achievers, being the school the best place where this process must be implemented.

Dyscalculia detection is carried out nowadays through screening tools based on a cognitive evaluation for numerical processing and calculation under the supervision of a professional, who is in charge of passing the test to the child. This methodology is a major handicap in the application of early detection/intervention of mathematical difficulties in schools.

Great advances in computer science and the extended use of mobile technology have made possible the use of video games for educational or health purposes (Serious Games). This type of tools gives the user a certain autonomy in its use, providing an immersive environment that is easy to use and fun, improving engagement and motivation and, therefore, increasing the effectiveness of the intervention. Several studies have been carried out evaluating mobile app games as tools for the remediation of dyscalculia [RSW*09] [KH14] [FCA*17] [AHA*19] [RSM20] [RBVP*20], as well as interventions based on Virtual Reality Serious Games (VRSG) and experimental Augmented Reality Serious Games (ARSG) [APVCA*18] [MZNN19], obtaining promising results.

Offering a fitted diagnose and personalized tools for detection and intervention is mandatory for dealing not only with dyscalculia (as involve different numeracy skills and comorbidity with other SLD's), but also with neurodevelopmental and specific learning disorders. The use of machine learning (ML) algorithms could be helpful to maximize adaptability and reliability in this sense. By one hand, these algorithms could be used to provide the system the ability to improve their predictions through the processing of the data obtained from all the previous tests passed to different children, gaining accuracy and reliability. On the other, use gathered data from a profile and other similar profiles to offer each player a fully adapted learning path.

Large interest in applying machine learning algorithms to different neurodevelopmental disorders was stated in [SMC*23], especially for ASD (135 studies) and ADHD (55) diagnosis, although none of these cases stated the use of serious games in combination with ML/DL. [RBR22] presented a study based on a web game combined with ML algorithms for dyslexia screening, obtaining an accuracy of 74% for German language and 69% for Spanish, using Random Forest and Extra Trees algorithms respectively. Some research have been conducted focusing on the design of reliable digital screening tests and/or intervention tools combining mobile games with machine learning (ML) techniques for dealing with mathematical difficulties. [KNH*19] used a previously trained model based on Support Vector Machines with Radial Basis Function kernel to improve dyscalculia detection process (obtaining an accuracy close to 90%), while [KBK*13] implemented a Bayesian Network model to determine students level in each moment offering personalized tasks in intervention (medium effect sizes were observed comparing pre-post training tests).

The main goal of this study was to validate a tool combining a serious game with machine learning algorithms in order to perform accurate prediction for cognitive assessment of numerical processing and calculation, facilitating autonomous early detection of mathematical difficulties at school. To assess accuracy and reliability, different algorithms were trained using both the results of the mathematical tasks performed during the serious game's sessions and the results of a validated dyscalculia screening test (which did not include any artificial intelligence support), previously performed by each participant.

2 Methods

2.1 Trial design

An uncontrolled open trial was conducted with a experimental group of children in 2nd year of primary education from different Spanish schools, aged between 7 and 8 years.

Trial was divided in three phases. Firstly, each student passed a cognitive assessment on numerical processing and calculation. In the next phase, each children played with the intervention serious game along two weeks. Finally, data collected from the previous phases was used to train and test different machine learning algorithms, analyzing which one/s offers the best reliability predicting signs of mathematical difficulties.

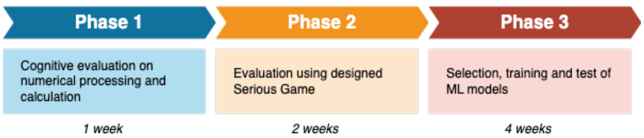


Figure 1: *Study design overview*

2.2 Participants

The sample size was 90 participants (see demographics in table 1). Recruitment period started in June 2023 and ended in August 2023, and the trial was conducted between October 2023 and the end of December 2023. As all participants in the study were children of primary school (aged 7-8 years), written informed consent was requested from both school staff and parents or legal guardians.

Corresponding study protocol was approved by the Research Ethics Committee (REC) at UVic-UCC (Vic, Spain) in May 2023.

2.3 Intervention

Predictive Artificial Intelligence (AI) methods, such as Machine Learning (ML), are used to obtain predictions by processing a cer-

N	Female	Male	Age range	Mean age	Age SD
90	49	41	7-8	7.354	0.483

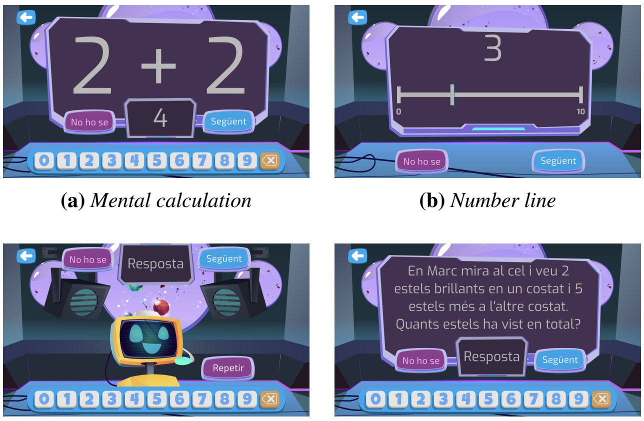
Table 1: *Overview of the participants (2nd course)*

tain amount of classified data. In this study we aimed to train different ML algorithms to predict the outcome of a numerical processing and calculation assessment using as training/test data the results obtained while children played a serious game.

Before starting the study, serious game apps had to be configured and installed on the devices of the school. All teachers received instructions on how to access the applications, how they worked, how the cognitive assessment has to be administered and how to carry out the sessions with the developed serious game. If additional questions arose, the research staff helped them to conduct the study correctly. Once the environment is properly prepared, the three phases of the study started.

2.3.1 Phase 1: Cognitive evaluation

Each student was evaluated with the cognitive assessment tool on numerical processing and calculation. For this, a digital dyscalculia risk screening test was used. This is a scientifically validated test to detect early signs of mathematical difficulties in children aged 5 to 12 [Neu]. Different mathematical skills and domains were assessed using 4 different activities: (1) Mental arithmetic (addition and subtraction); (2) Number line; (3) Transcoding; and (4) Arithmetic problems. (see figure 2).



(a) *Mental calculation* **(b)** *Number line*

(c) *Transcoding* **(d)** *Arithmetic problems*

Figure 2: *Activities evaluated in the screening test*

The evaluation process lasted between 10 and 15 minutes, and it was important to carry it out without interruptions, avoiding any kind of distraction. If necessary, the test could be restarted. Once the four activities were completed, results were calculated by the app using up to date scales of assessment for each course (no ML algorithms were involved in the evaluation). All the evaluation data was stored in each child's profile and a report with the results was generated, being available for consultation.

2.3.2 Phase 2: Sessions period

After completing the cognitive assessment, all children played the intervention serious game. This serious game, available for mobile devices (iOS and Android) and also as a web application, offered 8 different types of activities: operations, number line, number creation, comparison, sequence completion, subitizing, identification and numerical memory. Each of them assesses one or more mathematical concepts, and can be configured with different properties to adapt the level of difficulty: use of a numerical value or an addition/subtraction operation, definition of exposure time, number of repetitions of the activity and range of values to be assessed. For this study, a pre-established and equal configuration was used for each student. In order to cover all the mathematical domains [KN20] with the minimum number of activities, we used only 5 of the 8 activities (see Figure 4). This facilitates the evaluation of the same mathematical knowledge assessed in the cognitive test, excluding the assessment of arithmetic problems. A summary of the configuration and relationship between the planned session activities and the cognitive assessment of numerical processing and calculation is summarised in the table 2.

Figure 3: *Participant playing with the mobile serious game*

To collect the necessary data, children played between 10 to 20 minutes a day, doing 5 weekly sessions for 2 weeks. Sessions took place in the classrooms of the educational centres. Teachers were in charge of controlling the session schedule, helping the children to do it properly and solving any doubts that might arose about how each activity should be carried out, not helping children about how to solve the questions proposed in the activities, to not interfere in the results. At the end of the intervention period, it was estimated that each participant had completed at least 10 sessions.

Activity	Numerical ranges	Variant	Cognitive evaluation activity
Identification	1-20, 1-100	value	Transcoding
Operations	1-20, 1-100	addition	Operations (addition)
Operations	1-20, 1-100	subtraction	Operations (subtraction)
Number Line	1-20, 1-100	value	Number Line
Comparison	1-20, 1-100	value	Transcoding
Number Creation	1-20, 1-100	value	Transcoding

Table 2: *Activities and their configuration used for the intervention sessions*

2.3.3 Phase 3: ML training

The training of the different ML algorithms started once the data collection was completed. The first step was to analyse the raw data obtained from both the cognitive assessment and the serious game, in order to create a suitable set of clean data that could be used as input for the ML algorithms. Depending on the final amount of information and nature different alternatives and configurations could be selected and tested, obtaining a set of predictions that should be analysed to determine the effectiveness and reliability of each of them. The data processing and the training, fitting, testing and evaluation of models were carried out using the Caret package (version 6.0-94) [KM08] in the R environment [R C23].

3 Data Collection

It was necessary to collect a large amount of quality matched data from each child to properly train and test different ML algorithms. These data are mainly composed of the results obtained through the validated digital cognitive assessment on numerical processing and computation, and the data traces collected through all the activities that were included in the serious game. In addition, data such as date of birth and sex of the child were collected to assess whether other factors could influence the results of the study.

3.1 Data and related indexes

Mathematical impairments could affect one or more domains of mathematical knowledge: a) Magnitude (symbolic/non-symbolic learning of numerical processing and calculation); b) Memory/Verbal (processes related to repetition and/or memorization of knowledge and numerical coding); Reasoning (executive functions and how the related processes work) and Visuospatial (processes related to the spatial distribution of objects and their interactions). To detect if there could be any risk of deterioration, in both the numerical processing and calculation assessment and the serious game sessions different processes were evaluated:

- **Mental Calculation (addition and subtraction)**. Evaluates the ability to perform simple mathematical operations.
- **Knowledge of Number System**. Evaluates skills related with the different representations of numbers and related conversions.
- **Mental Representation of the Numerical System**. Evaluates the mastery of the spatial number distribution concept and its relations, represented by the numerical mental line.
- **Base 10 System**. Evaluates the knowledge about units, tens, hundreds and thousands for both concrete numbers and how to use them in problems.

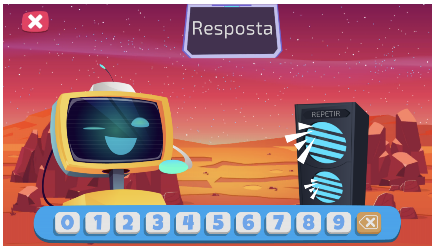
Hits or distance between question and response (for mental line) and response time are the main indexes that define a cognitive evaluation, being strongly associated with the different evaluated processes.

Sessions were divided into different activities. Each activity was designed to evaluate a key concept (domain) and was composed of a certain number of questions about them, so we were interested in storing traces of each repetition that included the following information: question, answer given, solution and response time. Information was also stored on the duration of the session (start/end time), the average percentage of correct answers and the response time for each activity, in order to have fully characterized sessions.

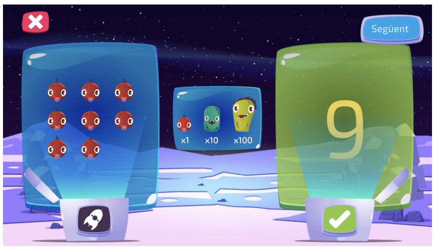
3.2 Participants sample size

Since this study was focused on determining the reliability of different ML algorithms trained with the data collected to predict the different indexes that compose a cognitive assessment on numerical processing and calculation, certain amount of paired and sorted data was necessary to adequately carry out the training and testing processes. Attending to the configuration of activities proposed for each session, the features (predictor variables) needed to feed our models and characterize each cognitive index were the number of repetitions proposed for each activity. One simple and extended way to estimate the number of samples (children) for predictive models is to suppose a minimum sample size of ten times the predictor variables [BDACS*22]. Although this approximation may not be generally applicable, as features are highly dependant on the context of the study, the prevalence of the outcome, the quality of the predictor variables chosen and the type of model to be developed, it could be a first approach, as this was an exploratory study. Based on the characteristics of the activities per session shown in the table 3, and taking into account that these predictors were obtained for each child, we took as a reference the minimum number of predictors to ensure that the rest of the indices could be processed correctly. This gave us a sample size of 540 children, since the minimum features per child were 54. However, due to some time constraints and some difficulties in the recruitment process, this study could only count on 90 participants.

(a) *Sessions dashboard*

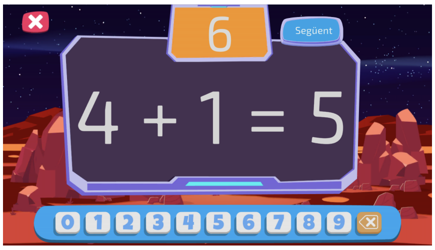


(b) *Identification*



(c) *Comparison*

(d) *Number line*



(e) *Operations*

(f) *Number creation*

Figure 4: *Screenshots of some activities of the serious game*

Activity	Variant	Cognitive evaluation	Repetitions (1-20)	Repetitions (1-100)	Total features
Identification	value	Transcoding	14	40	54
Operations	addition	Operations (addition)	10	50	60
Operations	subtraction	Operations (subtraction)	10	50	60
Number Line	value	Number Line	40	60	100
Comparison	value	Transcoding	20	60	80
Number Creation	value	Transcoding	20	60	80

Table 3: *Algorithm feature recount for proposed session's activities*

4 Analysis

4.1 Data preparation

The results yielded by the cognitive assessment on numerical processing and calculation are expressed as IQ scores (also known as Standard Scores), and is a standardization of Z-scores where a standard deviation of 15 and 100 is considered as the mean value, but these results can also be obtained as percentiles. These values can be expressed as ranges, classifying the different levels of impairment. For our study, and for the sake of simplicity, a two-class range was proposed to classify performance: (see table 4).

Standard score	Percentile	Performance
≥ 90	≥ 25	Average-Very superior
< 90	< 25	Low Average

Table 4: *Cognitive evaluation results. Performance classes*

It is important to notice that the 25th percentile was used because the lowest quartile of adults on tests of basic mathematics are functionally innumerate [Byn97] and children scoring below this cutoff on tests that assess basic mathematical competencies are at heightened risk for poor long-term outcomes in mathematics learning ([GHB*07] [GHNB12] [GHNB13] [MMHE07]) and later functional innumeracy [GHNB13]. On the other hand, information about each repetition done by index was expressed as a binary variable, where a hit was coded as a '1' and an error as a '0'.

4.2 Algorithm selection and metrics

As we had binary predictors and a two-class output, different ML algorithms based on classification were selected: K-Nearest-Neighbour (KNN), Random Forest (RF), Support Vector Machines (SVM, linear kernel), Gradient Boosting Machine (GBM), Classification and Regression Trees (CART), Naive Bayes (NB), Regularized Logistic Regression (GLMNET) and Linear Discriminant Analysis (LDA). This initial selection was based on the most used machine learning algorithms for classification [Ray24] [Tav24], being the first five highly used with good

Proceedings published by Eurographics - The European Association for Computer Graphics.

Activity	**Algorithm**	**Accuracy**	**P-Value (Acc >NIR)**	**Kappa**	**Recall**	**Precision**	**F1**	**Classification error rate**
Identification	***LDA***	0,654	0,589	0,170	0,333	0,5	0,4	0,346
Number Line	***LDA***	0,846	0,250	0,509	0,5	0,75	0,6	0,154
Operations: addition	***LDA***	0,962	0,736	0,649	1	0,5	0,667	0,038
Operations: subtraction	***GBM***	0,808	0,012	0,611	0,818	0,75	0,783	0,192
Number Creation	***KNN***	0,731	0,273	0,355	0,444	0,667	0,533	0,269
Comparison	***SVM Linear***	0,692	0,427	0,320	0,556	0,556	0,556	0,308

Table 5: *Analysis for different ML algorithms by activity and performance metrics*

performance metrics in (early) disease diagnosis and patient risk assessment [Pal24].

To determine which algorithms were more accurate and reliable, each trained model was tested for prediction, obtaining important information about each model behavior and performance metrics: accuracy (Acc), precision, recall and F1-score, and others that could help in the correct selection, like the the classification error rate and the p-value (significant if accuracy is higher than the largest proportion of the observed classes, known as No Information Rate or NIR). It was noticeable that the dataset used on the study of a mathematical disability could be relatively imbalanced, as the prevalence of this learning disorder is between the 3 and 6% of the children. For our study, it was important to avoid false negatives, since it implied not detecting a person with learning disabilities in mathematics. Thus, we could not only rely on accuracy, given that this metric could lead us to a wrong selection. Recall (compares detected positives over actual positives) must be considered a key metric when evaluating, finding the higher value (the lower false negatives are, the greater recall is). Also, we must check F1 score: high values indicate that the model is good at identifying both positive and negative cases (minimizing misdiagnosis). Finally, kappa could help us in order to determine if the real observations were close to the predicted (a value greater than zero and close to 1 denotes high fidelity).

All analyses were performed using the R environment (version 4.3.2)) [R C23].

4.3 Predictive model configuration

Dataset was divided in to sub-datasets: one for training (70% of the original data) and another for testing (30%). To guarantee that results could be repeatable, a random seed was selected to divide the original dataset (1324), and another seed (24424) was assigned to the training processes.

In order to improve the estimated performance of our machine learning model (reducing possible noise and errors) we performed a training based on the Repeated k-Fold Cross-Validation [Ber18] procedure. k-Fold Cross-Validation consists on splitting the data different times in order to obtain very different results. Repeating this process multiple times and reporting the mean result across all folds from all runs, it is expected to obtain a more accurate estimation of the mean performance of the model. For our case, we used a standard configuration: k=10 with 3 repetitions. Additionally, for each selected model, a tuning was made by testing different hyperparameters and train control configurations, in order to obtain an accurate training that lead us to better performances. Detailed information for each model's configuration can be found in Supplementary Materials.

5 Results

When the data collection was finished, different ML models were trained using the acquired data and the configuration (for dataset and algorithms) specified in section 4. It is important to notice that for this study, only the precision (hits) has been taken into account, letting the evaluation of predictive models for response times for a larger study. A model was selected for each activity, being possible to determine the performance for each different cognitive area. This way, we had more control on the deficits of the child than giving an overall performance for all grouped activities.

Once all the models were trained, they were tested against the corresponding test sub-dataset in order to obtain all the indicators that could help us to select the best option for each case. All this information is resumed in table 3. Only the best models for each activity and its related domain were selected, attending to the values of the different performance metric and its importance for our case. Details about this selection can be found in table 5.

All the selected algorithms presented good accuracy values, going from 65.4% to 96.2%. The best recall values were obtained for operations activities (1 for addition operations and 0.82 for subtraction), denoting that for these cases the rate of false positives was reduced (one of the most important premises for our case). F1 score and Kappa presented also the best results at operations (F1=0.67, Kappa=0.65 for addition and F1=0.78, Kappa=0.61 for subtraction), which indicates selected models for these activities were good at identifying both positive and negative cases and, therefore, minimizing misdiagnosis, presenting a high fidelity diagnose. Finally, for subtraction operation activity, p-value is in the limit for significance (value near to 0.01), so this selected model was the one with best metrics and can be considered of high reliability.

Paying attention to selected models, LDA was the best algorithm for three of the activities (Identification, Number Line and Addition Operations, associated to different cognitive domains). The other three activities presented best results using different ML algorithms: GBM for Subtraction Operations, KNN for Number Creation and SVM (Linear Kernel) for Comparison.

6 Discussion

This study tried to determine if it was possible to perform accurate predictions for cognitive assessment of numerical processing and calculation using a serious game in combination with machine learning algorithms, in order to offer to the schools a tool that facilitates an autonomous early detection of mathematical difficulties. To do so, a videogame was used to collect data an different ML algorithms were put to test to determine which of them could offer enough reliability on predicting mathematical difficulties.

Activities that are related with transcoding (Identification, Comparison and Number Creation) presented worse results in terms of reliability, although are better than we expected (taking into account the sample size). Features for these activities are quite different (54 for the first, 80 for the others), so this might not be related with results. Mental Calculation related activities (addition and subtraction), by the other hand, had a small number of features similar than the used on Identification, but achieved more accurate predictions and robustness against false positives, having very good detection indicators and rates. Finally, Number Line activities (with the higher number of features) presented similar (but a bit worse) results in terms of performance metrics and reliability than the obtained for the Operation cases. One possible cause behind a worse prediction for transcoding cognitive assessment could be related with the fact that we have separated related data in three different tasks. This have been done in order to know if this area presented more sensitivity on transcoding evaluation depending on the used activity. Grouping all data in order to train only one model for transcoding could offer better results.

7 Conclusions and future work

After training and testing different models for different cognitive areas, it is possible to conclude that the designed system (serious game combined with machine learning algorithms) could offer reliable predictions in order to detect mathematical difficulties on early stages, so we can validate the system.

It is important to remark that a larger study must be made in order to confirm the results of this preliminary study. First of all, reaction time performance must be implemented in order to get a complete assessment. Regression models could be a good option in order to detect difficulties evaluating response times, as both the results of the cognitive evaluation and predictor variables presented a continuous character. On the other hand, more children must be evaluated in order to have a good balance between features and samples. Also, it could be good to take into account the possibility of grouping all transcoding related activities, in order to know if the problem appointed in discussion could arose from this condition. Testing different number of features (using more sessions) could also be interesting to determine if models can be improved (taking care to not overfitting the system). Finally, it could be interesting to perform a better configuration of each concrete model (performing a deeper tuning on some hyperparameters) or even consider other ML algorithms, as that could provide more accurate predictions.

This preliminary study showed how to the combination of serious games and machine learning could be an interesting field of study to offer to schools reliable tools to detect not only mathematical difficulties, but also for other learning difficulties.

8 Acknowledgments and disclosure

This study was supported by a grant from *Fundación Cellnex*. We thank all children, their parents or legal guardians and teachers for participating in this study. The authors declare no conflict of interest.

References

[AHA*19] ARIFFIN M. M., HALIM F. A. A., ARSHAD N. I., MEHAT M., HASHIM A. S.: Calculic Kids© Mobile App: The Impact on Educational Effectiveness of Dyscalculia Children. *International Journal of Innovative Technology and Exploring Engineering 8*, 8S (6 2019). 2

[Ame22] AMERICAN PSYCHIATRIC ASSOCIATION: *Diagnostic and statistical manual of mental disorders. (5th ed., text rev.)*. 2022. doi:10.1176/appi.books.9780890425787. 1

[APVCA*18] AVILA-PESANTEZ D., VACA-CARDENAS L., AVILA R. M., PADILLA N., RIVERA L.: Design of an Augmented Reality Serious Game for Children with Dyscalculia: A Case Study. *Springer International Publishing eBooks* (8 2018), 165–175. doi:10.1007/978-3-030-05532-5_12. 2

[BDACS*22] BAEZA-DELGADO C., ALBERICH L. C., CAROT-SIERRA J. M., VEIGA-CANUTO D., DE LAS HERAS B. M., RAZA B., MARTI-BONMATI L.: A practical solution to estimate the sample size required for clinical prediction models generated from observational research on data. *European radiology experimental 6*, 1 (6 2022). doi:10.1186/s41747-022-00276-y. 4

[Ber18] BERRAR D.: *Cross-Validation*. 01 2018. doi:10.1016/B978-0-12-809633-8.20349-X. 6

[But18] BUTTERWORTH B.: The implications for education of an innate numerosity-processing mechanism. *Philosophical Transactions of the Royal Society B 373*, 1740 (2 2018), 20170118. doi:10.1098/rstb.2017.0118. 1, 2

[Byn97] BYNNER J.: Basic skills in adolescents' occupational preparation. *The Career Development Quarterly 45*, 4 (6 1997), 305–321. doi:10.1002/j.2161-0045.1997.tb00536.x. 5

[DDC*07] DUNCAN G. J., DOWSETT C. J., CLAESSENS A., MAGNUSON K., HUSTON A. C., KLEBANOV P. K., PAGANI L. S., FEINSTEIN L., ENGEL M., BROOKS-GUNN J., SEXTON H. R., DUCKWORTH K., JAPEL C.: School readiness and later achievement. *Developmental Psychology 43*, 6 (11 2007), 1428–1446. doi:10.1037/0012-1649.43.6.1428. 1

[FCA*17] FERRAZ F., COSTA A. P., ALVES V., VICENTE H., NEVES J. S., NEVES J.: Gaming in Dyscalculia: A Review on disMAT. *Advances in intelligent systems and computing* (4 2017). 252 doi:10.1007/978-3-319-56538-5\{_}25.

[GHB*07] GEARY D. C., HOARD M. K., BYRD-CRAVEN J., NUGENT L., NUMTEE C.: Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development 78*, 4 (7 2007), 1343–1359. doi:10.1111/j.1467-8624.2007.01069.x. 5

[GHNB12] GEARY D. C., HOARD M. K., NUGENT L., BAILEY D. H.: Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology 104*, 1 (1 2012), 206–223. doi:10.1037/a0025398. 5

[GHNB13] GEARY D. C., HOARD M. K., NUGENT L., BAILEY D. H.: Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLOS ONE 8*, 1 (1 2013), e54651. doi:10.1371/journal.pone.0054651. 5

[KBK*13] KÄSER T., BASCHERA G.-M., KOHN J., KUCIAN K., RICHTMANN V., GROND U., GROSS M., VON ASTER M.: Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition. *Frontiers in Psychology 4* (8 2013). doi:10.3389/fpsyg.2013.00489. 2

[KH14] KUHN J.-T., HOLLING H.: Number sense or working memory? The effect of two computer-based trainings on mathematical skills in elementary school. *Advances in Cognitive Psychology 10*, 2 (6 2014), 59–67. doi:10.5709/acp-0157-2. 2

[KM08] KUHN, MAX: Building predictive models in r using the caret package. *Journal of Statistical Software 28*, 5 (2008), 1–26. URL: https://www.jstatsoft.org/index.php/jss/article/view/v028i05, doi:10.18637/jss.v028.i05. 4

[KN20] KARAGIANNAKIS G., NOËL M.-P.: Mathematical Profile Test: A preliminary evaluation of an online assessment for mathematics skills of children in grades 1–6. *Behavioral sciences 10*, 8 (8 2020), 126. doi:10.3390/bs10080126. 3

[KNH*19] KARIYAWASAM R., NADEESHANI M., HAMID T. N. A. T. A., SUBASINGHE I., RATNAYAKE P.: A Gamified Approach for Screening and Intervention of Dyslexia, Dysgraphia and Dyscalculia. *2019 International Conference on Advancements in Computing (ICAC)* (12 2019). doi:10.1109/icac49085.2019.9103336. 2

[KS20] KUNWAR R., SHARMA L.: Exploring Teachers' Knowledge and Students' Status about Dyscalculia at Basic Level Students in Nepal. *Eurasia journal of mathematics, science and technology education 16*, 12 (10 2020), em1906. doi:10.29333/ejmste/8940. 1

[KVA14] KUCIAN K., VON ASTER M.: Developmental dyscalculia. *European Journal of Pediatrics 174*, 1 (12 2014), 1–13. doi:10.1007/s00431-014-2455-7. 1

[LGM13] LANDERL K., GÖBEL S. M., MOLL K.: Core deficit and individual manifestations of developmental dyscalculia (DD): The role of comorbidity. *Trends in Neuroscience and Education 2*, 2 (6 2013), 38–42. doi:10.1016/j.tine.2013.06.002. 1

[MDLL*21] MOLL K., DE LUCA M., LANDERL K., LANDERL K., ZOCCOLOTTI P., BANFI C., BANFI C., ZOCCOLOTTI P.: Editorial: Interpreting the Comorbidity of Learning Disorders. *Frontiers in Human Neuroscience 15* (12 2021). doi:10.3389/fnhum.2021.811101. 1

[MKN*14] MOLL K., KUNZE S., NEUHOFF N., BRUDER J., SCHULTE-KÖRNE G.: Specific Learning Disorder: Prevalence and Gender Differences. *PLOS ONE 9*, 7 (7 2014), e103537. doi:10.1371/journal.pone.0103537. 1

[MMHE07] MURPHY M. M., MAZZOCCO M. M., HANICH L. B., EARLY M. C.: Cognitive characteristics of children with Mathematics Learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities 40*, 5 (9 2007), 458–478. doi:10.1177/00222194070400050901. 5

[MZNN19] MIUNDY K., ZAMAN H. B., NORDIN A., NG K. H.: SCREENING TEST ON DYSCALCULIA LEARNERS TO DEVELOP A SUITABLE AUGMENTED REALITY (AR) ASSISTIVE LEARNING APPLICATION. *Malaysian Journal of Computer Science* (11 2019), 92–107. doi:10.22452/mjcs.sp2019no1.7. 2

[Neu] NEUREKALAB: NeurekaTEST para escuelas. URL: https://neurekalab.es/lang/es/neurekatests_escuelas.html. 3

[Pal24] PAL S.: A comparative analysis of machine learning algorithms for predictive analytics in healthcare. *Heritage Research Journal 72*, 3 (03 2024), 10 – 25. 6

[PP15] PETERSON R. L., PENNINGTON B. F.: Developmental dyslexia. *Annual Review of Clinical Psychology 11*, 1 (3 2015), 283–307. doi:10.1146/annurev-clinpsy-032814-112842. 1

[R C23] R CORE TEAM: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL: https://www.R-project.org/. 4, 6

[Rap16] RAPIN I.: Dyscalculia and the Calculating Brain. *Pediatric Neurology 61* (8 2016), 11–20. doi:10.1016/j.pediatrneurol.2016.02.007. 1

[Ray24] RAY S.: Top 10 machine learning algorithms to use in 2024, 3 2024. URL: https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms. 5

[RBR22] RAUSCHENBERGER M., BAEZA–YATES R., RELLO L.: A universal screening tool for Dyslexia by a Web-Game and machine learning. *Frontiers in computer science 3* (1 2022). doi:10.3389/fcomp.2021.628634. 2

[RBVP*20] RE A. M., BENAVIDES-VARELA S., PEDRON M., DE GENNARO M. A., LUCANGELI D.: Response to a specific and digitally supported training at home for students with mathematical difficulties. *Frontiers in Psychology 11* (8 2020). doi:10.3389/fpsyg.2020.02039. 2

[RH09] RUBINSTEN O., HENIK A.: Developmental dyscalculia: Heterogeneity might not mean different mechanisms. *Trends in cognitive sciences 13* (02 2009), 92–9. doi:10.1016/j.tics.2008.11.002. 1

[RSM20] ROHIZAN R., SOON L. D., MUBIN S. A.: MathFun: A Mobile App For Dyscalculia Children. *Journal of physics* (12 2020). doi:10.1088/1742-6596/1712/1/012030. 2

[RSW*09] RÄSÄNEN P., SALMINEN J., WILSON A., AUNIO P., DEHAENE S.: Computer-assisted intervention for children with low numeracy skills. *Cognitive Development 24*, 4 (1 2009), 450–472. doi:10.1016/j.cogdev.2009.09.003. 2

[SMC*23] SZKODO M. O. R. P. M., MICAI M., CARUSO A., FULCERI F., FAZIO M., SCATTONI M. L.: Technologies to support the diagnosis and/or treatment of neurodevelopmental disorders: A systematic review. *Neuroscience biobehavioral reviews/Neuroscience and biobehavioral reviews 145* (2 2023), 105021. doi:10.1016/j.neubiorev.2022.105021. 2

[Tav24] TAVASOLI S.: Top 10 Machine learning Algorithms for Beginners: Supervised, and more, 2 2024. URL: https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article. 5

[WDP*06] WILSON A., DEHAENE S., PINEL P., REVKIN S. K., COHEN L. D., COHEN D.: Principles underlying the design of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions 2*, 1 (5 2006). doi:10.1186/1744-9081-2-19. 1

[WMP*19] WILLCUTT E. G., MCGRATH L. M., PENNINGTON B. F., KEENAN J. M., DEFRIES J. C., OLSON R. K., WADSWORTH S. J.: Understanding Comorbidity Between Specific Learning Disabilities. *New Directions for Child and Adolescent Development 2019*, 165 (5 2019), 91–109. doi:10.1002/cad.20291. 1

[Wor04] WORLD HEALTH ORGANIZATION: ICD-10 : international statistical classification of diseases and related health problems : tenth revision, 2004. URL: https://apps.who.int/iris/handle/10665/42980. 1

[Zer15] ZERAFA E.: Helping Children with Dyscalculia: A Teaching Programme with three Primary School Children. *Procedia - Social and Behavioral Sciences* (6 2015). doi:10.1016/j.sbspro.2015.04.516. 1