Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 18 No. 15 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijet.v18i15.40957

PAPER

Towards an Intelligent Model for Evaluating Serious Games

Kamal Omari(✉)

Faculty of Sciences
Ben M'Sik, Hassan II
University of Casablanca,
Casablanca, Morocco

kamal.omari2013@gmail.com

ABSTRACT

Serious games are effective educational tools used in higher education to provide practical learning opportunities to students. However, few research works have focused on evaluating serious games as a project for developing a tool dedicated to use in a formative context. This document proposes an intelligent evaluation model that not only allows for the evaluation of serious games but also facilitates their integration into teaching practice. The model is designed around four dimensions, and their measurement criteria are well defined. Fuzzy decision-making methods were used to weight the criteria, and supervised machine-learning algorithms were considered to minimize the evaluator's bias. The proposed model provides a more objective and consistent solution for evaluating serious games, reducing the impact of evaluators' biases and subjective preferences on the weightings of the different evaluation dimensions. The multi-output support vector regression (M-SVR) model can be used flexibly and adapted to different contexts and applications, offering a more effective and reliable solution for evaluating serious games.

KEYWORDS

serious games, FMADM, Fuzzy AHP, Fuzzy TOPSIS, Fuzzy ELECTRE, machine learning, M-SVR

1 INTRODUCTION

Serious games, as a pedagogical resource in modern education, have gained increasing popularity due to their interactive and playful interfaces for learning [1]. These games are designed to complement traditional training methods by providing immersive and meaningful learning environments for learners [2]. Educational institutions are leveraging the high interest of current generations of students in video games by incorporating serious games as innovative educational resources [3].

To effectively use serious games in the learning process, it is crucial for trainers to be involved and for adequate materials and well-organized logistics to be available for smooth integration of the games [4]. Additionally, the appropriate choice of a serious game is essential to achieve the desired pedagogical objectives [5]. However,

Omari, K. (2023). Towards an Intelligent Model for Evaluating Serious Games. *International Journal of Emerging Technologies in Learning (iJET)*, 18(15), pp. 79–93. https://doi.org/10.3991/ijet.v18i15.40957

Article submitted 2023-04-29. Resubmitted 2023-06-08. Final acceptance 2023-06-08. Final version published as submitted by the authors.

evaluating serious games, not just as a training tool but as a culmination of a development project designed for formative use, can be challenging, as there are limited tools available for this purpose.

To address this gap, an intelligent evaluation tool has been proposed that measures four key dimensions—pedagogical, technological, playful and behavioral—according to well-defined criteria. This evaluation tool is expected to assist practitioners in evaluating serious games in various training contexts, facilitating their integration into teaching practice.

It is important to note that the human factor (evaluator) can significantly influence the weighting result of the evaluation dimensions. To minimize this impact and maintain correlation between the evaluation system variables, supervised machine-learning algorithms such as multi-output support vector regression (M-SVR) [6] have been considered.

By utilizing this evaluation tool, the integration of serious games in the learning process can be made easier and more effective. The ultimate goal is to facilitate the use of serious games in training contexts, enabling learners to acquire the skills they need for success in their future professional lives.

2 STATE OF THE ART

Once a serious game has been designed, it is imperative to evaluate it. The evaluation process should verify whether the serious game is in line with the objectives for which it was created. According to Liu and Ding [7], these serious games require an appropriate evaluation system because they are closely linked to educational objectives. Without adequate evaluation, integrating serious games as a training solution will not be possible. as Kevin Corti, CEO of PIXELearning, pointed out, “Evaluation is the future of Serious Games” [8]. In the literature, existing evaluation frameworks for serious games can be classified into three axes: evaluating the quality of serious games, evaluating their effectiveness, and holistic evaluation of serious games.

2.1 The quality evaluation of serious games

The evaluation of the quality of a serious game is an important aspect to ensure its effectiveness in various contexts, particularly in the field of education. Several approaches have been proposed to evaluate the quality of serious games, such as El Borji’s evaluation grid [9], which aims to evaluate games designed for teaching computer programming. The results showed that each game has its own strengths and weaknesses, but they all have pedagogical specification needs. Another approach, proposed by Abdellatif et al. [10], proposes an evaluation framework based on criteria such as usability, comprehensibility, motivation, engagement, and user experience. The results of a study conducted with students from Queen’s University Belfast showed that comprehensibility can affect other quality characteristics. To evaluate the quality of serious games in terms of motivation, user experience, and learning, Savi et al. [11] proposed an evaluation model called MEEGA. which provides a questionnaire to gather data on students’ perception after playing a serious game. The evaluation results of this model showed that the MEEGA questionnaire is reliable and valid, but it is necessary to group the criteria related to motivation and user experience to improve the feedback. Following this, Petri et al. [12] developed an enhanced version of the MEEGA model, named MEEGA+, which was validated on a large scale.

2.2 The effectiveness evaluation of serious games

Although serious games are recognized as effective educational tools in the educational field, some researchers have proposed reliable and automated methodologies to measure their effectiveness. Among these methods, Serrano-Laguna and colleagues [13] have developed a framework that allows for a systematic evaluation of the effectiveness of serious games using integrated computerized tracking within the game. This framework automatically converts the serious game into an evaluation tool and allows for instantaneous measurement of learning outcomes. It has been successfully tested on the game The Foolish Lady. De Freitas and Oliver [14] have also proposed a four-dimensional framework that allows teachers to assess the potential of serious games based on their needs. This framework has been successfully applied to two examples. However, some researchers [15] have criticized this framework for its lack of assistance to teachers in classifying serious games that meet their needs. Emmerich and colleagues [8] have also presented an evaluation-focused design framework for serious games that emphasizes the role of evaluating the game's objectives rather than usability testing or overall player experience issues. This framework provides guidance for planning and conducting an evaluation of a serious game and highlights the similarity of the game development process with scientific processes. According to its authors [8], this framework is not a means of measuring whether the serious game is effective in achieving its objective, but it provides a solid foundation for constructive discussion.

2.3 The holistic evaluation of serious games

Researchers have developed evaluation frameworks for serious games, recognizing their complexity. These frameworks aim to provide a toolbox to facilitate the evaluation of these games. The evaluation framework proposed by Ghannem [16] defines generic criteria to evaluate learning objectives and scenario specification. It also uses game theory to provide strategic options to players. The framework by Mayer et al. [17], consisting of eight essential steps, provides comprehensive recommendations for evaluating serious games. It has been tested on twelve games but is limited in terms of considerations related to game systems. The framework by Wilson et al. [18] focuses on evaluation in the four key conceptual foundations, while the SGDA Framework by Mitgutsch et al. [19] offers a structure for the formal conceptual study of serious games in relation to their explicit and implicit objectives. The latter has been successfully applied to online games such as Sweatshop and ICED to provide a structured discussion on design elements relative to the game's purpose.

In conclusion, it is possible to classify the evaluation objectives of serious games into three categories: quality evaluation, effectiveness evaluation, and holistic evaluation. However, there is no consensus on a specific evaluation approach, with each approach having advantages and disadvantages. For example, while the methodology proposed by Mayer et al. [17] is considered the most comprehensive, there is insufficient information on its applicability and validity. On the other hand, Petri's [12] MEEGA+ evaluation model is the most applied and evaluated in the scientific literature, but its creator has emphasized that this model is not suitable for all serious games and that specific criteria are necessary. This assertion is in line with Mayer et al.'s [20] observation on the lack of comprehensive frameworks and operationalized models. As a result, there is a need to identify a consistent and uniform approach to systematically evaluate serious games as the culmination of a development project dedicated to use in a formative context.

3 PROPOSAL OF AN INTELLIGENT MODEL FOR EVALUATING SERIOUS GAMES

In order to overcome the current scientific barriers, we will describe below our intelligent model for the evaluation of serious games, which aims to offer evaluators the possibility to adapt the tool to the context of use of the serious game they wish to evaluate. We have taken several steps to design this intelligent evaluation tool capable of evaluating serious games in different usage contexts.

3.1 Dimensions and criteria adopted

We have developed an intelligent evaluation model for serious games, based on four essential evaluation dimensions that every serious game must meet in order to fulfill its pedagogical mission: pedagogical (P), technological (T), ludic (L), and behavioral (B) dimensions.

We chose the pedagogical dimension because a serious game must meet one or more pedagogical objectives for which it was developed. Similarly, to ensure effective learning, a serious game must be attractive and benefit from the latest technological advances in the video game industry. The ludic dimension is also essential, as it allows learning in a fun and immersive environment, which maintains learners' attention and interest. Finally, the behavioral dimension allows testing the relevance of a serious game in its context of use, by measuring the motivation, engagement, and experience of learners.

These pedagogical, technological, ludic, and behavioral dimensions will be evaluated according to specific criteria that we present in the Table 1. These criteria have been selected based on an exhaustive literature review, including scientific articles [21, 17] and previous research in the field of serious games [22, 23]. The selection of criteria has been guided by their relevance and appropriateness to the objectives of our study, as well as their ability to rigorously evaluate each dimension. Adaptations and adjustments have been made to account for the specific context of our study, ensuring a comprehensive and accurate evaluation of the dimensions of the serious game.

Table 1. Dimensions and criteria

Dimensions	Pedagogical	Technological	Ludic	Behavioral
Criteria for measurement	– Targeted skills – Pedagogical consideration – Learning result – Error management	– Game design – Performance – User interface – Usability	– Challenge – Fun – Gameplay – Immersion	– Motivation – Engagement – User experience

The relative importance of each dimension depends on the context of use of the serious game. In a purely educational context, the pedagogical dimension is considered predominant compared with other dimensions. Therefore, it is crucial to adjust all the criteria adopted according to the context of use of the serious game. In our two subsequent studies [24, 25], we adopted methods based on the fuzzy multi-attribute decision-making approach (FMADM) to address this issue. In our first study [24], we decided to use the fuzzy analytic hierarchy process (FuzzyAHP) method [26] to ensure the validity of evaluator preferences. This approach ensures internal

consistency in the criteria-weighting process incorporated into the serious game evaluation tool.

For the second study [25], we also used two other fuzzy multi-attribute decision-making methods—namely, Fuzzy TOPSIS [27] and Fuzzy ELECTRE [28]—to evaluate the same criteria in the same study context as in our first study. The results obtained by these FMADM methods led to a convergent ranking of the selected criteria/dimensions. To confirm the effectiveness of our serious game evaluation model, we compared our tool with the MEEGA+ model [12] through a serious game called Leuco' War. The results showed the quality and relevance of our proposed evaluation tool.

These two contributions concluded that the influence of evaluators on the choices of preferences for serious game evaluation dimensions is an important factor to consider. Moreover, it was observed that there is a correlation between the dimensions of our serious game evaluation tool. Therefore, we must continue to refine our approach to take these factors into account and further improve our evaluation model.

In order to reduce human influence in the serious game evaluation process while maintaining the correlation between these dimensions, we proposed the use of a supervised self-learning algorithm (M-SVR) [6]. This algorithm allows for self-regulation of the weights of the dimensions adopted according to the context of use of the serious game being evaluated. Figure 1 presents our intelligent model for serious game evaluation.

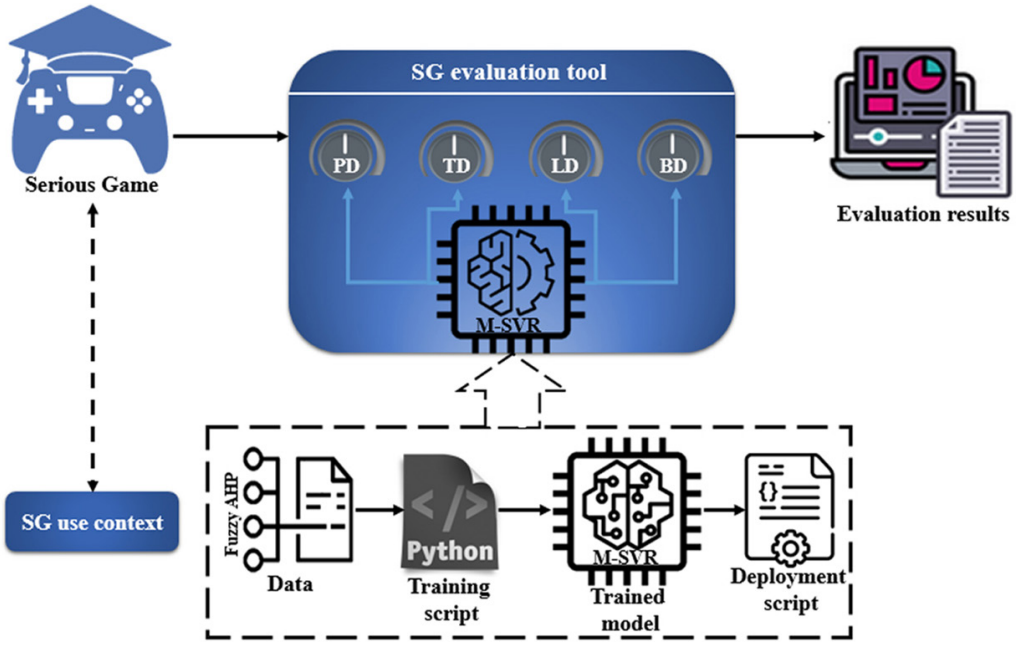

Fig. 1. Intelligent model for serious game evaluation

Figure 1 presents the proposed intelligent model for evaluating serious games in this study. This model is based on the use of the M-SVR algorithm to self-regulate the weighting of evaluation dimensions based on the specific context of serious game usage. The M-SVR algorithm minimizes human influence in the evaluation process while maintaining correlation among the different evaluation dimensions [6]. The model incorporates an automated approach to adjust the evaluation criteria based

on the specific context in which the serious game is being used, ensuring a more objective and contextually adapted evaluation. This innovative approach aims to enhance the quality and relevance of serious game evaluation.

3.2 M-SVR algorithms

M-SVR algorithms [6] are supervised machine-learning approaches that allow for simultaneous prediction of multiple continuous variables from a common set of attributes. This approach is particularly useful in systems with multiple factors and attributes, as it considers not only the relationships between attributes and targets but also the relationships among the targets themselves, thereby ensuring better predictive performance compared with single-output supervised machine-learning algorithms. The M-SVR algorithm proposed by Pérez-Cruz et al. [6] is an improvement of the basic SVR algorithm [29] that aims to handle correlated multiple outputs. It uses the iterative reweighted least-squares (IRWLS) method [30] to solve problems related to redefining the ε-insensitive loss function in the hyper-sphere space as well as obtaining the multi-output prediction model for each output. In practice, the M-SVR algorithm learns a correspondence between the multidimensional input space and the multidimensional output space [31], while capturing existing dependencies and internal relations to ensure optimal predictive performance. It is commonly used to achieve good predictive performance when predicting multiple outputs simultaneously [32], while speeding up calculations, obtaining a clearer representation (by avoiding the use of the same support vector multiple times) compared with problem transformation methods, and maintaining similar error rates to problem transformation approaches. We have chosen to use the M-SVR algorithm in our serious game evaluation system because it can confidently predict correlated multiple outputs, which is crucial for evaluating the overall quality of a serious game by considering multiple criteria at once [33]. Moreover, this algorithm is adaptable to the parameters of our system, making it a suitable choice for our application.

3.3 Description of the M-SVR algorithm

The goal of the M-SVR algorithm is to find the regressor w^j and b^j ($j = 1, ..., m$) for each output, that minimizes the following function (1):

$$\min_{W,b} L_p = \frac{1}{2}\sum_{j=1}^{m}\left\|w^j\right\|^2 + C\sum_{i=1}^{N} L(u_i) \tag{1}$$

where:

$$u_i = \left\|e_i\right\| = \sqrt{e_i^T e_i},\ e_i^T = y_i^T - \varphi(x_i)^T W - b^T$$

$W = [w^1, ..., w^m]$ is the vector of coefficients for multiple outputs,

$b = [b^1, ..., b^m]^T$ is the constant vector representing the bias for each output,

C is the regularization parameter that balances the model complexity and approximation accuracy,

$\varphi(.)$ denotes a nonlinear mapping from the n-dimensional input space to an m-dimensional feature space $\mathbb{R}^n \rightarrow \mathbb{R}^m$.

$L(u)$ is a quadratic-insensitive cost function, defined by the following equation (2):

$$L(u) = \begin{cases} 0, & u < \varepsilon \\ u^2 - 2u\varepsilon + \varepsilon^2, & u \geq \varepsilon \end{cases} \tag{2}$$

When $\varepsilon = 0$, in equation (2), the problem reduces to a regularized kernel least-squares regression for each component independently.

For $\varepsilon \neq 0$, it takes into account all outputs to construct regressors for each individual, and then produces a single support vector for all dimensions, aiming to obtain more robust predictions.

To solve equation (1), an iterative method called iteratively reweighted least squares (IRWLS) [30] has been used in these two studies [34, 30].

By introducing a first-order Taylor expansion of the cost function $L(u)$, the objective of equation (1) is approximated by the following equation (3):

$$Lp'(W,b) = \frac{1}{2}\sum_{j=1}^{m} \left\| w^j \right\|^2 + \frac{1}{2}\sum_{i=1}^{N} a_i u_i^2 + C \tag{3}$$

where:

$$a_i = \begin{cases} 0, & u_i^k < \varepsilon \\ \dfrac{2\gamma(u_i^k - \varepsilon)}{u_i^k}, & u_i^k \geq \varepsilon \end{cases} \tag{4}$$

C is a constant that does not depend on W and b, and the exponent k denotes the k-th iteration.

To optimize equation (3), an IRWLS procedure is constructed that linearly searches for the solution of the next step in the downward direction based on the previous solution [30].

According to the representative theorem [35], the best solution for minimizing equation (4) in the feature space can be expressed in the form $w^j = \sum_i \varphi(x_i)\beta^j$, so that the objective of M-SVR is transformed into searching for the best β and b.

The IRWLS method of M-SVR can be summarized by the following steps [36]:

Step 1: Define $k = 0$, $\beta^k = 0$, $b^k = 0$ and compute u_i^k and a_i.

Step 2: Compute the solution β^s *and* b^s according to the following equation:

$$\begin{bmatrix} K + D_a^{-1} & 1 \\ a^T K & 1^T a \end{bmatrix} \begin{bmatrix} \beta^j \\ b^j \end{bmatrix} = \begin{bmatrix} y^j \\ a^T y^j \end{bmatrix}, \; j = 1, \ldots, m \tag{5}$$

where $a = [a_1, ..., a_n]^T$, is the kernel matrix, D is a diagonal matrix of a. Define the descending direction corresponding to the direction of according P^k to the following equation:

$$P^k = \begin{bmatrix} w^s - w^k \\ (b^s - b^k)^T \end{bmatrix} \tag{6}$$

Note that equation (6) is not a vector but a matrix in which each column is a descending direction for each regressor.

Step 3: Use a recursive algorithm to compute β^{k+1} and b^{k+1} until convergence.

After this brief description of how the M-SVR algorithm works, we present in the next section its implementation in our serious game evaluation system.

4 MODELING PROCESS OF M-SVR IN THE SERIOUS GAME EVALUATION TOOL

The objective of machine learning is to enable the M-SVR algorithm implemented in our evaluation system to learn a mapping between:

- The input vector (X) $X^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$,
- And the output vector (Y) $Y^{(i)} = (y_1^{(i)}, \ldots, y_m^{(i)})$,
- From the training data set (D) $D = \{(X^{(i)}, Y^{(i)})\}_{i=1}^{N} \subset R^n \times R^m$ for N samples.

The goal is to find a function h that relates the input vector X to the output vector Y, $h(x) = Y$.

Thus, for a given new input vector $\hat{X}$, the model will be able to predict an output vector $\hat{Y}$.

$\hat{Y} = h(\hat{X})$ that best approximates the actual output vector Y.

The modeling process of M-SVR in the serious game evaluation tool consists of the steps shown in Figure 2.

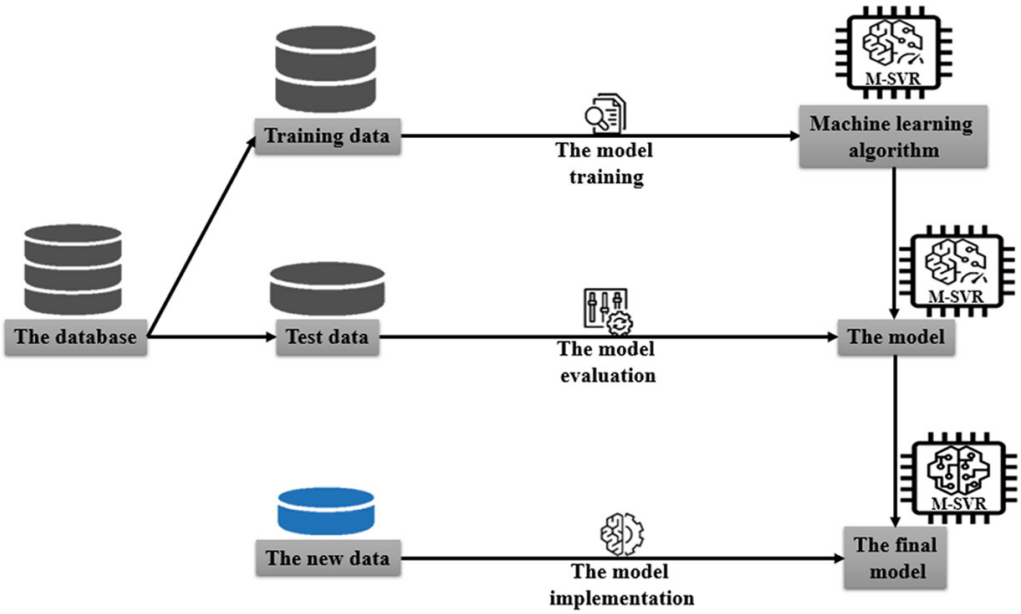

Fig. 2. Steps performed in the modeling process

Step 1: Data preparation—Collect and pre-process the data

During the data-preparation phase, we collected and preprocessed the data to create the training dataset (D). This step was performed using a MATLAB program that employed the Fuzzy AHP method as an approach for data extraction.

The data-collection process involved gathering input features, denoted as $X^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$, which were comprised of n = 6 features that represented the contexts of use for serious games from available sources. These input features were then processed using the Fuzzy AHP method, a fuzzy logic–based approach that incorporated expert knowledge to extract meaningful information from the data. The Fuzzy AHP method allowed for the aggregation of multiple criteria or dimensions of evaluation into a single representation, denoted as $Y^{(i)} = (y_1^{(i)}, \ldots, y_m^{(i)})$, which was comprised of m = 4 features, that could be used as the output labels for the M-SVR model.

Once the data was collected and processed, it was represented as $D = \{(X^{(i)}, Y^{(i)})\}_{i=1}^{N}$, where N = 2500, due to the time complexity of the M-SVR model adjustment, which becomes difficult to handle for datasets with more than 10,000 samples.

Data-preprocessing tasks such as normalization or scaling were performed as needed to ensure that the input features and output labels were on the same scale and had similar ranges. This was done to prevent any particular feature from dominating the model-training process due to differences in magnitudes.

Table 2 displays an example of the training vector pair (X, Y) used in this step.

Table 2. Training data

Input Vectors: Serious Games Usage Contexts						Output Vectors: Evaluation Dimension Weights			
x_1	x_2	x_3	x_4	x_5	x_6	y_1	y_2	y_3	y_4
0.5	3	0.2	0.2	0.5	0.5	0.135	0.133	0.154	0.578
7	1	7	3	5	3	0.508	0.234	0.196	0.062
5	1	3	7	9	5	0.387	0.371	0.182	0.06
3	5	5	0.142	7	7	0.514	0.134	0.307	0.045
7	3	0.125	5	3	5	0.307	0.293	0.189	0.212
7	3	3	5	7	0.333	0.527	0.279	0.076	0.118
2	5	7	3	5	3	0.511	0.310	0.122	0.057
0.333	5	0.125	3	0.142	1	0.138	0.214	0.104	0.544
0.2	1	0.2	1	0.2	3	0.100	0.223	0.290	0.386
0.5	3	0.2	5	0.2	0.5	0.128	0.247	0.076	0.549
3	5	7	5	5	7	0.540	0.293	0.124	0.043
5	9	5	5	3	1	0.631	0.219	0.064	0.086

Data-preprocessing tasks such as normalization or scaling were performed as needed to ensure that the input features and output labels were on the same scale and had similar ranges. This was done to prevent any particular feature from dominating the model-training process due to differences in magnitudes.

This data-preparation process was crucial to ensure that the training dataset (D) was well prepared and suitable for training the M-SVR model.

Step 2: Model configuration

In our experimentation, we implemented the M-SVR algorithm using the Python programming language, with the help of the Scikit-Learn machine-learning library [37]. This implementation step is the most time-consuming during the development of a machine-learning model.

It is important to parameterize the M-SVR algorithm in order to better adapt it to the requirements of our serious game evaluation tool. This helps minimize errors in the constructed model and takes into account the specificities present in our data. Thus, we need to examine the learning curve to see if the M-SVR algorithm is overfitting or underfitting the problem under study. This helps identify possible avenues for hyperparameter tuning.

For the M-SVR algorithm, there are three hyperparameters:

- The regularization parameter C, which controls the trade-off between empirical risk and the regularized term.

- The kernel parameter γ, which determines the similarity between samples in the feature space.
- The error parameter ε, which defines the width of the insensitive tube.

These hyperparameters are determined through a trial-and-error approach. In our case, the initial values assigned to these parameters with kernel = 'rbf' are as follows: C = 40, γ = 0.001, and ε = 0.001.

Step 3: Training the model

Following the recommendations from the scientific literature on machine learning [38], we employed an 80/20 split of the training dataset for model training and evaluation. This widely used approach allows us to assess the performance of the machine-learning model and ensure that it can generalize well to unseen data.

The training dataset, which was previously preprocessed and prepared during the data-preparation phase, as discussed earlier, was randomly divided into an 80% subset for training and a 20% subset for evaluation. The training subset was used to train the M-SVR model using the optimization process and iterative reweighted least=squares (IRWLS) algorithm. The objective was to minimize the cost function and learn the relationships between the input features and output labels in the data. The evaluation subset was kept separate and was not used during the training process.

Step 4: Model evaluation

After the training phase, the evaluation data is utilized to assess the performance of the model using the R^2 technique [39], which is a statistical measure indicating the quality of the regression model's fit. A higher R^2 value signifies a more accurate regression model.

$$R^2 = \frac{SS_{reg}}{SS_{tot}} = 1 - \frac{SS_{err}}{SS_{tot}} \tag{7}$$

where, SS_{reg} represents the sum of squares explained by the regression, SS_{tot} refers to the total sum of squares, and SS_{err} indicates the sum of squared errors.

Additionally, the root mean squared error (RMSE) [40] is calculated to estimate the standard deviation of errors that arise when making predictions on a dataset. A smaller RMSE value indicates fewer errors in the model's predictions.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left\| y(i) - \ddot{y}(i) \right\|^2}{N}} \tag{8}$$

where N denotes the number of data points, $y(i)$ represents the ith measurement, and $\ddot{y}(i)$ corresponds to its prediction.

Table 3. Model testing

Test Input Vectors						Actual Output Vectors				Model-Predicted Output Vectors			
x_1	x_2	x_3	x_4	x_5	x_6	y_1	y_2	y_3	y_4	$\ddot{y}_1$	$\ddot{y}_2$	$\ddot{y}_3$	$\ddot{y}_4$
7	3	0.125	5	0.125	3	0.271	0.117	0.147	0.465	0.270	0.126	0.140	0.462
7	1	7	3	5	1	0.524	0.241	0.153	0.082	0.519	0.237	0.146	0.094
3	3	7	7	7	0.333	0.492	0.357	0.065	0.087	0.444	0.409	0.070	0.077
1	1	3	7	9	9	0.242	0.515	0.195	0.049	0.249	0.497	0.190	0.060
9	5	0.125	3	0.142	3	0.317	0.098	0.144	0.441	0.304	0.102	0.138	0.454

Table 3 presents the results of the model evaluation phase. The R^2 value obtained is 0.98, indicating that the regression model explains 98% of the variance in the data. The RMSE is 0.016, which represents the standard deviation of errors in the model's predictions. A smaller RMSE value indicates better performance, and in this case, the model produces relatively accurate predictions with low error.

Step 6: Model deployment

After training and fine-tuning the M-SVR model with the training data, the next step is to deploy the model in the serious game evaluation tool. In this step, we use the intelligent model created with the M-SVR algorithm in the same evaluation context as presented in our initial contributions [24, 25], where the serious game is used purely for educational purposes and targets a university and scientific audience.

As such, we model the context of use of the serious game using the input vector $\boldsymbol{X}$ **= {3,5,7,3,5,3}**, which will be fed into the final model created by M-SVR to obtain the appropriate weights $\{y_1, y_2, y_3, y_4\}$ for each evaluation dimension of the serious game, as illustrated in Table 4.

Table 4. Results of obtained dimension weights

Context of use for serious games {3,5,7,3,5,3}	PD	TD	BD	LD
	0.557	0.270	0.123	0.052

The Table 4 shows that the results obtained from the intelligent weighting model indicate the same order of the four dimensions used in our two previous contributions [24, 25]. It should be noted that the influence of the human factor, i.e., the evaluator, during the M-SVR weighting process is neglected. This is because the intelligent process that was used creates a model utilizing the capabilities of supervised machine learning provided by the M-SVR algorithm. This allowed it to generate correlated and appropriate weights for each dimension with respect to the context of use of the serious game.

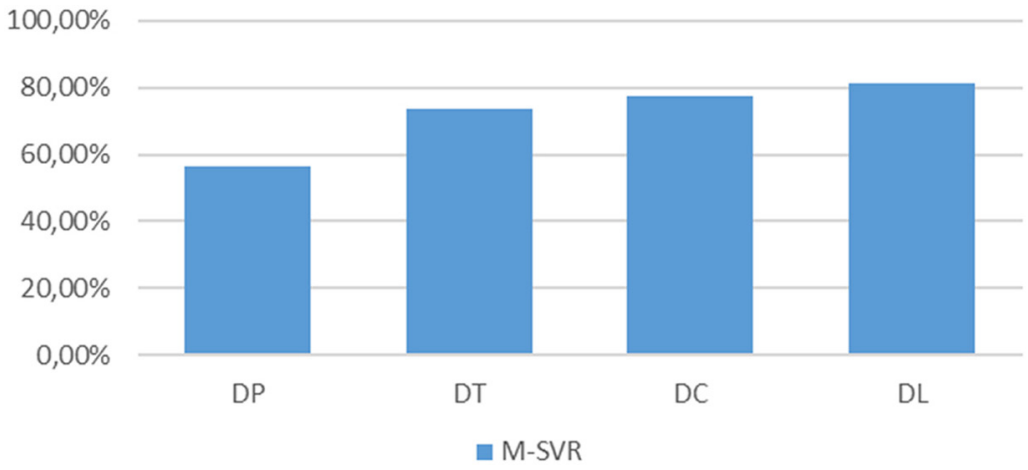

Fig. 3. Evaluation results of the serious game Leuco's war

As shown in Figure 3, these results confirm those obtained in the study [24], indicating that the serious game Leuco's War is more suitable for use in a gaming context rather than in a purely formative context like ours.

5 DISCUSSION

The results obtained in our study showed convergence of the FMADM methods used, allowing us to use the weights obtained from the Fuzzy AHP method to construct a database used by the multi-output machine-learning algorithm M-SVR for evaluating a serious game with our intelligent evaluation tool.

The implementation of the M-SVR algorithm was carried out using the Python programming language and the Scikit-Learn machine-learning library. This implementation step was time-consuming but necessary to adapt the M-SVR algorithm to the needs of our serious game evaluation tool by minimizing errors in the constructed model and taking into account the peculiarities of our data. We also examined the learning curve to detect potential overfitting or underfitting issues and identify possible avenues for hyperparameter tuning.

We consider our work as making a significant contribution to the scientific research on serious games evaluation. Indeed, our contributions have added a serious games evaluation model to the existing models and techniques, which are often applicable only to specific contexts.

However, it is important to acknowledge the limitations of our serious games evaluation model. Firstly, our feedback is based solely on a serious game in a specific domain, which may not be fully representative, given the diversity of serious games available on the market and designed for different contexts of use. Moreover, our analyses were conducted on a single sample—namely, the group of biology students at Hassan II University of Ben M'Sik—which limits the generalizability of our results to other target populations or higher education domains.

Finally, it should be noted that other technical and logistical parameters were not taken into account in our model, such as the size of the erious ame used in our tests and the hardware requirements to ensure smooth operation of the serious game. These aspects could influence the evaluation results and should be considered in future research.

In conclusion, while our serious games evaluation model is capable of fulfilling its intended function, it has certain methodological and contextual limitations that need to be considered in the interpretation of results and in future research in this field.

6 CONCLUSION

By incorporating an intelligent process such as M-SVR into our serious game evaluation system, we were able to minimize subjective evaluations introduced by human factors, as automatic weighting values were assigned based on the context of use of the chosen serious game. This helped to reduce bias and ensure a more objective evaluation process.

Furthermore, the performance of the algorithm was assessed using test parameters such as RMSE and accuracy, which were found to be 0.016 and 98.59%, respectively, indicating acceptable performance of the algorithm.

However, during our experimentation, we noticed some instability of the algorithm when dealing with very large dataset volumes. This suggests that the algorithm's performance may degrade with increasing dataset size, and further investigation is needed to address this issue.

To address this, we plan to compare the results obtained with M-SVR against other algorithms of similar nature that are more stable when handling datasets with over

10,000 samples. This is crucial, as the adjustment time complexity of the algorithm is more than quadratic, and finding a more stable alternative will be beneficial for large-scale evaluations.

In addition, we plan to implement an intelligent process in our system that establishes a direct link between the context of use of the serious game and the corresponding serious game itself. This will further enhance the accuracy and relevance of our evaluations, as the context of use is a critical factor in determining the suitability and effectiveness of serious games for specific purposes or target audiences.

Overall, the integration of M-SVR and the planned improvements to our system will contribute to a more robust and reliable serious game evaluation process, minimizing subjective bias and improving the accuracy of results.

7 REFERENCES

[1] Michael, D. R., & Chen, S. L. (2006). Serious games: Games that educate, train, and inform. Thomson Course Technology.

[2] Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. Personnel Psychology, 64(2), 489–528. https://doi.org/10.1111/j.1744-6570.2011.01190.x

[3] Lieberman, D. A. (2006). What can we learn from playing interactive games? In R. E. Ferdig (Ed.), Handbook of research on effective electronic gaming in education (pp. 1–16). Hershey, PA: Information Science Publishing.

[4] Gee, J. P. (2003). What video games have to teach us about learning and literacy. Palgrave Macmillan. https://doi.org/10.1145/950566.950595

[5] Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., de Freitas, S., Louchart, S., ... & Berta, R. (2015). Mapping learning and game mechanics for serious games analysis. British Journal of Educational Technology, 46(2), 391–411. https://doi.org/10.1111/bjet.12113

[6] Pérez-Cruz, Fernando & Camps-Valls, Gustau & Olivas, Emilio & Perez-Ruixo, Juan & Figueiras-Vidal, Aníbal & Artés Rodríguez, Antonio. (2002). Multi-dimensional function approximation and regression estimation. Lecture Notes in Computer Science—LNCS, 757–762. https://doi.org/10.1007/3-540-46084-5_123

[7] Liu, S., Ding, W. (2009). An approach to evaluation component design in building serious game. In: M. Chang, -R. Kuo, Kinshuk, G.-D. Chen and M. Hiroshe, Edutainment '09 Proceedings of 4th International Conference on E-learning and Games: Learning by Playing. Game-based Education System Design and Development, (pp. 141–148). Berlin: Springer-Verlag. https://doi.org/10.1007/978-3-642-03364-3_18

[8] Emmerich, Katharina & Bockholt, Mareike. (2016). Serious games evaluation: Processes, models, and concepts. https://doi.org/10.1007/978-3-319-46152-6_11

[9] El Borji, Yassine. (2014). Comparative study to develop a tool for the quality assessment of serious games intended to be used in education. International Journal of Emerging Technologies in Learning (iJET), 9. https://doi.org/10.3991/ijet.v9i9.4150

[10] Abdellatif, A. J., McCollum, B., & McMullan, P. (2018). Serious games: Quality characteristics evaluation framework and case study. In 2018 IEEE Integrated STEM Education Conference (ISEC): Proceedings (pp. 112–119). IEEE. https://doi.org/10.1109/ISECon.2018.8340460

[11] R. Savi, C. Gresse von Wangenheim, and A. F. Borgatto. "A model for the evaluation of educational games for teaching software engineering", Proc. of the 25th Brazilian Symposium on Software Engineering, 2011, pp. 194–203. São Paulo/SP, Brazil. https://doi.org/10.1109/SBES.2011.27

[12] Petri, Giani & Gresse von Wangenheim, Christiane. (2019). MEEGA+: A method for the evaluation of the quality of games for computing education. https://doi.org/10.5753/cbie.wcbie.2019.951

[13] Serrano-Laguna, Ángel & Manero, Borja & Freire, Manuel & Fernández-Manjón, Baltasar. (2018). A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes. Multimedia Tools and Applications, 77. https://doi.org/10.1007/s11042-017-4467-6

[14] De Freitas, Sara & Oliver, Martin. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? Computers & Education, 46, 249–264. https://doi.org/10.1016/j.compedu.2005.11.007

[15] Robertson, J., & Howells, C., (2008). Computer game design: Opportunities for successful learning. Computers & Education, 50(2), 559–578. https://doi.org/10.1016/j.compedu.2007.09.020

[16] Ghannem, Afef. (2014). Characterization of serious games guided by the educational objectives. https://doi.org/10.1145/2669711.2669904

[17] Mayer, Igor & Bekebrede, Geertje & Harteveld, Casper & Warmelink, Harald & Zhou, Qiqi & van Ruijven, Theo & Lo, Julia & Kortmann, Rens & Wenzler, Ivo. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. British Journal of Educational Technology, 45, 502–507. https://doi.org/10.1111/bjet.12067

[18] D. W. Wilson et al., (2016). Serious games: an evaluation framework and case study, in System Sciences (HICSS), 2016 49th Hawaii International Conference on, pp. 638–647: IEEE, Published. https://doi.org/10.1109/HICSS.2016.85

[19] Mitgutsch, Konstantin & Alvarado, Narda. (2012). Purposeful by design? A serious game design assessment framework. Foundations of Digital Games 2012, FDG 2012—Conference Program, 12. https://doi.org/10.1145/2282338.2282364

[20] Mayer, I. (2012). Towards a comprehensive methodology for the research and evaluation of serious games. Procedia Comput. Sci., 15, 233–247. https://doi.org/10.1016/j.procs.2012.10.075

[21] Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of the empirical evidence on computer games and serious games. Computers and Education, 59(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

[22] Calderon, A. and Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. Computers & Education, 87, 396–422. https://doi.org/10.1016/j.compedu.2015.07.011

[23] Bellotti, Francesco & Kapralos, Bill & Lee, Kiju & Moreno Ger, Pablo & Berta, Riccardo. (2013). Assessment in and of serious games: An overview. advances in human-computer interaction. https://doi.org/10.1155/2013/136864

[24] Omari, K., Moussetad, M., Labriji, E., & Harchi, S. (2020). Proposal for a new tool to evaluate a serious game. International Journal Of Emerging Technologies In Learning (IJET), 15(17), 238–251. https://doi.org/10.3991/ijet.v15i17.15253

[25] Omari, K., Harchi, S., Ouchaouka, L., Rachik, Z., Moussetad, M., & Labriji, E. (2021). Application the fuzzy topsis and fuzzy electre in the serious games evaluation tool. Journal of Theoretical and Applied Information Technology (JATIT), 99(9). http://www.jatit.org/volumes/Vol99No9/1Vol99No9.pdf

[26] Buckley, J. J. (1985). Fuzzy hierarchical analysis. Fuzzy Sets and Systems, 17, 233–247. https://doi.org/10.1016/0165-0114(85)90090-9

[27] Chen, C. T. (2000). Extensions of the TOPSIS for group decision-making under fuzzy environment. Fuzzy Sets and Systems, 114, 1–9. https://doi.org/10.1016/S0165-0114(97)00377-1

[28] Sevkli, M. (2010). An application of the fuzzy ELECTRE method for supplier selection. International Journal of Production Research, 48(12), 3393–3405. https://doi.org/10.1080/00207540902814355
[29] Vapnik, V. N. (1998). Statistical Learning Theory, John Wiley & Sons, Inc.
[30] Sánchez-Fernández, Matilde & De-Prado-Cumplido, Mario & Arenas-Garcia, Jerónimo & Pérez-Cruz, Fernando. (2004). SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems. Signal Processing, IEEE Transactions on, 52, 2298–2307. https://doi.org/10.1109/TSP.2004.831028
[31] Ny Hanitra, Ivan & Criscuolo, Francesca & Carrara, Sandro & Micheli, Giovanni. (2021). Multi-Ion-Sensing emulator and multivariate calibration optimization by machine learning models. IEEE Access, pp. 1–1. https://doi.org/10.1109/ACCESS.2021.3065754
[32] Borchani, Hanen & Varando, Gherardo & Bielza, Concha & Larranaga, Pedro. (2015). A survey on multi-output regression. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5. https://doi.org/10.1002/widm.1157
[33] Zhao, Wei & Liu, J. K. & Chen, Y. Y. (2015). Material behavior modeling with multi-output support vector regression. Applied Mathematical Modelling, 39. https://doi.org/10.1016/j.apm.2015.03.036
[34] Mao, W. T., et al. (2014a). A fast and robust model selection algorithm for multi-input multi-output support vector machine. Neurocomputing, 130, 10–19. https://doi.org/10.1016/j.neucom.2013.01.058
[35] Schölkopf, B., & Smola, A. J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press.
[36] Mao, W., Tian, M., & Yan, G. (2012). Research of load identification based on multiple-input multiple-output SVM model selection. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 226, 1395–1409. https://doi.org/10.1177/0954406211423454
[37] Pedregosa, F., Varoquaux, Gaël, Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
[38] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
[39] Windmeijer, F., & Cameron, A. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. Journal of Econometrics, 77, 329–342. https://doi.org/10.1016/S0304-4076(96)01818-0
[40] Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE) arguments against avoiding RMSE in the literature. Geosci Model Dev, 7(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

8 AUTHOR

Kamal Omari holds a Doctor's degree in computer science and serves as a professor in the field. He is a member of the Laboratory of Information Technology and Modelling at the Faculty of Sciences Ben M'Sik, University Hassan II of Casablanca, Morocco (e-mail: kamal.omari2013@gmail.com).