Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19614-w

A unified framework to catalogue and classify digital games based on interaction design and validation through clustering techniques

L. Cormio[1,2] · T. Agostinelli[1] · M. Mengoni[1]

Received: 11 December 2023 / Revised: 3 May 2024 / Accepted: 5 June 2024

Abstract
The digital games industry has grown exponentially due to the diversification of games and the increasing multiplicity of the user target base. The market explosion and the great variety make digital game cataloguing and classification challenging issues whose effectiveness can advance scientific research and address design, development and distribution. Firstly, the present study reviews previous cataloguing for video and serious games through systematic literature review and, joining together the findings from the literature review, develops a unified cataloguing model based on five definitions. This model can aid designers in tailoring their applications and contribute to disseminating game design knowledge in academic research. Then, a matrix that correlates design principles and the cataloguing model's metadata is applied to the cataloguing model, obtaining a unified classification system. Together, they offer a comprehensive framework for understanding the multifaceted landscape of digital games, addressing the limitations of existing domain-specific approaches and providing a versatile tool for game designers. Research validation exploits a two-stage cluster analysis using agglomerative hierarchical and k-means clustering on the data extracted from a sample of digital games. The results show the framework's effectiveness in categorizing digital games without a clear-cut distinction between video and serious games. The system's application in real-world scenarios suggests its potential to guide game development. Future work will refine the proposal based on feedback from digital game designers, expanding the research scope to include a broader range of games.

Keywords Digital games · Cataloguing · Classifying · Hierarchical clustering · K-means clustering · Design principles

✉ T. Agostinelli
t.agostinelli@pm.univpm.it

1 Department of Industrial Engineering and Mathematical Sciences, Polytechnic University of Marche, Via Brecce Bianche, 12, 60131 Ancona (AN), Italy

2 Department of Humanities, Letters, Cultural Heritage and Educational Studies, University of Foggia, Via Arpi, 176, 71121 Foggia (FG), Italy

1 Introduction

Over the years, the field of game design has gained increasing interest from both scientific and professional communities, particularly in relation to digital games. ‘Digital games’ is a broad term that encompasses videogames and serious games “played of any platform, online or offline” [1]. Videogames, as defined by Newman [2] have an entertainment purpose and distinguish themselves through interactivity, setting them apart from non-interactive media (e.g. books, films). Digital games are no longer merely for entertainment (videogames), as they also include serious games that serve practical purposes. Serious games have seen a surge in popularity since around 2002. Originating from educational purposes, they now incorporate diverse facets such as well-being, training, advertisement, and culture [3]. Examples of serious games include (i) Le Village Aux Oiseaux, a therapeutic game designed to train patients with Alzheimer’s disease to mitigate cognitive decline [4], and (ii) The Prepared Partner, an educational video game focusing on labour and childbirth preparation for parenthood [5, 6]. The digital games industry has experienced exponential growth [7], becoming an integral part of everyday life and seamlessly integrating into various sectors [3]. According to estimates, the global market for games generated between USD 81.5 billion and USD 93 billion in sales in 2015 [8]. By 2022, the market had grown to USD 197 billion [9]. The overall growth of the digital game industry partially depends on the increasing diversity of its user base and on the introduction of casual games [10], suitable for any user who is not necessarily a habitual digital game consumer (i.e., "gamer"). The ratio of male to female users has become more balanced [8], and user categories have expanded to include children, educators, corporate professionals, individuals in therapy, and those with various types of disabilities.

Due to the wide range of digital games available on the market, each with unique features, software solutions, interaction modalities, and targeted audiences, the attention of academic research has increasingly been captured by the dynamics of the game design process. Therefore, such variety entails to systematically arrange game design products to facilitate and foster the common understanding of this field for both academic and professional communities. Consequently, cataloguing and classifying digital games is a crucial and challenging activity in game design. Cataloguing is the process that grants users prompt access to information that aligns with their requirements. It involves identifying resources gathered by libraries, resulting in comprehensive metadata that serves various purposes [11]. Conversely, classification is a process that allows each entity to be systematically and methodically assigned to a single and distinct category [12]. Cataloguing models and classification systems are complementary: while the former provides a detailed description of the items, classification organises and divides them into categories based on common themes [13].

Some studies have attempted to catalogue or classify digital games, but they have focused solely on either videogames or serious games, rather than encompassing both. Other studies have included too many categories and parameters, which may be incompatible or cause confusion (see Section 2). Therefore, there is currently a lack of a comprehensive framework that provides a cataloguing model and classification system to facilitate the understanding of the multifaceted landscape of both video games and serious games. The present study aims to address this lack to define 1) a novel cataloguing model describing any digital game and 2) a classification system identifying the category it belongs.

Therefore, the Research Question driving the present paper is: "Is it possible to structure the empirical knowledge on game design identifying a unified model, helping game designers in collocating their products in well-defined classes with specific descriptors?".

To answer this RQ, the study conducts an analysis of the literature, employing a review methodology as suggested by [14]. This methodology entails to collect metadata from literature, systematically divide them, and categorise them on specific criteria.

The result of this process is a comprehensive cataloguing and classification framework for digital games (see subsection 3.1) that helps game designers in navigating the complexities of the current landscape, particularly during market analysis and conceptual design, but also enhances and fosters academic game design knowledge.

Specifically, the two main contributions of this work are:

- Firstly, we propose a cataloguing model that encapsulates the multitude of metadata describing digital games.
- Secondly, to overcome the problem of the technological advancements described by [15], we propose and validate a classification system that applies Norman's design principles [16] to the outlined cataloguing model.

As it is deeply analysed in the following sections, it is possible to correlate the metadata of the new cataloguing model with the five design principles by using a correlation matrix that allows every game to be classified. A sample of existing games on the market has been evaluated based on the proposed unified cataloguing model and classification system. Then, using statistical clustering techniques, the operation results are analysed to discover any natural groupings within the classification system.

The paper is arranged as follows: Section 2 discusses the related works. Section 3 illustrates all the methodological steps followed in this work, and Section 4 describes the application results. Section 5 presents the discussion and finally Section 6 presents concluding remarks on the research's limitations and future developments.

2 State of the art on cataloguing, classifications and clustering

In this section, we delve into the current state of research on cataloguing, classification, and clustering within digital games. Therefore, we explore a series of contributions that have attempted to devise frameworks and systems capable of encapsulating the vast diversity and complexity of video games. We begin by exploring the challenges and methodologies associated with digital game cataloguing and classification, highlighting the efforts to create comprehensive models with a critical discussion on the limitations of the current proposals.

Subsequently, we shift our focus to the application of clustering techniques as a method to refine and expand upon existing classification systems. This includes an examination of how these methodologies have been applied not only within the context of game design but also in broader media content analyses. Through this discourse, we aim to present a comprehensive overview of the current landscape and underscore the significance of these endeavours for the field of game studies.

2.1 Digital games cataloguing and classifications

Cataloguing and classifying video games are challenging issues introduced previously in several research works. Jantke et al. [17] proposed a three-dimensional taxonomy encompassing (i) the type of computer software, (ii) the game genre, and (iii) interaction aspects. Despite proposing an overarching framework to categorise any digital game, the three dimensions alone are not sufficient as they do not include features typical of serious games. In [3], the authors introduced a comprehensive multidimensional cataloguing of serious games, focusing on the essential design attributes such as the player's interaction style, mode, activity, environment, and application area. Although this paper provides a useful and thorough cataloguing, it is restricted to serious games only. In contrast, Natucci and Borges [18] discussed a theoretical framework to unify the Experience, Dynamics and Artifacts components of digital games, adopting a holistic view of games to include both edutainment and entertainment elements. However, they did not suggest any cataloguing or classification system.

The process of classifying video games needs to address inherent problems. First, finding unifying factors is difficult due to the significant diversity of video games. Prior classification attempts often led to confusion by including numerous, seemingly arbitrary, incompatible, or overlapping categories [19]. Moreover, as pointed out by [15], existing classifications—while being valuable references and thus essential to create novel knowledge—are inevitably bound to become outdated because of the rapid technological evolution that constantly introduces new perspectives and dimensions to the gaming experience. Thus, an "obsolescence paradox" emerges in the practice of classifying digital games, as classifications are essential to the creation of knowledge yet prone to rapid obsolescence [15]. It is therefore necessary to propose a unified classification system that is not affected by the problem of technological advancement, allowing on one hand to create a theoretical basis for future studies and on the other hand to consolidate the various proposals outlined in literature.

Building a unified digital game classification system is particularly relevant for five key reasons. First, game studies are gaining recognition as a legitimate academic discipline. Secondly, a unified system can support and enhance the academic legitimacy of this field. Thirdly, there is now enough foundational research to support a substantial effort towards consolidation. Fourthly, it can support the dissemination of information on game design theory. Finally, it could be the theoretical ground for designers and practitioners to systematically address game design [20].

2.2 Clustering

To validate and expand media content classifications, some researchers applied clustering techniques to their cataloguing models or classification systems. Fish et al. [21] proposed expanding traditional movie genre classification through clustering, identifying more detailed semantic information within the multi-modal content of movies. However, this paper is limited to the movie genre only. Zhou et al. [22] also presented a method for categorising movie genres but based on scene analysis. They applied k-means clustering to quantify features extracted from keyframes or movie trailers. Despite the authors' attempt to gain a deeper understanding of feature films, the classification is based solely on genre. Shifting to the specific context of game design, several studies employed cluster analysis to propose novel classification systems. Ramirez-Cano et al. [23] proposed a meta-classification approach consisting of three levels of analysis: 1) action/skill-based clustering, 2) preference-based clustering, and 3) socially based clustering. In contrast, Rodrigues and Brancher [24] explored how gameplay data can provide valuable insights into player profiles, beyond just demographic characteristics. Specifically, it proposed the use of clustering algorithms to find distinct players profiles, i.e. advanced, skilled, beginners, and intermediated.

However, these studies suggested classifications centred on game users rather than games. To the best of our knowledge, only Dahlskog et al. [25] employed clustering to classify games, using it as a descriptive tool to categorise and understand the different genres of digital games. Specifically, through their analysis, the authors pointed out that functional categories (e.g. player's perspective, challenges, mutability and savability) are essential to describe games. However, they only focused on game genre and did not consider other significant descriptive features such as the modality of interaction, gameplay or graphics.

3 Methodology

As highlighted in the previous sections, building a collective knowledge and shared understanding in the field of game design is crucial [20]. This paper intends to address this need by proposing a unified framework for digital games, not only to support scientific advancement in this field but also as a practical aid to designers in their work. The proposed framework includes a cataloguing model and a classification system as shown in Fig. 1. The cataloguing model provides a descriptive representation of digital games through metadata. By relating the metadata to fundamental design principles and going through mathematical steps, as it is thoroughly explained in the following paragraphs, a classification system that identifies the belonging of an item to a certain category is obtained. Such proposal could also contribute to the creation of a game design tool for the development of both video and

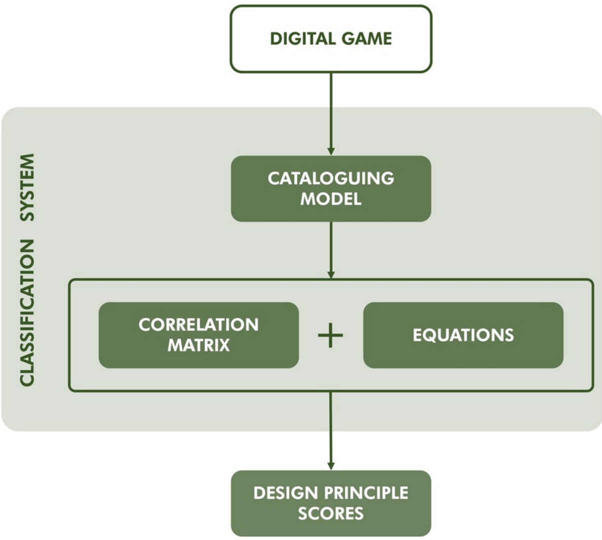

Fig. 1 The proposed framework composed by the classification system and its other components (cataloguing model, correlation matrix and equations)

serious games, enabling game designers to find their way through the vast scene of existing digital games and place their product on the market.

The proposed cataloguing model was obtained through a literature review of existing cataloguing. The metadata derived from the reviewed cataloguing were collected, divided, and categorised in the presented model. Then, we proposed a correlation matrix that relates the metadata to Norman's five design principles. Through the matrix, a classification system based on these five fundamental principles is obtained. Finally, the resulting framework (cataloguing model, correlation matrix, and classification system) was applied to a sample of existing digital games and validated by hierarchical and k-means clustering. The whole methodology is described in more detail in the following subsections.

3.1 Literature review

The literature review considered the collection and analysis of previously proposed cataloguing for video games and serious games, aiming to formulate a unified model that encompasses all the findings, to assist designers in tailoring their applications. The literature review was conducted by following the guidelines adapted from [14], namely:

- Formulating a review question;
- Developing a search string based on the review question;
- Defining which database to search and preliminary inclusion/exclusion criteria (type of papers included in the search, language, field and time frame);
- Preliminary screening based on title, abstract and keywords;
- Final selection based on inclusion/exclusion criteria related to the literature review goal.

The present research began by formulating the question: "How are video and serious games catalogued, and on which criteria?".

Subsequently, a search string was developed based on the research question: the search was conducted in electronic databases such as Scopus, IEEEXplore, and Web of Science, focusing on scientific papers, review articles, and conference proceedings. The search string was as follows: ("classification" OR "cataloguing" OR "cataloging") AND ("videogames" OR "digital" OR "serious") AND ("games"). The survey was confined to English-language articles from Computer Science and Engineering fields, resulting in 1135 records narrowed down to 911 after removing duplicates. Further filtering, including only publications from 2010 to 2023, led to 695 papers. A preliminary screening based on title, keywords, and abstracts narrowed the selection to 31 titles, further refined to 5 records based on a strict inclusion criterion stipulating that a screened record was eligible for review only if it proposed cataloguing for video and/or serious games.

The final selected papers proposed various catalogues. Nasution et al. [26] introduced cataloguing based on the G/P/S (gameplay, purpose, scope) concept, gaming platform, and user experience, providing insights into gamification studies in container terminal logistics. Although it considers both video games and serious games, it still overlooks important aspects like software-related game features and artistic choices related to video game genres. McMahon et al. [27] focused exclusively on cataloguing entertainment-oriented video games, while [28] presented a cataloguing for health rehabilitation games concentrating on UX, interaction, technology, and domain-specific criteria. However, both approaches cover only some digital games because both are limited to only one of the two categories of digital games (video games or serious games).

In addition, the cataloguing proposed by [28] has been tailored for health rehabilitation, thus it is also domain specific. The authors in [3] proposed a cataloguing exclusively of serious games. Finally, Zaki et al. [29] proposed a cataloguing for cognitively stimulating serious games. However, their cataloguing is limited to a specific domain, and it is valid only for serious games without broader categorisation.

3.2 Cataloguing model and classification system

The literature review process resulted in five distinct cataloguing [3, 26–29]. Although, individually, these findings adopted an excessively domain-specific or low-level approach, they incorporated several metadata attributes that are useful to build a unified model. For this reason, they constituted the basis for the next steps of the proposed methodology. Therefore, we analysed the five reviewed papers in order to pursue the overall goal of consolidating these cataloguing into a unified model. To achieve this purpose, we adopted a higher-level framework presented in [17]. This framework posits the classification of digital games based on three primary definitions:

- *D1*, "The game is seen as a computer program and all its related properties such as the admissible number of users, networking features and the like are considered."
- *D2,* "The game as a piece of artwork and media has a genre."
- *D3,* "Through intense interaction, players act and experience engagement such as building, fighting, or trading, e.g., interaction determines the psychological/social impact".

After excluding domain-specific parameters from the reviewed cataloguing systems, e.g., those on patient monitoring or assessment [28], we assigned the remaining metadata to one of the defined categories. The specific parameters of serious games, i.e., target audience and application area, were not covered by the three-definition framework that is more focused on video games rather than on all digital games.

Therefore, to encompass the omitted parameters, we introduced two ad-hoc additional definitions as follows:

- *D4,* digital games are inherently oriented towards a specific audience and market, necessitating consideration of the target user during the design phase.
- *D5,* serious games always serve a practical purpose, which, in conjunction with the target audience and market, guides the design phase.

From the analysis of the previously proposed cataloguing, we obtained 62 metadata that can be expressed as binary values, meaning that the game either has (1) or doesn't have (0) the specific attribute related to the metadata. For instance, if a game can be played in Augmented Reality, the tag D1.1.2—Augmented Reality is scored 1, otherwise it is scored 0. All tags are collected in the cataloguing model within the proper definition from D1 to D5. The comprehensive list of tags as well as their arrangement within the definitions is presented later, in the Results section.

Finally, a correlation matrix was defined between the tags of the proposed cataloguing model and the five fundamental design principles according to Norman. We chose to adopt these design principles as an analysis framework because if technologies, genres, and other inherent characteristics of digital games can change over the years, connecting them with Norman's design principles can help to make the proposed classification system persist over the time and avoid the influence of the temporal dimension. Such principles are:

- *Affordance*: refers to the relationship between a physical object and an entity (like a person, machine, or animal), highlighting how the object's properties align with the user's ability to determine its potential uses, like how a chair suggests sitting.
- *Signifier*: is the signal component of an affordance. It can be deliberate and intentional (like a 'push' sign on a door) or accidental and unintentional (like tracks in snow guiding a path), indicating where an action should be performed.
- *Constraints*: are powerful cues that limit the range of possible actions. Intelligent use of constraints aids individuals in choosing the appropriate course of action, even in novel situations.
- *Mapping*: is a technical term for the relationship between two elements, in this context, between controls, their operation, and their outcomes in the external world. Natural mapping, utilizing physical analogies or cultural models, facilitates immediate understanding.
- *Feedback*: is the return information that informs the user about their action and its resultant effect. It should be immediate and informative.

A three-point scale (1 – irrelevant correlation, 3 – medium correlation, 5 – high correlation) expressed the correlation between the tags of the cataloguing model (62 columns) and the design principles (5 rows). The matrix was created by three experts in the field of Human-Centred Design with gaming expertise, i.e., two from academia (two researchers with a background in interaction design and design studies) and one from industry (an experienced game designer). To obtain a shared consensus from the three experts, we followed a mini-Delphi approach [30]. We sent the experts a form with 310 closed-answer questions, one for every possible combination of tag and design principles (62 tags and 5 design principles, totalling 310 questions). The questions were formulated as "How much do you think design principle and tag are correlated?" and could be answered with "1 – irrelevant correlation", "3 – medium correlation" and "5 – high correlation". Each question was followed by a box in which experts were asked to leave a comment on their reasoning. After receiving the completed questionnaires, we collected the answers and comments and prepared a second-round questionnaire, in which we showed to the experts all the answers and comments from the previous round. In this second questionnaire we asked the experts whether, in the light of the answers of the first round, they intended to change their answers. This review phase was repeated in a third round, in which the experts reached a unanimous consensus. The procedure thus comprised three rounds and took place over a period of six months, during which the experts remained anonymous. The resulting correlation matrix of size 5×62 was the basis of the proposed classification system.

3.3 Cataloguing and classification procedure

The classification of each digital game started with cataloguing them according to the proposed model by assigning a binary value to the metadata. The metadata took the value of 1 if the feature it represented was present in the digital game, otherwise it took a value of 0. For each game we obtained a 62-dimensional vector (one vector component for each tag), named cataloguing vector, that is expressed in Eq. 1, with t_i the binary value of the i-th metadata:

$$\overline{T} = \left(t_1, t_2, t_3, \ldots, t_{62}\right) \tag{1}$$

The correlation matrix, whose values (1–3-5) were defined by the experts, is a 5×62 matrix visible in Eq. 2:

$$\overline{\overline{M}} = \begin{vmatrix} m_{1,1} & \cdots & m_{1,62} \\ \vdots & \ddots & \vdots \\ m_{5,1} & \ldots & m_{5,62} \end{vmatrix} \tag{2}$$

where the element m_{ij} represents the correlation value between the i-th metadata and the j-th design principle (e.g., m_{11} is the correlation value between metadata D1.1.1—Social presence and the first design principle, namely affordance). To classify the games according to the design principles, we multiplied each cataloguing vector and the correlation matrix using the matrix multiplication as in Eq. 3:

$$c_j = \sum\nolimits_{i=1}^{62} \left(m_{ji} \cdot t_i\right), \forall j = 1,2,3,4,5 \tag{3}$$

The result we obtained is a 5-dimension vector, named the classifier vector, one for each game. The classifier vector represents the values each game expresses for the five fundamental design principles. The classifying vector's components can be seen in Eq. 4:

$$\overline{C} = \left(c_1, c_2, c_3, c_4, c_5\right) \tag{4}$$

where c_j is the score obtained for the j-th design principle (e.g., c_1 is the affordance score).

3.4 Clustering algorithms

The ancient practice of categorising similar items into groups is a fundamental human endeavour present in various domains, including language, natural and social sciences [31, 32]. This involves the use of numerical techniques known as cluster analysis [33], a method of exploratory data analysis that seeks to reveal hidden structures within data [31, 32], unveiling previously unknown object groups that exhibit statistical similarity [34] and that could serve as the foundation for future research and hypotheses [32]. It can be divided into non-hierarchical (or partitional) and hierarchical. Two of the most popular and widely used algorithms are k-means (partitional) and agglomerative (hierarchical clustering) [31].

The k-means algorithm, recognized as the oldest and broadly used partitional method, has been extensively studied and applied across various fields due to its simplicity and efficiency. Its main purpose is to divide objects into k partitions or clusters, minimising the within-cluster sum of squares (WCSS) of the Euclidean distances between the elements assigned to a specific cluster and its centroid [31, 35].

The k-means algorithm involves selecting an a-priori k value and using it as the initial set of k centroids. The objects are then assigned to the cluster with the closest centroid. The new centroids of the k clusters are determined by averaging the cluster members. The last two steps are repeated until the criterion function stops changing [31]. The k-means algorithm necessitates a preliminary definition of the desired number k of clusters, a quantity that is typically unknown [32]. For this purpose, the empirical approach known as elbow analysis is one of the most widely adopted. It involves calculating the WCSS of the Euclidean distances for different k values, observing that as the k value increases, the average WCSS becomes smaller. The optimal k value is the one for which the diminishing returns effects occurs, i.e., further increasing the k value leads to a smaller benefit than the previous increment [33].

Like iterative partitional clustering, agglomerative hierarchical clustering stands out as one of the most well-known and frequently used clustering technique. The agglomerative hierarchical clustering algorithm typically follows these steps: (1) treat each observation as an initial cluster; (2) compute distances between clusters; (3) combine two clusters with the minimum distance (typically the Euclidean distance), replacing them with a single cluster; (4) repeat the last two steps until only a single cluster containing all observations remains.

The result is a nested sequence of partitions spanning from a single cluster containing all the individuals to n clusters each containing a single item [33]. The sequence is depicted as a dendrogram, a two-dimensional tree-like structure that illustrates the sequence of nested clusters. Selecting a partition (i.e., the number of cluster) involves cutting the dendrogram at a specified height. The vertical gap from a branch to the other represents the distance between the clusters, with the largest gap between fusion levels indicating the "best cut". Therefore, the dendrogram's appearance allows to identify the number of clusters [31, 33]. Currently, due to its recognized effectiveness, Ward linkage is the most frequently used hierarchical method [31]: it evaluates cluster adequacy by measuring distances between cluster centroids, and merges clusters based on minimal sums of squared dissimilarities to the centroid (a cohesion measure) [33].

It is also possible to combine the two aforementioned clustering methods, an approach known as two-stage clustering: the first stage makes use of hierarchical clustering, while the second stage makes use of the k-means algorithm. As a matter of fact, the use of the dendrogram obtained in the first stage (hierarchical clustering) to determine the number of clusters can be the starting point for the second stage (k-means algorithm), so that the advantages of both methods can be put to good use [31].

3.5 Cluster analysis

The first step for clustering was extracting data concerning 101 digital games. To this purpose, we aimed at establishing an objective metric to categorize data. Given the target of the study (i.e., game designers and industry professionals), we identified market share as a particularly useful measure for categorizing data. Accordingly, the sample included 96.04% video games (97 items) and 3.96% serious games (4 items), with the percentages chosen to reflect the market share of each category. Indeed, in 2022, serious games generated a market of USD 8.23 billion, or 4.17% of the total digital games market [36].

Additionally, it was not possible to identify a metric based on the total number of video games and serious games across any database consulted. This challenge is probably linked to the problem we propose to address: the absence of a unified framework for collectively cataloguing video games and serious games. Our investigation revealed that databases vary significantly in their coverage. For instance, Serious Games Classification [37] lists 3,426 titles, while another, MobyGames [38], includes over 270,998 digital games without distinguishing between video games and serious games. Based on these figures, serious games would represent an even smaller fraction, amounting to 1.26%, assuming MobyGames includes all 3,426 serious games (an assumption we cannot verify). Therefore, despite potential biases (i.e., the 96%/4% split may distort the data), we assert that market share remains the most appropriate metric for distinguishing between video games and serious games, reflecting the current landscape.

To select the digital games to be included in the sample, we generated 101 random numbers. We then selected the games from MobyGames [38] whose IDs corresponded to the generated numbers. To respect the 97/3 split as proposed, we made sure to select at least 4 serious games. To determine each digital games' metadata and thus be able to catalog and then classify the games, we needed to obtain information on them. Data on the selected games were extracted from the MobyGames and Game UI Database [39] databases, when available. We chose MobyGames and Game UI Database as our data sources because they are two of the most comprehensive databases on digital games. Additionally, MobyGames provides an open API to easily collect data on its content and covers some highly technical aspects of a digital game, useful for our research. Authors manually obtained the data not found in these databases through press reviews, gameplay, or their direct experiences.

Once obtained the 101 classifier vectors resulting from the above procedure (one for each of the sampled digital game), cluster analysis began, specifically applying a two-stage clustering approach.

Firstly, we determined the number of clusters k using agglomerative hierarchical clustering with Ward's algorithm. Then, we performed k-means clustering analysis with the aforementioned k value. To ensure the consistency between the clusters obtained with the two techniques, and thus to verify the validity of the approach, we calculated Cohen's kappa. Both clustering techniques were implemented with Python scripts: SciPy (v1.11.3) [40] library was adopted for the agglomerative hierarchical clustering, more precisely the linkage function from the scipy.cluster.hierarchy module. For k-means clustering, the scikit-learn (v1.3.2) [41] library was used, in particular the KMeans class from the sklearn. cluster module. Two graphs allowed to visualise the results. The first was a dendrogram, created using the dendrogram function from the scipy.cluster.hierarchy module for agglomerative hierarchical clustering. By examining this graph, the number of clusters was inferred by analysing the colour change in the dendrogram branches. The mergers between clusters, operated by the algorithm, can be retraced by going up the graph, with the jump between one level and another representing the distance between the fused and the resulting clusters. The colour change occurs when, going from a lower to a higher level, the leap exceeds a predetermined threshold.

The second graph was a scatterplot for the k-means clustering, obtained after a Principal Component Analysis (PCA). The PCA was solely used for dimensionality reduction to facilitate data visualisation in a 2D graph. It is important to note that both clustering phases were performed on raw, non-reduced data.

Finally, to deeply analyse the final k-means cluster's content, we computed a Shapiro–Wilk test to assess the distribution of the data regarding the tags and the publication year. Then, to assess the statistical significance of the differences between clusters we conducted a Mann–Whitney U test. All statistical analyses were performed using IBM SPSS Statistics v20.0.0.

4 Results

In this section we provide the results of our methodology. We start with presenting the results of the cataloguing model creation, highlighting how we joined together the findings from the literature review and presenting each tag within its definition. Then, we explain how the classifying system is built upon the cataloguing model and we provide an example to better illustrate how the framework is able to score a game. Finally, we present results of the clustering validation phase.

4.1 The unified cataloguing model

This subsection presents a comprehensive cataloguing model consisting of five distinct definitions, each offering crucial insights into the classification and cataloguing of software-related games. The subsection's focus is to depict each definition and describe all the respective categories and metadata contained within. The first four definitions (D1-D4) can be applied to all types of video games as they pertain to software, artistic choices, user interaction, and end-user experience features. Instead, the fifth definition (D5) is specific to serious games, focusing on the application area for which the serious game is intended. Each definition of the unified cataloguing model encompasses several categories, further subdivided into detailed metadata about the core issue.

The unified cataloguing model depicted in Fig. 2 solely shows the categories in order to create a concise an information-packed scheme.

4.1.1 Definition 1

Definition 1 (D1) sheds light on various dimensions of software-related games, encompassing the following categories: D1.1—Environment, D1.2—Game platform, D1.3—Interaction technology.

The "Environment" category enumerates various gaming system-supported environments, including D1.1.1—Social presence (whether the game has some social features or not), D1.1.2—Augmented reality (if the game has Augmented Reality features), D1.1.3—Virtual environment (if the game has Virtual Reality features), D1.1.4 – Location awareness (if the game allows players to be aware of their location in the game world, e.g., there is a map), and D1.1.5—Online (whether the game can be played online or not) [3, 29].

The second category, "Game Platform", includes D1.2.1 – PC (if the game can be played on a desktop or laptop pc), D1.2.2—Mobile (whether the game can be played on

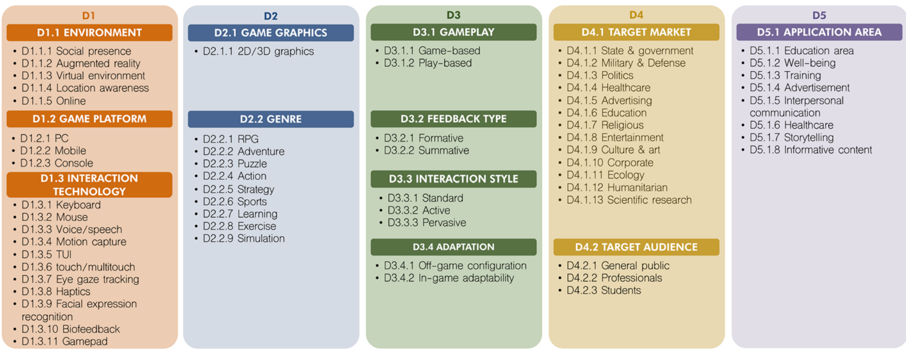

Fig. 2 The unified cataloguing model with the five definitions

mobile devices, e.g., smartphones, tablets) and D1.2.3 – Console (whether the game is played on gaming consoles, e.g., Microsoft Xbox, Sony Playstation, Nintendo Switch), covered by authors in [26, 28, 29].

Lastly, "Interaction Technology" encompasses numerous metadata, such as D1.3.1 – Keyboard (the game is played with a keyboard), D1.3.2 – Mouse (the game can be played with a mouse), D1.3.3—Voice/speech (the game has voice/speech recognition capabilities), D1.3.4—Motion capture (the game can be played with the movements of the body), D1.3.5 – TUI (the game can be played with Tangible User Interfaces, e.g., an everyday object that can act as a controller), D1.3.6—Touch/multitouch (whether the game accepts touch inputs), D1.3.7—Eye gaze tracking (if the game can be played with eye's movements), D1.3.8 – Haptics (if the game has some sort of haptic feedbacks), D1.3.9—Facial expression recognition (whether the game has facial expression recognition capabilities or not), D1.3.10 – Biofeedback (if the game can be played with biofeedback devices, e.g., EEG helmets), and D1.3.11 – Gamepad (if the game can be played with a traditional gamepad), discussed in works [3, 28]. Interaction technology is the category most likely to be affected by obsolescence and therefore the one whose tags should be updated as technology evolves.

4.1.2 Definition 2

Definition 2 (D2) delves into the design of game interfaces, with a particular focus on D2.1—Game graphics and D2.2—Genre.

The first category introduces only one metadata, D2.1.1—2D/3D graphics (whether the game has 2D or 3D graphics) [3, 29], while "Genre" contains a range of gaming genres, i.e., D2.2.1—RPG, D2.2.2—Adventure, D2.2.3—Puzzle, D2.2.4—Action, D2.2.5—Strategy, D2.2.6—Sports, D2.2.7—Learning, D2.2.8—Exercise, and D2.2.9—Simulation, with contributions from authors [27, 29]. All of the tags from D2.2 indicate what is the game's genre.

4.1.3 Definition 3

Definition 3 (D3) is the most comprehensive, featuring categories such as D3.1—Gameplay, D3.2—Feedback type, D3.3—Interaction style, and D3.4—Adaptation.

"Gameplay" distinguishes between D3.1.1—Game-based (the game has a final goal and/or a winning condition) and D3.1.2—Play-based approaches (the game is played for the pure fun of playing, without final goals or winning conditions), discussed by [26].

The category "Feedback type" embraces feedback during gameplay, D3.2.1 – Formative (the game provides players with feedback during gameplay), and after gameplay, D3.2.2 – Summative (the game provides players with feedbacks at the end of the gaming session, e.g., after finishing a level), both analysed by [28].

"Interaction style" describes player interaction methods: D3.3.1—Standard (traditional peripherals, such as mouse/keyboard or gamepad), D3.3.2—Active (alternative peripherals or body, e.g., a guitar-shaped gamepad or motion capture), and D3.3.3—Pervasive (a more involved interaction style, e.g., voice/speech recognition or facial expression recognition), with insights provided by author [28].

This category is a different way of looking at gaming peripherals as compared to Interaction technologies: whereas the latter describe peripherals from a technological point of view, and are therefore subject to ageing and obsolescence, Interaction style describes the way the user interacts with the peripheral, expressing a concept that does not change over time.

Finally, “Adaptation” encompasses D3.4.1—Off-game configuration (whether the game can only be configured before the gameplay actually begins or not), and D3.4.2—In-game adaptability (if the game can automatically adapt to the players and their skills’ levels), outlining how games adapt to player preferences and abilities [26, 28, 29].

4.1.4 Definition 4

Definition 4 (D4) underscores the importance of targeting when designing digital games. It focuses on tailoring games to meet the needs and preferences of the target user during the design phase, encompassing metadata contained in two categories, D4.1—Target market (the market for which the game is intended) and D4.2—Target audience (the audience for which the game is intended).

“Target market” includes D4.1.1—State & government, D4.1.2—Military & Defense, D4.1.3—Politics, D4.1.4—Healthcare, D4.1.5—Advertising, D4.1.6—Education, D4.1.7—Religious, D4.1.8—Entertainment, D4.1.9—Culture & art, D4.1.10—Corporate, D4.1.11—Ecology, D4.1.12—Humanitarian, D4.1.13—Scientific research [26].

“Target audience” instead embraces D4.2.1—General public, D4.2.2—Professionals, D4.2.3 – Students [29].

4.1.5 Definition 5

Lastly, Definition 5 (D5) refers to serious games’ application areas, crucial for game design, as each serious game is shaped by practical purposes aligning with the audience and market. The sole category of this definition, D5.1—Application area (the scope of the game, whether is intended to, e.g., advertise, train, educate), incorporates the following metadata: D5.1.1—Education area, D5.1.2—Well-being, D5.1.3—Training, D5.1.4—Advertisement, D5.1.5—Interpersonal communication, D5.1.6—Healthcare, D5.1.7—Storytelling, and D5.1.8—Informative content [3, 26].

Finally, in regard to the classification system, the correlation matrix consists of five rows (one for each design principle) and 62 columns (one for each tag in the cataloguing model). The tags are labelled D.X.Y, where X represents the number of the definition to which that tag belongs to, and Y is an incremental number representing the specific tag. The matrix is publicly available on Github, and the link to the repository can be found in the Data availability section, at the end of the paper. Classifier vectors classify digital games according to their affordance, signifier, constraint, mapping and feedback scores. For example, Super Mario Bros. [42] released in 1983, it’s an action (D2.2.4) console game (D1.2.3) played with a gamepad (D1.3.11), thus with a standard interaction style (D3.3.1). It is an entertainment (D4.1.8) game-based (D3.1.1) video game, oriented towards the general public (D4.2.1). It provides players with formative (D3.2.1) and summative (3.2.2) feedback. With these characteristics, according to the correlation matrix, it achieves an Affordance and Signifier score of 22, a Constraints score of 16, a Mapping score of 12 and, finally, a Feedback score of 24.

Overall, these definitions and associated metadata as well as the correlation matrix provide a structured and comprehensive framework of digital games. This framework, on one hand, serves as a valuable resource to deepen designers’, researchers’, and enthusiasts’ understanding of the digital games’ field. On the other hand, it enables designers to tailor their products more effectively to their intended audience, easily identifying the dimension characterizing their products and hence where to focus their design efforts. For instance, designers are encouraged to work backwards from this conceptual premise when formulating a game, e.g., a fantasy-themed serious game, to educate teenagers about the dangers of smoking. Having established the users and the target market, designers could analyse the placement of digital games that are comparable with the one they are developing. The tags activated by the comparable digital games (i.e., tags with a value of 1) resulting from the technical solutions adopted by competitors can direct game designers towards appropriate design choices. At this point, designers would have guidelines leading their work and highlighting areas of possible improvement. The active metadata can be considered as suggested specifications which will guide the game's development (e.g., a high-engaging game developed in VR and with social networking features to capture the attention of teenagers).

4.2 Clustering results

The cataloguing model and the classifier vectors obtained by applying the proposed framework are publicly available on Github, at the same link of the correlation matrix. The clusters with the included games are available on the same repository. Based on the classifier vectors, the agglomerative hierarchical clustering identified two clusters, as shown in Fig. 3. The threshold value used to determine the number of clusters is the default setting for the *scipy.cluster.hierarchy.dendrogram* function, which is 70% of the maximum distance among all distances between clusters. The subsequent phase of k-means clustering (k = 2) produces the two clusters visible in Fig. 4. The coordinates of the clusters' centroids are:

- Cluster0 = (26.41, 27.14, 21.59, 17.36, 29.23), contains 43 digital games.
- Cluster1 = (35.89, 37.89, 30.63, 24.25, 41.26), contains 58 digital games.

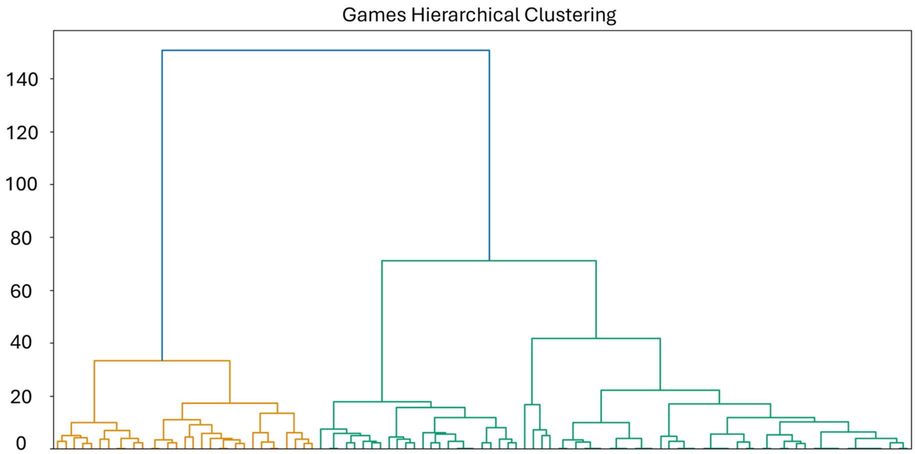

Fig. 3 The dendrogram resulting from the hierarchical clustering. At the bottom there are the 101 digital games. Moving up the graph from bottom to top, the mergers performed by the algorithm can be observed. The change of colour highlights the clusters

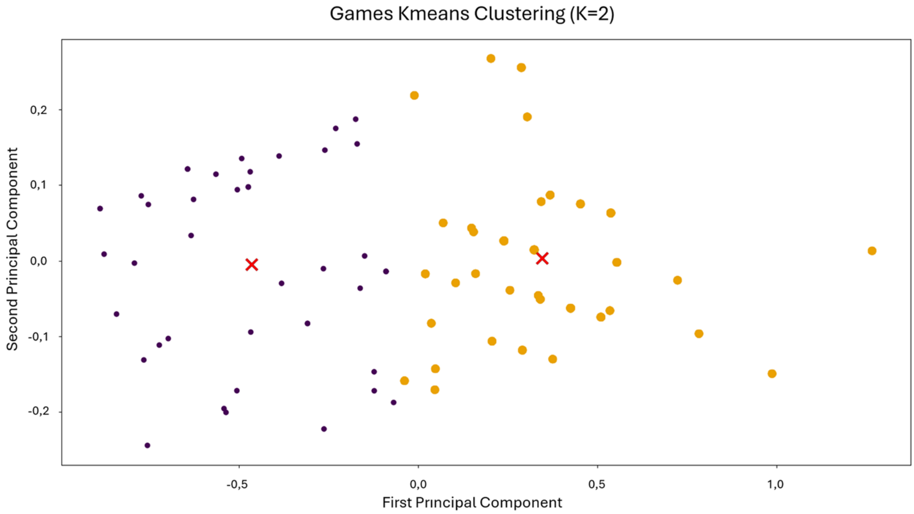

Fig. 4 The scatter plot resulting from the k-means clustering. Each dot represents a digital game in a 2D space obtained from PCA. The red crosses denote the centroids of the clusters. Clusters are highlighted in different colours

The correlation analysis between the clusters identified by the two methods reports a Cohen's kappa = 0.75, indicating a substantial agreement between the two techniques. Finally, the absolute percentage differences between each design principle between Cluster0 and Cluster1 are as follows: Affordance 26,43%, Signifier 28,39%, Constraint 29,51%, Mapping 28,38%, Feedback 29,17%. The average of these absolute percentage difference is 28.38%.

From the Shapiro–Wilk test it emerged that the distribution of the variable "publication year" and the distribution of all the tags within the clusters are not normal, with $p < 0.05$.

The results of the Mann–Whitney U test showed that the variable "Publication year" has statistically different distributions between clusters, with $p = 0.004 < 0.05$. The Cluster0's average publication year is 2013, while for Cluster1 it is 2017. The same test applied to the tags showed that D1.1.1, D1.1.4, D1.1.5, D.1.2.1, D1.2.3, D1.3.1, D1.3.2, D1.3.8, D1.3.11, D3.1.1, D3.1.2, D3.3.1 and D3.4.1 have statistically different distributions, with $p < 0.05$. Table 1 reports the results of the tags' Mann–Whitney U test. For each tag we reported the number of games that have the attribute related to that tag (i.e., the value of the tag is 1) for both clusters, with Cluster0 in the third column ("Cluster0 numerosity") and Cluster1 in the fourth column ("Cluster1 numerosity"). The fifth ("Cluster0%") and sixth ("Cluster1%) columns shows the percentage corresponding to the values reported in "Cluster0 numerosity" and "Cluster1 numerosity".

5 Discussion

The five cataloguing resulted from the literature review [3, 26–29], although providing useful insights to the scope of building a unified cataloguing model, are all too specific to a certain application field and they did not consider all the aspects of a digital game.

Table 1 Results of the Mann–Whitney U test for the tags

Code	Tag	Cluster0 numerosity	Cluster1 numerosity	Cluster0%	Cluster1%	*p*
D1.1.1	Social presence	14	40	32,56	68,97	0,000
D1.1.4	Location awareness	8	28	18,60	48,28	0,002
D1.1.5	Online	14	36	32,56	62,07	0,004
D1.2.1	PC	16	58	37,21	100,00	0,000
D1.2.3	Console	27	55	62,79	94,83	0,000
D1.3.1	Keyboard	14	58	32,56	100,00	0,000
D1.3.2	Mouse	15	58	34,88	100,00	0,000
D1.3.8	Haptics	15	55	34,88	94,83	0,000
D1.3.11	Gamepad	23	57	53,49	98,28	0,000
D3.1.1	Game-based	35	56	81,40	96,55	0,012
D3.1.2	Play-based	8	2	18,60	3,45	0,012
D3.3.1	Standard	40	58	93,02	100,00	0,042
D.3.4.1	Off-game configuration	21	52	48,84	89,66	0,000

Therefore, to address the lack of a unified cataloguing model describing any digital game (the gap that this paper aims to fill), we built a novel cataloguing model encompassing both video and serious games. This cataloguing comprises several digital games attributes, ranging from technical (e.g., D1.2 – Game platform) to artistic aspects (e.g., D2.1 – Game graphics), from player interaction (e.g., D3.3 – Interaction style) to market aspects (e.g., D5.1 – Application area). Additionally, with the aid of a correlation matrix and the five Norman's principles, we also proposed a classification system identifying the category any digital game belongs. The resulting comprehensive framework, representing our proposed structuring of the empirical knowledge on game design in a unified model, was then applied to a sample of existing digital games and validated through a cluster analysis, which led to the identification of two clusters (Cluster0 and Cluster1).

By examining the clusters' centroid statistics, a comparison between clusters reveals that the characteristics of games in Cluster0 are, on average, 28.38% lower than those in Cluster1. This disparity may stem from several factors.

Firstly, games belonging to Cluster0 are generally older than those in Cluster1, as their average publication year is 2013 against 2017 of the latter. Results show that, on average, more games belonging to Cluster1 present Social Presence and Online features and are generally configurable (more games with Off-game configuration), compared to Cluster0. Moreover, Cluster1's games enable the player to be aware of the game world, presenting Location Awareness features. These results suggest that games in Cluster1 are generally more complex and probably more challenging, allowing the player to interact with the off-game world and the player's community, while also being configurable to the player's preferences.

On the other hand, all the games in Cluster1 can run on a PC and almost all can run on a gaming console, being played with traditional interfaces such as mouse and keyboard or gamepads and presenting haptic feedback. The fact that most of the games contained in Cluster1 has a standard interaction style seems to suggest that they are more traditional compared to games in Cluster0.

Finally, games in Cluster1 are more oriented towards a game-based gameplay, while Cluster0 contains more play-based titles, suggesting that games in Cluster0 may be more casual games than those in Cluster1. By observing representative titles from Cluster1 we found games such as “Apex Legends” [43], “Grand Theft Auto V” [44] and “Elden Ring” [45]. These games are known for their demanding and challenging gameplay or the vastness of their game worlds. Elden Ring has a notoriously high difficulty, being listed among “hardcore games”, while Apex Legends is a game known for global esports competitions. Finally, Grand Theft Auto V, while not being a competitive or inherently challenging game, was the gold standard for open-world games presenting limitless game options to the players. In contrast, games such as “The Sims 2” [46], “flOw” [47] and “Planet Zoo” [48], which are representative of Cluster0’s play-based nature, exhibits simpler and less demanding gameplay due to the implementation of fewer mechanics.

Moreover, in Cluster0 we find three out of four serious games (i.e., “Learn to Brace” [49], “Le Village aux Oiseaux” [4] and “Smart Wheel & Train” [50]) that, compared to the other serious game in our sample (i.e., “Moonbase Alpha” [51]), show similar features to games representative of Cluster0. More importantly, rather than grouping all the serious games into one cluster, the framework created clusters based on games’ features, without distinguishing between serious games and video games. Thus, the framework seems to be able to unify all the digital games.

6 Conclusions

This research work was aimed to assess whether it would be possible to propose a unified cataloguing model and a classification system that could be applied to both video games and serious games, gathering existing knowledge and underscoring commonalities across various cataloguing systems.

To reach such a goal, this paper proposes a unified cataloguing model for digital games, encompassing both video and serious games. The cataloguing model is based on five definitions containing all the essential metadata, commonly called "tags," found in the existing literature on digital game cataloguing. Moreover, the paper proposes a classification system based on a correlation matrix and the five fundamental design principles, thus overcoming the obsolescence paradox. As technological advances occur, the cataloguing model can be kept up to date by simply adding new tags, framing them within the five definitions, and accordingly adjusting the correlation matrix. The fundamental design principles, however, remain valid even as technology advances and, with it, the classification system. Finally, to verify whether the cataloguing model and the classification system is able to unify video and serious games, we performed a two-stage clustering analysis. The analysis resulted in two clusters, each containing both video and serious games, showing that it is possible to have a framework that does not operate a clear-cut distinction.

This framework functions as an asset for enhancing the comprehension of the digital gaming domain among designers, researchers, and enthusiasts. Concurrently, it empowers creators to refine their products for their target audience, by facilitating the identification of the dimensions that define their products, thereby directing their design endeavours more strategically.

6.1 Limitations

The 101 digital games selected represent only a sample of the vast landscape of existing titles. Thus, the clustering analysis suffers from the selected sample's limited extension. Moreover, the selection may not be generalizable to the entire digital games market, possibly leading to findings that do not completely reflect the broader industry trends or the diversity within the market. Primarily relying on MobyGames and Game UI Database for metadata extraction could have introduced bias if they have their own collection biases (e.g., an overrepresentation of certain types of games). Moreover, some aspects considered by our cataloguing model were not covered by these databases. The process of manually obtaining data not found in the databases through press reviews, gameplay, or direct experiences may have introduced subjectivity and potential biases. An extension of the sample along with the inclusion of more and diverse databases could help researchers to overcome these limitations, enriching the findings.

6.2 Future works

An important path for future research involves extensive testing of our framework with game developers, aiming to evaluate the practical applications and usability of our framework within the industry. Engaging directly with game developers will provide critical insights into how our theoretical model aligns with real-world challenges and requirements in game development. Such collaboration will be essential for validating the effectiveness of our framework and identifying potential areas for refinement. To undertake this endeavour, it is advisable to conduct a series of structured interviews and usability tests with game developers from various segments of the industry. This will not only allow researchers to gather qualitative and quantitative data on the framework's performance in practical scenarios but will also make possible to incorporate direct feedback from professionals who are actively involved in the creation and design of digital games. The outcomes of this future research should aim at enhancing the framework's relevance and applicability, ensuring that it serves as a valuable tool for both academic researchers and game development practitioners alike. Another possible path for future works could be the integration of the framework with existing game development tools and engines (e.g., Unity, Unreal Engine), to streamline the game development with insights and analytics.

Author contributions Conceptualization: Laura Cormio, Thomas Agostinelli, Maura Mengoni; Data Curation: Thomas Agostinelli; Formal Analysis: Thomas Agostinelli; Investigation: Laura Cormio, Thomas Agostinelli; Methodology: Laura Cormio, Thomas Agostinelli, Maura Mengoni; Supervision: Maura Mengoni; Visualization: Laura Cormio; Writing – original draft: Laura Cormio, Thomas Agostinelli; Writing – review & editing: Laura Cormio, Thomas Agostinelli, Maura Mengoni.

Funding Open access funding provided by Università Politecnica delle Marche within the CRUI-CARE Agreement.

Data availability Data generated during the work is publicly available on a GitHub repository, at the link https://github.com/boso94/gamesCataloguingClassification. Access to the data is open.

Declarations

Conflict of interest Authors declare that they have no conflict of interest.

References

1. Vorderer PA, Ritterfeld U (2009) Digital games. In: Nabi R, Oliver MB (eds) The Sage handbook of media processes and effects. Sage, pp 455–467
2. Newman J (2004) Videogames. Routledge, London. https://doi.org/10.4324/9780203642900
3. Laamarti F, Eid M, El Saddik A (2014) An overview of serious games. Int J Comput Games Technol 2014:1–15. https://doi.org/10.1155/2014/358152
4. Mader S, Natkin S, Levieux G (2012) 'How to analyse therapeutic games: the player / game / therapy model'. In Entertainment Computing - ICEC 2012, vol. 7522. Herrlich M, Malaka R, Masuch M Eds. In Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 193–206. https://doi.org/10.1007/978-3-642-33542-6_17
5. Holloway A, Kurniawan S (2010) Human-centered design method for serious games: a bridge across disciplines. In: Proceedings from UCSC-SOE, pp 10–36
6. Holloway A, Kurniawan S (2010) System design evolution of the prepared partner: how a labor and childbirth game came to term. Meaningful play, Oct 2010
7. Jiang Y, Zheng L (2023) Deep learning for video game genre classification. Multimed Tools Appl 82(14):21085–21099
8. Ivory JD (2015) A brief history of video games. In: The video game debate. Routledge, pp 1–21
9. Writer JRS 'Global video games market predicted to hit $197bn in 2022', GamesIndustry.biz. Available: https://www.gamesindustry.biz/global-video-games-market-predicted-to-hit-usd197bn-in-2022. Accessed 12 Feb 2024
10. Parkkila J et al (2017) An ontology for videogame interoperability. Multimed Tools Appl 76:4981–5000
11. Esse UC (2013) Current trends in cataloguing and the challenges of a cataloguer in the digital age. Inf Impact J Inf Knowl Manag 4(2):16–23
12. Jacob EK (2004) Classification and categorization: a difference that makes a difference. Libr Trends 52(3):515–540
13. Joudrey DN, Taylor AG, Miller DP (2015) Introduction to cataloging and classification. Bloomsbury Publishing USA
14. Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering version 2.3. Engineering 45(4ve):1051
15. Alvarez J, Djaouti D, Ghassempouri R, Jessel JP, Methel G (2006) Morphological study of the video games. In: ACM international conference proceeding series, vol 207, pp 36–43
16. Norman D (2013) The design of everyday things: revised and expanded. Basic books
17. Jantke KP, Gaudl S (2010) Taxonomic contributions to digital games science. In: 2010 2nd international IEEE consumer electronics society's games innovations conference. IEEE, pp 1–8
18. Natucci GC, Borges MA (2021) The experience, dynamics and artifacts framework: towards a holistic model for designing serious and entertainment games. In: 2021 IEEE conference on games (CoG). IEEE, pp 1–8
19. Aarseth E, Smedstad SM, Sunnanå L (2003) A multidimensional typology of games. In: Proceedings of DiGRA 2003 conference: level up
20. Ralph P, Monu K (2015) Toward a unified theory of digital games. Comput Games J 4:81–100
21. Fish E, Weinbren J, Gilbert A (2020) Rethinking movie genre classification with fine-grained semantic clustering. arXiv preprint arXiv:2012.02639

22. Zhou H, Hermans T, Karandikar AV, Rehg JM (2010) Movie genre classification via scene categorization. In: Proceedings of the 18th ACM international conference on multimedia. ACM, Firenze Italy, pp 747–750. https://doi.org/10.1145/1873951.1874068
23. Ramirez-Cano D, Colton S, Baumgarten R (2010) Player classification using a meta-clustering approach. In: Proceedings of the 3rd annual international conference computer games, multimedia & allied technology. pp 297–304
24. Rodrigues LA, Brancher JD (2018) Improving players' profiles clustering from game data through feature extraction. In: 2018 17th Brazilian symposium on computer games and digital entertainment (SBGames). IEEE, pp 177–17709
25. Dahlskog S, Kamstrup A, Aarseth E (2009) Mapping the game landscape: locating genres using functional classification. In: Digital games research association (DiGRA). DiGRA, West London
26. Nasution NKG, Jin X, Singgih IK (2022) Classifying games in container terminal logistics field: a systematic review. Entertain Comput 40:100465
27. McMahon N, Wyeth P, Johnson D (2013) 'Exploring the role of activity in genre'. In proceedings of the 9th Australasian conference on interactive entertainment: matters of life and death. ACM, Melbourne Australia, pp 1–3. https://doi.org/10.1145/2513002.2513023
28. Rego PA, Moreira PM, Reis LP (2018) Proposal of an extended taxonomy of serious games for health rehabilitation. Games Health J 7(5):302–309. https://doi.org/10.1089/g4h.2017.0138
29. Zaki NAA, Wook TSMT, Ahmad K (2015) Analysis and classification of serious games for cognitive stimulation. In: 2015 international conference on electrical engineering and informatics (ICEEI). IEEE, pp 612–617
30. Pan SQ, Vega M, Vella AJ, Archer BH, Parlett GR (1996) A mini-Delphi approach: an improvement on single round techniques. Prog Tour Hosp Res 2(1):27–39. https://doi.org/10.1002/(SICI)1099-1603(199603)2:1%3c27::AID-PTH29%3e3.0.CO;2-P
31. Govender P, Sivakumar V (2020) Application of k-means and hierarchical clustering techniques for analysis of air pollution: a review (1980–2019). Atmos Pollut Res 11(1):40–56
32. Sinaga KP, Yang M-S (2020) Unsupervised K-means clustering algorithm. IEEE Access 8:80716–80727
33. Landau S, Ster IC (2010) Cluster analysis: overview. Á Á 11(x12):x1p
34. Khullar V (2024) K Means clustering and descriptive analytics based performance recommending system for Kabaddi team and player. Multimed Tools Appl 83(10):29897–29914
35. Halkidi M, Batistakis Y, Vazirgiannis M (2001) On clustering validation techniques. J Intell Inf Syst 17:107–145
36. R. and Markets 'Serious games global market report 2022'. Available: https://www.prnewswire.com/news-releases/serious-games-global-market-report-2022-301717373.html. Accessed 12 Feb 2024
37. 'Serious games classification'. Available: http://serious.gameclassification.com/. Accessed 12 Feb 2024
38. 'Video game database', MobyGames. Available: https://www.mobygames.com/. Accessed 12 Feb 2024
39. 'Game UI database', Game UI database. Available: https://www.gameuidatabase.com/. Accessed 12 Feb 2024
40. 'SciPy'. Available: https://scipy.org/. Accessed 12 Feb 2024
41. 'scikit-learn: machine learning in Python — scikit-learn 1.4.0 documentation'. Available: https://scikit-learn.org/stable/. Accessed 12 Feb 2024
42. Nintendo R&D4 (1983) 'Super Mario Bros'
43. Respawn Entertainment (2019) 'Apex Legends'. Electron Arts
44. Rockstar North (2013) 'Grand Theft Auto V'. Rockstar Games
45. FromSoftware (2022) 'Elden Ring'. Bandai Namco Entertainment
46. Maxis (2004) 'The Sims 2'. EA Games, Aspyr
47. Thatgamecompany (2006) 'flOw'. Sony Comput Entertain
48. Frontier Developments (2019) 'Planet Zoo'. Front Dev
49. Chittaro L (2015) Designing serious games for safety education: "Learn to Brace" versus traditional pictorials for aircraft passengers. IEEE Trans Vis Comput Graph 22(5):1527–1539
50. RMB Games 'Smart wheel & train'. RMB Investment Company
51. Virtual Heroes and Army Game Studio (2010) 'Moonbase Alpha'. NASA Learn Technol