

Article

# A Machine-Learning-Based Motor and Cognitive Assessment Tool Using In-Game Data from the GAME2AWE Platform

Michail Danousis \* and Christos Goumopoulos \* 

Information and Communication Systems Engineering Department, University of the Aegean, 83200 Samos, Greece

\* Correspondence: mdanousis@aegean.gr (M.D.); goumop@aegean.gr (C.G.)

**Abstract:** With age, a decline in motor and cognitive functionality is inevitable, and it greatly affects the quality of life of the elderly and their ability to live independently. Early detection of these types of decline can enable timely interventions and support for maintaining functional independence and improving overall well-being. This paper explores the potential of the GAME2AWE platform in assessing the motor and cognitive condition of seniors based on their in-game performance data. The proposed methodology involves developing machine learning models to explore the predictive power of features that are derived from the data collected during gameplay on the GAME2AWE platform. Through a study involving fifteen elderly participants, we demonstrate that utilizing in-game data can achieve a high classification performance when predicting the motor and cognitive states. Various machine learning techniques were used but Random Forest outperformed the other models, achieving a classification accuracy ranging from 93.6% for cognitive screening to 95.6% for motor assessment. These results highlight the potential of using exergames within a technology-rich environment as an effective means of capturing the health status of seniors. This approach opens up new possibilities for objective and non-invasive health assessment, facilitating early detections and interventions to improve the well-being of seniors.

**Keywords:** machine learning; exergames; motor assessment; cognitive assessment; elderly; GAME2AWE; motion tracking



**Citation:** Danousis, M.; Goumopoulos, C. A

Machine-Learning-Based Motor and Cognitive Assessment Tool Using In-Game Data from the GAME2AWE Platform. *Informatics* **2023**, *10*, 59. <https://doi.org/10.3390/informatics10030059>

Academic Editor: Jiang Bian

Received: 24 May 2023

Revised: 4 July 2023

Accepted: 7 July 2023

Published: 9 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Background and Motivation

With the progressive aging of the global population, the increase in the number of elderly individuals has had profound implications for health and social care systems worldwide [1]. An inevitable consequence of aging is a decline in both motor and cognitive functionality, which can significantly impact the quality of life and independent living capabilities of the elderly. Deterioration in motor function manifests in reduced coordination, strength, balance, and endurance, leading to mobility limitations and increased fall risk [2]. Concurrently, cognitive decline is characterized by a decrease in mental capabilities such as memory, attention, and problem-solving skills, which can escalate into severe conditions like dementia and Alzheimer's disease [3].

Given these implications, it is of paramount importance to devise innovative, engaging, and efficient interventions through which to counteract the aging-induced deterioration in motor and cognitive functionality [4]. Exercise is a well-established, non-pharmacological intervention that is used to promote overall health and well-being in aging populations [5,6]. However, traditional exercise programs can sometimes be mundane and challenging to adhere to—particularly for the elderly [7]. In this context, exergames (i.e., exercise games) have emerged as a promising tool through which to bridge this gap, providing both physical and cognitive stimulation while enhancing user engagement through a fun, interactive medium [8,9].

Exergames combine the benefits of physical exercise and cognitive training, promoting a dual-tasking environment that can significantly improve both motor and cognitive outcomes [10]. The incorporation of real-time feedback, goal-setting, and progressively challenging levels can stimulate cognitive functions such as attention, memory, and executive functions [11]. Simultaneously, the physical activity components of these games can enhance motor skills such as balance, strength, and coordination [12]. Furthermore, the interactive and entertaining nature of exergames can promote long-term adherence to exercise regimes among the elderly, which is a significant challenge in conventional physical and cognitive interventions [13].

In addition to their primary benefits, exergames offer significant auxiliary advantages to healthcare practitioners by enabling the development of personalized health intervention plans for each senior participant [14,15]. The real-time and granular data derived from the elderly users' interactions with the exergames provide an invaluable resource for tailoring a customized, user-centric approach through which to improve their motor and cognitive functions. This approach allows for the adjustment of the interventions based on the individual's progress and response, thereby optimizing the therapeutic benefits while minimizing potential risks or discomfort.

Furthermore, exergames introduce an innovative way for gathering in-game data [16]. In the context of exergames, the in-game data capture the spontaneous and intuitive interactions of the users within the gaming environment. Such data can offer insights into real-world performance and behavior, going beyond traditional health markers to provide a more holistic understanding of the user's health status. It can also shed light on user engagement and adherence patterns, which are both critical factors for the long-term success of any intervention program.

## 1.2. Related Work

The application of games extends beyond their health benefits, offering potential utility as non-invasive, ecologically valid assessment tools. A number of research studies have explored the possibility of using in-game performance metrics for the early detection of cognitive or motor decline, as well as for predicting responses to interventions. The following subsections offer an overview of studies that emphasize cognitive screening and physical health assessments based on in-game data collection.

### 1.2.1. Cognitive Assessment

The study conducted by Tong et al. investigated the viability of a serious game as a cognitive evaluation tool for senior patients in an emergency department [17]. The game, adapted for various elderly demographics, was evaluated against established cognitive tests like the Mini-Mental State Examination (MMSE) [18] and Montreal Cognitive Assessment (MoCA) [19]. Findings revealed a high patient acceptance of the game and a significant correlation between game performance and traditional cognitive test scores, suggesting its potential as a cognitive impairment screening tool in the elderly population.

In a similar study, a serious game named Smartkuber was developed. It demonstrated high concurrent, predictive, and content validity as a cognitive health screening tool, providing an engaging gaming experience for elderly players without significant learning effects, as well as showing a strong correlation with MoCA test scores [20].

Zygouris et al. evaluated the Virtual Supermarket Test, a serious game-based self-administered test, and found it to be a potent tool for detecting mild cognitive impairment (MCI) in older adults with subjective memory complaints [21]. The test demonstrated superior accuracy with a correct classification rate of 81.91% when compared to traditional tests such as the MoCA and the MMSE.

The COGNIPLAT platform, which features serious games that are designed to improve cognitive functions, provides machine learning models trained on performance data as a cognitive screening tool through which to differentiate between healthy individuals and those with cognitive impairments [22].

Another study described a serious game called AlzCoGame, which utilizes gamification strategies and machine learning to achieve high performance in detecting mild Alzheimer's disease or MCI [23]. This research demonstrates the potential of serious games as engaging and ecologically valid alternatives to traditional methods for neuropsychological assessment and cognitive screening.

### 1.2.2. Physical Health Assessment

While numerous studies have explored the impact of exergames on balance and physical capabilities in older adults [24–26], less attention has been given to the assessment of physical health states through using in-game data [27]. An overview of the most relevant research in this under-explored area is provided next.

A systematic review evaluated the reliability and validity of the Wii Balance Board (WBB) for assessing static standing balance [28]. The study confirmed the WBB as a reliable tool, but highlighted the impact of factors such as the reference criterion, test duration, outcome measure, data acquisition platform, and sample size on its validity.

In another study, researchers employed multidimensional exploratory data analysis and visualization techniques to extract meaningful measures from exergaming data for real-time balance quantification, highlighting speed, curvature, and a turbulence measure. These measurements are particularly promising due to their reflection of age-related changes in balance performance [29].

Mahboobeh et al. examined the use of the Personalized Serious Game Suite and Intelligent Motor Assessment Tests from the i-PROGNOSIS platform for the objective assessment of Parkinson's disease stages [30]. The study demonstrated that machine learning classifiers can effectively utilize data from these tools to accurately infer the stage of the disease with a high accuracy rate (>90%). These findings suggest a promising approach for remote monitoring and personalized interventions in Parkinson's disease.

Villegas et al. used time series analysis on the data collected from 32 participants using a WBB platform to distinguish between healthy individuals, those with diabetes, and those with diabetic neuropathy [31]. Utilizing statistical techniques and machine learning, a probabilistic model was created, yielding over 98% accuracy. This suggests that time series analysis could potentially supplement or replace questionnaire-based diagnosis in physical therapy.

### 1.3. Contribution

Through leveraging the capabilities of modern motion-sensing interfaces like the Microsoft Kinect sensor [32] and the concept of exergames, endeavors have been initiated through which to develop personalized exergaming platforms, like the GAME2AWE platform [33]. GAME2AWE serves both as an intervention and an assessment tool for tracking changes in motor skills and cognitive function in elderly individuals. Furthermore, through a participatory design approach for game scenarios, it offers an engaging gamified environment that encourages seniors to adhere to physical training routines as preventive measures through which to decrease the risk of falls [34].

This paper discusses the potential of the GAME2AWE platform in assessing senior motor and cognitive conditions based on their in-game performance and behavioral data. The platform records various data metrics during gameplay to facilitate the analysis of its impact on seniors and the development of assessment models. For instance, how quickly a senior reacts to stimuli in the game might be indicative of their cognitive speed, while the precision of their movements could provide information about their motor skills.

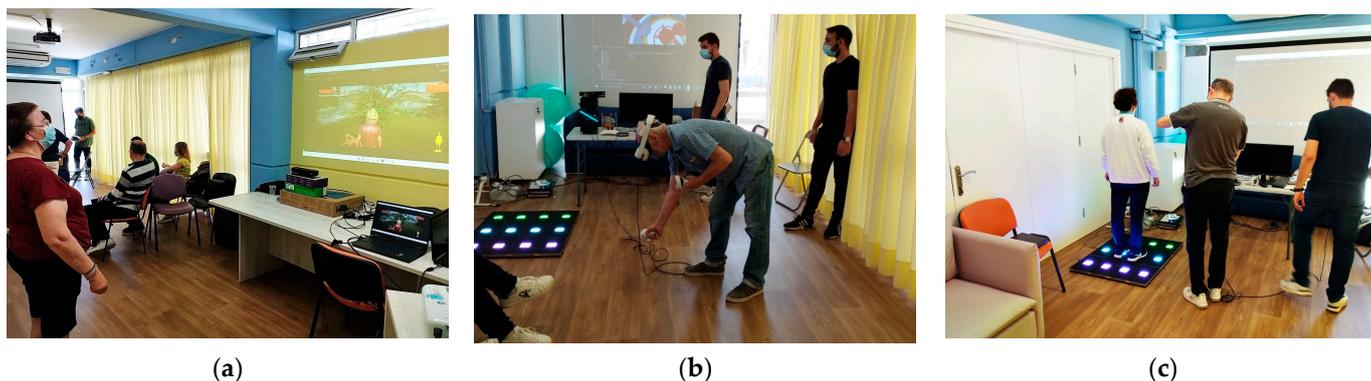
This paper specifically emphasizes the development of machine learning models based on in-game data in order to efficiently estimate the motor and cognitive states of the elderly. Throughout the course of this research, several machine learning methodologies were deployed. However, it was the Random Forest algorithm that yielded superior performance over alternative models, attaining a classification accuracy that spanned from 93.6% for cognitive screening to 95.6% for motor evaluation.

Implementing these models to elderly individuals who use this platform could facilitate widespread screening and could identify those who require additional care. It could also provide a valuable tool for health experts, enabling them to identify individuals who might need assistance or intervention remotely. To the best of our knowledge, this is the first study that concurrently examines the assessment of motor and cognitive states via the in-game data collected from an exergame platform and from machine learning classification models.

## 2. Materials and Methods

### 2.1. GAME2AWE Platform

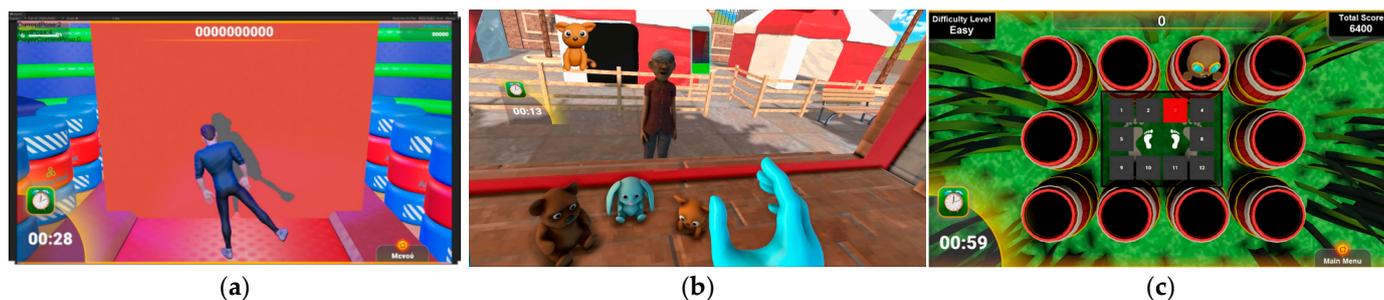
GAME2AWE is an innovative exergaming platform designed specifically for the elderly. It integrates advanced features through a combination of cutting-edge hardware and intelligent software, creating gaming experiences aimed at enhancing their physical and cognitive functions [35]. A particular goal of GAME2AWE is to utilize algorithmic techniques to analyze the in-game data, thereby enabling a dynamic adaptation of exergame activities by adjusting the level of difficulty based on the data analysis [36]. The platform uses a combination of three technologies, i.e., motion sensing, virtual reality (VR), and smart floor, in order to provide various kinds of gaming experiences and exercises (Figure 1). The platform provides a simple authentication method; one that is suitable for the elderly so that it can store user performance data and game preferences during playing sessions. In total, the platform provides 16 exergames that are organized into two themes: “Life on a Farm” and “Fun Park Tour”.



**Figure 1.** The deployment of the GAME2AWE platform in the research study: (a) demonstration of Kinect-based exergaming; (b) demonstration of VR-based exergaming; and (c) demonstration of smart floor exergaming.

For Kinect-based exergames, seniors need to stand in a suitable spot with enough surrounding space for the sensor to track their body posture and to accommodate the necessary game-related movements. In these types of exergames, the player’s physical movements control an avatar within the game world that mimics their movements. In GAME2AWE, the avatar can be a character (e.g., a farmer) or an object (e.g., an air balloon, or a pair of shoes). Some space is also necessary for VR-based exergames, given that the corresponding game scenarios involve a certain amount of movement from the player. For the development of VR games, Facebook’s Oculus Quest 2 technology is used, which includes a VR mask worn by the player and the controllers held in their hands. The smart floor, comprising tiles with integrated force sensors and an LED screen, features a modular, puzzle-like structure that simplifies assembly and allows for an easy adaptation to any space [37]. The smart floor is essentially used as a large touch screen to project the movement patterns that seniors should follow in a game scenario. Physiotherapists and geriatric specialists have guided the selection and integration of movements into the gameplay, which is aimed at enhancing balance and strength—thereby potentially reducing fall risk [34].

Figure 2 shows the snapshots of representative exergames that are featured in the GAME2AWE platform. The game “Hole in the Wall” is an example of a Kinect-based exergame. In this game, players are required to adopt the appropriate body posture, as indicated by the schematic representation on the wall. This wall swiftly approaches the player’s avatar, thus creating a dynamic and engaging gaming experience. In each round, the player aims to accumulate as many points as possible by successfully passing through as many walls as they can. The game encompasses a variety of poses that engage all parts of the body, including hand, foot, and body positions.



**Figure 2.** Snapshots from representative exergames on the GAME2AWE platform: (a) the “Hole in the Wall” exergame, an example of a Kinect-based game where players must match their body posture to a rapidly approaching wall; (b) the “Bazaar” exergame, a VR-based game where players interact with virtual objects and customers in a marketplace setting; and (c) the “Whack a Mole” exergame, a smart floor-based game where players must quickly step on illuminated tiles to score points.

An example of a game in the VR category is “Bazaar”. In this game, the player is presented with various items, such as toys, which they are expected to hand over to customers approaching their counter. Upon a customer’s arrival, an image of the requested item appears on the player’s left, and a timer bar on their right indicates the time remaining for the player to select the correct item. To hand an item to a customer, the player simply needs to extend their hand and touch the corresponding item.

Finally, an example of an exergame utilizing smart floor technology is “Whack a Mole”. In this game, the player’s objective is to quickly step on illuminated tiles, which represent the emergence of moles, in order to score points. The player begins by standing on the central tiles of the smart floor and then must monitor the surrounding tiles. When a tile lights up, indicating the appearance of a mole, the player must quickly step on that tile to ‘repel’ the mole and gain points. If the player’s reaction is too slow, the mole disappears, and the scoring opportunity is lost. The frequency of mole appearances and the duration for which they stay visible determines the game’s difficulty level. The game is designed to simultaneously exercise both physical and cognitive abilities.

For each game, four difficulty levels are established, with a fifth option provided for dynamic difficulty adjustment. These levels are defined by a variable number of parameters, the values of which dictate the difficulty levels. For instance, in the game “Hole in the Wall”, the difficulty level is determined by several parameters such as the speed of the wall’s movement toward the player, the number of walls to pass through in a round, and the variety of postures that can be requested during a game round. These parameters can be adjusted by a specialist using the platform’s user interface (Figure 3). Alternatively, these parameters can be dynamically adjusted by the software, based on an analysis of the player’s performance data.



**Figure 3.** Difficulty level adjustment for the GAME2AWE exergames: (a) difficulty level selection screen; (b) difficulty level parameters adjustment screen for the game “Hole in the Wall”.

#### 2.1.1.1. Fruit Harvest Exergame

This study collects data from the deployment of the GAME2AWE exergames, with a specific emphasis on the “Fruit Harvest” game (Figure 4), which is designed to primarily engage and train muscles in the upper part of the body. In the game, players control an avatar that is used to gather ripe fruits and deposit them into baskets within a set time frame. The challenge increases with the speed of fruit ripening and the addition of a second basket. Meanwhile, points are deducted for improper harvesting or allowing fruit to rot. Therefore, the game offers not only the potential for a range of exercises and movements, but also the opportunity for cognitive training.



**Figure 4.** Screenshot from the “Fruit Harvest” game, which is a part of the “Life on a Farm” theme and is based on the Kinect motion sensor.

In general, the goal of the “Fruit Harvest” game is to accumulate points by gathering a maximum number of fruits, where the player’s body movements play a crucial role in controlling the harvesting action. In this way, seniors can train their upper body muscles and increase or maintain their balance skills in a gamified manner. The game’s user interface is kept minimal, displaying only the relevant information. Consistency is also maintained, with the earned points always displayed on the top right of the screen, while the vitality bar

is shown on the left side. The vitality bar visually represents the level of physical activity achieved through gameplay actions.

The parameters that define the game's difficulty level are as follows:

- i. *State change speed*: this is the average time interval, measured in seconds, that elapses for a fruit to transition between its states (green, ripe, and rotten).
- ii. *Appearance mode*: this involves an equal distribution of the harvest target on both sides (left/right), with fruits first appearing on one side and then on the other, or the side where the appearance is continuously alternating, or where fruits appear randomly on either side.
- iii. *Harvest target*: the number of fruits that must be collected in a round of the game.

### 2.1.2. Data Management

The interaction of the elderly with the GAME2AWE platform is captured through the use of multiple supporting technologies. An example of a technology used in this context is the Kinect sensor, which focuses on detecting the movements of seniors and acts as a means of capturing their body position and gestures while they engage with the platform. This is achieved by extracting and analyzing the skeletal structure and outline of the body. By actively observing and perceiving the state of the seniors, valuable insights and feedback are obtained and conveyed through the data that are collected.

Moreover, the protection of senior privacy is a priority, and this requires the data that are collected from the GAME2AWE platform to be anonymized. The following data can be entered and accessed through the GAME2AWE platform:

1. Senior demographic information and account login details.
2. Past data derived from the game scenarios played by the seniors such as the following: date, user ID, game ID, game name, difficulty level, score, vitality level, missed sessions, completed movements, performed movement timestamp, and confidence.
3. Statistical information on the seniors' performance in the game, their accomplishment of goals, and their level of engagement.
4. Motor and cognitive clinical assessment test results.

The Game2AWE platform employs an SQLite database system for the efficient storage and management of data. The data are structured into various tables, including user identity and attributes, difficulty levels, static and dynamic game information, and the performance data that are associated with the required player movements. For example, the *Game\_log* table stores the data pertaining to player behavior during a game round, such as score points earned for achieving game goals, timestamps, and the vitality bar level (which signifies successful user movements and provides an additional incentive within the game). As another example, the *Actions* table captures game events, including successful actions, faults, and instances of no action when one was needed or instances of actions when none was needed. It includes fields for action type, reaction time, and the associated round, providing valuable insights into user performance and behavior in game tasks. The *Movement* table stores information such as movement type, detection confidence percentage, side of the movement, idle time prior to the movement, various timestamps, and secondary keys to other tables. It provides insights into user behavior and physical interactions during gameplay, supporting game mechanics, improving user experience, and facilitating the assessment of the user's motor condition. The *Round* table, on the other hand, provides a summary of each game round, including the player's accumulated points, round status, timestamps, difficulty level, and a reference to the *Game* table for additional game and player information. A part of the GAME2AWE database schema is provided in Figure 5, whereby the focus is on the tables that are most relevant to this study.

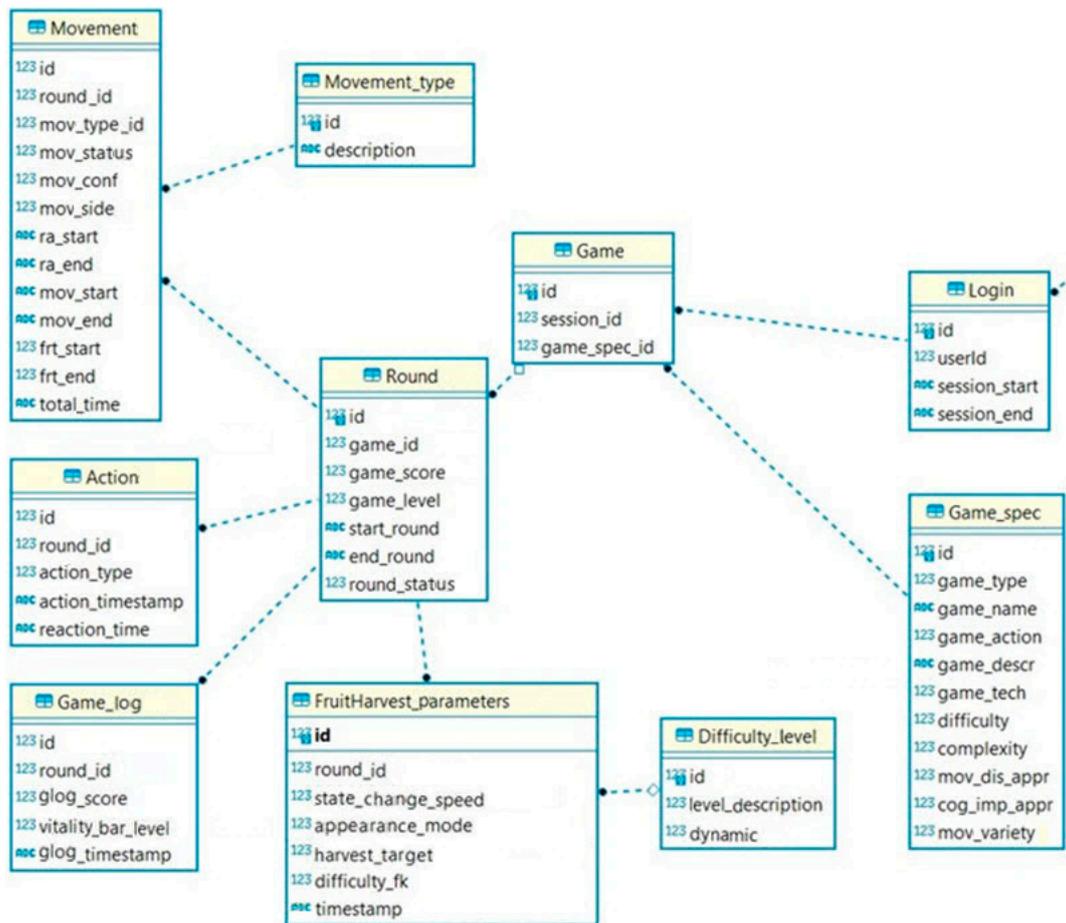


Figure 5. Part of the database schema of the GAME2AWE platform.

## 2.2. Study Design

### 2.2.1. Data Characteristics

The proposed study utilizes the data collected from seniors who interacted with the “Fruit Harvest” game. In this game, the user’s performance data are stored at the end of each performed movement or when specific game events occur, such as a fruit rotting. The stored data includes movement metrics, such as the time required to perform a movement, game metrics like the score and vitality bar level, and action metrics including the reaction time to a game stimulus and the type of action (i.e., harvesting too early, on time, or too late).

In addition, to assess the seniors’ actual motor and cognitive levels, specific clinical tests were conducted prior to their engagement with the GAME2AWE platform. The participants were classified into normal and abnormal states based on specific cut-off values, as defined in accordance with the relevant scientific literature. In the cognitive domain, the selected cut-off values denote the presence of mild cognitive impairment (MCI), which indicate the abnormal cognitive state within the context of this study. In the motor domain, the selected cut-off values indicate an increased risk of falling, representing an abnormal motor state. These tests were conducted by experienced research assistants.

For cognitive assessment, the MoCA test [19] was employed. The MoCA was designed to serve as a rapid tool for detecting MCI, providing an initial insight into the cognitive state of the examinee. The test includes exercises organized into different modules that cover cognitive functions such as attention, concentration, abstract thinking, memory, language, orientation, and executive functions. The maximum possible score is 30 points, and the time to complete the test is approximately 10 min. To differentiate between normal and abnormal cognitive states, cut-off scores of 26 and 23 were utilized for participants with an

educational level above primary and up to primary, respectively. These scores are based on MoCA normative data that are specific to the Greek population [38]. Our previous experience involving the evaluation of a cognitive training platform, where expert neurologists acknowledged these thresholds, also informed the selection of these thresholds [39].

Three motor assessment tests were administered to evaluate the participants' motor abilities: the Berg Balance Scale (BBS) [40], the 30 Second Sit to Stand Test (30SST) [41], and the Time Up and Go (TUG) [42] test. BBS is a widely used clinical assessment tool that evaluates an individual's static and dynamic balance abilities, and it is particularly suitable for older adults. It consists of 14 simple activities that assess various aspects of balance, ranging from transitioning from a sitting position to maintaining balance on one foot. Each item is graded on a scale of 0 to 4, with a maximum score of 56 indicating optimal balance. Lower scores on the BBS indicate a higher risk of falling. In this study, specific cut-off scores were employed: a score of 45 was used to identify individuals at an increased risk of falling [40], while a score of 51 can predict falls in individuals with a history of falls (with a sensitivity of 91% and specificity of 82%) [43].

The 30SST is a reliable and valid assessment tool for evaluating lower limb functional mobility. It measures the participant's ability to repeatedly transition from a seated position to a standing position and back within a 30 s timeframe. This test is commonly used to assess lower limb strength, flexibility, balance, and endurance, and has demonstrated a predictive value for fall risk in older individuals. The test outcome is quantified by the number of completed sit-to-stand repetitions within the given time, with a higher count indicating greater lower limb functional mobility. Table 1 shows the average number of repetitions by age and gender. Scores below the average suggest lower limb weakness, which is considered a risk factor for falls.

**Table 1.** The 30SST cut-off scores [44].

| Age Category | Male | Female |
|--------------|------|--------|
| 65–64        | <14  | <12    |
| 65–69        | <12  | <11    |
| 70–74        | <12  | <10    |
| 75–79        | <11  | <10    |
| 80–84        | <10  | <9     |
| 85–89        | <8   | <8     |
| 90–94        | <7   | <4     |

Lastly, the TUG test measures the time it takes for an individual to rise from a chair, walk a distance of three meters, turn around, walk back to the chair, and sit down again. The duration is measured in seconds, where a shorter time signifies improved mobility and balance, as well as a reduced risk of falling. Each subject performs the TUG test twice, and the best recorded time is used for scoring. If the completion time of the TUG test exceeds 12 seconds, it is indicative of an increased risk of falling [45].

### 2.2.2. Demographics of Participants

The research sample used in this study consisted of fifteen (N = 15) participants who were classified based on the specified cut-off values for the cognitive and motor clinical tests; this was conducted in normal and abnormal states, as shown in Table 2. The study included seniors aged between 60 and 79 years, with the highest number of participants being 66 years old. The mean age of the participants was 68.1 ( $\pm 4.9$ ) years. The gender composition of the study participants followed a ratio of 1:3, with 5 male seniors and 10 female seniors. Data collection occurred over a period spanning from November 2022 to March 2023 through regular sessions of the GAME2AWE platform, which were conducted at an elderly care center. Each participating senior played approximately twenty rounds of the "Fruit Harvest" game.

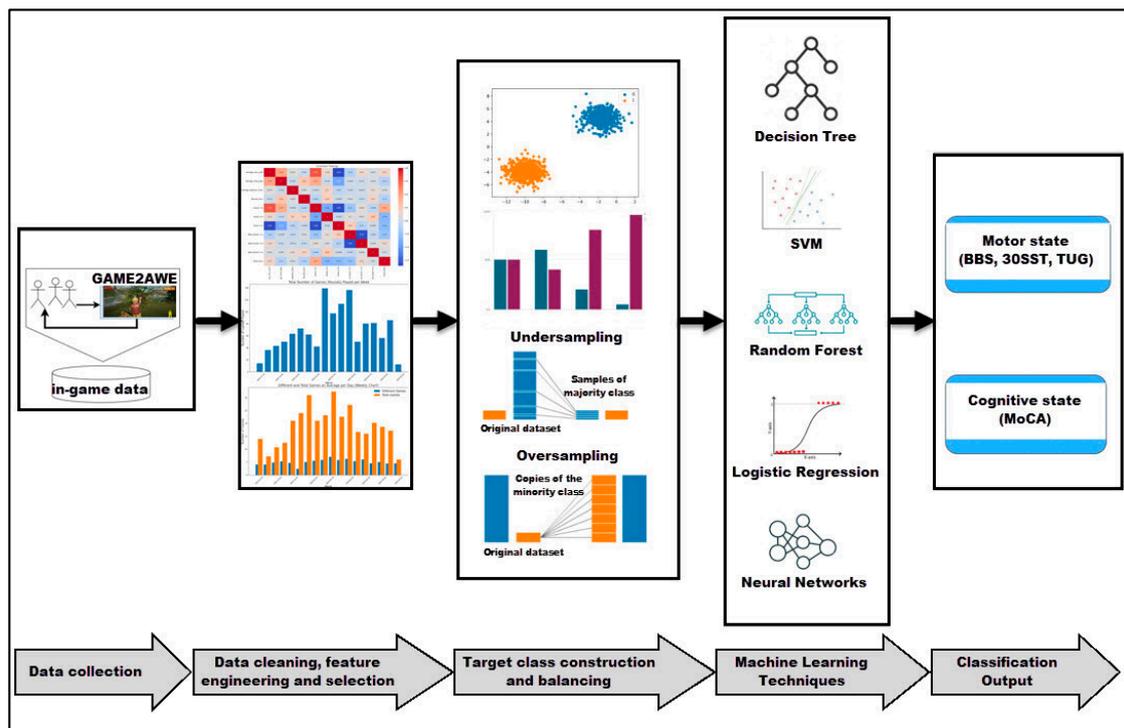
**Table 2.** Distribution of the cognitive and motor states (normal/abnormal) among the study participants.

| State    | MoCA | BBS | 30SST | TUG |
|----------|------|-----|-------|-----|
| Normal   | 11   | 12  | 10    | 11  |
| Abnormal | 4    | 3   | 5     | 4   |

The study adhered to the ethical principles outlined in the Helsinki Declaration, and obtained approval from the Ethical Research Committee of the University of the Aegean (reference number 70682). Written informed consent was obtained from all participants involved in the study. More details regarding the selection process of the seniors and the wider context of evaluating the GAME2AWE platform can be found in the work of Goumopoulos et al. [35].

### 2.3. Machine Learning Methodology

The aim of the proposed methodology is to identify and utilize the relevant features extracted from in-game data to accurately classify seniors into normal and abnormal motor and cognitive states, as was defined earlier. This approach enhances the potential of the GAME2AWE platform as a non-clinical tool for monitoring the status of seniors and for detecting any deterioration, including the transition from a normal to an abnormal stage. The stages of the proposed methodology are illustrated in Figure 6 and elaborated upon below.



**Figure 6.** Schematic representation of the sequential stages involved in the proposed methodology.

#### 2.3.1. Data Collection

Acquiring high-quality and sufficient data is crucial for obtaining reliable predictions in machine learning models. The GAME2AWE platform ensures this by capturing extensive and representative data from user gameplay and by implementing effective data management mechanisms. The data characteristics have been previously discussed, and the dataset used in the analysis was obtained from the GAME2AWE platform. This analysis incorporated in-game data collected from the “Fruit Harvest” game along with the results of clinical motor and cognitive tests, as presented in Table 3. It should be noted that only a subset of the attributes (such as *game\_score* and *game\_length*) were individually applicable

for subsequent analysis. The remaining attributes, defined at a granularity that is lower than a round (e.g., the confidence level of movement detection), were used to generate statistically based engineered features (for instance, the average value of the *move\_conf* attribute within the context of a round).

**Table 3.** Original dataset derived from the in-game data as a base for the proposed analysis.

| Attribute       | Type        | Description   |
|-----------------|-------------|---|
| game_score      | Number      | The score of a game round   |
| game_level      | Categorical | The difficulty level of a game round<br>1: Beginner, 2: Easy, 3: Medium, 4: Hard, and 5: Dynamic  |
| game_length     | Number      | The duration of a game round in seconds   |
| dlp_appearance  | Categorical | Appearance mode (it is a difficulty level parameter):<br>1: One by one side, 2: Continuous alternation, and 3: Random   |
| dlp_speed       | Number      | State change speed: the average time to change harvest state in seconds   |
| dlp_target      | Number      | Harvest target: number of fruits to collect in a game round   |
| vitality_level  | Number      | The ratio of expected successful movements in a game round, which is also represented in the user interface as the vitality bar level                           |
| move_conf       | Number      | The confidence level of movement detection  |
| move_side       | Categorical | The side of the performed movement:<br>−1: left side, 1: right side   |
| idle_time       | Number      | The idle time between two consecutive movements   |
| action_type     | Categorical | The type of action performed in the gameplay:<br>1: harvest too early, 2: harvest on time, and 3: harvest too late  |
| reaction_time   | Number      | The time between a game stimulus and the detection of the corresponding action  |
| MoCA            | Number      | The score of the MoCA test per participant  |
| BBS             | Number      | The score of the BBS test per participant   |
| 30SST           | Number      | The score of the 30SST test per participant   |
| TUG             | Number      | The score of the TUG test per participant   |
| motor_state     | Categorical | The target binary class for the motor level assessment based on the cut-off values of the motor clinical tests (BBS, 30SST, and TUG):<br>0: normal, 1: abnormal |
| cognitive_state | Categorical | The target binary class for the cognitive level assessment based on the cut-off values of the MoCA test:<br>0: normal, 1: abnormal                              |

### 2.3.2. Data Cleaning, Feature Engineering, and Selection

After data collection, the subsequent steps fall under the category of “data pre-processing”, which involves transforming the data into a suitable format for machine learning model training. Outliers, which are observations that significantly deviate from the rest of the data, can impact the accuracy of statistical analysis and machine learning algorithms. Null values, indicating missing data, can also lead to biased or inaccurate results.

In this study, the outliers were identified and eliminated via the interquartile range (IQR) method, which calculates the range between the 75th and 25th percentiles, as well as identifies the data points outside of this range as outliers. The IQR method is popular for outlier detection as it considers the data distribution, resulting in a smoother outcome [46]. Furthermore, null values were addressed by inputting the most plausible value, which predominantly was zero (0), thus ensuring data integrity while minimizing the introduction of biases.

As part of the data transformation process, feature engineering was conducted to capture the underlying data patterns and to enhance the understanding of players’ performances and overall game experience. This involved performing arithmetic, as well as cumulative and statistical transformations through which to generate new features. The newly created features were then evaluated for their significance and correlation with the target classification class.

The performance of players in terms of correct actions may vary during a game round due to factors such as the learning curve, environment, and fatigue. By analyzing the

trend of correct movements throughout the game, valuable information about the player’s condition can be obtained. Three basic functions were introduced to measure the relevant performance metrics, contributing to a better understanding of the player’s overall game experience. These functions are the *Rate of Early (RoE) harvest*, the *Rate of on Time (RoT) harvest*, and the *Rate of Late (RoL) harvest*, which are defined as follows:

$$RoE(Round_t) = \frac{N_E}{N_A} \tag{1}$$

$$RoT(Round_t) = \frac{N_T}{N_A} \tag{2}$$

$$RoL(Round_t) = \frac{N_L}{N_A} \tag{3}$$

where  $Round_t$  is a specific round played by the player ( $t$  refers to the current round,  $t - 1$  to the previous, etc.),  $N_E$  is the number of “harvest too early” actions (type 1),  $N_T$  is the number of “harvest on time” actions (type 2),  $N_L$  is the number of “harvest too late” actions (type 3), and  $N_A$  is the number of all the action logs on a specific round. After establishing these functions, we employed them within smaller timeframe windows during the progression of a game round to gain insights into the players’ performances. We divided the game rounds into five equal parts based on action logs and computed the above functions for each part. By applying the skewness formula [47] to the resulting five values, obtained by calculating the mean, median, and standard deviation, we derived a single value representing the skewness of  $RoE$ ,  $RoT$ , and  $RoL$  for each round. As an example, we provide the skewness of  $RoL$ :

$$Skew(RoL_t) = \sum_{i=1}^5 \frac{mean(RoL(Round_{ti})) - median(RoL(Round_{ti}))}{stdev(RoL(Round_{ti}))} \tag{4}$$

where  $Round_{ti}$  represents the  $i$ th part of  $Round_t$  round. As previously mentioned, when the harvest target for the round exceeds the player’s capability, there is a decline in their performance throughout the round. Therefore, by applying the  $RoL$  function to each of the five parts mentioned earlier, we calculate the asymmetry of  $RoL$  with Equation (4).

Through using attributes specified in Table 3 and by applying statistical transformations, such as Equation (4), new features were derived in order to summarize the information pertaining to each round, as presented in Table 4. These features serve as complements to round-level attributes, including *game\_score*, *game\_length*, *dlp\_appearance*, *dlp\_speed*, and *dlp\_target*.

**Table 4.** Additional features defined and explored in the developed models.

| Predictor           | Description   |
|---------------------|---|
| avg_move_conf       | The average value of the move_conf attribute within the context of a round.     |
| avg_move_side       | The average value of the move_side attribute within the context of a round.     |
| avg_idle_time       | The average value of the idle_time attribute within the context of a round.     |
| avg_reaction_time   | The average value of the reaction_time attribute within the context of a round. |
| action1_%           | The percentage of type 1 actions within the context of a round.                 |
| action2_%           | The percentage of type 2 actions within the context of a round.                 |
| action3_%           | The percentage of type 3 actions within the context of a round.                 |
| action1_Q1/Q2/Q3/Q4 | The percentage of type 1 actions in the 1st/2nd/3rd/4th quarter of a round      |
| action1_skew        | The skewness of $RoE$ (type 1 actions within the context of a round).           |
| action2_skew        | The skewness of $RoT$ (type 2 actions within the context of a round).           |
| action3_skew        | The skewness of $RoL$ (type 3 actions within the context of a round).           |
| game_logs           | The number of game logs within the context of a round.                          |

In this study, feature selection was conducted using visualization tools such as correlation heatmaps to assess the correlation between features and the target class. A correlation heatmap provides insights into the relationships between variables, thus aiding in the variable selection for machine learning models, as well as providing a visual representation of patterns and trends in the dataset.

An analysis of pairwise correlations between features was performed using Pearson’s correlation coefficient to quantify the strength and direction of the relationships. The resulting correlation matrix was visualized in a heatmap (Figure 7) to identify the highly correlated features. Features are deemed redundant when they exhibit a highly significant and strong correlation, typically with an absolute value of 0.8 or higher ( $p < 0.05$ ). In such cases, the variable with the highest correlation to the remaining variables is chosen for exclusion. Based on this analysis, the following features were removed from further analysis: *action1\_Q2/Q4*, *action3\_%*, *dpl\_appearance*, and *dpl\_speed*.

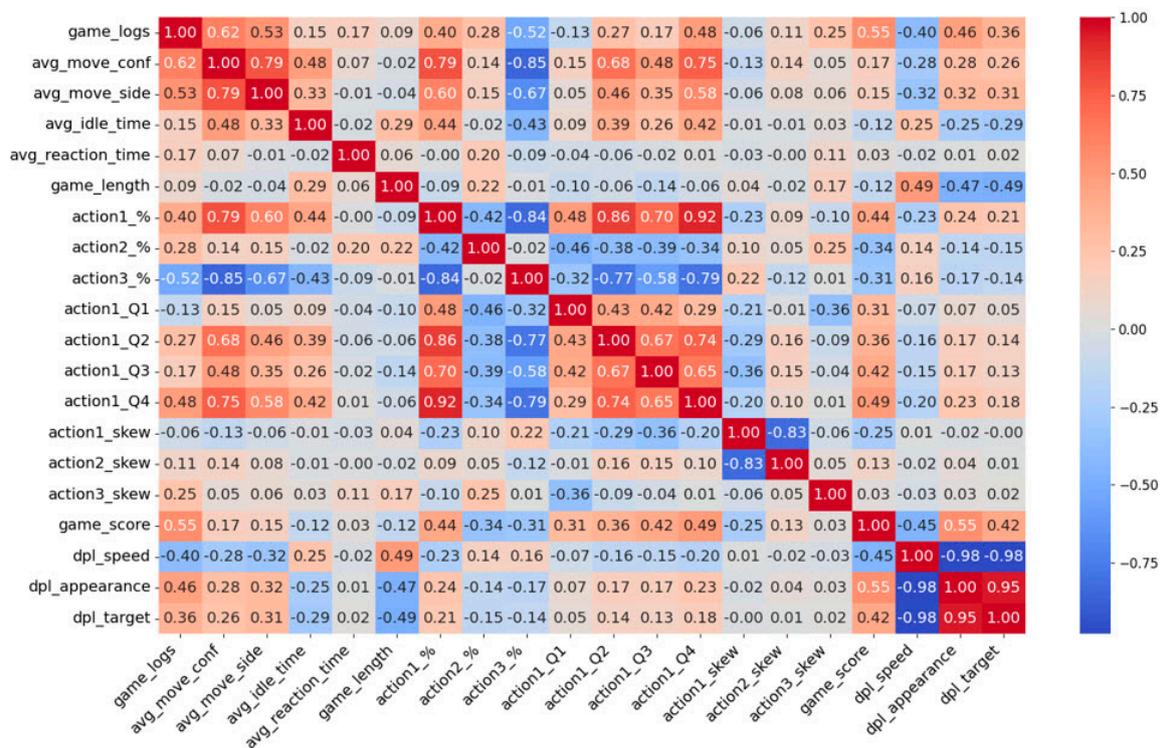
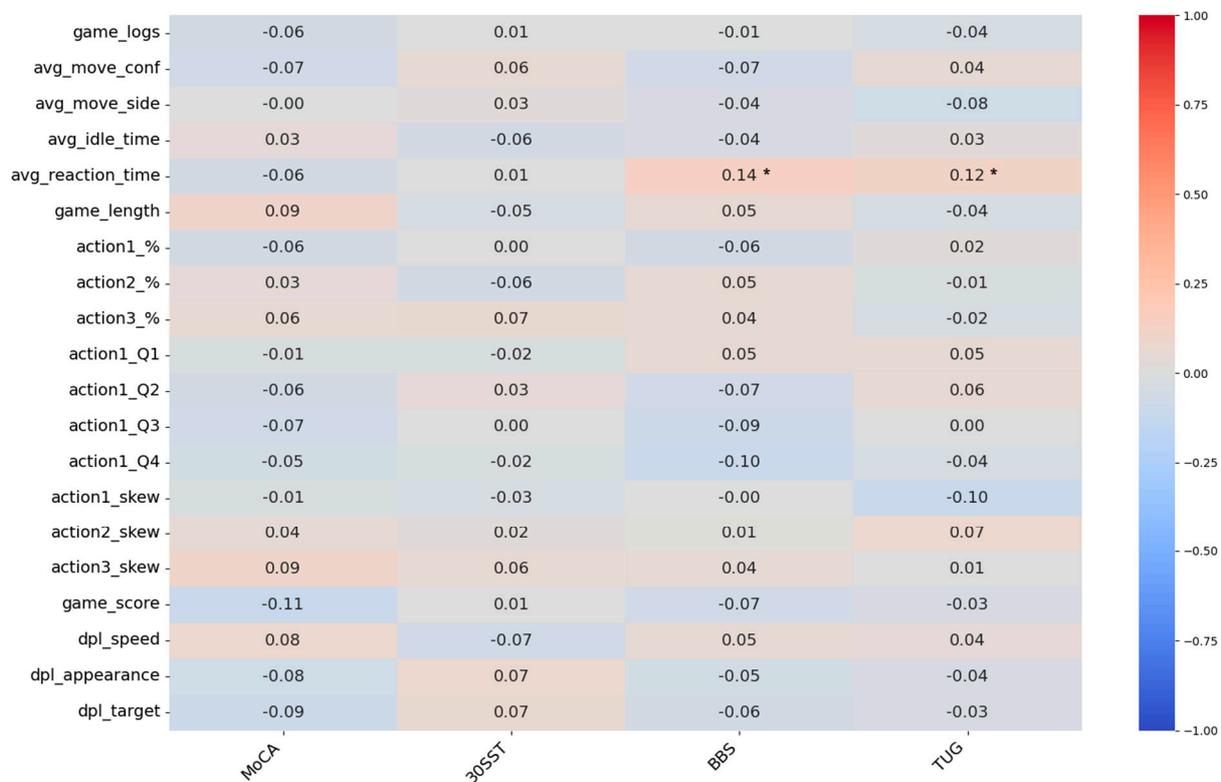


Figure 7. Heatmap of the pairwise correlation between features that were calculated using Pearson’s correlation coefficient.

Subsequently, a correlation matrix encompassing a comprehensive set of features and the four different target classes was provided, and it was constructed with the aim of visualizing and understanding the relationships between all predictors and each target class (Figure 8). This process is instrumental in identifying which predictors are most strongly associated with each target class—a key insight for enhancing the accuracy of the predictive model. Furthermore, an analysis of variance (ANOVA) test was conducted between the features and the four different classes with the aim of determining whether significant differences could be detected between the ‘Normal’ and ‘Abnormal’ outcomes, which was based on any of the predictors in the dataset. It was concluded that most feature-target pairs did not demonstrate many statistically significant variations between the two outcomes. In fact, statistically significant variations were only found for two pairs: *avg\_reaction\_time*—*BBS* (F-statistic: 5.919;  $p$ : 0.02) and *avg\_reaction\_time*—*TUG* (F-statistic: 4.093;  $p$ : 0.04). This observation led to the conclusion that, given the lack of significant variations of this nature, the deployment of predictive models should utilize the majority

of the predictor set. This approach provides the models with the opportunity to uncover the hidden correlations between combinations of features and classes.



**Figure 8.** Heatmap of the correlations between features and target classes, which was calculated using Pearson’s correlation coefficient. An analysis of variance (ANOVA) test was performed to identify the statistically significant differences between features and target classes (\*  $p < 0.05$ ).

### 2.3.3. Target Class Construction and Balancing

Selecting or constructing the target class of a dataset is a crucial step in the overall process as it determines the goal of the analysis. The construction of target classes and the rationale behind their boundaries were discussed in the data characteristics section. It is important to address the issue of class imbalance before deploying machine learning algorithms to prevent overfitting and to ensure optimal generalization performance; this is achieved by considering the disproportionate number of samples in one class.

To ensure unbiased results in our model, we needed to address the issue of imbalanced participants and, consequently, game rounds between the Normal and Abnormal target classes (Table 2). Two commonly used methods to tackle this problem are undersampling and oversampling. Considering the relatively small dimensionality of the dataset, undersampling appeared to be a viable option. However, due to the limited number of entries in the dataset, we opted for the oversampling method to avoid losing valuable information. Specifically, to address the issue of the imbalanced dataset, the data are balanced within each target class by utilizing the SMOTE approach [48]. This technique creates new synthetic samples by interpolating between existing samples of the minority class. By adding these synthetic samples to the dataset, we effectively increase the representation of the minority class, resulting in a more balanced dataset. The synthetic samples are carefully generated to preserve the underlying patterns and characteristics of the minority class. This approach helps to mitigate the potential bias caused by the class imbalance and enhances the performance and generalization of the machine learning model.

Table 5 summarizes the augmentation of the original dataset, which was achieved by using the SMOTE for the different classes examined. As is shown, SMOTE significantly increased the number of instances in each class, effectively addressing the issue of class

imbalance. As a specific example, in the case of the MoCA samples, there were initially a total of 300 instances (220 designated as ‘Normal’ and 80 as ‘MCI’). Following augmentation, the total number of instances rose to 440, with the minority class being increased until it matched the number of samples in the majority class, thus signifying an increase of 46.7%.

**Table 5.** Summary of the dataset augmentation when using SMOTE for the different classes examined.

| Dataset                    | MoCA   | BBS    | 30SST   | TUG    |
|----------------------------|--------|--------|---------|--------|
| Original (Normal/Abnormal) | 220/80 | 240/60 | 200/100 | 220/80 |
| SMOTE augmented (total)    | 440    | 480    | 400     | 440    |
| Increase (%)               | 46.7   | 60.0   | 33.3    | 46.7   |

#### 2.3.4. Machine Learning Techniques

The selected features were used as the input for the machine learning (ML) algorithms to carry out the classification of the motor/cognitive states. The Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Artificial Neural Networks (ANNs) ML methods were employed. Some studies have investigated the selection of algorithms and parameters based on the nature of the problem, including automated techniques like Landmarking and Sampling-based Landmarking [49]. However, in our case, given the manageable size of our data, the selection of ML models was primarily based on a combination of intuition and a trial-and-error approach.

The DT model is a tree-based algorithm that partitions the data based on feature conditions to make predictions [50]. The SVM model is a binary classifier that separates data points using a hyperplane in a high-dimensional space [51]. The SVM model employs a polynomial kernel of four degrees. This kernel enables the handling of nonlinear data by mapping it into a higher-dimensional space, which is where it becomes linearly separable. The RF model is an ensemble of decision trees that combines multiple individual trees to make predictions [52]. The RF model was implemented with a population of 100 trees, as it was observed that the model’s performance did not show significant improvement by adding more trees. LR is a linear classification algorithm that models the relationship between the features and the target variable [53]. This model was included in the algorithm set for comparison, serving as a baseline for the other algorithms.

The last two models in our set of ML classifiers are ANNs with different architectures. ANNs are a set of interconnected nodes that simulate the behavior of the human brain to learn and make predictions [54]. We chose ANNs because, when properly parameterized, they tend to perform well in supervised ML problems. The first network is a simple three-layer (3L) neural network, with sigmoid activation in the second layer and softmax activation in the output layer for decision making. The second network has seven layers (7L), including max-pooling and dropout layers, and uses a combination of sigmoid and rectified linear unit activations, with a softmax activation in the output layer for probability estimation. During the training stage, early stopping based on validation accuracy was implemented, with a learning rate of 0.001 and with the cross-entropy loss function (which is commonly used in binary classification). The optimizer used was the adaptive moment estimation, a variant of the stochastic gradient descent algorithm that adapts learning rates and incorporates momentum [55].

To assess the performance of the proposed classification models, we employed the leave-one-subject-out cross-validation (LOSOCV) protocol, which is commonly used when dealing with relatively low sample sizes [56] (as is the case in our study). This approach allows us to evaluate the model’s robustness to unseen data at the subject level, the highest level of the sampling hierarchy. In cross-validation, the data instances are divided into non-overlapping segments (folds), with one segment serving as the test dataset while the remaining segments are used for training the classification model. The performance of each model is then averaged across all folds.

For each classifier (DT, SVM, RF, LR, and ANNs), the overall classification accuracy was computed using the LOSOCV scheme. Additionally, the metrics of precision, recall,

F1 score, and Matthew’s Correlation Coefficient (MCC) were calculated, as presented in Table 6. In unbalanced datasets, where there is an unequal distribution of positive and negative instances, the MCC provides a comprehensive measure of the model’s prediction performance [57]. MCC ranges from −1 to +1, where a value of +1 indicates a perfect classifier, 0 indicates a random classifier, and −1 indicates a completely inaccurate classifier.

**Table 6.** Performance metrics used in this study for the ML classification models.

| Metric    | Description  | Formula  |
|-----------|--|--|
| Accuracy  | A measure of the overall correctness of a classification model | $\frac{TP+TN}{TP+TN+FP+FN}$  |
| Precision | A measure of the accuracy of positive predictions              | $\frac{TP}{TP+FP}$   |
| Recall    | A measure to correctly identify positive instances             | $\frac{TP}{TP+FN}$   |
| F1 score  | A single metric that combines precision and recall             | $2 * \frac{Precision * Recall}{Precision + Recall}$                      |
| MCC       | A measure of the quality of binary classifications             | $\frac{TP * TN - FP * FN}{\sqrt{(TP+FP) * (TP+FN) * (TN+FP) * (TN+FN)}}$ |

TP: true positives; TN: true negatives; FP: false positives; FN: false negatives; and MCC: Matthew’s Correlation Coefficient.

### 3. Results

The classification metrics obtained from LOSOCV for the scenarios outlined in Table 2—and which correspond to the DT, SVM, RF, LR, and ANNs classifiers—are depicted in Tables 7–10.

**Table 7.** Classification performance results for the MoCA-based cognitive assessment.

| ML Model | Accuracy        | Precision | Recall | F1 Score | MCC    |
|----------|-----------------|-----------|--------|----------|--------|
| DT       | 0.8480 ± 0.0581 | 0.7413    | 0.6425 | 0.6875   | 0.4227 |
| SVM      | 0.7957 ± 0.0264 | 0.9303    | 0.6426 | 0.7580   | 0.6235 |
| RF       | 0.9232 ± 0.0398 | 0.9362    | 0.9065 | 0.9144   | 0.8450 |
| LR       | 0.6371 ± 0.0395 | 0.6490    | 0.5988 | 0.6223   | 0.2755 |
| 3L ANN   | 0.8227 ± 0.0000 | 0.8743    | 0.7661 | 0.8110   | 0.6574 |
| 7L ANN   | 0.8712 ± 0.0017 | 0.9142    | 0.8194 | 0.8642   | 0.7465 |

**Table 8.** Classification performance results for the BBS-based motor assessment.

| ML Model | Accuracy        | Precision | Recall | F1 Score | MCC    |
|----------|-----------------|-----------|--------|----------|--------|
| DT       | 0.9091 ± 0.0161 | 0.9388    | 0.8745 | 0.9095   | 0.8142 |
| SVM      | 0.9200 ± 0.0281 | 0.9840    | 0.8538 | 0.9140   | 0.8478 |
| RF       | 0.9561 ± 0.0182 | 0.9671    | 0.9498 | 0.9593   | 0.9044 |
| LR       | 0.7555 ± 0.0387 | 0.7825    | 0.7084 | 0.7433   | 0.5136 |
| 3L ANN   | 0.9044 ± 0.0016 | 0.9247    | 0.8809 | 0.9021   | 0.8100 |
| 7L ANN   | 0.9075 ± 0.0047 | 0.9458    | 0.8652 | 0.9035   | 0.8183 |

**Table 9.** Classification performance results for the 30SST-based motor assessment.

| ML Model | Accuracy        | Precision | Recall | F1 Score | MCC    |
|----------|-----------------|-----------|--------|----------|--------|
| DT       | 0.8304 ± 0.0304 | 0.8720    | 0.7890 | 0.8172   | 0.6877 |
| SVM      | 0.8296 ± 0.0126 | 0.9228    | 0.7237 | 0.8096   | 0.6769 |
| RF       | 0.9119 ± 0.0366 | 0.9471    | 0.8859 | 0.9057   | 0.8143 |
| LR       | 0.6437 ± 0.0602 | 0.6931    | 0.5260 | 0.5917   | 0.3000 |
| 3L ANN   | 0.8010 ± 0.0329 | 0.8555    | 0.7369 | 0.7888   | 0.6117 |
| 7L ANN   | 0.8391 ± 0.0156 | 0.8778    | 0.7889 | 0.8308   | 0.6819 |

**Table 10.** Classification performance results for the TUG-based motor assessment.

| ML Model | Accuracy        | Precision | Recall | F1 Score | MCC    |
|----------|-----------------|-----------|--------|----------|--------|
| DT       | 0.8317 ± 0.0433 | 0.8088    | 0.8700 | 0.8399   | 0.6677 |
| SVM      | 0.8042 ± 0.0352 | 0.7520    | 0.9071 | 0.8218   | 0.6236 |
| RF       | 0.9133 ± 0.0632 | 0.9066    | 0.9400 | 0.9214   | 0.8522 |
| LR       | 0.6633 ± 0.0382 | 0.6507    | 0.7133 | 0.6790   | 0.3301 |
| 3L ANN   | 0.8483 ± 0.0150 | 0.7976    | 0.9433 | 0.8618   | 0.7158 |
| 7L ANN   | 0.8717 ± 0.0150 | 0.8094    | 0.9733 | 0.8837   | 0.7594 |

The results for the MoCA-based cognitive assessment are presented in Table 7. The assessment aims to predict the cognitive state of the participants, determining whether they fall into the normal or MCI category based on predefined MoCA cut-off values. The LR algorithm performed the worst in predicting the cognitive state, with an accuracy of 0.6371. This suggests that the relationship between the predictive features and the target class may not be easily captured by a simple function. The SVM model achieved an accuracy of 0.7957, while the DT and RF models achieved accuracy means of 0.8480 and 0.9232, respectively. The RF model achieved the highest accuracy, followed by the 7-layer ANN (0.8712). The models' F1 scores ranged from 0.6223 to 0.9144, indicating varying levels of recall and accuracy in distinguishing between the two states. It is worth noting that the 7-layer ANN, despite performing worse than the RF model, exhibited a lower standard deviation in its performance, making its behavior more predictable.

The results for the BBS-based motor assessment are presented in Table 8. The assessment aims to predict the motor state of the participants by determining whether they belong to the normal or fall risk category based on predefined BBS cut-off values. The accuracy of the models ranged from 0.7555 to 0.9561, with the RF model achieving the highest accuracy mean. The standard deviations of the accuracy means ranged from 0.0016 to 0.0387. The precision means of the models varied from 0.7825 to 0.9840, with the SVM model achieving the highest precision mean, followed closely by the RF model. In terms of F1 score, the RF model had the highest performance, ranging from 0.7433 to 0.9593 across the models. Overall, the results indicate that the RF model outperformed the other models in terms of accuracy mean, recall, F1 score, and MCC. However, the SVM model exhibited a strong performance characterized by high mean accuracy and precision, while the 7-layer ANN model also performed robustly, with a notably low standard deviation.

The results for the 30SST-based motor assessment are presented in Table 9. The assessment aims to predict the motor state of the participants by determining whether they belong to the normal or fall risk category based on predefined 30SST cut-off values. The accuracy means of the models ranged from 0.6437 to 0.9119. The 7-layer ANN model achieved an accuracy mean of 0.8391, while the RF model had the highest accuracy mean of 0.9119. The standard deviations of the accuracy means varied from 0.0126 to 0.0602. The RF model also had the highest F1 score, ranging from 0.5917 to 0.9057 across the models. Overall, these results indicate that the RF model performed the best among the tested models, demonstrating high accuracy, recall, F1 score, MCC, and precision mean. The 7-layer ANN model also performed well, with a high accuracy mean and relatively low standard deviation in its performance. The SVM model also demonstrated similar performance with the lowest standard deviation and with high precision means.

The results for the TUG-based motor assessment are presented in Table 10. The assessment aims to predict the motor state of the participants by determining whether they belong to the normal or fall risk category based on predefined TUG cut-off values. The accuracy means of the models ranged from 0.6633 to 0.9133. The 7-layer ANN model achieved an accuracy mean of 0.8717, which closely followed the RF model (which had the highest accuracy mean of 0.9133). The standard deviations of the accuracy means varied from 0.0150 to 0.0632. The RF model also had the highest F1 score, ranging from 0.6790 to 0.9214 across the models. The models' precision varied from 0.6507 to 0.9066, with the RF model achieving the highest precision mean. Based on its high accuracy mean, F1 score,

MCC, and precision mean, the RF model demonstrated the best overall performance among the tested models. However, the 7-layer ANN model is also a strong contender, exhibiting a high accuracy mean and a relatively low standard deviation.

#### 4. Discussion

By assessing the individuals while they engaged in the GAME2AWE exergames, we can capitalize on the benefits of maintaining their flow state, reducing test anxiety, increasing motivation, and obtaining more realistic results through extended sampling periods. This approach enables the periodic assessments of users' performance on the game platform over an extended duration, thus facilitating the capture of more accurate and representative data. Previous research has primarily focused on utilizing in-game data from serious games for the purpose of cognitive screening [17,20–23], and exergames for motor training [24–26], with occasional applications in physical health assessments [30,31]. In contrast, our proposed work addresses the challenge of designing a game platform that functions as both an intervention and an assessment tool for evaluating motor and cognitive abilities in elderly individuals.

The typical approach to investigating an algorithm's ability to predict the physical and cognitive conditions of the elderly is to use two separate games, each targeting a specific aspect. Indeed, while this approach was considered during our study design, the data used in this study came from a single exergame, i.e., Fruit Harvest, due to two primary reasons. Firstly, the game involves both cognitive and motor skills in a balanced manner. It thus provides a comprehensive set of data that allows us to concurrently assess both aspects. This strategy enabled us to extract a holistic set of features representing both the cognitive and motor states of the elderly participants. Secondly, considering the size and scope of this pilot study, we emphasized the importance of data quality and consistency. Focusing on a single game allowed us to streamline the feature engineering process, thereby making the problem more manageable. This approach was especially beneficial considering the intricate nature of interpreting various types of data.

In the process of feature engineering, the features derived from the initial dataset were essentially aggregates of information. Notably, statistical measures like averages and skewness were chosen for attributes such as "Reaction time" and "Movement confidence". This selection was not arbitrarily made, but rather, it was predicated on their significance in capturing critical information from each game round in statistically useful measures. The specific statistical formulas employed in this task were determined based on a combination of empirical and observational methods, along with careful consideration of relevant research works [58,59]. In relation to the transformation of original features, it should be clarified that these were not validated through a particular study, but they instead represent a product of our research design, which was devised to optimally capture the information needed for this study.

The majority of the results indicate that most of the compared models achieved a prediction accuracy higher than 80% in a balanced sample. This leads us to the conclusion that there is a significant predictive power between the selected features and the target classes. In all assessment scenarios, the RF model outperformed the other models, achieving an accuracy ranging from 93.6% for cognitive screening to 95.6% for motor assessment. However, it is worth noting that despite the common belief that a well-tuned ANN usually outperforms simpler models due to its ability to represent complex functions, the observation that a RF algorithm—with a max tree depth of 5 and a total of 100 trees—performs better than the tested ANN models suggests that further optimization of the neural network architectures or the addition of more layers, even at the expense of increased computing overhead, may be necessary.

Another noteworthy point is that these models were trained on balanced datasets, despite the initial target class distribution being highly imbalanced. This was conducted because our primary objective was to establish a baseline for assessment prediction models of this nature. Given the nature of the problem, which involves identifying abnormal

motor or cognitive states in a dataset, the goal of training these ML models could also be shifted toward achieving higher recall rather than precision and accuracy. By doing so, the GAME2AWE platform as an assessment tool would be able to identify the elderly who are at a higher risk more frequently, thus allowing medical professionals to examine more cases and reducing the likelihood of misclassifying individuals as normal when they actually require assistance.

Along the same lines, the performance of the ML models in this study offers a deep insight into their predictive capacities when working with physical and cognitive targets. Particularly noteworthy is the performance of the DT model. As indicated in Table 7, despite exhibiting high accuracy (0.8480), it presents relatively lower values in metrics that are closely associated with individual predictions, such as the F1 score (0.6875) and recall (0.6425). This disparity in the DT model's performance could potentially be attributed to the reduced dimensionality of the sample size. Machine learning models tend to perform optimally when provided with a large, diverse dataset that covers a broad spectrum of possible input scenarios. A smaller sample size might lead to models that are overfitted to the training data and which thus perform poorly on unseen data. Improving the performance in such cases may necessitate an in-depth discussion and investigation into various approaches. These could include enhancing the dimensionality and diversity of the data, refining the model's hyperparameters, considering different tree-building algorithms, or even leveraging ensemble methods to augment model performance.

Extracting a person's MoCA score is a time-consuming process that requires expertise, but assessing it through a less resource-intensive method could greatly assist individuals with limited healthcare access. Utilizing the Game2AWE platform may be a feasible solution, potentially leading to the improved detection of early stage MCI and better preventions of dementia. Along the same lines, the proposed methodology of utilizing ML models to predict motor states based on the in-game data from Game2AWE presents a promising solution. This approach has the potential to improve accessibility in balance assessments, to facilitate the early detection of balance impairment, and mitigating falls among older adults—all while harnessing the power of digital platforms to extend its reach and impact.

#### *Limitations and Future Work*

The restricted sample size of elderly participants in the study evidently limits the generalizability of the results. Nevertheless, the study's credibility is reinforced by conducting data collection in a controlled setting and by incorporating clinical validated metrics. Therefore, the suggested study allows for the extension of the current analysis to encompass a more extensive group of senior individuals. Additionally, this can facilitate the collection of comprehensive data from elderly individuals across different motor and cognitive stages, thereby transforming the binary classification problem explored in this study into a multi-class classification task. Furthermore, longitudinal data acquisition from the same subjects over an extended period of time would provide an opportunity to further validate and generalize the performance of the classifiers, especially when considering the variations among different individuals and those within the same individual over time.

Along the same lines, categorizing cognitive ability into just two categories may oversimplify the complexities of cognitive health conditions. This approach is reductive, as it potentially fails to capture the full spectrum and nuances of cognitive abilities or stages of cognitive impairment. Additionally, it could introduce bias due to the selection of the cutoff point between these two categories. Considering these points, a more nuanced approach, such as dividing cognitive ability into multiple severity ranges (for example, normal, MCI, and dementia) may provide a more accurate representation of cognitive conditions in future studies. This multi-class target approach would capture a wider array of cognitive abilities, providing a more comprehensive understanding of cognitive health in elderly populations.

In this work, the MoCA test was employed as the benchmark for cognitive assessment. While utilizing MoCA measurements and data is valuable, it is crucial to acknowledge

the potential of other metrics that can further enrich the study and model development. Neuropsychological tests such as the Trail Making Test A and B [60], as well as the Stroop Color-Word test [61], provide additional insights into executive functioning, processing speed, visual search ability, and mental agility, thereby significantly enhancing our capability to predict cognitive decline and its related issues.

The study examines the “Fruit Harvest” exergame from the GAME2AWE platform. However, future analysis will involve the inclusion of the data from extended pilot studies, encompassing diverse interactions of seniors with the GAME2AWE platform for an extended duration (such as a period of six months). This will enable investigation into the impact of exergame characteristics on the assessment of motor and cognitive skills in elderly individuals over time. Expanding the scope of the study to encompass additional evaluations of motor skills and non-motor symptoms, such as deteriorating voice quality, psychological distress, poor dietary habits, and sleep disturbances, has the potential to create a comprehensive assessment tool. Integration with the full range of exergames in the GAME2AWE platform would enhance information from multiple sources, and deep learning classification schemes could be employed to effectively represent the health status of seniors in the embedded space. Additionally, applying more advanced machine learning algorithms, such as long short-term memory networks or transformer networks, directly to the original round records is a promising avenue for future exploration.

The proposed gaming platform utilizes hardware such as the Kinect sensor and smart floor to track user movements, as well as to assess their physical and cognitive conditions. While the smart floor is custom-made and faces challenges in cost and distribution, the Kinect sensor could be replaced with a simple commercial video camera to enhance the platform’s accessibility and its mainstream utilization. While replacing the Kinect sensor with a camera poses challenges in recognizing user posture, the use of libraries like Google’s ARCore can address this issue [35]. Despite its potential performance differences when compared to the Kinect sensor, advancements in computer vision, as highlighted in recent research [62], offer promising solutions for body tracking problems and can be implemented either through custom implementations or through available libraries.

## 5. Conclusions

This paper presents a solution through which to address the time-consuming process of assessing the health condition of elderly individuals. Our research focuses on exploring the predictive power of the features derived from the GAME2AWE platform during gameplay in predicting the motor and cognitive states of the elderly. The experimental findings demonstrate the potential for achieving a classification accuracy that exceeds 90% when predicting the motor and cognitive states via in-game data sources. These results strongly support the use of exergames within a technology-rich environment as a means through which to effectively capture the health status of seniors via machine learning techniques. This viewpoint broadens the scope of health assessment by incorporating additional quantitative approaches that offer detailed measurements for monitoring symptoms remotely, specifically the deterioration of elderly skills. It involves a shift in the location of assessing motor and cognitive states from clinical settings to the seniors’ more familiar environments.

**Author Contributions:** Conceptualization, C.G.; methodology, C.G.; software, M.D.; validation, C.G., and M.D.; investigation, C.G. and M.D.; resources, M.D.; data curation, M.D.; writing—original draft preparation, M.D.; writing—review and editing, C.G., and M.D.; visualization, C.G.; supervision, C.G.; project administration, C.G.; funding acquisition, C.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was co-financed by the European Regional Development Fund of the European Union, and Greek national funds through the Operational Program Competitiveness, Entrepreneurship, and Innovation, under the call RESEARCH-CREATE-INNOVATE (project code: T2EDK-04785).

**Institutional Review Board Statement:** The study was conducted in accordance with the guidelines in the Declaration of Helsinki, and was approved by the Ethical Research Committee of the University of the Aegean (reference number 70682).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are available upon reasonable request to the corresponding authors.

**Acknowledgments:** We would like to thank our colleagues in the GAME2AWE project for their support in the development of the game platform, as well as all the elderly adults and the healthcare experts who participated in this study. We are grateful to Giannis Matsoukas for his support in the pilot study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Prince, M.J.; Wu, F.; Guo, Y.; Robledo, L.M.G.; O'Donnell, M.; Sullivan, R.; Yusuf, S. The burden of disease in older people and implications for health policy and practice. *Lancet* **2015**, *385*, 549–562. [\[CrossRef\]](#)
2. Demanze Laurence, B.; Michel, L. The fall in older adults: Physical and cognitive problems. *Curr. Aging Sci.* **2017**, *10*, 185–200.
3. Glisky, E.L. Changes in cognitive function in human aging. In *Brain Aging: Models, Methods, and Mechanisms*; CRC Press: Boca Raton, FL, USA; Taylor & Francis: Boca Raton, FL, USA, 2007; pp. 3–20.
4. Schoene, D.S.; Sturnieks, D.L. Cognitive-Motor Interventions and Their Effects on Fall Risk in Older People. In *Falls in Older People: Risk Factors, Strategies for Prevention and Implications for Practice*; Cambridge University Press: Cambridge, UK, 2021; pp. 287–310.
5. Gremeaux, V.; Gayda, M.; Lepers, R.; Sosner, P.; Juneau, M.; Nigam, A. Exercise and longevity. *Maturitas* **2012**, *73*, 312–317. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Levin, O.; Netz, Y.; Ziv, G. The beneficial effects of different types of exercise interventions on motor and cognitive functions in older age: A systematic review. *Eur. Rev. Aging Phys. Act.* **2017**, *14*, 20. [\[CrossRef\]](#)
7. Schutzer, K.A.; Graves, B.S. Barriers and motivations to exercise in older adults. *Prev. Med.* **2004**, *39*, 1056–1061. [\[CrossRef\]](#)
8. Larsen, L.H.; Schou, L.; Lund, H.H.; Langberg, H. The physical effect of exergames in healthy elderly—A systematic review. *Games Health J. Res. Dev. Clin. Appl.* **2013**, *2*, 205–212. [\[CrossRef\]](#)
9. Kappen, D.L.; Mirza-Babaei, P.; Nacke, L.E. Older Adults' Physical Activity and Exergames: A Systematic Review. *Int. J. Hum.-Comput. Interact.* **2019**, *35*, 140–167. [\[CrossRef\]](#)
10. Kamnardsiri, T.; Phirom, K.; Boripuntakul, S.; Sungkarat, S. An Interactive Physical-Cognitive Game-Based Training System Using Kinect for Older Adults: Development and Usability Study. *JMIR Serious Games* **2021**, *9*, e27848. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Wang, R.-Y.; Huang, Y.-C.; Zhou, J.-H.; Cheng, S.-J.; Yang, Y.-R. Effects of Exergame-Based Dual-Task Training on Executive Function and Dual-Task Performance in Community-Dwelling Older People: A Randomized-Controlled Trial. *Games Health J.* **2021**, *10*, 347–354. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Yang, C.M.; Hsieh, J.S.C.; Chen, Y.C.; Yang, S.Y.; Lin, H.C.K. Effects of Kinect exergames on balance training among community older adults: A randomized controlled trial. *Medicine* **2020**, *99*, e21228. [\[CrossRef\]](#)
13. Uzor, S.; Baillie, L. Investigating the long-term use of exergames in the home with elderly fallers. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 2813–2822.
14. Velazquez, A.; Martínez-García, A.I.; Favela, J.; Ochoa, S.F. Adaptive exergames to support active aging: An action research study. *Pervasive Mob. Comput.* **2017**, *34*, 60–78. [\[CrossRef\]](#)
15. Guimarães, V.; Oliveira, E.; Carvalho, A.; Cardoso, N.; Emerich, J.; Dumoulin, C.; Swinnen, N.; De Jong, J.; de Bruin, E.D. An Exergame Solution for Personalized Multicomponent Training in Older Adults. *Appl. Sci.* **2021**, *11*, 7986. [\[CrossRef\]](#)
16. Konstantinidis, E.I.; Bamidis, P.D.; Billis, A.; Kartsidis, P.; Petsani, D.; Papageorgiou, S.G. Physical Training In-Game Metrics for Cognitive Assessment: Evidence from Extended Trials with the Fitforall Exergaming Platform. *Sensors* **2021**, *21*, 5756. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Tong, T.; Chignell, M.; Tierney, M.C.; Lee, J. A Serious Game for Clinical Assessment of Cognitive Status: Validation Study. *JMIR Serious Games* **2016**, *4*, e7. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Folstein, M.F.; Folstein, S.E.; McHugh, P.R. "Mini-Mental State": A Practical Method for Grading the Cognitive State of Patients for the Clinician. *J. Psychiatr. Res.* **1975**, *12*, 189–198. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Nasreddine, Z.S.; Phillips, N.A.; Bédirian, V.; Charbonneau, S.; Whitehead, V.; Collin, I.; Cummings, J.L.; Chertkow, H. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *J. Am. Geriatr. Soc.* **2005**, *53*, 695–699. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Boletsis, C.; McCallum, S. Smartkuber: A Serious Game for Cognitive Health Screening of Elderly Players. *Games Health J.* **2016**, *5*, 241–251. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Zygouris, S.; Iliadou, P.; Lazarou, E.; Giakoumis, D.; Votis, K.; Alexiadis, A.; Triantafyllidis, A.; Segkouli, S.; Tzouvaras, D.; Tsiatsos, T.; et al. Detection of mild cognitive impairment in an at-risk group of older adults: Can a novel self-administered serious game-based screening test improve diagnostic accuracy? *J. Alzheimer's Dis.* **2020**, *78*, 405–412. [\[CrossRef\]](#)

22. Karapapas, C.; Goumopoulos, C. Mild Cognitive Impairment Detection Using Machine Learning Models Trained on Data Collected from Serious Games. *Appl. Sci.* **2021**, *11*, 8184. [[CrossRef](#)]
23. Mezrar, S.; Bendella, F. Machine learning and Serious Game for the Early Diagnosis of Alzheimer's Disease. *Simul. Gaming* **2022**, *53*, 369–387. [[CrossRef](#)]
24. van Diest, M.; Lamothe, C.J.; Stegenga, J.; Verkerke, G.J.; Postema, K. Exergaming for balance training of elderly: State of the art and future developments. *J. Neuroeng. Rehabil.* **2013**, *10*, 101. [[CrossRef](#)] [[PubMed](#)]
25. Wüest, S.; Borghese, N.A.; Pirovano, M.; Mainetti, R.; Van De Langenberg, R.; de Bruin, E.D. Usability and Effects of an Exergame-Based Balance Training Program. *Games Health J. Res. Dev. Clin. Appl.* **2014**, *3*, 106–114. [[CrossRef](#)]
26. Shih, M.C.; Wang, R.Y.; Cheng, S.J.; Yang, Y.R. Effects of a balance-based exergaming intervention using the Kinect sensor on posture stability in individuals with Parkinson's disease: A single-blinded randomized controlled trial. *J. Neuroeng. Rehabil.* **2016**, *13*, 78. [[CrossRef](#)] [[PubMed](#)]
27. Staiano, A.E.; Calvert, S.L. The promise of exergames as tools to measure physical health. *Entertain. Comput.* **2011**, *2*, 17–21. [[CrossRef](#)] [[PubMed](#)]
28. Clark, R.A.; Mentiplay, B.F.; Pua, Y.H.; Bower, K.J. Reliability and validity of the Wii Balance Board for assessment of standing balance: A systematic review. *Gait Posture* **2018**, *61*, 40–54. [[CrossRef](#)] [[PubMed](#)]
29. Soancatl Aguilar, V.; van de Gronde, J.J.; Lamothe, C.J.C.; van Diest, M.; Maurits, N.M.; Roerdink, J.B.T.M. Visual data exploration for balance quantification in real-time during exergaming. *PLoS ONE* **2017**, *12*, e0170906. [[CrossRef](#)]
30. Mahboobeh, D.J.; Dias, S.B.; Khandoker, A.H.; Hadjileontiadis, L.J. Machine Learning-Based Analysis of Digital Movement Assessment and ExerGame Scores for Parkinson's Disease Severity Estimation. *Front. Psychol.* **2022**, *13*, 857249. [[CrossRef](#)]
31. Villegas, C.M.; Curinao, J.L.; Aqueveque, D.C.; Guerrero-Henríquez, J.; Matamala, M.V. Identifying neuropathies through time series analysis of postural tests. *Gait Posture* **2023**, *99*, 24–34. [[CrossRef](#)]
32. Zhang, Z. Microsoft Kinect Sensor and Its Effect. *IEEE MultiMedia* **2012**, *19*, 4–10. [[CrossRef](#)]
33. Goumopoulos, C.; Karapapas, C. Personalized Exergaming for the Elderly through an Adaptive Exergame Platform. In *Intelligent Sustainable Systems: Selected Papers of Worlds4 2022*; Springer Nature: Singapore, 2023; Volume 2, pp. 185–193.
34. Goumopoulos, C.; Chartomatsidis, M.; Koumanakos, G. Participatory Design of Fall Prevention Exergames using Multiple Enabling Technologies. In *ICT4AWE*; SciTePress: Setubal, Portugal, 2022; pp. 70–80. [[CrossRef](#)]
35. Goumopoulos, C.; Drakakis, E.; Gklavakis, D. Feasibility and Acceptance of Augmented and Virtual Reality Exergames to Train Motor and Cognitive Skills of Elderly. *Computers* **2023**, *12*, 52. [[CrossRef](#)]
36. Danousis, M.; Goumopoulos, C.; Fakis, A. Exergames in the GAME2AWE Platform with Dynamic Difficulty Adjustment. In *Entertainment Computing–ICEC 2022: 21st IFIP TC 14 International Conference, ICEC 2022, Bremen, Germany, 1–3 November 2022, Proceedings*; Springer International Publishing: Cham, Switzerland, 2022; pp. 214–223.
37. Goumopoulos, C.; Ougkrenidis, D.; Gklavakis, D.; Ioannidis, I. A Smart Floor Device of an Exergame Platform for Elderly Fall Prevention. In *Proceedings of the 2022 25th Euromicro Conference on Digital System Design (DSD), Maspalomas, Spain, 31 August–2 September 2022*; IEEE: New York, NY, USA, 2022; pp. 585–592.
38. Poptsi, E.; Moraitou, D.; Eleftheriou, M.; Kounti-Zafeiropoulou, F.; Papisozomenou, C.; Agogiatou, C.; Bakoglidou, E.; Batsila, G.; Liapi, D.; Markou, N.; et al. Normative Data for the Montreal Cognitive Assessment in Greek Older Adults with Subjective Cognitive Decline, Mild Cognitive Impairment and Dementia. *J. Geriatr. Psychiatry Neurol.* **2019**, *32*, 265–274. [[CrossRef](#)]
39. Goumopoulos, C.; Skikos, G.; Frounta, M. Feasibility and Effects of Cognitive Training with the COGNIPLAT Game Platform in Elderly with Mild Cognitive Impairment: Pilot Randomized Controlled Trial. *Games Health J.* **2023**, *ahead of print*. [[CrossRef](#)]
40. Berg, K.O.; Wood-Dauphinee, S.L.; Williams, J.I.; Maki, B. Measuring balance in the elderly: Validation of an instrument. *Can. J. Public Health Rev. Can. Sante Publique* **1992**, *83*, S7–S11.
41. Rikli, R.E.; Jones, C.J. Development and Validation of a Functional Fitness Test for Community-Residing Older Adults. *J. Aging Phys. Act.* **1999**, *7*, 129–161. [[CrossRef](#)]
42. Haines, T.; Kuys, S.S.; Morrison, G.; Clarke, J.; Bew, P.; McPhail, S. Development and Validation of the Balance Outcome Measure for Elder Rehabilitation. *Arch. Phys. Med. Rehabil.* **2007**, *88*, 1614–1621. [[CrossRef](#)] [[PubMed](#)]
43. Shumway-Cook, A.; Baldwin, M.; Polissar, N.L.; Gruber, W. Predicting the probability for falls in community-dwelling older adults. *Phys. Ther.* **1997**, *77*, 812–819. [[CrossRef](#)]
44. Rikli, R.E.; Jones, C.J. Functional Fitness Normative Scores for Community-Residing Older Adults, Ages 60–94. *J. Aging Phys. Act.* **1999**, *7*, 162–181. [[CrossRef](#)]
45. Bischoff, H.A.; Stähelin, H.B.; Monsch, A.U.; Iversen, M.D.; Weyh, A.; von Dechend, M.; Akos, R.; Conzelmann, M.; Dick, W.; Theiler, R. Identifying a cut-off point for normal mobility: A comparison of the timed 'up and go' test in community-dwelling and institutionalised elderly women. *Age Ageing* **2003**, *32*, 315–320. [[CrossRef](#)]
46. Iglewicz, B.; Hoaglin, D.C. *Volume 16: How to Detect and Handle Outliers*; Quality Press: Milwaukee, WI, USA, 1993.
47. Groeneveld, R.A.; Meeden, G. Measuring skewness and kurtosis. *J. R. Stat. Soc. Ser. D (Stat.)* **1984**, *33*, 391–399. [[CrossRef](#)]
48. Blagus, R.; Lusa, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinform.* **2013**, *14*, 106. [[CrossRef](#)]
49. Luo, G. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. *Netw. Model. Anal. Health Inform. Bioinform.* **2016**, *5*, 18. [[CrossRef](#)]
50. Maimon, O.Z.; Rokach, L. *Data Mining with Decision Trees: Theory and Applications*; World Scientific: Singapore, 2014; Volume 81.

51. Wang, L. (Ed.) *Support Vector Machines: Theory and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2005; Volume 177.
52. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
53. LaValley, M.P. Logistic regression. *Circulation* **2008**, *117*, 2395–2399. [[CrossRef](#)] [[PubMed](#)]
54. Jain, A.; Mao, J.; Mohiuddin, K. Artificial neural networks: A tutorial. *Computer* **1996**, *29*, 31–44. [[CrossRef](#)]
55. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2016**, arXiv:1609.04747.
56. Foster, K.R.; Koprowski, R.; Skufca, J.D. Machine learning, medical diagnosis, and biomedical engineering research-commentary. *Biomed. Eng. Online* **2014**, *13*, 94. [[CrossRef](#)]
57. Boughorbel, S.; Jarray, F.; El-Anbari, M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS ONE* **2017**, *12*, e0177678. [[CrossRef](#)]
58. Savadkoobi, M.; Oladunni, T.; Thompson, L. A machine learning approach to epileptic seizure prediction using Electroencephalogram (EEG) Signal. *Biocybern. Biomed. Eng.* **2020**, *40*, 1328–1341. [[CrossRef](#)]
59. Parsapoor, M.; Alam, M.R.; Mihailidis, A. Performance of machine learning algorithms for dementia assessment: Impacts of language tasks, recording media, and modalities. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 45. [[CrossRef](#)]
60. Corrigan, J.D.; Hinkeldey, N.S. Relationships between Parts A and B of the Trail Making Test. *J. Clin. Psychol.* **1987**, *43*, 402–409. [[CrossRef](#)]
61. Stroop, J.R. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **1935**, *18*, 643–662. [[CrossRef](#)]
62. Xie, K.; Wang, T.; Iqbal, U.; Guo, Y.; Fidler, S.; Shkurti, F. Physics-based human motion estimation and synthesis from videos. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 11532–11541.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.