Serious Game Development for the Diagnosis of Major Depressive Disorder Cases Using Machine Learning Methods

Athanasios Tsionas
Aristotle University of Thessaloniki
School of Informatics
Thessaloniki, Greece
atsionas@csd.auth.gr

Aristotelis Lazaridis
Aristotle University of Thessaloniki
School of Informatics
Thessaloniki, Greece
arislaza@csd.auth.gr

Ioannis Vlahavas
Aristotle University of Thessaloniki
School of Informatics
Thessaloniki, Greece
vlahavas@csd.auth.gr

ABSTRACT

Major Depressive Disorder (MDD) is a serious mental disorder that affects millions of adults, occasionally leading to life-threatening results. Current diagnostic tools for MDD mostly consist of questionnaires and/or long, specialized therapy sessions. In this work we present a serious game called "The Delivery", developed for diagnosing MDD in players. The video game has the players immerse into a realistic scenario, the development of which depends on their actions, that is, through conversations with in-game characters, completion of quests, and interactions with the environment. All in-game features and mechanics are designed to correspond to specific diagnostic criteria for MDD. We recorded gameplay data from labeled players (both MDD and non-MDD cases) in order to train Machine Learning models that can accurately distinguish gameplay behaviors MDD-positive and MDD-negative players.

CCS CONCEPTS

• Computing methodologies → Machine learning; **• Applied computing → Life and medical sciences**.

KEYWORDS

serious games, major depression disorder, machine learning, artificial intelligence

1 INTRODUCTION

Major Depressive Disorder (MDD) is the most common type of depression in adults. The characteristics of MDD can be really hard to identify and manage, especially due to the fact that they are triggered by psychological factors. Since identifying the exact pathological and psychological features in a patient is difficult, alternative methodologies for this disorder's diagnosis should be explored.

This poses a challenge that many scientists attempt to tackle in various ways. Serious games (i.e. games with a specific purpose other than entertainment) have also made a step towards the improvement of mental health issues [6]. In this paper, we present a serious game developed for the diagnosis of MDD in young adults, using Artificial Intelligence (AI) and Machine Learning (ML) methods. In particular, we developed a story-based video game, where players are able to alter the scenario with their actions and conversations with game characters, all of which are carefully designed to correspond and portray different scientific criteria used in the diagnosis of MDD. Using a small dataset, collected through the gameplay of few players diagnosed with and without MDD, we were able to developed a prototype system using Machine Learning models which lead to promising results on MDD detection.

Other serious games exist for the treatment or prevention of depression [5, 7], but, to the authors' knowledge, this is the fist serious game developed for the diagnosis of MDD that uses AI and ML methods, and does not rely on any wearables to collect data about the player. For instance, in [10], the authors use wearables to sample physiological activities (EEG, ECG, etc.) from potential patients with depression during a gameplay session of a serious game developed for this reason. Statistical measures were then used to make correlations between extracted signals and players with negative moods.

Moreover, our work is in line with the directions proposed by [9], since the serious game we developed gathers substantial data related to the player's cognitive behavior during the gameplay session. Our findings indicate that this proof-of-work concept can be scaled to a highly-accurate, non-interventional system used for the diagnosis of MDD.

In [11], the authors use the AVEC2016 dataset [14] to train classifiers for the purpose of diagnosing depression from voice data. Even though the results indicate a better-than-chance classification ability, the feasibility and implementation details of extracting voice data from users through a proposed application for smartphone devices is not considered, and the model is not tested on real-world scenarios.

2 METHODOLOGY

This section covers the basic steps taken in regard to extracting the resulting conclusions. These steps include data processing operations, as well as algorithm selection, implementation and tuning carried out for the experiments.

2.1 Gameplay scenario and Major Depressive Disorder characteristics

Initially we developed a serious, story-based game called "The Delivery"[1], for the diagnosis of Major Depressive Disorder (MDD) using Artificial Intelligence and Machine Learning methods. "The Delivery" follows a relatively realistic scenario, in which the player is found inside a building, playing as a character who has been asked to deliver medical supplies to a friend. Shortly afterwards, a huge earthquake shall trap the building and its residents, introducing an

[1] https://www.dropbox.com/s/g4c872t9a6uwcfh/theDelivery.zip?dl=0

GAITCS2020, September 02–04, 2020, Athens, Greece

Figure 1: *The Delivery* serious game

unexpected turn of events, while the player will have to deal with unsettling situations.

Throughout the game, the player is given the opportunity to have conversations with in-game characters that lead to different outcomes. Additionally, the player is able to interact with in-game surroundings and objects. Even though many of the given choices and actions may seem insignificant to the player, they have been carefully devised to correspond to different criteria used in the diagnosis of MDD. These criteria have been extracted from the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) [2].

More specifically, the conversations with in-game characters are implemented in such ways that, different player answers signal different probability (weight) of the player meeting a particular criterion of MDD. Moreover, with the help of a commonly used AI method in video games termed *Behavior Trees*, conversation flows are highly flexible and are adjusted suitably to maximize not only information extraction, but also the feeling of a realistic dialogue.

After gathering data from gameplay sessions of MDD and non-MDD players, we used Machine Learning methods to develop classifiers that can predict with a probability whether a gameplay session belongs to a player with MDD.

2.2 Gameplay mechanics

As in any serious game, the purpose is dual: entertainment for the player, and achievement of the specific "serious" goal. The first element includes all entertaining features, which are related to the player's perspective about the game.

The second element includes the mechanisms that make this game a serious game. For this project, it was essential to have a good understanding of both dimensions.

In "The Delivery", the player has to look after his in-game statistics, i.e. *Health*, *Fatigue* and *Sanity*. For the player to keep these statistics within limits, he has to explore the environment, find hidden items or complete side-quests.

To achieve our goals, the player has to imagine himself within the scenario, in order to make the diagnosis more accurate, therefore the game was designed with this in mind. In detail, it belongs to a first-person video game category, so that the player can only imagine the character's look, and the character's voice is never heard.

The most important statistic within the game is the *final score*, which is an indication of the MDD level, according to the in-game mechanics. The player is not aware of the existence of this metric, so as not to become influenced by it.

The final score is computed in various ways. For instance, it changes every time the player interacts with an in-game character. Every answer during a discussion has a different weight relative to a particular diagnostic criterion extracted from DSM-V (e.g. the player admits that he/she does not sleep well at nights usually). When a given answer is related to a specific DSM-V criterion, the final score changes accordingly.

Another way to increase the final score is through interaction with the environment. In order to keep statistics high, the player needs to interact with items within the surroundings. Showing sensitivity to characters and circumstances alters the final score in favor of a non-MDD label.

2.3 Data collection and processing

During a gameplay session, the player's actions and their outcomes are recorded. This data includes: changes in the player's basic statistics (Health, Fatigue, Sanity), the player's answers during discussions (in chronological order), and the side quests successfully completed.

We collected 26 gameplay sessions from two groups: players who were diagnosed with MDD (24%) and players who were not diagnosed with MDD and were highly unlikely to suffer from MDD (76%). The second group included cases who had never received an official MDD diagnosis in the past, and were asked to fill in the Patient Health Questionnaire (PHQ-9) [12]. This questionnaire corresponds to an official questionnaire used for the diagnosis of MDD, and was used in order to filter cases that had non-trivial probability of suffering from MDD, so as to keep only cases with a small probability of suffering from this disorder. This was a clear necessity in order to label players accurately as MDD-positive or MDD-negative. It should be noted that the members of the second group are not considered as "healthy", but they are considered as non-MDD cases (i.e. they could be suffering from another disorder). Unfortunately, practical restrictions did not allow us to develop a larger dataset.

2.4 Algorithms

Initially, we performed a visualization procedure in order to visualize the game's capability to distinguish MDD-positive and MDD-negative behaviors within the game, using Principal Component Analysis (PCA) [13].

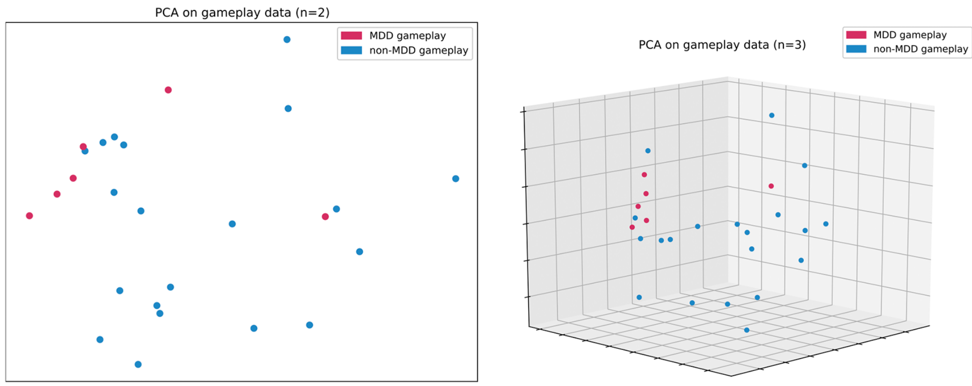

Figure 2: Data visualization after performing PCA with $n = 2$ and $n = 3$ components. MDD-positive players show a non-trivial difference in gameplay behavior than non-MDD players.

Then, we applied 4 classification methods: Decision Trees [4], k-Nearest Neighbors (kNN) [1] and Support Vector Machines (SVM) [15] with linear kernel, and Random Forests (RF) [3] in order to create a prediction model that would be able to classify a player as MDD-positive or MDD-negative, using our labeled data for the training procedure.

The RF classifier used $n = 20$ trees and the kNN classifier used $k = 6$ neighbors with the Euclidean distance metric. For the kNN classifier, we attempted to use an odd number of neighbors to avoid ties, but the optimal results were produced with $k = 6$.

3 RESULTS

In this section we present visualization results using the PCA dimensionality reduction method and performance evaluation results of Machine Learning models trained using gameplay data. Principal Component Analysis was applied for $n = 2$ and $n = 3$ (Figure 2) components to the data. Visualization of the resulting data points indicates that the game design has employed in a relatively accurate manner diagnostic criteria for MDD. More particularly, in both experiments, results show a dense area of MDD-positive cases, which represent their in-game behavior, with only a few outlier cases. On the contrary, MDD-negative cases are more sparse and do not follow particular behavioral patterns, which is expected since the game mechanics do not target such behaviors.

Additionally, we trained various classification models based on the gameplay data, with the purpose of predicting whether a gameplay session belongs to an MDD-positive or MDD-negative player. Even with such a small, imbalanced dataset such ours, all models after training showed promising results in managing to distinguish gameplay sessions of MDD-positive and MDD-negative players.

More particularly, we measured the performance of each model using the Accuracy, Precision, Recall, F1-score and F1-weighted score metrics, computed using a Leave-One-Out Cross Validation [8] scheme. Fine-tuning of the hyper-parameters was performed using a simple grid-search procedure. The results are presented in Table 1. The SVM classifier with linear kernel seems to have overall best results among the three classifiers, indicating that there is a hyperplane that can separate accurately the data. Random Forests also have a relatively good performance. The graph produced from the Decision Tree classifier is depicted in Figure 3. A baseline random labeling model (without stratification) was also applied for comparison purposes.

	Random	Decision Trees	SVM (linear)	kNN	RF
Accuracy	46.2%	65.4%	**80.8%**	77%	73.1%
Precision	48.2%	53.7%	**72.8%**	38.5%	63.5%
Recall	47.5%	54.2%	**70%**	50%	65%
F1	43.1%	53.8%	**71.2%**	43.5%	64.1%
F1-w	50.2%	66.3%	**80.1%**	66.9%	73.8%

Table 1: Performance of different classifiers in predicting MDD cases from gameplay data.

4 CONCLUSION

In this paper, we developed a serious game prototype system called "The Delivery" for the diagnosis of Major Depressive Disorder (MDD). This video game, which features Artificial Intelligence methods, was designed with special care taken when implementing in-game mechanics, so as to correspond to official diagnostic criteria of the particular disorder. After recording gameplay sessions of players who were diagnosed with MDD and players who were highly unlikely to have MDD, we were able to train Machine Learning models that showed promising performance in distinguishing positive from negative MDD cases in gameplay behaviors.

Our proposed method is a novel concept, since no other work has touched the subject of MDD diagnosis in such perspective. It is crucial that new, more accurate and efficient techniques are developed, which are better suited to current and future societal needs. Additionally, this prototype can be further improved by extending the current scenario and gathering more data, which will lead to higher diagnostic accuracy of the system. A larger amount of gameplay data will also give us the opportunity to try other Machine Learning methods as well (e.g. Neural Networks). Moreover, this system can be adapted accordingly, for the purpose of identifying other types of depression, or mental health issues in general. Lastly, the video game industry is an ideal stakeholder for incorporating such diagnostic methodologies in video games and services.

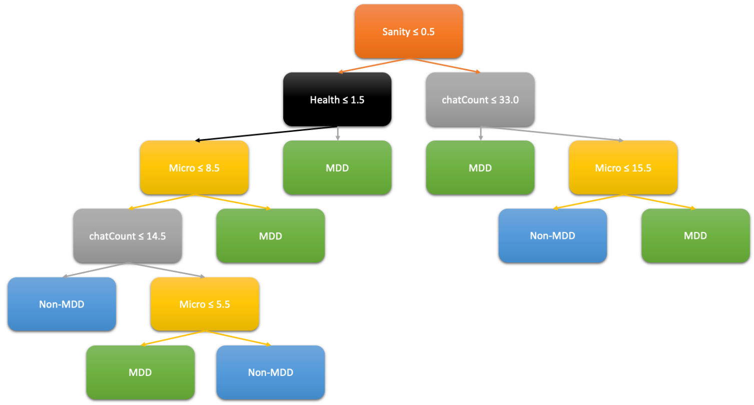

Figure 3: Decision Tree classifier for gameplay data. The root node (*Sanity*) has a strong role in the final labeling of a player.

REFERENCES

[1] N. S. Altman. 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician* 46, 3 (1992), 175–185. https://doi.org/10.1080/00031305.1992.10475879 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/00031305.1992.10475879

[2] American Psychiatric Association et al. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5®).* American Psychiatric Pub.

[3] Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (01 Oct 2001), 5–32. https://doi.org/10.1023/A:1010933404324

[4] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. 1984. Classification and regression trees. Belmont, CA: Wadsworth. *International Group* 432 (1984), 151–166.

[5] Lucas Pfeiffer Salomão Dias, Jorge Luis Victória Barbosa, and Henrique Damasceno Vianna. 2018. Gamification and serious games in depression care: a systematic mapping study. *Telematics and Informatics* 35, 1 (2018), 213–224.

[6] Theresa M Fleming, Lynda Bavin, Karolina Stasiak, Eve Hermansson-Webb, Sally N Merry, Colleen Cheek, Mathijs Lucassen, Ho Ming Lau, Britta Pollmuller, and Sarah Hetrick. 2017. Serious games and gamification for mental health: current status and promising directions. *Frontiers in psychiatry* 7 (2017), 215.

[7] Theresa M Fleming, Colleen Cheek, Sally N Merry, Hiran Thabrew, Heather Bridgman, Karolina Stasiak, Matthew Shepherd, Yael Perry, and Sarah Hetrick. 2014. Serious games for the treatment or prevention of depression: a systematic review. (2014).

[8] Seymour Geisser. 1993. *Predictive inference.* Vol. 55. CRC press.

[9] Regan Lee Mandryk and Max Valentin Birk. 2019. The potential of game-based digital biomarkers for modeling mental health. *JMIR mental health* 6, 4 (2019), e13485.

[10] Rytis Maskeliūnas, Tomas Blažauskas, and Robertas Damaševičius. 2017. Depression behavior detection model based on participation in serious games. In *International Joint Conference on Rough Sets*. Springer, 423–434.

[11] Alexandros Roniotis and Manolis Tsiknakis. 2017. Detecting depression using voice signal extracted by Chatbots: A feasibility study. In *Interactivity, game creation, design, learning, and innovation.* Springer, 386–392.

[12] Robert L Spitzer, Janet BW Williams, Kurt Kroenke, Raymond Hornyak, Julia McMurray, Patient Health Questionnaire Obstetrics-Gynecology Study Group, et al. 2000. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *American journal of obstetrics and gynecology* 183, 3 (2000), 759–769.

[13] Michael E Tipping and Christopher M Bishop. 1999. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61, 3 (1999), 611–622.

[14] Michel F. Valstar, Jonathan Gratch, Björn W. Schuller, Fabien Ringeval, Denis Lalanne, Mercedes Torres, Stefan Scherer, Giota Stratou, Roddy Cowie, and Maja Pantic. 2016. AVEC 2016 - Depression, Mood, and Emotion Recognition Workshop and Challenge. *CoRR* abs/1605.01600 (2016). arXiv:1605.01600 http://arxiv.org/abs/1605.01600

[15] Vladimir Naumovich Vapnik. 2000. *The Nature of Statistical Learning Theory, Second Edition.* Springer.