**E-Health Methods and Applications: Report Part 2**

Contents

[***Building Dataset with Pubmed Papers*** 1](#_Toc86402646)

[***Reliability of Pubmed Papers*** 2](#_Toc86402647)

[***Building a Dataset of Similar Validated Applications per Non-Validated Serious Game*** 2](#_Toc86402648)

[***Benchmark for Game Validation*** 2](#_Toc86402650)

[***Benchmark Results*** 3](#_Toc86402651)

[***Dashboard*** 3](#_Toc86402652)

## ***Building Dataset with Pubmed Papers***

Starting from the dataset of serious games selected by our filters, we built a dataset considering only app name, learning category and associated papers found on Pubmed through Pymed library. Our purpose is to clinically evaluate the serious games of our dataset on the basis of scientific publications.

Moreover, we choose the learning category as a feature that can subsequently relate clinically validated games with serious ones that do not have scientific validation on Pubmed.

## ***Reliability of Pubmed Papers***

For each paper on our dataset, we extracted, among others, the title, the abstract, the methods, conclusions, results and associated keywords. In order to measure the robustness and reliability of these papers, we used the Natural Language Processing method.

Indeed, we searched, in the papers information, for keywords related to specific study types such as observational studies, systematic reviews, meta-analyses... Then, we assigned a score to each paper according to the number of keywords found and their associated study types. Accordingly, keywords associated to meta-analyses, systematic reviews and RCTs give a higher score than keywords linked to observational studies.

Once the scores are computed, we defined a validation threshold to determine whether or not a game is considered as “validated”. This threshold coincides with the degree of reliability of an observational study which we have choosen as a comparator for reliability.

## ***Building a Dataset of Similar Validated Applications per Non-Validated Serious Game***

Our goal was to provide any user interested in a non-validated serious game the possibility to see other similar applications that are actually validated. Therefore we implemented a function that relates each non-validated application with a short list of validated serious games from the same learning category. This way, the user will have a general idea of the level of validation of each learning field.

***Benchmark for Game Validation***

We want to compare the clinical validation of a serious game found by our algorithm with the one found by a human being. To begin with, we found 34 mobile games mentioned on PubMed:

* 29 serious games judged as clinically validated by the members of the group
* 5 serious games judged as not clinically validated by the members of the group because of the lack of reliable information on them.

This set of serious games was used to benchmark the ability of our algorithm to validate the scientific evidence on them. We based the evaluation on the following statistical parameters:

* Accuracy: the ability to distinguish a validated serious game from a non-validated

one;

* Sensitivity: the ability to recognize a validated serious game;
* Specificity: the ability to recognize a non-validated serious game;

## ***Benchmark for Game Validation results***

Accuracy:

Sensitivity:

Specificity:

## ***Dashboard***

In order to display all the above-mentioned information, we chose to create a dashboard, using the Python framework Dash. This dashboard meets different needs by answering the following questions:

* Given a learning category and an age range, what are the corresponding serious games available on the market?
* Given a specific application: what are the main details about it? How many papers are associated to this application? Even if it has been the subject of clinical studies, what is its level of validation and can we consider it as “validated”?
* Given a specific paper: what is the main information about it? What is its level of validation, according to the type of study it refers to?
* Given a specific application whose level of validation is insufficient: is there any similar scientifically validated application? If yes, what are their names?

This dashboard is divided in four sections (tabs): field overview, details per application, papers per application and similar applications per non-validated applications.

The first tab allows the user to have a visual summary of the number of applications per learning category and age range, thanks to bar graphs. Then, using the dropdown menus, he’s able to select a specific learning category and age range to display the associated list of serious games. This list contains general information about these applications, such as their rating, the number of ratings, their price…

The second tab enables him to display more detailed information about all the applications: he can read their description, the number of installs, the date of the last update… By selecting one specifically, he can see the list of associated papers and their information. He’s also able to visualize its level of validation: if it’s green, it means that the validation score is above the defined threshold and the app is considered as “validated”. On the contrary, the level of validation will be displayed in red.

The third tab permits him to visualize the number of papers per application, thanks to a colored pie chart. He’s then able to select one specific application and the title of one of its associated papers to display information about it.

Finally, as several applications were not the subjects of any scientific publication in PubMed, we wanted to give to the user the possibility to see the names of similar validated applications. This way, he can search for these applications in PubMed and find papers that can permit him to assess the level of validation of the learning category he’s interested in.