



## Towards Rapid Quality Reviews for Digital Health Apps: A Machine Learning Approach

Hyzy, M., Mulvenna, M., Bond, RR., Bai, L., Daly, R., & Leigh, S. (in press). Towards Rapid Quality Reviews for Digital Health Apps: A Machine Learning Approach. In *Towards Rapid Quality Reviews for Digital Health Apps: A Machine Learning Approach*

[Link to publication record in Ulster University Research Portal](#)

### Published in:

Towards Rapid Quality Reviews for Digital Health Apps: A Machine Learning Approach

### Publication Status:

Accepted/In press: 09/03/2023

### Document Version

Author Accepted version

### General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [pure-support@ulster.ac.uk](mailto:pure-support@ulster.ac.uk).

# Towards Rapid Quality Reviews for Digital Health Apps: A Machine Learning Approach

MACIEJ HYZY, Ulster University, UK

RAYMOND BOND, Ulster University, UK

MAURICE D. MULVENNA, Ulster University, UK

LU BAI, Ulster University, UK

ROBERT DALY, ORCHA, UK

SIMON LEIGH, ORCHA, UK and Warwick University, UK

The Organisation for the Review of Care and Health Apps (ORCHA) is a UK based digital health compliance company. This study examines the data set of more than 2000 digital health apps assessed with ORCHA assessment tool. ORCHA assessment tool consists of approximately 300 questions and is used to assess digital health apps for quality (defined as "compliance with best practice standards"). The objective of the study was to select a small number (1-5) of important questions from the tool and determine if the app will pass or fail assessment. Accuracy was 76% with logistic regression and 78% with random forest.

CCS Concepts: • **Human-centered computing**; • **General and reference**; • **Applied computing** → **Health informatics**;

Additional Key Words and Phrases: Digital health, mHealth, Quality assessment, Machine learning

## ACM Reference Format:

Maciej Hyzy, Raymond Bond, Maurice D. Mulvenna, Lu Bai, Robert Daly, and Simon Leigh. 2023. Towards Rapid Quality Reviews for Digital Health Apps: A Machine Learning Approach. In . ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/nmnnnnn.nmnnnnn>

## 1 INTRODUCTION

This work was done in partnership with UK based digital health compliance company 'The Organisation for the Review of Care and Health Apps' (ORCHA). In order to thoroughly assess digital health apps, ORCHA uses approximately 300 questions in its ORCHA Baseline Review (OBR) assessment tool. A data set of the assessment of 2053 digital health apps with OBR was provided by ORCHA to aid this research. The data set used included digital health app assessments that were published between 18th January 2021 and 6th January 2022.

Each assessment has been carried out by at least 2 trained reviewers, where in the case of a dispute, a third reviewer would resolve it. All reviewers have undergone the same training to use the OBR assessment tool. A pre-assessment review determining whether a digital health app is likely to pass or fail an ORCHA quality assessment could be useful to developers who seek to have their app assessed. This would allow them to better prepare for the actual ORCHA quality assessment. The objective of this study was to analyse the provided data set and determine if it's possible to predict whether a digital health app will pass or fail ORCHA assessment with good accuracy with up to 5 questions from the ORCHA assessment tool OBR.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2023 Association for Computing Machinery.

Manuscript submitted to ACM

2 METHODS

2.1 The data set and analysis

R language and R studio has been used for machine learning and generation of figures for data visualisation. 2053 of digital health apps have been assessed using ORCHA assessment tool. The score of 65 is a starting point, based on answers this score will increase, decrease, or stay the same. Each assessment has been carried out by at least two trained reviewers, where in the case of a dispute, a third reviewer would resolve it. All reviewers have undergone the same training to use the assessment tool. The assessments were carried out between 18th January 2021 and 6th January 2022.

The questions that best predict the ORCHA score has been determined using 4 methods: boruta [2], random forest [3] IncMSE and incNodePurity, and logistic regression [1] odds ratios.

2.2 Ethical approval

This secondary data analysis study gained ethical approval by Ulster University (ethics filter committee, Faculty of Computing, Engineering and the Built Environment). The process undertaken by ORCHA ensures that digital health apps’ developers are aware of their score and are given time to contest findings of the assessment which may be amended if developers provide additional relevant information. All reviews, unless explicitly asked to be removed by the developer, are covered as suitable for research in ORCHA’s privacy policy [5].

3 RESULTS

Figure 1 depicts a graphical presentation of the question importance with box-plots when using boruta method. In total 63 of the ORCHA fundamental questions (questions that were answered for all the digital health apps in the assessment) ere used in the feature extraction process.

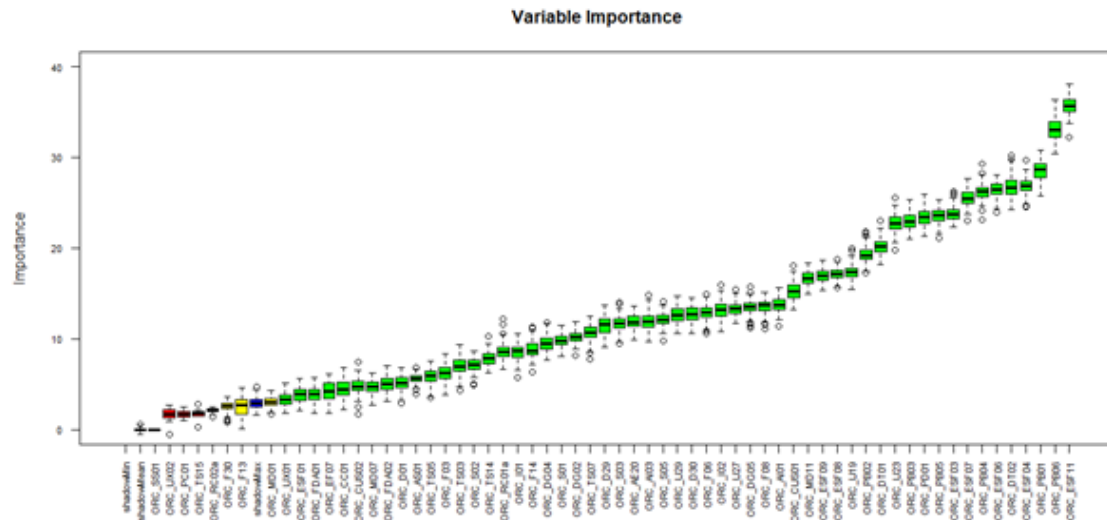


Fig. 1. Box-plots depicting question importance using boruta method.

Table 1 depicts how different methodologies for feature extraction ranked importance of different ORCHA questions (depicted by Question code). The decision column depicts the question and their corresponding importance that has been determined by examining the results of all the feature extraction methods used in this study.

Table 1. Question importance by question code

| Importance | Boruta    | IncMSE    | IncNodePurity | Odds ratio | Decision  | Correlation <sup>1</sup> | P-value <sup>1</sup> |
|------------|-----------|-----------|---------------|------------|-----------|--------------------------|----------------------|
| 1          | ORC_ESF11 | ORC_PB06  | ORC_ESF11     | ORC_ESF11  | ORC_ESF11 | .761                     | P<.001               |
| 2          | ORC_PB06  | ORC_ESF11 | ORC_ESF06     | ORC_RC02a  | ORC_PB06  | .342                     | P<.001               |
| 3          | ORC_PB01  | ORC_DT02  | ORC_ESF07     | ORC_ESF01  | ORC_PB01  | .516                     | P<.001               |
| 4          | ORC_ESF04 | ORC_ESF03 | ORC_PB01      | ORC_F30    | ORC_DT02  | .405                     | P<.001               |
| 5          | ORC_DT02  | ORC_PB01  | ORC_ESF08     | ORC_UX02   | ORC_ESF06 | .671                     | P<.001               |

Table 2 depicts ranked question importance with question code and question text. This table shows what were the 5 most important questions when considering 4 different feature extraction methods.

Table 2. Question code and question text

| Importance | Question code | Question text  |
|------------|---------------|--|
| 1          | ORC_ESF11     | Does the app have appropriate evidence for the ESF tier?   |
| 2          | ORC_PB06      | Is there evidence within the app that the developer has validated any guidance with relevant reliable information sources or references? |
| 3          | ORC_PB01      | Is there a suitably qualified Professional involved in the Development team of the App?  |
| 4          | ORC_DT02      | Is there a statement either in the app or store about user involvement in testing?   |
| 5          | ORC_ESF06     | Has the app met Tier 1 minimum requirements?   |

Using logistic regression and random forest with 80/20 split for training and testing revealed that question "Does the app have appropriate evidence for the ESF tier?"(Question code: ORC\_ESF11), is the most relevant question in predicting whether an app will pass or fail ORCHA quality assessment. For training and testing, ORCHA score  $\geq 65$  as been denoted as 1 (meaning a pass) and ORCHA  $< 65$  (meaning a fail), has been denoted as 0. Accuracy of 89% has been achieved with both algorithms used in the study. Increasing the number of questions to include the next 4 most important ranked questions did not increase the accuracy with logistic regression and only increased by .25 of a percent with random forest. When excluding ESF tier questions, 76% accuracy has been achieved with logistic regression using only ORC\_PB01, ORC\_PB06 and ORC\_DT02. And 78% with random forest.

<sup>01</sup>Decision question's correlation and its P-value with ORCHA score (ORCHA score  $\geq 65$  as been denoted as 1 (meaning a pass) and ORCHA  $< 65$  (meaning a fail), has been denoted as 0).

## 4 DISCUSSION

This study revealed that it is possible to predict whether a digital health app will pass or fail ORCHA quality assessment with 89% accuracy just by answering one question "Does the app have appropriate evidence for the ESF tier?" (Question code: ORC\_ESF11), where 'ESF' refers to National Institute for Health and Care Excellence's (NICE) 'evidence standard framework' [4]. However, ORC\_ESF11 and ORC\_ESF06 would be expected to be good predictors of whether a health app passes or fails quality assessment, since they are about compliance with national guidelines set by health professionals. Questions ORC\_PB01, ORC\_PB06 and ORC\_DT02 achieved 76% accuracy with logistic regression and 78% with random forest. These questions are more descriptive on what is needed to pass quality assessment.

The analysis conducted in this study shows that the advantage of using a short review can allow for a good estimation if a digital health app will pass or fail an assessment. This can be a useful pre-assessment questions that can be used by digital health apps developers to increase their chances of passing the ORCHA quality assessment. The benefit of a long, more rigorous, 300 question ORCHA review, allows for assignment of a score from 0 to 100. ORCHA assessment consists of 4 section: scene setters, professional/clinical assurance, data privacy and usability and accessibility. Long review can identify drawbacks of the digital health app assessed in regards to professional/clinical assurance, data privacy and usability and accessibility. This does provide developers with greater feedback on how the digital health app can be improved.

### 4.1 Limitations

A pre-assessment review with three question can only predict if the digital health app will pass or fail an assessment, it will not predict the score from 0 to 100 and it has 76% accuracy with logistic regression and 78% with random forest.

## 5 ACKNOWLEDGMENTS

This research is done in partnership with ORCHA, a UK-based digital health compliance company. This work is supported by a Northern Ireland DfE CAST award / PhD scholarship. We would like to acknowledge the contribution of the many digital health app reviewers and developers who worked with ORCHA that allowed for the review of digital health app and consented for their data to be used for the purposes of research.

## REFERENCES

- [1] R Documentation. 2022. glm: Generalized Linear Models (R-compliant). (2022). Retrieved January 31, 2023 from <https://www.rdocumentation.org/packages/SparkR/versions/2.1.2/topics/glm>
- [2] Miron Bartosz Kursa. 2022. Package 'Boruta'. (2022). Retrieved January 31, 2023 from <https://cran.rproject.org/web/packages/Boruta/Boruta.pdf>
- [3] Andy Liaw and Matthew Wiener. 2022. Package 'randomForest'. (2022). Retrieved January 31, 2023 from <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
- [4] NICE. 2022. *Evidence standards framework for digital health technologies*. Retrieved February 3, 2023 from <https://www.nice.org.uk/corporate/ecd7>
- [5] ORCHA. 2023. *Privacy policy*. Retrieved January 31, 2023 from <https://appfinder.orchaco.uk/privacy-policy/>