Check for updates

SOFTWARE TOOL ARTICLE

Data analytics for research on complex brain disorders

[version 1; peer review: 2 approved with reservations]

Michail Kontoulis, George Doukas, Theodosios Pountridis, Loukas Ilias, George Ladikos, Vaggelis Lamrpou, Kostantinos Alexakis, Dimitris Askounis, Christos Ntanos

Decision Support System Laboratory, National Technical University of Athens, Zografou, Attica, 157 80, Greece

v1 **First published:** 08 Nov 2024, **4**:246
https://doi.org/10.12688/openreseurope.18505.1
Latest published: 08 Nov 2024, **4**:246
https://doi.org/10.12688/openreseurope.18505.1

Abstract

There exist many research tools for complex brain disorders both proprietary and open source. Those tools are mostly focused on specific use cases and don't offer a holistic view of the research and diagnostic processes. The MES-CoBraD platform offers a holistic tool that participates in all the phases of the process from planning to execution. In it researchers can create workflows of the diagnostic and research procedures. The Data Analytics module provides the functionality described in these workflows, mainly the analytical functions. The tools are integrated with the rest of the platform in a seamless manner. They are capable of analyzing different categories of data including imaging data, electroencephalogram (EEG)/polysomnography (PSG), medical device data, and consumer technology data. In this paper we present a summary of how the MES-CoBraD platform operates, the role of the data analytics in it, the actual functions implemented and integrated in the Data Analytics module and a number of use cases which show how each specific function can be used during the research and clinical procedures.

Keywords

Analytics, MRI, EEG, Actigraphy, Artificial Intelligence, Statistical Analysis, Researchers, Clinicians

H2020 This article is included in the Horizon 2020 gateway.

Open Peer Review

Approval Status ? ?

	1	2
version 1 08 Nov 2024	? view	? view

1. **Ali Husnain**, Chicago State University, Chicago, USA

2. **Yu-Fang Yang**, Freie Universität Berlin, Berlin, Germany

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Michail Kontoulis (mkontoulis@epu.ntua.gr)

Author roles: Kontoulis M: Conceptualization, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Doukas G**: Conceptualization, Investigation, Methodology, Project Administration, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation; **Pountridis T**: Investigation, Software, Validation, Visualization, Writing – Original Draft Preparation; **Ilias L**: Conceptualization, Data Curation, Investigation, Methodology, Software, Validation, Writing – Original Draft Preparation; **Ladikos G**: Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation; **Lamrpou V**: Data Curation, Investigation, Methodology, Software, Validation; **Alexakis K**: Methodology, Software; **Askounis D**: Conceptualization, Funding Acquisition, Project Administration; **Ntanos C**: Conceptualization, Funding Acquisition, Project Administration, Supervision

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the [European Union's Horizon 2020 research and innovation programme under grant agreement No 965422(Multidisciplinary Expert System for the Assessment & Management of Complex Brain Disorders [MES-CoBraD]).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Kontoulis M *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Kontoulis M, Doukas G, Pountridis T *et al.* **Data analytics for research on complex brain disorders [version 1; peer review: 2 approved with reservations]** Open Research Europe 2024, **4**:246 https://doi.org/10.12688/openreseurope.18505.1

First published: 08 Nov 2024, **4**:246 https://doi.org/10.12688/openreseurope.18505.1

Introduction

The Data Analytics tools of the Multidisciplinary Expert System for the Assessment & Management of Complex Brain Disorders (MES-CoBraD) is a module aiming towards providing a versatile toolkit that will support and expand the capabilities of researchers and clinicians. They are mostly aimed towards the analysis and handling of, imaging data, electroencephalogram (EEG)/polysomnography (PSG), medical device data, and consumer technology data.

There already exists a number of tools that can be utilised when working with this type of data, (Ilias *et al.*, 2023a) but those tools can often be difficult to install, the user may lack the necessary hardware or software and they don't offer a holistic view of the process. The MES-CoBraD platform in general aims to provide researchers and clinicians with semi-automated workflows of the diagnostic and research proceadures they follow during their work, from the acquisition of data to the displaying of results. This holistic funcitonality is not available in most tools which focus on specific functions. The Data Analytics module in particular offers the actual implementation of those functions, in the integrated environment of MES-CoBraD.

Methods

Operation

The data analytics tool requires a version of Docker that supports docker compose version 3.7. Other system requirements depend on the size of data that will be uploaded and the modules used. For most modules a system with 8gb of ram should suffice. For the AI module, a system with at least 64gb and a powerful GPU is recommended. All files can be accessed for download at https://github.com/mescobrad-project/data-analytics-and-visualisation-backend/tree/dev and https://github.com/mescobrad-project/data-analytics-and-visualisation-frontend.

The operation flow consists of two versions depending on how the application is used integrated with the rest of the MES-CoBraD platform or in Standalone mode.

In the first case the user, i) uploads data to the platform, ii) creates workflows, iii) run each workflow selecting data for each step, iv) executes the analytics function in the application, repeat iii – iv until workflow complete

In the second case the user, i) Creates folders with data for each function, ii) Accesses a specific function from the main page of the application, iii) executes the analytics function

Implementation

The MES-CoBraD platform purpose is to offer an environment that gives clinicians and researchers the capability to create workflows that describe in detail treatment and research methods. These workflows are composed of different steps such as acquiring data, editing data, displaying results, changing the flow based on the results, and finally performing various analytics functions. These functions are focused to the needs and wants of the users and will be focused on handling and analysing neurological data and general statistical analysis. (Figure 1)

Most functions of the analytics modules need some parameterisation which can be done through the provided UI. The functions themselves and their accompanying results are provided to the platform both through a user interface and through an API where the functions can be run semi-automatically after being parameterised once. This secondary semi-automated functionality can be used to run numerous tests at the same time without constant user input.

For the MES-CoBraD platform to fulfil its goal of aiding specifically with complex brain disorders, the analytics module should offer these things.

a) General analytics functions that are commonly used either in research or in clinical workflows. Each function should be deconstructed into the smallest parts so they can be freely combined and ordered in the workflow manager.

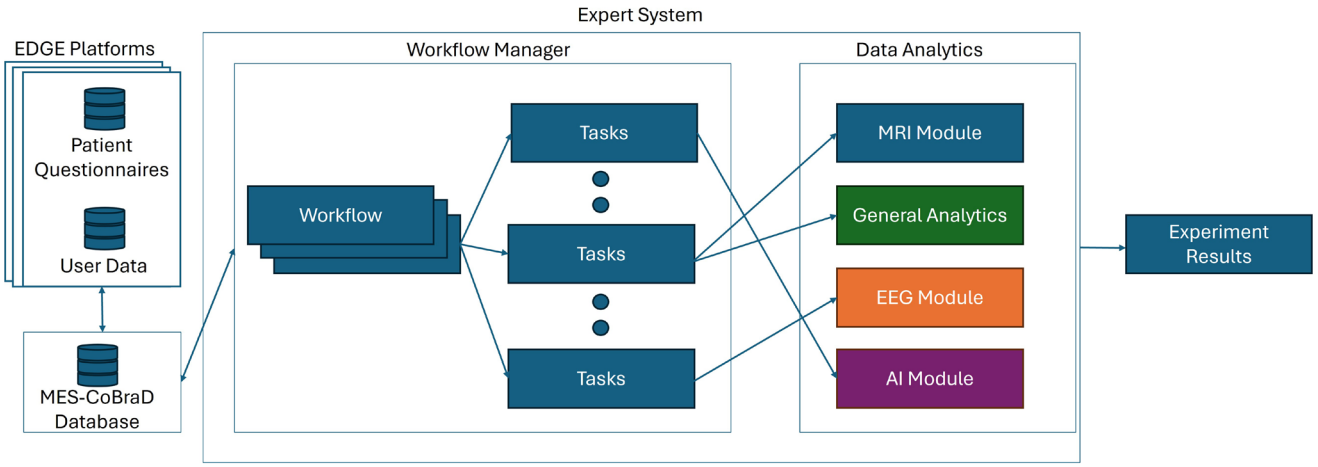

Figure 1. Simple summary of MES-CoBraD platform functionality.

b) Analytics functions of EEG signals. Each function should be deconstructed into the smallest parts so they can be freely combined and ordered in the workflow manager.

c) Analytics functions of MRI signals. Each function should be deconstructed into the smallest parts so they can be freely combined and ordered in the workflow manager.

d) Complex functions that combine multiple steps, that always need to be completed together or are aimed at more specific analytics methods.

Standalone mode. The application can also be used without the rest of the MES-CoBraD platform on its own. The users can create folders with their data manually in their system and access the function pages from the main page of the application. More details can be found here https://github.com/mescobrad-project/data-analytics-and-visualisation-backend/blob/dev/README.md.

Choosing which functions to implement. In the beginning we needed a method to decide on which functions to include in the module and how to deconstruct them so they can be used along the workflow manager. There is significant number of tools, methods, functions and not all of them are used equally nor the offer the same value in the field of neurophysiology. Additionally, there is a finite amount of time and budget allocated to the project which furthers enhances the need for choosing and implementing only the most relevant and important functionality.

A preliminary analysis of available tools was completed in a previous paper. (Ilias *et al.*, 2023b) In addition we took advantage of the participation of scientific partners in the MES-CoBraD project and requested from each partner for a list of the analytics functions they use most often either during research or clinical methods. Specifically, we focused on the type of data that will be available in the platform and how they are utilised. We proceeded to decompose some of the provided functions to smaller discrete steps while deciding to keep other more complex ones, depending on how they are used. The initial list of functions and potential use cases is presented below:

Data analytics available functions

Due to the high number of different functions offered by the project, they have been categorised based on the type of data they are using as input.

EEG analysis

Functions in these sections are focused on EEG related functions. Input for most functions is one or more files in EDF Format.

Table 1. EEG – Timeseries analysis implemented functions and use cases (Extended data)

MRI analysis

Table 2. MRI analysis implemented functions and use cases (Extended data)

Actigraphy Analysis

Table 3. Actigraphyi analysis implemented function and use cases (Extended data)

Statistical Analysis

Table 4. Statistical analysis implemented function and use cases (Extended data)

Artificial Intelligence Analysis

Table 5. Artificial Intelligence analysis implemented function and use cases (Extended data)

Use cases

A number of specific use cases have already been described in the previous chapters for each function of the analytics module. But these use cases are usually only a part of the whole process and do not give an accurate image of the aims of the module. The general use case that describes the whole process can be found in the pilot use cases of the MES-CoBraD project and specifically the use case:

PUC 2: MES-CoBraD Advanced Analytics Deep-Phenotyping and Prognostication

This use case describes in detail the real-world data related that need to be analysed by the Data Analytics modules and provides several KPIs as guide and measure for the success of the module that related to the percent of real world data analysed, number of analytic functions, and comparison of results of other tools with the ones provides the Data Analytics module.

Discussion/conclusions

The MES-CoBraD Data Analytics module incorporates many functionalities from different fields to accomplish the following:

1) Ease of installation for users

Not every medical researcher has the necessary technical knowledge or the available hardware for executing these applications. The module provides all these functions ready to use through a browser window with no other requirements.

2) Extending usability of open-source tools and frameworks

The module extends how open-source tools are used by both providing UI for most of the functions and handling tertiary operations like data handling between the modules.

3) Enhancing reproducibility of research tests

By providing a stable environment and manner for reproducing the functions with the same or different datasets, tests can be rerun and validated again easily.

4) Complete the created workflows of research and diagnostic processes created from other modules of the MES-CoBraD platform.

As previously described, the other MES-CoBraD modules can create workflows of the research and diagnostic processes and the analytics module provides the final step that runs the described functions.

Future steps will lead to two different axis, depth, and breadth. First the increase of the TRL level of the application. The aim of the final TRL is X, which means it's not ready for release to the public. Several improvements need to be made in the UX design. In addition, the created functions can be further elaborated on to either be more specific for the research and diagnostic specific use cases of the users of the platform or more generic in some cases where the developed functions aren't easily used outside the specific use cases around which it was designed. The second one is that the application can be expanded with new functions based on new use cases. Due to the nature of the module most parts of a diagnostic or research procedure can be added as a new function, deciding which ones are needed to be implemented and implementing them is the next step for the module.

Data and software availability statement

Underlying data

No data are associated with this article

Extended data

Title: **Data Analytics for Research on Complex Brain Disorders Extended Data**

Repository: Zenodo

DOI: https://zenodo.org/records/14050475

This project contains:
Data Analytics for Research on Complex Brain Disorders Extented Data.docx

License: Data are available under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0)

Software availability statement

The software described in this article is hosted in two different repositories on Github where the source code along with guides for the setup can be found. The identifiers and links to the repositories are:

Data analytics backend

Source code available from: https://github.com/mescobrad-project/data-analytics-and-visualisation-backend

Archived source code at time of publication: 10.5281/zenodo.11098616

License: MIT License

Data are available under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0)

Data analytics frontend

Source code available from: https://github.com/mescobrad-project/data-analytics-and-visualisation-frontend

Archived source code at time of publication: 10.5281/zenodo.11098676

License: MIT License

Data are available under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0)

References

Ilias L, Askounis D, Psarras J: **Multimodal detection of epilepsy with deep neural networks.** *Expert Syst Appl.* 2023a; **213**: 119010.
Publisher Full Text

Ilias L, Doukas G, Kontoulis M, *et al.*: **Overview of methods and available tools used in complex brain disorders [version 1; peer review: 2 approved].** *Open Res Eur.* 2023b; **3**: 152.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 16 December 2024

https://doi.org/10.21956/openreseurope.20000.r46140

© **2024 Yang Y.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Yu-Fang Yang
Freie Universität Berlin, Berlin, Germany

Summary:
The article introduces the MES-CoBraD platform, a tool aimed at bridging diagnostic and research workflows for complex brain disorders through integrated analytics and reproducible workflows. This is an exciting development as it targets the analysis and handling of imaging data, electroencephalogram (EEG)/polysomnography (PSG), medical device data, and consumer technology data, fostering best scientific practices. Building on prior studies, the platform seeks to combine functionalities across diverse data modalities, including EEG and MRI, while prioritising accessibility and extensibility through open-source frameworks. However, despite its clear potential and ambitious goals, I have concerns about where the platform may fall short, particularly in terms of practical accessibility, workflow clarity, and real-world validation.

#1: The authors stress the importance of ease of installation, which is a fantastic goal for a platform like this. But I couldn't help noticing the reliance on Docker for deployment, which seems at odds with the idea of being user-friendly. Docker is great for reproducibility, but not everyone will be comfortable with command-line tools or manual configurations. I wonder if the authors have thought about alternatives, like a web-based interface or hosting the platform on something like Binder. These options could make the platform far more accessible for non-technical users. Adding step-by-step, beginner-friendly instructions with screenshots could also lower the barrier to entry. These changes would really make a difference for users less familiar with technical setups.

#2: The platform's described functionalities are impressive, but I kept wondering: how does this all work in practice? For example, could the authors include a case study showing the entire process of analysing EEG data, from uploading the raw data to generating diagnostic insights? Such an example would do a lot to bring the platform's capabilities to life. Even a walkthrough of a basic workflow, with a sample dataset, would give users a clearer sense of how the platform handles different data modalities and where its strengths really shine. Without this, it's hard to grasp how effectively it integrates into real-world research or clinical scenarios.

#3: The platform's ability to handle multiple data types, like EEG and MRI, is a real strength, but I struggled to visualise how these workflows differ in practice. Previous studies (Botvinik et al., 2019; Trubutschek et al., 2024) highlight that workflows for analysing EEG and MRI data are often complex and difficult to replicate due to their distinct requirements and modular design. This raises an important question: how does MES-CoBraD integrate these pipelines into its unified platform? How do users create and customise these workflows to ensure both flexibility and reproducibility?

4: I was intrigued by the mention of KPIs, such as the percentage of real-world data analysed and the comparison of results with other tools. These are great ideas, but I felt the paper left me wanting more detail. How are these KPIs being measured or validated? Are there benchmarks or datasets that were used to assess the platform's performance? Including examples of these comparisons—or even some preliminary results—would really help showcase the platform's value. It would also reinforce the scientific credibility of the analytics module.

#5: The authors clearly designed this platform with a wide range of users in mind, which is commendable. But I wonder—has the platform been tested with its intended audience? For example, have clinicians or experimental researchers been involved in pilot studies to evaluate its usability? If so, what kind of feedback was gathered, and how has it shaped the platform's development?

#6: One way MES-CoBraD could strengthen its position is by explicitly comparing itself to existing tools in the field, such as EEGLAB, SPM, FieldTrip, or FSL. While these tools are widely used for EEG or imaging data, they typically focus on specific data types or analysis tasks and often lack the integrated, multi-modality approach MES-CoBraD aims to provide. For example, EEGLAB is excellent for EEG analysis but doesn't handle imaging data, and tools like SPM or FSL are primarily designed for neuroimaging with no support for EEG. A comparison would make it clearer how MES-CoBraD stands out, particularly through its integrated workflows and broader scope.

The MES-CoBraD platform is an exciting step forward for research on complex brain disorders and builds on important prior work. But there are areas where it could be improved to better serve its target audience. Making deployment easier, providing practical demonstrations, clarifying workflows, and incorporating user feedback would bring the platform closer to its ambitious goals. These changes would make the platform not only more usable but also more credible and impactful in the field.

Refer 1 and 2.

References

1. Botvinik-Nezer R, Iwanir R, Holzmeister F, Huber J, et al.: fMRI data of mixed gambles from the Neuroimaging Analysis Replication and Prediction Study.*Sci Data*. 2019; **6** (1): 106 PubMed Abstract | Publisher Full Text
2. Trübutschek D, Yang YF, Gianelli C, Cesnaite E, et al.: EEGManyPipelines: A Large-scale, Grassroots Multi-analyst Study of Electroencephalography Analysis Practices in the Wild.*J Cogn Neurosci*. 2024; **36** (2): 217-224 PubMed Abstract | Publisher Full Text

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?
Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Open Science, EEG, Many Analysts

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 09 December 2024

https://doi.org/10.21956/openreseurope.20000.r47470

© 2024 Husnain A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Ali Husnain
Chicago State University, Chicago, Illinois, USA

The paper introduces the MES-CoBraD platform, designed to aid researchers and clinicians working on complex brain disorders by providing a holistic environment for data analysis and workflow management. The Data Analytics module, central to the platform, integrates analytical functionalities for various data types such as imaging, EEG/PSG, and medical device outputs. It supports seamless integration with workflows, offering ease of installation and usability for users with diverse technical backgrounds. The module also emphasizes reproducibility and extensibility by leveraging open-source frameworks.

Strengths of the Article:

- **Holistic Approach:** The platform fills a critical gap by integrating diagnostic and research workflows into a unified tool.
- **Versatility:** Functions for EEG, MRI, and statistical analyses are described, catering to diverse data modalities.
- **Accessibility:** The tool simplifies installation and usability, addressing challenges faced by non-technical users.

- **Reproducibility:** Emphasis on stable, replicable workflows enhances the scientific rigor of the tool.

Areas for Improvement:

1. **Details for Replication:**
   - While the article mentions the use of Docker and provides GitHub links for code access, the level of detail regarding setup and configuration is insufficient. Explicit instructions on data preparation and deployment processes should be included.
   - Example datasets and workflows should be provided to enable replication and testing of the described functionalities.
2. **Interpretation of Output:**
   - The paper does not elaborate enough on how users can interpret outputs from various analyses. Detailed descriptions, including sample outputs and their relevance to research or clinical decisions, would be beneficial.
   - Metrics or validation methods to assess the accuracy and utility of results generated by the tool should be included.
3. **Technical Soundness:**
   - While technically robust, the article lacks clarity in explaining the underlying algorithms or statistical methods used in key functions. Providing references or supplementary materials for these methodologies would strengthen the technical foundation.
4. **User Feedback:**
   - The platform's functionality could be enhanced by incorporating feedback from potential end-users during its development phase. A section detailing usability testing or pilot studies would add depth to the evaluation.

Recommendations for Revision:

- Include a step-by-step guide for setting up the software, along with test cases and sample data for replication.
- Provide detailed examples of expected output datasets and their interpretation.
- Elaborate on the algorithms and statistical methods underpinning the tool's functions.
- Integrate validation results or user feedback to substantiate the tool’s utility in real-world scenarios.

Conclusion:

The MES-CoBraD platform demonstrates significant potential to advance research on complex brain disorders. Addressing the gaps in replication details and output interpretation will enhance its scientific value and usability.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets

and any results generated using the tool?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Artificial IntelligenceData AnalyticsSoftware EngineeringHealthcare Technology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.