This project has received funding from SMALL GRANT SCHEME Call under grant agreement NOR/SGS/BIPOLAR/0239/2020-00.

Bipolar disorder prediction with sensor-based semi-supervised learning

D4.2 – Initial BIPOLAR package Architecture

Deliverable No.	D4.2	**Due date**	30-SEP-2022
Type	Report	**Dissemination Level**	Public
Version	1.0	**WP**	WP1
Description	Initial architecture of the BIPOLAR software package		

Systems Research Institute Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland

Authors

Name	Email
Katarzyna Kaczmarek-Majer	k.kaczmarek@ibspan.waw.pl
Olga Kamińska	o.kaminska@ibspan.waw.pl
Kamil Kmita	k.kmita@ibspan.waw.pl
Jakub Małecki	jakub.malecki@ibspan.waw.pl

History

Date	Version	Change
12-MAY-2022	0.1	Task assignments and integrated version of the document
18-SEP-2022	0.2	Version for internal review
28-SEP-2022	0.3	Draft ready for submission
30-SEP-2022	1.0	Initial architecture version

This project has received funding from SMALL GRANT SCHEME Call under grant agreement NOR/SGS/BIPOLAR/0239/2020-00.

Executive summary

This deliverable outlines the results of Task 4.1 activities of the BIPOLAR project, which were driven by the BIPOLAR team in cooperation with advisors representing the medical and computer science fields.

To this end, this document describes the initial package architecture and the information identified up to the delivery date of this deliverable are included, but the work will continue to complete and update necessary information throughout the project. Final version including feedback from WP2-WP3 will be delivered on M18 as D4.4 (Final BIPOLAR package Architecture).

This project has received funding from SMALL GRANT SCHEME Call under grant agreement NOR/SGS/BIPOLAR/0239/2020-00.

Table of Contents

This project has received funding from SMALL GRANT SCHEME Call under grant agreement NOR/SGS/BIPOLAR/0239/2020-00.

1.Introduction

In this document, we describe the technical aspects of BIPOLAR package. We concentrate more on how the eventual solution will look in the future since we're at the early phase of development. We're going to describe what we want our code to do and how we will manage it. We see the final product as open-source software so we give a brief explanation how the code will be shared.

2. Overview

Our goal is to create open-source software helping researchers and/or psychiatrists in assessment of uncertainty in their diagnoses. We'd like our software to be easily accessible, free and relatively easy to apply on custom datasets. The software will offer dedicated algorithms for semi-supervised learning as well as visualization tools to make interaction with users smooth and friendly. All these features are described in more detail in sections below.

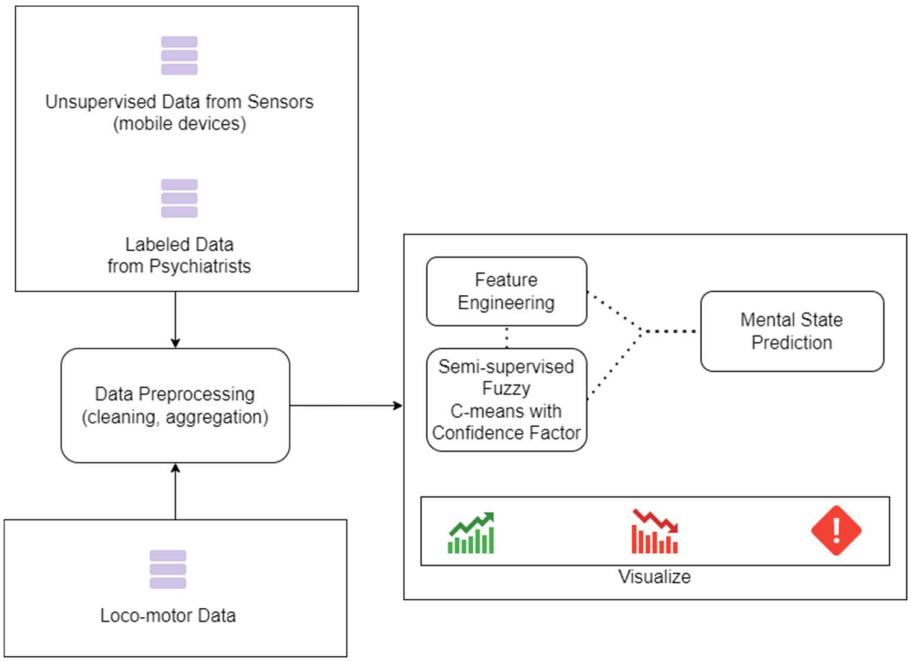

Figure 1 Initial architecture of BIPOLAR software

3. Python/R packages

We use well-known programming languages which such as the two most popular languages in data science, namely Python or R. Users will be able to get the code from GitHub. We're going to have our solution in a form of an R package. This will allow users to install the software easily on their local computers with no (or minimal) effort associated with configuration and dependencies management.

Another part of the solution is an R package demonstrating potential pipeline in data transformation before the proper clustering algorithms can be applied. For this purpose, we use

This project has received funding from SMALL GRANT SCHEME Call under grant agreement NOR/SGS/BIPOLAR/0239/2020-00.

sample mobile recordings data. The dataset contains information about voice parameters of users who agreed on collecting such data on their devices (smartphones). The R code shows potential use case how to check data consistency, clean the dataset and calculate selected statistics. Eventually, the data is in a form that can be further consumed by algorithms provided in the main Python package. As mentioned above, the R package will be maintained on GitHub under the following link: **https://github.com/ITPsychiatry/bipolar**

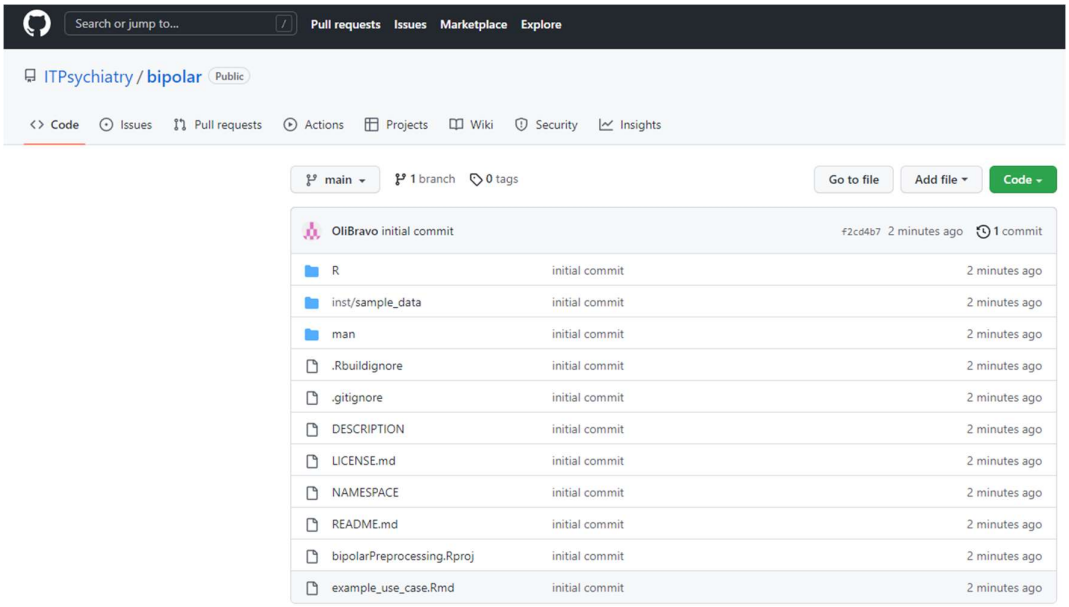

Figure 2 Preprocessing mobile recording data - code repository on GitHub

```
## # A tibble: 0 × 2
## # … with 2 variables: dw_mobilerecording_id <dbl>, chunks_count <dbl>
```

Calculate aggregates

```
sample_call_parameters <- get_sample_call_parameters()

stats <- tryCatch({
  mobile_chunks %>%
    tidyr::nest(data = -dw_mobilerecording_id) %>%
    mutate(stats = lapply(data, function(i) {
      i %>%
        select(all_of(sample_call_parameters)) %>%
        summarise_at(vars(sample_call_parameters), list(mean = mean, sd = sd))
    }))
})

stats <- stats %>%
  select(dw_mobilerecording_id, stats) %>%
  tidyr::unnest(cols = c(stats))  %>%
  select(dw_mobilerecording_id, ends_with("mean"), ends_with("sd"))

stats
```

```
## # A tibble: 10 × 173
##    dw_mobilere…¹ pcm_L…² pcm_z…³ voice…⁴ f0_sm…⁵ f0env…⁶ pcm_f…⁷ pcm_f…⁸ pcm_f…⁹
##            <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1         21987   -20.6   0.222   0.541  159.    189.   1.57e-9 4.04e-7    891.
##  2         21927   -14.0   0.398   0.239    0       0    1.34e-6 1.27e-5   1612.
##  3         21929   -20.5   0.351   0.221    0       0    7.94e-9 1.10e-7   1535.
```

Figure 3 Code fragment for data aggregation applied on sample data

This project has received funding from SMALL GRANT SCHEME Call under grant agreement NOR/SGS/BIPOLAR/0239/2020-00.

The main idea of our software is to give researchers a tool that help them automatically quantify the uncertainty level of diagnosis of their patients given only a small fraction of labeled data. Labeled means a patient has been diagnosed and at that time we were certain about her mental condition. Our software will allow to extrapolate this knowledge on periods where medical diagnosis is unavailable or is obsolete, according to commonly honored rules in psychiatry. As a result, we'll get a precise quantification of how sure we can be when assessing the state of the patient at particular time.

A detailed explanation of the first version of the uncertainty-aware annotation and Dynamic Incremental Semi-Supervised FCM algorithm can be found in the following paper:

Kmita., K., Casalino, G., Castellano, G., Hryniewicz, O., Kaczmarek-Majer, K., (2022) **Confidence path regularization for handling label uncertainty in semi-supervised learning: use case in bipolar disorder monitoring**, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).

Here we only highlight that it's been implemented as our authorial solution. In particular, up to day, there was no other software publicly available offering semi-supervised fuzzy clustering methods taking into account the factor of uncertainty.

Another feature of the package will be the ability to work with locomotor data[1]. Analysis of collected data will help to assess the mental state of a patient.

Users of our software will be able to generate reports presenting data associated with their patients. Thanks to that they will be able to analyze data quickly and effectively. We plan the reports to be accessible in a web browser, so no additional software will be required to display them. Another advantage of this approach is the fact that we can use JavaScript language to make the documents interactive. Precise description on visualizations and layouts will be provided in the future.

3.Future work

Revised version including feedback from WP2-WP3 will be delivered on M18 as D4.4 (Final BIPOLAR package Architecture).

[1] See the psychiatric scenarios LDP-PS4 and LDP-PS5 defined in deliverable D1.1. for details.

This project has received funding from SMALL GRANT SCHEME Call under grant agreement NOR/SGS/BIPOLAR/0239/2020-00.