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

D4.4 – Final BIPOLAR package Architecture

Deliverable No.	D4.4	**Due date**	30-Jun-2023
Type	Report	**Dissemination Level**	Public
Version	1.0	**WP**	all
Description	Final 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
17-JUN-2023	0.1	Task assignments and integrated version of the document
29-JUN-2023	0.2	Version for internal review
30-JUN-2023	1.0	Final 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 describes the final architecture for the BIPOLAR package and is continuation of the Deliverable 4.2. (Initial BIPOLAR package architecture).

This deliverable outlines the results of mainly Task 4.1 activities ("Requirement analysis and software package prototype architecture"). This activities were performed in close collaboration with Task 4.2. ("Development of fuzzy semi-supervised learning models") and Task 4.3 ("Software package implementation and modules integration of the BIPOLAR project"), which were driven by the BIPOLAR team in cooperation with advisors representing the medical and computer science fields.

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 the BIPOLAR package. The main data flow, technological stack as well as features of the software are described in the remaining of this document.

2. Overview

Our **BIPOLAR** software is intended to help 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 offers dedicated algorithms for semi-supervised learning and visualization tools to make interaction with end users smooth and friendly. To the best of our knowledge, this is the first open software package that enables **SEMI-SUPERVISED UNCERTAINTY-AWARE CLUSTERING.**

Main features of the BIPOLAR package are described in more detail in the Figure 1 which gives an overview of the architecture of BIPOLAR software.

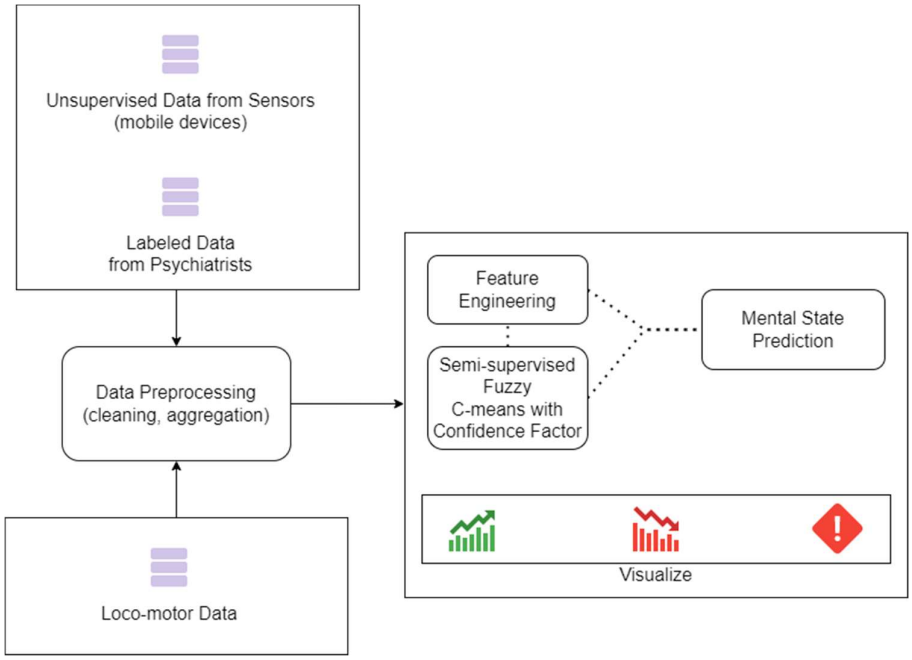

Figure 1 Overview of the architecture of BIPOLAR software

BIPOLAR is comprised of two main components: preprocessing and semi-supervised uncertainty-aware learning that consists of feature engineering and modeling aiming at

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

prediction. Moreover, there is a user-friendly interface for importing datasets and also a set of user-friendly dashboards visualizing the outcomes of the whole pipeline including modeling.

3. R packages

Our software is delivered as R packages. R is one of the most popular programming language used commonly in machine learning, data analytics and its great visualization capabilities make it a language that data scientists and researchers use commonly. It also allows building interactive dashboards easily.

The code is stored on GitHub platform in two repositories:

- https://github.com/ITPsychiatry/bipolar
- https://github.com/ITPsychiatry/ssfclust

The repositories are public meaning end users can download the code or install the packages directly from github. Figure 2 provides an illustrative example of the history of commits to the BIPOLAR repository from github.

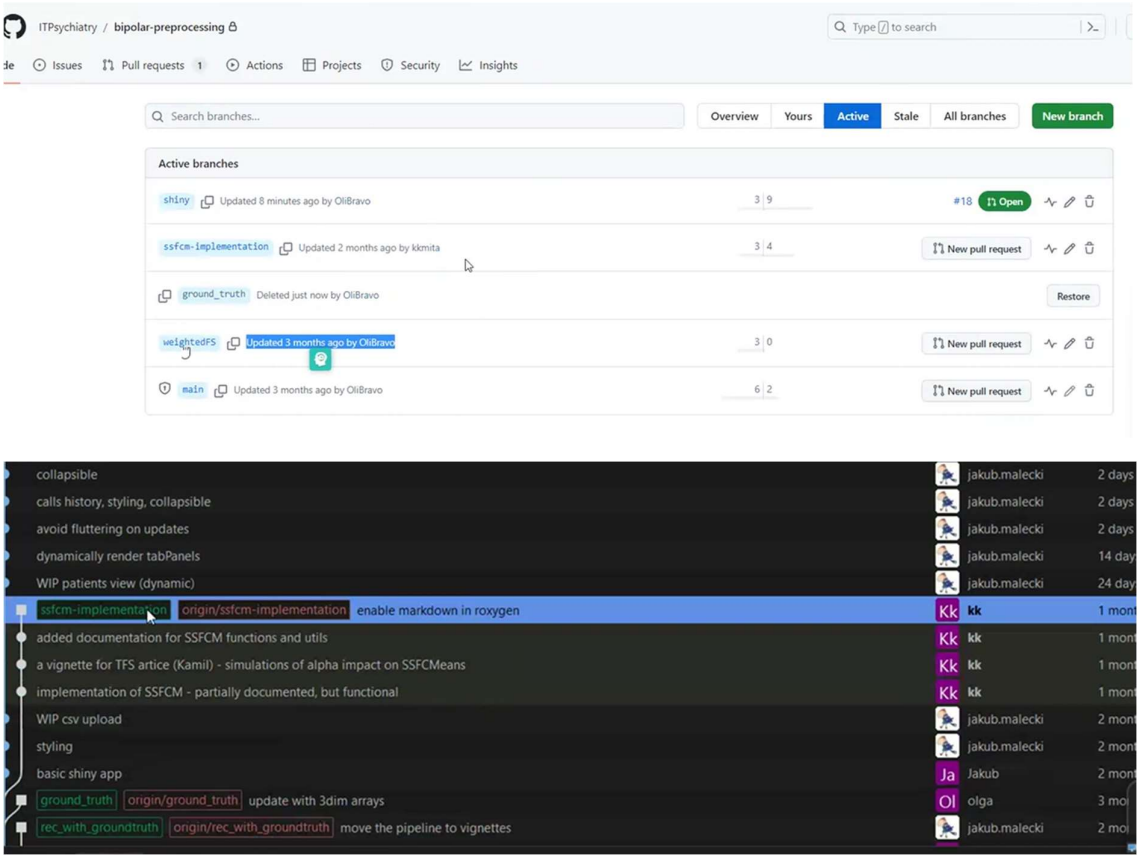

Figure 2 Example of the history of commits to the BIPOLAR repository from github.

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

The **bipolar** repository contains relevant functions for data pre-processing before the modelling algorithms can be applied, see Figure 3 for an example of the preprocessing workflow of the mobile recording data.

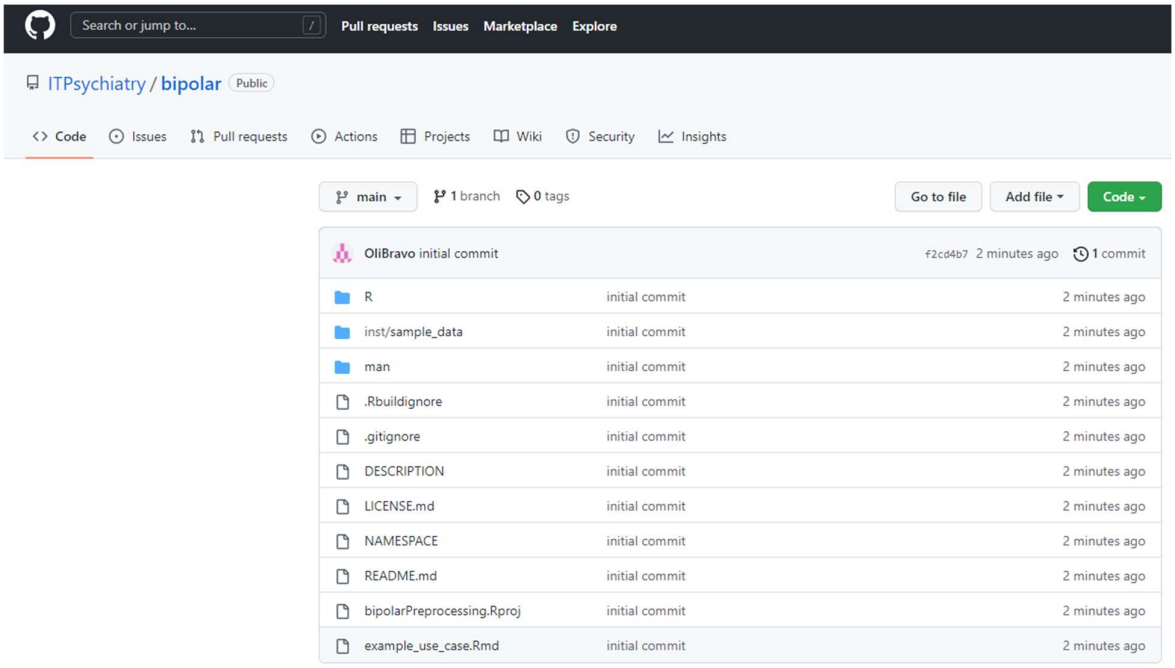

Figure 3 Preprocessing mobile recording data - code repository on GitHub

The second main repository – **ssfclust** – aims at the implementation of SSFCM (semi-supervised fuzzy c-means) algorithm; what is worth noticing, our implementation is a matrix-based one and this distinguishes it from similar packages available.

We give researchers a tool that helps 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 allows for extrapolation of 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 a particular moment in 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:

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

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 today, there's been no other software publicly available offering semi-supervised fuzzy clustering methods taking into account the factor of uncertainty.

Furthermore, apart from the semi-supervised and uncertainty-aware learning, we provide uncertainty-aware feature selection as detailed in the Deliverable (2.6. "Guidelines for selecting features in BD scenarios").

Another feature of the package is the ability to work with locomotor data. Our underlying goal is that this software is able to intelligently combine (uncertainty-aware way) sensor data (either acoustic from smartphone or about activity from locomotor sensors) with the outcomes of the psychiatric assessment (e.g., rating scales assessing the intensity of the depressive or manic symptoms).

Finally, Vignettes help users get started with the package. Figure 4 depicts one of BIPOLAR vignettes presenting selected features to end users.

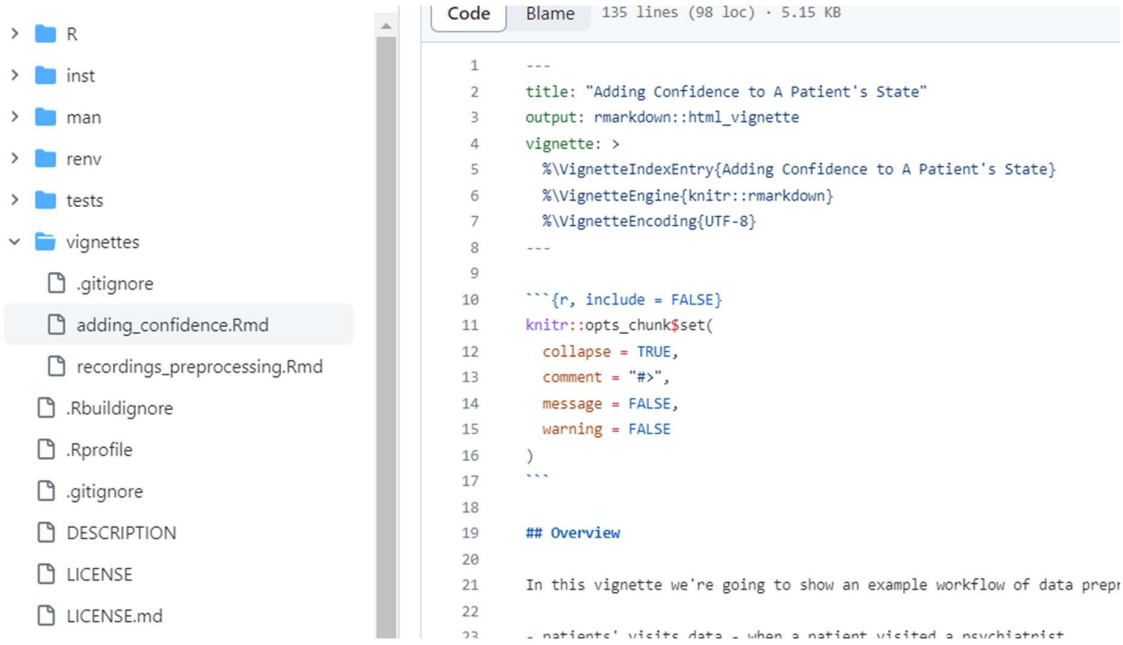

Figure 4 One of BIPOLAR vignettes presenting selected features to end users. Vignettes help users get started with the package.

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

4. Dashboards and Visualizations

Users of our software are able to generate reports presenting data associated with their patients. Thanks to that they are able to analyze data quickly and effectively. The reports are accessible via a web browser, so no additional software is required to display them. We use shiny package in R, so no extra configuration is required on users' side. Figure 5 illustrates data import view in the shiny application from our package.

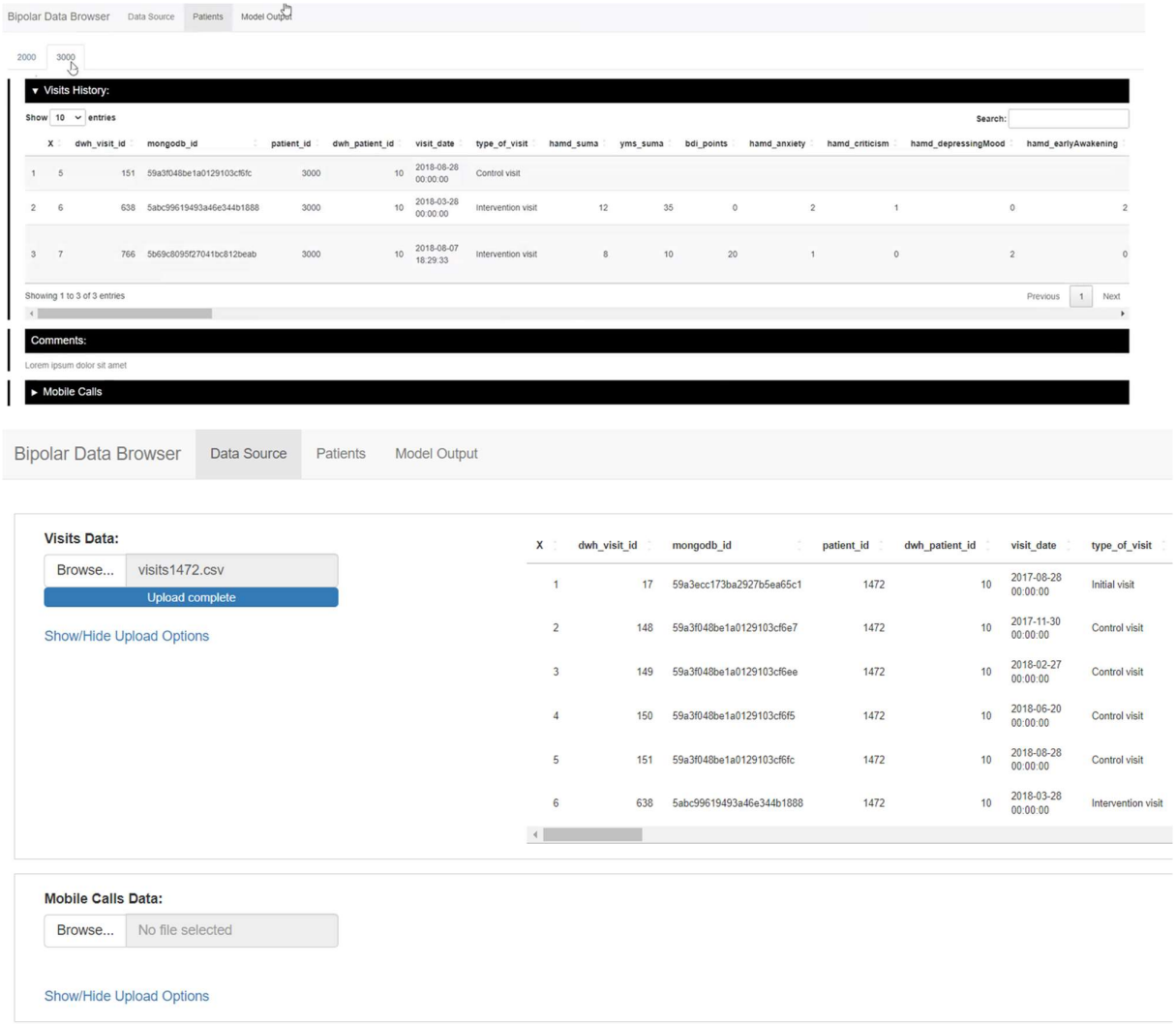

Figure 5 Data import view in the shiny application from our package.

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