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

D1.4 – Initial use case evaluations

Deliverable No.	D1.4	**Due date**	30-DEC-2022
Type	Report	**Dissemination Level**	Public
Version	1.0	**WP**	WP2
Description	Initial evaluation in use cases		

.

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	j.malecki@ibspan.waw.pl
Izabella Zadrożna	izabella.zadrozna@ibspan.waw.pl

History

Date	Version	Change
30-SEP-2022	0.1	Task assignments and integrated version of the document
16-DEC-2022	0.2	Description of datasets added
30-DEC-2022	0.3	Version ready for submission

Executive summary

This deliverable outlines the preliminary results of Task 1.3 activities dedicated to the evaluation of BIPOLAR in use cases. Thus, two scenarios from D1.1. will be tested in initial report with standard metrics, such as e.g., accuracy, recall, AUC. Structured evaluation methods and KPIs as developed in T.1.2. will be added and applied for final evaluation that will be summarized with deliverable D1.5.

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.

List of acronyms

Acronym	Explanation
BIPOLAR	Bipolar disorder prediction with sensor-based semi-supervised learning project
BD	Bipolar disorder

1.Introduction

Prediction of Bipolar disorder state is demanding and not trivial task even for psychiatrist. Having such specific data like acoustic features from mobile recordings and locomotors sensor received during examination we can start with classification of patients' state. For that case, we are using both unsupervised and supervised methods to deal with missing and unlabeled data. For a discussion about the related work, please see deliverable 4.3. with the recent survey of the related approaches.

2.Initial evaluation of Pilot 1: Predict bipolar disorder with Acoustic Data collected from smartphones Pilot (ADP) - ADP-PS2: Semi-supervised prediction of the mental state

The experiments begin with retrieving all available multiple non-stationary data streams representing voice from the real-life dataset [1]. All frames are then aggregated to the patient's phone call level with the mean function. Psychiatric assessments obtained during visits, represented as labels are spread around the day of the patient's visit. Within the feature selection, the top 10 most important voice parameters are selected using labeled dataset. Additionally, all available patient data are clustered to include unlabeled data as well. Simultaneously, the classification of the patient's condition is carried out on the data containing clusters membership and without this information. Finally, results are evaluated with multiple metrics.

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

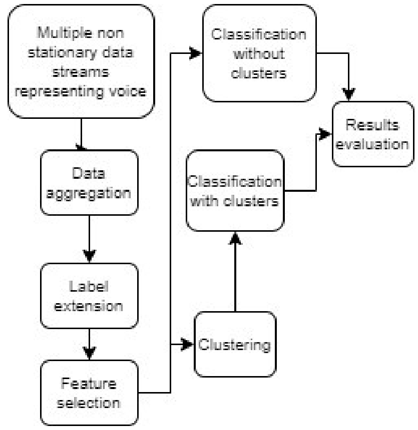

Figure 1. Overview of the selected semi-supervised approach for prediction of bipolar disorder [1].

In the first step of selecting most important features, current markers were chosen as depicted in Table 1.

Table 1. Most important features for dataset of an illustrative patient

"f0final_sma"
"pcm_fftMag_mfcc_8_"
"pcm_fftmag_spectralminpos_sma"
"slope0-500_sma3"
"pcm_fftMag_mfcc_6_"
"pcm_fftMag_mfcc_10_"
"pcm_fftMag_mfcc_9_"
"slope500-1500_sma3"
"pcm_fftmag_spectralharmonicity_sma_compare"
"pcm_fftMag_mfcc_2_"

Figures 2-6 present clusters' membership values received by Fuzzy C-mean algorithm used on complete dataset (including labeled and unlabeled data). On the y-axis, we can observe patients state, colour of points indicates the actual patients BD state. Observations for all of the selected patients were assigned to more than 2 different clusters for their BD state. Unfortunately, there is plenty of missing data, that makes such analysis challenging.

Let us now consider an example of patient with ID 1472. This patient had 3 visits (labelled data) with a doctor. First visit that took place in March 2018 is surrounded by mobile calls grouped to all of available clusters, next two visits are surrounded by mobile calls only by 2 clusters. Thus, the mixed state which was assigned by the doctor as a BD state received 3 different clusters (called respectively as number one, two and three). BD state called euthymia was assigned with 2 different clusters. In the ideal scenario, we could receive for each available BD state a different number of clusters. However, since the psychiatric states naturally overlap (for example the mixed state is a mix of depressive and manic symptoms), such ideal scenario is hardly observed for real data.

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

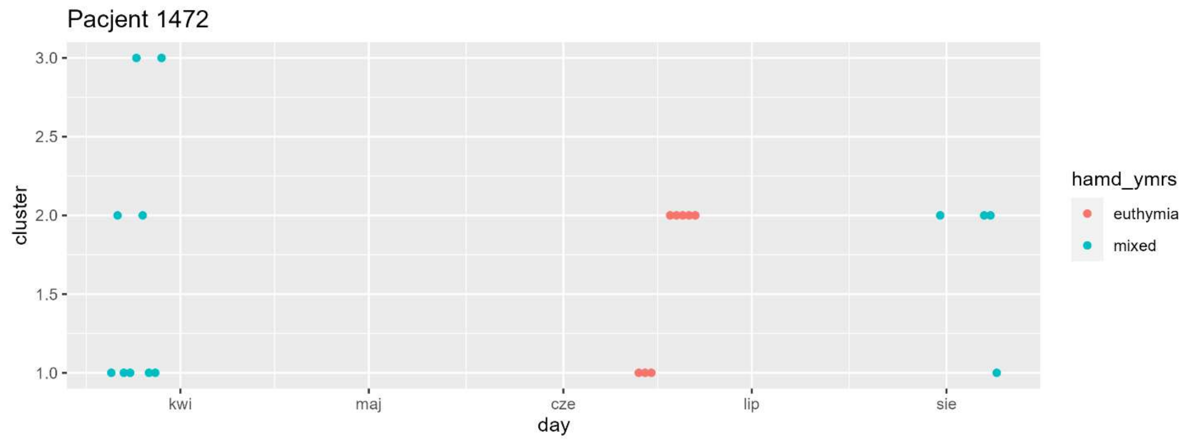

Figure 2. Clusters membership for patient 1472

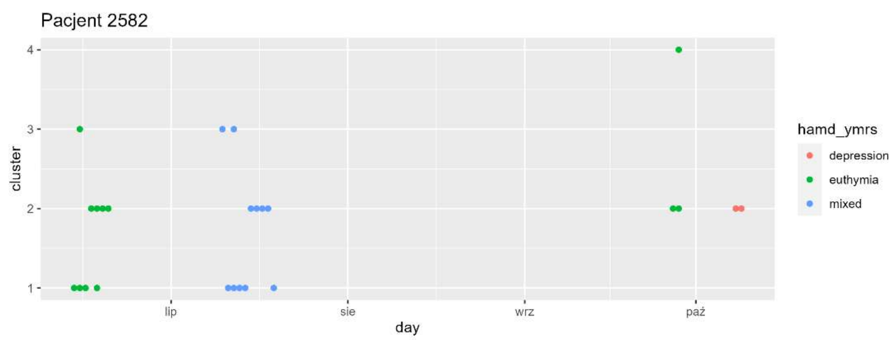

Figure 3. Clusters membership for patient 2582

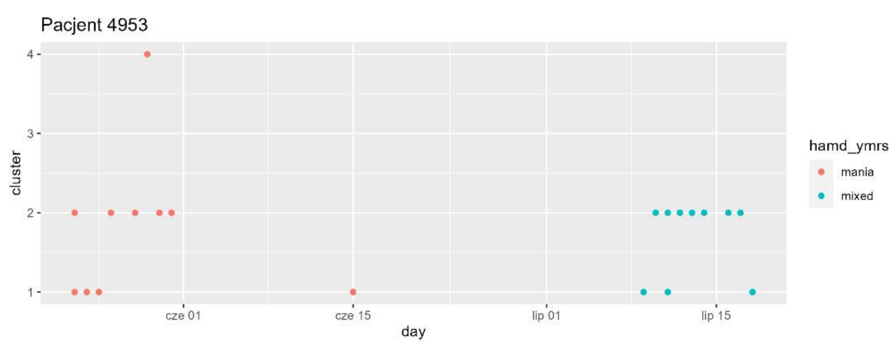

Figure 4. Clusters membership for patient 4953

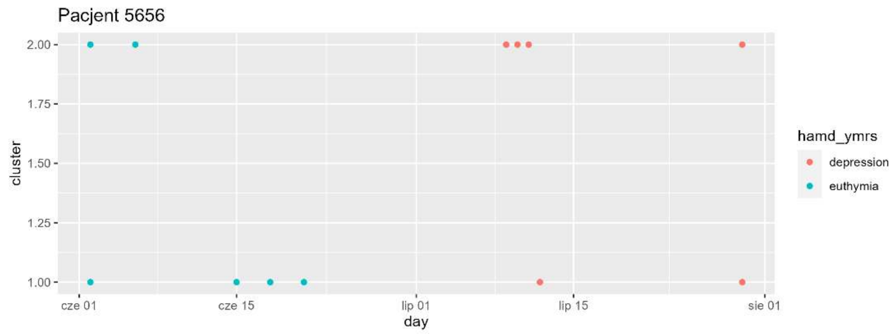

Figure 5. Clusters membership for patient 5656

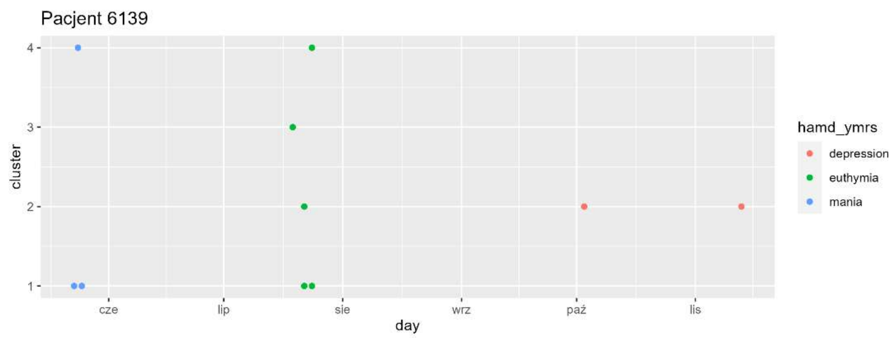

Figure 6. Clusters membership for patient 6139

Let us now see the statistics, regarding to algorithm [1] all of the received cluster were used for classification patients' states. The Random Forest model were selected as a classifier and Leave-One-Visit-Out (for each run we are testing on different number of visit) cross Validation method was selected as a performance indicator. Results of best round are presented in Table below:

	REFERENCES			
PREDICTION	**depression**	**euthymia**	**mania**	**mixed**
depression	91	11	0	2
euthymia	24	18	2	2
mania	0	0	0	0
mixed	5	42	10	50

Table 2. Confusion matrix for best round from cross-validation with clusters membership

Accuracy = 0.618

Multiclass AUC = 0.818

To compare an influence of clusters another classification model (Random Forest algorithm) was used (without information about clusters). Received Confusion matrix is presented below in table 3.

	REFERENCES			
PREDICTION	**depression**	**euthymia**	**mania**	**mixed**
depression	91	11	0	2
euthymia	24	16	3	0

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

mania	0	0	0	0
mixed	6	44	9	48

Table3. Confusion matrix for best round from cross-validation without clusters membership

Accuracy = 0.603

Multiclass AUC = 0.811

Adding information on clusters membership does not significantly improve the results of the model. However, received results seems to be promising and will be developed in the next stage of the project.

2. Initial evaluation of Pilot 2: Predict depression with Locomotor Data collected from sensors Pilot (LDP) - Evaluation of LDP-PS4: Prediction of depression using loco-motor sensors

Experiments described in this Section were based on the data collected from loco-motor sensors received from patients with depression and health control. The main predictive algorithm is similar to the one cited in Section 1 and in [1]. The first step was to specify the most important features available in the received dataset. The top 5 features are presented in the table below.

Table5. Most important features

Feature name	**Description**
AGE	patients age
CE_LAT_SWAY:	closed-eyes lateral sway
CE_STEPS_I	closed-eyes number of steps during test
OE_LAT_SWAY	open-eyes lateral sway
OE_SL	open-eyes step length

The next step of algorithm is to cluster all of the data. Received clusters for depression and healthy control are presented in Figure 7.

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

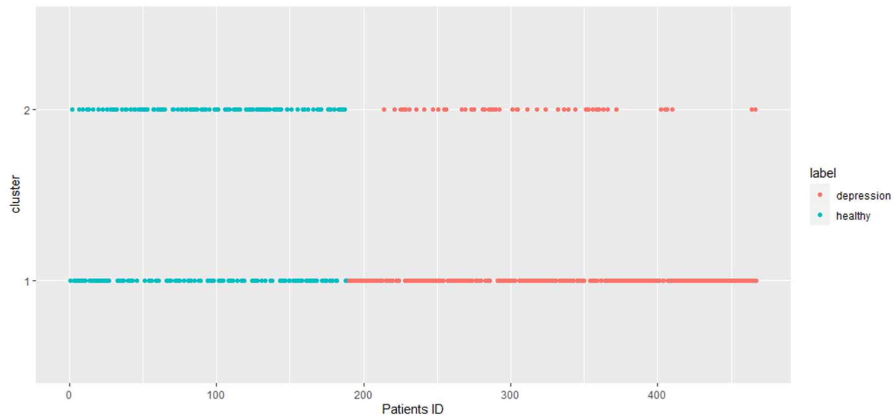

Figure 7. Clusters membership for all available patients

Clustering was executed on the complete dataset. Aiming to assign a cluster to a specific label, we start with comparing of the data-driven basic characteristics of the clusters for the considered BD states, as gathered in Table 6. It is observed that for patients with depression, most observations are assigned to cluster 1. Healthy patients are distributed almost equally to both clusters. Furthermore, overwriting the label accordingly 1 for depression and 2 for healthy control we could calculate RAND index- that is a measure of the similarity between two data clusters. Received results for patients with depression is equal to 0.7 which means that there is a light similarity between clusters and for healthy patients is equal to 0.49 which means that similarity cannot be determined.

Table6. Average membership

state	**RAND index**	**Cluster 1**	**Cluster 2**
depression	0.703	228	50
healthy	0.497	95	94

Last step of the algorithm is the classification. Similarly as in the previous example, all of the received cluster were used for classification of patients' states. Classification was performed on datasets contain top 5 features and clusters received by Fuzzy C-means algorithm with 5-fold Cross Validation. Results of that validation scenario are presented in Table 7.

Table7. Confusion matrix for best round from cross-validation with clusters membership

state	REFERENCES	
PREDICTION	**depression**	**healthy**
depression	229	49
healthy	59	130

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

Accuracy = 0.768 [+ 0.05 SD]

AUC = 0.756 [+ 0.05 SD]

To compare an influence of clusters on classification, another classification model (Random Forest algorithm) was used (without information about clusters). The resulting confusion matrix is presented below in Table 8.

	REFERENCES	
PREDICTION	**depression**	**healthy**
depression	230	48
healthy	59	130

Table8. Confusion matrix for best round from cross-validation without clusters membership

Accuracy = 0.771 [+ 0.05 SD]

AUC = 0.757 [+ 0.05 SD]

As in the previous Section, adding information on clusters membership does not significantly improve the results of the model. However, the received results seems to be promising and will be developed in the next stage of the project.

4.Future

In the future, final evaluation will be performed for all five psychiatric scenarios. Experiments will be significantly extended with next KPIs and other validation approaches.

5.References

[1] Impact of clustering unlabeled data on classification: case study in bipolar disorder, FedCSIS, https://doi.org/10.15439/2022F210

[2] PLENARY Explaining black-box models in natural language through fuzzy linguistic summaries, Information Sciences, 614, 2022, 374-399, https://doi.org/10.1016/j.ins.2022.10.010

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