A Simple and Effective Classifier for the Detection of Psychotic Disorders based on Heart Rate Variability Time Series

Krisztian Buza[1,2,*,†], Kamil Książek[3,†], Wilhelm Masarczyk[4], Przemysław Głomb[3], Piotr Gorczyca[4] and Magdalena Piegza[4]

[1] Artificial Intelligence Laboratory, Institute Jozef Stefan, Jamova cesta 39, 1000 Ljubljana, Slovenia
[2] BioIntelligence Group, Department of Mathematics-Informatics, Sapientia Hungarian University of Transylvania, Targu Mures, Romania
[3] Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Bałtycka 5, 44-100 Gliwice, Poland
[4] Department of Psychiatry, Faculty of Medical Sciences in Zabrze, Medical University of Silesia, Pyskowicka 49, 42-612 Tarnowskie Góry, Poland

Abstract

In this paper, we focus on automated detection of schizophrenia and bipolar disorder. For this task, we describe a simple and effective classifier, i.e. convolutional nearest neighbor. It provides a data-driven and objective approach for the detection of schizophrenia and bipolar disorder based on heart rate variability time series. According to our results, our approach is able to distinguish whether the selected person belongs to the patient group with an accuracy of 85% and area under receiver-operator characteristic curve of 0.92.

Keywords

heart rate variability (HRV), convolutional nearest neighbor, classification, schizophrenia, bipolar disorder

1. Introduction

Psychotic disorders are a group of psychiatric disorders that are characterized by a disruption of the ability to distinguish between the internal experience of the mind and the external reality. The two most common psychotic disorders are schizophrenia and bipolar disorder (BD) [1]. The prevalence of schizophrenia as well as BD in the Western population can be estimated at 1% each [2, 3]. Due to the fact that most of the human population has little to no access to effective mental healthcare [4] and due to the large impact that these diseases have on quality of life [5], it is important to search for new automated solutions in the field of disease diagnostics and supervision.

The diagnostic criteria for schizophrenia and BD are well defined by the International Classification of Diseases (ICD-11) and the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [6, 7]. These criteria enable a trained physician to decide on the diagnosis based on the interview with the patient. This process can be subjective and time consuming as it depends on the physicians own skills, mood and the ability to establish contact with the patient. There are questionnaire tools such as the Positive and Negative Symptom Scale (PANSS) that enable for a certain degree of objectivity by giving a more closed structure to the psychiatric interview [8, 9]. A structured interview is not a potential full solution to this problem because it does not take into account the context in which the patient lives and acts. Therefore it may potentially lead to diagnostic and disease severity assessment errors [10]. This situation urges for tools that would give a higher degree of objectivity and repeatability in the diagnostic process.

When searching for data-driven, objective methods for the diagnosis of schizophrenia and BD, heart rate variability (HRV) could be a potential biomarker that is easily obtainable with consumer grade wearable devices that function as electrocardiographs such as the Polar H10 [11, 12, 13]. HRV can be interpreted as an indirect information on the function of the centrally regulated autonomic nervous system (ANS) and its parasympathetic/sympathetic balance [14, 15]. There is a documented correlation between this ANS balance and the psychotic process severity, where a lack of proper autonomic regulation constitutes itself as lower HRV in the psychotic patient [16].

Thanks to the advancements in machine learning and data processing methods, heart rate and HRV could potentially be integrated into the clinical diagnostic process. Attempts have been made for the continuous assessment of psychotic symptoms in schizophrenia through HR anal-

ITAT'23: Information technologies – Applications and Theory, September 22–26, 2023, Tatranské Matliare, Slovakia

*Corresponding author.
†These authors contributed equally.
buza@biointelligence.hu (K. Buza)
http://www.biointelligence.hu/ (K. Buza)
0000-0002-7111-6452 (K. Buza); 0000-0002-0201-6220 (K. Książek); 0000-0001-9516-0709 (W. Masarczyk); 0000-0002-0215-4674 (P. Głomb); 0000-0002-9419-7988 (P. Gorczyca); 0000-0002-8009-7118 (M. Piegza)

ysis, although the assessment of HR and accelerometry instead of HRV raises the question whether this is an ANS or physical activity biomarker [17]. Attempts for continuous assessment of HRV in BD used low precision photoplethysmography (PPG) sensors that are inferior to ECG devices [11, 18].

To avoid the beforementioned problems in this paper we focus on HRV and use ECG sensors. We propose an automated, data-driven and therefore objective method that may assist the diagnosis of schizophrenia and bipolar disorder. Our approach is based on machine learning techniques. In particular, we aim to recognize the aforementioned diseases based on HRV time series. Thus, we consider the diagnosis of schizophrenia and bipolar disorder as a time series classification task. To the best of our knowledge, ours is the first work that uses convolutional nearest neighbor for the classification of HRV data in the context of the diagnosis of schizophrenia and bipolar disorder.

2. Materials and Methods

2.1. Data

For this study, we recruited[1] 30 adult patients diagnosed and hospitalized for schizophrenia or bipolar disorder in the Psychiatric Department in Tarnowskie Góry, Poland.

We decided to consider schizophrenia and BD together as *treatment group* for the classifier due to their similarities in genetic and neuroanatomical features as well as the fact that both are defined as clinical syndromes or disorders without regard to their pathophysiology [19]. Moreover, HRV can be understood as a transdiagnostic biomarker of psychopathology which further justifies our decision [20]. The *control group* consisted of 30 adults without any current psychiatric condition. The youngest persons in the treatment and the control were 20 and 24, respectively, while the oldest in both groups were 69.

For data collection, we used a wearable device of high quality, particularly Polar H10. It collects ECG signals, detects R peaks in the ECG and calculates the time between consecutive R peaks resulting in RR time series, i.e., each value in the RR time series corresponds to the time between two consecutive R peaks (i.e., the length of an RR interval). The RR time series acquired by the device is referred to as the *raw RR time series* in the remainder of the paper.

After removing artifacts, we obtain the *preprocessed RR time series*. The length of both raw and preprocessed RR time series varies between 7.000 and 13.000 with around 10.000 on average. This corresponds to approximately 2-hour recordings of RR intervals. We made the data publicly available with its documentation to which we refer for further details [21].

[1]Participants were informed and asked for written consent prior to admission to the experiment. The experimental procedure in the present study received approval from the local Bioethics Committee. Approval nr.: BNW/NWN/0052/KB1/135/I/22/23.

2.2. Convolutional Nearest Neighbor

As mentioned previously, our data contains 60 persons in total. In both versions of the data, i.e., both in the raw and preprocessed data, each of the participants is associated with a RR time series, the length of which is 10 000 on average. Given these relatively long time series and the fact that the disease is reflected in relatively short segments of the RR time series, we consider short segments of the long time series. The length of each short segment is 25. In particular, we extract overlapping segments according to a moving window schema with a step size of 5 and 1 for training and test data, respectively. For example, in case of a step size of 1, a RR time series with length of 9500 results in $9500 - 25 + 1 = 9451$ segments, whereas in case of a step size 5, the number of resulting segments is $\lceil (9500 - 25 + 1)/5 \rceil = 1891$.

We use k-nearest neighbor with $k = 25$ and cosine distance to assign a score to each of these segments. The score of a segment corresponds to the ratio of its neighbors that belong to the treatment (positive) class. For example, if 7 out of 10 nearest neighbors of a segment belong to the treatment class, the score of the segment is 0.7.

In order to classify a person, we consider all segments of her/his time series. The final score associated with the evaluated person is the average of the scores associated with the segments. This score may directly be used to assess the likelihood of the disease for the particular person, while all the scores associated with a set of experiment participants may be used to assess the quality of the model, for example, in terms of the area under the receiver-operator characteristic (ROC) curve.

Whenever a clear decision is needed on whether the model considers a particular person to belong to the treatment group or control group (for example, in order to calculate accuracy, i.e., the ratio of correctly classified persons), we use a decision threshold of 0.5.

As our approach incorporates a rolling window technique with overlapping windows, it resembles a 1-dimensional convolutional kernel. Therefore, we call it *convolutional nearest neighbor*.

3. Results and Discussion

We performed experiments according to the leave-one-person-out cross-validation protocol. That is: in each round of the cross-validation, we considered all the segments belonging to one of the persons as test data,

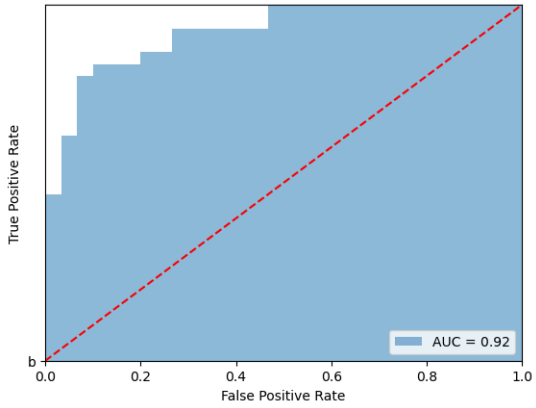

Figure 1: Receiver-operator characteristic (ROC) curve in case of preprocessed RR time series.

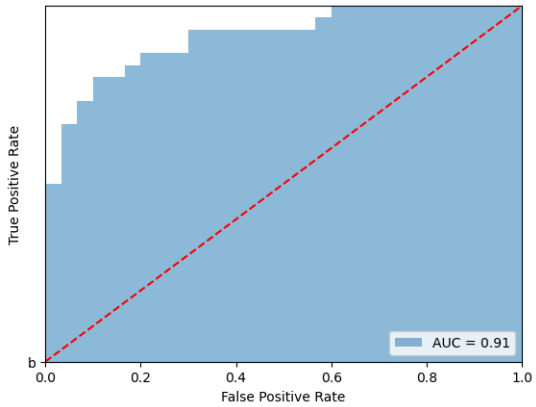

Figure 2: Receiver-operator characteristic (ROC) curve in case of raw RR time series.

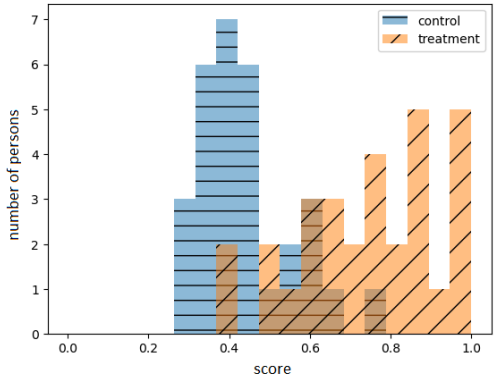

Figure 3: The distribution (histogram) of predicted scores in case of preprocessed RR time series for subjects in the control and treatment groups.

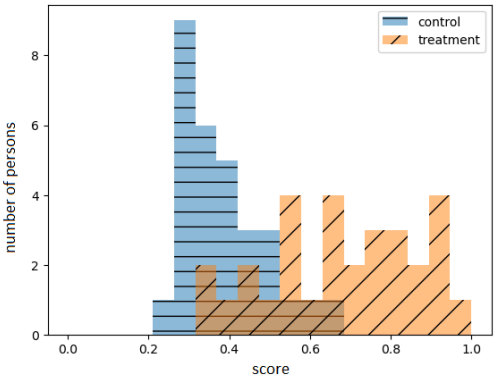

Figure 4: The distribution (histogram) of predicted scores in case of raw RR time series for subjects in the control and treatment groups.

whereas the segments of all the other persons were considered as training data.

Our classifier achieved an accuracy of 0.850 and 0.833 using the preprocessed and raw data, respectively. The corresponding receiver-operator characteristic curves are shown in Fig. 1 and Fig. 2. The area under the receiver-operator characteristic curve (AUC) is 0.92 and 0.91.

As one can see, the performance on the preprocessed data is slightly better (one more person is classified correctly). The fact that convolutional nearest neighbor is able to achieve relatively high accuracy, even in the case of raw data, indicates that convolutional nearest neighbor is a promising candidate in applications where (semi-automated) preprocessing is not feasible, but enough data is available. Furthermore, we mention that we experimented with other versions of convolutional nearest neighbor, in particular with various other distance metrics (Euclidean, Manhattan, dynamic time warping) and various numbers of nearest neighbors used for classification and they achieved similar results, indicating the stability of convolutional nearest neighbor in this domain.

The distributions of participants' scores predicted by our convolutional nearest neighbor classifier are shown in Fig. 3 and Fig. 4 in the case of the preprocessed and raw RR data, respectively. In both cases, the distributions of persons in the treatment (positive) and control (negative) groups are clearly different. As one can see, 0.5 may serve as an appropriate threshold value that allows classifying most of the persons correctly.

3.1. Diagnosis using less data

Next, we simulate the situation in which the ECG of the patient is observed for a shorter period of time and examine the performance of our classifier in this case.

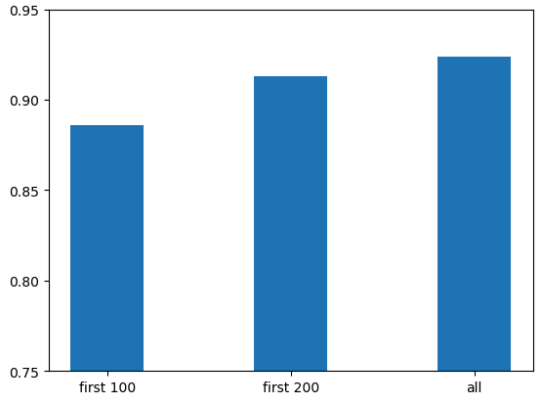

Figure 5: AUC in case of classifying persons based on the first 100, first 200 and all the segments of the preprocessed RR time series.

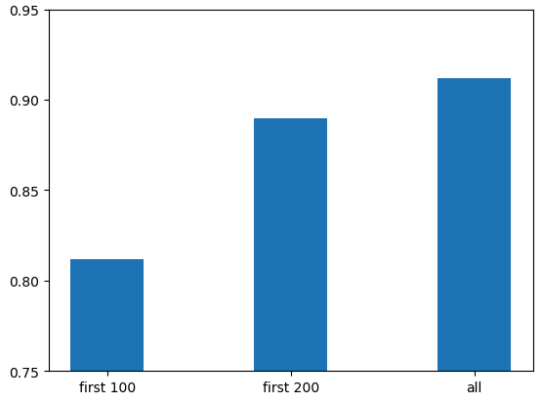

Figure 6: AUC in case of classifying persons based on the first 100, first 200 and all the segments of the raw RR time series.

In particular, when classifying a given person, we only consider the (i) first 100 and (ii) first 200 segments of the RR time series. This corresponds to a few minutes of RR time series and roughly 1% and 2% of the entire RR signal we obtained for the given person.

Fig. 5 and Fig. 6 show the AUC in case of preprocessed and raw RR time series when using the first 100, first 200 and all the segments. As expected, when using first 100 or 200 segments only, the classifier is less accurate compared with using the entire signal. Nevertheless, the difference is marginal. Therefore the results show that accurate classification may be achieved even if the ECG is observed for a relatively short time.

3.2. Other classifiers

Most of the "classic" time series classifiers are based on dynamic time warping, see e.g. [22] for an introductory survey. In contrast, recent approaches are based on deep learning [23] or the combination of deep learning techniques and dynamic time warping, see e.g. [24].

We run an initial experiment with such classifiers. As for dynamic time warping, we tried to use it as a distance measure instead of cosine distance within our convolutional nearest neighbor classifier. As expected, the resulting approach was orders of magnitude slower (i.e., more expensive computationally). However, the accuracy was roughly the same as in the case of other distances, such as cosine, Manhattan and Euclidean. Our finding is in line with the observations reported by Ding et al. [25], according to which the performance of Euclidean distance converges to the performance of dynamic time warping with increasing size of training data. We note that in our case, the training data contained roughly 100.000 segments in each cross-validation round which may explain why similar results were achieved with various distance metrics.

Regarding the approaches based on deep learning, fully convolutional neural networks (FCNs) were reported to be a strong baseline [23]. According to our initial observations, it was difficult to find appropriate hyperparameters (such as learning rate, batch size, etc.) for training FCNs and their performance on the test data was unstable.

Last, but not least, we mention that our AUC of 0.92 is slightly higher than the AUC achieved by Reinersten et al. [17], in case of two-day-long signals which is substantially longer than the RR time series we considered. Furthermore, in terms of AUC, our results seem to be competitive with many other approaches from the literature as well [26]. Nevertheless, we emphasize that both Reinersten et al. [17] and the works surveyed by Montazeri et al. [26] used different datasets.

4. Conclusion

In this paper, we focused on the automated, data-driven and therefore objective detection of schizophrenia and bipolar disorder based on heart rate variability (HRV) time series. We presented convolutional nearest neighbor, a simple, but effective approach for this task. We provided a detailed analysis of the predictions under various conditions, such as raw and preprocessed RR signals. We point out that, according to our observations, convolutional nearest neighbor is a robust classifier w.r.t. the settings of its hyperparameters, such as the distance metric or the number of nearest neighbors. Additionally, we showed that convolutional nearest neighbor is able to classify persons with a reasonable accuracy even if the

HRV is only observed for a relatively short time of a few minutes.

As for our future work, we will examine convolutional nearest neighbor more systematically under various conditions (e.g. different number of nearest neighbors, segment length) and we plan to perform more experiments with neural networks.

References

[1] J. A. Lieberman, M. B. First, Psychotic disorders, New England Journal of Medicine 379 (2018) 270–280.

[2] R. S. Kahn, I. E. Sommer, R. M. Murray, A. Meyer-Lindenberg, D. R. Weinberger, T. D. Cannon, M. O'Donovan, C. U. Correll, J. M. Kane, J. van Os, T. R. Insel, Schizophrenia, Nature Reviews Disease Primers 1 (2015) 15067. URL: http://www.nature.com/articles/nrdp201567. doi:10.1038/nrdp.2015.67.

[3] A. Jain, P. Mitra, Bipolar Disorder, 2023. URL: http://www.ncbi.nlm.nih.gov/pubmed/36579354.

[4] World Health Organization, Mental disorders, 2022. URL: https://www.who.int/news-room/fact-sheets/detail/mental-disorders.

[5] S. Evans, S. Banerjee, M. Leese, P. Huxley, The impact of mental illness on quality of life: A comparison of severe mental illness, common mental disorder and healthy population samples, Quality of Life Research 16 (2007) 17–29. doi:10.1007/s11136-006-9002-6.

[6] D. J. Stein, P. Szatmari, W. Gaebel, M. Berk, E. Vieta, M. Maj, Y. A. De Vries, A. M. Roest, P. De Jonge, A. Maercker, et al., Mental, behavioral and neurodevelopmental disorders in the icd-11: an international perspective on key changes and controversies, BMC medicine 18 (2020) 1–24.

[7] American Psychiatric Association, Diagnostic and statistical manual of mental disorders: DSM-5, American Psychiatric Association Washington, DC, 2013.

[8] N. C. Andreasen, S. Arndt, D. Del Miller, M. Flaum, P. Nopoulos, Correlational Studies of the Scale for the Assessment of Negative Symptoms and the Scale for the Assessment of Positive Symptoms: An Overview and Update, Psychopathology 28 (1995) 7–17. URL: https://www.karger.com/Article/FullText/284894. doi:10.1159/000284894.

[9] N. C. Andreasen, The Scale for the Assessment of Negative Symptoms (SANS): Conceptual and theoretical foundations, British Journal of Psychiatry 155 (1989) 49–52. doi:10.1192/s0007125000291496.

[10] J. Nordgaard, L. A. Sass, J. Parnas, The psychiatric interview: Validity, structure, and subjectivity, European Archives of Psychiatry and Clinical Neuroscience 263 (2013) 353–364. doi:10.1007/s00406-012-0366-z.

[11] K. Hinde, G. White, N. Armstrong, Wearable devices suitable for monitoring twenty four hour heart rate variability in military populations, Sensors 21 (2021) 1061.

[12] K. E. Speer, S. Semple, N. Naumovski, A. J. McKune, Measuring heart rate variability using commercially available devices in healthy children: A validity and reliability study, European Journal of Investigation in Health, Psychology and Education 10 (2020) 390–404.

[13] R. Gilgen-Ammann, T. Schweizer, T. Wyss, Rr interval signal quality of a heart rate monitor and an ecg holter at rest and during exercise, European Journal of Applied Physiology 119 (2019) 1525–1532. doi:10.1007/s00421-019-04142-5.

[14] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, et al., Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology, Circulation 93 (1996) 1043–1065.

[15] G. G. Berntson, J. Thomas Bigger Jr, D. L. Eckberg, P. Grossman, P. G. Kaufmann, M. Malik, H. N. Nagaraja, S. W. Porges, J. P. Saul, P. H. Stone, et al., Heart rate variability: origins, methods, and interpretive caveats, Psychophysiology 34 (1997) 623–648.

[16] B. R. Benjamin, M. Valstad, T. Elvsåshagen, E. G. Jönsson, T. Moberget, A. Winterton, M. Haram, M. C. Høegh, T. V. Lagerberg, N. E. Steen, L. Larsen, O. A. Andreassen, L. T. Westlye, D. S. Quintana, Heart rate variability is associated with disease severity in psychosis spectrum disorders, Progress in Neuro-Psychopharmacology and Biological Psychiatry 111 (2021) 110108. URL: https://www.sciencedirect.com/science/article/pii/S0278584620304243. doi:https://doi.org/10.1016/j.pnpbp.2020.110108.

[17] E. Reinertsen, M. Osipov, C. Liu, J. M. Kane, G. Petrides, G. D. Clifford, Continuous assessment of schizophrenia using heart rate and accelerometer data, Physiological measurement 38 (2017) 1456.

[18] A. Stautland, O. B. Fasmer, P. Jakobsen, B. Osnes, J. Torresen, T. Nordgreen, K. J. Oedegaard, P178. heart rate variability as biomarker for bipolar mania, Biological Psychiatry 91 (2022) S159.

[19] C. Dacquino, P. De Rossi, G. Spalletta, Schizophrenia and bipolar disorder: the road from similarities and clinical heterogeneity to neurobiological types,

Clinica chimica acta 449 (2015) 49–59.

[20] T. P. Beauchaine, J. F. Thayer, Heart rate variability as a transdiagnostic biomarker of psychopathology, International journal of psychophysiology 98 (2015) 338–350.

[21] K. Książek, W. Masarczyk, P. Głomb, M. Romaszewski, I. Stokłosa, P. Ścisło, P. Dębski, R. Pudlo, P. Gorczyca, M. Piegza, HRV-ACC: a dataset with R-R intervals and accelerometer data for the diagnosis of psychotic disorders using a Polar H10 wearable sensor, 2023. URL: https://doi.org/10.5281/zenodo.8171266. doi:10.5281/zenodo.8171266, The study was approved by the Bioethics Committee of the Medical University of Silesia, No.: BNW/NWN/0052/KB1/135/I/22/23.

[22] K. Buza, Time series classification and its applications, in: Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics, 2018, pp. 1–4.

[23] Z. Wang, W. Yan, T. Oates, Time series classification from scratch with deep neural networks: A strong baseline, in: 2017 International joint conference on neural networks (IJCNN), IEEE, 2017, pp. 1578–1585.

[24] K. Buza, M. Antal, Convolutional neural networks with dynamic convolution for time series classification, in: Advances in Computational Collective Intelligence: 13th International Conference, ICCCI 2021, Kallithea, Rhodes, Greece, September 29–October 1, 2021, Proceedings 13, Springer, 2021, pp. 304–312.

[25] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, E. Keogh, Querying and mining of time series data: experimental comparison of representations and distance measures, Proceedings of the VLDB Endowment 1 (2008) 1542–1552.

[26] M. Montazeri, M. Montazeri, K. Bahaadinbeigy, M. Montazeri, A. Afraz, Application of machine learning methods in predicting schizophrenia and bipolar disorders: A systematic review, Health Science Reports 6 (2023) e962.