

# Monitoring of COVID-19 Pandemic-related Psychopathology using Machine Learning

Kenneth C. Enevoldsen<sup>1,2</sup>, Andreas A. Danielsen<sup>2,3</sup>, Christopher Rohde<sup>2,3</sup>, Oskar H. Jefsen<sup>2,3</sup>,  
Kristoffer L. Nielbo<sup>1,2</sup>, Søren D. Østergaard<sup>2,3</sup>

<sup>1</sup>Center for Humanities Computing Aarhus, Aarhus University, Aarhus, Denmark

<sup>2</sup>Department of Affective Disorders, Aarhus University Hospital - Psychiatry, Aarhus, Denmark

<sup>3</sup>Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

*Keywords:* Coronavirus; COVID-19; Pandemic; Mental Disorders; Machine Learning; Natural  
Language Processing

## **Corresponding author:**

Kenneth C. Enevoldsen  
Center for Humanities Computing Aarhus  
Aarhus University  
Jens Chr. Skous Vej 4,  
8200 Aarhus  
Denmark  
Phone: +45 20824404  
E-mail: [kenneth.enevoldsen@cas.au.dk](mailto:kenneth.enevoldsen@cas.au.dk)

## Abstract

The COVID-19 pandemic has been shown to have major impact on global mental health. We show that the developments in COVID-19 pandemic-related psychopathology can be meaningfully monitored using machine learning methods providing invaluable insights to psychiatric services and of the management of the psychological consequences of the COVID-19 pandemic. The COVID-19 pandemic-related psychopathology among patients with mental illness was found to covary with the pandemic pressure, however, was less pronounced during the second wave compared to the first wave of the pandemic - possibly due to habituation.

The COVID-19 pandemic is believed to have a major negative impact on global mental health due to the viral disorder itself as well as the lockdowns, social distancing, isolation, fear, and increased uncertainty associated with the pandemic (Brooks et al., 2020; Sønderskov et al., 2021; Szcześniak et al., 2021). Individuals with preexisting mental illness are likely to be particularly vulnerable to the psychological stress associated with the COVID-19 pandemic (Jefsen et al., 2020; Rohde et al., 2020). Indeed, based on screening of clinical notes from electronic health records, we recently showed that many patients with mental illness appeared to develop COVID-19 pandemic-related psychopathology – i.e., symptoms of mental illness that seemed to be either directly or indirectly caused by the pandemic – during the first wave of the pandemic in the spring of 2020 (Rohde et al., 2020). Here, we continue this effort in an attempt to monitor the development of COVID-19 pandemic-related psychopathology over the extended course of the pandemic.

Based on the clinical notes assessed for COVID-19 pandemic-related psychopathology by Rohde et al., (2020), two types of supervised machine learning models were trained to classify notes as “pandemic-related psychopathology” and “not pandemic-related psychopathology”. Rohde et al. (2020) manually screened a total of 11 072 clinical notes from the electronic health record system of the Psychiatric Services of the Central Denmark Region from the period from February 1<sup>st</sup> to March 23<sup>rd</sup>, 2020 (the first confirmed case of COVID-19 in Denmark was on February 27<sup>th</sup> 2020). The 11 072 clinical notes were extracted from all 412 804 clinical notes from the period in question using COVID-19 related search queries: “corona” (corona), “covid” (covid), “virus” (virus), “epidemi” (epidemic), “pandemi” (pandemic), “smitte” (contaminate/contamination), including compound words. The 11 072 notes were labeled

manually with regard to COVID-19 pandemic-related psychopathology by CR and OHJ with discrepancies resolved through discussion and consultation with SDØ. 1 357 of the 11 072 clinical notes described COVID-19 pandemic-related psychopathology. For more details, please see Rohde et al. (2020).

Using the 11 072 labeled clinical notes (1357 describing COVID-19 pandemic-related psychopathology and 9715 without COVID-19 pandemic-related psychopathology), two types of machine learning models (XGBoost and Naïve Bayes) were trained using a 5-fold cross-validated grid search on the training set. The training set consisted of 70% of the 11 072 clinical notes (randomly drawn). The models were trained on the natural text in the clinical notes, the text field indicator (e.g. ‘observations of the patient’, ‘current mental status’, ‘plan’, ‘conclusion’) patient’s sex and diagnosis, and inpatient or outpatient status. Subsequently, the trained models were validated on a test set consisting of the remaining 30% of the clinical notes. The best performing model was then applied to the clinical notes (selected using the same COVID-19 related search queries as in Rohde et al. (2020)) from the period from April 1<sup>st</sup>, 2020 to March 23<sup>rd</sup>, 2021. These notes were labeled dichotomously (pandemic-related psychopathology; yes/no) based on a threshold calculated from the test set to result in 95% specificity. Finally, to validate model performance in this later period, 500 clinical notes from this period were manually labeled by CR and OHJ (similar approach as that used by Rohde et al. (2020)), and compared to the labeling performed by the model. For more information on model selection, preprocessing, and validation, see the Supplementary Material. This project was approved by the Chief Medical Officer of Psychiatry in the Central Danish Region as part of

quality development aimed at monitoring the level of pandemic-related psychopathology during the COVID-19 pandemic.

The best performing model was XGBoost (see the Supplementary Material), which achieved an area under the receiver operating characteristic curve (AUC) of 0.88 on the test set. The results obtained when applying the XGBoost model to the clinical notes from the period from April 1<sup>st</sup>, 2020 to March 23<sup>rd</sup>, 2021 are available in Figure 1, which shows the development in pandemic-related psychopathology alongside the number of deaths due to COVID-19 in Denmark. There was a positive correlation between the level of pandemic-related psychopathology and the number of deaths due to COVID-19 in Denmark ( $R^2 = 0.13$ , 95% *CI* [0.06, 0.20],  $p < 0.0001$ , smoothed  $R^2 = 0.20$ , 95% *CI* [0.12, 0.29],  $p < 0.0001$ ). A total of 78 540 clinical notes were labeled using the model of which 7125 was found to describe pandemic-related psychopathology. The validation of the model revealed a slight performance decrease to AUC=0.83, but no systematic decline in performance over time (see Figure S1 in the Supplementary Material) as might have been expected if there was a distribution shift, i.e. a substantial change in the content of the clinical notes over time.

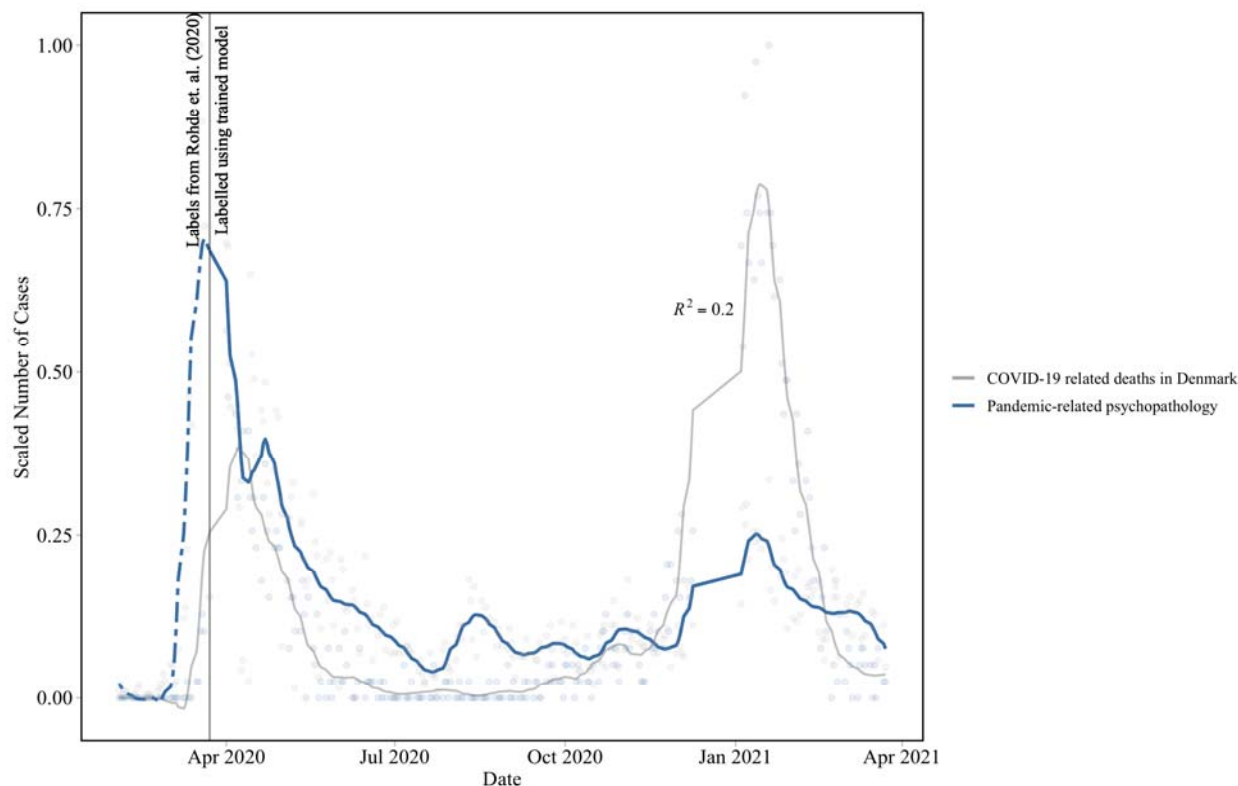


Figure 1: COVID-19 pandemic-related psychopathology and deaths due to COVID-19 over time. Both signals are smoothed using a local polynomial regression with an  $h = 10$  days and holidays and weekends are excluded. The scale is normalized between 0 and 1 to allow for comparison. The dotted line indicates the transition from the clinical notes labeled by Rohde et al. (2020) to the trained model. Data points representing less than 5 cases of pandemic-related psychopathologies are not visualized.

The correlation between the number of clinical notes describing pandemic-related psychopathology and the number of deaths due to COVID-19 in Denmark indicates that the mental health impact of the COVID-19 pandemic among patients with mental illness covaries with the pandemic pressure. This finding is clinically meaningful and compatible with the

covariation between the pandemic pressure and psychological well-being observed at the general population level in Denmark (Sønderskov et al., 2021). The correlation, however, seems to be less pronounced for the second wave compared to the first wave of the pandemic. As there was no systematic decline in model performance over the observation period, the weakened correlation could likely be due to habituation – i.e. that patients are less sensitive to the development in the pandemic as time passes, possibly because the situation appears less insecure compared to the initial phase of the pandemic. Both the correlation in itself and its weakening over time are important observations that may aid psychiatric services in the planning of the management of the psychological consequences of the COVID-19 pandemic. Furthermore, the results are testament to the potential of applying machine learning on structured and natural text data from electronic health records in clinical psychiatry.

### **Acknowledgements**

The authors thank Bettina Nørremark from the Business Intelligence Office in the Central Denmark Region for her assistance with extraction of data.

### **Funding**

This project is supported by an unconditional grant from the Novo Nordisk Foundation to SDØ (Grant number: NNF20SA0062874). CR is supported by the Danish Diabetes Academy, funded by the Novo Nordisk Foundation (grant number NNF17SA0031406) and the Lundbeck Foundation (grant number: R358-2020-2342). OHJ is supported by the Health Research Foundation of Central Denmark Region (grant number: R64-A3090-B1898). SDØ is supported by the Lundbeck Foundation (grant numbers: R358-2020-2341 and R344-2020-1073), the

Danish Cancer Society (grant number: R283-A16461), the Central Denmark Region Fund for Strengthening of Health Science (grant number: 1-36-72-4-20), the Danish Agency for Digitisation Investment Fund for New Technologies (grant number 2020-6720) and Independent Research Fund Denmark (grant number: 7016-00048B).

### **Declaration of interest**

CR received the 2020 Lundbeck Foundation Talent Prize. SDØ received the 2020 Lundbeck Foundation Young Investigator Prize. The remaining authors declare no conflicts of interest.

### **Data Availability**

Due to the sensitive nature of the data, they are only available for quality development projects to employees in the Central Denmark Region upon application to, and approval by, the Chief Medical Officer of Psychiatry.



## References

- Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *The Lancet*, *395*(10227), 912–920. [https://doi.org/10.1016/S0140-6736\(20\)30460-8](https://doi.org/10.1016/S0140-6736(20)30460-8)
- Jefsen, O. H., Rohde, C., Nørremark, B., & Østergaard, S. D. (2020). COVID-19-related self-harm and suicidality among individuals with mental disorders. *Acta Psychiatrica Scandinavica*, *142*(2), 152–153.
- Rohde, C., Jefsen, O. H., Nørremark, B., Danielsen, A. A., & Østergaard, S. D. (2020). Psychiatric symptoms related to the COVID-19 pandemic. *Acta Neuropsychiatrica*, *32*(5), 274–276. <https://doi.org/10.1017/neu.2020.24>
- Sønderskov, K. M., Dinesen, P. T., Vistisen, H. T., & Østergaard, S. D. (2021). Variation in psychological well-being and symptoms of anxiety and depression during the COVID-19 pandemic: Results from a three-wave panel survey. *Acta Neuropsychiatrica*, 1–4. <https://doi.org/10.1017/neu.2020.47>
- Szczęśniak, D., Gładka, A., Misiak, B., Cyran, A., & Rymaszewska, J. (2021). The SARS-CoV-2 and mental health: From biological mechanisms to social consequences. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *104*, 110046. <https://doi.org/10.1016/j.pnpbp.2020.110046>