Mental Health problem prediction of Tech Employees Using Machine Learning

Article · January 2024

DOI: 10.5281/zenodo.8337193

5 authors, including:

Rachna Narula
Bharati Vidyapeeth College of engineering new delhi
25 PUBLICATIONS **24** CITATIONS

SEE PROFILE

HBRP PUBLICATION

Advancement of Computer Technology and its Applications
Volume 6 Issue 3
DOI: https://doi.org/10.5281/zenodo.8337193

Mental Health problem prediction of Tech Employees Using Machine Learning

Siddharth Gupta [1],
PratibhaBarua[1], Akanksha Kochhar[2], Vijay kumar[1*], Rachna narula[3],
[1, 1]Students, Department of Computer Science and Engineering, Bharati Vidyapeeth's College OfEngineering, Affiliated to GGSIPU, Delhi, India
[2,1,3]Assistant Professor, Department of Computer Science and Engineering, Bharati Vidyapeeth'sCollege of Engineering, Affiliated to GGSIPU, Delhi, India*

***Corresponding Author**
***E-Mail Id:**-vijay.kumar@bharatividyapeeth.edu*

ABSTRACT

In this era rapid societal changes, in addition to technology improvements, might pose problems and stress for the future generations. Individuals and society as a whole must place a high priority on mental health and well-being in order to reduce the detrimental effects of these developments. Individuals should give more importance to their own particular ideals and ambitions than just keeping up with society's pace. This kernel's goal is to identify the factors that affect someone's mental health based on this dataset. In 2014, attitudes towards mental health as well as the prevalence of mental health issues in the tech industry were assessed. This kernel seeks to create a methodical approach to comprehending mental Health in the workplace, in contrast to the other kernels. Is there a preliminary action that must be taken? There are typically many creative kernels on many issues, but only a small number are dedicated to addressing the issues of how to start on issues in medicine, particularly those relating to local knowledge. This section outlines prostate cancer and how to identify it. Additionally, we have completed Electronic Design Automation, created a dataset, andestablished a dataset.

***Keywords:-**Mental health, Logistic Regression, K-nearest Neighbor Classifier, Decision Tree Classifier, Random Forest Classifier, Support Vector Machine, Artificial Neural Network*

INTRODUCTION

The term "mental wellbeing" describes a person's overall wellbeing. It describes a person's ability to explore new ideas, bring comfort to his surroundings, etc. Our daily lives depend on our mental health since it helps us cope with challenges and suggest methods toaddress difficulties. [1, 2]

When a person struggles with mental health concerns, it can have a significant impact on their lives. This may incorporate:

Passionate trouble: Significant emotional distress can be brought on by mental health problems, including feelings of sympathy, unease, or melancholy.

Disabled working: A person's ability to carry out daily responsibilities, such as attending work or school or maintaining personal connections, might be affected by mental health disorders.

Physical wellbeing issues: Physical well-being problems, such as headaches, stomachaches, or fatigue, can be exacerbated by mental health problems.

Social separation: Social isolation,

Advancement of Computer Technology and its Applications
Volume 6 Issue 3
DOI: https://doi.org/10.5281/zenodo.8337193

difficulties forming and maintaining connections with others, and mental health disorders can all be caused by one another.

Expanded chance of habit: Issues with mental health can make it more likely that someone will turn to alcohol, drugs, or other substances as a coping mechanism.

Candidates who have completed Mental Health First Aid training are taught how to identify and assist people who are experiencing crises related to their mental health and/or substance misuse, as well as how to connect them with the right staff resources. [3,4]

LITERATURE REVIEW

The research paper Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges written by the authors Jetli Chung, Jason Teo, Aniello Minutol [5] used a technology of machine learning named as PRISMA. It was proposed that recent research and studies exhibits that machine learning has proven to be a convenient tool for deciphering psychiatric problems. Additionally, it helps in detecting and categorizing patients' psychiatric issues so that they can receive further therapy. The conclusion and the result may not be entirely true because different data models produce varying findings.

The paper Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment written by the authors Matthew Squires, Xiaohui Tao,Soman Elangovan, Raj Gururajan, Xujuan Zhou, U Rajendra Acharya & Yuefeng Li[6] used a technology of Decision Tree Classifier. All phases of the treatment cycle are supported by cutting-edge algorithms in this paper. These technologies have the ability to identify people with mental health disorders, enabling them to receive the care they require and allowing doctors to customize therapies for specific patients who will most likely benefit from them. The report claims that there are now a number of barriers hindering the development of data-based paradigms in healthcare. Uncertainty quantification, model validation, the creation of interdisciplinary research teams, improved access to a wide range of data, and standardized definitions should all be given more thought in this discipline.

The paper named Behavioral Modeling for Mental Health using Machine Learning Algorithms written by the authors M. Srividya,S. Mohanavalli and N. Bhalaji[7] utilized the technologies of Naive Bayes, K-Nearest Neighbour, Decision Trees, Vector Machines, and Logistic Regression. It is now feasible to track different forms of anxiety, depression, and mental health disorders and deliver the most precise findings thanks to a vast array of machine learning models.

Mental Health and Personality Determination using Machine Learning is a research paper written by Kimaya Raut, Jui Patil, Siddhi Wade, Jisha Tinsu[8]. The research paper has used a technology Logistic Regression, SVM, Random Forest, Naive Bayes KNeighbors and XGB. The Key benefit of this study is that Logistic Regression accurately predicts mental health to the maximum of 96.05, whereas XGBoost classifier only provides an accurate prediction of67.68 percent for personality.

Another Research paper written by Jina Kim, Daeun Lee, Eunil Park[9] was titled Machine Learning for Mental Health in Social Media: Bibliometric Study gave the conclusion that In binary and multiclass classification tasks, the model's outcomes demonstrated that the accuracy of the CNN classifier was 91.8% and 79.8%. A total of 197,615 posts have been collected from 418 people. By utilizing

Advancement of Computer Technology and its Applications
Volume 6 Issue 3
DOI: https://doi.org/10.5281/zenodo.8337193

10-fold cross-validation techniques and pre-trained GloVe embeddings to feed word sequences into the bidirectional LSTM model, an AUC of 0.94 was achieved.

Diagnosis and prognosis of mental disorders by means of EEG and deep learning: A systematic mapping study was a research paper written by Manuel J. Rivera, Miguel A. Teruel, Alejandro Maté and Juan Trujillo [10] proposed a major advantage i.e. by using EEG with DL, it was able to provide a clear insight into the most relevant research topics on diagnosis and prognosis of mental health problems. It offered several suggestions for further research. Several outcomes are: The majority of the experiments employ convolutional neural networks, with epilepsy being the most prevalent mental illness.
A Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection is a research paper written by Piyush Kumar, Rishi Chauhan, Thompson Stephan, Achyut Shankar, Sanjeev Thakur[11] used a combination of machine learning models namely K-nearest Neighbor, NB, SVM, Decision Tree, LR. In order to provide appropriate results for this study on unemployment and mental illness, algorithms were applied to 334 sample sizes over 31 distinct fields.Social Determinants of Mental Health: Where We are and Where We Need to Go is a research paper written by Margarita Alegría, Amanda NeMoyer, Irene Falgàs Bagué, Ye Wang & Kiara Alvarez [12]. In this research paper, the technology Linear Progression has been used. It has been possible to find out the accurate result for the data for how social determinants have resulted in.

METHODOLOGY

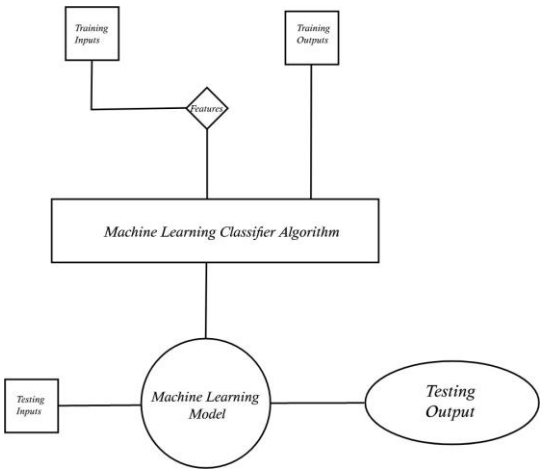


Fig.1:-*Flowchart: working of a machine learning model*

Figure 1 provides a thorough representation of the machine learning model's underlying workflow, covering the full procedure from start to finish. It explains how the algorithm efficiently processes the orderly succession of train inputs and their related outputs. Additionally, it illustrates how the trained model is then applied to evaluate fresh data, known as testing data, making it simpler to obtain the desired result. Models we tested as follows:

Logistic Regression: This Machine Learning algorithm is a very popular supervised neural algorithm that can be used to forecast either a categorical or a discrete dependent variable

Advancement of Computer Technology and its Applications
Volume 6 Issue 3
DOI: https://doi.org/10.5281/zenodo.8337193

using a set of independent variables. Because it represents the link between input and output variables, a logistic regression yields probabilistic values in the range of 0 to 1. These measurements suggest that there is a very high likelihood that specific outcomes, such as 0 or 1,yes or no, and true or false, will occur. In our research, a logistic regression was proven to be 80% accurate.

K-nearest Neighbor (K-NN) Classifier: The kNN classifies input data into coherent clusters based on the similarity to previously trained data. The class that the input data shares the most distant neighbors with is the one that will get it [13].In our study, the KNN classifier was able to obtain an accuracy of 66.31 percent.

Decision Tree Classifier: It employs a tree-like structure to express the statistical likelihood or sequence of events, action, and result. The decision tree classifier methods separate the attributes at each node in the tree to determine whether splitting is optimal in each class. [14] Our investigation discovered an accuracy rating of 72.14 percent when using the Decision Tree Classifier.

Random Forest Classifier: Random Forest classifier is a method frequently used in the context of supervised machine learning to address classification and regression issues. Each tree in an irregular backwoods is subject to the upsides of an arbitrary vector that was tested freely and with a uniform conveyance among the woodland's trees [15]. It merges several decision trees to form an amalgamation known as "the forest." All trees will be trained usingVarious, randomly chosen subsets of the input features. In our test, the Random Forest classifier succeeded in obtaining an accuracy of 72.7%.

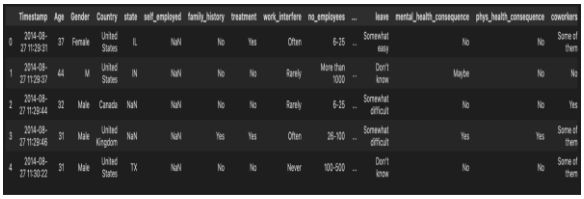
Support Vector Machine (SVM): Support Vector Machines (SVM) are supervised ML models that analyze data for classification and regression problems using a network of connected learning algorithms. The SVMs seek an optimal hyperplane with the greatest amount of data point separation between classes [16]. During our analysis, the SVM classifier had a documented accuracy of 82.5%.

Ensemble (Bagging): By using a training sample selected at random with replacement, bagging creates subgroups from the data [17].They use each set of data to train their decision trees. We now have an assortment of different models as a result. It is more trustworthy to use a decision tree classifier composed of the average of these predictions from different trees. On the XGBoost Classifier, we get 92.5 percent accuracy by using the Bagging Classifier.[17]

RESULT AND DISCUSSION

A. Data Acquisition

Our dataset was acquired by an open source platform kaggle.



	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	...	leave	mental_health_consequence	phys_health_consequence	coworkers
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	...	Somewhat easy	No	No	Some of them
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	...	Don't know	Maybe	No	No
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	Rarely	6-25	...	Somewhat difficult	No	No	Yes
3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes	Often	26-100	...	Somewhat difficult	Yes	Yes	Some of them
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	...	Don't know	No	No	Some of them

Fig.2:*-Dataset from kaggle*

B. Data description

Advancement of Computer Technology and its Applications
Volume 6 Issue 3
DOI: https://doi.org/10.5281/zenodo.8337193

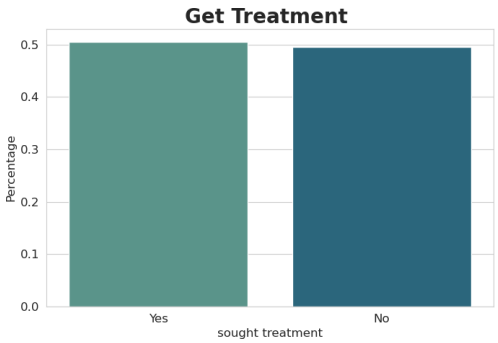


Fig.3:-*Representation of the percentage of employees getting treatment for mental health*

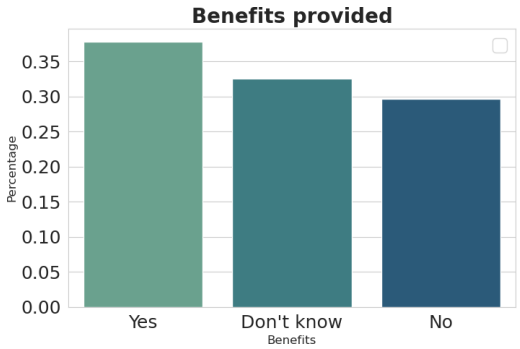


Fig.4:-*Representation of the percentage of employees getting mental health benefits from employer*

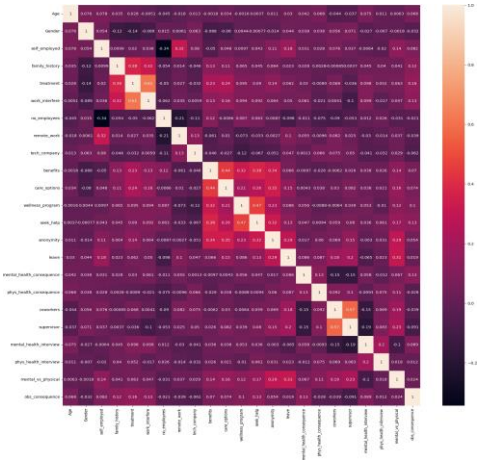


Fig.5:-*Correlation Matrix*

C. Data preprocessing

We have deleted from this paper the bizarre and irreconcilable data that we thought impossible for practical life as a negative age. To convert the text into numerical data so that our machine learning model could interpret this information, we have also

Advancement of Computer Technology and its Applications
Volume 6 Issue 3
DOI: https://doi.org/10.5281/zenodo.8337193

employed a method known as Label Encoding.

D. Machine learning model research

We have experimented with numerous machine learning models, as described by the methods in this work.

The Ensemble model implementation using XGBoost Classifier Machine Learning was our best machine learning model. We've divided the data into two parts:

The train data is the first part, and the test datais the second part. The ratio in which we split the data in the ratio 70:30 respectively and the accuracy we noted was in the test data.

Below is the table of all the models and their respective testing accuracy.

The obtained results are as follows:

Table 1:-*Result table*

Model	Accuracy
Logistic Regression	80%
K-nearest Neighbor (K-NN)	68.70%
Decision Tree	76.12%
Random Forest	81.69%
Support Vector Machine	82.50%
Ensemble(Bagging)	94.42%

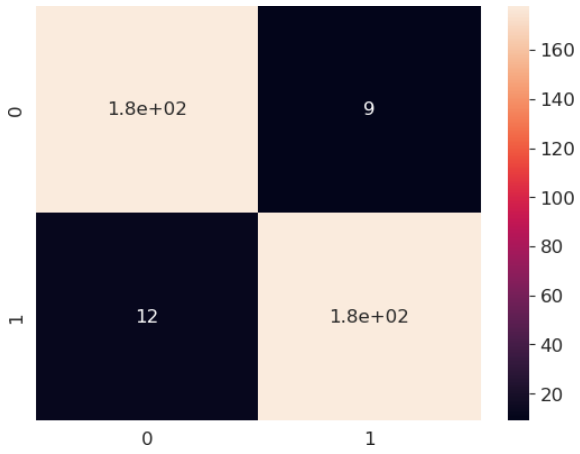


Fig.6:-*Correlation matrix for Ensemble (Bagging) model*

Advancement of Computer Technology and its Applications
Volume 6 Issue 3
DOI: https://doi.org/10.5281/zenodo.8337193

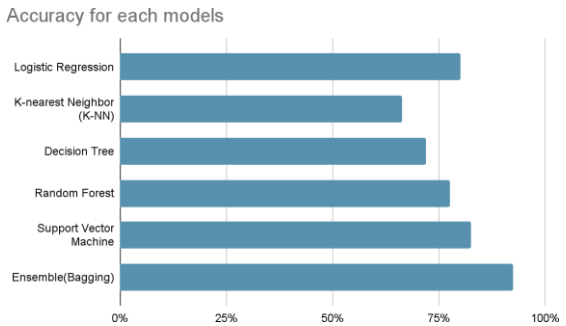


Fig.7:-*Accuracy chart for machine learning models tested on our dataset*

Receiver Operating Characteristic (ROC) Curve:

The notion of "separator" and choice variables is a basis for the ROC curve concept. If the "criterion" or "cut off" of positive results on the decision axis are changed, the frequency of positive and negative diagnostic test results will change. A ROC curve in unit square is generated by the plot of TPF sensitivity versus FPF (1, 1) Specificity with respect to each different cutoff.

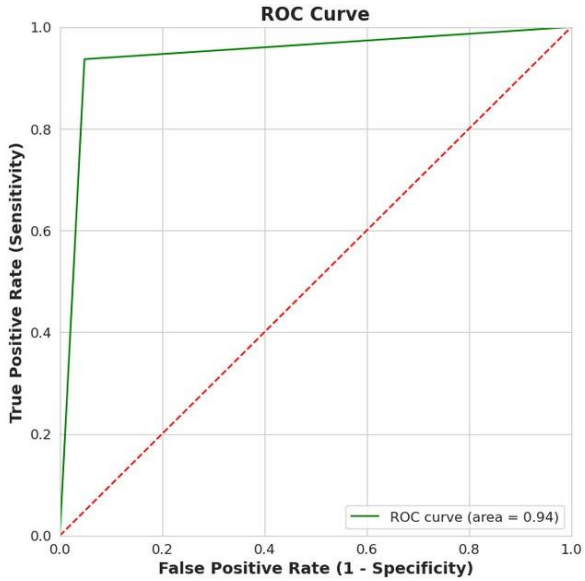


Fig.8:-*ROC curve [18] for Ensemble (Bagging) model*

We are pleased to report that our classifier is performing exceptionally well, as evidenced by the area under a curve AUC of 0.94. This high AUC indicates that our model demonstrates strong discriminatory power and accurately distinguishes between individuals with and without mental health conditions. These encouraging findings show that our classifiers can facilitate mental health assessments and intervention which could lead to more accurate diagnosis and better treatment outcomes.

CONCLUSION

A comparison of various types of ML algorithms for mental health prediction is

DOI: https://doi.org/10.5281/zenodo.8337193

offered in this research report. The outcomes show that, with an accuracy of 94.42 percent, the Bagging model performed better than competing methods. The potential for improved diagnostic accuracy and customized treatment strategies is highlighted in these findings, which add to the expanding body of research on machine learning-based mental health prediction. Exploring the fundamental causes of the Bagging model's superior performance as well as its practical use in clinical settings should be the main goal of future study. Further study should concentrate on understanding the underlying mechanisms driving the higher performance of the Bagging model and its practical implementation in clinical practice.

1. Future Scope

A number of other future possibilities are to be taken into account in order to enhance the effectiveness as well as reliability of your Machine Learning Model onBinary Classification for Technical Workers' Mental Health. In the first place, a larger and broader set of examples for training and evaluating these models will be provided by increasing the size of this data base beyond its present small size. Additionally, adding features that address work-life balance, stress levels, social support, or workplace policies can help us understand the aspects that affect the mental health of IT employees better. Implementing cross-validation methods like k-fold or stratified cross-validation can assist evaluate the model's performance on various subsets of data to ensure resilience. Finally, adding assessment variables like F1-score, recall, and precision in addition to accuracy would provide a more detailed analysis of the model's performance as well as its benefits and drawbacks in the binary classification task.

REFERENCES

1. Kumar, P., Chauhan, R., Stephan, T., Shankar, A., & Thakur, S. (2021, January). A machine learning implementation for mental health care. Application: smart watch for depression detection. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 568-574). IEEE.
2. Teles, A., Rodrigues, I., Viana, D., Silva, F., Coutinho, L., Endler, M., & Rabêlo, R. (2019, June). Mobile mental health: A review of applications for depression assistance. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 708-713). IEEE.
3. Deshmukh, V. M., Rajalakshmi, B., Dash, S., Kulkarni, P., & Gupta, S. K. (2022, January). Analysis and characterization of mental health conditions based on user content on social media. In *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-5). IEEE.
4. Chung, J., & Teo, J. (2022). Mental health prediction using machine learning: taxonomy, applications, and challenges. *Applied Computational Intelligence and Soft Computing*, *2022*, 1-19.
5. Squires, M., Tao, X., Elangovan, S., Gururajan, R., Zhou, X., Acharya, U. R., & Li, Y. (2023). Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment. *Brain Informatics*, *10*(1), 1-19.
6. Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). Behavioral modeling for mental health using machine learning algorithms. *Journal of medical systems*, *42*, 1-12.

7. Raut, K., Patil, J., Wade, S., & Tinsu, J. (2022, June). Mental Health and Personality Determination using Machine Learning. In *2022 7th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1231-1236). IEEE.
8. Kim, J., Lee, D., & Park, E. (2021). Machine learning for mental health in social media: Bibliometric study. *Journal of Medical Internet Research*, *23*(3), e24870..
9. Rivera, M. J., Teruel, M. A., Mate, A., & Trujillo, J. (2022). Diagnosis and prognosis of mental disorders by means of EEG and deep learning: a systematic mapping study. *Artificial Intelligence Review*, 1-43.
10. Kumar, P., Chauhan, R., Stephan, T., Shankar, A., & Thakur, S. (2021, January). A machine learning implementation for mental health care. Application: smart watch for depression detection. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 568-574). IEEE.
11. Alegría, M., NeMoyer, A., Falgàs Bagué, I., Wang, Y., & Alvarez, K. (2018). Social determinants of mental health: where we are and where we need to go. *Current psychiatry reports*, *20*, 1-13.
12. Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019, May). A brief review of nearest neighbor algorithm for learning and classification. In *2019 international conference on intelligent computing and control systems (ICCS)* (pp. 1255-1260). IEEE..
13. Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering*, *6*(10), 74-78..
14. Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.
15. Sammut, C., & Webb, G. I. (Eds.). (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.
16. Gupta, G., Adarsh, U., Reddy, N. S., & Rao, B. A. (2022). Comparison of various machine learning approaches uses in heart ailments prediction. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012010). IOP Publishing..
17. Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine*, *4*(2), 627.

Cite as: Siddharth Gupta, Pratibha Barua, Akanksha Kochhar, Vijay kumar & Rachna narula, (2023). Mental Health problem prediction of Tech Employees Using Machine Learning. Advancement of Computer Technology and Its Applications, 6(3), 67–75.
https://doi.org/10.5281/zenodo.8337193