

Proposal details

Title of the project: Picture perfect? Predicting Psychological Resilience in Youth by Machine Learning

Duration of the project (in months): 12

Public summary

English public summary (100/100): Up to 50% of Dutch youth (aged 16-30) experience psychological problems such as anxiety and depressive symptoms, and suicide is the leading cause of death among Dutch youth. In this longitudinal experience sampling study, we will investigate to what extent risk factors and protective factors for psychological problems can be derived from photos shared on social media. We will apply scraping techniques and Machine Learning to extract facial expressions from photos - as well as relevant contextual factors (e.g., being with friends) - and test whether these factors can be used to predict youth's resilience or vulnerability to psychological problems.

Dutch public summary (97/100): Bijna de helft van de Nederlandse jongeren (16-30 jaar) heeft psychische problemen, zoals angst- en depressieve symptomen, en zelfdoding is de belangrijkste doodsoorzaak onder de Nederlandse jongeren. In deze longitudinale gestructureerde dagboekstudie onderzoeken we in hoeverre risicofactoren en beschermende factoren voor psychische problemen kunnen worden afgeleid uit foto's die op sociale media worden gedeeld. We zullen scraping-technieken en Machine Learning toepassen om gezichtsuitdrukkingen uit foto's te extraheren - evenals relevante contextuele factoren (bijv. samenzijn met vrienden) - en testen of deze factoren kunnen worden gebruikt om de veerkracht en kwetsbaarheid van jongeren voor psychische problemen te voorspellen.

Budget

Type of costs	Clarification	Costs
Postdoc	0.2 FTE, 12 months	€ 16.751
Non-scientific personnel MBO	0.3 FTE, 12 months	€ 18.716
Other costs	Data collection; participant reimbursement; Open Access publishing	€ 12.000
Travel and accommodation costs	Presentation of findings at conference	€ 2.500
Total request		€ 49.967

Budget justification (148/150)

Personnel

During the first month of the project, the postdoc will write a Research Data Management Protocol and start the local ethics procedure. The research assistant will provide practical support, and scrape data for 90 days per participant. Towards the end of the project, the postdoc will write an Open Access article.

Other costs

The data is collected by Research Agency Flycatcher (€6200), an academic Support Center that assists researchers in their data collection process. To improve response, three Apple iPads will be allotted by means of a lottery (€1800). The local Center for Information Technology will assist in scraping the relevant data from the social media platform BeReal, and money is reserved for Open Access publishing (€4000).

Travel and accommodation costs

Conference, travel and accommodation expenses for the postdoc to present results towards the end of the project. For example, the European Conference on Developmental Psychology 2024.

Project proposal

The COVID-19 pandemic has had a major impact on the mental health of youth. Already before the pandemic, psychological problems like depression and anxiety had been gradually increasing within this group.^{1,2} **Many Dutch youth experience psychological problems**, such as loneliness (49%), stress (49%), and suicidal ideation (16%).³ In fact, suicide is the leading cause of death among youth.⁴ Psychological problems typically start before the age of 24, and can have devastating long-term effects like school dropout, decreased social engagement, severe psychiatric disorders, and ultimately suicide. In Europe, the total costs of mental ill-health are over EUR 600 billion (4% of GDP).⁵

So far, **Machine learning**, a branch of Artificial Intelligence, has **successfully been implemented to predict psychological problems**. For example, suicidal ideation has been accurately predicted from audio recordings (e.g.,⁶) and sentiment of texts shared in social media posts (e.g.,⁷). However, most studies use Facebook data - a social media platform mostly used by adults - and focus on written text. Social media is widely used by youth to express their thoughts and emotions in their daily life. Not only in written text but also in the form of photos – for example, in the new app “BeReal”.

BeReal asks their users to show who they really are by sharing a **daily snapshot of their daily life** without using filters or other enhancing methods. Once a day, at a random time point, users are surprised with a notification starting a 2-minute window to post a picture using the phone’s front-facing and back-facing cameras simultaneously. Friends can respond to the post with comments and emojis. Photos disappear from BeReal after 24 hours. BeReal is a hype among youth, and approximately 32%-37% of Dutch youth aged 16-24 years have the app installed on their smartphone.⁸

The photo materials uploaded in BeReal might be **eminently suited for research** purposes because - unlike Instagram and Snapchat - real emotions are shared. Mental health problems can be predicted by **facial expressions**⁹, as well as by self-reported contextual cues such as being in **company** of friends¹⁰, exposure to **nature** (e.g., visiting a park)¹¹, and **being active** or doing sports¹². Interestingly, all aforementioned characteristics can be determined in photos in an **innovative way** using Machine Learning techniques. If proven feasible, this would be **groundbreaking for patients and clinicians** because it will **alleviate the burden** on patients if they can take photos instead of repeatedly answering questions that are perceived as burdensome. In addition to the impact on clinical practice, being able to deduce risk and protective factors for mental health from photos will also have a **tremendous scientific impact**. That is, it will not only allow community health service institutes to screen for mental health problems on a **drastically larger scale**, but it will also **open up new exciting possibilities** for population-based research on the mental health of youth. Therefore, the objectives of this project are to:

1. Use scraping and machine Learning techniques to extract relevant characteristics from photos;
2. Validate these photo-characteristics using Experience Sampling Methods (i.e., daily self-reports on what participants were doing, with whom, and whether they were in urban areas);
3. Identify which valid photo-extracted factors can be reliably used to predict youth’s mental health and resilience to psychological problems.

Data collection tools and procedures

A minimum of 300 participants aged between 16 and 30 years will be recruited via Flycatcher, a Dutch internet research organization with its own panel representative of the Dutch population (FlyCatcher Panel). The eligibility criteria include being fluent in Dutch, currently living in The Netherlands, having a smartphone and being active on BeReal. Informed consent will be obtained from all participants for scraping photos and analyzing their content. Participants will be reimbursed €20,- for their participation. In addition, participants will be offered the opportunity to enter a draw to win one of three iPads.

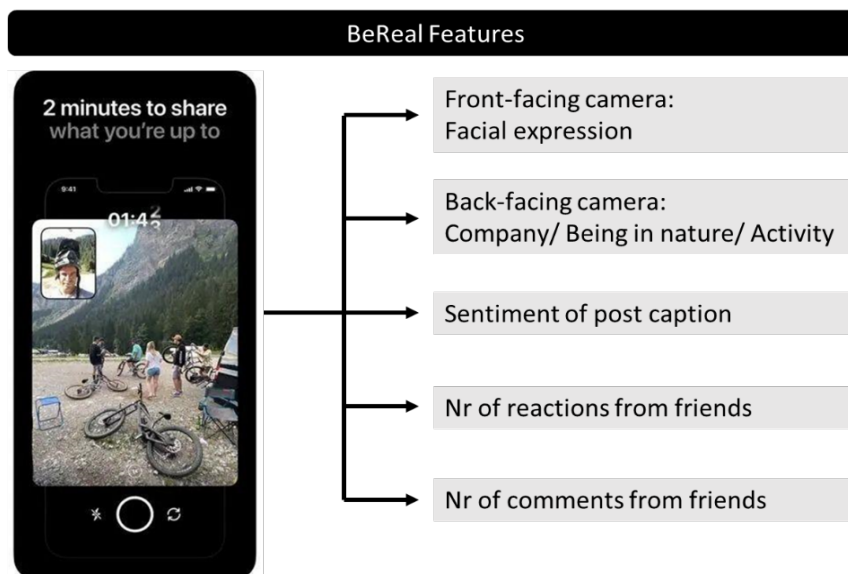
Eligible participants will be asked to befriend the new BeReal account that will be created for the project. A recruitment poster will be uploaded to the account. Participants can click on the link in the post or scan the QR code in the poster which directs to an online questionnaire with baseline questions on socio demographics (e.g., age, sex, socioeconomic status), mental health (e.g., depression, anxiety), and constructs related to psychological resilience (e.g., coping, cognitive flexibility).

After participants have become friends with the newly created BeReal account, open source scraping software will be used to extract their daily BeReal activity before it disappears for the upcoming 90 days. The time of posting will be scraped; the caption of the post; the picture taken by the front-facing and back-facing cameras; the number of reactions and the number of comments on their posts.

After scraping all relevant information, machine learning will be employed to enable sentiment analysis. That is, facial expressions will be extracted from the front-facing photo using the open-source toolkit “OpenFace”¹³, and the content of the photo taken with the back-facing camera will be described using a screen reader that is typically used by blind people (e.g.,

“Biking trip with friends”). The caption of each post will be subject to sentiment analysis as well using the free version of DiscoverText (academics get one-year free access).

To validate the data extracted from the BeReal photos, participants will be sent daily experience sampling questionnaires with questions about how they felt, how much stress, loneliness and psychological complaints they experienced, what they were doing (e.g., biking), whether they were in urban areas and with whom they were during the BeReal snapshots that day (e.g., with friends). This will give us an indication of the precision of the photo-derived characteristics.



Finally, multilevel analyses will be used to identify which photo-extracted factors can be reliably used to predict youth’ mental health and resilience to, for example, anxiety and depressive symptoms. That is, by means of multilevel regression analyses with cross-level interactions in R, we will investigate to what extent change in participants’ facial expressions or relevant contextual factors (e.g., company of friends; being in nature; being active) will predict mental health indicators (e.g., depressive symptoms) at time T, while controlling for level of that mental health indicator at the previous day T-1 (i.e., $Depression_{ij} = \beta_{0j} + \beta_{1j}(Depression^{T-1}) + \beta_{2j}(\text{facial expressions and contextual factors}) + r_{ij}$). By controlling for mental health indicators on the previous day, we will test whether *change* in facial expressions or context is associated with a *change* in mental health.

Ethical considerations

The applicant is an expert in mental health, psychological resilience and multilevel modeling. Given the sensitive nature of the data that will be collected for this project, in line with the GDPR, a Data Protection Impact Assessment (DPIA) will be conducted in which all data flows will be properly mapped and possible privacy risks will be assessed together with a privacy officer from the Digital Competence Centre (DCC). All digital materials will be stored on a secured drive, and in the informed consent we will fully disclose which data we will harvest to help potential participants decide whether they want to apply.

Innovation, risk, and groundbreaking potential

This innovative application of Machine Learning has groundbreaking potential because it allows for scaling up data collection in a way that is perceived as less burdensome than diary studies. Moreover, determining the critical factors for the development of psychological problems from everyday photos can help protect youth, and can inform prevention programs for mental illness and suicide. However, there are several small risks: youth may not want to share their photo materials or fail to complete the daily diary, and the analyses may not yield factors that are strong enough predictors of youths’ mental health.

Work plan

Month:	1	2	3	4	5	6	7	8	9	10	11	12
RDMP & Ethics review	█	█										
Study preparation			█	█								
Data collection					█	█	█	█	█			
Peer-reviewed article										█	█	█

The study will be pre-registered and all materials will be made FAIR and open on Open Science Framework (OSF). Its results will be presented on at least one conference and published Open Access in a relevant peer-reviewed journal.

Literature references

1. Bor, W., Dean, A. J., Najman, J., & Hayatbakhsh, R. (2014). Are child and adolescent mental health problems increasing in the 21st century? A systematic review. In *Australian and New Zealand Journal of Psychiatry* (Vol. 48, Issue 7, pp. 606–616). SAGE Publications Ltd. <https://doi.org/10.1177/0004867414533834>
2. Inchley, J., Currie, D., Budisavljevic, S., Torsheim, T., Jåstad, A., Cosma, A., Kelly, C., Arnarsson, Á. M., Barnekow, V., & Weber, M. M. (2020). Spotlight on adolescent health and well-being: Findings from the 2017-2018 Health Behaviour in School-Aged Children (HBSC) survey in Europe and Canada. World Health Organization.
3. RIVM. (2021). "Mentale gezondheid jongeren." Retrieved on March 20th 2023 from: <https://www.rivm.nl/gezondheidsonderzoek-covid-19/kwartaalonderzoek-jongeren/mentale-gezondheid>.
4. NJI. (2023) <https://www.nji.nl/cijfers/overlijden-kind#:~:text=In%20de%20leeftijdsgroep%2015%2D20,2020%20naar%201.001%20in%202021>.
5. OECD/European Union. (2022). Health at a glance: Europe 2022: State of health in the EU cycle.
6. Dhelim, S., Chen, L., Ning, H., & Nugent, C. (2022). Artificial intelligence for suicide assessment using Audiovisual Cues: a review. *Artificial Intelligence Review*, 1-28.
7. Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, 10, 1178222618792860. doi:10.1177/1178222618792860
8. Het Social Media onderzoek - Newcom. (2023, 31 januari). Newcom Research & Consultancy Onderzoeksbureau. <https://www.newcom.nl/product/nationale-social-media-onderzoek-basis-2023/>
9. Laksana, E., Baltrušaitis, T., Morency, L. P., & Pestian, J. P. (2017, May). Investigating facial behavior indicators of suicidal ideation. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 770-777). IEEE.
10. Cleary M, Lees D, Sayers J. Friendship and Mental Health. *Issues Ment Health Nurs*. 2018 Mar;39(3):279-281. doi: 10.1080/01612840.2018.1431444. Epub 2018 Feb 21. PMID: 29465280.
11. Birch, J., Rishbeth, C., & Payne, S. R. (2020). Nature doesn't judge you—how urban nature supports young people's mental health and wellbeing in a diverse UK city. *Health & Place*, 62, 102296.
12. Rodriguez-Ayllon M., Neumann A., Hofman A., Voortman, T., Lubans, D.R., Yang-Huang, J., Jansen, P.W., Raat, H., Vernooij, M.W., & Muetzel, R.L. Neurobiological, Psychosocial, and Behavioral Mechanisms Mediating Associations Between Physical Activity and Psychiatric Symptoms in Youth in the Netherlands. *JAMA Psychiatry*. Published online March 29, 2023. doi:10.1001/jamapsychiatry.2023.0294
13. Baltrušaitis, T., Robinson P., & Morency, L.P. "OpenFace: An open source facial behavior analysis toolkit," 2016 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, 2016, pp. 1-10, doi: 10.1109/WACV.2016.7477553
14. Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H.U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research*, 21(3), 169–184. <https://doi.org/10.1002/mpr.1359>
15. Vongtangton, P., Kiatrungrit, K., Treebupachatsakul, T., & Poomrittigul, S. Predicting Self-Harm and Suicidal Behaviors by Machine Learning Models Using Facebook Data. *Available at SSRN 4331994*.