



HAL
open science

Predicting suicidal ideation from irregular and incomplete time series of questionnaires in a smartphone-based suicide prevention platform: a pilot study

Gwenolé Quellec, Sofian Berrouiguet, Margot Morgiève, Jonathan Dubois, Marion Leboyer, Guillaume Vaiva, Jérôme Azé, Philippe Courtet

► To cite this version:

Gwenolé Quellec, Sofian Berrouiguet, Margot Morgiève, Jonathan Dubois, Marion Leboyer, et al.. Predicting suicidal ideation from irregular and incomplete time series of questionnaires in a smartphone-based suicide prevention platform: a pilot study. *Scientific Reports*, 2024, 14 (1), pp.20870. 10.1038/s41598-024-71760-1 . hal-04696762

HAL Id: hal-04696762

<https://u-paris.hal.science/hal-04696762v1>

Submitted on 13 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



OPEN

Predicting suicidal ideation from irregular and incomplete time series of questionnaires in a smartphone-based suicide prevention platform: a pilot study

Gwenolé Quéllec¹✉, Sofian Berrouiguet^{1,2}, Margot Morgiève^{3,4,5,6}, Jonathan Dubois⁷, Marion Leboyer^{8,9,10}, Guillaume Vaiva^{11,12,13}, Jérôme Azé¹⁴ & Philippe Courtet^{4,7,8}

Over 700,000 people die by suicide annually. Collecting longitudinal fine-grained data about at-risk individuals, as they occur in the real world, can enhance our understanding of the temporal dynamics of suicide risk, leading to better identification of those in need of immediate intervention. Self-assessment questionnaires were collected over time from 89 at-risk individuals using the EMMA smartphone application. An artificial intelligence (AI) model was trained to assess current level of suicidal ideation (SI), an early indicator of the suicide risk, and to predict its progression in the following days. A key challenge was the unevenly spaced and incomplete nature of the time series data. To address this, the AI was built on a missing value imputation algorithm. The AI successfully distinguished high SI levels from low SI levels both on the current day (AUC = 0.804, F1 = 0.625, MCC = 0.459) and three days in advance (AUC = 0.769, F1 = 0.576, MCC = 0.386). Besides past SI levels, the most significant questions were related to psychological pain, well-being, agitation, emotional tension, and protective factors such as contacts with relatives and leisure activities. This represents a promising step towards early AI-based suicide risk prediction using a smartphone application.

With over 700,000 suicides annually and 20 times that number in suicide attempts, suicidal behavior remains a significant global health issue¹, which resists efforts in prevention and treatment. Clinicians are still facing the current impossibility to predict the occurrence of suicidal thoughts and behavior in at risk patients². Recently, smartphone-based solutions have emerged to monitor suicide risk^{3,4}, with the aim to detect in real time the potential for suicidal gesture within a short period of time. These technologies are widely available and easily leveraged to collect real-time ecological momentary assessment (EMA) data, which refers to actively asking questions via smartphone. When an imminent risk is detected, the patient can be offered an immediate preventive intervention called Just-in-time adaptive interventions⁵. During the suicidal crisis, JITAs would deliver an intervention such as a safety planning intervention, which has been largely proven to prevent suicidal behavior⁶.

By providing an accurate depiction of the patient's symptoms⁷, EMA enhances understanding of the temporal dynamics of suicide risk⁸. Indeed, in recent years, a growing number of studies provided relevant new findings about the nature and short-term predictors of suicidal thoughts and behavior using smartphone-based EMA⁹. Recent reviews of papers focusing on intensive longitudinal data and suicidal ideation (SI) revealed that suicidal

¹Inserm, UMR 1101, LaTIM, IBRBS building, 22 avenue Camille Desmoulins, 29200 Brest, France. ²Department of Psychiatry, CHU Brest, Brest, France. ³Université Paris Cité, CNRS, Inserm, Cermes3, Paris, France. ⁴Department of Emergency Psychiatry and Acute Care, Lapeyronie Hospital, CHU Montpellier, Montpellier, France. ⁵ICM - Paris Brain Institute, Hôpital de la Pitié-Salpêtrière, Paris, France. ⁶GEPS - Groupement d'Étude et de Prévention du Suicide, Paris, France. ⁷IGF, Univ. Montpellier, CNRS, INSERM, Montpellier, France. ⁸Fondation Fondamental, Hôpital Albert-Chenevier, Créteil, France. ⁹Faculté de Médecine, Institut National de la Santé et de la Recherche Médicale, Université Paris-Est Créteil, Créteil, France. ¹⁰Assistance Publique Hôpitaux de Paris, Pôle de Psychiatrie et Addictologie, Hôpitaux Universitaires Henri Mondor, Créteil, France. ¹¹CHU Lille, Hôpital Fontan, Department of Psychiatry, Lille, France. ¹²Centre National de Ressources and Résilience pour les Psychotraumatisme, Université de Lille, Lille, France. ¹³CNRS UMR-9193, SCALab - Sciences Cognitives et Sciences Affectives, Université de Lille, Lille, France. ¹⁴LIRMM, CNRS, Univ Montpellier, Montpellier, France. ✉email: gwenole.quellec@inserm.fr

thinking is associated with negative affect, impulsivity, sleep disturbances, daily social support, and adaptive coping strategies, both contemporaneously (measured at the same time point) and temporally (negative affect is measured before suicidal thinking)^{10,11}.

The best ally of these technical advancements in improving suicide prediction is the use of so-called machine learning techniques. The rapid growth of studies using this statistical methodology in the field of suicide is promising. While studies using all available data in the medical record show generally strong accuracy in detection of suicidal thoughts and behaviors¹², machine learning might be applied to intensive longitudinal data in order to identify which factors, among a large set of factors assessed over time, are most relevant to imminent suicide risk¹⁰. Noteworthy, Czyz et al.¹³ recently reported that a four-times-daily EMA assessment enabled good prediction of next-day SI (AUC: 0.84 [0.02]) and Lei et al.¹⁴ reported that a 25-day EMA yielded acceptable prediction performance for SI at 1, 3, and 8 months.

EMMA (Ecological Mental Momentary Assessment) is a smartphone-based application that integrates EMA and EMI (ecological momentary intervention)^{15,16}. It was developed in recent years to improve the identification of high-risk individuals needing immediate interventions¹⁵. The EMA module of EMMA includes questionnaires submitted daily, weekly and monthly to evaluate the user's current and recent mental state and inquire about their activities; spontaneous questionnaires can also be submitted. The EMI module provides access to emergency phone numbers and preset contact numbers of relatives, offering timely suggestions to the user. Additionally, it can propose mindfulness sessions, soothing images, or music to help the user.

As stated by Kleiman et al., ecological momentary assessment and machine learning “might become solutions to this problem” of “the ability to predict and prevent suicidal thoughts and behaviours”¹⁰. We hypothesize that automatically analyzing of data collected through the EMA module could help identify early risk factors of suicide, starting with suicidal ideation (SI). Detecting high levels of SI or predicting the onset of high SI levels automatically could trigger immediate interventions via the EMI module. This study aims to automatically estimate current SI levels and predict future SI levels using past and current answers to the EMA questionnaires.

However, analyzing EMA data poses several challenges. First, participants receive different sets of questionnaires at different intervals (e.g., daily, weekly and monthly). Second, participants do not always complete all questionnaires, and when they do, they may submit them late. Third, not all questions within a questionnaire are answered, with completion rates varying by the position of the question within the questionnaire¹⁶. Consequently, the time series of variables (i.e., answers) to analyze are unevenly spaced (i.e., irregular) and highly incomplete (i.e., partially observed).

To address these challenges, we investigate multivariate time series imputation methods^{17,18}. Given a time series of partially observed variables, the goal is to fill the missing values using the observed values (for this variable and others). Traditional solutions include replacing any missing value with the last value observed for that variable (LOCF: Last Observation Carried Forward) or with the median value of that variable¹⁷. These solutions do not leverage multivariate time series analysis. With the advent of deep learning, new solutions have emerged that consider past (and optionally future) observations of all variables. One set of solutions is based on recurrent neural networks (RNN)^{19,20}. For instance, BRITS imputes missing values according to hidden states from bidirectional RNNs while considering correlations between variables¹⁹. Other solutions utilize generative models, such as generative adversarial networks (GAN)^{21,22} and variational autoencoders (VAE)^{23,24}. These solutions offer improved speed over previous RNN-based solutions but training is more challenging. The latest generation of solutions relies on transformers^{18,25,26}, which alleviates both issues. An example is SAITS, which captures both temporal dependencies and feature correlations between time steps using diagonally-masked self-attention¹⁸.

Theoretically, all those deep learning algorithms can address the target task: imputing missing levels of SI using 1) past answers to the SI question and 2) past (and optionally present) answers to other questions. Three algorithms are investigated in this paper: BRITS, the original Transformer and SAITS. However, these approaches tackle a more general problem: imputing missing answers to any question. We hypothesize that a better imputation model for the variable of interest, namely the self-reported SI level, can be developed if training is focused specifically on that variable. Therefore, in this paper, we also propose variations on these algorithms with loss functions tailored to accurately impute SI levels. These modified algorithms are referred to as mBRITS, mTransformer and mSAITS, respectively. The source code was shared on GitHub under the GPL-v3 license (<https://github.com/leto-atreides-2/pyopts-1-feature-imputation>).

Material and methods

Participant recruitment

Participants were recruited from one of four University Hospitals centers in France (Brest, Montpellier, Lille, and Creteil). They were either under psychiatric consultation or received in psychiatric emergency services. The inclusion criteria were: 18 years of age or older, having experienced suicidal ideation or suicide attempt within the last week, the ability to understand the study's nature, purpose, and methodology, and owning an Android or iOS smartphone. Participants were monitored for six months, with four scheduled visits with a psychiatrist (at inclusion, and after 1, 3 and 6 months). This is a prospective, longitudinal, and multicentric research study authorized by the French Health Ministry (ANSM, November 30, 2017) and approved by the Est IV Ethical Committee for the Protection of Patients (October 10, 2017). Trial registration: NCT03410381. The study was conducted in accordance with the guidelines of the Declaration of Helsinki. Written informed consent was obtained from all participants for the inclusion in the study.

Questions

Participants were asked to complete EMA questionnaires at different frequencies: daily, weekly, monthly, or spontaneously. A socio-demographic questionnaire was also required at the time of inclusion. The data was collected

through the EMMA smartphone application^{15,16}. Different questions were asked in each series of questionnaires (series 1: daily or spontaneous questionnaires, series 2: weekly questionnaires, series 3: monthly questionnaires). The SI level was asked (but not necessarily answered) daily and weekly. Participants were invited to subjectively assess this level using items 7–11 of the Columbia-Suicide Severity Rating Scale (C-SSRS)²⁷; the sum of these 5 items was then converted to a scale from 0 (nonexistent) to 10 (extreme). Similar subjective questions about appetite, agitation, emotional tension, pleasure in activities, physical pain, etc., were asked daily, weekly, and monthly. Single-choice and multiple-choice questions addressed topics like places visited and current relationships. Yes or no questions were also asked: have you experienced a particularly negative event that upset or distressed you? Have you experienced any conflicts? Etc. Some questions were conditional; for example, if a participant experienced a conflict, they were asked to specify with whom (family, friends, colleagues, or others). Certain questions pertained to the current day, while others related to the past week. Excluding free-text questions, which were not analyzed in this paper, a total of 93 questions (potentially redundant) were asked across the different types of questionnaires.

The number Q of questions used in the analysis is defined as a hyperparameter that requires tuning (refer to the “Hyperparameter analysis” section). By convention, the Q -th question is defined as the SI level, which is the variable we aim to predict. All other questions are sorted by decreasing number of answers in the dataset, and the first $Q - 1$ are selected. These are defined as the $Q - 1$ input variables for the SI level predictor.

Time series preprocessing (see Fig. 1)

To allow analysis by imputation algorithms, EMA data collected from each participant were resampled. One multivariate signal X_n was built for the n -th participant as follows. The signal starts ($t = 0$) at the participant’s inclusion date. X_n is designed to be evenly spaced by (1) defining constant time intervals of Δ days and (2) aggregating all answers to the same question within each interval. The way answers are aggregated depends on the type of questions:

- It is done through averaging for subjective scales,
- Through “logical OR” for yes/no questions,
- Through majority voting for single-choice questions;
- Multiple-choice questions were treated as multiple yes/no questions.

The Δ parameter thus controls the temporal resolution of the analysis. Although evenly spaced, signal X_n may be partially observed: there is no guarantee that each question received at least one answer in each time interval. Therefore, an observation mask M_n was also defined. For each question q , if no answer was recorded in the t -th time interval, then the q -th variable is considered missing: $M_n(t, q) = 0$. Otherwise, the variable is observed: $M_n(t, q) = 1$ and $X_n(t, q)$ is computed (by aggregating multiple answers when applicable, as explained above). Examples of resampled SI level signals are reported in Fig. 1.

SAITS algorithm (see Fig. 2)

Next, those time series are processed by an imputation algorithm to predict SI. In this paper, we investigate the original and a modified version of SAITS (Self-Attention-based-Imputation for Time Series) for this purpose¹⁸. In SAITS, missing value imputation is supervised: (1) during training, a percentage of randomly-selected observed values (20%) are artificially masked in each multivariate signal and (2) a neural network is trained to correctly predict masked values. This task is called the *Masked Imputation Task* (MIT).

Let I_n indicate artificially masked variables: $I_n(t, q) = 1$ if $X_n(t, q)$ is artificially masked, 0 otherwise. The time series fed to the neural network, after artificial masking, is denoted by \hat{X}_n . The MIT is solved by minimizing the mean absolute error (MAE) between the original signal X_n and the predicted signal \tilde{X}_n , for variables indicated by I_n , i.e., $E(\tilde{X}_n, X_n, I_n)$, where:

$$E(\tilde{X}, X, \delta) = \frac{\sum_{t=1}^T \sum_{q=1}^Q |[\tilde{X}(t, q) - X(t, q)] \times \delta(t, q)|}{\sum_{t=1}^T \sum_{q=1}^Q \delta(t, q)}, \quad (1)$$

and $\delta \in \{0, 1\}^{T \times Q}$ is a binary signal ($\delta = I_n$ in this case).

The neural network predicts a value for each variable, including the observed variables. Therefore, for consistency, it is also trained to solve an *Observed Reconstruction Task* (ORT): the predicted value of observed variables (i.e., their reconstructions) are instructed to be as close as possible to the actual observed values. Let \hat{M}_n indicate the observed value signal: $\hat{M}_n(t, q) = 1$ if the value of $\hat{X}_n(t, q)$ is observed, 0 if it is either missing or artificially masked. The ORT task is achieved by minimizing the MAE between the original signal X_n and the predicted signal \tilde{X}_n , for variables indicated by the observed value signal \hat{M}_n , i.e., $E(\tilde{X}_n, X_n, \hat{M}_n)$.

As for the neural network architecture used in SAITS (illustrated in Fig. 2), the main building blocks are two Diagonally-Masked Self-Attention (DMSA) blocks. DMSA is a variation on the conventional self-attention mechanism²⁵, where the attention term for features extracted at the same time step is set to zero. Therefore, input values at the same time step are prohibited from contributing to their own estimations. The goal is to explicitly capture both the temporal dependencies and feature correlations between time steps, which is intended to improve training speed and imputation performance. In SAITS, two DMSA blocks are cascaded to progressively learn more accurate representations: X'_n and then X''_n . A final representation X'''_n is obtained by dynamically

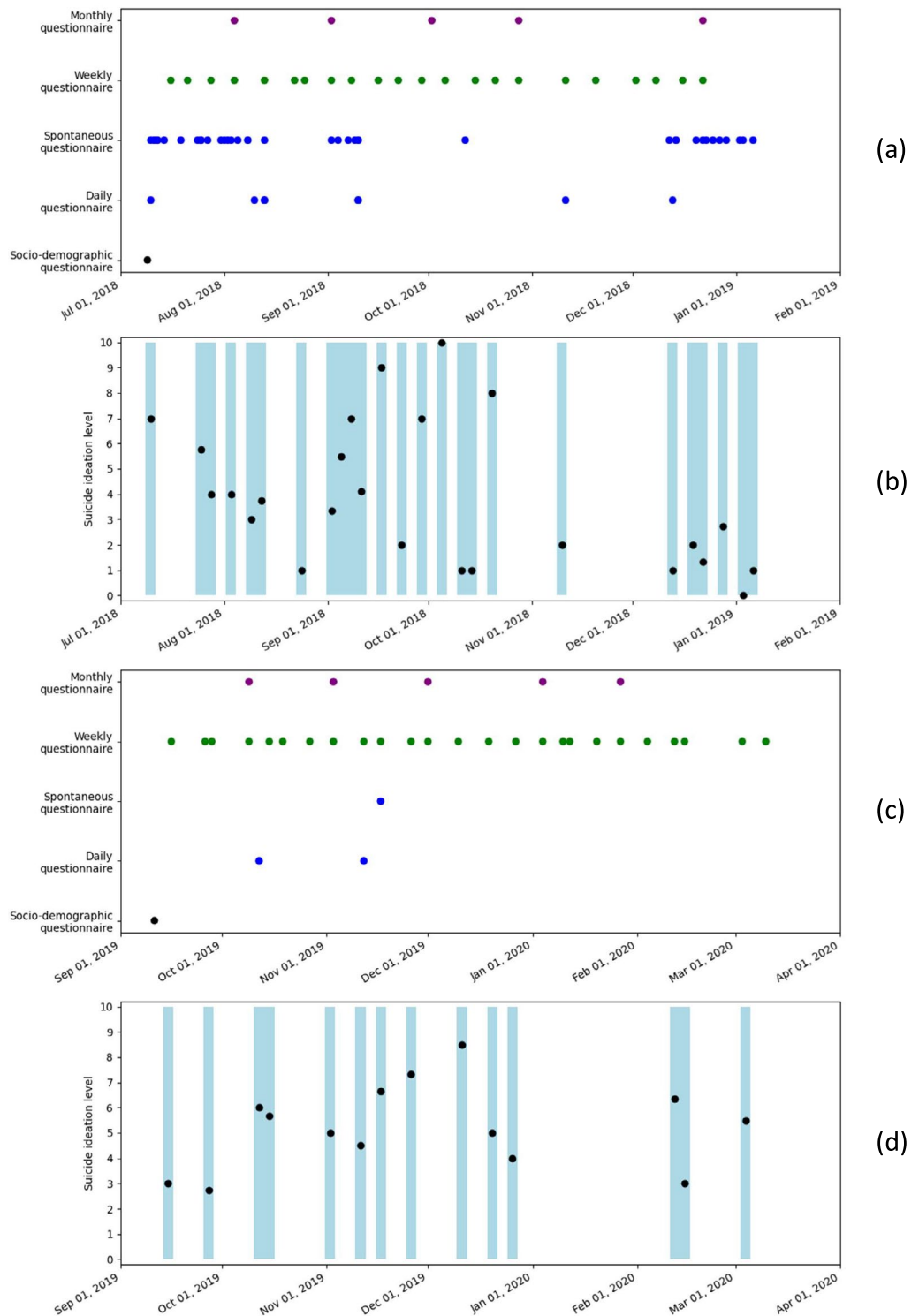


Fig. 1. Examples of questionnaire timelines (a and c) with suicidal ideation signals (b and d, respectively). In (a) and (c), dots indicate dates where questionnaires were filled. In b and d, they indicate average suicidal ideation levels $X_n(t, Q)$; the observation masks $M_n(t, Q)$ are in blue.

combining X'_n and X''_n . \tilde{X}_n is obtained by replacing missing values in the input signal by the corresponding values in X''_n .

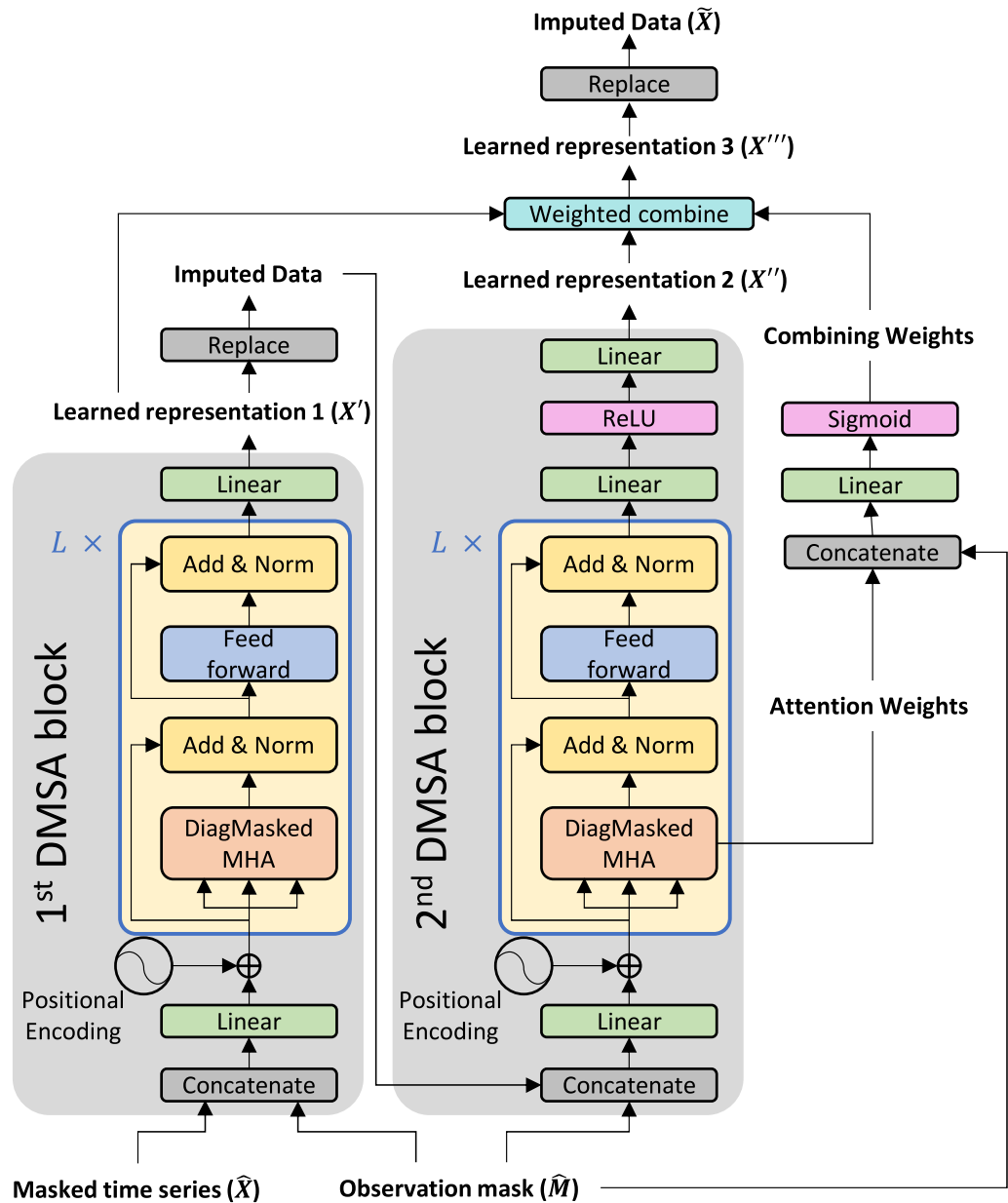


Fig. 2. SAITS neural network architecture, adapted from Du et al.¹⁸.

Suicidal ideation (SI) imputation

Estimating the Current Level of SI: Once the SAITS algorithm is trained, it can be used to estimate the current level of SI at time t , namely $X_n(t, Q)$, all other variables currently or previously inputted by the participant being known. For performance evaluation, $X_n(t, Q)$ is masked out, as well as all variable values $X_n(t', q), t + 1 \leq t' \leq T, 1 \leq q \leq Q$. The final learned representation $\tilde{X}_n(t, Q)$ is the estimated current level of SI.

Predicting the Next Level of SI: The same process is followed to predict the level of SI in the next time interval, namely $X_n(t + 1, Q)$. In that case, for performance evaluation, all variable values $X_n(t', q), t + 1 \leq t' \leq T, 1 \leq q \leq Q$ are masked out: $\tilde{X}_n(t + 1, Q)$ predicts the future SI level.

Focused Imputation: In this paper, we are not primarily interested in imputing all missing variables: we simply want to estimate SI (variable Q). For improved performance, we propose to compute the MAE for variable Q only, both in the MIT and ORT losses. This can be achieved by simply replacing Eq. (1) with the following:

$$E'(\tilde{X}, X, \delta) = \frac{\sum_{t=1}^T |[\tilde{X}(t, Q) - X(t, Q)] \times \delta(t, Q)|}{\sum_{t=1}^T \delta(t, Q)} \quad (2)$$

The modified algorithm is referred to as mSAITS.

Baseline algorithms

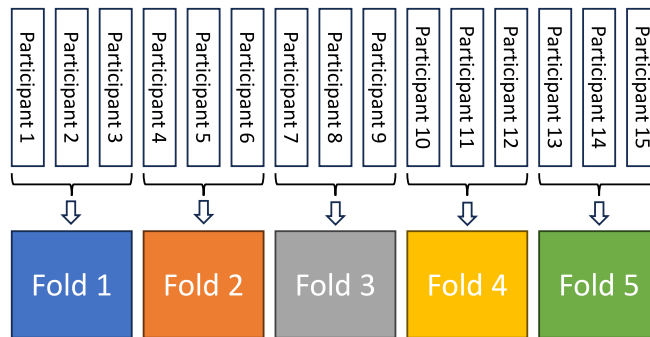
For comparison, two other variable imputation algorithms were evaluated: BRITS, based on bidirectional RNN¹⁸, and the encoder part of the original Transformer²⁵, trained through the MIT and ORT losses. Modified versions of BRITS and Transformer, namely mBRITS and mTransformer, were also evaluated: they are based on the focused MAE function of Eq. (2), plus a similarly modified forward/backward consistency loss for mBRITS. Additionally, two traditional imputation techniques were evaluated: LOCF and Median (i.e., filling missing values with median values of each variable).

Training procedure (see Fig. 3)

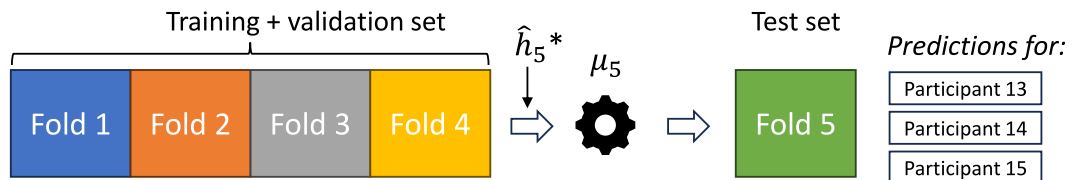
Given the limited size of the dataset, a nested cross-validation procedure, illustrated in Fig. 3, was adopted. In that purpose, the dataset was split into $K = 5$ folds: each participant was assigned at random to one of the K folds.

- *Outer loop:* K independent models $\mu_k, 1 \leq k \leq K$, were trained: each model μ_k was tested on the k -th fold; the remaining $K - 1$ folds were used for training and validation, as described hereafter.
- *Inner loop:* To train and validate model μ_k , a validation set is needed for hyperparameter optimization. To assess a set of hyperparameters, $K - 1$ submodels $\mu'_{k,l}, l \neq k$, were trained: each submodel $\mu'_{k,l}$ was validated using the l -th fold; the remaining $K - 2$ folds were used for training. Hyperparameters maximizing the mean

1. Data assignment



2. One iteration in the outer loop: training model μ_5 (for instance)



* Finding the optimal set of hyperparameters for model μ_5 (for instance)

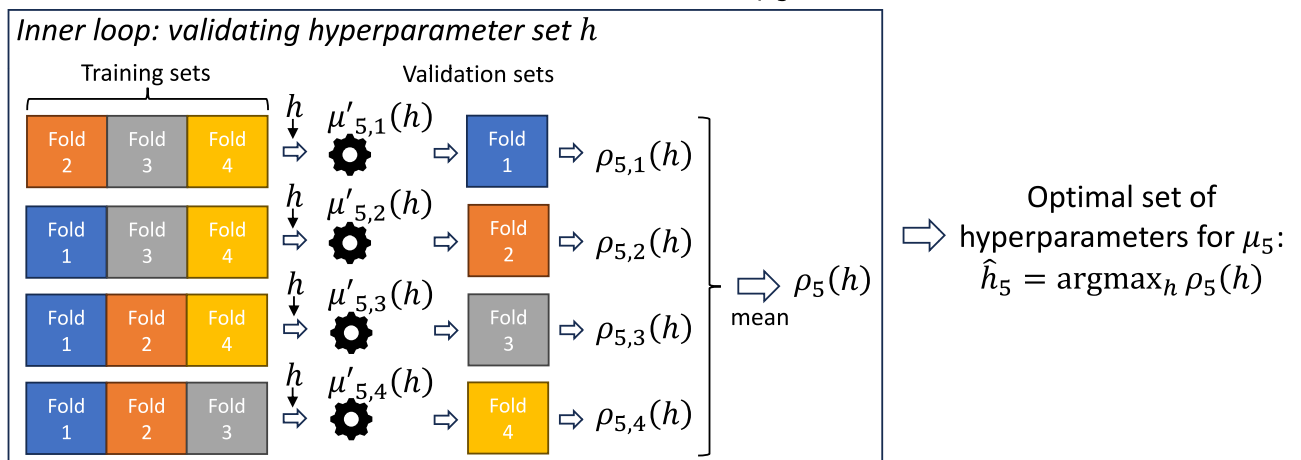


Fig. 3. Training procedure with nested cross-validation. The figure highlights the complete independence between test participants (assigned to data fold $k = 5$ in this example) and model hyperparameter optimization and training (relying solely on participants assigned to data folds $l \neq k$). Imputation models are indicated by letter μ . Hyperparameter sets are indicated by letter h . Validation scores (Spearman's ρ) are indicated by letter ρ .

of the l validation scores are selected for model μ_k . Spearman correlation (ρ) between the estimated and self-reported SI levels is used as validation score.

- *Outer loop (retraining)*: Once the optimal hyperparameters are found for the k -th split, through the inner loop, μ_k is trained with those hyperparameters using the $K - 1$ folds (all but the k -th).

K different sets of hyperparameters are thus obtained, which complicates the discussion. However, independent predictions are obtained for each participant, which gives more statistical power.

Results

Collected EMA data

A total of 99 participants were recruited. Participants were mostly female (75%). They were 30 years old on average at baseline (standard deviation: 11.5 years, minimum: 18, maximum: 58). One third of the participants were recent suicide attempters (≤ 8 days) at baseline. Smartphone application logs, containing answers to these questions, could be retrieved and analyzed for 89 participants; 10 participants did not use the application or failed to provide their logs. Each of those 89 participants submitted exactly one socio-demographic questionnaire at baseline. On average, participants submitted 8.24 daily or spontaneous questionnaires per month (standard deviation: 16.55), 3.04 weekly questionnaires per month (standard deviation: 1.74), and 0.64 monthly questionnaires per month (standard deviation: 0.31). The completion rate (i.e., the percentage of questions answered in each questionnaire) was 60% for daily or spontaneous questionnaires, 82% for weekly questionnaires and 45% for monthly questionnaires. Completion rates varied according to the question position in the questionnaire¹⁶. Finally, the distribution of self-reported SI levels in those questionnaires is reported in Fig. 4. A low (respectively high) SI level is reported in 67.8% (respectively 32.2%) of the questionnaires.

Receiver operating characteristic (ROC) analysis

Next, the ability of each imputation algorithm to detect high self-reported SI levels was assessed. Prior to the study, a subjective level ≥ 5 was defined as high by psychiatrists: this cutoff was used to classify self-reported SI levels as high (≥ 5) or low (< 5). For this binary classification problem, a receiver operating characteristic (ROC) analysis was conducted. A ROC curve is built by plotting sensitivity as a function of specificity, for various cutoffs

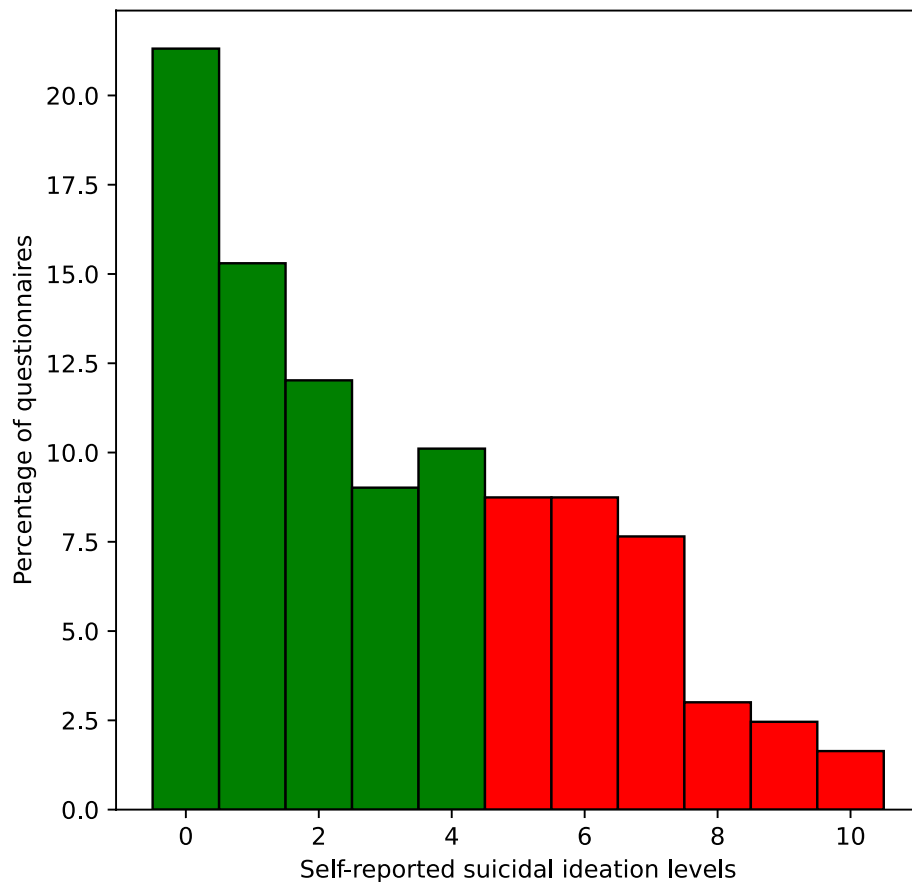


Fig. 4. Distribution of self-reported suicidal ideation (SI) levels across all questionnaires in the dataset. Green indicates low SI levels. Red indicates high SI levels.

on SI level predictions. The area under the ROC curve (AUC) was used as main score to assess the predictions of each algorithm.

Figure 5 compares these algorithms for the task of estimating the current level of SI, using a time interval of $\Delta = 3$ days. It appears that SAITS is more efficient than a simpler transformer, which is itself more efficient than the RNN-based BRITS. It also appears that the modified versions of SAITS and Transformer are more efficient than their original versions.

A detailed evaluation of the modified SAITS algorithm, for the task of estimating the current SI level and the task of predicting the SI level in the next time interval, is reported in Fig. 6. We observe that using time intervals of $\Delta = 3$ days leads to a better performance than using time intervals of $\Delta = 1$ day or $\Delta = 7$ days. The explanations could be that 1) time series have too many missing values when $\Delta = 1$, and 2) participants' state of mind varies too much in $\Delta = 7$ days. As expected, we also observe that estimating the current SI level is easier than predicting the next level, whatever the time interval, although the performance decrease is limited.

Overall, the best performance was obtained with the modified SAITS algorithm and time intervals of $\Delta = 3$ days. The performance at detecting high self-reported SI levels (≥ 5) with this algorithm was $AUC=0.804$; an interesting operating point provides a sensitivity of 86.4% and a specificity of 65.0%. The performance at predicting high self-reported SI levels, 3 days in advance, was $AUC=0.769$; however, for an equal sensitivity of 86.4%, specificity drops to 53.2%.

Additional metrics

For a detailed analysis of the predictions, independent of any high/low SI level cutoff, performance was evaluated in terms of Spearman correlation between the estimated SI levels and the self-reported SI levels. For the best algorithm, namely the modified SAITS algorithm ($\Delta = 3$ days), the following Spearman correlations were calculated: $\rho=0.535$ for current SI level estimation and $\rho=0.500$ for next SI level prediction. Scatter plots are reported in Fig. 7. The distribution of high/low SI levels is imbalanced: low self-reported SI levels are more frequent (67.8%) than high self-reported SI levels (32.2%). Therefore, two additional classification metrics, often regarded as more suitable for imbalanced datasets, were reported: the F1-score and Matthews Correlation Coefficient (MCC). The F1-score is the harmonic mean between precision and recall. MCC is a Pearson correlation coefficient computed between binary variables. Unlike ROC's AUC, a cutoff on predictions is needed to compute

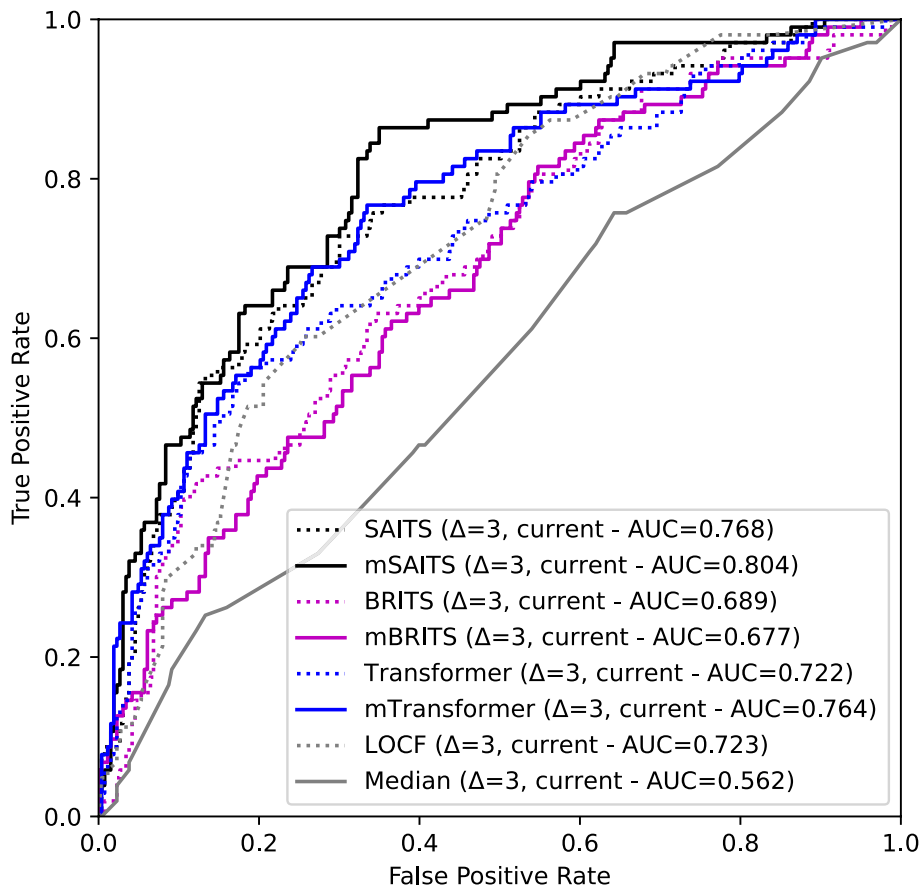


Fig. 5. Receiver-operating characteristic (ROC) analysis of various algorithms for estimating the current self-reported level of suicidal ideation (SI). Time intervals of $\Delta=3$ days are used. The 'm' prefix in method names indicates the use of modified loss functions focusing on the SI variable.

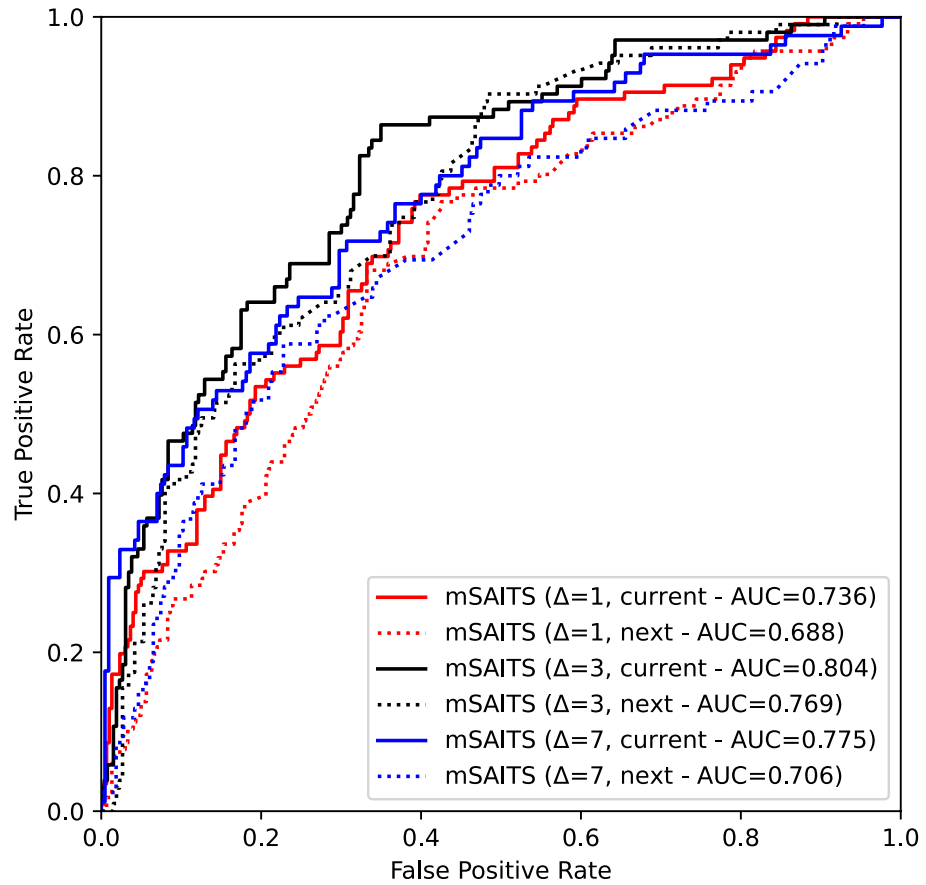


Fig. 6. ROC analysis of the modified SAITS algorithm (mSAITS) for predicting the current level of SI, as well as the SI level in the next time interval. Time intervals of $\Delta = 1, 3$ and 7 days are considered.

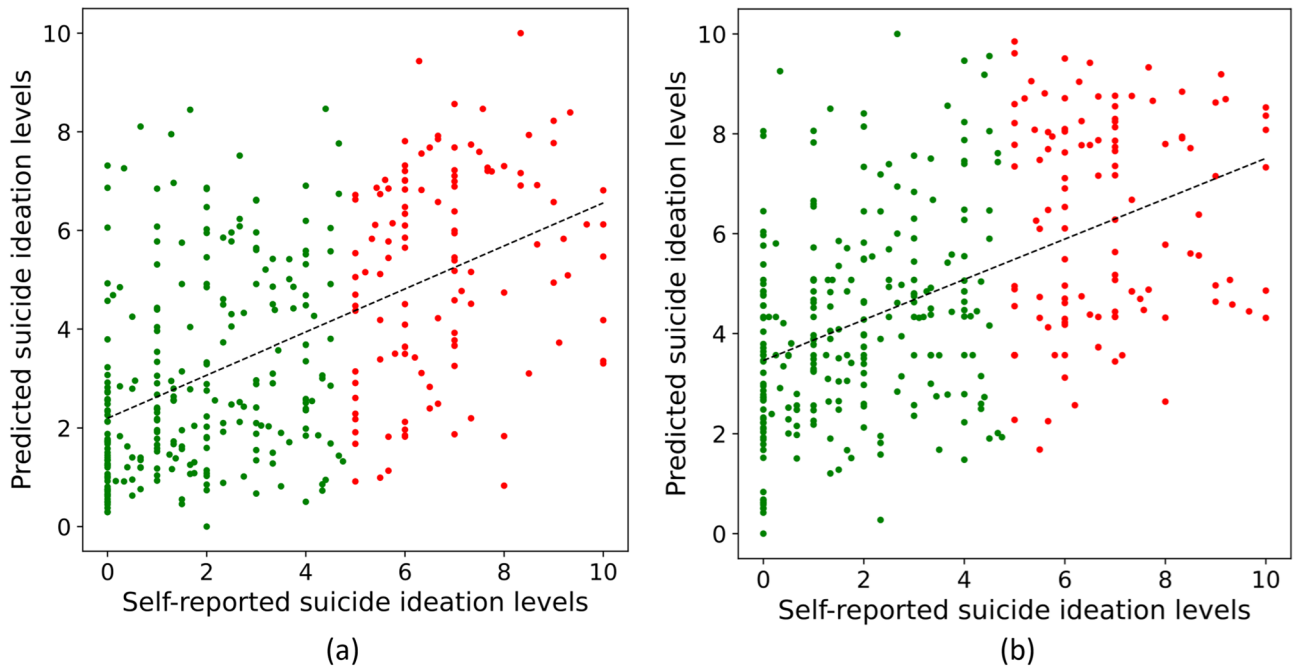


Fig. 7. Scatter plots of predicted versus self-reported SI levels, for two tasks: current SI level estimation (a) and next SI level prediction (b). Time intervals of $\Delta = 3$ days are used. Colors indicate high or low self-reported SI levels. The dashed lines represent linear regressions.

those metrics (predicted SI level ≥ 5). All metrics are reported in Tables 1 and 2 for estimating the current SI level or predicting the future SI level, respectively. It can be observed that the ranking of imputation models is not significantly impacted by the choice of a metric.

Hyperparameter analysis

As detailed in the Methods section, different hyperparameters had to be tuned, through cross-validation, to define the algorithms: (1) the number L of layers and the number H of attention heads per block, which control model complexity, and (2) the number Q of questions analyzed by the algorithm (the Q most frequent). Interestingly, more complex models were obtained for future SI prediction ($3 \leq L \leq 5$ layers per block, $4 \leq H \leq 8$ attention heads) than for current SI estimation ($1 \leq L \leq 2$ layers per block, $2 \leq H \leq 4$ attention heads). The number of questions used in both tasks, however, is similar ($12 \leq Q \leq 15$). Those questions concern the level of agitation, of emotional tension, of psychological pain, the existence of recent contacts with friends, with family, the time spent with important people, the time spent on leisure activities, and the perceived quality of life. It should be noted that a larger number of questions is used for SI estimation and prediction when considering intervals of $\Delta = 1$ day or $\Delta = 7$ days, indicating more difficult tasks.

Relevance of each question

Although the most frequent questions were considered for model development, not all of these questions are equally relevant. To assess the relevance of each question for estimating the current level of self-reported SI, the permutation importance of each variable²⁸ was assessed. The permutation importance measures how much the prediction scores decrease when a variable is missing. Table 3 reports the questions impacting Spearman's ρ the most with this method. Most of these questions relate to the subjects feel, and are answered on a 0–10 subjective scale. The relationships between those questions are the self-reported SI level, also on a 0–10 subjective scale, are illustrated in Figs. 8 and 9.

Discussion

We have presented preliminary results for estimating the current level of suicidal ideation (SI) and predicting future SI levels using time series data from a smartphone-based ecological momentary assessment (EMA) application. The classification results for distinguishing between high and low SI levels are promising, with an AUC of 0.804 for current estimation (F1 = 0.625, MCC = 0.459) and an AUC of 0.769 for three-day ahead future SI prediction (F1 = 0.576, MCC = 0.386). These results are based on a modified version of SAITS, a transformer-based algorithm for missing value imputation.

The permutation importance study (Table 1) shows that, as expected, previous SI levels are the most significant contributors to predicting the current SI level. However, this variable accounts for only half of the performance

Model	Spearman's ρ	AUC	F1-score	MCC
SAITS	0.412	0.768	0.585	0.435
mSAITS	0.535	0.804	0.625	0.459
BRITS	0.285	0.689	0.502	0.236
mBRITS	0.265	0.677	0.508	0.249
Transformer	0.337	0.722	0.552	0.373
mTransformer	0.470	0.764	0.583	0.387
Median	0.116	0.562	0.317	0.144
LOCF	0.442	0.723	0.534	0.327

Table 1. Performance of various algorithms for estimating the current self-reported level of suicidal ideation (SI). Time intervals of $\Delta = 3$ days are used. The 'm' prefix in method names indicates the use of modified loss functions focusing on the SI variable. Median and LOCF are traditional imputation techniques. Maximal values are in bold.

Time interval	Prediction horizon	Spearman's ρ	AUC	F1-score	MCC
$\Delta = 1$	current	0.360	0.736	0.552	0.338
$\Delta = 3$	current	0.535	0.804	0.625	0.459
$\Delta = 7$	current	0.467	0.775	0.590	0.343
$\Delta = 1$	next	0.328	0.688	0.536	0.316
$\Delta = 3$	next	0.500	0.769	0.576	0.386
$\Delta = 7$	next	0.384	0.706	0.561	0.311

Table 2. Performance of the modified SAITS algorithm (mSAITS) for predicting the current level of SI, as well as the SI level in the next time interval. Time intervals of $\Delta = 1, 3$ and 7 days are considered. Maximal values, for each prediction horizon, are in bold.

Question	Importance (score decrease)
<i>(Previous) level of suicidal ideation?</i>	0.277
What was your average level of psychological pain this week?	0.041
Do you currently have contact with your family?	0.029
What was your level of well-being this week?	0.021
What is your current level of psychological pain?	0.021
Did you enjoy your activities this week?	0.018
Did you have a good week?	0.013
How does your quality of life look this week?	0.011
What is your current level of agitation or emotional tension?	0.008
What was your maximum level of agitation or emotional tension this week?	0.007
What was your maximum level of psychological pain this week?	0.006
How much time (in hours) did you spend on your leisure activities this week?	0.004
How much time (in hours) did you spend with the people important to you over the past week?	0.002

Table 3. Most discriminant questions for the task of suicidal ideation estimation – permutation importance. Significant values are in italics.

score (a decrease of 0.277 out of $\rho = 0.535$). Other important questions relate to psychological pain, well-being, agitation, emotional tension, and protective factors like contacts with relatives and leisure activities. As illustrated in Figs. 8 and 9, none of these questions, when considered individually, strongly correlates with the SI level. This underscores the need for a more complex artificial intelligence (AI) model, such as the proposed mSAITS, which takes multivariate time series as input.

The reported prediction performance (AUC = 0.804 for current estimation and AUC = 0.769 for 3-day ahead prediction) is lower than that reported by Czyz et al. for next-day prediction (AUC = 0.84)¹³. This discrepancy can be partly attributed to differences in data acquisition frequency: Czyz et al.'s study involved four EMA measurements per day, while ours involved at most one per day. Another factor is participant adherence: in Czyz et al.'s study, participants were followed for only 8 weeks and were financially compensated for adherence, whereas our study observed low participant adherence, averaging 8.24 daily questionnaires per month, possibly due to longer follow-up and absence of compensation. The data missingness in our study complicates the prediction task, necessitating the use of imputation algorithms. However, this likely also makes our findings more representative of real-world conditions. The ability to predict SI in these time intervals (e.g., three days in advance) is interesting because it opens up the reasonable possibility of either intervening immediately (JITAs) or connecting the patient with care systems to offer a crisis intervention or hospital admission.

We report here that the variables contributing to the prediction of the occurrence of SI are psychological pain, well-being, agitation, emotional tension, and protective factors such as contacts with relatives and leisure activities. This is consistent with previous EMA studies, which have shown that the occurrence of suicidal ideation is associated with negative affect, psychological pain, and social support¹⁰. It also aligns with a study on patients post-discharge following hospitalization for a suicide attempt, which found that controlling for prior SI, being in the home of close others or in festive or leisure environments decreased SI probability²⁹. Our data highlights the need to include psychological pain in EMA assessments, to predict the occurrence of suicidal thoughts and behaviors. This finding aligns with a recent study by Baryshnikov et al., which reported that psychological pain is an independent risk factor for SI in depressed patients, varying within a timescale of days alongside depression and hopelessness, based on a three-times-daily EMA assessment³⁰.

One limitation of this study is the limited sample size. A total of 89 signals from 89 participants was available for training and evaluating an AI algorithm, which is quite small. The challenge was therefore to avoid overfitting. One solution to this problem was to reduce the number of parameters to train. Specifically, instead of considering all available questionnaire questions, we focused solely on the Q most common ones, resulting in more compact neural architectures. Additionally, we implemented a nested cross-validation procedure (see Fig. 3): while overly complex neural architectures might perform well on one validation set by chance, it is unlikely they would perform well on four independent validation sets, thus reducing the risk of overfitting. Despite the limited training data, the investigated AI-based approaches outperformed traditional ones that do not require any training data (AUC = 0.562 with median value imputation and AUC = 0.723 with LOCF, versus AUC = 0.804 with mSAITS). However, the performance of AI models is known to increase with the number of training samples, so we expect a significant performance increase with a larger training dataset.

In conclusion, this pilot study validates that the proposed approach to processing irregular and incomplete time series of questionnaire data works in principle, but additional work is needed to develop a clinically useful application. Although the model was trained on multicentric data, its generality on an independent population also needs to be evaluated. Once trained and evaluated on a larger and more diverse population, we anticipate that such a solution could help identify participants at risk of attempting suicide more timely, enabling ecological momentary intervention (EMI) when needed most. Mobile health solutions, enhanced by AI, are promising tools

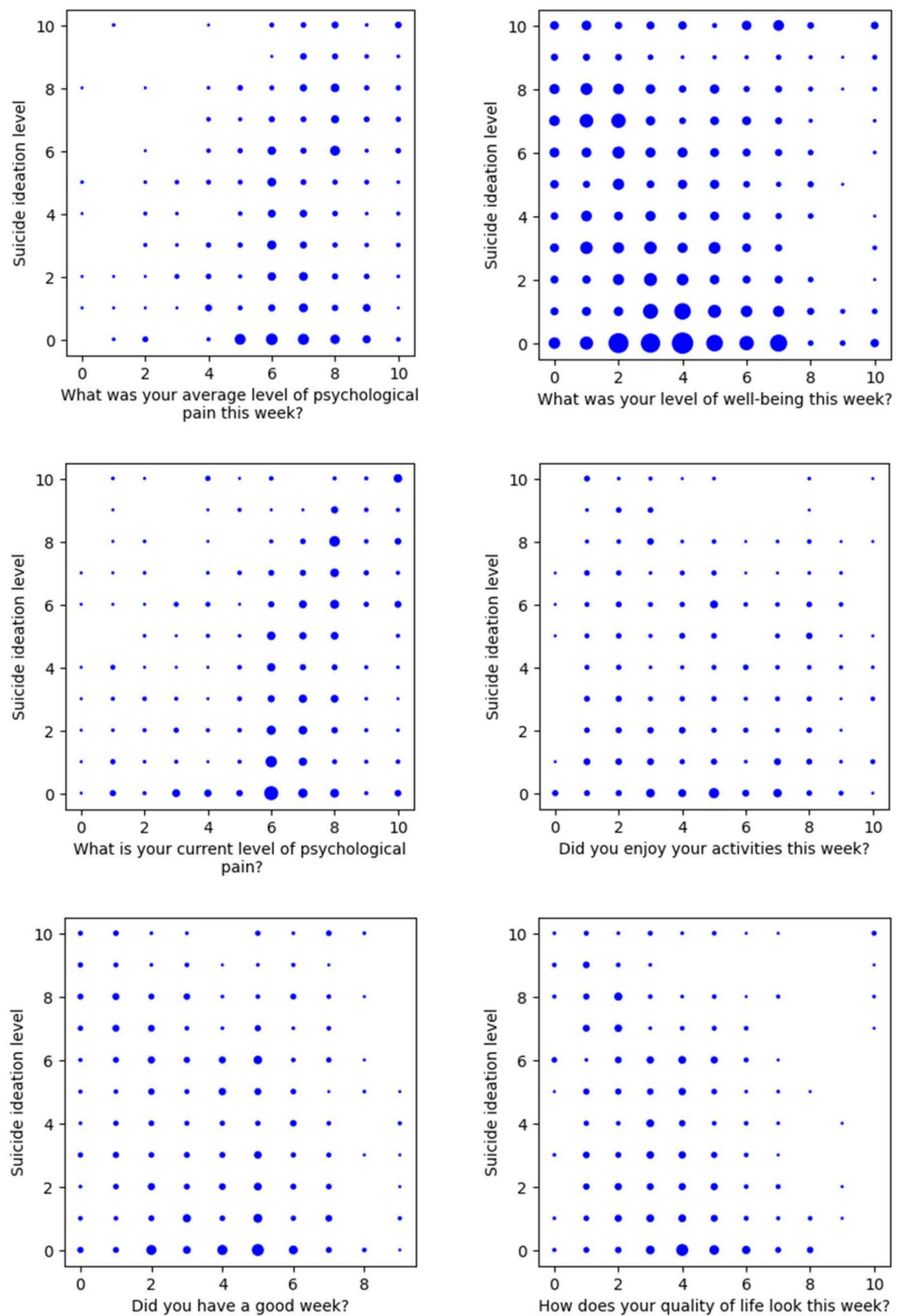


Fig. 8. Impact of the most discriminant subjective questions on the self-reported SI level – part 1. Answers are on a 0–10 subjective scale. Circle size is proportional to the number of occurrences.

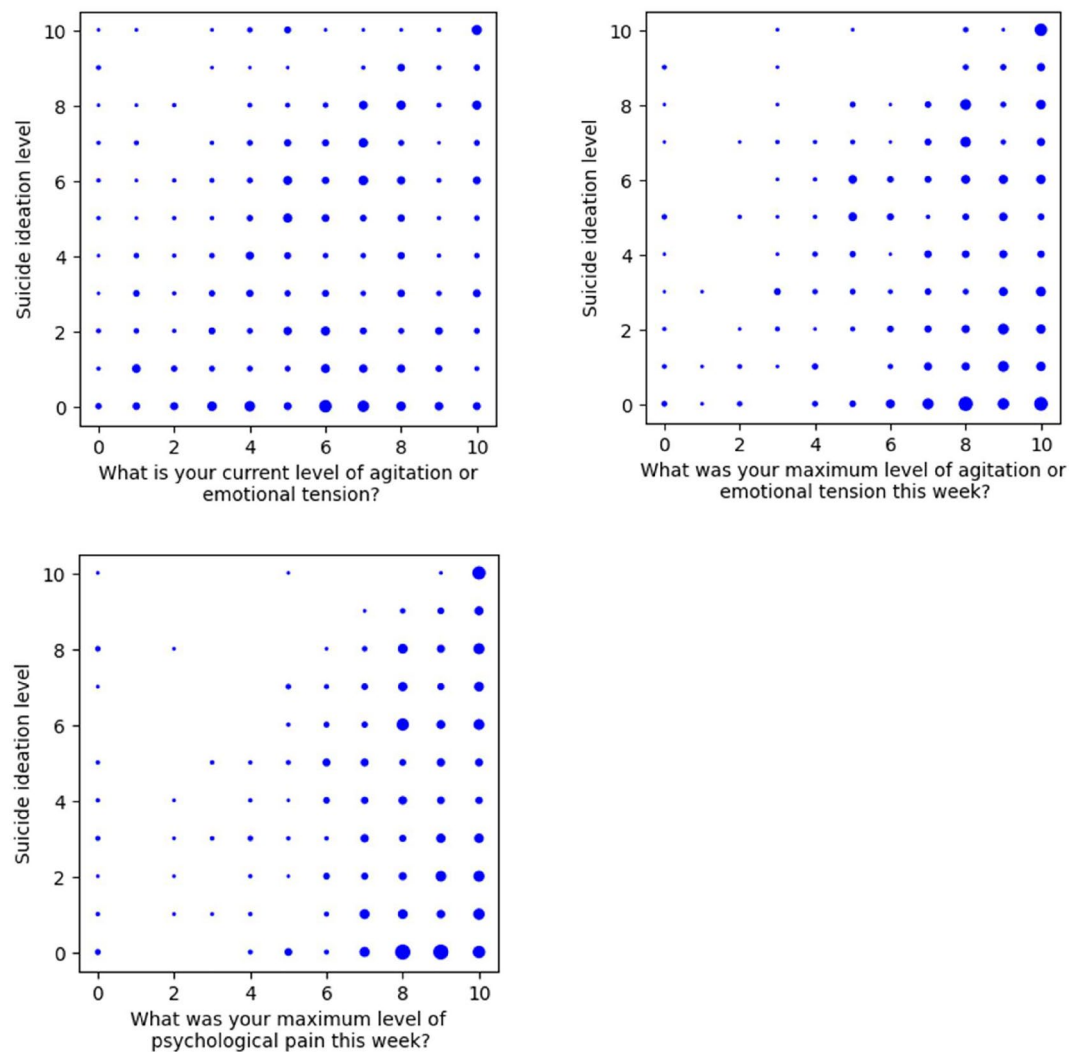


Fig. 9. Impact of the most discriminant subjective questions on the self-reported SI level – part 2.

to assist health services (i.e., the emergency departments, crisis centers, hospital services, outpatient services, and general practitioners) in preventing suicide.

Data availability

Data that were analyzed in this study will be made readily available by the corresponding author upon reasonable written request.

Received: 5 December 2023; Accepted: 30 August 2024

Published online: 06 September 2024

References

1. Suicide prevention. <https://www.who.int/health-topics/suicide> (2023). Accessed 28-08-2024.
2. Chan, M. K. Y. *et al.* Predicting suicide following self-harm: Systematic review of risk factors and risk scales. *Br. J. Psychiatry J. Ment. Sci.* **209**, 277–283 (2016).
3. Martinengo, L. *et al.* Suicide prevention and depression apps' suicide risk assessment and management: A systematic assessment of adherence to clinical guidelines. *BMC Med.* **17**, 231 (2019).
4. Torous, J. *et al.* Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: A review of current progress and next steps. *Curr. Psychiatry Rep.* **20**, 51 (2018).
5. Nahum-Shani, I. *et al.* Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Ann. Behav. Med. Publ. Soc. Behav. Med.* **52**, 446–462 (2018).
6. Nuij, C. *et al.* Safety planning-type interventions for suicide prevention: Meta-analysis. *Br. J. Psychiatry J. Ment. Sci.* **219**, 419–426 (2021).
7. Hilty, D. M., Chan, S., Hwang, T., Wong, A. & Bauer, A. M. Advances in mobile mental health: Opportunities and implications for the spectrum of e-mental health services. *mHealth* **3**, 34 (2017).
8. Selby, E. A., Yen, S. & Spirito, A. Time varying prediction of thoughts of death and suicidal ideation in adolescents: Weekly ratings over six month follow-up. *J. Clin. Child Adolesc. Psychol.* **42**, 481–495 (2013).
9. Barrigon, M. L., Courtet, P., Oquendo, M. & Baca-García, E. Precision medicine and suicide: An opportunity for digital health. *Curr. Psychiatry Rep.* **21**, 131 (2019).

10. Kleiman, E. M., Glenn, C. R. & Liu, R. T. The use of advanced technology and statistical methods to predict and prevent suicide. *Nat. Rev. Psychol.* **2**, 347–359 (2023).
11. Kivelä, L., van der Does, W. A. J., Riese, H. & Antypa, N. Don't miss the moment: A systematic review of ecological momentary assessment in suicide research. *Front. Digit. Health* **4**, 876595 (2022).
12. Kusuma, K. *et al.* The performance of machine learning models in predicting suicidal ideation, attempts, and deaths: A meta-analysis and systematic review. *J. Psychiatr. Res.* **155**, 579–588 (2022).
13. Czyz, E. K. *et al.* Ecological momentary assessments and passive sensing in the prediction of short-term suicidal ideation in young adults. *JAMA Netw. Open* **6**, e2328005 (2023).
14. Lei, C., Qu, D., Liu, K. & Chen, R. Ecological momentary assessment and machine learning for predicting suicidal ideation among sexual and gender minority individuals. *JAMA Netw. Open* **6**, e2333164 (2023).
15. Morgiève, M. *et al.* A digital companion, the emma app, for ecological momentary assessment and prevention of suicide: Quantitative case series study. *JMIR MHealth UHealth* **8**, e15741 (2020).
16. Morgiève, M. *et al.* Acceptability and satisfaction with emma, a smartphone application dedicated to suicide ecological assessment and prevention. *Front. Psychiatry* **13**, 952865 (2022).
17. Acuña, E. & Rodríguez, C. The treatment of missing values and its effect on classifier accuracy. In *Classification, Clustering, and Data Mining Applications* (eds Banks, D. *et al.*) 639–647 (Springer, 2004).
18. Du, W., Cote, D. & Liu, Y. SAITS: Self-attention-based imputation for time series. Preprint at <https://doi.org/10.48550/arXiv.2202.08516> (2022).
19. Cao, W. *et al.* BRITS: Bidirectional recurrent imputation for time series. In *Advances in Neural Information Processing Systems* (eds Bengio, S. *et al.*) (Curran Associates Inc, 2018).
20. Yoon, J., Zame, W. R. & van der Schaar, M. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. *IEEE Trans. Biomed. Eng.* **66**, 1477–1490 (2019).
21. Liu, Y., Yu, R., Zheng, S., Zhan, E. & Yue, Y. NAOMI: Non-autoregressive multiresolution sequence imputation. In *Advances in Neural Information Processing Systems* (eds Wallach, H. *et al.*) (Curran Associates Inc, 2019).
22. Luo, Y., Cai, X., Zhang, Y., Xu, J. & Xiaojie, Y., Multivariate time series imputation with generative adversarial networks. In *Advances in Neural Information Processing Systems* (eds Bengio, S. *et al.*) (Curran Associates Inc, 2018).
23. Fortuin, V., Baranchuk, D., Raetsch, G. & Mandt, S. 2020, GP-VAE: Deep probabilistic time series imputation. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (eds Chiappa, S. & Calandra, R.), PMLR, 1651–1661.
24. Ramchandran, S., Tikhonov, G., Kujanpää, K., Koskinen, M. & Lähdesmäki, H. 2021, Longitudinal variational autoencoder. In *Proceedings of The Twenty Fourth International Conference on Artificial Intelligence and Statistics*, (Banerjee, A. & Fukumizu, K. eds.), PMLR, 3898–3906.
25. Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems* (eds Guyon, I. *et al.*) (Curran Associates Inc, 2017).
26. Shan, S., Li, Y. & Oliva, J. B. NRTSI: Non-recurrent time series imputation. Preprint at <https://doi.org/10.48550/arXiv.2102.03340> (2021).
27. Posner, K. *et al.* The Columbia-suicide severity rating scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am. J. Psychiatry* **168**, 1266–1277 (2011).
28. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
29. Husky, M. *et al.* Predictors of daily life suicidal ideation in adults recently discharged after a serious suicide attempt: A pilot study. *Psychiatry Res.* **256**, 79–84 (2017).
30. Baryshnikov, I., Rosenström, T. H. & Isometsä, E. T. Psychological pain as a risk factor for suicidal ideation: An ecological momentary assessment study on inpatients with depression with and without comorbid borderline personality disorder. *J. Clin. Psychiatry* **85**, 234926 (2024).

Acknowledgements

This study was supported by the FondaMental Chair for the prevention of suicidal behavior created by the FondaMental Foundation, led by Philippe Courtet and SNCF, SNCF réseau, and SNCF mobilité.

Author contributions

Study design: G.Q., S.B., M.M., J.A., P.C. Data acquisition: S.B., J.D., M.L., G.V., P.C. Algorithm development: G.Q. Data analysis and interpretation: P.C., S.B. Statistical analysis: G.Q. Manuscript preparation: G.Q. Manuscript review: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.Q.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024