

Leveraging Longitudinal Lifelog Data Using Survival Models for Predicting Risk of Relapse among Patients with Depression in Remission

Felan Carlo C. Garcia^{*1}, Ayumi Hirao¹, Aran Tajika², Toshi A. Furukawa²,
Kazushi Ikeda¹, and Junichiro Yoshimoto¹

Abstract—Managing depression relapse is a challenge given factors such as inconsistent follow-up and cumbersome psychological distress evaluation methods which leaves patients with a high risk of relapse to leave their symptoms untreated. In an attempt to bridge this gap, we proposed an approach on the use of personal longitudinal lifelog activity data gathered from individual smartphones of patients in remission and maintenance therapy (N=87) to predict their risk of depression relapse. Through the use of survival models, we modeled the activity data as covariates to predict survival curves to determine if patients are at risk of relapse. We compared three models: CoxPH, Random Survival Forests, and DeepSurv, and found that DeepSurv performed the best in terms of Concordance Index and Brier Score. Our results show the possibility of utilizing lifelog data as a means of predicting the onset of relapse and towards building eventual tools for a more coherent patient evaluation and intervention system.

Index Terms— Depression Prediction, DeepSurv, Random Survival Forest, Lifelog Data, Mobile Computing, Survival Analysis

I. INTRODUCTION

Depression is a severe mental disorder characterized by persistent low mood and lack of interest, which leads to a progressive diminished function and quality of daily life [1]. Another aspect of depression is its high frequency of recurrence, further exacerbated by factors such as non-adherence to medication, cumbersome evaluation methods, and inconsistent communication between physician and patient [2]. Consequently, given these factors, patients who are at risk of relapse or who forayed into depression are more likely to have their symptoms untreated [3].

Several works have been attempted to address some of these issues by providing depression evaluation through capitalizing on an individual's smartphone as a resource for predicting their mental state. For instance, Buck et al [4] has shown that significant changes in activity logs preceded schizophrenia relapse events, whereas Li et al [5] has shown that recorded subtle changes in daily behaviour can be used for predicting an individual's well-being through deep learning. Recently, Chikersal et al [6] has shown that its possible to detect post-semester depression from students from their smartphone logs such as Bluetooth activity, sleep, location, and step. Most closely related to our work is done by Kumagai et al [7], in which they established the relationship between daily activity lifelogs of patients under

depression maintenance therapy and their onset of depression relapse through a Panel VAR analysis.

These studies have shown that relating activity data to mental health state is possible. However, there are several limitations that these studies failed to address: First, non-linearity have not been considered in the dataset. Second, there is a gap in how to predict the risk of an onset or relapse of a disease for an individual and how this risk evolve over time using activity logs. Finally, a limited observation window introduces the problem of data censorship where events of interest (e.g depression relapse) may not have been observed for several patients. The issue lies in that unobserved relapse events do not imply non-occurrence for a patient's lifetime. Therefore, using classic machine learning approaches results in biased estimates and false interpretation as these methods handle censored data differently [8].

In this study, we present an approach for predicting a patient's likelihood of having a relapse by leveraging their smartphone longitudinal lifelog data using survival models to determine their risk of relapse onset. Lastly, we tested these models on unseen patient data and compared their performance on how well they predict patient-specific survival curves.

II. METHODS

This is a second analysis of the observational study in [7]

A. Ethics Statement

The study protocol was approved by the ethics committees/institutional review boards of Kyoto University Graduate School of Medicine (R0591-6), Nagoya City University Hospital (50-16-0001), Kochi Medical School (28-67), Hiroshima University Hospital (C-105-3), Toho University School of Medicine (A17062), and Nara Institute of Science and Technology (2017-M-5, 2017-M-6). All participants provided written informed consent.

B. Dataset

87 eligible patients with depression in remission participated in this study of which their data was collected in approximately one year (52 weeks) using the following devices: Silmee W20 wristband sensor, and Kurashi lifelog application [7]. Specifically in this study, we focus on the Kurashi application which records a semi-automatic label and self-report of 16 daily activities (sleep, bath, work, commuting, chores, shopping, hospital visit, socializing, exercise, hobby, reading, watching TV, others). For ground

¹ Nara Institute of Science and Technology
^{*}garcia.felan_carlo.gdl@is.naist.jp

² Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health

truth values, patients perform a weekly self-report for the K6 Scores and a once-a-month evaluation of their PH-9 score via a medical professional to evaluate their depression severity. See [7] for detailed protocols for data collection.

C. Survival Analysis

Survival Analysis refers to a set of methods that is concerned with modeling and analysing data involving the occurrences of events of interest observed over an interval of time [9]. One aspect that makes Survival Analysis distinct from a typical regression problem is it is able to handle censored data [8]. Common usage of survival analysis involved the use of the non-parametric Kaplan-Meier estimator to measure the survival function of an observed population given by the following:

$$\hat{S}(t) = \prod_{j:T_j < t} \left(1 - \frac{d_j}{n_j}\right). \quad (1)$$

where d_j is the events occurrences at time T_j , and n_j are the number of population instances that are observed to have survived [10]. One thing to note is that the Kaplan-Meier estimator does not have the capability to perform survival estimation of a particular instance from a given observed population and how its specific features correlates to this estimate. Given that we aim to relate activity lifelog of patients in order to predict their likelihood of relapse, we used the following survival models (Cox Proportional Hazard, Random Survival Forest, DeepSurv) that can fit a survival function given an observed population and predict an individual survival curves using patient specific features.

D. Cox Proportional Hazard

The Cox Proportional Hazard (CoxPH) [11] is one of the most used methods for predicting individual survival curve by taking the hazard function given by:

$$\lambda(t|x) = \lim_{\delta t \rightarrow 0} \frac{P(t < O \leq t + \delta t \mid O > t|x)}{\delta t}. \quad (2)$$

and modelling it as changes of risk of event occurrence over a period of time in relation to a series of feature vectors $x = (x_1, \dots, x_2)^\tau$:

$$\lambda(t \mid x) = \lambda_0(t) \exp(x^\tau \beta). \quad (3)$$

CoxPH does have some limitations, the hazard function is assumed as proportional for each individual patients' hazard function. However, on non-linear data, this assumption is often violated. For this work, we used the following parameters after performing grid search for CoxPH: learning rate = 0.00001, l2 regularization = 0.01.

E. Random Survival Forest

Random Survival Forest (RSF) and has been introduced as an alternative non-parametric model that automatically detects feature interactions and able to extend to high-dimensional and non-linear data for survival analysis [12].

In RSF, the model's objective is to split into left and right daughter nodes with dissimilar survival probability using a

Log-rank split rule which is defined with the following: given a time-to-event data of x_i, t_i, δ_i , where x_i is a feature vector, t_i is the observation time, and δ_i is the censoring indicator, a proposed split for a specific instance x is given by $x \leq c$ and $x \geq c$ which serves as a criterion for splitting the head node into left and right daughters (L and R) respectively.

Given these variable definition $Y_j = Y_{j,L} + Y_{j,R}$, and $\delta_j = \delta_{j,L} + \delta_{j,R}$, the log-rank splitting rule R is defined by the following:

$$LR(x, c) = \frac{\sum_{j=1}^m (d_{j,L} - Y_{j,L} \frac{d_j}{Y_j})}{\sqrt{\sum_{j=1}^m \frac{Y_{j,L}}{Y_j} \left(1 - \frac{Y_{j,L}}{Y_j}\right) \left(\frac{Y_j - d_j}{Y_j - 1}\right) d_j}}. \quad (4)$$

where $t_1 < \dots < t_m$ is the death times, $(d_{j,L}, d_{j,R})$ and $(Y_{j,L}, Y_{j,R})$ represents the number of deaths and individuals at risk at time t_j in the L, R daughter nodes [12].

For this model, we optimized the model with grid search, the final hyperparameters we used are the following: number of trees = 350, max depth = 5, minimum node size = 16.

F. DeepSurv

DeepSurv [13] is a multilayer perceptron architecture for survival analysis that takes an input features derive from the lifelog activity data as x and estimates non-linear proportional hazards log-risk function similar to Eq. (4) through the use of a modified log partial likelihood loss function $l(\theta)$:

$$l(\theta) = -A \sum_{i:E_i=1} \left(\hat{h}_\theta(x_i) - \log \sum_{j \in \mathcal{R}(T_i)} e^{\hat{h}_\theta(x_j)} \right). \quad (5)$$

where $A = -\frac{1}{N_{E=1}}$, $N_{E=1}$ is the number of patients with on observed relapse event.

The network is built with modern implementation of fully-connected layers and uses a Rectified Linear Unit (ReLU) activation functions which allows the network to be constructed as shallow (1 hidden layer) or as deep as possible (n -hidden layers) to accommodate a specific survival analysis application. For this study we employed a 2-layer network each followed by a dropout layer with 0.10 value, we performed grid search on the number of units in the hidden layer each with range (10,200) and iterated with a step increase of 10 units. The final layer used consisted of two-layer network with the following number of hidden units [70,50].

G. Data Processing

In this study, we specifically focused on the data from the lifelogging application as features. In addition, several features have been derived as they have been identified as risk factors [14] for depression: frequency of specific meal times (Breakfast, Lunch, and Dinner), Continuous Long Sleep Time, and Awake Time. Moreover, we processed these features and computed their sum, maximum, minimum, and median values. Patient's K6-Score were used as a measure of psychological distress with 0-7 indicating no distress, 8-12

as mild distress, and 13 and above as serious distress [15]. Given the following levels, we used a threshold value to tag patients as under relapse when their K6 Score ≥ 10 .

We processed the data into a partly conditional time-to-event form (x_i, t_i, δ_i) [16]. x_i is lifelog activity features and their derived values for a given observation window, the censorship indicator δ_i is used to indicate if relapse event has occurred, the time of observation t_i is computed as the number of days starting from a remission stage until a relapse event occurs or until the observation ends, if there are no relapse event occurring for a particular patient instance, then value is set as the whole observation time.

III. EXPERIMENT

Our task is to predict a survival curve for evaluating a patient's risk of relapse through their daily activity lifelog data, we compare the performance using DeepSurv, Random Survival Forest, CoxPH and a baseline model that predicts a flat 0.5 probability of survival at every time point.

Our training strategy involves the use of a 10-fold cross-validation in order to partition individual patients into training and testing set. We aim to consider how well do these models generalize for unseen patients without having to train their historical data. Both CoxPH and Random Survival Forest were implemented using PySurvival [18] while DeepSurv was implemented using Pytorch [13]. Hyperparameters of the models were tuned via grid search. Evaluation is performed using the following metrics:

1) *Concordance Index* : The Concordance Index (C-Index) evaluates a survival model's ability to discriminate if the predicted survival times have the same order as the true survival times. C-Index is similar to the area under the ROC (AUC) for binary classification[10]

and can be computed as:

$$\hat{c} = \frac{1}{num} \sum_{i:\delta_i=1} \sum_{j:y_i < y_j} I[S(\hat{y}_j | X_j) > S(\hat{y}_i | X_i)]. \quad (6)$$

where y observed time of a given sample, \hat{y} is the predicted time, $I()$ the indicator function, num denotes the number of comparable pairs, and denotes $S()$ corresponds to the estimated survival probabilities [17].

2) *Brier Score*: The Brier Score measures the averaged squared distance between the predicted and observed survival functions [17]. For a given population N with the following instance of (x_i, δ_i, T_i) , and the predicted survival function $\hat{S}(t, \vec{x}_i)$ from the population, the Brier Score that takes into account censored data can be computed as:

$$A \left(\frac{(0 - \hat{S}(t, \vec{x}_i))^2 \cdot 1_{T_i \leq t, \delta_i=1}}{\hat{G}(T_i^-)} + \frac{(1 - \hat{S}(t, \vec{x}_i))^2 \cdot 1_{T_i > t}}{\hat{G}(t)} \right). \quad (7)$$

where $A = \frac{1}{N} \sum_{i=1}^N$, and $\hat{G}(t) = P[C > t]$ is the observed survival function calculated from the Kaplan-Meier estimator and C is the censoring time [17]. Also, we also used the Integrated Brier Score (IBS)

$$IBS(t_{max}) = \frac{1}{t_{max}} \int_0^{t_{max}} BS(t) dt. \quad (8)$$

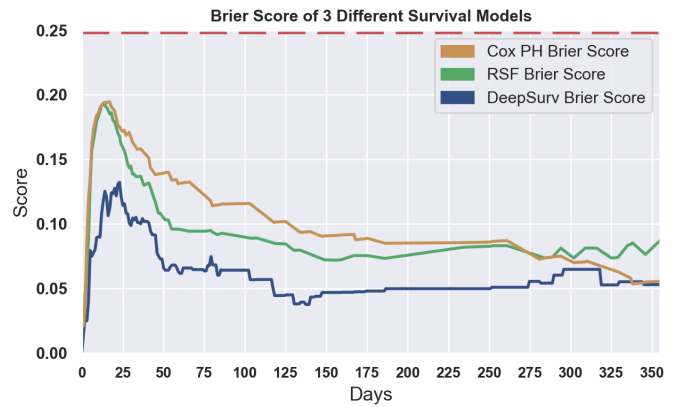


Fig. 1. Brier Score comparison between CoxPH, RSF, and DeepSurv

to measure the performance for a given time interval. A model that perfectly predicts the survival probabilities of every instances will have a score of 0 and a score of 0.25 for a random prediction model.

IV. RESULTS & DISCUSSION

Our results shows in Table I that for both C-Index and IBS metrics. DeepSurv outperformed both RSF and CoxPH in predicting a survival curve for individual patients. We further examine the resulting Brier Score of the models and found out that all of them followed the same trend in that they have higher brier score in predicting relapses that occur within 0-50 days and a lower Brier Score in predicting relapses occurring in the middle and towards the end of year. This suggests that the models have relatively lesser accuracy in predicting patients who are immediately at risk of relapse in contrast to patients who have relapse in the middle and towards the end of a year. The results in Fig. 1 also demonstrates the relative strength and consistency of DeepSurv as compared to both the CoxPH and RSF performing near of a random baseline model at early relapse onset predictions.

TABLE I
C-INDEX AND BRIER SCORE COMPARISON

Model	Concordance Index	Integrated Brier Score
DeepSurv	0.89	0.04
RSF	0.84	0.07
CoxPH	0.73	0.16
Baseline	0.5	0.25

Further inspection of the model, we computed the whole survival curve of the group using Kaplan-Meier method and measured the prediction error when compared to the prediction of DeepSurv, CoxPH, Random Survival Forest. Fig. 2 shows how the errors are distributed, a negative error implies that the model underestimated survival probability prediction, on the contrary a positive error value indicates that the model overestimated its prediction. Our results show that CoxPH consistently underestimates the prediction with its negative error mean and distribution spread as compared

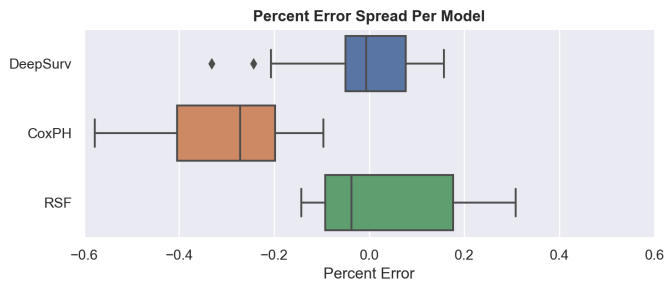


Fig. 2. Prediction MSE spread comparison

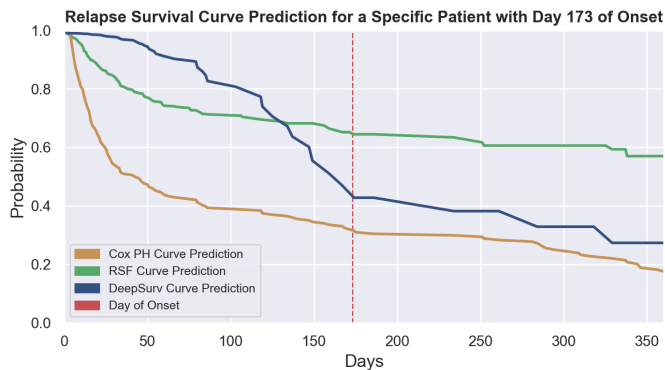


Fig. 3. Comparison of predictions for a given randomly selected patient with a relapse onset at 173 days.

to DeepSurv and RSF. This implies that CoxPH predicts patients to be more at risk for relapse as oppose to DeepSurv and RSF. Among the models, DeepSurv has shown to be consistent in predicting relapse of patients with minimal error without overestimating or underestimating the prediction.

Lastly, Fig 3. shows an example resulting survival curve prediction of the three models using a patient's lifelog features. DeepSurv shows a steep decline on day 173 indicating that the patient has a greater likelihood of relapse for that day. RSF shows an optimistic prediction that the patient has higher chance of not having a relapse with the survival probability staying above 0.5. On the other hand, CoxPH demonstrates a more conservative prediction with steeper chances of surviving as the days progress for the patient.

V. CONCLUSIONS

In this paper, we tackled the problem of predicting a patients likelihood of depression relapse through the use of their personal activity lifelog data using survival analysis and have shown that DeepSurv performed the better in terms of predicting survival probabilities of relapse occurrences.

The study does have limitations, for instance, daily activity behaviours of people are greatly affected by their geography and culture and as such the model might not be able to generalize predictions for other patient populations. In future works, we aim to utilize models that are able take into account fine-grain changes in longitudinal data as well as properly model sequences such as LSTM's. In the long run, the goal is to develop a alerting and intervention system for

patients given their predicted risk through the monitoring of their daily activity lifelog and the results shown here is a step towards demonstrating the feasibility of deploying these kinds of application.

REFERENCES

- [1] R. Kessler and E. Bromet, "The Epidemiology of Depression Across Cultures", *Annual Review of Public Health*, vol. 34, no. 1, pp. 119-138, 2013.
- [2] S. Nuggerud-Galeas, L. Sáez-Benito Suescun, N. Berenguer Torrijo, A. Sáez-Benito Suescun, A. Aguilar-Latorre, R. Magallón Botaya, and B. Oliván Blázquez, "Analysis of depressive episodes, their recurrence and pharmacologic treatment in primary care patients: A retrospective descriptive study," *PLOS ONE*, vol. 15, no. 5, 2020.
- [3] A. Thapar, G. Hammerton, S. Collishaw, R. Potter, F. Rice, G. Harold, N. Craddock, D.J. Smith, "Detecting recurrent major depressive disorder within primary care rapidly and reliably using short questionnaire measures". *Br J Gen Pract.* 64(618):e31-7, 2014.
- [4] B. Buck et al., "Relationships between smartphone social behavior and relapse in schizophrenia: A preliminary report", *Schizophrenia Research*, vol. 208, pp. 167-172, 2019.
- [5] B. Li, H. Yu and A. Sano, "Toward End-to-end Prediction of Future Wellbeing using Deep Sensor Representation Learning", 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2019.
- [6] P. Chikersal et al., "Detecting Depression and Predicting its Onset Using Longitudinal Symptoms Captured by Passive Sensing", *ACM Transactions on Computer-Human Interaction*, vol. 28, no. 1, pp. 1-41, 2021.
- [7] N. Kumagai et al., "Predicting recurrence of depression using lifelog data: an explanatory feasibility study with a panel VAR approach", *BMC Psychiatry*, vol. 19, no. 1, 2019.
- [8] D. Vock et al., "Adapting machine learning techniques to censored time-to-event health record data: A general-purpose approach using inverse probability of censoring weighting", *Journal of Biomedical Informatics*, vol. 61, pp. 119-131, 2016.
- [9] Y. Li, J. Wang, J. Ye and C. Reddy, "A Multi-Task Learning Formulation for Survival Analysis", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [10] P. Wang, Y. Li and C. Reddy, "Machine Learning for Survival Analysis", *ACM Computing Surveys*, vol. 51, no. 6, pp. 1-36, 2019.
- [11] C. Englebert, T. Quinn, and I. Bichindaritz, "Feature Selection for Survival Analysis in Bioinformatics", *Proceedings of the Workshop on Advances in Bioinformatics and Artificial Intelligence: Bridging the Gap co-located with 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, Vol. 1942, pp 30-35
- [12] H. Ishwaran and M. Lu, "Random Survival Forests", *Wiley StatsRef: Statistics Reference Online*, pp. 1-13, 2019.
- [13] J. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang and Y. Kluger, "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network", *BMC Medical Research Methodology*, vol. 18, no. 1, 2018.
- [14] S. Burcusa and W. Iacono, "Risk for recurrence in depression", *Clinical Psychology Review*, vol. 27, no. 8, pp. 959-985, 2007.
- [15] S. Dhingra, M. Zack, T. Strine, B. Druss, J. Berry and L. Balluz, "Psychological Distress Severity of Adults Reporting Receipt of Treatment for Mental Health Problems in the BRFSS", *Psychiatric Services*, vol. 62, no. 4, pp. 396-403, 2011.
- [16] Q. Gong and D. Schaubel, "Estimating the average treatment effect on survival based on observational data and using partly conditional modeling", *Biometrics*, vol. 73, no. 1, pp. 134-144, 2016.
- [17] H. Kvamme, Ø. Borgan, I. Scheel, "Time-to-Event Prediction with Neural Networks and Cox Regression", *JMLR* vol. 20 pp. 1-30, 2019
- [18] "PySurvival", 2021. [Online]. Available: https://square.github.io/pysurvival/metrics/brier_score.html