IEEE ICHI 2023
Houston, TX, USA • 26–29 June 2023

From Digital Phenotype Identification To Detection Of Psychotic Relapses

N. Efthymiou [1] , *G. Retsinas*[1], *P. Filntisis*[1], *C. Garoufis*[1], *A.Zlatintsi*[1], *E.Kalisperakis*[2,3], *V. Garyfalli*[2,3], *T. Karantinos*[2], *M. Lazaridi*[2,3], *N. Smyrnis*[2,3], *P. Maragos*[1]

[1] *School of ECE, National Technical University of Athens, Greece*
[2] *Laboratory of Cognitive Neuroscience, University Mental Health Research Institute, Athens, Greece*
[3] *Medical School, National & Kapodistrian University of Athens, Greece*

e-Prevention
Advanced Support System for Treatment Monitoring and Relapse Prevention in Patients

Psychotic Disorders

- **Millions** of people experience symptoms of psychotic disorders
- **Schizophrenia** & **bipolar disorder** are the most common
- Chronic diseases with **repeated periods of relapse**
- **Predicting** psychotic relapses early remain a major clinical issue

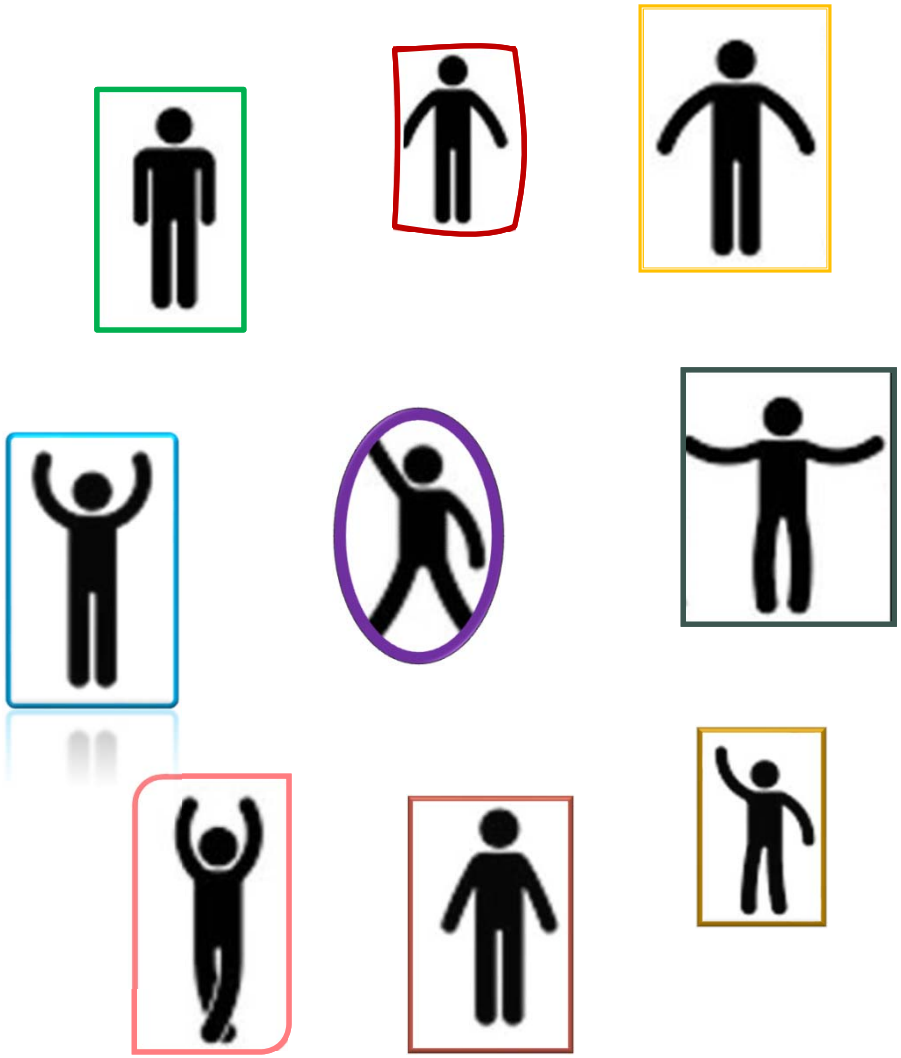
Employ wearable consumer products

Reliable, unobtrusive, remote collection of behavioral and biometric data

From Digital Phenotype Identification ...

Is it possible to use digital phenotype using smartwatches for person identification?

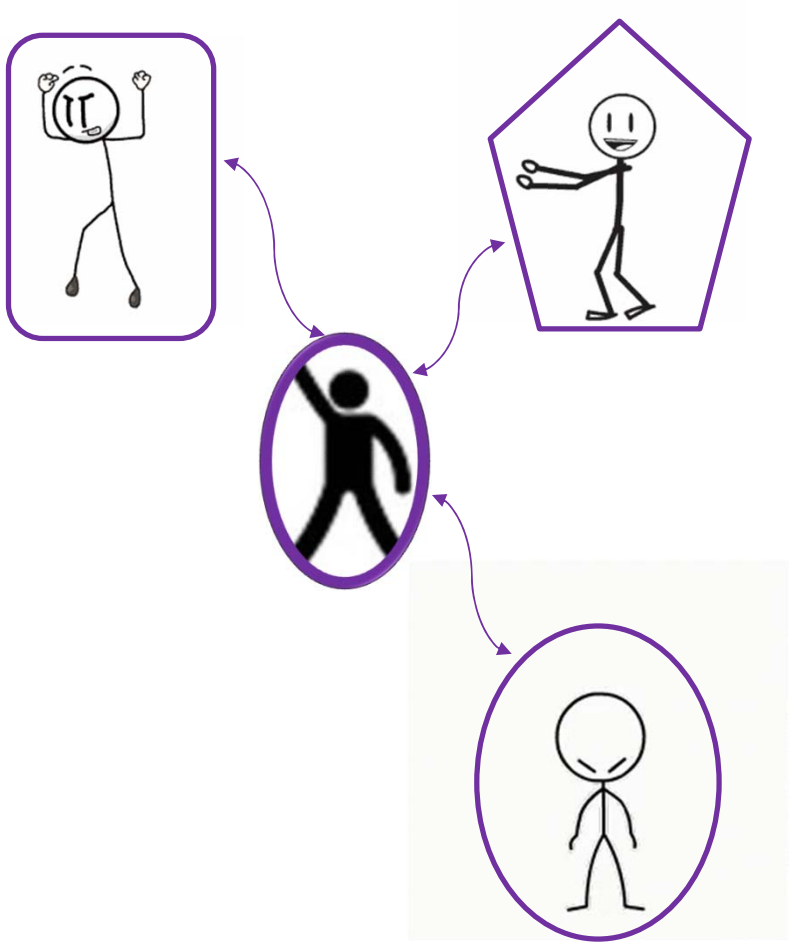
From Digital Phenotype Identification ...



Is it possible to use digital phenotype using smartwatches for person identification?

Yes, there are numerous works to exploit multimodal data from smartwatches such as heart rate variability, acceleration, step, sleep etc.

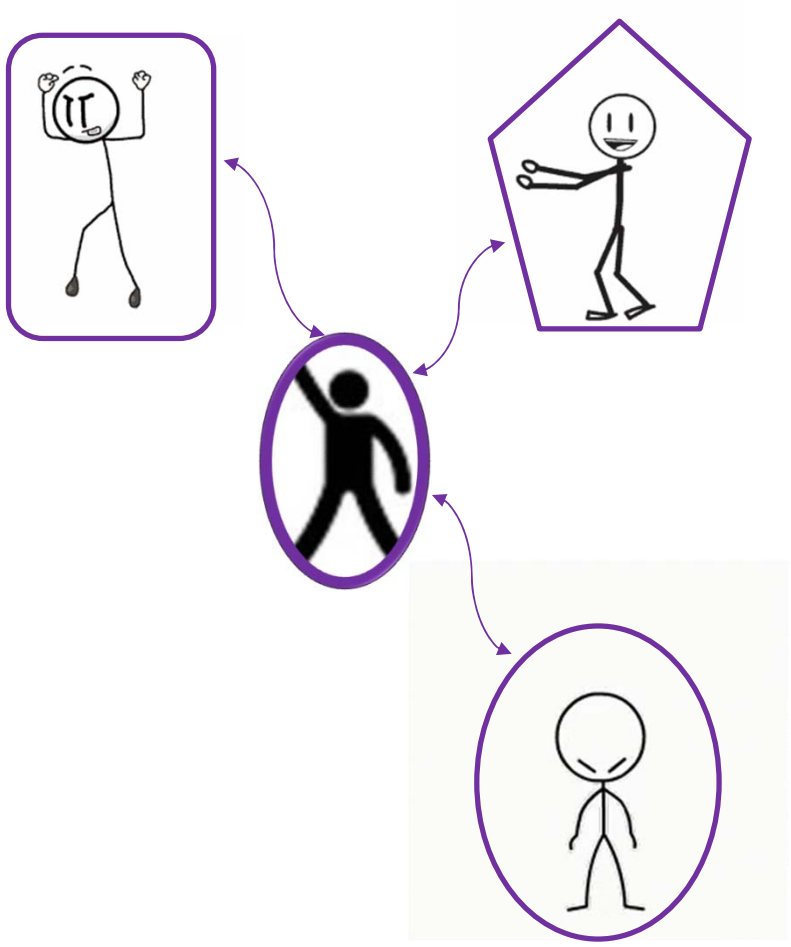
… to Detection of Psychotic Relapses



Does a person identification system which use digital signals could help to detect psychotic relapses?

During a psychotic relapse, the person tends to adopt different behavioral patterns.

… to Detection of Psychotic Relapses



Does a person identification system which use digital signals could help to detect psychotic relapses?

***Our hypothesis*:**

Leveraging deep neural networks trained for digital phenotype classification we could detect diverging behavioral patterns during a psychotic relapse

Employing person identification for the discovery of psychotic relapses

What we do?

- Introduce a **novel method for detecting psychotic relapses** casting the problem of relapse detection as a **missclassification problem**
- **Verify the changes in the output probability distribution scores** for each person in different periods of a psychotic disorder with thorough statistical analysis.

How?

- **Design and build deep learning architectures for person identification** using behavioral and cardiovascular signals from smartwatch
- **Perform extensive exploratory studies** (architecture networks, input features, augmentation)

On what data?

Train and validate the assumptions on collected dataset of biometric wearable data

- including a total of **29 patients**
- ∼ **18k days** of recording data
- **spanning up to 2.5 years** of continuous monitoring

Background – Collected Data during the e-Prevention Project

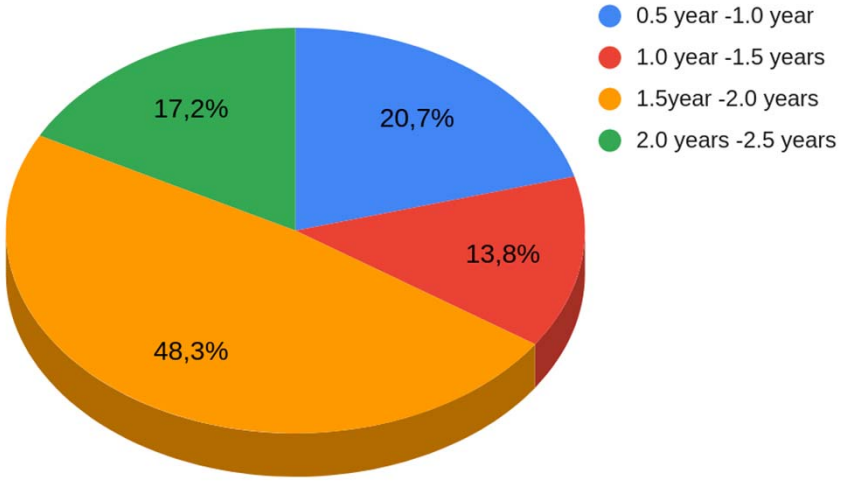
- 60 people were recruited at *University Mental Health, "Costas Stefanis" (UMHRI)*
  - 37 patients in the psychotic spectrum (schizophrenia and bipolar disorder)
  - 23 healthy controls
- Subjects wore a Samsung Gear S3 smartwatch which monitored:
  - linear acceleration (20 Hz) from accelerometer
  - angular velocity (20 Hz) from gyroscope
  - heart rate and heart rate variability (5 Hz) using photoplethysmography
  - sleep/wake cycles
  - steps and distance covered
- Patients had monthly assessments and communication with their attending physician – more 40 psychotic and non-psychotic relapses were annotated

In total, 20,000 human-days were collected among all participants – one of the largest datasets of this kind ever recorded !

Used e-Prevention biosignals database

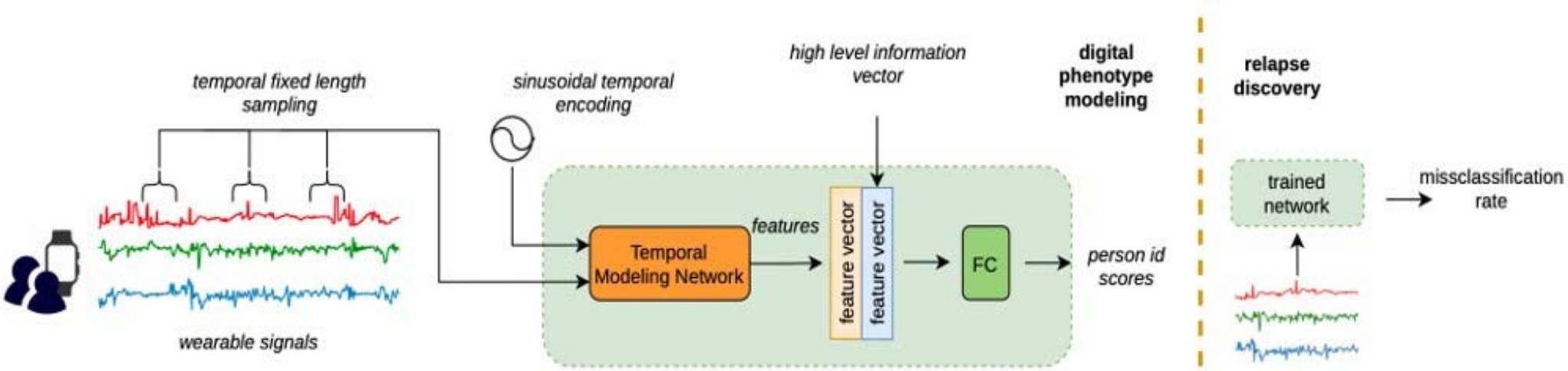
- 29 patients with more than 180 recording days after the preprocessing
- 11 patients experienced 16 psychotic relapses

	Mean	STD	Sum
Days of participation	736.70	168.90	21,365
Days in database	613.00	187.71	17,777
Days in database (relapsed)	29.27	43.59	849
Hours per day	14.53	1.94	-

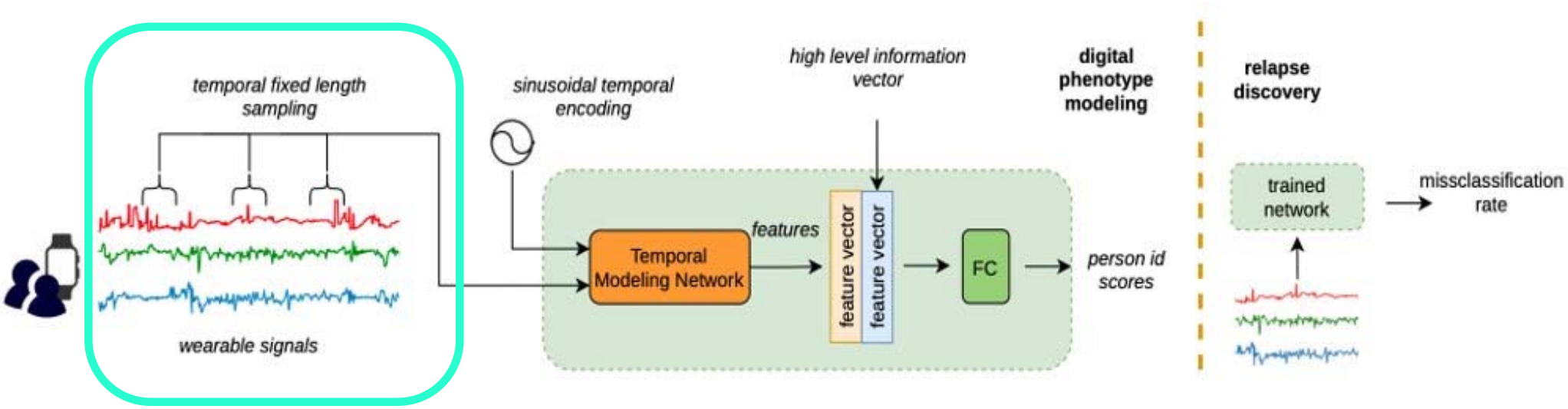


Participation in years

Our proposed system



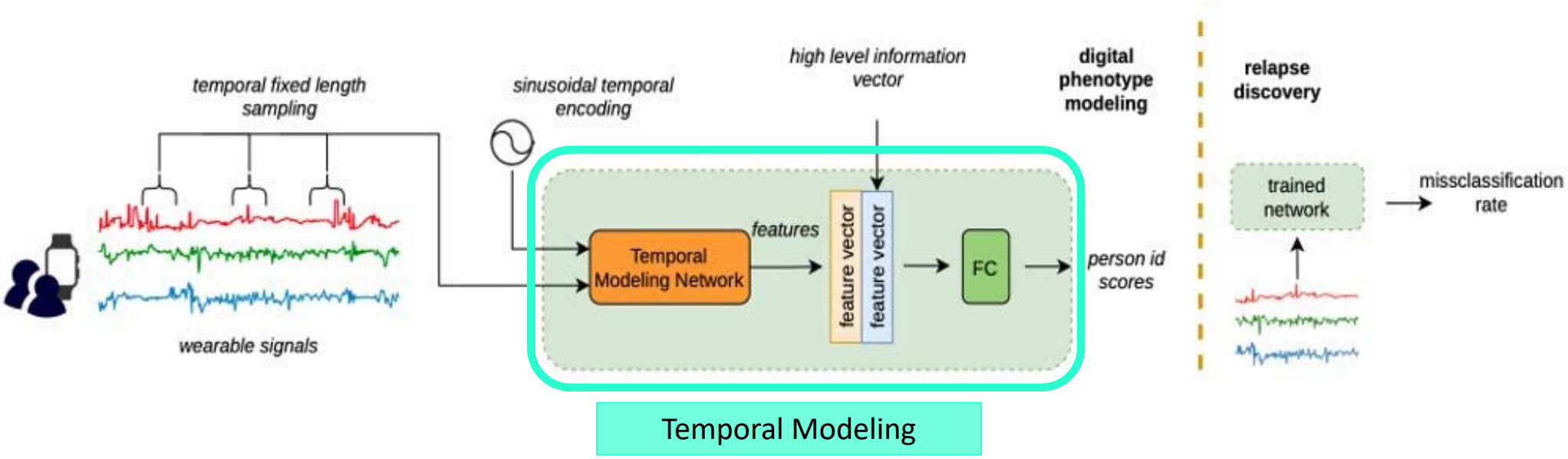
Our proposed system



Random sampling of daily biosignals

- Extract sequences of features from data collected continuously from smartwatch aggregate them in small temporal scale
- Perform random sampling through temporal dimension to reduce the dimensionality

Our proposed system



Designed 2 different architectures:

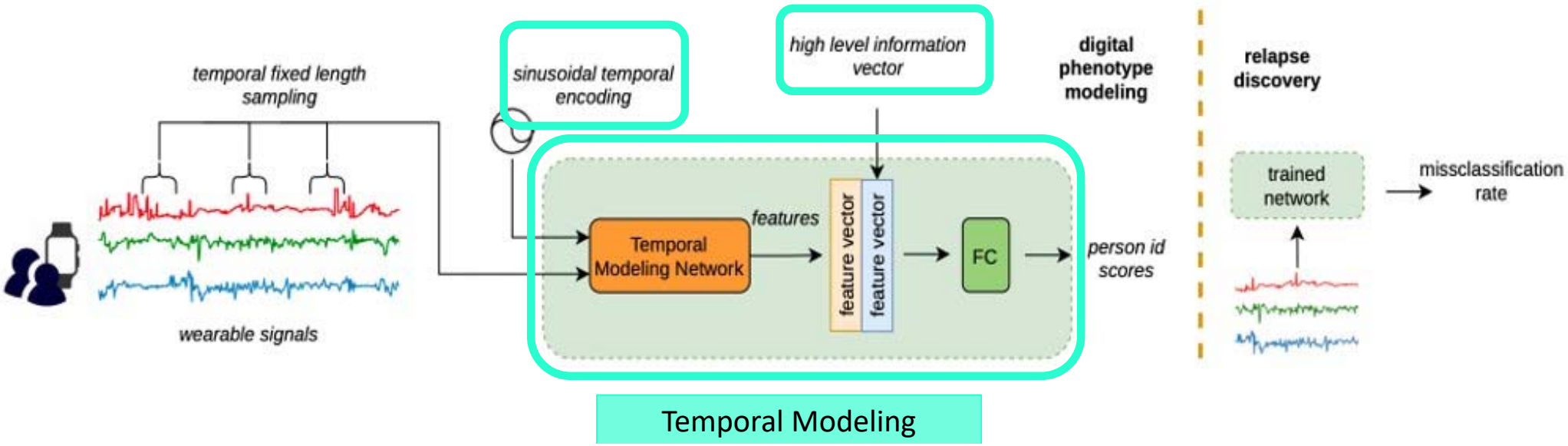
- CNN-based with 1D convolutional kernels
- LSTM-based

Data augmentation techniques to avoid overfitting:

- Sampling a normal distribution and applying this noisy signal to the initial signal in a multiplicative manner
- Applying a random mask (2D sparse binary matrix) on the input signal
- Employing the mixup[1] technique

[1] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in ICLR 2018

Our proposed system



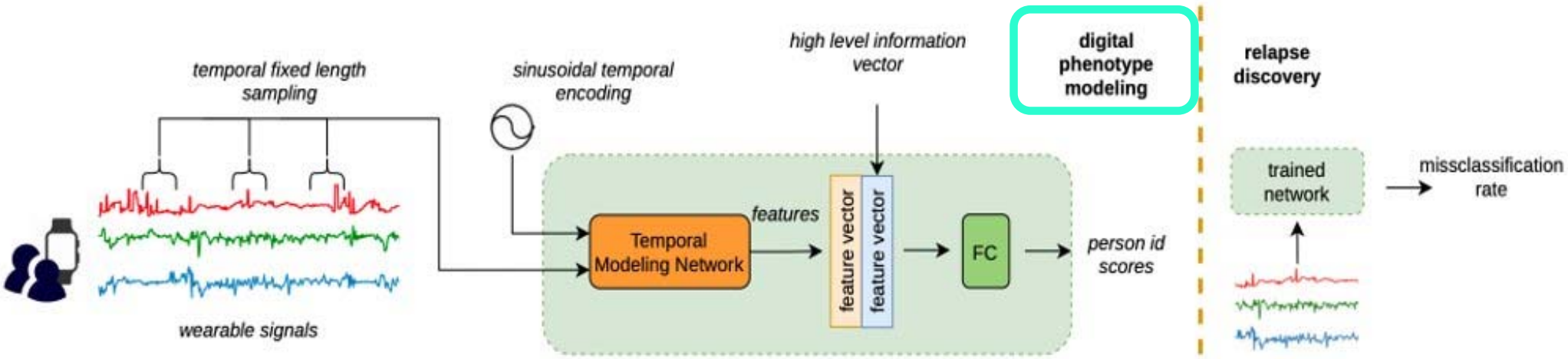
Concatenate our multidimensional signal - sequence with:

- 2 sinusoidal features which denote the daily temporal cycle
- A high level information vector (hours of sleep + temporal weekly cycle information)

The augmented vector is fed into 2 fully connected layers, which output the identification scores; cross-entropy loss is used for training.

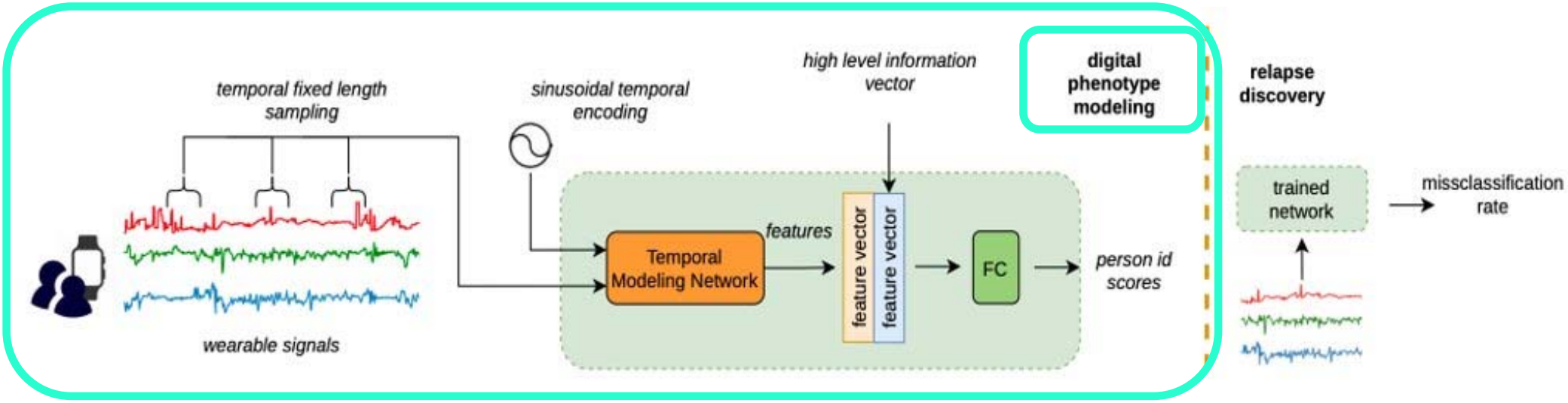
During inference: we perform random sampling for each day 5 times and sum the unnormalized scores to obtain the final score.

Our proposed system



We obtain the user's digital phenotype

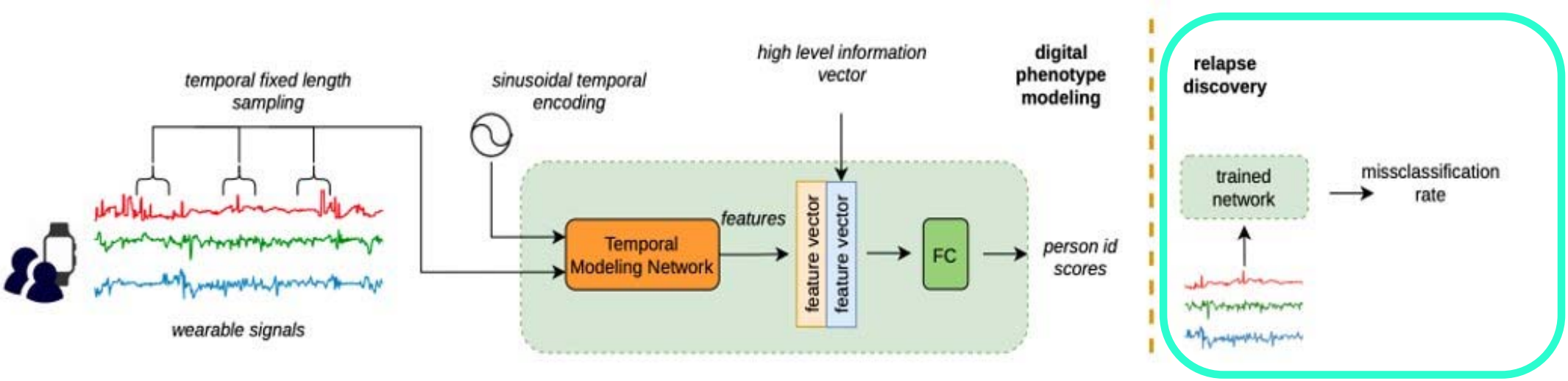
Our proposed system



We obtain the user's digital phenotype

Note: Training is carried out with data from patients who are in remission!

Our proposed system



Investigate how the behavioral shifts are reflected on the output distribution probability & classification accuracy

Challenges for creating the experimental dataset

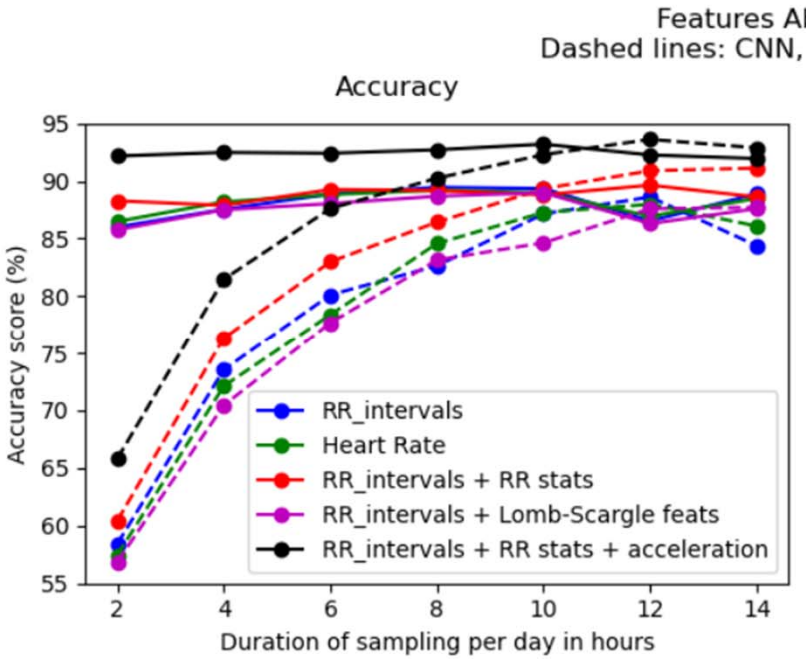
Mean recording length is ~ 14 hours per day << 18-20 hours in patients with good compliance

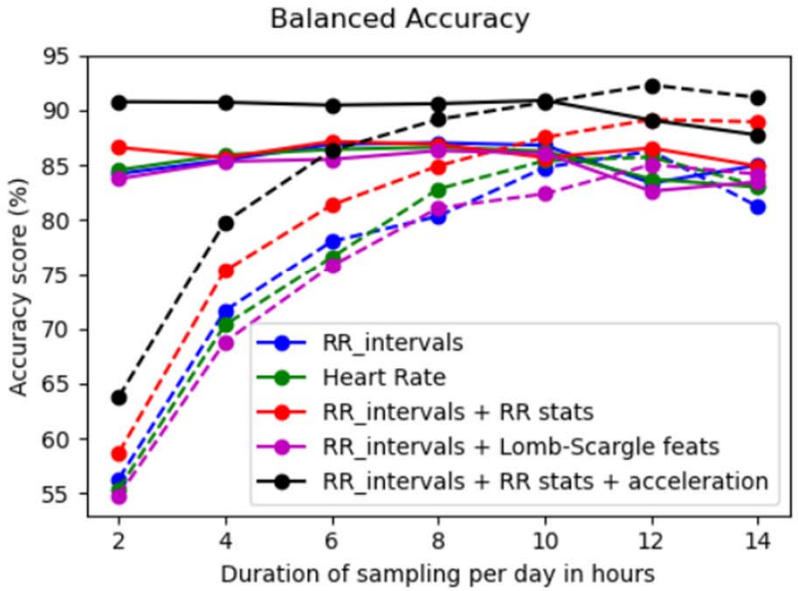
- ineffective photoplethysmography provided by the watch (sweaty wrists),
- failure of the recording watch application
- typical errors in use (not charging before the battery runs out, forgetting to wear)
- poor patient compliance
- network issues.

Features exploration

Two important decisions for getting a robust representation of the day:

- Feature vector length (random sampling, sequential sampling, padding)
- Feature types (features are extracted from 5 minute aggregation of raw signals)





RR_intervals : Heart rate variability statistics

Heart Rate: Heart rate average statistics (beats per minute)

RR_stats: SDRR, RMSSD

Lomb-Scargle feats: Lomb-Scargle LF, HF

Acceleration: Norm of linear acceleration

Architecture ablation

- CNN are much more sensitive than the LSTM variant
- LSTM shows stable behavior across all settings
- Noticed a slight decrease in most settings when temporal encoding is used
- But this information is important for the final task, since it can be better model the behavior cycle within a day

			Without Augmentation	Adding Noise	+ Applying Random Mask	+ Mixup Samples
CNN	Base	Acc.	**92.29**	89.58	89.19	70.36
		Bal.Acc.	**90.72**	88.42	87.58	65.41
	+ temporal encoding	Acc.	**91.56**	88.87	87.83	69.79
		Bal.Acc.	**89.81**	86.78	85.60	64.01
LSTM	Base	Acc.	**93.20**	91.32	92.89	91.54
		Bal.Acc.	**90.90**	88.64	90.46	89.18
	+ temporal encoding	Acc.	92.03	90.27	92.31	**92.57**
		Bal.Acc.	88.97	87.08	89.58	**90.06**

Distinguishing different periods of psychotic disorders

	Metrics	Balanced Acc. (29 patients)	Balanced Accuracy (11 patients)			Mean Probability (11 patients)			Median Probability (11 patients)		
	Extra Features	Normal	Normal	Pre-Rel.	Relapse	Normal	Pre-Rel.	Relapse	Normal	Pre-Rel.	Relapse
Base	None	90.13	87.53	75.15	74.83	0.8685	0.8243	0.8227	0.9708	0.9096	0.9433
	Hours of sleep (HS)	**90.19**	**88.37**	77.44	72.52	0.8794	0.8215	0.8124	0.9873	0.9109	0.9472
	Day of Week (DW)	89.40	86.46	75.88	70.12	0.8632	0.8156	0.7756	0.9859	0.9332	0.8898
	HS + DW	89.76	87.65	76.88	70.35	0.8744	0.8076	0.7902	0.9909	0.8809	0.8980
+ Temporal encoding	None	89.04	85.93	75.02	71.53	0.8564	0.8288	0.8027	0.9465	0.9120	0.8947
	Hours of sleep (HS)	88.85	85.65	76.81	70.20	0.8579	0.8076	0.7812	0.9895	0.8832	0.8724
	Day of Week (DW)	88.07	85.16	75.20	68.84	0.8486	0.8032	0.7712	0.9687	0.8506	0.8791
	HS + DW	88.67	**86.00**	**74.28**	**69.86**	**0.8634**	**0.7954**	**0.7593**	**0.9690**	**0.8734**	**0.8498**

- Simplest version of the network (base) + HS achieves highest results
- Considering 11 patients, the addition of high-level features widens the gap between normal and relapse phases in terms of accuracy
- Mean and median classification probability scores are more indicative of the models' outcomes.
- All possible supplementary features contribute to the detection of relapses. They intuitively capture a full profile of a person's daily activities

Per person & per period identification scores

	Probability			Absolute Change		
ID	normal	pre-rel	relapse	normal → pre-rel	normal → relapse	pre-rel → relapse
1	0.971	0.944	0.826	**−0.027**	**−0.145**	−0.118
2	0.952	0.928	0.872	**−0.024**	**−0.080**	**−0.056**
3	0.963	0.913	0.997	**−0.051**	0.034	0.085
4	0.853	0.784	0.726	**−0.070**	**−0.127**	**−0.058**
5	0.905	0.718	0.765	**−0.187**	**−0.140**	0.047
6	0.905	1.000	0.794	0.095	**−0.112**	−0.206
7	0.966	0.983	0.974	0.017	0.008	−0.009
8	0.659	0.672	0.496	0.014	**−0.163**	**−0.176**
9	0.693	0.407	0.541	**−0.286**	**−0.152**	0.134
10	0.961	0.986	1.000	0.025	0.039	0.014
11	0.703	0.406	0.358	**−0.297**	−0.346	−0.049
all	0.866	0.795	0.759	−0.0723	−0.107	−0.035

- **Significant drop in the mean identification** probability of the network during the relapse and pre-relapse periods.
- The motif of absolute drop can be seen in **most users** with some exceptions.

Stiatistical Analysis of Scores

	Probability			Absolute Change		
ID	normal	pre-rel	relapse	normal → pre-rel	normal → relapse	pre-rel → relapse
1	0.971	0.944	0.826	**−0.027**	**−0.145**	−0.118
2	0.952	0.928	0.872	**−0.024**	**−0.080**	**−0.056**
3	0.963	0.913	0.997	**−0.051**	0.034	0.085
4	0.853	0.784	0.726	**−0.070**	**−0.127**	**−0.058**
5	0.905	0.718	0.765	**−0.187**	**−0.140**	0.047
6	0.905	1.000	0.794	0.095	**−0.112**	−0.206
7	0.966	0.983	0.974	0.017	0.008	−0.009
8	0.659	0.672	0.496	0.014	**−0.163**	**−0.176**
9	0.693	0.407	0.541	**−0.286**	**−0.152**	0.134
10	0.961	0.986	1.000	0.025	0.039	0.014
11	0.703	0.406	0.358	**−0.297**	−0.346	−0.049
all	0.866	0.795	0.759	−0.0723	−0.107	−0.035

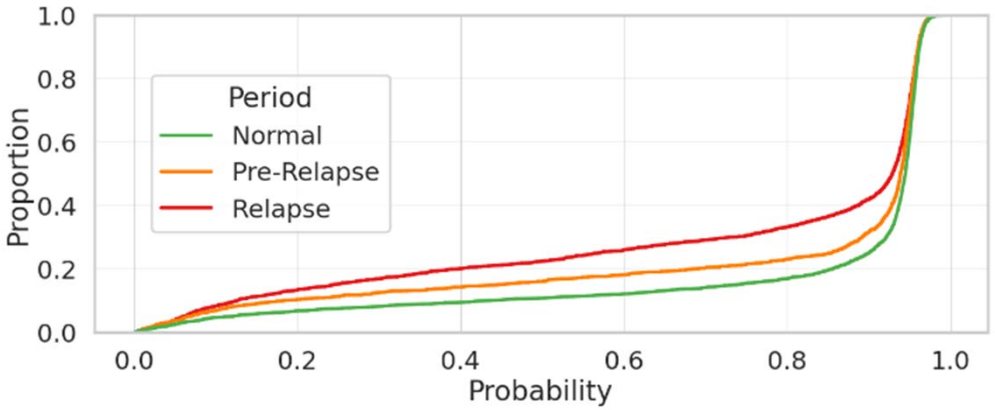
- Collected for **each patient across all folds** (5 fold cross-validation) **the identification scores** for each period and conducted single tail Mann-Whitney U-tests with Bonferroni correction.
- **Values in bold denotes** that scores during the phase at the left were **significantly grater** than scores during the phase at the right of the arrow.

Per person & per period identification scores

	Probability			Absolute Change		
ID	normal	pre-rel	relapse	normal → pre-rel	normal → relapse	pre-rel → relapse
1	0.971	0.944	0.826	**−0.027**	**−0.145**	−0.118
2	0.952	0.928	0.872	**−0.024**	**−0.080**	**−0.056**
3	0.963	0.913	0.997	**−0.051**	0.034	0.085
4	0.853	0.784	0.726	**−0.070**	**−0.127**	**−0.058**
5	0.905	0.718	0.765	**−0.187**	**−0.140**	0.047
6	0.905	1.000	0.794	0.095	**−0.112**	−0.206
7	0.966	0.983	0.974	0.017	0.008	−0.009
8	0.659	0.672	0.496	0.014	**−0.163**	**−0.176**
9	0.693	0.407	0.541	**−0.286**	**−0.152**	0.134
10	0.961	0.986	1.000	0.025	0.039	0.014
11	0.703	0.406	0.358	**−0.297**	−0.346	−0.049
all	0.866	0.795	0.759	−0.0723	−0.107	−0.035

- **Significant drop in the mean identification** probability of the network during the relapse and pre-relapse periods.
- The motif of absolute drop can be seen in **most users** with some exceptions.

Stiatistical Analysis of Scores



empirical cumulative distribution (eCDF)

- Collected all the identification scores **across all patients and folds**.
- Identifications scores

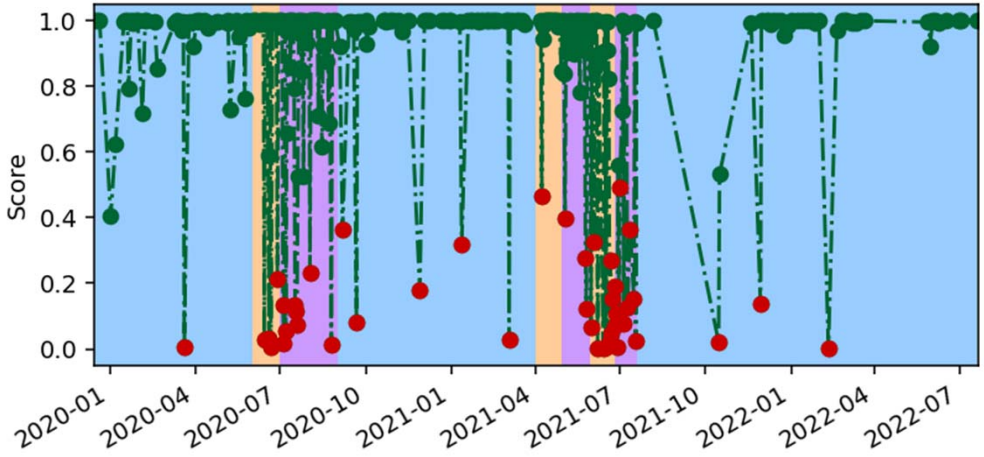
Relapse < Pre-relapse < Normal period

- Conducted pairwise single tail Mann-Whitney U-tests with Bonferroni correction for the overall scores

Per person results

- Visualizing identification predictions of patient #2

	start_date	stop_date	period	predictions	total
	2019-12-14	2020-06-01	normal	49	50
	2020-06-01	2020-07-01	pre-relapse	21	25
	2020-07-01	2020-09-01	relapse	45	53
	2020-09-01	2021-04-01	normal	51	56
	2021-04-01	2021-04-29	pre-relapse	19	20
	2021-04-29	2021-05-30	relapse	26	29
	2021-05-30	2021-06-25	pre-relapse	12	21
	2021-06-25	2021-07-19	relapse	7	18
	2021-07-19	2022-07-20	normal	41	44



	Probability			Absolute Change		
ID	normal	pre-rel	relapse	normal → pre-rel	normal → relapse	pre-rel → relapse
2	0.952	0.928	0.872	**−0.024**	**−0.080**	**−0.056**

- Missclassifications of the user
- Correct prediction of the user

Limitations & Conclusions

- Our experimental analysis shows that there are **significant changes in the output distribution scores during relapse and pre-relapse periods** as well as a drop in the classification rate of the network.
- Further aspects of our framework should be **examined more thoroughly to draw safe conclusions and assist clinicians**
- i.e. network architectures can be enhanced, create personalized models, working more on pre-relapse periods
- We hope that our work can stimulate further research in this direction.

IEEE ICHI 2023
Houston, TX, USA • 26–29 June 2023

Thank you!

Phenotype = Φαινότυπος

from Ancient Greek φαίνω *(phaínō)* 'to appear, show'
+ τύπος *(túpos)* 'mark, type'

https://robotics.ntua.gr/eprevention-sp-challenge/