A comparative study of autoencoder architectures for mental health analysis using wearable sensors data

M. Panagiotou, A. Zlatintsi, P. P. Filntisis, A. J. Roumeliotis, N. Efthymiou, and P. Maragos
School of ECE, National Technical University of Athens, 15773 Athens, Greece
{el16703, aroumeliot}@mail.ntua.gr, {filby, nefthymiou}@central.ntua.gr, {nzlat, maragos}@cs.ntua.gr

***Abstract*—In this study, the application of deep learning models for the detection of relapses in patients with psychotic disorders (i.e., bipolar disorder and schizophrenia) is examined, using physiological signals, collected by smartwatches. In order to tackle the problem of relapse detection, which in our case is handled as an anomaly detection task, four different autoencoder architectures, based on Transformers, Fully connected Neural Networks (FNN), Convolution Neural Networks (CNN) and Gated Recurrent Unit (GRU), are implemented as personalized and global models. In this work, time-scaled data of total duration of 1569 days, segmented into five minutes intervals, from ten patients suffering from psychotic disorders have been examined yielding encouraging results. Furthermore, since the patients' relapses were appropriately annotated by clinicians as low, moderate or severe, we conducted a post hoc analysis using the models that performed best, to examine the importance of the severity level among three participants who relapsed multiple times with different severity level, providing important evidence.**

***Index Terms*—Psychotic Disorders, Anomaly Detection, Time-series, Autoencoder Architectures, Neural Network**

I. Introduction

Digital phenotyping [1] is a nascent exciting interdisciplinary field motivated by the broad adoption of wearable devices, such as smartwatches, in our daily lives. The term encompasses the quantification of human behavior and traits, by utilizing the sensors included in these devices. Such wearables collect multimodal data, often including accelerometers, gyroscopes and heart rate monitors, for measuring the user's physical activity, as well as autonomic function [2], [3].

This abundance of sensory data has kickstarted the development of applications focused in general user and health monitoring, and other predictive analytic tasks, e.g., emotional well-being [4], sleep tracking [5], and physical activity detection [6] among others. Many works have also focused in identifying behavioral and biometric markers, which can be extracted from such data and provide insights into general medicine [7].

The success of these applications has increased the interest in psychological health for human wellness, as well, showing that such data could also be introduced into clinical psychiatry [8]; employing them to examine depression [9], bipolar disorder [10], and schizophrenia [11]. Especially for people having mental disorders, sensing biometric markers of interest unobtrusively, through passive, continuous and long-term monitoring could be proven effective in improving both wellness and the course of the disorder.

Previous works have mostly used smartphones [12] and focused mainly on social features such as text messages, call duration, and sleep duration among others [13], [14]; lasting from some hours to a few weeks [13] with some exceptions [14]. Compared to smartphones, wearable sensors are unobtrusive, lightweight and can be used for monitoring of daily activities [15]; it has been shown that people suffering from psychotic disorders are comfortable and willing to integrate them into their daily life, something that supports the fact that by using smartwatches, we could go beyond feasibility and underscore the novel physiological and activity data that can be easily collected with low cost [16].

An approach that could be used for relapse detection in patients suffering from psychotic disorders is the sensor-based anomaly detection, the importance of which has been highlighted during the last years, through the clinical mass adoption of telehealth [17]. In previous works [14], encoder-decoder models were built, so as to predict behavioral anomalies that could indicate early warning signs of relapses. Such methods are largely applied in natural language processing, audio [18], video and time-series data [19].

This study is an ongoing work of the ***e-Prevention*** project (http://eprevention.gr), targeting to innovative e-Health services for patients' effective monitoring [20]. In contrary to aforementioned works, we use smartwatch data that provides long-term continuous data, collected during two years. Specifically, we use physiological signals of movement and heart rate to identify biomarkers that can assist the detection of psychotic relapses. Therefore four architectures are built, such as Fully Connected Neural Network (FNN), Convolution Neural Network (CNN), Recurrent Neural Network (RNN) autoencoders and Transformers, which are evaluated in both personalized and global schemes showing promising results. Additionally, we perform a post hoc analysis to evaluate the importance of the relapse severity, providing important evidence.

The rest of this paper is organized as follows: in Sec. II data collection, the preprocessing and the final dataset used for the evaluation of relapse detection in psychotic patients are presented. In Sec. III our unsupervised deep learning approach is described; while in Sec. IV a thorough analysis

This research has been financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH–CREATE–INNOVATE (project acronym: e-Prevention, code:T1EDK-02890/ MIS: 5032797)

TABLE I
DEMOGRAPHICS INFORMATION

Male/Female	6/4
Age (years)	30.60 ± 7.31
Education (years)	13.8 ± 1.99
Illness dur. (years)	7.3 ± 7.06

for the performance of the examined architectures is presented. Finally, Sec. V concludes our work and gives future directions.

II. Data Collection and Preprocessing

A. Data Collection

Twenty-four (24) patients with a disorder in the psychotic spectrum (12 with Schizophrenia, 8 with Bipolar I disorder, 2 with Schizoaffective disorder, 1 with Schizophreniform disorder and 1 with Brief Psychotic episode) were recruited at the University Mental Health, Neurosciences and Precision Medicine Research Institute "Costas Stefanis" (UMHRI) in Athens, Greece. The protocol regarding the recruitment of the patients in the project is detailed in [18].

The raw data of patients are recorded on 24/7 basis by a Samsung Gear S3 Frontier smartwatch (which includes an accelerometer, a gyroscope and a PPG-based (Photoplethysmography) non-invasive heart rate monitor) and are stored to a cloud-based platform [20]. Linear acceleration and angular velocity are sampled at 20Hz, while heart rate at 5Hz. The total data collection period lasted from 11/2019 to 09/2021, varying for each patient, with a total of 1569 days after preprocessing.

The patients undergo monthly assessments of their mental health symptoms by the clinicians, providing this way valuable annotations indicating their condition as either healthy or relapsing (also denoting the specific period of the relapse and its severity as low, moderate or high). Depending on these annotations, we split the data into three categories: **normal** data, where the patient was stable; **relapse** data corresponding to time periods when a relapse had occured and **near-relapse** data, thus data recorded up to 21 days prior to the appearance of each relapse. The determination of the relapse was multifaceted and was evaluated by the clinicians through the following: 1) Monthly assessments that assisted in quantifying the duration and severity of the relapse 2) the use of questionnaires that gives information for the general psychopathology and the relapse itself, and through communication with 3) the attending physician, 4) the family and 5) the hospital, when the patient had to be hospitalized.

B. Data Preprocessing

Before training, we created sequences of features found to contain significant information as shown in [21]. The extracted features are: the mean energy of the accelerometer and gyroscope norm, the mean heart rate and R-R interval, the mean frequency in the LF and HF bands of the heart ($0.04-0.15$Hz and $0.15-0.40$Hz, respectively), and the value of the width of the ellipse in the Poincare recurrence plot. Three additional features are included: the percentage of correctly identified pulses in the given interval (which corresponds to a measure of how well the patient is wearing the watch) and the sine and cosine representations of corresponding seconds to model the chronological order of the time-series. Afterwards, the features

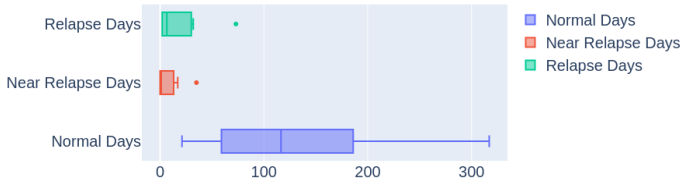

Fig. 1. Number of days used in our experiments after preprocessing.

are transformed into a more dense representation of 5-minutes intervals. The chosen interval length is small enough to capture micro-scale patterns [19], and it also allows us to have an adequate amount of data for our deep learning architectures.

In cases of missing data up to 10 consecutive hours (e.g., when the patient was charging or did not wear the watch), we filled the data with the median values of a given feature. Two other filling methods were examined, i.e., mean, or neighbor; however, we experimentally confirmed that "median" yields better results. In cases that we had more than 10 consecutive hours of missing data, we completely disregarded the interval, since we experimentally found that this is the maximum of hours to be filled, while not inserting false values.

After preprocessing, the data are considered as a multivariate time-series $\mathbf{X}_{L\times d}$, where L denotes the total length (varying for each patient) and d the number of features. Each point in the timeseries corresponds to a 5-minute timestamp. Then, we apply an l-length rolling window with stride 1, creating a total of $N = L-l+1$ subsequences, thus, resulting in a $M_{N\times l\times d}$ tensor for the data of each patient.

As a final step, we split the data in subsequences of 24 hours examining this way the patients' daily patterns, as in [14]. After preprocessing the number of relapsed subjects with adequate data, meaning missing values for less than 10 hours and consecutive data to construct the subsequences was reduced to 10, of which 2 with Schizoaffective Disorder, 4 with Bipolar I Disorder, 1 with Brief Psyhcotic Episode, 1 with Schizophreniform Disorder and 2 with Schizophrenia (see Table I). We have to note that for the analysis in this work, we did not include NR data. Figure 1 shows the number of days used in our experiments after the preprocessing.

III. Methodology

A. Autoencoder Architectures

We created four different architectures based on encoder-decoder schema [22], [23] that learned to reconstruct an input time series. The models were implemented as follows.

In the transformer model as seen in Fig. 2, the input sequence is first imported into a positional encoding layer where information about their chronological order is kept through the application of sine and cosine functions. Afterwards, four transformer encoder layers are stacked. Each encoder layer is made up of two sub-layers. Both sub-layers are followed by a normalization layer. After encoding the input sequence is reversed and piped into the decoder, which consists of four decoder layers, with similar architecture with the encoder. Finally, we apply a linear layer after the decoder.

For the gated recurrent unit (GRU) sequence-to-sequence model with attention (Fig. 3), we input a subsequence of data

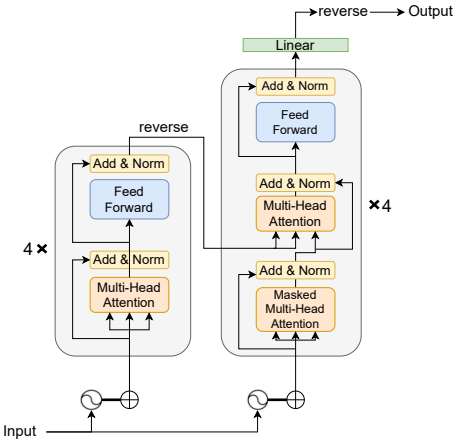

Fig. 2. Transformer architecture.

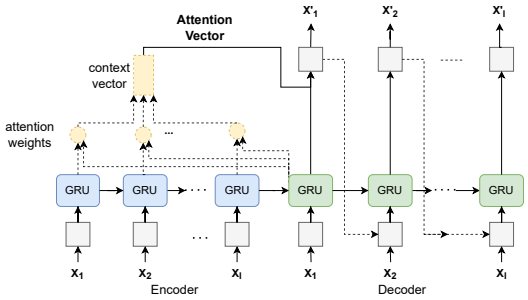

Fig. 3. GRU with attention AE architecture.

into an encoding layer of a GRU with a specified hidden unit of size 100. We then pass the weighted average of all encoded outputs (attention vectors) from all time-steps as inputs into a GRU decoder layer that reconstructed the subsequence.

The fully connected neural network autoencoder (FNN) model (Fig. 4), comprised 2 fully connected encoder and decoder layers that compressed an input subsequence to a lower dimension and then endeavor to reconstruct the initial subsequence. Each fully connected layer is followed by a ReLu non-linearity, where the last layer contains also a dropout layer in order to avoid over-fitting.

The convolutional neural network autoencoder (CNN) (Fig. 5) consists of an encoder that learns to map a high-dimensional input into a low-dimensional latent representation and a decoder, which attempts to reconstruct the original input from a latent representation. The encoder consists of 4 successive downsampling convolutional blocks, each of which includes a 1D-Convolution layer, a Bach Normalization layer and a LeakyReLu activation function. Similarly, the decoder consists of 4 successive upsampling convolutional blocks, which in turn incorporates Upsampling layers, so as to restore the initial dimensions of the input time-series, followed by 1D Convolution layers. LeakyReLu activations were applied after each block as well, except for the final layer. Also, we apply a dropout layer in order to avoid over-fitting. At the top of the network, we apply a fully connected layer.

B. Training and Evaluation of Anomalies

In our analysis, we separate the normal data in three sets, i.e., train, validation and test set, with a split of 60–20–20(%), respectively. We normalize all data in the $[0, 1]$ range, except

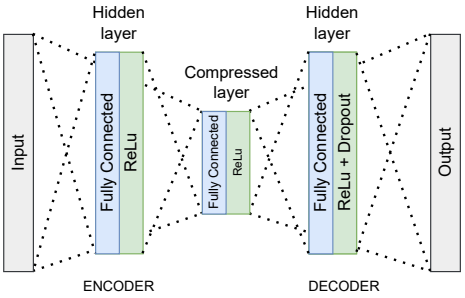

Fig. 4. FNN AE architecture.

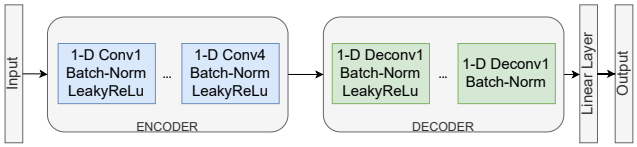

Fig. 5. CNN AE architecture.

from the time features that are already in $[-1, 1]$. The train set, contains only "normal" data, while the validation and test sets contain both data without and with anomalies (relapses).

First, we train each architecture using only "normal" data. To access the performance of our models, we evaluate it to "unseen" normal and relapse data. Intuitively, we expect to observe small reconstruction errors in "unseen" normal data and larger errors in relapse data. The reconstruction error vector with size $l \times d$ of each point i is calculated as the mean absolute error (MAE), between the predictions $\hat{x}^{(i)}$ and the given data $x^{(i)}$. The error vectors $e^{(i)}$ in the validation set, are used to compute the mean (μ) and covariance (Σ) of a multivariate normal distribution that is the expected error distribution. Then, we extract the Mahalanobis distance referred to as the "anomaly score" between the predicted points in the test set and the Gaussian distribution that was calculated in the validation set, as follows [24]:

$$a^{(i)} = \sqrt{(e^{(i)} - \mu)^T \Sigma^{-1} (e^{(i)} - \mu)} \quad (1)$$

In our work, the anomaly scores and the performance are day-averaged similarly to [14]. According to related literature [14], [24] the maximization of a performance metric, such as F-beta score or ratio of True Positive and False Positive Rates, i.e., TPR/FPR, can be used for choosing the threshold. In our study, we evaluate the performance of our models under multiple thresholds using Receiver Operating Characteristic Area Under Curve (ROC AUC) and Precision-Recall Area Under Curve (PR AUC). In our experiments we performed a 5-fold cross validation and the median of this procedure is reported as the final score.

This evaluation is performed for each patient separately, aiming in a **personalized scheme**, since it is a logical starting point for the specific task [13]. However, we explored the potential of the suggested models to scale in a **global scheme** as well; thus, we train our models on data corresponding to all patients and evaluating them either *globally* (global scheme, tested to all patients) or *individually* (global scheme evaluated individually, thus per patient).

Finally, we used Pytorch for the implementation of the architectures, which were trained using the Mean Square Error

TABLE II
RESULTS FOR PR-AUC (PERSONALIZED SCHEME).

Patients	FNN	CNN	Transformer	GRU	Random
#1	0.94	0.95	**0.97**	0.91	0.91
#2	**0.05**	0.04	0.02	0.03	0.03
#3	**0.54**	0.46	0.43	0.44	0.53
#4	0.26	**0.34**	0.18	0.19	0.18
#5	**0.63**	0.57	0.60	0.61	0.63
#6	0.70	**0.72**	0.63	0.67	0.68
#7	0.82	0.86	**0.87**	0.85	0.86
#8	0.83	**0.87**	0.65	0.81	0.85
#9	0.79	**0.80**	0.45	0.75	0.68
#10	0.97	0.95	0.94	**0.97**	0.97
Median	0.75	**0.76**	0.61	0.71	0.68

TABLE III
RESULTS FOR ROC-AUC (PERSONALIZED SCHEME).

Patients	FNN	CNN	Transformer	GRU	Random
#1	0.94	0.96	**0.97**	0.93	0.91
#2	**0.49**	0.40	0.22	0.36	0.28
#3	**0.57**	0.53	0.49	0.49	0.52
#4	**0.39**	0.39	0.35	0.29	0.22
#5	0.44	0.28	**0.45**	0.40	0.42
#6	**0.49**	0.49	0.39	0.42	0.48
#7	0.56	**0.69**	0.69	0.64	0.62
#8	0.72	**0.78**	0.64	0.60	0.72
#9	**0.78**	0.75	0.28	0.58	0.42
#10	0.91	0.88	0.81	**0.94**	0.91
Median	0.57	**0.61**	0.47	0.54	0.50

(MSE) as a loss function and a batch size of 64. For the Transformer and the FNN AE we used Adam optimizer [25], while RMSprop optimizer was used for CNN AE and GRU AE. All models had learning rate of 0.0001, except for the Transformer that had 0.001. The training has been performed for 50 epochs, while early stopping was applied to monitor the model's performance, using the validation loss.

IV. RESULTS AND DISCUSSION

Tables II and III depict the results of our experiments for the personalized scheme. With bold are shown the best results for each patient and each architecture. We observe that the best performance is obtained by the CNN AE model, while patient #1, who suffered a moderate relapse of about 11 days, has the best performance with PR and ROC AUC score at 0.97 in the Transformer model. In order to have a baseline for our experiments we implemented a random classifier, where we classify the data without training the models. Specifically, we calculate the mean and the covariance in the validation set and then the anomaly scores in the test sets. The last column (Random) of the aforementioned tables presents this baseline results for each patient, showing that all personalized models' results surpass it. In Fig. 6, we present the anomaly score of the test set for patient #1. The anomaly score to the right of the dividing line regards the relapse days and to the left the normal days. We observe that the anomaly score during relapse days is higher than in normal days. Note that the Days on the x-axis are not continuous.

In Table IV we present the results of our experiments for the global scheme and the global scheme evaluated individually. We observe that the FNN AE model, which was evaluated individually has the best performance with PR and ROC AUC of 0.77 and 0.62, respectively. As observed, the global scheme

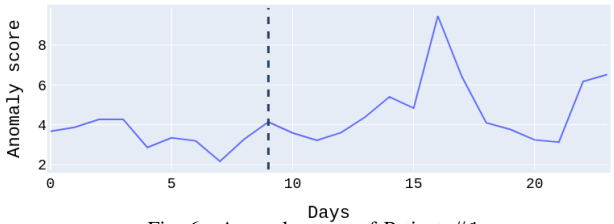

Fig. 6. Anomaly score of Patient #1.

TABLE IV
RESULTS FOR PR AND ROC AUC (GLOBAL SCHEME (GLOBAL) AND GLOBAL SCHEME EVALUATED INDIVIDUALLY (MEDIAN)).

	FNN	CNN	Transformer	GRU	Random
PR AUC					
Median	**0.77**	0.71	0.76	0.73	0.68
Global	0.48	0.49	0.47	**0.52**	0.50
ROC AUC					
Median	**0.62**	0.58	0.52	0.57	0.50
Global	0.47	0.51	0.45	**0.53**	0.50

in general presents lower performance than the models that were evaluated individually.

Concluding, we note that the global scheme has lower performance than the personalized schemes. This may be due to the fact that several of the relapses (regarding all patients that were evaluated) had low severity, causing difficulty to the global models in detecting them. To verify this claim, we performed a post hoc analysis, presented next. Finally, we observe that the patients that had moderate and severe relapses yielded better results than the others patients. The difference of the personalized CNN AE and the global FNN AE model that was evaluated individually is relatively small (by 0.01).

Generally, we observe that in the personalized CNN AE and FNN AE models that were evaluated individually, patients #1 and #10 had the best performance with a moderate and a severe relapse, respectively. Both patients also recorded an increased performance in the case of the baseline, since the features had significant differences between the normal and the relapse periods in comparison to other patients with lower performance. We have to note that patient #2 had only 2 relapse days of low severity thus, our models assume that almost all data are normal, and cannot detect the relapses. Moreover, in cases of limited data of normal and relapse periods, the performance of our approach is slightly low. Finally, we contacted t-tests for statistical significance to observe differences between the results of our architectures and the random classifier, where we noticed statistical significance for 6/10 patients, with p-value lower than 0.05, while patients with more relapse days and data, i.e., patients #6 and #7 had better results related to the statistical significance tests.

In our study, a total of 3 patients relapsed multiple times with different severity levels (Low and Moderate). We performed a post hoc analysis using the best-performing models (personalized CNN AE model and global FNN AE model that was evaluated individually) to compare the reconstruction errors between the severity levels. Table V presents the median of the reconstruction error of the personalized CNN AE model and the global FNN AE model that was evaluated individually. Specifically, on both models, we observe that for patient #7

TABLE V
RECONSTRUCTION ERROR OF THE BEST PERFORMING MODELS FOR LOW AND MODERATE RELAPSES.

	CNN AE		FNN AE	
Patients	**Low**	**Moderate**	**Low**	**Moderate**
#5	0.007824	0.006617	0.004493	0.004342
#6	0.005098	**0.005481**	0.003960	0.003824
#7	0.006257	**0.006515**	0.005915	**0.005951**

TABLE VI
RECONSTRUCTION ERROR OF THE BEST PERFORMING MODELS (GLOBAL SCHEME) FOR LOW, MODERATE AND SEVERE RELAPSES.

Patients	**FNN AE**	**CNN AE**
Low	0.002598	0.003650
Moderate	0.002629	0.003759
Severe	**0.002796**	**0.004076**

(who had 148 normal days and 38 and 35 low and moderate relapse days, respectively) the moderate relapse reconstruction error is higher than the low relapse. On the other hand, for patient #5 the reconstruction error of the moderate relapse is slightly lower compared to the reconstruction error of the low relapse days. This is probably due to the fact that for this patient we had limited low relapse data (2 hours), 6 moderate relapse and 73 normal days, respectively, thus, our results may not be reliable. Lastly, for patient #6 (having 143 normal days, 12 hours and 32 days of low and moderate relapse, respectively) the moderate relapse reconstruction error is higher than the low relapse reconstruction error for the personalized CNN AE model, while for the global FNN AE model that was evaluated individually the reconstruction error for the low relapse is slightly higher than the moderate.

We further examined the reconstruction errors of the global FNN AE and CNN AE for low, moderate and severe relapses, separately. Specifically, we collected relapse data per severity category for all patients together. As shown in Table VI in both models there is a gradual increase of the reconstruction error in relation to the severity of the relapse.

V. CONCLUSIONS

In this work, we provide promising results for relapse detection of patients suffering from various psychotic disorders, exploiting smartwatch data. Personalized autoencoder models, which are more appropriate due to the peculiarity of the specific task, have been evaluated for each patient obtaining encouraging results for the CNN AE model. Moreover, we experimented with global models as well as global models evaluated individually, obtaining the best performance for the latter and specifically the FNN AE model. In the proposed work, CNN AE and FNN AE architectures have shown to perform more robustly compared to the others. In addition, we notice statistical significance of our results between our architectures and the random classifier for most of the patients. Furthermore, we observed that the more severe the relapse is, the easier it is to detect. Concluding, for future work we aim to include more patients, experiment with other and possible more informative feature representations and examine the differentiations that could be found when the patients are awake or asleep. Finally, we intent to utilize and experiment with more advanced architectures.

REFERENCES

[1] J. Torous et al., "New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research," *JMIR Mental Health*, vol. 3, no. 2, pp. e5165, 2016.

[2] S. Patel et al., "A review of wearable sensors and systems with application in rehabilitation," *J. Neuroengineering and Rehabilitation*, vol. 9, pp. 21, 2012.

[3] C. Boletsis, S. McCallum, and B. F. Landmark, "The use of smartwatches for health monitoring in home-based dementia care," in *Proc. Int. Conf. Human Aspects IT Aged Population*, 2015.

[4] P. Schmidt, A. Reiss, R. Duerichen, et al., "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proc. of the 20th ACM Int. Conf. on Multimodal Interaction*, 2018.

[5] A. Supratak et al., "Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg," *IEEE Trans. on Neural Systems and Rehabil. Engin.*, vol. 25, pp. 1998–2008, 2017.

[6] A. Saeed, T. Ozcelebi, and J. Lukkien, "Multi-task self-supervised learning for human activity detection," *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, pp. 1–30, 2019.

[7] A.K. Yetisen et al., "Wearables in medicine," *Advanced Materials*, vol. 30, pp. 1706910, 2018.

[8] M. H. Aung et al., "Sensing behavioral symptoms of mental health and delivering personalized interventions using mobile technologies," *Depression and Anxiety*, vol. 34, pp. 603–609, 2017.

[9] S. Saeb, M. Zhang, C. Karr, et al., "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study," *Journal of Medical Internet Research*, vol. 17, pp. e175, 2015.

[10] F. Gravenhorst, A. Muaremi, J. Bardram, et al., "Mobile phones as medical devices in mental disorder treatment: an overview," *Personal and Ubiquitous Computing*, vol. 19, pp. 335–353, 2015.

[11] R. Wang et al., "Crosscheck: toward passive sensing and detection of mental health changes in people with schizophrenia," in *Proc. of the ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing*, 2016.

[12] J. L. Reyes-Ortiz, L. Oneto, A. Ghio, et al., "Human activity recognition on smartphones with awareness of basic activities and postural transitions," in *Proc. Int'l Conf. on Artificial Neural Networks*, 2014.

[13] I. Barnett et al., "Relapse prediction in schizophrenia through digital phenotyping: A pilot study," *Neuropsychopharmacology*, 2018.

[14] D. A. Adler et al., "Predicting early warning signs of psychotic relapse from passive sensing data: an approach using encoder-decoder neural networks," *JMIR mHealth and uHealth*, vol. 8, no. 8, 2020.

[15] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors J.*, vol. 15, 2014.

[16] D. Robotham et al., "Do we still have a digital divide in mental health? A five-year survey follow-up," *J. Medical Internet Research*, vol. 18, pp. e309, 2016.

[17] V.A. Canady, "Psychiatrists, patients favor telemedicine; challenges exist," *Mental Health Weekly*, vol. 30, no. 20, pp. 4–5, 2020.

[18] C Garoufis, A Zlatintsi, P.P. Filntisis, N Efthymiou, E Kalisperakis, V Garyfalli, T Karantinos, L Mantonakis, N Smyrnis, and P Maragos, "An unsupervised learning approach for detecting relapses from spontaneous speech in patients with psychosis," in *Proc. BHI*, 2021.

[19] G. Retsinas, P. P. Filntisis, N. Efthymiou, E. Theodosis, A. Zlatintsi, and P. Maragos, "Person identification using deep convolutional neural networks on short-term signals from wearable sensors," in *Proc. Int'l Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, 2020.

[20] I. Maglogiannis, A. Zlatintsi, A. Menychtas, D. Papadimatos, P.P. Filntisis, N. Efthymiou, G. Retsinas, P. Tsanakas, and P. Maragos, "An intelligent cloud-based platform for effective monitoring of patients with psychotic disorders," in *Proc. AIAI*, 2020.

[21] P. P. Filntisis, A. Zlatintsi, N. Efthymiou, E. Kalisperakis, T. Karantinos, M. Lazaridi, N. Smyrnis, and P. Maragos, "Identifying differences in physical activity and autonomic function patterns between psychotic patients and controls over a long period of continuous monitoring using wearable sensors," *arXiv*, vol. abs/2011.02285, 2020.

[22] A. Vaswani et al., "Attention is all you need," in *in Proc. Int'l Conf. on Neural Information Processing Systems*, 2017.

[23] M.-T. Luong et al., "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[24] P. Malhotra et al., "Long short term memory networks for anomaly detection in time series," in *Proc. ESANN*, 2015.

[25] D.P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *in Proc. Int'l Conf. on Learning Representations, ICLR*, 2015.