ARIMA-based Motor Anomaly Detection in Schizophrenia Inpatients

Talia Tron[1,4] Yehezkel S. Resheff[1,4] Mikhail Bazhmin[2] Daphna Weinshall[4] Abraham Peled[2,3]

***Abstract*— Motor alteration is an important aspect of the elusive schizophrenia disorder, manifested both throughout the various phases of the disease and as a response to treatment. Tracking of patients' movement, and especially in a closed ward hospital setting, can therefore shed light on the dynamics of the disease, and help alert staff to possible deterioration and adverse effects of medication. In this paper we describe the use of ARIMA-based anomaly detection for monitoring of patient motor activity in a closed ward hospital setting. We demonstrate the utility of the approach in several intriguing case studies.**

I. INTRODUCTION

Monitoring of motor behavior is part of the regular assessment of schizophrenia patients and is vital to diagnosis, progress assessment and to the monitoring of medication response. Various alterations of motor behavior are evident throughout the phases of the disease, and as a response to treatment. The psychotic acute phase of schizophrenia is typically accompanied by restlessness, including occasional bizarre movements and gestures, while post psychotic deficiency negative symptoms are related to reduced activity, slowness and even freezing. Antypsychotic medications may cause Parkinsonism, i.e., tremor, rigidity, and slowness, which usually pass after the first week of treatment.

Despite its clinical and diagnostic value, to date, motor monitoring is done in a descriptive non etiological manner based on subjective clinical scales, which may result in biased, inaccurate and typically non quantifiable assessments. This kind of assessment requires expert staff and the availability of resources, and may not be frequent enough to capture significant changes in spontaneous and drug-induced conditions. These issues can be alleviated by carrying out objective, continuous quantifiable monitoring [1], the investigation of which is the goal of this study. Accelerometers and gyroscopes, commonly embedded in smart-watches and other wearable devices, have been extensively used over the last decades in medical applications ranging from human physiology [2] to movement disorders [3] and mental healthcare [4]. These cheap and widely available sensors may be used for continuous qualitative patient monitoring in natural clinical settings. Accelerometer data have already been shown to effectively provide insights into patients clinical state, and motor features were successfully used for clinical sub-typing in a closed ward mental hospital setting [5], [6]. Here we focus on detecting acute abnormal behaviors which are either the result or the cause of drug modifications or changes in patients' clinical conditions. Our approach employs forecasting models widely used in statistics and econometrics, applied to step-count data. We demonstrate the utility of this approach with 4 schizophrenia case studies, in which we evaluate monitoring performance based on medical and clinical records.

[1]The Edmond and Lily Safra center for Brain Science, Hebrew University, Jerusalem 91904, Israel talia.tron@mail.huji.ac.il

[2]Rappaport Faculty of Medicine, Technion Institute of Technology, Haifa 3200003, Israel

[3]Sha'ar Menashe Mental Health Center, Sha'ar Menashe 38706, Israel

[4]The Rachel and Selim Benin School of Computer Science and Engineering, Hebrew University, Jerusalem 91904, Israel

II. MATERIALS AND METHODS

A. Study Design

Four inpatients from the closed ward at Shaar-Menashe mental health center, diagnosed with schizophrenia according to the DSM-5, participated in the study. One patient (patient B) was diagnosed with paranoid schizophrenia. Participants' age varied from 24 to 54 (average 36.9), with course of illness varying from 7 to 35 years (average of 13.5 years). After signing the appropriate Helsinki legal consents, participants were tracked for a period of approximately one month (27-31 days) in natural settings. During this period, patients were monitored for medication use (type, dosage, and frequency) by the nurses and the physicians. In addition, every patient underwent a clinical evaluation of Positive and Negative Syndrome Scale (PANSS [7]) and Neurological Evaluation Scale (NES [8]) by a trained psychiatrist twice a week. The neurological evaluation was only utilized to confirm that no psycho-motor deficits were evident in any of the participants during the experiment.

All procedures performed in the study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

B. Data Acquisition

At study onset, participants were given smart-watches with embedded accelerometers (GeneActiv[1]). These watches were worn on the wrist throughout the experiment. The output ($50Hz$) of the sensors was stored on internal memory cards. The study was conducted in natural settings, where patients were *not* required to change any personal or medical procedure. None of the patients expressed any discomfort or disturbance from wearing the device.

III. DATA ANALYSIS

A. Building personal ARIMA Models

Analysis focused on the walking pattern of patients, aiming to detect significant quantitative changes. Stepping

[1]https://www.activinsights.com/products/geneactiv/

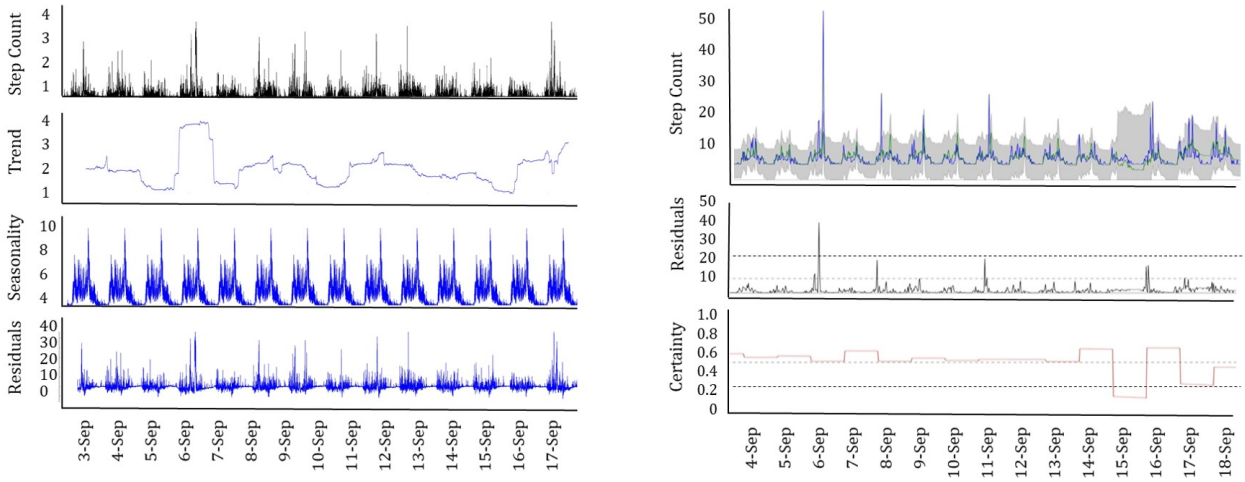

Fig. 1. **Left**- Decomposition of daily steps (top) of a single patient to trend (smoothed series calculated using centered moving average), seasonality (regularly repeating data patterns calculated as the average of the smoothed series for each period) and noise. **Right**- Demonstration of the ARIMA model for patient A. The model returns the predicted mean and a 95% confidence interval (CI) around it. Abnormal behavior is detected when (a) the observed step count value lies outside the CI predicted by the model, (b) the residuals are higher than threshold (e.g. September 6), or (c) when certainty is lower than threshold (e.g. September 15).

behavior was detected as large maxima of the smoothed square norm of the 3-axial $50Hz$ point-wise acceleration, and the number of steps (step count) was averaged over 1 minute intervals (see [5] for further details).

We used AutoRegressive Integrated Moving Average (ARIMA) models to detect abnormal walking patterns. One week of data was used to predict the step count for the following day, together with the associated confidence interval. Repeating this in a rolling window design produced the predicted step count for the entire duration of available data, around 3 weeks for each participant excluding the first week. Predicted values were then compared to those observed in practice for the purpose of anomaly detection [9].

We began by decomposing the step-count data into trend, seasonality and noise components, as shown in the left side of Fig. 1. As expected, strong daily seasonality was seen in the data. It is interesting to note that the trend component, to the extent that it exists, may potentially be used for direct real-time monitoring of patients symptom severity over time.

Next, we aggregated each patient's step-count data in windows of 10-30 minutes (this was done to smooth the data on the one hand, and reduce computation on the other). Both regular and daily seasonal differentiation were computed to obtain a stationary signal. We applied 4 different ARIMA models to all patients, and evaluated them using AIC criteria with mean and absolute errors. The emerging preferred model was ARIMAX(1,1,1) seasonal (1,1,2), which had a consistent lower error and lower AIC over all patients.

B. Abnormal behavior detection

For each patient separately, we ran an ARIMAX(1,1,1) seasonal (1,1,2) model, which was based on 7 days of data in order to predict the following day. The model provided the predicted mean and a 95% confidence interval (CI) around it. Model residuals were calculated as the squared difference between the model predicted values and the observed values during the test period.

A measure of prediction certainty was calculated based on the normalized CI size ($|CI_z|$) as follow:

$$Certainty = 0.95 \times 2 \times \frac{std(data)}{|CI_z|} \tag{1}$$

This is a measure of model confidence, with low values indicating that the model hasn't been able to accurately predict future values based on the patient's history. The multiplier of 0.95 sets the maximum certainty value to 0.95 (model confidence level). Although certainty is somewhat correlated with residuals size, this is an important independent measure. Specifically, it covers cases where the observed value is lower than the predicted value, which is not always expressed in CI range or high residuals.

Abnormal behavior is defined as one the following (see right side of Fig. 1): (a) The predicted value is not in the model CI; (b) the residuals between model prediction and observed values are higher than threshold (set to be 3 times the mean residuals on train data); (c) the $certainty$ of the model is lower than threshold (0.3). In order to avoid trailing errors and secure robustness, when abnormal behavior is detected, the observed values of the training period are replaced with predicted values. On repeated detections (more than twice) the model is adjusted back to observed values.

C. Evaluating model performance

In order to evaluate our model we systematically studied the patients clinical records and drug charts, and compared them with model anomaly detections. No clear abnormal event, such as an outburst of violence or riot, was recorded during the experiment period. We therefore used the PANSS clinical records in order to identify *abnormal* events, which are time stamps corresponding with a steep increase or

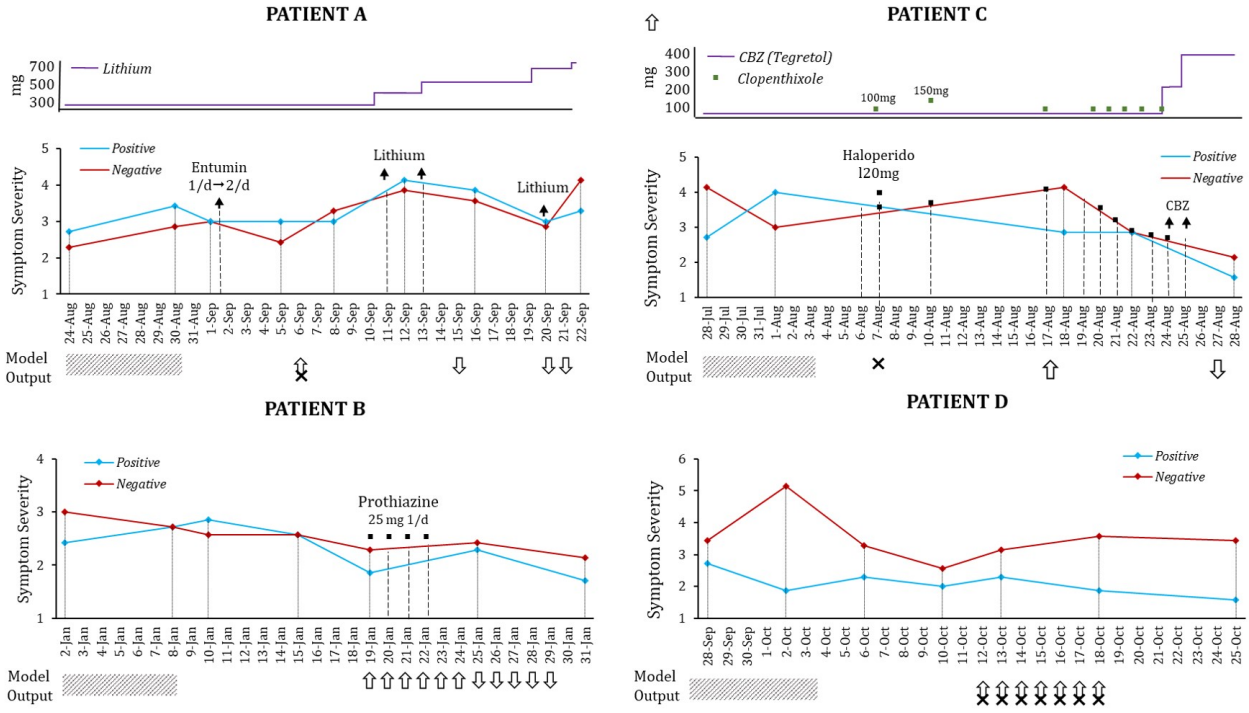

Fig. 2. Description of model prediction vs. clinical and medication records monitoring for all four patients. The direction of the white arrows in the bottom part of each graph indicates whether increased activity (up) or decreased activity (down) has been detected. A cross under the arrow indicates unexplained detection, while a cross without an arrow indicates an event that wasn't detected by the model. The dashed rectangle marks the training period of the model. In the line chart above, the mean severity of positive (red) and negative (blue) symptoms is shown. The black symbols indicate a change in drug dosage (arrow) or a single administration (square). In case of dosage change, the top graph (in patients A and C) indicates its amount (in mg).

decrease in symptom severity (more than 2 degrees on the PANSS scale) between two clinical sessions. Results are summarized in Fig. 2.

In an effort to capture some larger scale dynamics, we took note of the general positive and negative symptoms trend. Every change in drug dosage was also considered an abnormal event, since these changes are rare and usually indicate a change in a patient's clinical condition. It should be noted that increased drug dosage may be either a response to abnormal activity (when the detected event took place prior to drug adjustment) or its trigger (when the detected event followed a drug adjustment). Decreased dosage, on the other hand, is usually followed by continuous improvement in symptom severity, but may still cause side effects. Therefore, in order to obtain a coherent picture, both timing and the direction of the dosage change were taken into account.

For each abnormal event detected by our model, we looked for an explanation (as defined above) in the clinical records (drug dosage and PANSS scores); an event which did not have a satisfactory explanation, was labeled as 'unexplained'. Likewise, a drug change event or a steep change in the clinical evaluation data which was not detected by our model was labeled as 'undetected'. The number of unexplained and undetected events was used to roughly estimate the accuracy and sensitivity of our model. Events in consecutive days were counted as one continuous event.

1) Patient A: Abnormal increased walking behavior was detected on September 6th. On the same day, the dosage of *entumin* (a.k.a *clotiapine*), an atypical anti-psychotic drug, was increased from 40mg 1/day to 40mg 2/day.

On September 15th, and then again during September 20-22, our model detected lower than expected activity. In the clinical records, we see a significant increase in both positive and negative symptoms during September 5-12, with a steep rise in active social avoidance, hostility and social withdrawal. Possibly this behavioral change has resulted from the increased entumin dosage, although we cannot rule out other possible triggers.

Following this deterioration in the patient's condition, on September 11th the dosage of *lithium* was increased, and again on the 13th. Both positive and negative symptoms were reduced in subsequent days, with active social avoidance and hostility returning to normal values. We also see the emergence of increased negative symptoms, including blunted affect and passive apathetic social withdrawal.

Lithium is known to take effect within 1-3 weeks, so the lower activity found by our model during September 20-22 may be the result of the September 11th dosage increase. The September 15th detection remains unexplained by drug records but is congruent with clinical data.

In summary, 2/3 detected events for this patient had a co-found explanation in the clinical and medication records. One event had only a weak co-found in the clinical data. No clinical trend or drug changes remained undetected.

2) Patient B: The model detected a period of extreme increased activity during January 19-24, followed by decreased

activity during January 25-31. On January 19th, this patient was given *prothiazine*, a neuroleptic medication used as a sedative and weak anti-psychotic, for a period of 4 days. We found no significant change in symptom severity for this patient prior to the sedative drug administration, with only a small decrease in overall negative symptoms at that time. This is probably because clinical evaluation was not frequent enough to capture the change. The fact that our model detected this event while the clinical data did not, can be used as evidence for the potential benefit of continuous automated monitoring.

On January 25th, two days after the patient has stopped receiving the medication, we see a small improvement in his clinical condition with normal level of motor activity. In the model this is expressed by a detected 'lower than expected' activity, based on the increased activity in the previous days.

In summary, for this patient all detected events (2) had a co-found explanation in the medication records but no co-found (or a minor one) in the clinical records. No clinical trend or medication alteration remained undetected.

3) Patient C: Increased activity level was detected by the model on August 17th. Clinical data together with medical records clearly suggest that around this period there was an aggravation in the patient's condition. On August 17th, he was injected with 100mg of *clopenthixole acetate* (anti-psychotic and acute sedative medication), and once again in the following days (August 20-25). The drug's effect seems to have been dimmed unsatisfactory, since during August 24-25 the patient was also prescribed 200mg and then 400mg of *carbamazepine* (CBZ), an off label medication used in combination with anti-psychotics when the treatment with anti-psychotics alone has failed [10]. In the clinical data we see a decrease in both negative and positive symptoms severity around August 18-22, with a steep decrease in hallucinations, poor attention, and motor retardation. This improvement is most probably the result of the massive drug treatment. On August 27th, after the patients symptoms were reduced and drug treatment was stabilized, the model detected a significant reduction in patient's activity.

In August 7 the patient received two types of typical anti-psychotic medications (clopenthixole and *haloperidol*), and then again in August 10 (only *clopenthixole*). Since these drugs act on a short term basis, it is not probable that the the worsening in the patient's condition in subsequent days was triggered by this medication change. The most probable explanation is that there was some acute event at that time, which was not detected by our model.

In summary, all detected events (2) had a co-found explanation in the clinical and medication records, while one likely clinical event remained undetected.

4) Patient D: The model reported a period of decreased activity during October 12-18, with low certainty. No medication change was registered in this time period, and no substantial evidence was found in the clinical data (only a steep increase in stereotyped thinking). The overall trend of symptoms' change around that period leaned towards increased negative symptoms and reduced positive symptoms. This happened following approximately a week of steep decrease in negative symptoms.

In summary, the event detected by our model had no co-found explanation in the medication records. No clinical trend or medication alteration remained undetected.

TABLE I
SUMMARY OF ANOMALY DETECTION RESULTS AND PATIENTS' DATA.

	Days	Sessions	Explained	Missed
Patient A	31	10	2/3	0
Patient B	29	7	2/2	0
Patient C	31	11	2/2	1
Patient D	27	7	0/2	0

As summarized in Table I, when aggregating data from all patients, 6/8 anomaly events detected by our model had a co-found explanation in the medication and clinical records (precision of 75%). 6/7 events were detected by our model, with one certain mis-detection in patient C (recall of 85%).Other detected events may have alternative explanation not available to our experimental design.

IV. CONCLUSIONS

Our study demonstrates the benefits of using forecasting models in conjunction with accelerometer data for the continuous monitoring of schizophrenia patients. In three out of four case studies, we found a direct link between detected behavioral events and changes in the patient's clinical condition or drug regime.

REFERENCES

[1] Christopher G Goetz, John G Nutt, and Glenn T Stebbins. The unified dyskinesia rating scale: presentation and clinimetric profile. *Mov. Dis.*, 23(16):2398–2403, 2008.

[2] John Staudenmayer, Shai He, Amanda Hickey, Jeffer Sasaki, and Patty Freedson. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *Journal of Applied Physiology*, 119(4):396–403, 2015.

[3] Avishai Wagner, Naama Fixler, and Yehezkel S Resheff. A wavelet-based approach to moniotring parkinson's disease symptoms. *International Conference on Acoustics, Speech, and Signal Processing*, 2017.

[4] Rui Wang, Fanglin Chen, et al. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM Int Joint Conf on Pervasive and Ubiquitous Computing*, pages 3–14. ACM, 2014.

[5] Talia Tron, Yehezkel S Resheff, Mikhail Bazhmin, Abraham Peled, and Daphna Weinshall. Real-time schizophrenia monitoring using wearable motion sensitive devices. 2016.

[6] Talia Tron, Yehezkel S Resheff, Mikhail Bazhmin, Alexander Grinshpoon, Abraham Peled, and Daphna Weinshall. Topic models for automated motor analysis in schizophrenia patients*. In *Body sensor networks (BSN), 2013 IEEE international conference on*. IEEE, 2017.

[7] Stanley R Kay, Abraham Flszbein, and Lewis A Opfer. The positive and negative syndrome scale (panss) for schizophrenia. *Schizo. bulletin*, 13(2):261, 1987.

[8] Robert W Buchanan and Douglas W Heinrichs. The neurological evaluation scale (nes): a structured instrument for the assessment of neurological signs in schizophrenia. *Psychiatry research*, 27(3):335–350, 1989.

[9] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[10] Daniela Ceron-Litvoc, Bernardo Garcia Soares, Geddes, et al. Comparison of carbamazepine and lithium in treatment of bipolar disorder: a systematic review of randomized controlled trials. *Human Psychopharmacology: Clinical and Experimental*, 24(1):19–28, 2009.