Monitoring Motor Activity Data for Detecting Patients' Depression Using Data Augmentation and Privacy-Preserving Distributed Learning

Amin Aminifar[1], Fazle Rabbi[1,2], Violet Ka I Pun[1,3], and Yngve Lamo[1]

***Abstract*— Wearable devices are currently being considered for collecting personalized physiological information. Lately, such information is being used to provide healthcare services to individuals. One application is detecting depression by utilization of motor activity signals collected by the ActiGraph wearable wristbands. However, to develop an accurate classification model, we require to use a sufficient volume of data from several subjects, taking the sensitivity of such data into account. Therefore, in this paper, we present an approach for extracting classification models for predicting depression based on a new augmentation technique for motor activity data in a privacy-preserving fashion. We evaluate our approach against the state-of-the-art techniques and demonstrate its performance based on the mental health datasets associated with the Norwegian INTROducing Mental health through Adaptive Technology (INTROMAT) Project.**

I. INTRODUCTION

Mental health disorders are the primary contributor to chronic diseases in Europe [1]. Twenty-five percent of people develop at least one mental or behavioral disorder in their life [2]. Depression is the most prevalent among mental health disorders and this is expected to increase in the following years [3], [4], [5]. Therefore, addressing and controlling depression is necessary for society as it affects individuals' physical, emotional, and economic aspects [6].

Wearable devices provide the opportunity to monitor patients on a long-term basis to detect and prevent health disorders in earlier stages [7], [8], [9], [10]. Wearable technologies offer pervasive healthcare solutions at an affordable price by removing time and location restrictions [11]. The data collected by such devices has attracted a lot of attention for mental health applications [12]. One such application is detecting depression in patients based on motor activity data collected from ActiGraph wristband [13]. The motor activity is captured by the accelerometry signals acquired by the ActiGraph wristband. Figure 1 explains a scenario for the analysis of sensor data generated by wearable devices. In this figure, the individuals' activity signal is collected by a wristband. The signal is transferred to the personal mobile phone. Then, the raw data may be preprocessed and prepared for the analysis task locally on the phone or analyzed in a distributed fashion [14], [15].

Monitoring mental health and, in particular, depression by using signals collected by wearable devices involve several challenges. First, sharing healthcare data for analysis purposes is not always feasible due to privacy and legal concerns [16], [17], [18], [19]. In particular, privacy and security are among the most concerning challenges in real-time health monitoring using mobile health technologies [20], [21], [22]. Privacy-preserving data sharing, e.g., [23], [24], [25], and privacy-preserving data mining [26], [27], [28], [29], [30], [15] approaches offer a solution for data analysis without the raw data leaving the individuals' devices. Secondly, although there is a connection between mental health problems and disturbance in internal biological systems, relations between mood and physiological signals are not well-identified [31], [13]. Therefore, finding the correlation between physiological signals and mental health problems is challenging.

This paper addresses analyzing motor activity data collected by the ActiGraph wristband. We use the Depresjon (Depresjon is the Norwegian word for depression) dataset [13] which contains motor activity signals of patients from control (non-depressed) and condition (depressed) groups. Our goal is to predict depression in patients based on such data. Previous studies [13] have considered a feature-based approach for the detection of depression. However, as we show in this paper, the prediction performance may be improved by further exploiting the information carried in the signals (beyond the basic statistical attributes, e.g., the mean and standard deviation).

In this paper, we propose an augmentation approach for generating new records from the Depresjon dataset [13] to improve the classification performance. In other words, our approach produces new data records from the raw data in order to use them for the learning and evaluation process. We show that adopting our augmentation approach leads to learning classification models with higher performance. However, the motor activity raw data that is required for the analysis is generated on each patient's wearable device and inherently distributed. Such data cannot be transferred to a center for further analysis due to personal and/or legal privacy concerns (e.g., to infer mental health status from the data). To address this privacy issue, we investigate the possibility of using our recently proposed privacy-preserving distributed machine learning approach, PPD-ERT [32], for sensor data based on the Depresjon dataset, which paves the way for the real-world applications of our approach for wearable technology in the described settings.

The remainder of this article is structured as follows: The approach and details about generating records from the Depresjon dataset are described in Section II. The evaluation of the approach and experimental results are presented in Section III. Finally, Section IV concludes the paper.

[1]Western Norway University of Applied Sciences, Bergen, Norway firstname.lastname@hvl.no
[2]University of Bergen, Bergen, Norway
[3]University of Oslo, Oslo, Norway

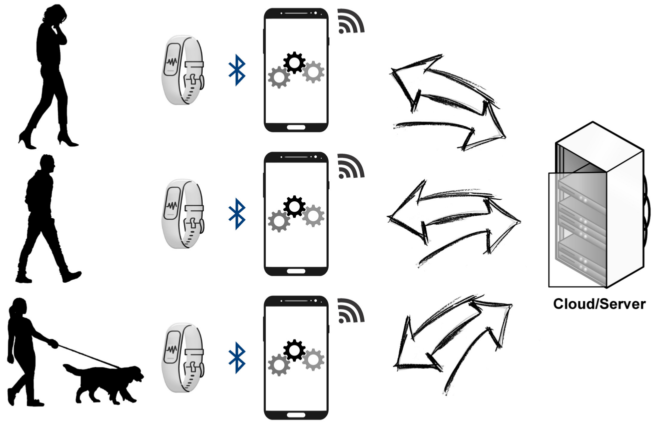


Fig. 1: Analysis of sensor data generated by wearable devices

II. Approach

This section presents our approach to detecting depression based on the motor activity data. First, we describe our approach for generating data records from the Depresjon dataset. Then, we discuss how PPD-ERT [32] is utilized for privacy-preserving distributed learning of classification models from generated data records and in the context of real-world wearable devices.

Let us first give an overview of the Depresjon dataset. The Depresjon dataset consists of motor activity data of 55 patients (30 and 25 for females and males) collected by ActiGraph wristband worn at the right wrist of patients. In this dataset, 23 of the patients are diagnosed with depression, including both unipolar and bipolar patients, and the remaining 32 belong to the control group. Each patient wore the ActiGraph wristband for an arbitrary number of days, between 5 to 20 days. The total number of days for the condition group is 291 days, and for the control group is 402.

The recorded values (samples) for each patient in each minute are proportionate to the quantity, duration, and strength of the patient's movements. Each patient has at least a sample value greater than or equal to zero for every minute of the day. It should also be noted that, on the first day for each patient, the recording started in the middle of the day. We refer to the data for each day of each patient as a record. Each record consists of several sample values (or samples in short).

The authors in [13] proposed the application of the mean and the standard deviation of the activity level along with the proportion of minutes with no activity in a day as the data attributes for depression classification. In addition, a normalization between zero and one is performed for attribute values. Therefore, this approach leads to only 693 records (291 for the condition group and 402 for the control group), one for each day in the raw data.

Although adopting the proposed approach in [13] extracts a representation of the raw data that results in a fair classification performance, it may still lead to suboptimal results. In this dataset, the total number of recorded data for patients is limited, i.e., only 693 days. Therefore, if we generate one record for each day, the volume of the data that the algorithm is trained on will be small, which in turn limits the detection performance. Moreover, the motor activity signal on certain days are shorter, where we do not have a recorded sample for every minute of the day. In this way, the mean, standard deviation, and the proportion of zero activity for that data are affected and will be very different from the days with complete recording. On the other hand, the number of recorded days for each patient is different. We have less than one week of recorded activity for some of the patients, while we have almost three weeks for some others. Therefore, the approach presented in [13] makes the data more imbalanced, which may eventually lead to poor classification performance.

This paper adopts a data augmentation approach for generating data records from the original Depresjon dataset. Data augmentation is a functional approach for increasing the diversity and volume of data by augmenting records at random [33], [34]. The majority of machine learning algorithms, e.g., deep neural networks, learn higher performance classification models when they are trained on larger datasets. Moreover, data augmentation can lead to better generalization and robustness by learning models invariant to the transformation of the data, e.g., learning an object classifier model that can classify objects correctly even if the images are rotated.

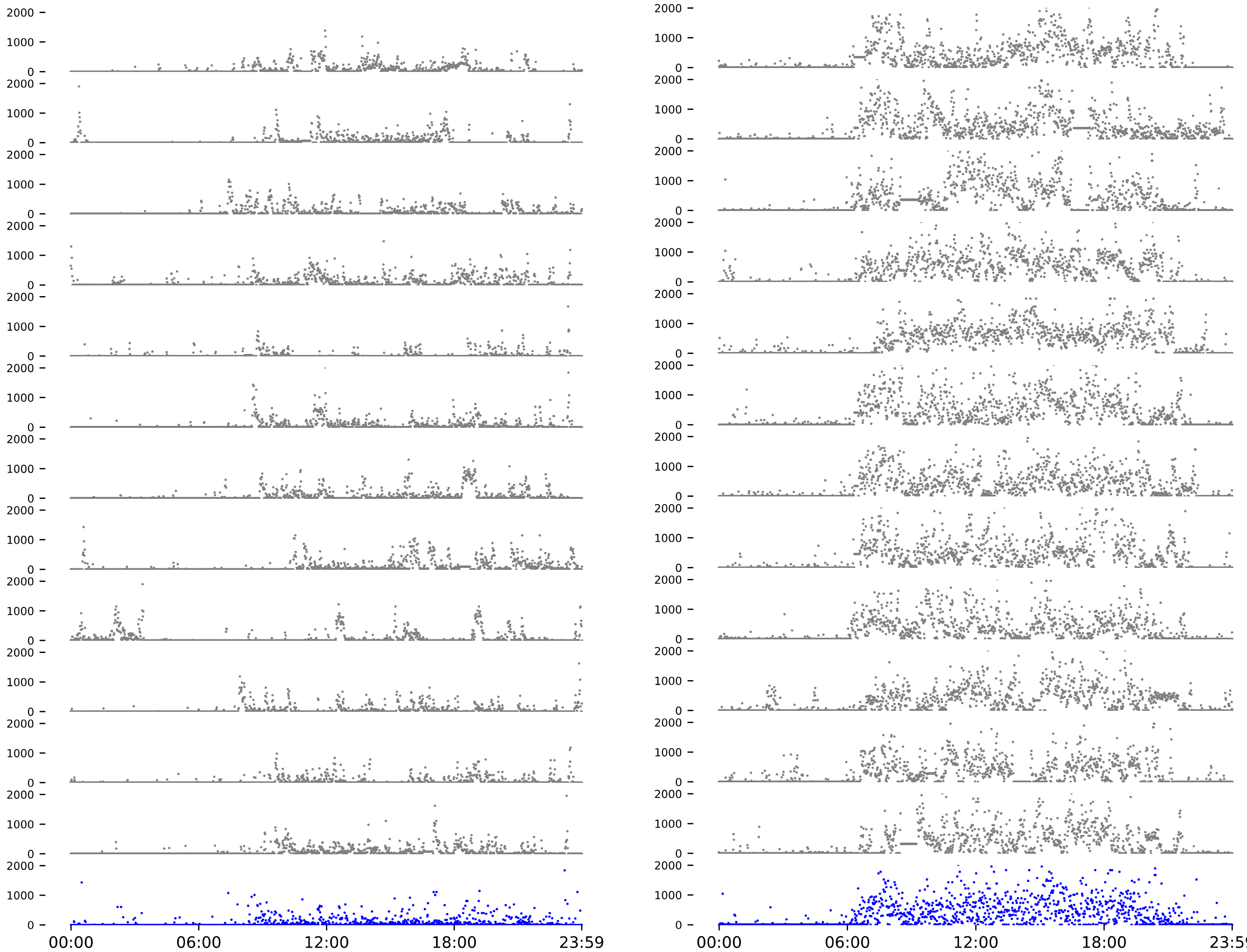
By adopting a data augmentation approach, we generate an equal number of records for each patient. All the generated records will have a unique size equal to the number of minutes in a day. For each patient, we generate n records, where n can be adjusted based on the user needs. The length of each record, l, is equal to the number of minutes in a day, i.e., $l = 1440$ (60×24), representing the patient activity level in one day.

Let us denote the set of all samples for patient i by S_i and define it as: $S_i = \{s_{ijk} \in R_{ij}, \forall j, k\}$. R_{ij} captures the j-th record of patient i and s_{ijk} is the sample k in record R_{ij}. For every minute t during the day, we check the available samples for this patient and for this specific time in the day, e.g., t = *12:00*. The recorded samples for different days of this patient around this particular time, i.e., $t \pm \delta$, are the candidates for being selected as the new (augmented) sample for this timestamp, where $2 \cdot \delta$ is the duration of this interval.

$\hat{R}_{ij}$ captures the j-th augmented record of patient i and is defined as: $\hat{R}_{ij} = [\hat{s}_{ij1}, ..., \hat{s}_{ijl}]$. $\hat{s}_{ijk}$ denotes the k-th sample for the generated record $\hat{R}_{ij}$, where $1 \leq k \leq l$. $\hat{s}_{ijk}$ is the sample at time t and is selected at random from set S_i and in the time interval $[t - \delta, t + \delta]$. This is formally defined as $\hat{s}_{ijk} \in \{s \in S_i | t - \delta \leq t(s) \leq t + \delta\}$, where $t(s)$ is the time of sample s. This process is repeated until we have n records for each patient.[1]

The augmented record reflects the patients' activity level in a day and is proportionate to the original data since its

[1]The source code of our approach is available at https://github.com/AminAminifar/dataprep



(a) Subject 20 from condition group

(b) Subject 20 from control group

Fig. 2: The figure presents two examples of generated records based on the raw data for individuals in condition and control groups. Each signal/record showed by gray dots represents one recorded day of the patients. The last signal shown by blue dots represents the generated record by our proposed approach ($l = 1440$ and $\delta = 10$). The examples show the correspondence of raw data and generated records and that the subject in the condition group usually has less activity, after the sleeping time, compared to the one from the control group.

samples for all timestamps (t) are randomly chosen from samples for the close timestamps ($t \pm \delta$) in the reported days. Therefore, the approach preserves the changes in the activity level of the patients in the data. This is particularly important, as studies found evidence that suggests a relationship between decreased daytime motor activity and increased nighttime activity and a depressive state, compared to healthy individuals [35], [13]. In similar studies, the decreased motor activity and more diversity in the activity level are reported for patients suffering from bipolar depression [36]. That being said, this means that preservation of activity level changes during the day for a patient is one of the main advantages of our augmentation approach.

Figure 2 shows the generation of records from two patients' raw data. The horizontal axis represents the time in a day, and the vertical axis shows the activity level. All patient's activity levels at different timestamps are shown in the figure by gray bubbles, and the blue bubbles are the samples for the generated record by our augmentation technique. Figures 2a and 2b show the raw signals/records (twelve days) and the augmented record for Subject 20 from the condition group and Subject 20 from the control group, respectively. The figure shows the association of the generated record and the raw data. In the intervals that the patient usually has a low level of activity, e.g., from midnight to the morning, the generated record also shows a low level of activity and vice versa.

The augmented dataset can then be used by the machine learning algorithms for the detection of depression. The raw data generated from each patient's activity are stored on patients' personal devices. Due to privacy and legal issues, such data cannot be transferred to a center for analysis in such healthcare applications. A practical solution in such situations is performing analysis through privacy-preserving distributed data analysis methods. Therefore, here we adopt our proposed PPD-ERT algorithms [32], [37] for analyzing

TABLE I: Classification performance (leave one patient out) of different classification algorithms for the approach in [13] and the generated data records based on our approach

Algorithms		Our approach			Approach in [13]		
		F1-score	ACC	MCC	F1-score	ACC	MCC
Distributed	PPD-ERT	**76.3%**	**76.8%**	**0.518**	66.3%	67.0%	0.310
	Distributed ID3	65.1%	65.0%	0.286	65.6%	66.5%	0.296
Centralized	ERT	76.3%	76.8%	0.518	66.3%	67.0%	0.310
	Random forest	74.4%	75.1%	0.481	64.3%	64.7%	0.266
	XGBoost	76.2%	76.3%	0.510	64.3%	64.7%	0.265
	Decision Tree	65.7%	65.8%	0.293	60.6%	60.7%	0.191
	Linear SVM	69.5%	69.5%	0.375	68.4%	68.6%	0.349

the Depresjon dataset and learning the classification model. The ensemble learning procedure adopted by PPD-ERT reduces the risks of overfitting.

The described approach for generating data instances (augmenting data) is compatible with our privacy-preserving distributed methods. This is because the new records are generated merely based on one patient's raw data and are independent of other patients' data. Therefore, each patient generates the instances on its own device locally. Then, the generated records are the data that is used for training the PPD-ERT algorithm. By employing the PPD-ERT approach, we learn high-performance classification models without sharing raw data or sensitive information. The learned models will then be used for detecting depression by each individual.

III. Evaluation and Discussion

In this section, we evaluate our proposed augmentation technique for motor activity data. We consider several classification algorithms to assess and compare the results obtained from our proposed approach and the approach in [13]. Moreover, we use our recently proposed method, PPD-ERT [32], for the described problem, i.e., detection of depression in patients based on motor activity data, to investigate the possibility of applying this method for such data from wearable devices.

The objective here is to learn classification models to detect depression based on the motor activity signals collected by the ActiGraph wristband. The trained model will later be used to detect depression in individuals based on their activity levels. The target categories for classification are two, i.e., normal/control category and depressed/condition category.

As described in Section II, [13] proposes using a dataset (obtained from original data) which contains three attributes (i.e., mean, standard deviation, and zero activity ratio) and one label for each record, and each record represents one day of collected data for one patient. This is while our approach generates records that contain a sample for each minute during the day, i.e., 1440 attributes for each record ($l = 1440$). In our experiments, we generate 100 records for each patient ($n = 100$). Every record is generated based on the samples collected at different days of a patient's collected signals. Each timestamp's sample for the record is selected among the available samples in 10 minutes time span around it ($\delta = 10$). Therefore, in both approaches, each generated record belongs to one and only one patient. This provides the possibility for leave-one-patient-out evaluation, which in turn enables the adoption of our privacy-preserving distributed learning framework.

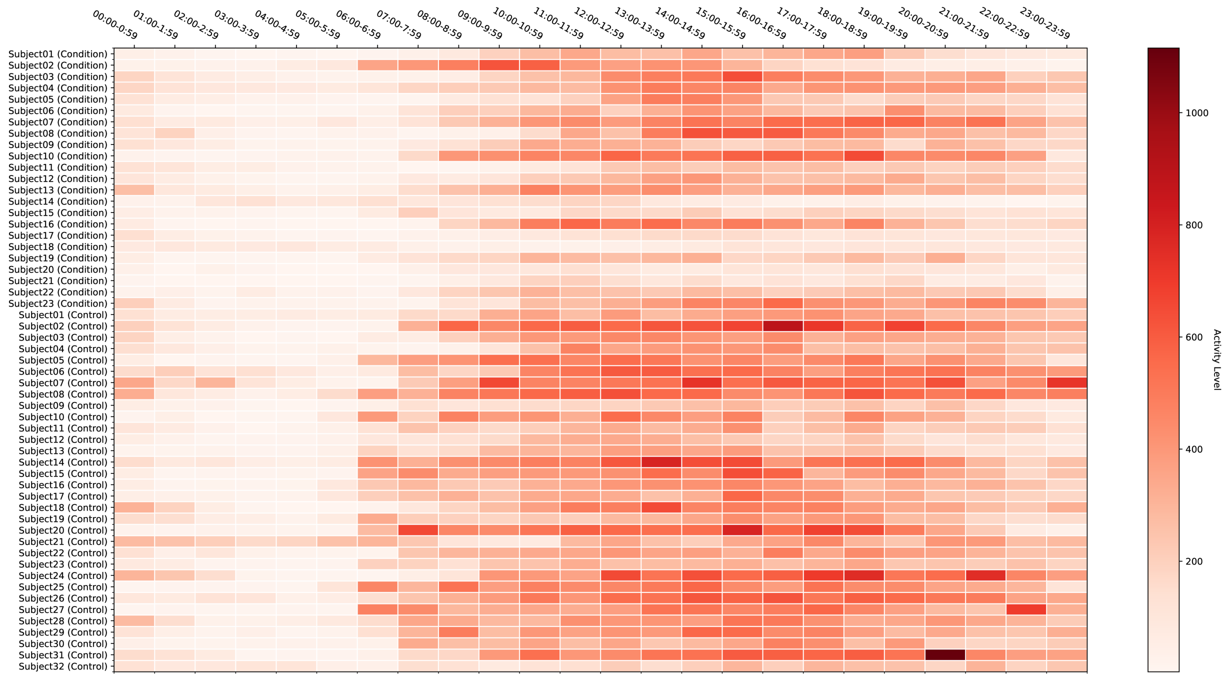
In our experiments, we measure the classification performance of several learning algorithms on data generated based on the two approaches, with leave-one-patient-out evaluation. We perform the leave-one-out evaluation for each patient, where the target patient's data is considered as the test set and the remaining data from other patients is considered as the training set. We use F1-score (weighted average), Accuracy (ACC), and Matthews Correlation Coefficient (MCC) to measure the quality of classification, which are the metrics used for performance evaluation on this dataset [13].

We perform the learning process on both the data with attributes proposed in [13] and generated data records by our approach, based on five centralized and two privacy-preserving distributed machine learning algorithms. Table I exhibits these results.

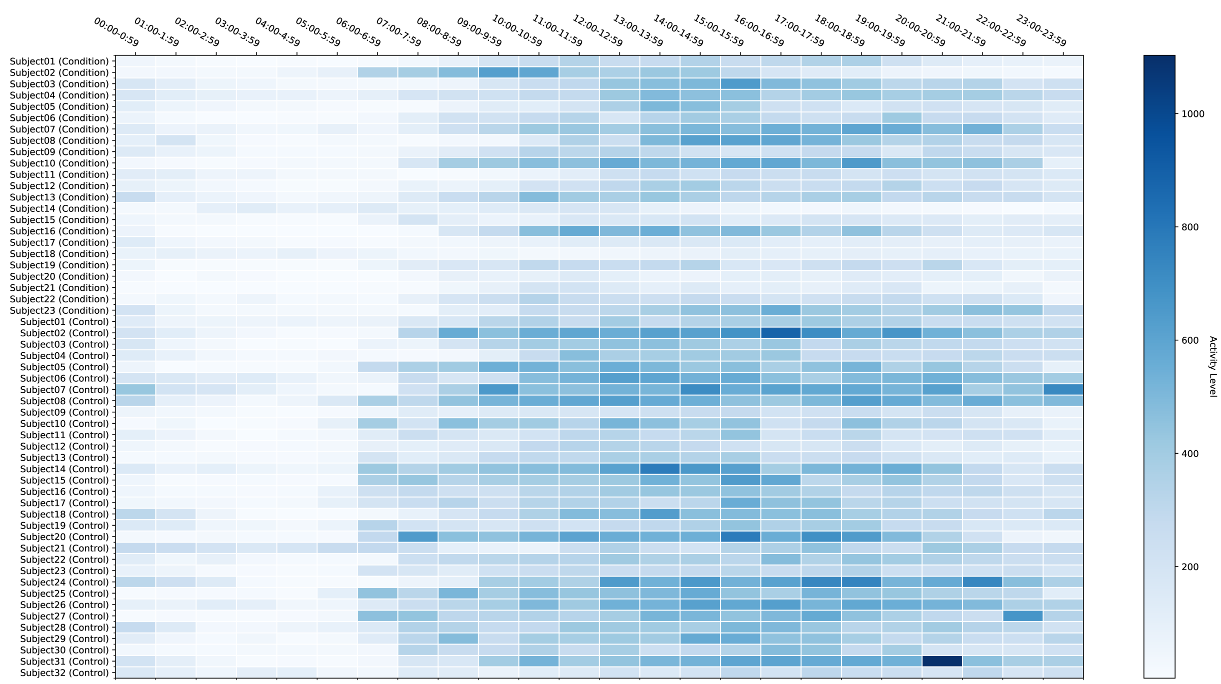
The results show a substantial improvement in the classification performance by employing our approach for generating data records from raw data. Particularly, tree-based ensemble learning approaches, i.e., PPD-ERT, ERT [38], random forest [39], and XGBoost [40], present more accurate results when trained on data generated by our augmentation approach. This is while training on the data with attributes proposed in [13] yields the best results when employing the linear SVM algorithm [41]. Comparing the best results for both approaches shows that applying our approach leads to learning more accurate classification models, i.e., models with 7.9% higher F1-score, 8.2% higher ACC, and 0.169 higher MCC. The PPD-ERT and ERT algorithms follow the same learning procedure and have the same classification performance [32].

Figures 3a and 3b show the heat-map for the raw and generated data, respectively. Each box represents the average activity level of one patient in one-hour intervals in a day. For the raw data, Figure 3a represents the average activity level of each patient based on all recorded days for him/her. Figure 3b shows the heat-map for the average activity of patients based on the generated data by adopting our approach.

The heat-maps of activity level based on the raw dataset and the generated dataset in Figures 3a and 3b are visually similar. This similarity explains the association of the gener-



(a) Heat-map for the raw data



(b) Heat-map for the augmented data

Fig. 3: Heat-map for averaged activity level in one-hour intervals for each patient

ated records and the original dataset. In order to measure the similarity between the generated data by our approach and the raw data, we calculate the relative difference among the corresponding values for each cell, averaged over the entire heat-map. The average relative difference is calculated as follow:

$$D[R, A] = \frac{1}{n \cdot m} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{|r_{ij} - a_{ij}|}{r_{ij}}, \qquad (1)$$

where n is the number of patients in each group, and m is the number of one-hour intervals in a day. $R = \{r_{11}, r_{12}, \ldots, r_{nm}\}$ is the set of average activity level of one patient in one-hour intervals in a day calculated from raw dataset. The r_{ij} is the average activity level of patient i in one-hour interval j in the raw data. Moreover, $A = \{a_{11}, a_{12}, \ldots, a_{nm}\}$ is the set of average activity level from augmented dataset. The average activity level of patient i in one-hour interval j in the augmented data is captured by a_{ij}. The value of D for the condition group is 3.3%. The value of D for the control group is equal to 3.5%.

In summary, the evaluation results in this section indicate the preservation of the activity-level information in the augmented data for the detection of depression from motor activity data. Our experimental results show that modern techniques, e.g., tree-based ensemble learning algorithms, learn more accurate classifier models given such extensive information compared to learning from the few basic statistical attributes in previous studies.

IV. Conclusion

In this paper, we propose an approach based on data augmentation to analyze the Depresjon dataset and improve the performance of detecting depression in subjects. We introduced an approach for augmenting data records from the Depresjon dataset, which leads to higher detection performance when employing modern learning algorithms. Employing our approach leads to learning more accurate models with up to 7.9% higher F1-score, 8.2% higher ACC, and 0.169 higher MCC. Moreover, the possibility of employing privacy-preserving data analysis for such data is investigated. We demonstrate the possibility of using our privacy-preserving distributed data analysis technique, PPD-ERT, for wearable devices/sensors to ensure the preservation of the privacy of sensitive information for the patients in the context of depression and mental health disorders.

Acknowledgment

This research is supported by INTROducing Mental health through Adaptive Technology (INTROMAT) project. The paper is partially supported by SIRIUS: Centre for Scalable Data Access.

References

[1] "Mental health: data and resources," https://www.euro.who.int/en/health-topics/noncommunicable-diseases/mental-health/data-and-resources, accessed: 2021-02-17.

[2] W. H. Organization *et al.*, "Mental and neurological disorders," in *Mental and neurological disorders*, 2001.

[3] M. Olfson, B. G. Druss, and S. C. Marcus, "Trends in mental health care among children and adolescents," *New England Journal of Medicine*, 2015.

[4] G. V. Polanczyk, G. A. Salum, L. S. Sugaya, A. Caye, and L. A. Rohde, "Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents," *Journal of child psychology and psychiatry*, 2015.

[5] J. M. Twenge, "Time period and birth cohort differences in depressive symptoms in the us, 1982–2013," *Social Indicators Research*, 2015.

[6] M. A. Vammen, S. Mikkelsen, Å. M. Hansen, J. P. Bonde, M. B. Grynderup, H. Kolstad, L. Kærlev, O. Mors, R. Rugulies, and J. F. Thomsen, "Emotional demands at work and the risk of clinical depression: a longitudinal study in the danish public sector," *Journal of occupational and environmental medicine*, 2016.

[7] D. Sopic, A. Aminifar, A. Aminifar, and D. Atienza, "Real-time classification technique for early detection and prevention of myocardial infarction on wearable devices," in *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2017.

[8] ——, "Real-time event-driven classification technique for early detection and prevention of myocardial infarction on wearable systems," *IEEE transactions on biomedical circuits and systems*, 2018.

[9] D. Sopic, A. Aminifar, and D. Atienza, "e-glass: A wearable system for real-time detection of epileptic seizures," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018, pp. 1–5.

[10] F. Forooghifar, A. Aminifar, L. Cammoun, I. Wisniewski, C. Ciumas, P. Ryvlin, and D. Atienza, "A self-aware epilepsy monitoring system for real-time epileptic seizure detection," *Mobile Networks and Applications*, pp. 1–14, 2019.

[11] U. Varshney, "Pervasive healthcare and wireless health monitoring," *Mobile Networks and Applications*, 2007.

[12] F. J. Penedo and J. R. Dahn, "Exercise and well-being: a review of mental and physical health benefits associated with physical activity," *Current opinion in psychiatry*, 2005.

[13] E. Garcia-Ceja, M. Riegler, P. Jakobsen, J. Tørresen, T. Nordgreen, K. J. Oedegaard, and O. B. Fasmer, "Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients," in *Proceedings of the 9th ACM multimedia systems conference*, 2018.

[14] F. Forooghifar, A. Aminifar, and D. Atienza, "Resource-aware distributed epilepsy monitoring using self-awareness from edge to cloud," *IEEE transactions on biomedical circuits and systems*, 2019.

[15] S. Baghersalimi, T. Teijeiro, D. Atienza, and A. Aminifar, "Personalized real-time federated learning for epileptic seizure detection," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[16] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Mobile sensor data anonymization," in *Proceedings of the international conference on internet of things design and implementation*, 2019.

[17] D. Pascual, A. Aminifar, D. Atienza, P. Ryvlin, and R. Wattenhofer, "Synthetic epileptic brain activities using generative adversarial networks," *Machine Learning for Health (ML4H) at Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

[18] D. Pascual, A. Amirshahi, A. Aminifar, D. Atienza, P. Ryvlin, and R. Wattenhofer, "Epilepsygan: Synthetic epileptic brain activities with privacy preservation," *IEEE Transactions on Biomedical Engineering*, 2020.

[19] S. D. Lustgarten, Y. L. Garrison, M. T. Sinnard, and A. W. Flynn, "Digital privacy in mental healthcare: current issues and recommendations for technology use," *Current Opinion in Psychology*, 2020.

[20] A. Aminifar, P. Eles, and Z. Peng, "Optimization of message encryption for real-time applications in embedded systems," *IEEE Transactions on Computers*, 2017.

[21] A. Aminifar, "Minimal adversarial perturbations in mobile health applications: The epileptic brain activity case study," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1205–1209.

[22] ——, "Universal adversarial perturbations in epileptic seizure detection," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–6.

[23] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2002.

[24] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2007.

[25] A. Aminifar, Y. Lamo, K. I. Pun, and F. Rabbi, "A practical method-

ology for anonymization of structured health data," in *Proceedings of the 17th Scandinavian Conference on Health Informatics*, 2019.

[26] M. Kantarcioglu, "A survey of privacy-preserving methods across horizontally partitioned data," in *Privacy-preserving data mining*. Springer, 2008.

[27] J. Vaidya, "A survey of privacy-preserving methods across vertically partitioned data," in *Privacy-preserving data mining*. Springer, 2008.

[28] J. Konečnỳ, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *arXiv preprint arXiv:1610.02527*, 2016.

[29] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, *et al.*, "Communication-efficient learning of deep networks from decentralized data," *arXiv preprint arXiv:1602.05629*, 2016.

[30] M. Malekzadeh, B. Hasircioglu, N. Mital, K. Katarya, M. E. Ozfatura, and D. Gündüz, "Dopamine: Differentially private federated learning on medical data," *arXiv e-prints*, pp. arXiv–2101, 2021.

[31] E. M. Marco, E. Velarde, R. Llorente, and G. Laviola, "Disrupted circadian rhythm as a common player in developmental models of neuropsychiatric disorders," *Neurotoxin Modeling of Brain Disorders—Life-long Outcomes in Behavioral Teratology*, 2015.

[32] A. Aminifar, F. Rabbi, K. I. Pun, and Y. Lamo, "Privacy preserving distributed extremely randomized trees," in *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, 2021.

[33] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, 2012.

[35] C. Burton, B. McKinstry, A. S. Tătar, A. Serrano-Blanco, C. Pagliari, and M. Wolters, "Activity monitoring in patients with depression: a systematic review," *Journal of affective disorders*, 2013.

[36] J. Scott, G. Murray, C. Henry, G. Morken, E. Scott, J. Angst, K. R. Merikangas, and I. B. Hickie, "Activation in bipolar disorders: a systematic review," *JAMA psychiatry*, 2017.

[37] A. Aminifar, F. Rabbi, and Y. Lamo, "Scalable privacy-preserving distributed extremely randomized trees for structured data with multiple colluding parties," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.

[38] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, 2006.

[39] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*. IEEE, 1995.

[40] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016.

[41] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, 1995.