Bennett CC, et al. (2024) Digital Health Sensor Data in Autism: Developing Few Shot Learning Approaches for Traditional Machine Learning Classifiers. *IEEE-EMBS International Conference on Body Sensor Networks (BSN),* In Press.

Digital Health Sensor Data in Autism: Developing Few Shot Learning Approaches for Traditional Machine Learning Classifiers

Casey C. Bennett
School of Computing
DePaul University
Chicago, USA
cbenne33@DePaul.edu

Cedomir Stanojevic
Dept. of Recreation and Tourism
Clemson University
Clemson, USA
cstanoj@clemson.edu

Jiyeong Oh
Dept. of Data Science
Hanyang University
Seoul, Korea
ojyhi010309@hanyang.ac.kr

Junyeong Ahn
Dept. of Data Science
Hanyang University
Seoul, Korea
shewas1230@naver.com

Yeeun Jeon
Dept. of Data Science
Hanyang University
Seoul, Korea
esthyj@naver.com

Kanghee Son
Dept. of Data Science
Hanyang University
Seoul, Korea
skhh0321@naver.com

Nikki Abbott
Dept. of Recreation and Tourism
Clemson University
Clemson, USA
nmabbot@clemson.edu

***Abstract*—There is great interest in applying artificial intelligence (AI) techniques to healthcare issues such as Autism, particularly in combination with digital health technologies (robots, wearables, smartphones, etc.) in user homes. However, a critical challenge is that modern AI techniques like deep learning (DL) typically require large datasets with millions of samples, yet in healthcare we are often working with smaller clinical samples (<50 participants). To address that challenge, we need to develop new approaches that can learn *more efficiently from less data*. In this paper, we propose a novel approach to few-shot learning (FSL) called SMOTE_FSL, which is applicable to traditional machine learning (ML) models, allowing them to work with smaller sample sizes as well as various types of healthcare data (not only image or text data). We compare SMOTE_FSL on two healthcare sensor datasets gathered using robots and wearables, with results showing SMOTE_FSL performs comparably to state-of-the-art DL-based FSL methods (e.g. autoencoders, generative adversarial networks [GAN]). That indicates such an approach holds potential to expand utilization of FSL to a broad range of healthcare data derived from smaller clinical sample sizes.**

***Keywords*—Autism, Sensor Data, Few Shot Learning, Machine Learning, SMOTE, Deep Learning**

I. Introduction

One of the great challenges right now in applying artificial intelligence (AI) techniques to healthcare is the need for massive datasets for many popular forms of AI, such as deep learning (DL) models. However, in healthcare we often have to deal with clinical populations for a *specific* health problem, which are of limited size in any given region, especially after considering the variability of how the same disease can manifest in different individuals. Autism and related spectrum disorders (ASD) is a perfect example of this, given the heterogeneity in spectrum disorders, as well as the challenges of not only recruiting participants but also having them adhere to research protocols during a study [1,2]. For those reasons, many ASD studies are often done with smaller sample sizes (<50 participants), outside of broad survey-based research or meta-analyses [3,4]. From our own experience, we know that any study of neurological disorders (autism, dementia, bipolar, etc.) that enrolls even a couple hundred individuals would be considered "massive" … but in the big data world of modern DL where millions of samples are the norm, such clinical datasets would be considered tiny and of limited utility.

A different approach is to try to learn more efficiently from less data, which is where few-shot learning (FSL) comes in. FSL attempts to extrapolate generalized patterns from a small number of samples. FSL modeling can be loosely broken into 4 types: data augmentation, multimodal learning, transfer learning, and meta-learning [5,6]. **However, one problem with FSL is that it is primarily designed to work with perceptual data (e.g. images, text data) via DL models.** That problem limits the applicability of FSL in healthcare, because there are many types of non-image/non-text data we may want to work with, such as physiological measurements, vital signs, clinical observations (e.g. symptoms), prescribed medication data, patient demographics, lab tests, family history, genetic data, etc., not to mention emerging digital health data sources from *beyond the clinic* in user homes and workspaces [7,8]. Moreover, overcoming that limitation may allow us to use FSL with traditional machine learning (ML) classifier methods (e.g. random forests, gradient boosting, support vector machines, etc.), which can enable "feature selection" in order to identify important factors related to a particular health problem.

To that end, we have been exploring use of FSL with traditional ML classifiers in ASD, using sensor data generated by multiple digital health devices (wearables, robots, smartphones) in participant's homes. Given the lack of existing FSL methods for ML, **that entailed developing a new method for FSL with traditional ML classifiers, which we call SMOTE_FSL**. We describe results of that work here. Additionally, to further validate that our approach holds potential beyond autism specifically (e.g. for other health conditions), we replicated the same analysis on a similar set of sensor data gathered from a robotic pet system used with neurotypical individuals from the same age group (college students).

II. METHODS

A. Dataset and Data Collection Procedures

There were two datasets we collected for this study: 1) college-aged autistic individuals transitioning to adulthood, and 2) similar college-aged neurotypical individuals without autism. The first dataset comprised 5 individuals, and the second comprised 12 individuals. The goal was to study these smaller number of users intensively over a longer period of time, rather than study many users briefly. Each individual was given a robotic pet (Joy-for-All) to use in their own home for 2 weeks. The robot was fitted with a custom sensor collar, that could detect activity data in the vicinity of the robot, including light, sound, motion, and indoor environmental conditions (see Figure 1) [9, 10]. Sensor collar data was collected roughly 9 times per second, every minute of every day, across the in-home deployment period. Participants were then "pinged" 5-7 times per day via smartphone app about their activities near the robot, following an ecological momentary assessment (EMA) approach [11]. Those activity modalities included Petting the robot, Talking, Eating/Cooking, Listening to TV/Media (e.g. YouTube), Playing with the Robot, and Moving locations (e.g. from one room to another). Participants could select more than one activity each time when pinged. That process produced roughly 11.7 million data points per participant, with the EMA data providing "ground truth" labels for later ML/DL modeling. In other words, the EMA pings became our "**targets**" in the dataset, with the sensor collar data as "**features**" for predicting the target. For the Autism portion of the study, participants also wore Empatica EmbracePlus wristbands during the 2-week deployment phase (https://www.empatica.com/embraceplus/), so that data was additionally available for each participant as features beyond the collar data. That provided additional digital biomarker data on heart rate, electrodermal activity, sleep patterns, and body temperature. This study, including the study design and sample sizes, were approved by the IRBs of Hanyang University, DePaul University, and Clemson University.

Fig. 1. Robot Sensor collar

We note that beyond autism, the robot system was originally designed to track activities in older adults with chronic illness, e.g. Alzheimer's and related dementias (ADRD), depression and loneliness, and so forth. However, here the data was gathered from younger individuals (college-aged 18-30 years old) with and without autism.

B. Analysis Methods

We employed two primary forms of analysis in this study: 1) our novel SMOTE_FSL algorithm for ML classifiers, and 2) common DL methods for FSL including autoencoders (AE), variational autoencoders (VAE), and generative adversarial networks (GAN). Those DL methods are widely used for tasks in computer vision and natural language processing. The goal here is to compare the different methods, including FSL for ML and FSL for DL, on the same dataset.

1) SMOTE_FSL for ML: To enable FSL to be used with ML methods, we developed a new data augmentation algorithm, based on custom variations of the existing SMOTE algorithm used to correct class imbalance in the target class in datasets [12]. Class imbalance refers to when there are many examples of the majority case (e.g. *NOT* having a disease) but few examples of the minority case (e.g. having a disease). However, one of the problems with the original SMOTE algorithm is that it only samples *within* the existing data distribution of the minority case , i.e. it assumes that existing distribution represent the total "universe" of possible samples and simply "fills in" the data with new synthetic samples from that same distribution (see [13] for examples). New synthetic samples are not created outside that range, e.g. we assume that the individuals with ASD in our dataset represent all individuals with ASD in existence. That is obviously a problematic assumption, which is likely untrue when working with smaller clinical datasets. That problem led to the development of several variations of SMOTE, such as SMOTE_OUT [13], that attempt to correct that problem in various ways. Several other variations can be found here: https://pypi.org/project/smote-variants/

The problem for FSL is slightly more complicated than the existing variations though, in that in order to perform data augmentation, we need to generate synthetic data for *both* the minority and majority cases (with representatively-sized distributions), while also accounting for any possible skewness due the small sample size in FSL problems. As such, we developed our novel SMOTE_FSL algorithm, after creating and experimenting with dozens of potential approaches. The algorithm can be described in two steps, as pseudocode shown in Figure 2:

```
[1] <setup synthetic sample distribution>
For c in (target classes):
    synth_n_c = synth_n * (1 – (n_c / n_total))

[2] <create synth samples for each target class>
For synth in range(synth_n_c)
    idx = random(sample)
    k_array[] = select k nearest neighbors of idx

    <assign feature values for synth sample>
    For j in range(feature_count):
        Calculate average_j, stdDev_j, median_j of all synth_n_c data
        skew_j = (average_j  - median_j)/ stdDev_j
        gap = random.uniform(-1.0,1.0)
        diff_j = (Σ_{i=0}^{k} abs(idx_j – k_array[i]_j)) / k
        skew_shift_j = average_j * ((skew_j * skew_x) - skew_j) * abs_value(gap)
        synth_j =  value_j+ (((diff_j + (stdDev_j * pad)) * gap) + skew_shift_j
```

Fig. 2. SMOTE_FSL Psuedocode

Where:

- **synth_n** = number of synthetic samples to create
- **k** = number of nearest neighbor samples to utilize
- **diff**$_j$ = average absolute difference (of feature *j*) between index sample (idx) and its "k" nearest neighbors
- **gap** = random value between -1 and 1 that is reset every iteration, so that synthetic samples vary in nearness to existing index sample
- **pad** = padding, i.e. percentage of std dev to add to each feature, so that new synthetic samples can sometimes exceed the distribution range of the original samples
- **skew_x** = skew factor, where >1 increases skew, <1 decreases skew, =1 no effect

The SMOTE_FSL process can be summarized as follows. First, it figures out how many samples to synthesize for the majority/minority classes in order to augment the dataset sample size while also partially rebalancing it, if applicable. It then creates each synthetic sample iteratively, using an index sample (*idx*) and some *k* nearest neighbors to calculate the average "difference" for each feature in their local neighborhood. That difference is combined with several statistical measures (e.g. average, std dev) to create feature values of the new synthetic sample, where the measures are "inverted" so that the new synthetic sample can sometimes (but not always) exceed the range of the feature values in the original dataset. The inversion process and its frequency are controlled by parameters such as pad, k, and skew_x. The parameters can be varied automatically, allowing for hyperparameter tuning within SMOTE_FSL to search for optimal parameters on a given dataset.

2) Common DL Methods for FSL: To compare with SMOTE_FSL, we employed several common methods for DL-based FSL, including autoencoders (AE), variational autoencoders (VAE), and generative adversarial networks (GANs). Those all have been described at length elsewhere in the literature [14-16], but we provide a brief description here. All three can be broadly seen as types of generative models.

Autoencoders are essentially neural network models that have similarly sized input and output layers along with smaller-sized hidden layers between, that attempt to learn encodings of the original input data by "compressing" it. Autoencoder models have an encoder component and decoder component, where the decoder can be used to generate new data from the encodings. VAEs are an extension of autoencoders that inserts a probabilistic latent space into the hidden layers of the network, adding "noise" to the encodings to try to avoid overfitting to the original input dataset (a common problem with autoencoders). That noise serves a similar purpose to our "statistical inversion" in SMOTE_FSL (see Section II.B.1 above), in creating new synthetic samples beyond the original distribution. GANs are another type of generative model, which has some similarities to autoencoders but a fundamentally different approach to learning using an adversarial feedback loop. In place of an encoder and decoder, we instead have a generator and discriminator, where the generator attempts to create new samples with some "noise" added in, passes those synthetic samples along with unchanged original samples to the discriminator, and the discriminator must determine which samples are real or fake. The idea is that the generator will gradually learn how to generate more plausible synthetic samples over time through direct competition with the discriminator (its "adversary").

C. ML/DL Model Validation

Models in this study were constructed using Python Scikit (for ML) or TensorFlow via the Python Keras package (for DL). For simplicity here, we utilized the same ML model (Random Forests) on all datasets to test the effectiveness of our SMOTE_FSL method. To evaluate model performance, we utilized standard techniques in the machine learning field, following established guidelines [17]. For ML, that entailed 5-fold cross-validation based on multiple performance metrics, such as accuracy and AUC (area-under-curve). For DL, that entailed a standard 20% hold-out validation set for testing models trained on the other 80% of data.

III. Results

The main results of modeling with the autism dataset can be seen in table 1. For brevity, we show the average performance across all activity modalities. As can be seen in the table, the performance of our SMOTE_FSL for ML method was comparable to the performance of the VAE and GAN methods using Deep Learning. That indicates that SMOTE_FSL holds potential to extend few shot learning approaches to traditional ML classifiers (e.g. random forests, gradient boosting, support vector machines, etc.) beyond DL.

TABLE I. Model Result Comparison – Autism

Method	Accuracy	AUC
ML (SMOTE_FSL)	80.0	0.8880
AE	75.0	0.8001
VAE	79.5	0.8514
GAN	81.6	0.8277

TABLE II. Model Result Comparison – Neurotypical

Method	Accuracy	AUC
ML (SMOTE_FSL)	79.7	0.8665
AE	77.6	0.8315
VAE	80.0	0.8429
GAN	80.1	0.8392
Original Results	*83.6*	*0.8847*

We do note there were lower scores for a couple activity modalities than others, in particular Petting and Listening to TV/Media, which we have reported previously [9]. That was likely due to the available sensors onboard the robot, which makes some modalities easier to detect than others. That is a sensor hardware issue though, rather than a modeling issue.

To further validate SMOTE_FSL beyond autism, we tested it on a neurotypical population using the same in-home robot sensor system. The goal was two-fold: 1) to see if SMOTE_FSL would still work on an entirely different population, and 2) to mimic what would happen after FSL if we later obtained data from additional patients for testing, e.g. if we created a FSL model on 5 patients and then later applied it to 10 new patients. As such, we divided our neurotypical data up so 50% of it was used for the usual training/testing of models and 50% was held out as an "unseen" final validation set. That unseen data is intended to simulate some group of later new patients. Results are shown in Table 2.

One interesting aspect of the neurotypical data is that it has been subject to intense study previously, with ML/DL modeling results (*without* FSL) being reported in multiple prior publications [9,10,18]. As such, we have ground truth "original results" to compare to the FSL methods, which are shown at the bottom of Table 2. In comparison to the original results, we can see again that the FSL models performed fairly well, with slightly lower performance scores. More critically, we note that the SMOTE_FSL model was again able to closely match the accuracy and AUC of the VAE and GAN models, further validating their potential to apply to a broad range of healthcare data and opening the door to FSL methods being applied to a broader range of non-image and/or non-text data.

IV. Discussion

A major problem with AI in healthcare is that some techniques that we may want to use are currently only applicable to certain kinds of data. For instance, FSL methods that are useful for modeling smaller-sized clinical datasets are limited to perceptual data (e.g. images, text data) via DL models [5,6]. However, there are many types of non-image/non-text data we may want to work with, such as physiological measurements, vital signs, clinical observations (e.g. symptoms), prescribed medication data, patient demographics, lab tests, family history, genetic data, etc., not to mention emerging digital health data sources from *beyond the clinic* in user homes and workspaces [7,8].

In this paper, we addressed that limitation by developing the SMOTE_FSL method specifically to apply FSL to more traditional ML methods (e.g. random forests, gradient boosting, support vector machines, etc.), making it applicable to any sort of healthcare data including those listed above. We tested the method on both a population of individuals with autism as well one with neurotypical individuals, to assess its performance in different healthcare scenarios. The data collection included sensor data generated by multiple digital health devices (wearables, robots, smartphones) in participant's homes. **We found that SMOTE_FSL worked comparatively well in both populations as several state-of-the-art DL methods for FSL, including VAE and GAN models.** That indicates such an approach holds potential to expand utilization of FSL a broad range of healthcare data, while also opening up opportunities to use ML-based feature selection methods in the future to identify important factors related to a particular health problem.

In the *digital health* space, this also can expand the types of modeling we do with sensor data collected from various devices (robots, smartphones, wearables). That is significant, in that we are often dealing with smaller clinical samples sizes in healthcare [2], but moreover collecting real-world sensor data from many participants is often costly and fraught with logistical challenges (not to mention hardware failures). As such, it is critical we develop methods like SMOTE_FSL that can enhance our ability to ***learn more efficiently from less data***. Unfortunately many modeling methods do the opposite, which limits their applicability in healthcare.

Acknowledgment

Funding for this work was provided by Clemson University's R-Initiative Program. The authors would also like to thank our various collaborators who contributed to this work.

References

[1] M.V. Lombardo, M.C. Lai, and S. Baron-Cohen, "Big data approaches to decomposing heterogeneity across the autism spectrum." *Molecular Psychiatry*, 24(10), 1435-1450, 2019.

[2] R.A. Kadam, S.U. Borde, S.A. Madas, S.S. Salvi, and S.S. Limaye, "Challenges in recruitment and retention of clinical trial subjects." *Perspectives in Clinical Research*, 7(3), 137, 2016.

[3] M.V. Lombardo, "Prototyping as subtyping strategy for studying heterogeneity in autism." Autism Research, 14(10), 2224-2227, 2021.

[4] S. Healy, A. Nacario, R.E. Braithwaite, and C. Hopper, "The effect of physical activity interventions on youth with autism spectrum disorder: A meta‐analysis." *Autism Research*, 11(6), 818-833, 2018.

[5] Y. Song, T. Wang, P. Cai, S. Mondal, and J.P. Sahoo, "A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities." *ACM Computing Surveys*, 55(13), 1-40, 2023.

[6] Y. Wang, Q. Yao, J.T. Kwok, and L.M. Ni, "Generalizing from a few examples: A survey on few-shot learning." *ACM Computing Surveys*, 53(3), 1-34, 2020.

[7] B. Shickel, P.J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis." *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589-1604, 2017.

[8] Y. Si, J. Du, Z. Li, X. Jiang, T. Miller, F. Wang, W.J. Zheng, and K. Roberts, "Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review." *Journal of Biomedical Informatics*, 115, 103671, 2021.

[9] C.C. Bennett, S. Sabanovic, C. Stanojevic, Z. Henkel, S. Kim, J. Lee, et al., "Enabling robotic pets to autonomously adapt their own behaviors to enhance therapeutic effects: a data-driven approach." *32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1625-1632, 2023.

[10] C.C. Bennett, C. Stanojevic, S. Kim, S. Sabanovic, J. Lee, J.A. Piatt, et al., "Comparison of in-home robotic companion pet use in South Korea and the United States: A case study." *9th IEEE International Conference on Biomedical Robotics & Biomechatronics (BIOROB)*, pp. 1-7, 2022.

[11] C.C. Bennett, C. Stanojevic, S. Sabanovic, J.A. Piatt, and S. Kim, "When no one is watching: Ecological momentary assessment to understand situated social robot use in healthcare." *ACM International Conference on Human-Agent Interaction (HAI)*, pp. 245-251, 2021.

[12] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique." *Journal of Artificial Intelligence Research*, vol.16, pp.321–357, 2002.

[13] F. Koto, "SMOTE-Out, SMOTE-Cosine, and Selected-SMOTE: An enhancement strategy to handle imbalance in data level." *IEEE International Conference on Advanced Computer Science and Information System*, pp. 280-284, 2014.

[14] J. Zhai, S. Zhang, J. Chen, and Q. He, "Autoencoder and its various variants." *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 415-419, 2018.

[15] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability." *Special Lecture on IE*, 2(1), 1-18, 2015.

[16] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A.A. Bharath, "Generative adversarial networks: An overview." *IEEE Signal Processing Magazine*, 35(1), 53-65, 2018.

[17] J. Siebert, L. Joeckel, J. Heidrich, K. Nakamichi, K. Ohashi, I. Namba, et al., "Towards guidelines for assessing qualities of machine learning systems." *International Conference on the Quality of Information and Communications Technology (QUATIC)*. 17-31, 2020.

[18] S. Kim, C.C. Bennett, Z. Henkel, J. Lee, C. Stanojevic, K. Baugus, et al., "Generative replay for multi-class modeling of human activities via sensor data from in-home robotic companion pets." *Intelligent Service Robotics*, 17(2), 277-287, 2024.