ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096235

TOWARD PRIVACY-ENHANCING AMBULATORY-BASED WELL-BEING MONITORING: INVESTIGATING USER RE-IDENTIFICATION RISK IN MULTIMODAL DATA

Ravi Pranjal[*], *Ranjana Seshadri*[*], *Rakesh Kumar Sanath Kumar Kadaba*[*],
Tiantian Feng[†], Shrikanth S. Narayanan[†], and Theodora Chaspari[*]

[*]Texas A&M University
[†]University of Southern California
{ravipranjalp, ranjana.seshadri, rakeshkumar5165}@tamu.edu,
tiantiaf@usc.edu, shri@sipi.usc.edu, chaspari@tamu.edu

ABSTRACT

The sensitivity of data collected via ambulatory monitoring, which regularly involve the recording of speech signals and sensor information, can cause strong privacy concerns. We investigate user re-identification risk in a corpus of such data collected to observe the interplay between behavior, physiology, and well-being of healthcare workers in their daily life. We then develop a user anonymization approach that preserves well-being information (i.e., anxiety), but eliminates user identify (ID) information. We formulate this via an auto-encoder that learns a transformed version of the original feature set in an adversarial manner so that it minimizes the anxiety estimation loss and maximizes the user classification loss. Results indicate that the original features bear a large user re-identification risk, while also having a good ability to classify a user's anxiety. After removing the most prone features to user re-identification from the original feature set, the user classification accuracy decreases, while the anxiety classification performance is preserved. The final features transformed via the auto-encoder further reduce evidence of user ID and preserve anxiety classification ability. Findings from this study can contribute to the design privacy-aware bio-behavioral models that can be used for responsible ambulatory monitoring in healthcare and beyond.

1. INTRODUCTION

Advances in Internet of Things (IoT) and wearable devices are enabling the shift of in-clinic care to mobile and pervasive healthcare that can improve the access and efficiency of healthcare services and transform personalized prevention and delivery [1]. Yet, these technologies have been met with public skepticism, since they rely on individuals sharing their digital traces that can explicitly or implicitly contain personally identifiable information (PII). Consumers are overall skeptical about sharing their (sensitive) data with entities other than their doctor and even if they accept doing so in order to leverage the benefits of digital healthcare, the inherent sensitivity of the data complicates the sharing process [2]. Other studies found that while participants are overall supportive of sharing their data for biomedical research purposes, their perceived risk depends on the type of data that is under consideration [3]. For example, participants are less willing to share personal and location data. Moreover, regulations such as the Health Insurance Portability and Accountability Act (HIPPA), the California Consumer Protection Act (CCPA), and the European General Data Protection Regulation (GDPR), restrict the collection and use of sensitive user data. Thus, it is becoming more evident that machine learning (ML) systems that enable digital healthcare services should be inspected for potential PII leaks and should be subsequently designed so that they protect PII.

Conceptually PII can be leaked throughout the cycle of data collection, storage, and usage. Increasing evidence supports the existence of data privacy breaches and PII leakage in the way the data collection occurs, including the type of data, the context and purpose for which the data is collected, and the entity that conducts the data collection [4, 5, 6]. However, PII leakage can also occur while the data is utilized, either explicitly by the individuals who have access to the stored data, or implicitly by the ML models that can learn patterns based on the collected data [7]. Inference attacks are such an example, including model inversion attacks [8, 9, 10] and membership inference attacks [11, 12, 13, 14], in which PII might not be explicitly encoded in the data, but can inferred by the models that represent this data. For example, the panel of laboratory test results can be indicative of one's identity

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2017-17042800005 (Narayanan, PI), and the National Science Foundation (NSF) IIS-2046118 (Chaspari, PI). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. R. Seshadri and R.K.S.K. Kadaba contributed equaly to this work.

(ID) [15]. These can have serious societal implications and negatively impact the users of digital healthcare technologies, such as potentially causing discrimination in employment and insurance policy [16].

This paper investigates user re-identification risk for the task of anxiety classification using a set of multimodal data that was continuously collected from healthcare workers throughout the span of four weeks. First, we explore the types of features that are the most and least prone in leaking user ID information. Following that, we investigate anxiety classification performance using the original set of multimodal features, as well as the features that were found to be the least prone in leaking the user ID. Anxiety is chosen because of its correlation to well-being for healthcare workers [17]. In addition, we transform the features to remove user-dependent information while retaining anxiety-related information. For this purpose, we leverage an auto-encoder that learns an identity mapping between the original and transformed features, while imposing two additional constraints of: (1) minimizing user classification accuracy; and (2) maximizing anxiety classification accuracy. The weights of the autoencoder are learned in an adversarial manner, in which we sequentially freeze parts of the network and learn the weights for the other parts, which allows to achieve a stable solution. Results indicate that by removing from the feature set the features that depict the largest user re-identification risk, we are able to reduce the user classification accuracy and preserve the anxiety classification accuracy. The proposed feature transformation via the adversarial learning further benefits to the two aforementioned goals.

2. PRIOR WORK

The focus of the biomedical data analytics community has recently shifted to designing techniques that can reduce the risk of user ID leakage from health data while preserving utility information that might be important for health applications. El Emam *et al.* found that a perturbation strategy via noise addition makes it highly unlikely that one's record can be matched to a group of less than 10 individuals and at the same time, preserves clinical information [18]. Shahin et. al. [19] proposed adding Laplacian noise between the encoding and decoding layers to render speech utterances user independent. The concept of differential privacy, first introduced by Dwork & Pottenger [20], adds controlled noise to the data to render user re-identification hard. This approach has been also explored in biomedical health informatics for suppressing patient ID in a breast cancer dataset [21]. In the context of behavioral applications, such as emotion recognition, Feng & Narayanan explored the interplay between preserving emotion information and reducing evidence of speaker ID in acoustic measures [22]. Jaiswal & Provost found that demographic information can be preserved in emotion classification models that have been trained using acoustic and linguistic features, and investigated an adversarial learning approach to "unlearn" the sensitive information [23]. Narula *et al.* also empirically showcased the existence of user-specific information in the convolutional kernels of a convolutional neural network (CNN) that has been trained solely for the purpose of emotion classification, and leveraged an adversarial learning method for suppressing this information [24].

The contributions of this work in association to the existing studies are: (1) We examine the risk of user ID leakage on a set of multimodal data that were collected in real-life and with wearable sensors, a field that is relatively less explored compared to electronic health records (EHR) or clinic-based biomedical data (e.g., radiology images); (2) We examine the extent to which we can enhance user privacy in anxiety classification via removing the multimodal features with the highest user separability; and (3) We investigate an adversarial learning approach to suppress user ID information from the data and preserve anxiety information, that is relevant to ambulatory well-being monitoring in real-life.

3. DATA DESCRIPTION

We use the TILES 2018 data [25] that was collected in order to understand the dynamic relationships among individual differences, work and wellness behaviors, and the contexts in which they occur. Our data includes 212 hospital workers who were further equipped with wearable sensors used to collect vocal acoustic and physiological signals. Based on these, 69 ambulatory multimodal features were extracted, that include 25 measures of physiology and daily activity measures, 29 acoustic features, and 15 physiological features, as in [25]. Missing values were replaced with the mean of the corresponding feature computed over all users. Finally, each feature was normalized to depict zero mean and unity standard deviation. Anxiety was measured via an ecological momentary assessment (EMA) collected on a daily level using a 5-point Likert scale answer to the question "Please select the response that shows how anxious you feel at the moment." Anxiety responses were binarized using the median as the threshold between the low and high anxiety class. This resulted in 2,904 samples for all users, where each sample corresponds to a day of recording. User identification was a 212-classification task that included the user ID at its output.

4. METHODOLOGY

4.1. Characterizing user re-identification risk

We empirically measure user re-identification risk by capturing the effectiveness of the considered multimodal features in accurately classifying among users. For this purpose, we use a logistic regression (LR) classifier and a random forest (RF) classifier with 100 trees whose input features are the 69 multimodal features (Section 3). We approximate user re-identification risk by computing the user classification accuracy among the 212 participants. Large accuracies would reflect high user re-identification risk, and vice-versa. We use 80% of the samples for training and 20% for testing.

4.2. Establishing a baseline for anxiety classification

Similar to Section 4.1, we train a LR and a RF classifier based on the 69 multimodal features (Section 3) for the task of anxiety classification. The anxiety classification accuracy would serve as a baseline that indicates the ability of the features to preserve information related to user anxiety. We use same experimental parameters for this task as in Section 4.1.

4.3. Privacy preserving well-being estimation

Here, we discuss the two methods that we investigated to preserve information related to well-being (i.e., anxiety), while reducing the trace of speaker ID.

First, we find the multimodal features that are the most discriminative among users, thus, bear the largest user re-identification risk. For this purpose, we conduct an analysis of variance (ANOVA) for each feature, where the classes correspond to the different users. After calculating the F-value for each feature, we select the 20 lowest-ranked features that are the least differentiable among users with a threshold of 9.67. Out of these features, 11 were acoustic measures of jitter, shimmer, loudness, spectral energy/skewness/entropy, 8 were sleep-based measures, and 1 feature was the daily step count. We will compare this reduced feature set with the original features via user ID and anxiety classification experiments.

Second, we aim to transform the feature space so that it preserves information regarding a user's anxiety, while eliminating user ID. Grounded in prior work on adversarial learning that depicts promising results for this purpose [23, 24], we design an auto-encoder architecture that takes as an input the original feature vector, $\mathbf{x} \in \mathbb{R}^D$, and yields an output feature vector, $\hat{\mathbf{x}} \in \mathbb{R}^D$, via a non-linear transformation $\hat{\mathbf{x}} = g_{\mathbf{w_d}}(f_{\mathbf{w_e}}(\mathbf{x}))$, where $\mathbf{z} = f_{\mathbf{w_e}}(\mathbf{x})$ is the bottleneck layer of the auto-encoder (Figure 1). Functions f and g serve as the encoder and decoder parts, parameterized via weights $\mathbf{w_e}$ and $\mathbf{w_d}$, respectively. The autoencoder conducts an identity mapping to minimize the difference between the input and the output via the loss function $L_{ae}(\mathbf{w_e}, \mathbf{w_d}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2$. At the same time, the bottleneck layer goes through two additional transformations; the first transformation, $\mathbf{y_{id}} = d_{\mathbf{w_{id}}}(\mathbf{z})$, outputs the user ID $\mathbf{y_{id}}$ and the second transformation, $y_a = h_{\mathbf{w_a}}(\mathbf{z})$, outputs the anxiety label y_a. The loss functions $L_{id}(\mathbf{w_{id}})$ and $L_a(\mathbf{w_a})$ correspond to the categorical cross-entropy measures of the user ID and low/high anxiety outcome. Thus, the autoencoder learns a set of weights so that the following function is minimized:

$$\begin{aligned} L(\mathbf{w_e}, \mathbf{w_d}, \mathbf{w_i d}, \mathbf{w_a}) &= \alpha L_{ae}(\mathbf{w_e}, \mathbf{w_d}) \\ &+ \gamma L_a(\mathbf{w_a}) + \beta L_{id}(\mathbf{w_{id}}) \quad (1) \end{aligned}$$

where $\alpha, \gamma \geq 0 \, and \, \beta \leq 0$

This ensures that the autoencoder is trained in an adversarial manner so that the anxiety loss is minimized and the user ID loss is maximized. The training of the autoencoder occurs via sequentially freezing different parts of the network, as depicted in Table 1, where w_e is always trainable, a process that empirically resulted in increased convergence in adversarial learning, as demonstrated in prior work [26, 24]. The output $\hat{\mathbf{x}}$ of the autoencoder serves as the anonymized feature vector that is used for speaker classification (Section 4.1) and anxiety classification (Section 4.2).

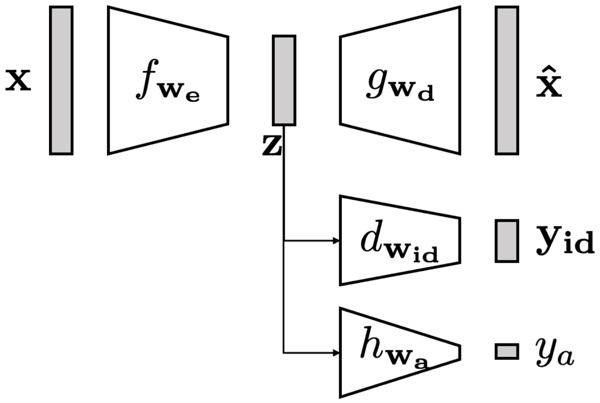


Fig. 1. Autoencoder architecture that learns an identity function of the original feature space while preserving anxiety information and suppressing user identity.

Table 1. Sequential freezing of autoencoder weights

Frozen	Trainable	α	β	γ	#Epochs	Learning Rate
$\mathbf{w_{id}}, \mathbf{w_d}$	$\mathbf{w_a}$	0	0	1	30	0.01
$\mathbf{w_a}, \mathbf{w_d}$	$\mathbf{w_{id}}$	0	-1	0	50	0.01
$\mathbf{w_{id}}, \mathbf{w_a}$	$\mathbf{w_d}$	1	0	0	50	0.0001
$\mathbf{w_{id}}, \mathbf{w_d}$	$\mathbf{w_a}$	1	0	3	50	0.00003
$\mathbf{w_{id}}$	$\mathbf{w_a}, \mathbf{w_d}$	1	0	1	1400	0.005

5. RESULTS

Results indicate that the original feature set bears considerable user re-identification risk, since it yields user classification accuracy of 81% and 93% with the LR and RF, respectively (Table 2). This is also visually demonstrated in Figure 3(a), where each user got clustered in different regions based on the original features. Using the least discriminative features of a user ID as the input to the classifiers reduces the user ID classification accuracy to 15% and 50%, for the LR and RF classifiers, respectively. At the same time, this method preserves the anxiety classification accuracy, yielding only a 4%-8% absolute decrease compared to the original features. When we include more features that are discriminative of the user ID to the feature space, as evaluated by the ANOVA, the user re-identification risk considerably increases (i.e., increased user ID accuracy), while the anxiety classification performance appears to increase to a smaller extent (Figure 2). Transforming the original feature set via the proposed auto-encoder also decreases the user re-identification risk, but not as much as the feature selection via ANOVA. However,

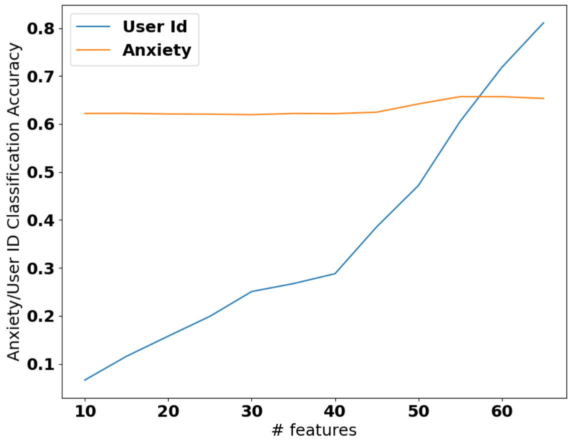
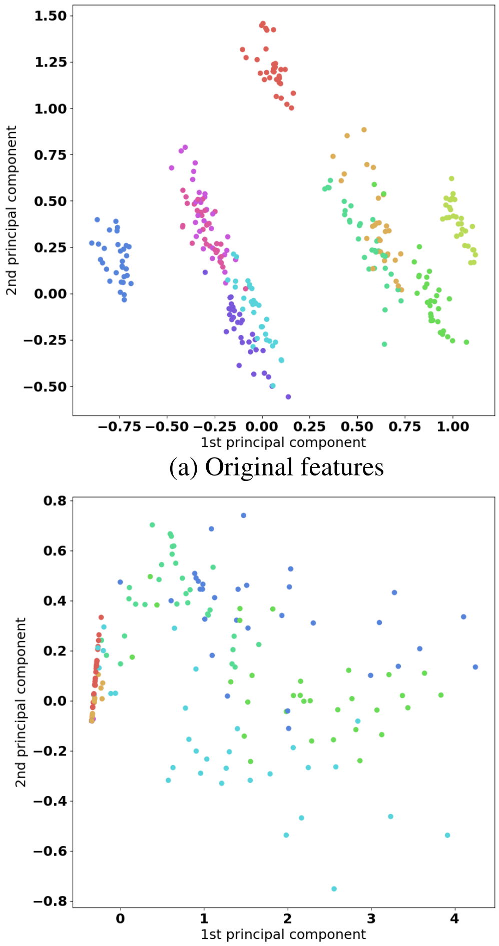


Fig. 2. User identification (ID) and anxiety classification accuracy with increasing number of features that are discriminative of a user, as selected via the analysis of variance (ANOVA).

when these two methods are combined, the user ID classification significantly drops compared to the original feature set (i.e., 10% and 35% absolute decrease for LR and RF, respectively), while the anxiety classification accuracy remains around 64-65% (Table 2). This is also visually demonstrated in Figure 3(b), that shows high overlap among users with respect to the transformed feature space.

Data Transform	Model	Feature Selection	User ID Accuracy	Anxiety Accuracy
No	LR	No	81.22%	66.45%
No	LR	ANOVA	15.71%	62.13%
No	RF	No	93.31%	72.36%
No	RF	ANOVA	50.22%	64.14%
Auto-encoder	LR	No	20.24%	64.01%
Auto-encoder	LR	ANOVA	10.55%	64.73%
Auto-encoder	RF	No	37.11%	64.70%
Auto-encoder	RF	ANOVA	35.29%	64.88%

Table 2. User identification (ID) and anxiety classification accuracy using logistic regression (LR) and random forests (RF) with different anonymization methods.



(a) Original features

(b) Transformed features via the autoencoder

Fig. 3. Scatter plot of the first two principal components of the original and transformed features for 10 example participants.

6. CONCLUSION

We investigated user re-identification risk in multimodal data that are collected from real-life work settings with wearable sensors for the purpose of well-being monitoring. Results indicate in an empirical manner the presence of significant user re-identification risk, since the considered multimodal features alone can be effectively used for building classifiers that estimate a user's ID. We explored two ways to mitigate this risk via selecting the features that are the least discriminative of the user ID and transforming the feature space so that we eliminate the sensitive information. Our findings indicate the effectiveness of both approaches in eliminating user ID information, while preserving information related to the user's well-being (i.e., anxiety). Since we have only employed a closed-set user classification task to evaluate the ability of our method to reduce evidence of information related to a user's ID, as part of our future work, we will consider an open-set user identification framework, in which the data for a target user are part of the training set. We will further explore additional well-being constructs for the same task, such as affect and stress, and examine privacy-enhancing feature transformations via federated learning techniques that assume the distribution of data across multiple agents.

7. REFERENCES

[1] Andreas K Triantafyllidis and Athanasios Tsanas, “Applications of machine learning in real-life digital health interventions: review of the literature,” *Journal of Medical Internet Research*, vol. 21, no. 4, pp. e12286, 2019.

[2] S Day, C Seninger, J Fan, K Pundi, A Perino, and M Turakhia, “Digital health consumer adoption report 2019.,” 2019.

[3] LaPrincess C Brewer, Karen L Fortuna, Clarence Jones, Robert Walker, Sharonne N Hayes, Christi A Patten, and Lisa A Cooper, “Back to the future: achieving health equity through health informatics and digital health,” *JMIR mHealth and uHealth*, vol. 8, no. 1, pp. e14512, 2020.

[4] Pierangelo Rosati, Peter Deeney, Mark Cummins, Lisa van der Werff, and Theodore Lynn, “Social media and stock price reaction to data breach announcements: Evidence from us listed companies,” *Research in International Business and Finance*, vol. 47, pp. 458–469, 01 2019.

[5] Naga Vemprala and Glenn B. Dietrich, “A social network analysis (SNA) study on data breach concerns over social media,” in *Hawaii International Conference on System Sciences (HICSS)*, 2019.

[6] Chen Wang, Xiaonan Guo, Yingying Chen, Yan Wang, and Bo Liu, “Personal pin leakage from wearable devices,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 3, pp. 646–660, 2017.

[7] Yoonyoung Park, Jianying Hu, Moninder Singh, Issa Sylla, Irene Dankwa-Mullan, Eileen Koski, and Amar K Das, “Comparison of methods to reduce bias from clinical prediction models of postpartum depression,” *JAMA network open*, vol. 4, no. 4, pp. e213909–e213909, 2021.

[8] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart, “Model inversion attacks that exploit confidence information and basic countermeasures,” in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, 2015, CCS ’15, p. 1322–1333, ACM.

[9] Zecheng He, Tianwei Zhang, and Ruby B. Lee, “Model inversion attacks against collaborative inference,” in *Proceedings of the 35th Annual Computer Security Applications Conference*, New York, NY, USA, 2019, ACSAC ’19, p. 148–162, ACM.

[10] Xi Wu, Matthew Fredrikson, Somesh Jha, and Jeffrey F. Naughton, “A methodology for formalizing model-inversion attacks,” in *2016 IEEE 29th Computer Security Foundations Symposium (CSF)*, 2016, pp. 355–370.

[11] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov, “Membership inference attacks against machine learning models,” in *2017 IEEE Symposium on Security and Privacy (SP)*, 2017, pp. 3–18.

[12] Jinyuan Jia, Ahmed Salem, Michael Backes, Yang Zhang, and Neil Zhenqiang Gong, “MemGuard: Defending against black-box membership inference attacks via adversarial examples,” New York, NY, USA, 2019, CCS ’19, p. 259–274, ACM.

[13] Jiacheng Li, Ninghui Li, and Bruno Ribeiro, “Membership inference attacks and defenses in classification models,” in *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy*, New York, NY, USA, 2021, CODASPY ’21, p. 5–16, ACM.

[14] Milad Nasr, Reza Shokri, and Amir Houmansadr, “Machine learning with membership privacy using adversarial regularization,” in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, 2018, CCS ’18, p. 634–646, ACM.

[15] Ravi V Atreya, Joshua C Smith, Allison B McCoy, Bradley Malin, and Randolph A Miller, “Reducing patient re-identification risk for laboratory results within research datasets,” *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 95–101, 2013.

[16] Louise M Slaughter, “The genetic information nondiscrimination act: why your personal genetics are still vulnerable to discrimination,” *Surgical Clinics of North America*, vol. 88, no. 4, pp. 723–738, 2008.

[17] Imrana Siddiqui, Marco Aurelio, Ajay Gupta, Jenny Blythe, and Mohammed Y Khanji, “COVID-19: Causes of anxiety and well-being support needs of healthcare professionals in the uk: A cross-sectional survey,” *Clinical Medicine*, vol. 21, no. 1, pp. 66, 2021.

[18] Khaled El Emam and Fida Kamal Dankar, “Protecting privacy using k-anonymity,” *Journal of the American Medical Informatics Association*, vol. 15, no. 5, pp. 627–637, 2008.

[19] Ali Shahin Shamsabadi, Brij Mohan Lal Srivastava, Aurélien Bellet, Nathalie Vauquier, Emmanuel Vincent, Mohamed Maouche, Marc Tommasi, and Nicolas Papernot, “Differentially private speaker anonymization,” in *Proceedings on Privacy Enhancing Technologies*.

[20] Cynthia Dwork and Rebecca Pottenger, “Toward practicing privacy,” *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 102–108, 2013.

[21] James Gardner, Li Xiong, Yonghui Xiao, Jingjing Gao, Andrew R Post, Xiaoqian Jiang, and Lucila Ohno-Machado, “Share: system design and case studies for statistical health information release,” *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 109–116, 2013.

[22] Tiantian Feng and Shrikanth Narayanan, “Privacy and utility preserving data transformation for speech emotion recognition,” in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2021, pp. 1–7.

[23] Mimansa Jaiswal and Emily Mower Provost, “Privacy enhanced multimodal neural representations for emotion recognition,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, pp. 7985–7993.

[24] Vansh Narula, Kexin Feng, and Theodora Chaspari, “Preserving privacy in image-based emotion recognition through user anonymization,” in *Proceedings of the 2020 International Conference on Multimodal Interaction*, New York, NY, USA, 2020, ICMI ’20, p. 452–460, ACM.

[25] Karel Mundnich, Brandon M Booth, Michelle l’Hommedieu, Tiantian Feng, Benjamin Girault, Justin L’hommedieu, Mackenzie Wildman, Sophia Skaaden, Amrutha Nadarajan, Jennifer L Villatte, et al., “Tiles-2018, a longitudinal physiologic and behavioral data set of hospital workers,” *Scientific Data*, vol. 7, no. 1, pp. 354, 2020.

[26] Clément Feutry, Pablo Piantanida, Yoshua Bengio, and Pierre Duhamel, “Learning anonymized representations with adversarial neural networks,” in *International Workshop on Machine Learning and Artificial Intelligence (MLAI)*, 2018.