Predicting Job Performance Using Mobile Sensing

Shayan Mirjafari, *Dartmouth College, Hanover, NH, 03755, USA*
Hessam Bagherinezhad, *University of Washington, Seattle, WA, 98195, USA*
Subigya Nepal, *Dartmouth College, Hanover, NH, 03755, USA*
Gonzalo J. Martinez, *University of Notre Dame, Notre Dame, IN, 46556, USA*
Koustuv Saha, *Georgia Institute of Technology, Atlanta, GA, 30332, USA*
Mikio Obuchi and Pino G. Audia, *Dartmouth College, Hanover, NH, 03755, USA*
Nitesh V. Chawla, *University of Notre Dame, Notre Dame, IN, 46556, USA*
Anind K. Dey, *University of Washington, Seattle, WA, 98195, USA*
Aaron Striegel, *University of Notre Dame, Notre Dame, IN, 46556, USA*
Andrew T. Campbell, *Dartmouth College, Hanover, NH, 03755, USA*

We hypothesize that behavioral patterns of people are reflected in how they interact with their mobile devices and that continuous sensor data passively collected from their phones and wearables can infer their job performance. Specifically, we study day-today job performance (improvement, no change, decline) of N=298 information workers using mobile sensing data and offer data-driven insights into what data patterns may lead to a high-performing day. Through analyzing workers' mobile sensing data, we predict their performance on a handful of job performance questionnaires with an F-1 score of 75%. In addition, through numerical analysis of the model, we get insights into how individuals must change their behavior so that the model predicts improvements in their job performance. For instance, one worker may benefit if they put their phone down and reduce their screen time, while another worker may benefit from getting more sleep.

The recent development of mobile computing technology with various built-in sensors has made it possible to sense human behavior continuously at a low cost. In the context of ubiquitous computing and human–computer interaction, mobile sensing data collected from individuals are often combined with different machine learning algorithms to build models that understand and predict behavioral patterns. These predictive models can be embedded in mobile devices and support people in every aspect of life. In this article, we focus on mobile sensing in the workplace, where people spend much of their time to make a living. We discuss methods to predict workers' *day-to-day job performance* using passive sensing data from phones and wearables.

We use a number of psychologically validated job-related surveys as a proxy for workers' job performance. Each of these surveys capture somewhat different aspects of the workers' job performance: 1) individual task proficiency (ITP)[1]: measures the proficiency of a worker with respect to their performing core activities related to their job; 2) in-role behavior (IRB)[1]: measures the behavior required of a worker to accomplish the duties assigned in the organization; 3) organizational citizenship behavior (OCB)[2]: measures worker's behavior that are toward the organizational culture; and 4) counterproductive work behavior (CWB)[2]: measures behaviors demonstrated by a worker that

1536-1268 © 2021 IEEE
Digital Object Identifier 10.1109/MPRV.2021.3118570

negatively affects the well-being of the organization. Our model, using mobile sensing, predicts the performance of the workers on each of these surveys.

Wrangling each worker's sensing data to understand their job performance represents a very challenging task requiring deep knowledge across a wide variety of domains and occupations (e.g., specific tasks, working conditions). Therefore, there is a need for scalable approaches capable to automatically capture behaviors associated with job performance across a spectrum of different workplace environments. Alternatively, hand-crafted features used in previous machine learning approaches require expert knowledge about the ins and outs of specific domains (e.g., software engineers in a specific sector) to be able to create high-level features that capture the key dimensions of the problem/occupation under study and therefore do not scale well. With neural network based representation learning, the costs of resources and the development time of high-level feature engineering is reduced as data analysts no longer have to rely on hand-crafted features by domain experts. We use an unsupervised autoencoder to automatically augment the feature-set for predictions. Specifically, we predict day-to-day job performance (improvement, no change, and decline) of information workers across a number of different domains (e.g., tech, consultancy) using mobile sensing. We then perform gradient analysis on the prediction model to get insights into how individuals must change their behavior so that the model predicts improvements in their job performance. For example, one worker may benefit if they put their phone down and reduce their screen time, while another worker may benefit from more sleep.

METHODOLOGY

Dataset

The dataset for this article contains four months of data from 298 workers, employed at various industries across the United States, who had enrolled to participate in the Tesserae study.[3] Each worker was given a wristband and an Android/iOS application to install on their phones. The mobile sensing app collected data from different sensors on the phone as well as streams from the wearable device, which was connected to the phone via Bluetooth. While we discuss important and relevant information from the Tesserae project as it relates to job performance, refer to the works of Mattingly *et al.*[3] and Martinez *et al.*[4] for complete details of the study. Regarding the demographic information of the participants in our study: there are four distinct types of organizations in our analysis, which we label as A (44%), B (13%), C (17%), and D (26%)—percentages indicate the distribution of how the participants are spread into each organization. Group A consists of workers from a big technology company, B includes workers from a well-known consultancy company, C refers to workforce across a university setting, and Group D is made up of workers from different small organizations spread throughout the United States. There are considerable differences in the style of working conditions of these different groups. We believe the groups represent a challenging set of different types of work and therefore represent different demands on workers in terms of job performance. Gender-wise, 48% of participants are women and the rest are men. Age-wise, 82% of participants are under 40. The occupations of the participants also vary within as well as across the organizations. The distribution of occupations is as follows: 38% administrative, 13% business, 27% math and computer, 10% engineering, and 12% others.

Workers completed a set of self-reports relevant to job performance three days a week within the first two months of the study. There were four job performance metrics self-reported by participants used as the supervision, i.e., ITP,[1] IRB,[1] OCB,[2] and CWB,[2] capturing different aspects of workers' job performance. These metrics were presented to workers in the format of daily questionnaires, which measure their job performance. There are three questions about work proficiency for the ITP metric with five possible answers: selecting one assigns a number from 1 (very little) to 5 (a great deal) as a score to each question. So the total ITP sum score is in the range of 3–15. Similarly, there are seven questions for the IRB metric asking the level of agreement about work behaviors, with seven possible answers to each question scored from 1 (strongly disagree) to 7 (strongly agree). The total IRB sum score ranges from 7 to 49. Each of the OCB and CWB metrics contains eight questions about contextual job performance, with binary answers (yes or no) to each question with a score of 1 or 0, respectively. So the total sum score for both the OCB and CWB surveys is a number from 0 to 8. Details about questions of each metric are discussed in Table 2 of Mirjafari *et al.*'s work.[5] Note that a higher score of ITP, IRB, and OCB along with a lower score of CWB are associated with an improved work performance.

Supervision

To predict job performance, we must label the data with respect to the original scores of each metric to identify whether a worker performs above, under, or at the level of their ability. To be able to do this, we first

TABLE 1. Distributions of the labels for each of job performance metrics.

Metric	Decline	No change	Improvement
ITP	19%	70%	11%
IRB	18%	73%	9%
OCB	21%	66%	13%
CWB	20%	68%	12%

normalize the scores for each worker, separately, using z-score transformation for each of ITP, IRB, OCB, and CWB scores—this transforms the original scores to new scores with a mean of 0 and a standard deviation of 1. We then label all these normalized scores for each worker based on the interval a score falls within. If the normalized score is within -1 to 1 (i.e., $-1 < y_i < 1$), we label that day for the worker as "no change" in job performance. If the score is less than or equal to -1 (i.e., $y_i \leq -1$), or greater than or equal to 1 (i.e., $1 \leq y_i$), we label the corresponding day as a "decline" or "improvement," respectively, for each of ITP, IRB, and OCB scores as they are positively associated with job performance. However, for CWB, we swap "decline" and "improvement" labels as CWB is negatively associated with job performance. This way of categorization based on the standard deviation is a standard approach that has also been used in prior work. For example, Mehrotra and Musolesi[6] categorized depressive states using a similar approach. Table 1, shows the distributions of the labels for every job performance metric.

Lower Level Features

We consider several passive and continuous sensing streams from workers' phones and wearables over a four month period. These streams were collected using different duty cycles that tradeoff signal for energy taxation on the devices. We aggregate these time series for each worker per day. We call these features "lower level features" representing close to raw and simple aggregations including duration, mean, median, and standard deviation of stream values. The lower level features from different streams are listed as follows.

Physical activity: duration of being active, duration of being highly active, sedentary duration, and number of steps;

Mobility: number of unique locations visited and total distance traveled;

Phone usage: unlock duration and number of lock/unlock;

Heart rate: mean, median, and standard deviation of heart rate/inter-beat intervals;

Stress: mean, median, and standard deviation of stress level and duration of feeling stressed (high/moderate/low);

Sleep quality: bedtime/wake-up time, duration of sleep, duration of light/deep sleep, and duration of wake-ups during sleep hours;

Weather: sunrise/sunset time, length of daylight, cloudiness, humidity, feels-like, and precipitation.

Autoencoder-Based Features (Higher Level)

Hand-crafting higher level features relevant to job performance requires specific knowledge about different workers' circumstances (e.g., industrial settings, work flexibility, and occupation) and expertise in the domain of organizational behavior. In addition, it is also possible that the data collected by mobile devices are noisy due to reasons, such as users' wearing behaviors and movements. Therefore, there is no guarantee that even hand-crafted features lead models to achieving higher performance. It is also time-consuming and challenging to handcraft predictive features. However, in our analysis we avoid these challenges by automatically augmenting the feature set using an unsupervised autoencoder.[7]

In our use of autoencoder, we consider the output of the encoder as higher level features. We use these features for two reasons: 1) *Having more discriminative features instead of hand-crafting features leads to higher performance of prediction model.* Later, we see that these higher level features play an important role in improving prediction performance. 2) *Reducing noise in the lower level features to achieve a reliable model.*

PREDICTION

Before training an autoencoder, we first perform two things: 1) normalizing the lower level features per worker ($\frac{x^i_j - \mu_i}{x^i_{\max} - x^i_{\min}}$ for the jth element in the ith feature), and 2) creating a prediction dataset from those days of interest that job performance surveys are self-reported on. We hold the prediction set out of the autoencoder training process for the final supervised prediction task. The purpose of Step 1 (within-worker normalization) is to cancel out the possible effect of workers' baselines on biasing the network. Different workers may have different baseline patterns of

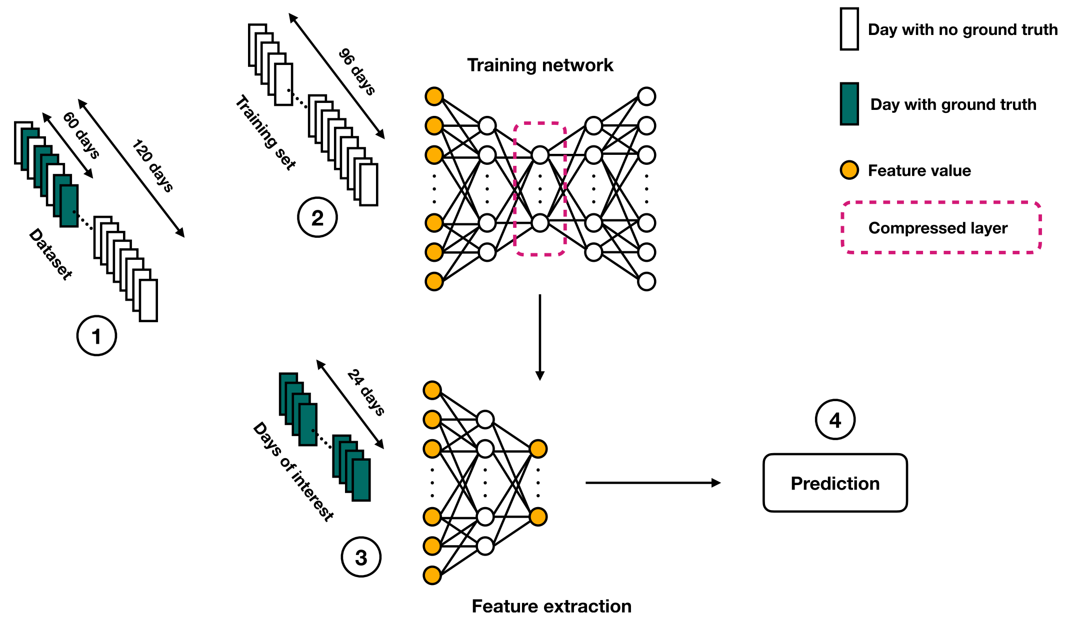

FIGURE 1. Modeling pipeline. (1) In total, there are 120 days of mobile sensing data per worker, 24 days with ground truth (i.e., job performance scores are self-reported) within the first 60 days. (2) An autoencoder is trained on the lower level features of 96 days with no job performance ground truth. (3) Lower level features of the days with ground truth are fed into the trained encoder to generate higher level features. (4) Finally, higher level feature values are extracted from the compressed layer for the final prediction task of job performance.

physical activity, heart rate, or phone usage. Therefore, these differences may bias the network toward high or low baselines.

To train the autoencoder, we split each worker's data with no reported ground truth into 80% and 20% chunks for training and validations, respectively. We thus train the autoencoder on the training set and fine-tune parameters on the validation set. The autoencoder is trained using different network architectures (e.g., various activation functions, different numbers of layers, and neurons). The loss function is mean squared error (MSE). Finally, the best architecture which gives us the lowest MSE on the validation set is selected. After the autoencoder is trained, we feed the data of the prediction set (days of interest) into the network to extract the higher level feature values from the middle layer (a.k.a. compressed layer) for the final prediction task of job performance. Figure 1 illustrates the entire modeling pipeline.

For the final stage of the job performance prediction task, we split the prediction dataset with augmented feature set into the training and test sets. The data of each worker are randomly split into 80% for training and 20% for test. We ultimately train an XGBoost[8] model on the training set and evaluate the its performance on the test set. The final stage of training is repeated five times with different random seeds. We later report the average performance in the following section.

EXPERIMENTAL ANALYSIS

Predictive Power

We evaluate the model performance on the test data of workers for each worker and each job performance metric. The performance measures are distributions across all workers (i.e., means and standard deviations of weighted precision, recall and F1 scores). Figure 2 shows the performance measures of the XGBoost models trained on different feature sets when predicting the job performance metrics. The performances are calculated on the test data of each worker separately. We compare the prediction accuracy once when only the original features are used, and again when both original features and higher level autoencoder features are used. As shown in the figure, for each job performance metric, the higher level features are given by the encoder with activation function that helps the prediction model obtain the highest F1 score. To identify whether the improvements in F1 scores are statistically significant, we perform Wilcoxon signed-rank paired test. The p-value of the test in all the comparisons for

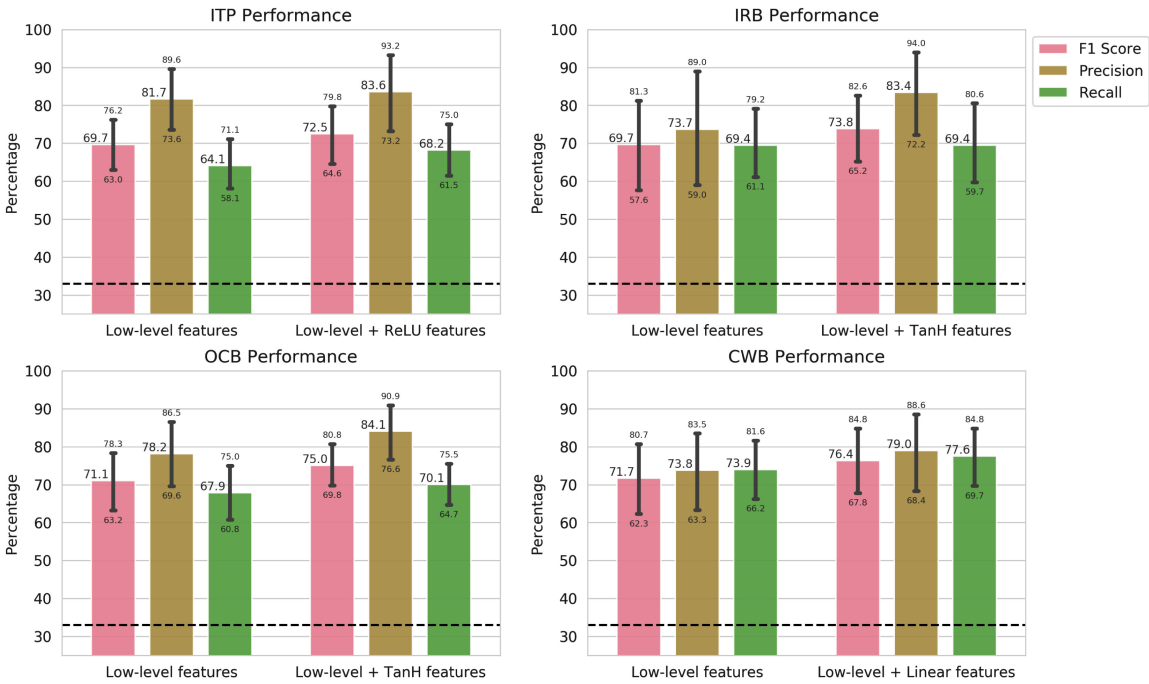

FIGURE 2. Model performance measures (weighted F1 score, precision, and recall) when predicting job performance metric using the best set of features, i.e., a combination of original features and those extracted from autoencoder with specific activation function (ReLU, tanh, or linear) that helps the model obtain the highest performance. The performances are calculated on the test data of each worker separately. So each value in the bar plots represents the performance of model when predicting job performance of N = 1 individual worker. The dashed line at 33.3% is the chance performance.

each job performance metric is less than 0.05 indicating that the improvements are significant. Also, given that the distribution of the test labels is imbalanced, we set a baseline model that always returns the category with the highest occurrence ("No change") for prediction regardless of the test example. We calculate the weighted F1 score of such a model on the test set as a performance baseline to compare with the reported F1 score in Figure 2. The baseline F1 scores for ITP, IRB, OCB, and CWB are 56%, 58%, 53%, and 56%, respectively.

Gradient Analysis

We run an "ablation" of the study on the trained model to learn what is the required change in the sample input for the model to predict a different category from its corresponding one. In the context of deep neural networks, ablation describes a procedure where certain parts of the network are changed or removed to gain a better understanding of the network's behavior. Ablation is therefore a simple way to look into causality. Understanding causality in the network is the most straightforward way to generate reliable knowledge. This kind of analysis and gradient computation is popular in computer vision domain.[9] Liu *et al.*[10] used the gradient analysis of the network to generate face images with different facial expressions from the original images. Inspired by this literature, we perform gradient analysis on the prediction model to get insights into how one individual may change their behavior so that the model predicts improvements in their job performance. To the best of our knowledge, this is the first time that the gradient analysis technique is applied to behavioral change suggestions using mobile sensing. Future work should investigate the broad capability of this approach while studying various mobile sensing problems.

Let us describe how we obtain knowledge from the numerical gradient analysis of the model. Let $F(x)$ be the function of the model that given an input feature vector x, it outputs the probability of x belonging to the relative high-performance category. By looking at $\frac{\partial F(x)}{\partial x}$ at a particular point x, we see the direction and the magnitude of the gradient that overall each input feature vector must be moved so that the model

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

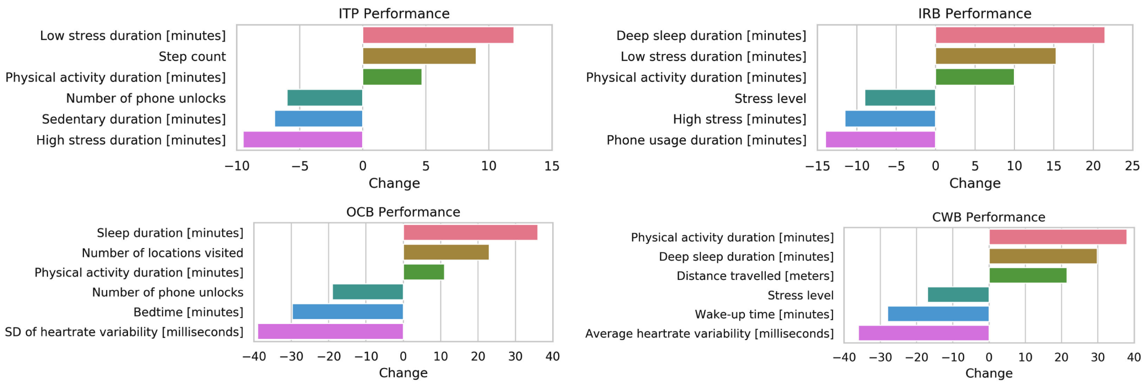

FIGURE 3. Sensing features that if linearly changed (increasing or decreasing) in a sample input, the model would predict the sample as belongs to the high-performance category for each job performance metric. The *x*-axis of each figure represents the amount of required change that the corresponding feature item should take so that the model predicts improvement in the job performance metric. For example, as shown for improvement in individual task performance (ITP), the low stress duration should increase by 12.3 minutes, and correspondingly the duration of high stressed feeling should decrease by 8.5 minutes. Similarly, the phone should be less used as the number of phone unlocks is decreased by 6. The change units are indicated next to each feature name. The results are aggregated over all workers.

would predict a higher $F(x)$. That is, a higher probability score for the higher performance category.

Each data point in our model is a feature vector representing a (worker, day) pair. By performing the gradient analysis on each test data point, we learn what features for that (person, day) pair would need to be modified so that the probability of higher performance prediction increases. We can accumulate this gradient information across all days of a worker to gain aggregated insight on what are some feature values that a worker could increase/decrease to get higher performance overall. We can also accumulate over all days and all participants to gain insights into how feature values could be increased/decreased to get the overall participants to show improvements in their job performance. Also, since $\frac{\partial(f+g)}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial x}$, summing the per (person, day) pair gradients is equivalent to calculating gradients across all participants in the dataset. Figure 3 shows some sensing features, their change direction, and the relative magnitude by which each feature should be linearly increased or decreased so that the model predicts higher probability score for improvement in job performance.

RELATED WORK

Mobile technologies offer the advantage of passively collected context based *in situ* data, which can be used to make a variety of inferences about the individuals' behaviors. Schaule *et al.*[11] showed that physiological data obtained from wearables can be used to detect office workers' cognitive load. Olguin *et al.*[12] used wearable sensing technology to assess human behavior in organizational settings. Using passively sensed heart rate and sleep data from wearable, Saha *et al.*[13] showed that greater role ambiguity is related to greater heart rate, greater stressful arousal, and decreased work hours and sleep. Robles-Granda *et al.*[14] proposed a framework that can extract meaningful predictors from noisy and incomplete data derived from wearable, mobile, and social media sensors to predict constructs of physical and physiological behavior, psychological states and traits, and job performance. Das Swain *et al.*,[15] studied how a feature set of activities complements effects of personality to explain a worker's performance. In our prior work,[16] we attempt to predict subjective performance measures from objective data that intend to find what objective data are related to individual success in organizations, i.e., getting promoted. Similarly, in our another work,[5] we find correlation of high and low workplace performance with several hand-crafted features obtained from mobile and wearable sensing. In this article, we advance our work by predicting the "day-to-day workplace performance" based on an approach potentially applicable to a wide variety of different jobs.

Much of the prior work uses hand-crafted features for modeling. However, there is an increasing number of studies using deep learning for predictive models. The advantage being that the deep models do not need hand-crafted features and can learn to find patterns from the raw data. Yao *et al.*[17] proposed DeepSense, a deep learning framework for processing time-series data

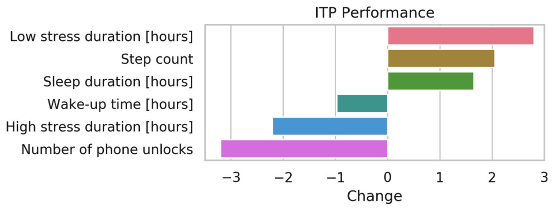

FIGURE 4. Features that can be modified as behavioral patterns of an information worker in a tech company to achieve a boost in task performance at work derived from our modeling. These suggestions can be given to the worker as personal tips.

collected from mobile sensing. Mehrotra *et al.*[6] used an autoencoder to automatically generate features from GPS-based mobility traces to predict depressive states of the participants. They state that potentially hidden significant and informative emerging patterns of human behavior in the data might not be characterized by engineered features, which are typically the result of a time-consuming trial-and-error process to generate. Therefore, such approach, if implemented in relation to workplace performance, in workplace performance would mean that we would not need to generate hand-crafted features for workers taking into account their different circumstances.

DISCUSSION AND FUTURE WORK

Through numerical analysis of the model gradients, we show how participants could change their behavioral patterns so that the prediction model would predict a higher performing day for them. This approach can be used for future mobile sensing and intervention systems that could be personalized in a way that provides workers with insightful individualized tips to improve their workplace performance. Figure 4 shows how a full-time worker in a technology company could change their behavior to improve their task performance at work as suggested by the model. As shown, the workers seem to perceive themselves as performing better on days when increased periods of lower stress are detected. Also, it would be better for them to increase their sleep duration and wake up earlier. In addition, these people may benefit from increasing step counts—perhaps their work related tasks keep them at their desk for long periods. However, note that these example suggestions derived from our model are not validated in empirical trials. Future work could investigate how such suggestions can affect a person's job performance by getting feedback and personal opinions from the participant. This will allow us to validate the efficacy of such suggestions with respect to each person's work behavior, flexibility, and preferences before our proof-of-concept could be used as a workplace application.

As mentioned in Section "Prediction," for personalized modeling purposes, the data of each person in our study are split into both the train and test sets. We also train the model using the leave-N-subjects-out approaches. However, the results are not as good as the ones reported in this article. We argue that the assessment of job performance using mobile sensing is a challenging task and is not simply generalizable from one person to another.

Future of work research is rapidly gaining momentum but we have little guidance with respect to the privacy and ethical aspects of it. While any form of sensing-based studies could be seen as potentially invasive, workplace sensing specifically needs to be considered with utmost importance. We believe that workers would be interested in participating in such studies if they feel that they have control over their data and importantly gain some utility from a workplace app, such as feedback to help improve their performance. However, such data could also be used by employers. This begs the question of how worker's data are going to be secured and protected against misuse. This demands deeper discussion regarding the guidelines of how the future of work technology is deployed and used. There is a clear need for workers to manage who accesses their data and who does not. A clear protocol needs to be established around the governance and ownership of behavioral data in the workplace for the future of work technologies to be adopted widely in the workforce.

Given the rapid trajectory of phone–computer architecture and ML support, we are convinced that the type of models presented in our work which are trained and run offline are capable of being trained and run online. Therefore, every part of the inference process presented in this article (i.e., collecting the streams, computing the lower level features, generating higher level features, and final prediction tasks) can be implemented and run locally on an individual's smartphone for personalized inferences. Such an evolution of implementing the complete ML pipeline on the phone will limit the exposure of the user's data and better maintain their privacy.

CONCLUSION

We envision that future mobile sensing systems will be capable of predicting workplace performance for workers across a wide variety of industries. These machines will be capable of adapting to a wide variety of tasks within a field and across different occupational domains, and be personalized to specific workers' environments. We advanced our understanding of the scaling and interpretability challenges by first applying an autoencoder to automatically augment our feature set and successively performing a numerical gradient analysis of the trained models to understand how an individual worker should change their behavioral patterns to improve their performance at work. By analyzing information workers' (N = 298) mobile sensing data, we predicted their job performance with an F1 score of 75%.

ACKNOWLEDGMENTS

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA under Contract 2017-17042800007.

REFERENCES

1. W. C. Borman and S. Motowidlo, "Expanding the criterion domain to include elements of contextual performance,"*Personnel Selection in Organizations*. San Francisco, CA, USA: Jossey-Bass, p. 71, 1993, doi: 10.5465/amr.1993.9402210158.
2. R. S. Dalal, H. Lam, H. M. Weiss, E. R. Welch, and C. L. Hulin, "A within-person approach to work behavior and performance: Concurrent and lagged citizenship-counterproductivity associations, and dynamic relationships with affect and overall job performance," *Acad. Manage. J.*, vol. 52, no. 5, pp. 1051–1066, 2009, doi: 10.5465/amj.2009.44636148.
3. S. M. Mattingly *et al.*, "The tesserae project: Large-scale, longitudinal, in situ, multimodal sensing of information workers," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–8, doi: 10.1145/3290607.3299041.
4. G. J. Martinez *et al.*, "On the quality of real-world wearable data in a longitudinal study of information workers," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, 2020, pp. 1–6, doi: 10.1109/PerComWorkshops48775.2020.9156113.
5. S. Mirjafari *et al.*, "Differentiating higher and lower job performers in the workplace using mobile sensing," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 2, pp. 1–24, 2019, doi: 10.1145/3328908.
6. A. Mehrotra and M. Musolesi, "Using autoencoders to automatically extract mobility features for predicting depressive states," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–20, Sep. 2018, doi: 10.1145/3264937.
7. G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006, doi: 10.1126/science.1127647.
8. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
9. J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," *IEEE Trans. Evol. Comput.*, vol. 23, no. 5, p. 828–841, Oct. 2019, doi: 10.1109/TEVC.2019.2890858.
10. Y. Liu, X. Hou, J. Chen, C. Yang, G. Su, and W. Dou, "Facial expression recognition and generation using sparse autoencoder," in *Proc. IEEE Int. Conf. Smart Comput.*, 2014, pp. 125–130, doi: 10.1109/SMARTCOMP.2014.7043849
11. F. Schaule, J. O. Johanssen, B. Bruegge, and V. Loftness, "Employing consumer wearables to detect office workers cognitive load for interruption management," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–20, Mar. 2018, doi: 10.1145/3191764.
12. D. O. Olguín, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland, "Sensible organizations: Technology and methodology for automatically measuring organizational behavior," *IEEE Trans. Syst., Man, Cybern., Part B.*, vol. 39, no. 1, pp. 43–55, Feb. 2008, doi: 10.1109/TSMCB.2008.2006638.
13. K. Saha, M. D. Reddy, S. Mattingly, E. Moskal, A. Sirigiri, and M. De Choudhury, "Libra: On LinkedIn based role ambiguity and its relationship with wellbeing and job performance," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. CSCW, pp. 1–3, Nov. 2019, doi: 10.1145/3359239.
14. P. Robles-Granda *et al.*, "Jointly predicting job performance, personality, cognitive ability, affect, and well-being," *IEEE Comput. Intell. Mag.*, vol. 16, no. 2, pp. 46–61, May 2021, doi: 10.1109/MCI.2021.3061877.
15. V. Das Swain *et al.*, "A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 4, pp. 1–27, 2019, doi: 10.1145/3369828.
16. S. Nepal, S. Mirjafari, G. J. Martinez, P. Audia, A. Striegel, and A. T. Campbell, "Detecting job promotion in information workers using mobile sensing," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–28, 2020, doi: 10.1145/3414118.

17. S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "DeepSense: A unified deep learning framework for time-series mobile sensing data processing," in *Proc. 26th Int. Conf. World Wide Web. Int. World Wide Web Conf. Steering Committee*, 2017, pp. 351–360, doi: 10.1145/3038912.3052577.

SHAYAN MIRJAFARI is currently a graduate student of computer science with Dartmouth College, Hanover, NH, USA. His interests include machine learning systems and ubiquitous computing. He is the corresponding author of this article. Contact him at shayan@cs.dartmouth.edu.

HESSAM BAGHERINEZHAD received the Ph.D. degree in computer science specialized in computer vision and machine learning from the University of Washington, Seattle, WA, USA, in 2020. Contact him at hessam@cs.washington.edu.

SUBIGYA NEPAL is currently a Ph.D. student in computer science with Dartmouth College, Hanover, NH, USA. His research focuses on passive sensing and its applications in understanding and modeling human behavior. Contact him at sknepal@cs.dartmouth.edu.

GONZALO J. MARTINEZ is currently a Ph.D. candidate with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA. His research interests include intersection of data science, human–computer interaction, and behavioral studies. Contact him at gmarti11@nd.edu.

KOUSTUV SAHA will be joining Microsoft Research, Montreal, QC, Canada, as a Research Scientist. His research interests include social computing and computational social science. He received the Ph.D. degree in computer science from the Georgia Institute of Technology, Atlanta, GA, USA, in 2021. Contact him at koustuv.saha@gatech.edu.

MIKIO OBUCHI is currently with Amazon Web Services, Japan. His interests include artificial intelligence, machine learning, and ubiquitous computing. He received the master's degree in computer science from Dartmouth College, Hanover, NH, USA, in 2020. Contact him at mikio.obuchi.gr@dartmouth.edu .

PINO G. AUDIA is currently a Professor of management and organizations with the Tuck School of Business, Dartmouth College, Hanover, NH, USA. His research interests include learning from performance feedback, self-enhancement in organizations, the social networks of individuals, and the influence of geographic communities on organizational outcomes. He received the Ph.D. degree in 1996 at the University of Maryland, USA. Contact him at Pino.G.Audia@tuck.dartmouth.edu.

NITESH V. CHAWLA is currently a Frank M. Freimann Professor of computer science and engineering with the University of Notre Dame, Notre Dame, IN, USA. He is the Founding Director of the Lucy Family Institute for Data and Society. His research interests include making fundamental advances in artificial intelligence, data science, and network science. He received the Ph.D. degree in 2002 at the University of South Florida, USA. He is a Senior Member of the IEEE. Contact him at nchawla@nd.edu.

ANIND K. DEY is currently a Professor and Dean of the Information School, University of Washington, Seattle, WA, USA, where he also an adjunct professor with the Department of Human-Centered Design and Engineering. His research interests include the intersection of human–computer interaction, machine learning, and ubiquitous computing. He received the Ph.D. degree in 2000 at the Georgia Institute of Technology, USA. He is a Member of the IEEE. Contact him at anind@uw.edu.

AARON STRIEGEL is currently a Professor of computer science and engineering with the University of Notre Dame, Notre Dame, IN, USA. His research interests include instrumenting the wireless networked ecosystem to gain insight with respect to user behavior and global network performance, interdisciplinary collaborations exploring the interplay of sociology, neuroscience, and wireless technology. He received the Ph.D. degree in 2002 at Iowa State University, USA. He is a Senior Member of the IEEE. Contact him at striegel@nd.edu.

ANDREW T. CAMPBELL is currently an Albert Bradley 1915 Third Century Professor with Dartmouth College, Hanover, NH, USA. In 2008, his group developed the first smartphone application to unobtrusively infer human activity and sociability for the iPhone. Since then, he has developed a number of fundamental sensing algorithms that have helped pioneer the use of smartphones as sensors. He received the Ph.D. degree in 1996 from Lancaster University, England. Contact him at andrew.t.p.campbell@gmail.com.