Journal of the American Medical Informatics Association, 29(8), 2022, 1400–1408
https://doi.org/10.1093/jamia/ocac071
Advance Access Publication Date: 18 May 2022
Research and Applications

Research and Applications

Identifying daily activities of patient work for type 2 diabetes and co-morbidities: a deep learning and wearable camera approach

Hao Xiong, Hoai Nam Phan, Kathleen Yin, Shlomo Berkovsky, Joshua Jung, and Annie Y.S. Lau

Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, New South Wales, Australia

Hao Xiong and Hoai Nam Phan are the co-first authors.

Corresponding Author: Hao Xiong, PhD, Centre for Health Informatics, Australian Institute of Health Innovation, Level 6, 75 Talavera Road, North Ryde, NSW 2109, Australia; hao.xiong@mq.edu.au

Received 17 December 2021; Revised 8 April 2022; Editorial Decision 20 April 2022; Accepted 28 April 2022

ABSTRACT

Objective: People are increasingly encouraged to self-manage their chronic conditions; however, many struggle to practise it effectively. Most studies that investigate patient work (ie, tasks involved in self-management and contexts influencing such tasks) rely on self-reports, which are subject to recall and other biases. Few studies use wearable cameras and deep learning to capture and classify patient work activities automatically.

Materials and Methods: We propose a deep learning approach to classify activities of patient work collected from wearable cameras, thereby studying self-management routines more effectively. Twenty-six people with type 2 diabetes and comorbidities wore a wearable camera for a day, generating more than 400 h of video across 12 daily activities. To classify these video images, a weighted ensemble network that combines Linear Discriminant Analysis, Deep Convolutional Neural Networks, and Object Detection algorithms is developed. Performance of our model is assessed using Top-1 and Top-5 metrics, compared against manual classification conducted by 2 independent researchers.

Results: Across 12 daily activities, our model achieved on average the best Top-1 and Top-5 scores of 81.9 and 86.8, respectively. Our model also outperformed other non-ensemble techniques in terms of Top-1 and Top-5 scores for most activity classes, demonstrating the superiority of leveraging weighted ensemble techniques.

Conclusions: Deep learning can be used to automatically classify daily activities of patient work collected from wearable cameras with high levels of accuracy. Using wearable cameras and a deep learning approach can offer an alternative approach to investigate patient work, one not subjected to biases commonly associated with self-report methods.

Key words: patient work, self-management, wearable camera, deep learning

INTRODUCTION

Increasingly, people with chronic conditions are expected to take care of their health outside of medical settings (ie, self-management). Self-management refers to actions taken by people to recognize, treat and manage their own health.[1] It is widely promoted to empower patients, improve health outcomes, and reduce constraints on the overstretched health system. However, many individuals living with chronic conditions struggle to practice self-management effectively.[2–4]

A major challenge to investigating the barriers in self-management is the difficulty in obtaining a detailed and unbiased picture of the 'work' involved from an individual perspective (ie, patient work). *Patient work*, a concept derived from health ergonomics, is a way to study self-management by breaking down the tasks, contexts and the work involved. It describes the physical and cognitive tasks conducted by the individual to manage one's health, as well as the holistic sum of contexts (physical, social, mental, and organizational) that influence the work conducted. While self-management focuses on the strategies people employ, patient work breaks down these strategies into day-to-day tasks and examines how the effort and time involved, as well as the contextual and ergonomic factors, affect the way self-management is practised and why some tasks are carried out while others are neglected.[5–9]

However, past attempts to study patient work (or self-management) rely primarily on self-reported data, which are subject to problems commonly associated with self-report methods, such as recall bias, data collected being unreliable or inconsistent, or using data collection approaches that are hard to execute.[10,11] Other approaches, which rely on wearable sensors, tend to perform poorly due to noise or can only identify simple activities (eg, walking) that may not be useful to understand the full spectrum of health behaviors. Visual logging (eg, via wearable cameras) has been proposed as an alternative approach to automatically capture the full spectrum of self-management activities, without relying on self-report methods nor being restricted by the narrow focus of sensor-based approaches.

Wearable cameras capture first person point-of-view recordings, providing details on how real-time contextual factors influence health behaviors, and offering a wider perspective on self-management as opposed to the narrow focus using other wearable approaches.[12–14] Wearable cameras have been used for dietary assessment,[15] travel and sedentary behavior assessment,[16,17] monitoring behavior changes in dementia,[18] and recognition of physical activities.[19]

To our knowledge, we are the first to propose a wearable camera to capture daily activities of people with their Type 2 diabetes and comorbidities. However, current approaches still rely on manual viewing and analysis of video footage captured by wearable camera, which is a daunting task due to the large volume of data collected. Thus, in this study, we report an original approach of using deep learning to automatically classify activities of self-management collected by wearable cameras.

METHODOLOGY

Patient work data, ethics, and privacy

The dataset was collected as a part of our previous study,[5] which investigated patient work of people living with type 2 diabetes and comorbidities (More details in Supplementary Appendix). Briefly, 26 participants were recruited with a median age of 72 years, with 16 male and 10 females, a mean period of living with a T2DM diagnosis for 19.5 years, where 16 were using insulin (10 using oral medications) at the time of the study.

In ref.,[5] we describe the study protocol of our mixed-methods study, which includes: (1) visiting participants at home to learn about their self-management routines through interviews and questionnaires; (2) asking participants to wear a body camera for an entire waking day (~16 h) and complete a time-use diary to document their daily activities; and (3) visiting participants the following day for a post-study interview and to go through the wearable camera footage together. Eventually, the data captured by wearable camera, diary, interviews, and questionnaires were utilized to analyze self-management routines and behaviors. Here, the participants were given a wearable camera (Edesix VB 300, Edinburgh, UK), which automatically recorded silent continuous video footage. It could be attached to clothing and/or worn on a lanyard, being located in front of the chest.

The study was approved by the Macquarie University Human Research Ethics Committee for Medical Sciences (reference number 5201700718) and informed participant consent was obtained. For full details on how we address these issues, please refer to our study protocol.[5] We elaborate on this in the Discussion section.

Deep-learning approach and experimental setup

Our approach of applying deep-learning to classify daily activities of patient work consists of 4 main steps: dataset preparation, *Sleeping* activity filtering, non-sleeping activities classification, and weighted ensemble network (a brief flowchart is shown in Figure 1). We filter out *Sleeping* as one of the initial steps because wearable cameras normally face the ceiling or empty walls during the recording, so that the captured images lack texture and cannot be effectively recognized. Consequently, sleeping activity images are filtered out before classifying other activities.

Our method was implemented using Pytorch, an open-source library for deep learning applications development. Some models in our ensemble network were pre-trained on the large classification dataset ImageNet.[20] Details of our deep learning environment setup are found in Table 1.

Step 1—dataset preparation

We conducted further data cleaning to optimize the screenshots for deep learning training. Some of the activity coding did not contain enough images for training, and we therefore aggregated images from similar activities into one category. By doing this, the dataset was collapsed from 23 daily activity classes into 12 classes (see Supplementary Appendix Table S2 for more details on we collapsed 23 classes into 12).

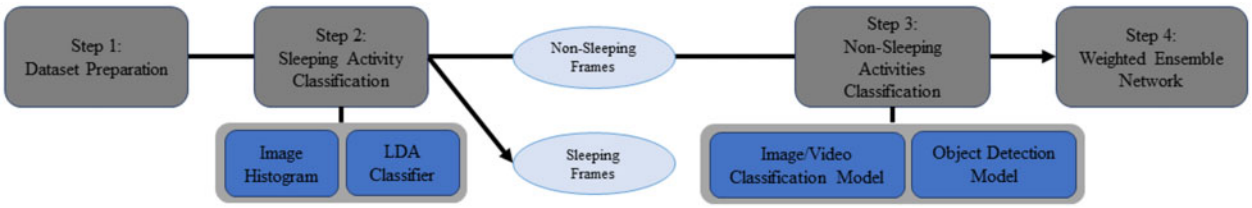

Figure 1. An overview of our approach. It consists of 4 steps: Dataset Preparation, Sleeping Activity Classification, Non-Sleeping Activities Classification, and Weighted Ensemble Network.

Table 1. Our weighted ensemble network consists of 2 models: image and video classification based networks

Hyper-parameter	Image classification models	Video classification models
Input size	640 × 368	640 × 368
Batch size	32	8 × 8
Inference batch size	4	64
Learning rate	0.0001	0.001
Momentum	0.09	0.9
Optimizer	SGD	SGD
Epochs	50	200
Early stopping	5	5
Pre-trained dataset	ImageNet	Kinetics-MomentsInTime

Note: The implementation parameters of these models are shown in the table.

In our method, the dataset was first split into training and test subset using the ratio of 7:3. The training data was then further split into a training and validation dataset with the ratio of 9:1. Here, the validation dataset is utilized to evaluate and select the optimal trained model in the training phase. The splitting process was manually implemented by analyzing all the images as a set of events, where images with similar events were grouped into the same category. Table 2 shows the training and testing data for each of the 12 activity classes.

Step 2—sleeping activity filtering

The *Sleeping* activity is separately recognized by a machine learning classifier since wearable cameras normally faced the ceiling or empty walls when participants were asleep. As a result, images captured during sleeping are highly likely to contain the ceiling/empty walls only. For these textureless sleeping images, no features can be effectively extracted, learned, and recognized by neural networks. In contrast, images from other non-sleeping activities contain various objects and textures, which can be exploited to differentiate from sleeping images.

Linear discriminant analysis

To classify whether an image is a *Sleeping* activity, we first convert the RGB images from all activities to image histograms. Here, image histogram counts the number of pixels for each tonal value in an image, which thus represents the distribution of entire tonal.[21] (shown in Step 2 of Figure 2) Then, these histograms are fed into Linear Discriminant Analysis (LDA)[22] classifier for training and recognition. As shown in Step 2 of Figure 2, the pixel values of sleeping images are normally distributed within the tonal range of 70–120 in our dataset, while the pixel values of the non-sleeping images are scattered across the full tonal range. Visually, the patterns of sleeping and non-sleeping image histograms differ substantially, which renders them distinguishable by the LDA classifier. After classifying all testing images using LDA, those classified as non-sleeping activities proceed to the subsequent Step 3 of activity classification.

Step 3—Non-sleeping activities classification

All video images of non-sleeping activities can be classified using several deep learning-based models: *image classification network* (ResNet152,[23] WideResNet,[24] and ResNeXt 50 and ResNeXt 100[25]), *video classification network* (3D-ResNet[26]), and *object detection network* (YoloV3[27]). All these techniques were deployed and compared to identify the best approach to classify each activity class.

Table 2. Dataset distribution

Activity	Training Events (images)	Validation Events (images)	Testing Events (images)
Managing health	31 (577)	5 (60)	7 (94)
Exercise	8 (1231)	1 (142)	3 (550)
Food related	130 (4491)	7 (488)	40 (1606)
Indoor	125 (7956)	17 (910)	66 (3614)
Outdoor	37 (1514)	6 (161)	31 (897)
Shopping	12 (693)	1 (87)	1 (291)
Electronic devices	52 (3893)	9 (417)	8 (1747)
Driving	52 (2392)	6 (263)	25 (1379)
Socializing	88 (2067)	13 (209)	15 (976)
Watching TV	27 (5214)	10 (408)	14 (2226)
Study	92 (1909)	7 (191)	30 (599)
Sleeping	**Training images** 1062		**Testing images** 353

Note: There are 12 different activities in our dataset and each activity is further broken down into training, validation, and testing images.

Image classification network

Deep convolutional neural networks have led to a series of breakthroughs for image classification. They naturally incorporate low/mid/high level features and classifiers into an end-to-end multilayer fashion.[23–26] The low/mid/high level features are extracted from images by convolutional layers of the network. ResNet152[23] first proposes a devised residual block to increase the network depth. As a result, it is possible to extract higher level features and achieve high image classification accuracy. Besides increasing the network depth, WideResNet also aims to augment the width of residual network and proves to be more accurate.[24] Other than network depth and width, the concept of cardinality (the size of the set of transformations) was first introduced in ResNeXt,[25] and showed that increasing cardinality is more effective for enhancing image classification accuracy than simply going deeper or wider.

Video classification network

Traditional ResNet exploits 2D convolutions to extract features for image classification.[23] However, 2D convolutions can only extract features within a single image. Normally, the extracted features are referred to as *spatial information*. Instead of 2D convolutions, 3D-ResNet[26] incorporates 3D convolutions that not only extract features within one image but also across several consecutive video images. Therefore, in addition to spatial information, the extracted features also contain *temporal information* across the video images, which is key to improving video classification accuracy.

Object detection network

In our dataset, the *Socializing* activity typically occurs in an indoor/outdoor setting, where images belonging to the *Socializing* activity are often misclassified as indoor or outdoor activities using image/video classification networks. Based on the observation that most *Socializing* images contain a person or body part of another person that the participant is interacting with, we apply YoloV3[27] object detection model to detect person objects (eg, arms, head, legs) in images, similar to[28] and.[29] This way, images containing the detected person/body parts are regarded as *Socializing*.

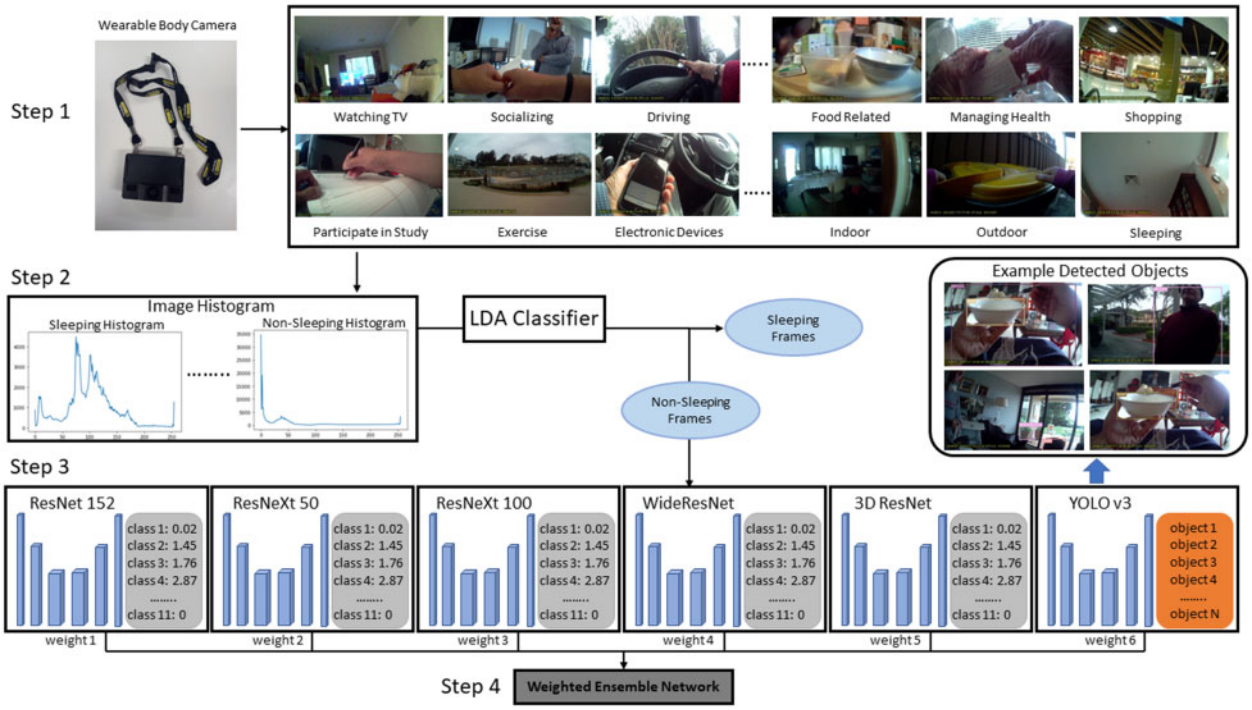

Figure 2. Flowchart and detailed illustration for each step involved in our approach.

Step 4—weighted ensemble network

In Step 3, the image, video classification and object detection networks are applied independently to all testing images. In this step, we designed a weighted ensemble network aggregating all individual predictions from these networks, in order to achieve more accurate classification.

Specifically, each prediction from the chosen network is assigned a weight that defines the importance and reliability of that network. Here, the weighted output of a network determines how much it contributes to the final prediction. Following this, we sum up the weighted outputs from all networks in the ensemble and an aggregated output with newly calculated probability scores is generated. Among the probability scores of the aggregated output, the one with highest score is referred to as the predicted activity class. Details of our proposed method are illustrated in Figure 2. An ablation study was also conducted to identify the optimal combination of weights for the best-performing weighted ensemble network.

Evaluation metrics

In our experiments, we utilize **Top-1** and **Top-5** metrics, commonly used to evaluate image/video classification performance,[23–26] to evaluate the classification accuracy of our model and baselines. **Top-1** accuracy reports conventional accuracy, assessing whether the highest probability prediction matches the ground truth activity label of an image. **Top-5** accuracy measures whether the ground truth label is included in the 5 highest probability predicted labels produced by the model.

Baseline methods

We compare the performance of our weighted ensemble network against several baselines:

ResNet152[23]—an image classification method that exploits the residual learning framework to avoid potential overfitting.

WideResNet[24]—an image classification method that focuses on increasing the width rather than the depth of the network.

ResNext50 and ResNext100[25]—an image classification method that devises a new dimension called cardinality, in addition to the depth and width of the network.

3D ResNet[26]—a video classification method that exploits temporal information by analyzing the relationships among category and instance.

RESULTS

Performance of baselines and weighted ensemble network

Table 3 shows the Top-1 and Top-5 scores (in %) obtained in the evaluation. We compare the proposed method with the aforementioned baselines. The highest score for each activity is highlighted in bold. First, we observe that the average Top-1 and Top-5 scores of WideResNet101 are the lowest among all methods. In fact, the overall performances of the whole image-based classification methods, including ResNext50, ResNext100, and ResNet152, are inferior to 3D ResNet. Second, we observe that by exploiting the temporal information for activity classification, the video-based 3D ResNet substantially enhances the accuracy compared to image classification methods, mainly for Top-1 predictions. Averaging Top-1 and Top-5 scores of all image-based classification methods, we observe an improvement of 41.9% and 3.8% for Top-1 and Top-5, respectively. More importantly, the average Top-1 and Top-5 scores of the proposed weighted ensemble network outperforms all the other methods. This justifies our assumption that integrating several baseline classification and object recognition methods into an ensemble network indeed enhances performance.

In addition to average accuracy comparison, the Top-1 scores of our method are highest for all activities and the Top-5 scores of our

Table 3. A comparison of our weighted ensemble network and its baseline methods (The best result is indicated in boldface)

Activity	ResNet152 Top-1/Top-5	WideResNet101 Top-1/Top-5	ResNeXt50 Top-1/Top-5	ResNeXt100 Top-1/Top-5	3DResNet Top-1/Top-5	Ours Top-1/Top-5
Socializing	3%/7%	0%/4%	8%/9%	8%/8%	1%/4%	**81%/83%**
Sleeping	7%/8%	7%/8%	0%/2%	1%/1%	0%/1%	**90%/90%**
Driving	93%/95%	82%/93%	87%/95%	89%/95%	100%/100%	**100%/100%**
Electronic device	63%/95%	44%/88%	44%/88%	63%/95%	82%/**100%**	**86%**/92%
Exercise	74%/97%	68%/95%	67%/91%	74%/97%	100%/100%	**100%/100%**
Food related	33%/85%	26%/86%	22%/**90%**	22%/85%	54%/83%	**71%**/89%
Indoor	57%/98%	49%/96%	57%/**98%**	57%/98%	64%/91%	**79%**/90%
Managing health	4%/31%	2%/14%	1%/**34%**	4%/31%	13%/13%	**13%**/15%
Outdoor	51%/78%	48%/74%	46%/74%	51%/78%	86%/**97%**	**90%**/96%
Shopping	91%/96%	85%/95%	84%/97%	91%/96%	100%/100%	**100%/100%**
Study	33%/78%	26%/78%	23%/74%	30%/78%	59%/86%	**74%/87%**
Watching TV	70%/97%	62%/95%	53%/94%	70%/97%	97%/100%	**99%/100%**
Average	59.9%/84.1%	50.6%/79.1%	50.0%/81.2%	58.9%/83.3%	74.5%/85.0%	**82.7%/87.1%**

method are highest for 7 activities out of the 12. Furthermore, the effect of applying ensemble technique is even more evident in some activities, such as *Socializing* and *Sleeping*. For these classes, the accuracy of our ensemble method substantially outperforms the other baselines. It can also be seen that for *Driving*, *Exercise*, and *Shopping* activities, our method achieved 100% accuracy in the Top-1 score. However, *Managing Health* demonstrated lowest accuracy scores than other activities, with 13% and 15% Top-1 and Top-5 scores, respectively. A confusion matrix of the detailed prediction by our method is shown in Figure 3. As can be seen, some activities are hardly distinguishable primarily due to these activities sharing similar visual contexts. For instance, *Food Related* activity normally occurs in an indoor setting and is often misclassified as *Indoor* activity. Likewise, *Indoor* activity is likely to be misclassified as *Food Related* and *Study* activities.

Ensemble model weights

Our weighted ensemble network enhances classification accuracy by combining several baseline classification networks. For each baseline network, a weight is assigned to define the importance and reliability of that model. The final prediction is then made by fusing all the predictions from baselines using their weights.

However, there is no underlying principle to determine the optimal combination of weights that generates the best result. We empirically tested several combinations and found that we were able to achieve promising performance by following these heuristic rules. In principle, the 4 image classification models, ResNet152, ResNeXt 50, ResNeXt 100, and WideResNet, share the same relatively low weights. Besides, the video classification model—3D ResNet has a higher weight than the image classification models owing to its higher predictive accuracy. Lastly, the object detection model YOLOv3 was allocated the highest weight in the ensemble network. Given the above rules, our weighted ensemble network sampled 5 combinations of weights shown in Table 4. Of the evaluated weighted combinations, set 5 achieved the highest accuracy, although this may not be the most optimal combination. However, the overall number of possible combinations is high and we are unable to test all the possible combinations to identify the most optimal combination. Despite that, we demonstrate that our weighted ensemble network with the selected combination (set 5) outperformed all the other sets as well as baseline methods.

Visualization of patient work classifications

Summarizations of daily activities for selected participants are visualized in Figures 4 and 5. In Figure 4, we illustrate classification examples generated by our method, alongside ground truth labels. It is noteworthy that some activities have relatively lower accuracy, due to similar physical contexts. For instance, the *Participate in Study* image (first image of second column) in Figure 4 was misclassified as *Indoor*, as it appeared like an indoor environment even from a human perspective. Another *Indoor* activity shown in the third row of first column was misclassified as *Food Related* Activities due to the kitchen-like scenario presenting in the image.

In Figure 5, each pair of bars represents a patient and contains a variety of daily activities that are encoded by different colors. For ease of comparison, each prediction bar of a participant (POxx) is associated with a corresponding ground truth bar (GTxx). It can be seen that the bars of our predictions and ground truths were tightly matched for our participants, which indicates a high level of prediction accuracy.

DISCUSSION

Main findings

To the best of our knowledge, we are among the first to apply deep learning models to identify daily activities of people with type 2 diabetes and chronic comorbidities collected from wearable cameras. Across the 12 daily activities, our approach of combining Linear Discriminant Analysis, Deep Convolutional Neural Networks, Object Detection algorithms, and Weighted ensemble network achieved the best Top-1 and Top-5 scores of 81.9 and 86.8, demonstrating the feasibility of automatically classifying images of patient work collected from wearable cameras with high accuracy.

Comparison with other studies

Past studies on self-management and patient work are highly dependent on self-reports from participants,[10,11] or rely on external sensors such as GPS, smart phones, and auditory and motion sensors.[30–32] However, self-report methods are subject to its own set of problems (eg, recall bias),[10,11] and wearable sensor-based methods tend to perform pooly due to noise[33] or can only identify simple activities (eg, walking) that may not be useful to understand the full spectrum of health behaviors.[34–36] Studies have demonstrated that images captured by wearable camera are more reliable than the written diary to assist with the recalling of past events.[37–39]

The approaches, which involve non-wearable or wearable devices, to capture daily activities of self-management amongst people

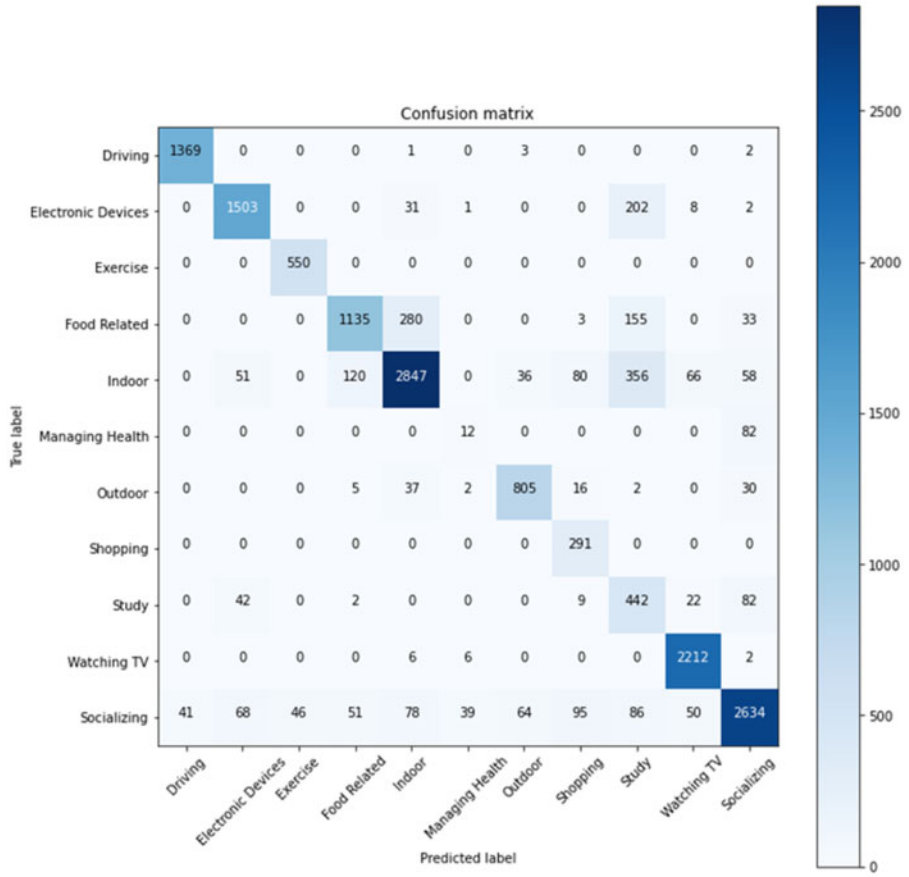

Figure 3. Confusion matrix of our weighted ensemble network for the 12 classes with rows as the predicted labels and columns as the actual labels.

Table 4. A comparison of 5 combinations of weights in the ensemble model (The best result is indicated in boldface)

	ResNet152	WideResNet101	ResNeXt50	ResNeXt100	3DResNet	YoloV3	Top-1/Top-5 accuracy
Set 1	0.1	0.1	0.1	0.1	0.4	0.2	70.3%/76.1%
Set 2	0.1	0.1	0.1	0.1	0.2	0.4	74.9%/81.3%
Set 3	0.09	0.09	0.09	0.09	0.3	0.34	76.2%/82.4%
Set 4	0.1	0.1	0.1	0.1	0.3	0.3	81.9%/85.8%
Set 5	0.1	0.1	0.1	0.1	0.25	0.35	**82.7%/87.1%**

with chronic conditions are briefly outlined here. For instance, some works have used non-wearable approaches to investigate self-management of Type 2 diabetes.[40–43] These works often involve open-ended interviews, survey-based assessment, questionnaires and booklets that require people with Type 2 diabetes to complete and reflect upon using their lived experiences.[44–46] In essence, these studies rely on self-reports which are also subject to biases previously mentioned.

Wearable devices attached to body parts including finger,[47] knee,[48] and heart,[49] have been used to capture human movement such as flexion and extension of fingers, passive flexion of knee, heart rate variability, in order to monitor disease progression for people with stroke, heart disease and Parkinson. In fact, these wearable devices can only capture simple body motions rather than providing diverse activity information for self-management analysis. Specific to diabetes, wearable devices, such as Gyroscope,[50] Infrared sensor,[51] Eversense Glucose Monitoring,[52] GPS and Wifi,[53] are also utilized to obtain diabetes related parameters including gait detection, change of temperature on the feet, glucose level monitoring, and quantification of sedentary behavior. Though these wearable sensors offer more accurate, objective and automatic measurements, they capture only small and narrow aspects of daily self-management for diabetes patients.

Strengths and limitations

The main strength of this study is that we have demonstrated, it is indeed possible to apply deep learning on images collected from wearable cameras to classify self-management routines amongst people with Type 2 diabetes and co-morbidities. Acknowledging the wide range of health behaviors captured in this study, as well as overcoming real-life camera image issues (such as blurring and low lighting), our method incorporates several networks into an ensemble network, demonstrating it is possible to study the spectrum of patient work by using wearable cameras and a deep-learning approach.

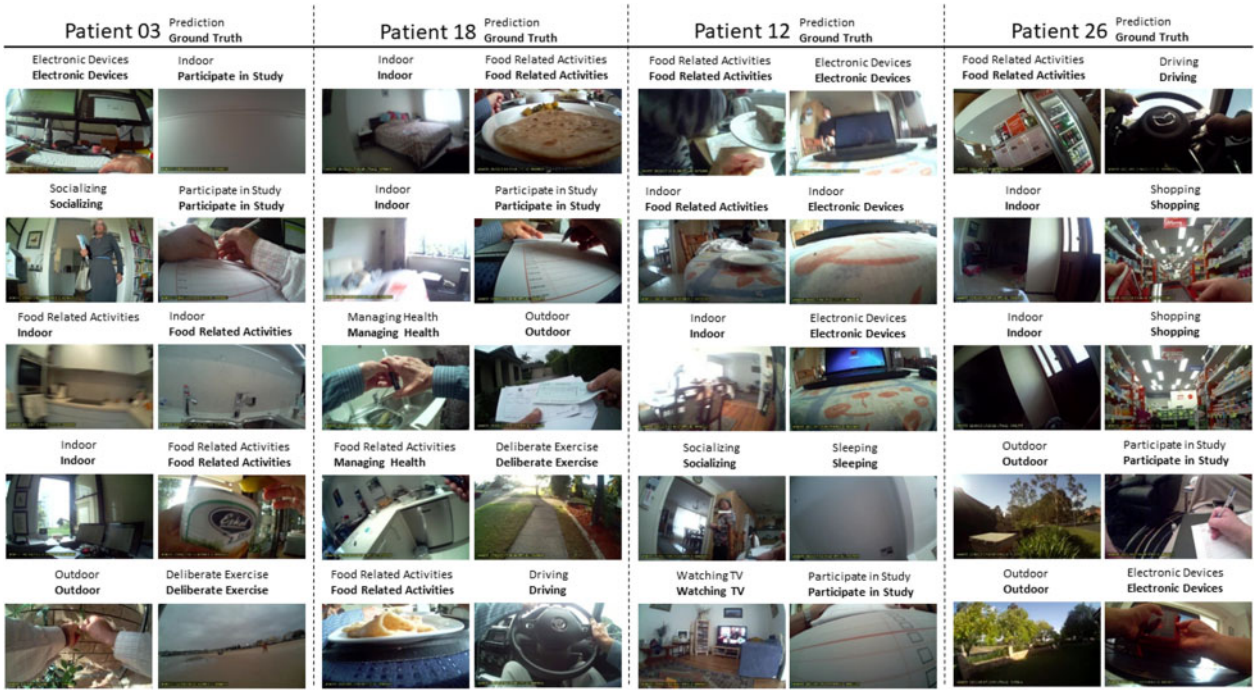

Figure 4. Example classification results of our weighted ensemble network for patients 03, 12, 18, 26. The class in bold corresponds to the ground truth labels.

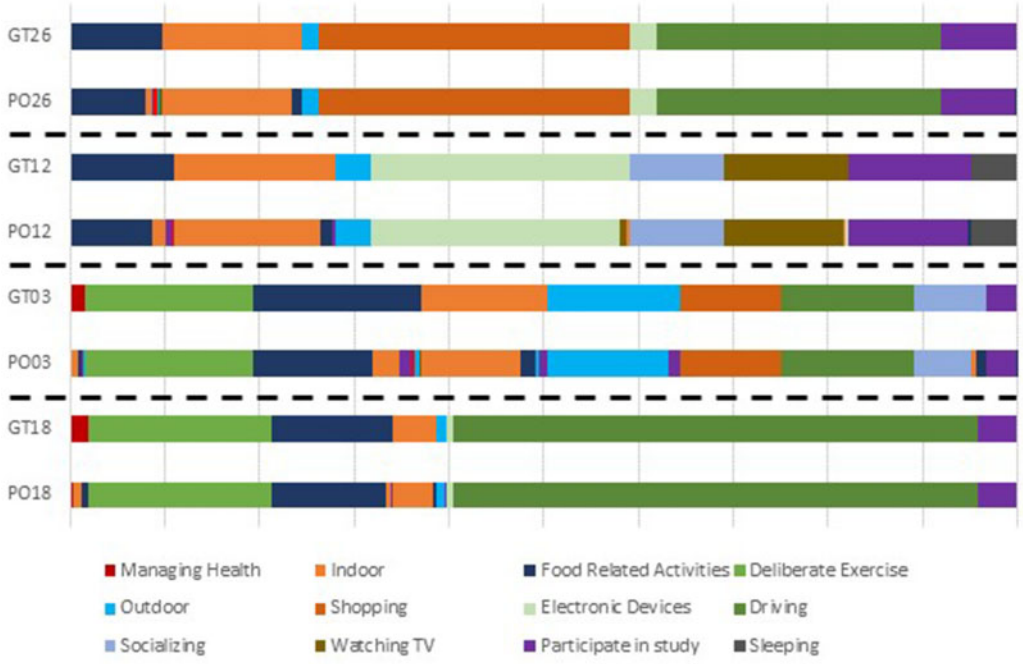

Figure 5. Complete illustrations of daily activities for patients 03, 12, 28, 26.

Another strength lies in combining image-based, video-based, and object detection networks into an ensemble-based model, where each method is used to target and classify a specific activity. As a result, such an ensemble network leverages the advantages of those individual networks, addressing the unique features of each activity. Our approach worked exceptionally well for specific activities, namely *Driving*, *Exercise*, and *Shopping*, which achieved Top-1 scores of 100%.

Privacy concerns are a crucial and important aspect of this study, which we address seriously and detailed strategies on ways to mitigate privacy risks are outlined in our study protocol.[5] Throughout the study, participants were reminded that they could withdraw consent at any time, and that participants should prioritize their privacy (eg, turn off camera) over the needs of the research. As such, it is possible that due to privacy considerations we did not capture the full range of participants' daily activities using the wearable camera approach.

To this date, we have not received any withdrawal or complaints from participants or third parties. Several approaches were used to protect participants' identities and to elicit bystander consent.[5] To recap, we asked participants to:

- Carry a wallet-sized card with description of the study, ethics approval, and contact details of the researchers shall they get approached by third parties about the study;
- Inform their household members, friends and acquaintances about the study, and seek permission from these parties prior to recording;
- Turn off video recording if participant (or any bystander) requests filming to be stopped;
- Take note of the time requests were made by bystanders that video recordings were to be deleted, and assure the enquirers that the related recordings will be removed; and
- Review video footage during post-study session and delete any recording they do not want others to see before sharing with researchers.

Furthermore, a potential technical solution to alleviate this issue is to apply face/object detection models, such as Yolo,[27] to detect the faces captured by the camera and de-identify them accordingly. Such de-identification is not expected to affect the performance of our model since our method does not rely on any face features for activity recognitions. Although this technical approach to de-identify participants' faces and protect their identity is available, we did not need to apply these techniques in this study.

Future research direction

From a methodological perspective, further research is required to investigate under what circumstances it is acceptable and appropriate to use wearable camera to study self-management behaviors. For example, we have found that wearable cameras have been particularly useful to understand what people do at home, especially when they are alone on their own. This approach has been important for understanding their daily routines, at what time and how much time they spent on each activity, and how their home environment may affect their abilities to conduct self-management. Such intrinsic knowledge is often ingrained in participants' daily routines that people experience difficulty in articulating certain aspects accurately. Wearable cameras, in this case, could offer a reliable alternative approach in studying self-management routines, especially for those who spend a lot of time alone at home, without causing additional burden to recall details, nor worrying about bystander involvement. From a deep learning perspective, classification accuracy can be further enhanced by collecting more training data, especially in the *Managing Health* class, to augment training samples. In addition, the overall classification may be improved by integrating metric learning, such as triplet loss.

It is noteworthy that our ensemble network can be considered a rather generic activity recognition model and has the potential to investigate daily activities of people with other chronic conditions. That is, the developed method is potentially a generalizable and robust approach to study self-management routines beyond type 2 diabetes. However, generalization of our deep learning approach beyond the diabetes context is outside the focus of this study, and future studies could utilize our approach to investigate non-diabetes context and examine whether our approach harnessing wearable cameras and deep learning to study other chronic conditions is replicable and generalizable.

CONCLUSION

In this work, we have successfully applied deep learning to classify images of daily living collected from wearable cameras worn by people with type 2 diabetes and chronic comorbidities with a high level of accuracy. Fine-turning these methods would facilitate an alternative approach to studying self-management, providing a more realistic and objective picture of the challenges involved. By offering a more robust approach to understand the 'work' involved in self-management, we hope to shed light into understanding areas where self-management may fail, and help people overcome the daily challenges they experience with managing their chronic conditions.

FUNDING

AYSL was supported by the New South Wales Health Early-Mid Career Fellowship, and her research was supported by the National Health and Medical Research Council grant APP1134919 (Centre of Research Excellence in Digital Health) and grant ID 1170937 (Centre of Research Excellence in Connected Health).

AUTHOR CONTRIBUTIONS

HX designed the deep learning models, analyzed research findings, and drafted the manuscript. HNP implemented ensemble network, conducted experiments, and drafted the manuscript. KY labeled the datasets, revised the manuscript. SB revised the manuscript. JJ labeled the datasets. AYSL conceptualized the study, supervised the work, provided domain knowledge, and revised the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY STATEMENT

Data cannot be shared for ethical/privacy reasons.

The data underlying this article cannot be shared publicly due to the sensitive nature of the data and for the privacy of individuals that participated in the study. The data will be shared on reasonable request to the senior author, subject to further ethics approval.

REFERENCES

1. World Health Organization. *Non-Comunicable Disease Factsheet*. Geneva, Switzerland: WHO; 2015. http://www. who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases.
2. Gardetto NJ. Self-management in heart failure: where have we been and where should we go? *J Multidiscip Healthc* 2011; 4: 39–51.
3. McManus RJ, Mant J, Franssen M, *et al.*; TASMINH4 investigators. Efficacy of self-monitored blood pressure, with or without telemonitoring, for titration of antihypertensive medication (TASMINH4): an unmasked randomised controlled trial. *Lancet* 2018; 391 (10124): 949–59.
4. Savard LA, Thompson DR, Clark AM. A meta-review of evidence on heart failure disese management progrmas: the challenges of describing and synthesizing evidence on complex interventions. *Trials* 2011; 12: 194.
5. Yin K, Harms T, Ho K, *et al.* Patient work from a context and time use perspective: a mixed-methods study protocol. *BMJ Open* 2018; 8 (12): e002163.

6. Yin K, Jung J, Coiera E, *et al*. How patient work changes over time for people with multimorbid type 2 diabetes: qualitative study. *J Med Internet Res* 2021; 23 (7): e25992.
7. Yin K, Jung J, Coiera E, *et al*. Patient work and their contexts: scoping review. *J Med Internet Res* 2020; 22 (6): e16656.
8. Valdez RS, Holden RJ, Novak LL, Veinot TC. Transforming consumer health informatics through a patient work framework: connecting patients to context. *J Am Med Inform Assoc* 2015; 22 (1): 2–10.
9. Valdez RS, Holden RJ, Novak LL, Veinot TC. Technical infrastructure implications of the patient work framework. *J Am Med Inform Assoc* 2015; 22 (e1): e213–15.
10. Gemming L, Utter J, Ni Mhurchu C. Image-assisted dietary assessment: a systematic review of the evidence. *J Acad Nutr Diet* 2015; 115 (1): 64–77.
11. Welk G. *Physical Activity Assessments for Health-Related Research*. Champaign, IL: Human Kinetics; 2002.
12. Free C, Phillips G, Galli L, *et al*. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. *PLoS Med* 2013; 10 (1): e1001362.
13. Gurrin C, Smeaton AF, Doherty AR. Lifelogging: personal big data. *Found Trends Inf Retrieval* 2014; 8 (1): 1–125.
14. Doherty AR, Hodges SE, King AC, *et al*. Wearable cameras in health: the state of the art and future possibilities. *Am J Prev Med* 2013; 44 (3): 320–3.
15. Gemming L, Rush E, Maddison R, *et al*. Wearable cameras can reduce dietary under-reporting: doubly labelled water validation of a camera-assisted 24 h recall. *Br J Nutr* 2015; 113 (2): 284–91.
16. Leask CF, Harvey JA, Skelton DA, Chastin SF. Exploring the context of sedentary behaviour in older adults (what, where, why, when and with whom). *Eur Rev Aging Phys Act* 2015; 12: 4.
17. Kelly P, Doherty A, Berry E, Hodges S, Batterham AM, Foster C. Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. *Int J Behav Nutr Phys Act* 2011; 8 (1): 44.
18. Piasek P, Irving K, Smeaton AF. Case study in SenseCam use as an intervention technology for early-stage dementia. *Int J Comput Healthc* 2012; 1 (4): 304.
19. Zhang H, Li L, Jia W, Fernstrom JD, Sclabassi RJ, Sun M. Recognizing physical activity from ego-motion of a camera. In: *International Conference of IEEE Engineering in Medicine and Biology*. 2010: 5569–72.
20. Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: a large-scale hierarchical image database. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2009: 248–55; Miami Beach, FL.
21. Sutton E. Histograms and the Zone System. *Illustrated Photography* 2015. https://web.archive.org/web/20150223082314/http://www.illustratedphotography.net/basic-photography/zone-system-histograms. Accessed March 17, 2022.
22. Friedman JH. Regularized discriminant analysis. *J Am Stat Assoc* 1989; 84 (405): 165–75.
23. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 770–8; Las Vegas, NV.
24. Zagoruyko S, Nikos K. Wide residual networks. In: *British Machine Vision Conference* 2016: 87.1–87.12; York, UK.
25. Xie S, Girshick R, Dollar P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 5987–95; Honolulu, HI.
26. Hirokatsu K, Wakamiya T, Hara K, Satoh Y. Would mega-scale datasets further enhance spatiotemporal 3d cnns? *arXiv preprint arXiv:2004.04968* 2020.
27. Redmon J, Farhadi A. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* 2018.
28. Fathi A, Hodgins JK, Rehg JM. Social interactions: a first-person perspective. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2012: 1–8; Providence, RI.
29. Aghaei M, Dimiccolo M, Radeva P. With whom do I interact? Detecting social interactions in egocentric photo-streams. In: *International Conference on Pattern Recognition*. 2016: 2959–2964; Cancun, Mexico.
30. Biagioni J, Krumm J. Days of our lives: assessing day similarity from location traces. In: *User Modeling*. 2013.
31. Eagle N, Pentland A. Reality mining: sensing complex social systems. *Personal Ubiquitous Computing* 2006; 10 (4): 255–68.
32. Clarkson BP. *Life Patterns: Structure from Wearable Sensors* [PhD thesis]. MIT; 2005.
33. Ross R, Kelleher J. A comparative study of the effect of sensor noise on activity recognition models. In: *Evolving Ambient Intelligence*. 2013: 151–162; Dublin, Ireland.
34. Shoaib M, Scholten J, Havinga PJM. Towards physical activity recognition using smartphone sensors. In: *10th IEEE International Conference on Ubiquitous Intelligence and Computing*. 2013: 80–87; Sorrento Peninsula, Italy.
35. Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. In: *International Workshop on Knowledge Discovery from Sensor Data*. 2010: 74–82; Washington, DC.
36. Janidarmian M, Fekr AR, Radecka K, Zilic Z. A comprehensive analysis on wearable acceleration sensors in human activity recognition. *Sensors* 2017; 17 (3): 529.
37. Browne G, Berry E, Kapur N, *et al*. SenseCam improves memory for recent events and quality of life in a patient with memory retrieval difficulties. *Memory* 2011; 19 (7): 713–22.
38. Hodges S, Williams L, Berry E, *et al*. SenseCam: a retrospective memory aid. In: *International Conference on Ubiquitous Computing*. 2006: 177–193; Orange County, CA.
39. Allen AL. Dredging up the past: Lifelogging, memory, and surveillance. *U Chi L Rev* 2008; 75: 47.
40. Huang ES, Gorawara-Bhat R, Chin M. Self-reported goals of older patients with type 2 diabetes mellitus. *J AM Geriatr Soc* 2005; 53 (2): 306–11.
41. Wabe NT, Angamo MT, Hussein S. Medication adherence in diabetes mellitus and self-management practices among type-2 diabetes in Ethiopia. *N Am J Med Sci* 2011; 3 (9): 418–23.
42. Yusuff KB, Obe O, Joseph BY. Adherence to anti-diabetic drug therapy and self management practices among type-2 diabetics in Nigeria. *Pharm World Sci* 2008; 30 (6): 876–83.
43. Weller SC, Baer R, Nash A, Perez N. Discovering successful strategies for diabetic self-management: a qualitative comparative study. *BMJ Open Diab Res Care* 2017; 5 (1): e000349.
44. Alhaiti AH, Senitan M, Dator WLT, *et al*. Adherence of type 2 diabetic patients to self-care activity: tertiary care setting in Saudi Arabia. *J Diabetes Res* 2020; 2020: 4817637.
45. Adhikari Baral I, Baral S. Self-care management among patients with type 2 diabetes mellitus in Tanahun, Nepal. *Arch Community Med Public Health* 2021; 7 (1): 03–042.
46. van Smoorenburg AN, Hertroijs DFL, Dekkers T, Elissen AMJ, Melles M. Patients' perspective on self-management: type 2 diabetes in daily life. *BMC Health Serv Res* 2019; 19 (1): 605.
47. Henderson J, Condell J, Connolly J, Kelly D, Curran K. Review of wearable sensor-based health monitoring glove devices for rheumatoid arthritis. *Sensors* 2021; 21 (5): 1576.
48. Ortiz M, Juan R, Val SL. Reliability and concurrent validity of the goniometer-pro app vs. a univeral goniometer in determining passive flexion of knee. *Int J Comput Appl* 2017; 173: 30–4.
49. Sayem ASM, Teay SH, Shahariar H, Fink PL, Albarbar A. Review on smart electro-clothing systems (SeCSs). *Sensors* 2020; 20 (3): 587.
50. Grewal GS, Bharara M, Menzies R, Talal TK, Armstrong D, Najafi B. Diabetic peripheral neuropathy and gait: does footwear modify this association? *J Diabetes Sci Technol* 2013; 7 (5): 1138–46.
51. Fraiwan L, AlKhodari M, Ninan J, Mustafa B, Saleh A, Ghazal M. Diabetic foot ulcer mobile detection system using smart phone thermal camera: a feasibility study. *Biomed Eng Online* 2017; 16 (1): 117.
52. Kumar HS. Wearable technology in combination with diabetes. *Int J Res Eng Sci Manag* 2019; 2: 1–4.
53. McLean A, Osgood N, Newstead-Angel J, Stanley K, Knowles D, *et al*. Building research capacity: results of a feasibility study using a novel mhealth epidemiological data collection system within a gestational diabetes population. *Stud Health Technol Inform* 2017; 234: 228–32.