JMIR Biomedical Engineering

Engineering for health technologies, medical devices, and innovative medical treatments and procedures
Volume 6 (2021), Issue 1 ISSN 2561-3278 Editor in Chief: Syed A. A. Rizvi, MD, PhD, MBA, MPH, BSN

Contents

Review

Monitoring of Sitting Postures With Sensor Networks in Controlled and Free-living Environments: Systematic Review (e21105)
Arpita Kappattanavar, Nico Steckhan, Jan Sachs, Harry Freitas da Cruz, Erwin Böttinger, Bert Arnrich. 2

Original Papers

Physical Activity Evaluation Using a Voice Recognition App: Development and Validation Study (e19088)
Hideyuki Namba. 16

Subspace Clustering of Physiological Data From Acute Traumatic Brain Injury Patients: Retrospective Analysis Based on the PROTECT III Trial (e24698)
Sina Ehsani, Chandan Reddy, Brandon Foreman, Jonathan Ratcliff, Vignesh Subbian. 27

Point-of-Care Quantification of Serum Alpha-Fetoprotein for Screening Birth Defects in Resource-Limited Settings: Proof-of-Concept Study (e23527)
Balaji Srinivasan, Julia Finkelstein, David Erickson, Saurabh Mehta. 35

Review

Monitoring of Sitting Postures With Sensor Networks in Controlled and Free-living Environments: Systematic Review

Arpita Mallikarjuna Kappattanavar[1], MSc; Nico Steckhan[1], MSc, PhD; Jan Philipp Sachs[1], MSc, MD; Harry Freitas da Cruz[1], MBA; Erwin Böttinger[1,2], MD, PhD; Bert Arnrich[1], PhD

[1]Hasso-Plattner-Institut, University of Potsdam, Potsdam, Germany

[2]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Corresponding Author:
Arpita Mallikarjuna Kappattanavar, MSc
Hasso-Plattner-Institut
University of Potsdam
Rudolf-Breitscheid-Straße 187
Potsdam, 14482
Germany
Phone: 49 33155094895
Email: Arpita.Kappattanavar@hpi.de

Abstract

Background: A majority of employees in the industrial world spend most of their working time in a seated position. Monitoring sitting postures can provide insights into the underlying causes of occupational discomforts such as low back pain.

Objective: This study focuses on the technologies and algorithms used to classify sitting postures on a chair with respect to spine and limb movements, using sensors and wearables such as inertial measurement units, pressure or piezoresistive sensors, accelerometers or gyroscopes, combined with machine learning approaches.

Methods: A total of three electronic literature databases were surveyed to identify studies classifying sitting postures in adults. Quality appraisal was performed to extract critical details and assess biases in the shortlisted papers.

Results: A total of 14 papers were shortlisted from 952 papers obtained after a systematic search. The majority of the studies used pressure sensors to measure sitting postures, whereas neural networks were the most frequently used approaches for classification tasks in this context. Only 2 studies were performed in a free-living environment. Most studies presented ethical and methodological shortcomings. Moreover, the findings indicate that the strategic placement of sensors can lead to better performance and lower costs.

Conclusions: The included studies differed in various aspects of design and analysis. The majority of studies were rated as medium quality according to our assessment. Our study suggests that future work for posture classification can benefit from using inertial measurement unit sensors, since they make it possible to differentiate among spine movements and similar postures, considering transitional movements between postures, and using three-dimensional cameras to annotate the data for ground truth. Finally, comparing such studies is challenging, as there are no standard definitions of sitting postures that could be used for classification. In addition, this study identifies five basic sitting postures along with different combinations of limb and spine movements to help guide future research efforts.

(JMIR Biomed Eng 2021;6(1):e21105) doi:10.2196/21105

KEYWORDS

classification; algorithms; sitting position; spine; technology; machine learning; back pain; movement; extremities

Introduction

Background

The proportion of people sitting for long hours during work and daily life has increased in recent years. Approximately 75% of employees in call centers, software companies, and other industrial jobs spend an average of 90% of their workday sitting on a chair [1,2]. Many individuals who sit for long hours in the same posture, or *bad* posture, experience musculoskeletal discomfort and pain at the ischiocrural muscle region [3]. Prolonged sitting behavior and spine-straining sitting postures

have been reported to act as negative factors, increasing the probability of developing low back pain (LBP) [1,2,4-6].

LBP has been identified as a significant cause of sick leaves and disability, leading to impairment in daily and occupational activities, reflecting a significant economic burden on the society [7-9]. The majority (90%) of LBP cases are nonspecific [8,10-12]. By definition, nonspecific LBP cases have an unknown origin, where mechanical factors and multifactorial etiology are suspected. There is still a gap in understanding whether mechanical factors are associated with nonspecific LBP, as it has not been verified in research studies [8]. Continuous monitoring of spine movements and daily activities would help understand the link between the various mechanical and psychosocial factors leading to LBP and differentiate them [12].

To implement appropriate intervention and prevention programs for LBP, especially within an office environment, identifying risk factors such as stress at work and sitting postures is of high importance according to the Bulletin of the World Health Organization (WHO) [9] and from the studies conducted by Bontrup et al [1] and Søndergaard et al [3]. Therefore, this systematic literature review focusses on the classification of sitting postures.

In traditional methods, sitting postures were analyzed by observing the seated subjects and self-reported answers to questionnaires [13]. However, these methods are biased and subjective, and vary for each doctor and patient. Therefore, the data were unreliable. With advancements in micro-electro-mechanical systems and nano-electro-mechanical systems, different types of miniaturized sensor technologies are readily available in the market. They can assess and classify sitting postures more objectively and accurately. In the last decade, studies have used miniaturized pressure sensors made from air bladders, piezoelectric materials, fibers coated with yarn materials, force sensors, and force-sensing resistors in the form of cushions, sensor array sheets, and mats or just as individual sensors to provide the necessary signals to classify sitting postures [1,13-24]. Such classifications should also preferably include limb movements, as these are suspected to be associated with musculoskeletal discomfort and pain [25-27].

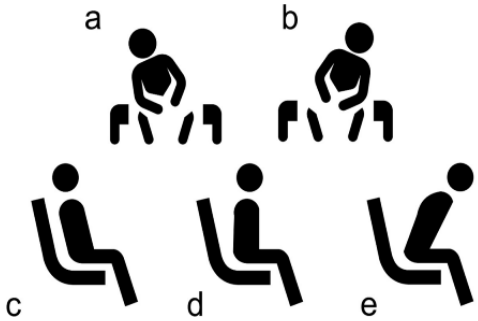
Objective

This study has been conducted to understand the state-of-the-art technologies for classifying sitting postures on a chair along with limb and spine movements. To achieve this goal, we (1) use a systematic database search approach using the Population or Problem, Intervention or Exposure, Comparison, and Outcome (PICO) scheme; (2) carry out a quality appraisal of the included papers; (3) summarize the algorithms and the number of postures classified; (4) investigate the study design and the type of environment in which these studies were conducted; and (5) identify the challenges in classifying sitting postures and critically assess the technological solutions employed.

The majority of the studies in this review used pressure sensors to measure sitting postures, whereas neural networks (NNs) were the most frequently used approaches for classification tasks in this context. In total, 5 main postures as shown in Figure 1, were presented in all studies along with different combinations of limb and spine movements.

The organization of this paper is as follows: In the *Methods* section, we present the search approach, the inclusion and exclusion criteria applied to shortlist the papers, checklists for bias assessment, and data extracted from the papers. In *Results* section, we present the summary of all the shortlisted articles and outline the details of the extracted data. In the *Discussion* section, we investigate and discuss the findings, and in the *Conclusions* section, we provide recommendations and an outlook on future work. Finally, in the *Limitations* section, we discuss the limitations of this study.

Figure 1. The 5 most common sitting postures: (a) lean right, (b) lean left, (c) lean backward, (d) upright sitting, and (e) lean forward.



Methods

Search Approach

A systematic search was conducted on PubMed, IEEE Xplore, and Web of Science databases until June 2019. The literature search strategy framework in systematic reviews is typically based on the PICO scheme [28]. The search in this study was partially based on PICO. In the keywords' formation, we included *sitting* for population or problem, the *tools or technology* was mentioned for intervention, and *algorithms* were mentioned for the outcome. As we did not have a comparison, it was excluded from PICO. The search focused on papers with the main terms and specific terms, as indicated in Table 1. Search engines' specific terms varied slightly (Multimedia Appendix 1). Additional papers were identified by manually searching and screening the reference lists of other papers to identify papers that have been overlooked by the electronic search. Retrieved papers were imported to Mendeley Desktop.

Table 1. Terms for literature database search. The specific terms of the 3 main categories have been joined by an AND condition.

Main terms	Specific search terms
Sitting	(sitting OR seating OR seated) AND (posture* OR position OR behaviour)
Sensor	sensor* OR "inertial measurement unit" OR IMU OR wearable OR pressure OR piezoresistive OR accelerometer OR gyroscope
Algorithm	"machine learning" OR "neural network*" OR algorithm* OR *supervised OR classif* OR detection OR identification OR recognition

Study Selection Method

The search terms were assessed by two authors independently and iteratively and then finalized. The final search terms were used to obtain papers from the aforementioned databases. The titles and abstracts of the obtained papers were carefully read and analyzed before shortlisting them based on the inclusion and exclusion criteria.

Papers that met the following criteria have been included:

1. More than 3 sitting postures were classified
2. Journal or conference papers were published in English language
3. The involved population was sitting on a chair
4. The study involved adult population (older than 18 years)

Papers have been excluded based on the following criteria, if:

1. Limb movements while sitting were not considered. Studies have revealed that leg movements affect musculoskeletal discomfort and pain [25-27].
2. The involved population was sitting in a wheelchair or driving a vehicle. The postures for a wheelchair subject were less dependent on the leg movements. Moreover, driving postures differ from those of sitting postures in occupational settings.
3. The methodology and classification accuracy of each posture was not mentioned or reported, as the classification accuracy provided the proof of the methodology for replication.
4. Duplicates were avoided if the same author mentioned the same methodology in a journal and a conference article, and then the conference article was excluded.
5. The same methodology is mentioned in 2 papers by the same authors with little variation. A paper that provided a higher level of details was included in this study.
6. A paper is not related to sitting postures.
7. Sensors were implanted inside the body, as our study focused on noninvasive methods.

Differences in the inclusion of specific papers were resolved by consulting with other authors of this study.

Study Quality Assessment

Quality appraisal checklists were developed to extract key details and identify the risk of bias in each study. This checklist was prepared based on consultation with other authors and using the studies by Papi et al [12] and Hagströmer et al [29] as a reference to include relevant points. The prepared checklist has questions related to 3 categories, that is, study description, study design, and robustness. Table 2 presents the study quality assessment checklist questionnaires based on the three categories to assess the risk of bias.

The customized checklist is provided in Table 3. The table has been further numbered as 0, 1, or 2 for each selected paper to rate it as no detail, limited detail, and good detail, respectively. The total score is based on the sum of those checkpoint scores (0-26). These papers were rated as low (low<10), moderate (10<moderate<18), or high (19<high<26) quality based on the total score of the paper. The score for each paper was based on the discussion with other authors.

Data Extraction

This study was conducted to investigate the technology and algorithms used to classify the sitting postures in different settings. Therefore, we extracted details concerning the technology, study design, classification algorithm, and algorithm performance from the shortlisted papers, as presented in Textbox 1. In addition to quality appraisal, these items will guide the remainder of this paper.

Table 2. The checklist questions to assess the risk of bias.

Categories	Checklist questions
Study description	• Q1. Are the research objectives or aims stated? • Q2. Is the study clearly described? • Q3. Were the main findings of the study stated? • Q4. Are the limitations of the study clearly described?
Study design	• Q5. Are appropriate subject information and anthropometric details provided? • Q6. Were the number of subjects studied justified? • Q7. Was prominence of leg crossing considered? • Q8. Were the eligibility criteria mentioned? • Q9. Were there ethics committee approval and written consent mentioned in the papers? • Q10. Was the justification for the sensor setup and location given?
Robustness	• Q11. Were measures of reliability or accuracy of the algorithm reported? • Q12. Were the classifications cross-validated? • Q13. Is the system robust in the wild (controlled or free-living environments)?

Table 3. Quality bias assessment table to rate the quality of the shortlisted papers based on the three question categories described previously. Q represents the checklist question number.

Study	Study description				Study design						Robustness			Total	Quality
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13		
Ma et al [30]	2[a]	1[b]	1	2	0[c]	0	1	1	0	0	1	0	1	10	Medium
Zemp et al [14]	2	2	2	2	2	2	1	1	2	0	1	1	1	19	High
Xu et al [13]	2	1	2	1	1	0	1	0	0	1	2	0	1	12	Medium
Martins et al [15]	2	1	2	2	2	0	2	0	0	2	2	1	1	17	Medium
Zemp et al [16]	2	2	2	2	2	0	1	0	2	0	2	1	2	18	Medium
Kamiya et al [17]	2	2	2	0	1	0	2	0	0	0	2	1	1	13	Medium
Liu et al [18]	2	2	2	0	0	0	1	0	0	0	1	2	1	11	Medium
Pereira et al [19]	2	2	1	0	2	0	2	0	0	2	2	0	1	14	Medium
Zhu et al [20]	2	1	2	0	1	0	2	0	0	1	1	0	1	11	Medium
Bontrup et al [1]	2	1	2	2	2	0	1	1	2	0	2	1	2	18	Medium
Mutlu et al [21]	2	1	2	2	1	0	2	0	0	2	2	1	1	16	Medium
Huang et al [22]	2	1	2	1	0	0	1	0	0	0	2	0	1	10	Medium
Wang et al [23]	2	2	2	1	2	0	1	0	0	1	2	0	1	14	Medium
Noh et al [24]	2	0	2	0	0	0	0	0	0	0	2	0	1	7	Low

[a]2: good detail.

[b]1: limited detail.

[c]0: no detail.

Textbox 1. Summary of the data extracted from each of the shortlisted papers.

Technology

- Sensor type
- The number of sensors
- Sensor location

Study design

- The environment in which these studies were performed
- The number of subjects recruited
- Study protocol

Classification algorithm

- Algorithms used
- The type of features extracted
- Number of postures classified

Algorithm performance

- Performance metrics
- Evaluation setup

Results

Shortlisted Papers

The shortlisting of the papers was based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart presented in Figure 2 [31]. The search terms given in Table 1 were used with minor modifications to identify and retrieve 1359 potentially relevant papers: 610 from Web of Science, 307 from IEEE Xplore, and 442 from PubMed. The elimination of duplicates from these retrieved papers resulted in 949 papers. A total of 3 additional papers were included from the reference search and other sources. After screening them by reading the titles and abstracts, 105 papers were shortlisted. The excluded papers were related to air embolism, human activity recognition, gait analysis, hypertension, and other topics unrelated to sitting postures. Only 14 of the 105 shortlisted papers were selected after reading the complete papers, based on the inclusion criteria. The reasons for excluding the remaining 91 papers are provided in Figure 2. In total, 5 of the selected papers are from journal publications. A summary of the shortlisted papers is presented in Tables 4 and 5.

Figure 2. The literature search strategy using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

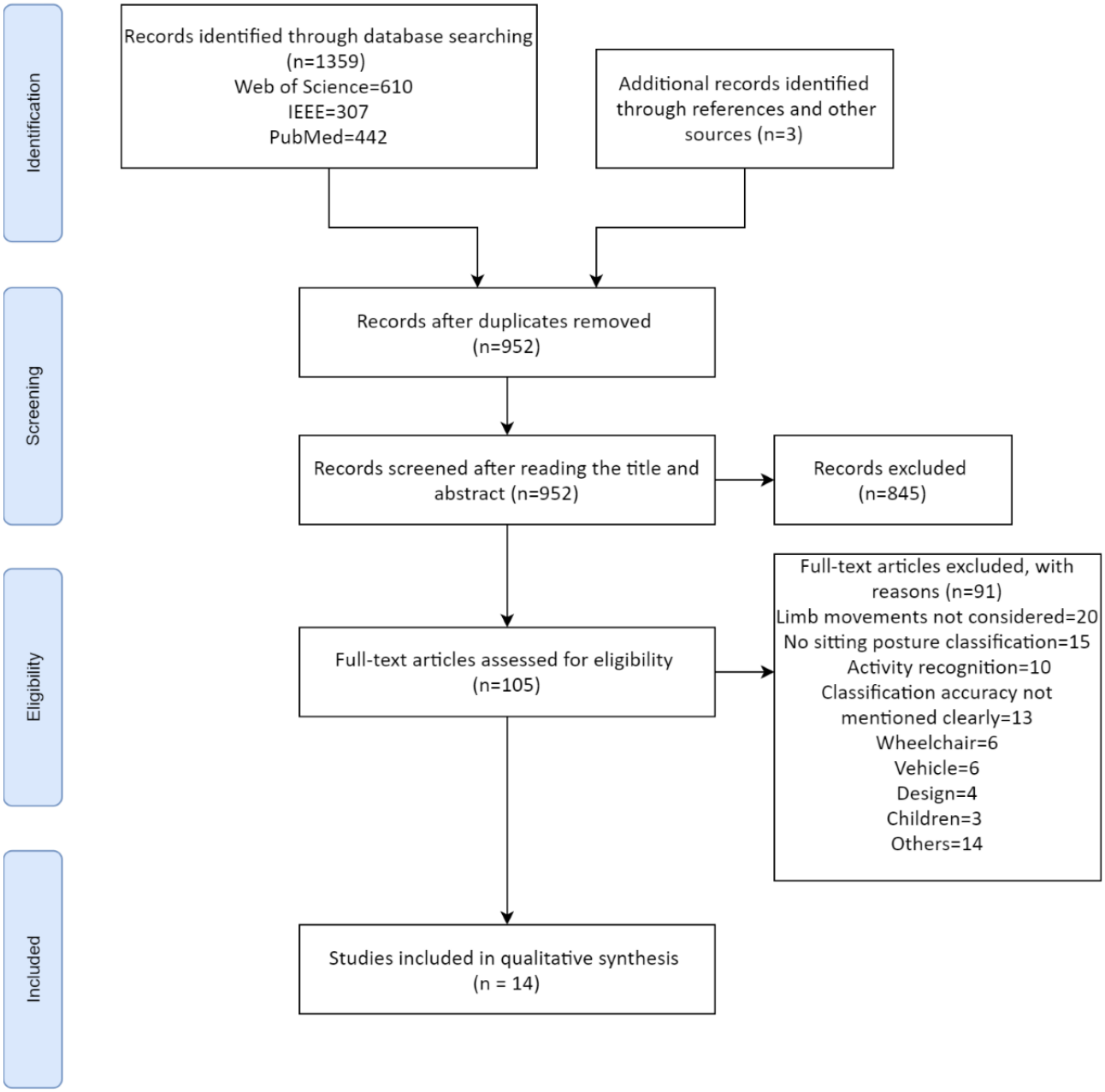


Table 4. Summary of the reviewed papers.

Study	Number and type of postures	Classification accuracy
Ma et al [30]	5 types: US[a], LF[b], LB[c], LR[d], and sitting cross-legged	SVM[e]: 95.33% accuracy; K-means clustering: 89.35% accuracy
Zemp et al [14]	7 types: US, LF, LB, LL[f], LR, the left leg crossed over the right, and the right leg crossed over the left	Multimodal regression: 90.4% accuracy; NN[g]: 90.4% accuracy; RF[h]: 90.9% accuracy; combination of boosting, NN, and RF: 90.8% accuracy
Xu et al [13]	7 types: US, LF, LB, LL, LR, right foot over left, and left foot over right	Dynamic time warping: 85.9% accuracy
Martins et al [15]	Experiment A: 11 types: US, LF, LB, LL, LR, LB with no lumbar support, the right leg crossed, the right leg crossed with LL, the left leg crossed, the left leg crossed with LR, and slouching; experiment B: 8 types: US, LF, LB, LL, LR, LB with no lumbar support, the right leg crossed, and the left leg crossed	Experiment A: artificial NN: 70% accuracy; experiment B: thresholding and artificial NN: 93.4% accuracy
Zemp et al [16]	7 types: US, LF, LB, LL, LR, crossed legs right over left, and crossed legs left over right	RF: 82.7% accuracy
Kamiya et al [17]	9 types: US, LF, LB, LL, LR, the right leg crossed, LR with the right leg crossed, the left leg crossed, and LL with the left leg crossed	SVM: 98.9% accuracy known subject; SVM: 93.9% accuracy unknown subject
Liu et al [18]	8 types: US, LF, LB, LL, LR, crossed legs right over left, crossed legs left over right and slouching	Convolutional NN: 98% accuracy; back propagation NN: 92.8% accuracy
Pereira et al [19]	12 types: US, LF, LB, LL, LR, LB with no lumbar support, the right leg crossed, the right leg crossed with LL, the left leg crossed, the left leg crossed with LR, left leg over right, and the right leg over left	Artificial NN: 80.9% accuracy
Zhu et al [20]	10 types: US, LF, LB, LL, LR, the right leg crossed, the left leg crossed, LL with the right leg crossed, LR with the left leg crossed, and slouching	k-nearest neighbor: 81% accuracy; principal component analysis: 86% accuracy; linear discriminant analysis: 81% accuracy; sliced inverse regression: 86% accuracy; NN: 80% accuracy
Bontrup et al [1]	7 types: US, LF, LB, LL, LR, crossed legs right over left, and crossed legs left over right	RF: 90% accuracy
Mutlu et al [21]	10 types: US, LF, LB, LL, LR, the left leg crossed with LR, the right leg crossed, slouching, the left leg crossed, the right leg crossed with LL	Tekscan: 31 sensor SimpleLogistic: 87% accuracy; Prototype Sensor System: 19 sensors SimpleLogistic: 78% accuracy
Huang et al [22]	8 types: US, LF, LB, LL, LR, slumped sitting, the right leg crossed, and the left leg crossed	Artificial NN: 92.2% accuracy
Wang et al [23]	6 types: US, LF, LB, the left leg crossed, the right leg crossed, and astride sitting	Decision tree: 99% accuracy
Noh et al [24]	9 types: US, LF, LB, LL, LR, left leg trembling, right leg trembling, left leg twisted, and the right leg twisted	Triangle center: 98% accuracy

[a]US: upright sitting.

[b]LF: lean forward.

[c]LB: lean backward.

[d]LR: lean right.

[e]SVM: support vector machine.

[f]LL: lean left.

[g]NN: neural network.

[h]RF: random forest.

Table 5. Summary of the reviewed papers.

Study	Population	Duration of the study	Number and type of sensor(s) and location
Ma et al [30]	6 subjects	Each posture was held for 5 min, but data were collected after 1 to 2 min.	One triaxial accelerometer and cervical spine
Zemp et al [14]	41 subjects	Each posture was held for 5 s.	17 pressure sensors: 10 pressure sensors were fixed within the seat pan, 4 were fixed on the backrest, 3 were fixed on each armrest, and 1 accelerometer sensor at the rear of the backrest
Xu et al [13]	25 subjects	—[a]	256 pressure sensors in a cushion placed on the seat of the chair
Martins et al [15]	Experiment A: 30 subjects; experiment B: 30 subjects	Experiment A: each subject held each posture for 20 s; experiment B: each subject held each posture for 15 s.	8 pressure sensors or cells: 4 in the seat pad and 4 in the backrest
Zemp et al [16]	20 subjects	Free-living environment recording was for 330 min.	64 pressure sensors mat placed on the seat pan
Kamiya et al [17]	10 subjects	Each posture was maintained for 2 to 3 s.	64 pressure sensors sheet placed on the seat of the chair
Liu et al [18]	25 subjects	—	1024 pressure sensors array placed on the chair
Pereira et al [19]	72 subjects	Each subject had each posture for 20 s.	8 pressure sensor (air bladder): 4 in the seat pad and 4 in the backrest
Zhu et al [20]	50 subjects	—	Two 2016 pressure sensor sheets mounted on the seat pan and the backrest of the chair
Bontrup et al [1]	64 call center employees	Data were collected from each participant for almost 6.2 (SD 1.5) h	196 pressure sensors mat fixed to the seat pan of an office chair
Mutlu et al [21]	Tekscan: 52 subjects; Prototype Sensor System: 20 subjects	—	Tekscan: 2016 pressure sensor mat each placed on the backrest and the seat; Prototype Sensor System: 19 pressure sensors optimally placed on the backrest and on the seat of the chair.
Huang et al [22]	—	Each posture maintained for 5 s.	2288 pressure sensor (a piezoresistive sensor) on the seat
Wang et al [23]	5 subjects	Data were collected with each posture for 30 s, followed by 10 s rest.	8 pressure sensor (capacitive proximity sensor): 4 sensors on the seat and 4 sensors on the backrest
Noh et al [24]	10 subjects	Data were collected for 10 min.	8 pressure sensors on the seating area

[a]Data not available.

Study Quality Assessment

The quality assessment of papers was performed based on the evaluations of the questionnaires in Table 2. All papers had a low bias in their aims (Q1), in their reliability in reporting the accuracy of the algorithms (Q11), and when stating their findings (Q3). Most papers were rated low regarding the justification of the number of subjects enrolled (Q6) and the eligibility criteria (Q8) used to recruit them into the studies. Other factors for the lower rating were that most papers did not mention ethics approval and written informed consent from subjects (Q9), the justification of sensor positioning (Q10), the use of cross-validation to evaluate the algorithms (Q12), and the study's limitations (Q4). Due to the aforementioned factors, most papers were rated as medium quality, and only 1 study [14] was as rated high quality. Therefore, for all the included papers' total median assessment score was rated as 13.5 (on a scale of 0 to 26).

Technology

This section describes and investigates the types of sensors used, their quantity, and their placement. All studies, except for 1 study out of the 14 shortlisted studies [30], have used pressure sensors to classify sitting postures. These pressure sensors were used in the form of smart cushions, pressure mats, pressure sensor sheets, or as individual pressure sensors. In the experiment conducted by Zemp et al [14], an additional triaxial accelerometer sensor was used that was placed in the backrest of the seat to access the global chair movement and angle of the backrest. The experiment conducted by Ma et al [30] is the only study in which a triaxial accelerometer sensor was used to measure the sitting posture. An accelerometer was placed on the cervical spine to study the seated posture.

In previous studies [1,13,16-18,20-22], pressure sensors were used in the form of an array of sensors and placed on the chair's seat pan. In 2 papers, the sensor arrays were placed on the backrest of the chair [20,21]. The sizes of these sensor arrays varied between 64 and 2288 sensors.

In some papers [14,15,19,21,23,24], sensors were sparsely placed, varying from 7 to 17. In previous studies [15,19,23], the sensors were placed such that the ischial tuberosity, the thigh region, the lumbar region of the spine, and the scapula had better contact with the sensors. The variations in the pressure distributions in these regions were distinct for different postures.

In 1 of the papers, pressure sensors were additionally placed in the armrest [14]. In the study by Mutlu et al [21], ideal positions for the sensors were identified using approximation algorithms. These placements were based on the classes and features extracted from the pressure sensors. A total of 7 studies used commercial sensors in their experiments [1,14-16,20,21,24]. These commercial pressure sensors were either from Tekscan, Interlink, Sensomative, or Honeywell. The rest of the authors had designed custom-made pressure sensors.

Study Design

This section describes the environment of the study, the duration of the study's recording, and the number of subjects recruited. Most of the studies were conducted in a controlled environment, and the subjects were asked to follow the protocol designed for that study. In these studies, the recording duration for each posture was between 5 seconds to 3 minutes. A total of 4 papers did not mention the duration of the study [13,18,20,21]. Two studies [1,16] were designed in an occupational (free-living) setting with a sitting duration from 3 hours to 6.2 hours. In these 2 studies, the subjects were free to choose their postures as they worked. All studies in this paper had recruitment numbers varying between 6 and 72 subjects. Except for 1 study [22], all mentioned the number of subjects.

Classification Algorithms

Different algorithms applied to differentiate and identify the various sitting postures, the type of features, and the number of postures classified are investigated in this section. NNs with varying parameters of neurons, layers, transfer function, and backpropagation methods were used for classifying the sitting postures in the majority of the studies [14,15,18-20,22]. Support vector machines [17,21,30] and random forest (RF) [1,14,16] were the second most used models. Furthermore, algorithms such as K-means [30], multimodal regression [14], boosting [14], dynamic time warping [13], k-nearest neighbors [20], sliced inverse regression (SIR) [20], decision tree [23], Naive Bayes [21], SimpleLogistic [21], linear discriminant analysis [20], principal component analysis (PCA) [20], and triangle center [24] were deployed in the papers to classify sitting postures. In this study, 6 papers [14,15,18,20,21,30] compared the performance of classification algorithms using more than one classifier. Please refer to Table 4 for more information on the classification accuracy.

Training of the classification algorithm and its accuracy depends on the type of features used and its sample size. Different types of features have been used in papers to train and test classification algorithms. Ma et al [30] extracted features from the accelerometer using PCA. Zemp et al [1,14,16] used the median of the sensor data's 1-second duration as the features. In the study by Xu et al [13], the two-dimensional pressure data were converted into one-dimensional data, and the similarity between the signals was used as the feature. In 2 papers [15,19], data collected for each posture from the pressure sensors were divided into groups, and the average of each group was used as a feature. Mutlu et al [21] obtained the position and size of the bounding box; distance of the bounding boxes; the distance and angle between the centers of the pressure areas of the seat and backrest; the centers, radii, and orientations of 2 ellipses from the seat; and the pressure applied to the bottom area as the features to train the algorithm. Huang et al [22] used 40 frames of collected pressure data for each position as the training data. Wang et al [23] used the average and SD of the pressure of the sensors as features. Noh et al [24] used the distance between the center points, intensity, and frequency size of the pressure sensor movements between the current frame center and the previous to train the algorithm.

In the reviewed papers, the number of sitting postures classified varied between 5 and 12. The most common postures observed in Table 4 were upright sitting lean forward, lean backward, lean right, and lean left, as shown in Figure 1. The rest of the postures mentioned in the papers are slight variants of these postures and include different limb movements. Noh et al [24] has also considered in their studies trembling and twisting of the right and left legs.

Algorithm Performance

The evaluation of the algorithms must understand the accuracy, sensitivity, and positive predictive value and check if there is overfitting of the algorithm. Most classification algorithms use confusion matrices to evaluate the posture's classification accuracy. The confusion matrix gives the accuracy of the algorithm and helps interpret the data and the posture that has been misclassified as some other posture. The confusion matrix helps analyze misclassification and rectify errors using each posture's sensitivity and positive predictive value. Of the 14 papers [1,13,15-17,19,21-24], 10 used a confusion matrix to evaluate the performance of the classification algorithm.

Overfitting is another challenge when training an algorithm for machine learning. This occurs when the algorithm fits the training data set and not on a new data set. To check that the algorithm is not overfitting, 6 papers [1,14-17,21] used either 10-fold or leave-one-out cross-validation (internal validation). External validation was performed in the study by Liu et al [18] using 5 external test data sets instead of just cross-validation. Thus, the evaluation and overfitting of the algorithms need to be checked for each classification algorithm.

Discussion

Principal Findings

The WHO and other authors [1,3,9] emphasize the monitoring of sitting postures in free-living environments as a way to understand the mechanical factors involved in musculoskeletal discomfort and pain such as LBP. Different types of sensors (eg, pressure and triaxial accelerometer sensors) and algorithms could be used to accomplish this task. Therefore, this study intends to reveal the current state-of-the-art and the involving algorithms and sensors to classify sitting postures. In each of the included studies, we investigated the type of sensors, the algorithms used, the number of postures classified, the study design, and the environment in which these studies were conducted.

Study Quality Assessment

In summary, most of them had an overall medium quality, with a median score of 13.5. Most of the included papers were

conference papers, and only 5 papers were published as journal papers. Among the 5 journal papers, we rated only 1 as high quality. Most of these papers neither had a straightforward study design nor were their results cross-validated. Thus, the sitting posture classification research is still preliminary and requires further investigation to evaluate the findings. Therefore, to evaluate the findings more systematically in future studies, we recommend that the study be designed carefully. The classification results need to be cross-evaluated such that newer studies can replicate the findings.

Technology

The majority of the papers in this study reported pressure sensors positioned on a chair to classify sitting postures. In these studies, the authors could distinguish between postures because of changes in pressure intensity at different body locations. Arrays of such pressure sensors were used in the backrest and seat [1,13,16-18,20-22] in these studies. However, this is an inefficient use of sensors. As we know from the other studies [15,19,23], the sensors can be strategically placed on the chair in direct contact with the ischial tuberosity, thigh, lumbar, and scapular regions of the subject's body. Such positioning can show the changes in the pressure intensity by using fewer sensors. Nevertheless, while using pressure sensors, researchers must ensure that the subjects empty their pant pockets, as this could hamper the results by changing the pressure intensity further [15,19]. Furthermore, the study conducted by Mutlu et al [21] concluded that the number of pressure sensors should be decreased to reduce the cost of hardware and improve the classification accuracy. Therefore, to improve the classification accuracy and reduce the hardware cost, we suggest the strategic placement of sensors. Thus, the type of sensor and its location must be carefully considered while performing the study.

Regarding the classification of postures, as the number of distinct postures increased, the accuracy of the algorithm decreased [1,16], as the differentiation between the postures was challenging. Therefore, in some studies that used pressure sensors, specific postures were merged into a single posture, resulting in the loss of similar postures. One possible solution to overcome this challenge is to measure the spine movement. Spine movements can be used to differentiate between similar but distinct postures. In the study conducted by Ma et al [30], a triaxial accelerometer was placed at a random location on the cervical spine to measure spine movements to classify sitting postures. However, this was not enough to measure the spine movement, as the upper and lower parts of the spine are independent in motion [32,33]. Accordingly, we propose using multiple sensors, which measure the orientation (eg, inertial measurement unit [IMU] sensors) of the upper and lower spine to be employed to differentiate and classify similar postures. IMUs used in movement trackers in clothing is another option that could be considered if the sensors are placed at the right location on the spine.

Study Design

Most studies were performed in a controlled environment, with subjects being asked to sit in specific postures. However, in free-living conditions, the number of postures could vary and might not match those performed in controlled environments. Moreover, sitting postures depend on the subject's spinal curvature, intradiscal pressure, tissue stress, and muscle activation [34,35]; however, most studies did not investigate these factors in depth. Therefore, we strongly recommend that communities perform sitting posture classification in free-living environments. The sitting posture classification is personalized and can be translated into real-life sitting.

Classification Algorithms

The accuracy of the classification algorithms depends on the type of features, location of the sensors, number of subjects, sample size of the data used for training the algorithm, and number of postures classified. For example, the studies that had the highest number of subjects (72, 64, 52, and 50) had maximum accuracies of 80.9% (Artificial NN), 90% (RF), 87% (SimpleLogistic), and 86% (SIR), respectively [1,19-21]. On the basis of the comparison of the number of classes and the duration of the experiment, the RF algorithm appears to be suitable for classification using pressure sensors. However, the conclusion that the RF algorithm performs well could be biased; it is still too early to conclude that the RF algorithm has the best performance, as there were only 7 postures involved; and the information regarding evaluation using labels was not provided. On the basis of these findings, we advise that the assessment of the predictions of these algorithms should not be based only on the overall accuracy of the system but also on the classification accuracy of each posture and the sample size.

After analyzing the *Type of postures* column in Table 4, we can infer that there are 5 main sitting postures: upright sitting, lean forward, lean backward, lean left, and lean right. Other similar postures are the combination of these postures along with spine and limb movements. Limb movements are an essential aspect of understanding musculoskeletal discomfort and pain [25,27]. Studies have revealed that cross-legged sitting results in asymmetries in the spine and pelvic shapes and increases external oblique muscle activities [26]. Therefore, limb movements must be considered while performing the classification. Furthermore, the postures can be subclassified based on the spine movement. For example, there are 2 types of sitting postures in upright sitting: thoracic upright sitting and lumbopelvic upright sitting. Therefore, in the subclassification, the spine curvature should also be identified along with the posture type, as the curves provide insight into the type of strain the lower back undergoes [36]. Thus, we would point out that future researchers perform subclassification of postures based on spine and limb movements and 5 basic sitting postures.

Static sitting postures are associated with musculoskeletal discomfort and pain. Hence, it is important to understand how frequently a subject moves while monitoring sitting postures [1,16]. Therefore, static sitting postures must be differentiated from dynamic sitting postures. Hence, 2 studies measured transitional periods [1,16], representing the change of one posture to other. Transitional periods also indicate whether a person with static sitting posture has made changes. Therefore, transitional periods must be considered when classifying sitting postures. However, the types of static postures that indicated the presence of musculoskeletal discomfort and pain were not mentioned. Therefore, to understand the cause of

musculoskeletal discomfort and pain, we urge future studies to unambiguously classify the type of static sitting posture while considering the transitional periods for differentiating between static and dynamic postures.

Algorithm Performance

In the shortlisted papers, there was no mention of the use of labels for evaluating the accuracy of the classification. An exclusive annotation of labels is not required in studies using a defined protocol. However, when the study is performed in a free-living environment, exclusive annotation of labels is needed to evaluate the performance of the classification algorithm. One solution for labels is the self-reporting of sitting postures by the subject. However, this was limited by the duration of the measurements. Over a longer time, the subjects may become unaware of many movements while concentrating on their work. The challenge of self-reporting of labeled postures can be overcome using video cameras [37]. In the future, we will use a three-dimensional camera as a ground truth that can be used to verify the predicted activity at any particular time instant.

During the evaluation, it is also important to check if there is an overfitting of the algorithm. In this study, most of the included studies prevented overfitting by performing cross-validation or by using a new data set for testing. In one of the studies, when the algorithm's training and testing were performed on a data set collected from a known subject, the classification accuracy was 98.9% after cross-validation. However, when training and testing were performed with 10-fold cross-validation using 9 subjects' data for training and 1 subject's data for testing, the accuracy decreased to 93.9% [17]. Therefore, it is crucial to evaluate the robustness of an algorithm even when the same subject is not used to train the algorithm.

Conclusions

This study has been conducted to understand the types of sitting postures in the context of spine and leg movements, as sitting for long hours is related to musculoskeletal discomfort and pain.

The quality appraisal shows that future studies need to provide a more precise description of the study design and validation to replicate the studies. The following 5 main sitting postures were present in most of the studies evaluated: upright sitting, lean forward, lean backward, lean left, lean right, and different combinations of limb and spine movements. However, a deeper understanding of spine orientation and variations in those sitting postures are still needed for a more personalized assessment in the context of musculoskeletal discomfort and pain. This is because the individual posture relies on spine curvature, that is, even upright sitting differs from person to person. Even the same person does not always maintain the same posture as the postures instructed in the laboratory. Therefore, it is essential to perform these studies in a free-living environment to understand people's actual postures and reduce the bias from experiments in a controlled environment. To accomplish these studies in a free-living environment in the future, we recommend using multiple sensors that can measure three-dimensional spine movement and angle, such as IMUs. Furthermore, we suggest using labels to evaluate the classification of sitting postures and cross-validation of the algorithms to avoid overfitting to a specific data set. Three-dimensional cameras could be recommended for initial studies to obtain labels. Finally, we recommend the measurement of transitional periods to shed light on more factors affecting musculoskeletal discomfort and pain.

Limitations

In this study, we only included papers written in English, excluding papers written in other languages. Furthermore, with the strict exclusion criteria, there is a possibility that this study missed some additional methods for sitting posture classification. Moreover, we did not include a quantitative analysis (eg, meta-analysis) because of the high heterogeneity in subject characteristics, experimental design, and algorithms. Finally, systematic technological reviews would benefit from standardized methods to assess the risk of bias and systematic content creation, similar to the PRISMA guidelines used in medical and life sciences. Therefore, we believe that the research community must invest in more standardized systematic reviews in such interdisciplinary areas.

Acknowledgments

The authors would like to thank Abhishek M Kappattanavar and Shireen Parasa for helping with the shortlisting of the papers and discussing them in the crucial stage of writing this paper. The authors would also like to thank Ariane Morassi Sasso for reviewing the paper. This study was generously funded by the Hasso Plattner Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The search terms for the three database search engines.
[PDF File (Adobe PDF File), 388 KB - biomedeng_v6i1e21105_app1.pdf]

References

1. Bontrup C, Taylor WR, Fliesser M, Visscher R, Green T, Wippert P, et al. Low back pain and its relationship with sitting behaviour among sedentary office workers. Appl Ergon 2019 Nov;81 [FREE Full text] [doi: 10.1016/j.apergo.2019.102894] [Medline: 31422243]

2. Standing or walking versus sitting on the job in 2016. US Bureau of Labor Statistics. 2017. URL: https://www.bls.gov/opub/ted/2017/standing-or-walking-versus-sitting-on-the-job-in-2016.htm [accessed 2020-11-12]
3. Søndergaard KHE, Olesen CG, Søndergaard EK, de Zee M, Madeleine P. The variability and complexity of sitting postural control are associated with discomfort. J Biomech 2010 Jul 20;43(10):1997-2001. [doi: 10.1016/j.jbiomech.2010.03.009] [Medline: 20399433]
4. Pronk NP, Katz AS, Lowry M, Payfer JR. Reducing occupational sitting time and improving worker health: the Take-a-Stand Project, 2011. Prev Chronic Dis 2012;9:E154 [FREE Full text] [doi: 10.5888/pcd9.110323] [Medline: 23057991]
5. Lis AM, Black KM, Korn H, Nordin M. Association between sitting and occupational LBP. Eur Spine J 2007 Feb;16(2):283-298 [FREE Full text] [doi: 10.1007/s00586-006-0143-7] [Medline: 16736200]
6. Cho CY, Hwang YS, Cherng RJ. Musculoskeletal symptoms and associated risk factors among office workers with high workload computer use. J Manipulative Physiol Ther 2012 Sep;35(7):534-540. [doi: 10.1016/j.jmpt.2012.07.004] [Medline: 22951267]
7. Wieser S, Horisberger B, Schmidhauser S, Eisenring C, Brügger U, Ruckstuhl A, et al. Cost of low back pain in Switzerland in 2005. Eur J Health Econ 2011 Oct 5;12(5):455-467 [FREE Full text] [doi: 10.1007/s10198-010-0258-y] [Medline: 20526649]
8. Balagué F, Mannion AF, Pellisé F, Cedraschi C. Non-specific low back pain. Lancet 2012 Feb 04;379(9814):482-491. [doi: 10.1016/S0140-6736(11)60610-7] [Medline: 21982256]
9. Ehrlich GE. Low back pain. Bull World Health Organ 2003;81(9):671-676 [FREE Full text] [Medline: 14710509]
10. Hernandez A, Gross K, Gombatto S. Differences in lumbar spine and lower extremity kinematics during a step down functional task in people with and people without low back pain. Clin Biomech (Bristol, Avon) 2017 Aug;47:46-52. [doi: 10.1016/j.clinbiomech.2017.05.012] [Medline: 28600994]
11. Clays E, De Bacquer D, Leynen F, Kornitzer M, Kittel F, De Backer G. The impact of psychosocial factors on low back pain: longitudinal results from the Belstress study. Spine (Phila Pa 1976) 2007 Jan 15;32(2):262-268. [doi: 10.1097/01.brs.0000251884.94821.c0] [Medline: 17224824]
12. Papi E, Koh WS, McGregor AH. Wearable technology for spine movement assessment: a systematic review. J Biomech 2017 Nov 07;64:186-197 [FREE Full text] [doi: 10.1016/j.jbiomech.2017.09.037] [Medline: 29102267]
13. Xu W, Huang MC, Amini N, He L, Sarrafzadeh M. eCushion: a textile pressure sensor array design and calibration for sitting posture analysis. IEEE Sensors J 2013 Oct;13(10):3926-3934. [doi: 10.1109/jsen.2013.2259589]
14. Zemp R, Tanadini M, Plüss S, Schnüriger K, Singh NB, Taylor WR, et al. Application of machine learning approaches for classifying sitting posture based on force and acceleration sensors. Biomed Res Int 2016;2016 [FREE Full text] [doi: 10.1155/2016/5978489] [Medline: 27868066]
15. Martins L, Lucena R, Belo J, Santos M, Quaresma C, Jesus A, et al. Intelligent chair sensor. In: Engineering Applications of Neural Networks. Berlin, Heidelberg: Springer,; 2013:191.
16. Zemp R, Fliesser M, Wippert P, Taylor WR, Lorenzetti S. Occupational sitting behaviour and its relationship with back pain - a pilot study. Appl Ergon 2016 Sep;56:84-91 [FREE Full text] [doi: 10.1016/j.apergo.2016.03.007] [Medline: 27184315]
17. Kamiya K, Kudo M, Nonaka H, Toyama J. Sitting posture analysis by pressure sensors. In: Proceedings of the 19th International Conference on Pattern Recognition. 2008 Presented at: 19th International Conference on Pattern Recognition; Dec 8-11, 2008; Tampa, FL, USA. [doi: 10.1109/icpr.2008.4761863]
18. Liu W, Guo Y, Yang J, Hu Y, Wei D. Sitting posture recognition based on human body pressure and CNN. In: AIP Conference Proceedings. 2019 Presented at: AIP Conference Proceedings (Vol. 2073, No. 1,); 2019 Feb 13; Wuhan, China. [doi: 10.1063/1.5090747]
19. Pereira H, Martins L, Almeida R, Ribeiro B, Quaresma C, Ferreira A, et al. System for posture evaluation and correction: development of a second prototype for an intelligent chair. In: Proceedings of the International Conference on Biomedical Electronics and Devices - Volume 1: BIODEVICES. 2015 Presented at: International Conference on Biomedical Electronics and Devices - Volume 1: BIODEVICES; 2015; Lisbon, Portugal p. 204-209. [doi: 10.5220/0005286002040209]
20. Zhu M, Martinez A, Tan H. Template-based recognition of static sitting postures. In: Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop. 2003 Presented at: 2003 Conference on Computer Vision and Pattern Recognition Workshop; 2003; Madison, WI, USA. [doi: 10.1109/cvprw.2003.10049]
21. Mutlu B, Krause A, Forlizzi J, Guestrin C, Hodgins J. Robust, low-cost, non-intrusive sensing and recognition of seated postures. In: Proceedings of the 20th annual ACM symposium on User interface software and technology. 2007 Presented at: 20th annual ACM symposium on User interface software and technology; 2007; Newport Rhode Island USA p. 149-158. [doi: 10.1145/1294211.1294237]
22. Huang M, Gibson I, Yang R. Smart chair for monitoring of sitting behavior. In: DesTech Conference Proceedings. 2017 Feb 09 Presented at: The International Conference on Design and Technology; December 5–8, 2016; Geelong, Australia p. 274. [doi: 10.18502/keg.v2i2.626]
23. Wang C, Kim Y, Min SD. A preliminary study on implementation of sitting posture analysis system using a conductive textile. Adv Sci Lett 2017 Oct 01;23(10):10399-10403. [doi: 10.1166/asl.2017.10461]

24. Noh Y, Seo J, Jeong D. Development of distraction limit estimation index using posture change monitoring system. In: Information Science and Applications 2018. Singapore: Springer; 2019:29.
25. Al-Eisa E, Egan D, Deluzio K, Wassersug R. Effects of pelvic asymmetry and low back pain on trunk kinematics during sitting: a comparison with standing. Spine (Phila Pa 1976) 2006 Mar 01;31(5):E135-E143. [doi: 10.1097/01.brs.0000201325.89493.5f] [Medline: 16508537]
26. Ahn S, Kim S, Kang S, Jeon H, Kim Y. Asymmetrical change in the pelvis and the spine during cross-legged sitting postures. J Mech Sci Technol 2013 Nov 21;27(11):3427-3432. [doi: 10.1007/s12206-013-0865-5]
27. Sorensen CJ, Johnson MB, Norton BJ, Callaghan JP, Van Dillen LR. Asymmetry of lumbopelvic movement patterns during active hip abduction is a risk factor for low back pain development during standing. Hum Mov Sci 2016 Dec;50:38-46 [FREE Full text] [doi: 10.1016/j.humov.2016.10.003] [Medline: 27744105]
28. Santos CMDC, Pimenta CADM, Nobre MRC. The PICO strategy for the research question construction and evidence search. Rev Lat Am Enfermagem 2007 Jun;15(3):508-511 [FREE Full text] [doi: 10.1590/s0104-11692007000300023] [Medline: 17653438]
29. Hagströmer M, Ainsworth B, Kwak L, Bowles H. A checklist for evaluating the methodological quality of validation studies on self-report instruments for physical activity and sedentary behavior. J Phys Act Health 2012 Jan;9 Suppl 1(s1):S29-S36. [doi: 10.1123/jpah.9.s1.s29] [Medline: 22287445]
30. Ma S, Cho W, Quan C, Lee S. A sitting posture recognition system based on 3 axis accelerometer. In: Proceedings of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). 2016 Presented at: 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB); Oct 5-7, 2016; Chiang Mai, Thailand. [doi: 10.1109/cibcb.2016.7758131]
31. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7) [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
32. Bao L, Intille S. Activity recognition from user-annotated acceleration data. In: Pervasive Computing. Berlin, Heidelberg: Springer; 2004:1-17.
33. O'Sullivan K, McCarthy R, White A, O'Sullivan L, Dankaerts W. Lumbar posture and trunk muscle activation during a typing task when sitting on a novel dynamic ergonomic chair. Ergonomics 2012 Sep 25;55(12):1586-1595. [doi: 10.1080/00140139.2012.721521] [Medline: 23009637]
34. Wong AYL, Chan TPM, Chau AWM, Cheung TH, Kwan KCK, Lam AKH, et al. Do different sitting postures affect spinal biomechanics of asymptomatic individuals? Gait Posture 2019 Jan;67:230-235. [doi: 10.1016/j.gaitpost.2018.10.028] [Medline: 30380507]
35. Castanharo R, Duarte M, McGill S. Corrective sitting strategies: an examination of muscle activity and spine loading. J Electromyogr Kinesiol 2014 Feb;24(1):114-119. [doi: 10.1016/j.jelekin.2013.11.001] [Medline: 24295543]
36. O'Sullivan PB, Dankaerts W, Burnett AF, Farrell GT, Jefford E, Naylor CS, et al. Effect of different upright sitting postures on spinal-pelvic curvature and trunk muscle activation in a pain-free population. Spine (Phila Pa 1976) 2006 Sep 01;31(19):E707-E712. [doi: 10.1097/01.brs.0000234735.98075.50] [Medline: 16946644]
37. Winters M, Brugarolas R, Majikes J, Mealin S, Yuschak S, Sherman B, et al. Knowledge engineering for unsupervised canine posture detection from IMU data. In: Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology. 2015 Presented at: 12th International Conference on Advances in Computer Entertainment Technology; 2015; Iskandar, Malaysia. [doi: 10.1145/2832932.2837015]

Abbreviations

IMU: inertial measurement unit
LBP: low back pain
NN: neural network
PCA: principal component analysis
PICO: Population or Problem, Intervention or Exposure, Comparison, and Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RF: random forest
SIR: sliced inverse regression
WHO: World Health Organization

Edited by G Eysenbach; submitted 23.07.20; peer-reviewed by K Pradeep, D Pförringer; comments to author 09.10.20; revised version received 18.12.20; accepted 08.02.21; published 01.03.21.

Please cite as:
Kappattanavar AM, Steckhan N, Sachs JP, Freitas da Cruz H, Böttinger E, Arnrich B
Monitoring of Sitting Postures With Sensor Networks in Controlled and Free-living Environments: Systematic Review
JMIR Biomed Eng 2021;6(1):e21105
URL: http://biomedeng.jmir.org/2021/1/e21105/
doi:10.2196/21105
PMID:

Original Paper

Physical Activity Evaluation Using a Voice Recognition App: Development and Validation Study

Hideyuki Namba[1], PhD, MEng

Physical Education Lab., College of Science and Technology, Nihon University, Chiba, Japan

Corresponding Author:
Hideyuki Namba, PhD, MEng
Physical Education Lab.
College of Science and Technology
Nihon University
7-24-1 Narashinodai-Funabashi
Chiba, 274-8501
Japan
Phone: 81 47 469 5518
Fax: 81 47 469 5518
Email: nanba.hideyuki@nihon-u.ac.jp

Abstract

Background: Historically, the evaluation of physical activity has involved a variety of methods such as the use of questionnaires, accelerometers, behavior records, and global positioning systems, each according to the purpose of the evaluation. The use of web-based physical activity evaluation systems has been proposed as an easy method for collecting physical activity data. Voice recognition technology not only eliminates the need for questionnaires during physical activity evaluation but also enables users to record their behavior without physically touching electronic devices. The use of a web-based voice recognition system might be an effective way to record physical activity and behavior.

Objective: The purpose of this study was to develop a physical activity evaluation app to record behavior using voice recognition technology and to examine the app's validity by comparing data obtained using both the app and an accelerometer simultaneously.

Methods: A total of 20 participants (14 men, 6 women; mean age 19.1 years, SD 0.9) wore a 3-axis accelerometer and inputted behavioral data into their smartphones for a period of 7 days. We developed a behavior-recording system with a voice recognition function using a voice recognition application programming interface. The exercise intensity was determined from the text data obtained by the voice recognition program. The measure of intensity was metabolic equivalents (METs).

Results: From the voice input data of the participants, 601 text-converted data could be confirmed, of which 471 (78.4%) could be automatically converted into behavioral words. In the time-matched analysis, the mean daily METs values measured by the app and the accelerometer were 1.64 (SD 0.20) and 1.63 (SD 0.20), respectively, between which there was no significant difference (P=.57). There was a significant correlation between the average METs obtained from the voice recognition app and the accelerometer in the time-matched analysis (r=0.830, P<.001). In the Bland-Altman plot for METs measured by the voice recognition app as compared with METs measured by accelerometer, the mean difference between the two methods was very small (0.02 METs), with 95% limits of agreement from –0.26 to 0.22 METs between the two methods.

Conclusions: The average METs value measured by the voice recognition app was consistent with that measured by the 3-axis accelerometer and, thus, the data gathered by the two measurement methods showed a high correlation. The voice recognition method also demonstrated the ability of the system to measure the physical activity of a large number of people at the same time with less burden on the participants. Although there were still issues regarding the improvement of automatic text data classification technology and user input compliance, this research proposes a new method for evaluating physical activity using voice recognition technology.

(JMIR Biomed Eng 2021;6(1):e19088) doi:10.2196/19088

KEYWORDS

voice recognition; smartphone; physical activity; accelerometer; application

Introduction

There has been remarkable progress in wearable devices (the collective name for information devices that are worn and carried) in recent years. Particularly, devices that objectively record lifelog data, such as physical activity and sleep, have attracted attention [1]. Many companies have developed smartphones and wristwatch-type behavior-recording sensors. Wearable devices such as these are expected to contribute to the promotion of physical activity for many people [2]. According to the global market forecast for mobile health solutions (2015-2022), the global mobile health market is expected to reach $90.4 billion in 2022 from $21.1 billion in 2017 [3]. The global health market, along with the Internet of Things and big data analysis, has become an important part of Japan's growth strategy for the fourth industrial revolution [4]. Despite the fact that deficient physical activity is the third leading cause of death [5], physical activity in Japan has reportedly decreased slightly over the past 10 years [6].

The questionnaire method has excellent cost performance and is suitable for large-scale physical evaluation surveys, but it has been pointed out that there is a problem with its validity [7]. An epidemiological study [8] used a smartphone app (the Argus app) to demonstrate that step count is negatively associated with obesity. However, most wearable devices with built-in accelerometers have been shown to underestimate the physical activity energy expenditure value of the doubly labeled water [9]. Although accelerometers are highly useful, data bridging might be necessary in epidemiologic studies due to underestimation (depending on the wearable device used to measure energy expenditure). In a previous study [10], we developed and validated a physical activity measurement system based on a 24-hour activity recording method, to be used as an accessible physical activity evaluation system. The accuracy of the energy expenditure measurement was determined using the doubly labeled water method. The average energy expenditure measurement value and the average value of the doubly labeled water nearly matched, and the two were very highly associated [10]. In another study [11], we collected data from thousands of people using this physical activity measurement system and conducted epidemiological studies such as regional comparisons. We found that our system can evaluate physical activity in a large number of people simultaneously, with substantial accuracy and at a low cost. However, the users are tasked with inputting information about the activities in which they have engaged every 15 minutes for the 24 hours in a day. As a result, users are required to make an effort and input time while stopping physical activity.

To address these problems, we attempted, in this study, to develop a behavior-recording method using voice recognition. Voice recognition technology enables the recognition and translation of verbal information into text, which can then be used in automated data processing systems. Using voice recognition technology enables behavior recording without requiring users to touch electronic devices. Information and communication technology–based telehealth programs with voice recognition technology show the potential to improve the health of patients with chronic heart failure by self-care management behaviors [12]. With the improvement of voice recognition technology accuracy, such programs are used in the medical field to enhance the adherence of health behavior of patients with chronic diseases [13,14]. The possibility of a voice-based mobile nutrition monitoring system that uses voice processing has been reported [15]. However, we do not know of any research that has, to date, used voice recognition technology to evaluate physical activity. Because it has been shown that physical activity can be estimated with high accuracy from 33 types [16] and 66 types [10] of behavior records, we hypothesized that physical activity could be evaluated by using voice input that could then be converted into behavioral text.

If this study could prove the validity of physical activity evaluation using a behavior recording method that relies on voice recognition technology, it might be possible to evaluate physical activity more simply in the future. This study may improve the validity of physical activity evaluation and the usability of voice recognition devices.

The purpose of this study was to develop a physical activity evaluation app that records behavior using voice recognition and to examine the validity of the method by using the app and accelerometer at the same time.

Methods

Participants

A total of 20 healthy students (14 men and 6 women) with an average age of 19.1 (SD 0.9) years participated in the study. The developed voice recognition app was downloaded to their smartphones and a 1-week behavior record was inputted. We explained the purpose of the survey to the participants and the voluntary nature of their participation. We explained that the privacy and anonymity of the participants would be strictly observed, confirmed that there were no related health concerns, and obtained consent. This research project was approved by the Nihon University College of Sports Science for Research Ethics Review (2019-012).

Voice Recognition App Development

We developed a behavior-recording system with a voice recognition function (Figures 1 and 2) using a voice recognition application programming interface. In this research, we commissioned the development of a voice recognition app called ACTRA (Yonefu International Inc, Fukuoka) using the SIRI application program interface (Apple Inc, California). SIRI's speech recognition technology uses an acoustic model and a language model to convert speech data into text via a server. Attempts have been made to utilize various speech recognition technologies, including SIRI, in the field of health medicine [17-19]. The basic structure of our physical activity evaluation system is based on the Web-Based Physical Activity Measurement System [10] and Web-Based Physical Activity Records [11] developed with reference to a simplified physical activity record [16]. By accumulating data in a cloud server and analyzing big data, we were able to analyze behavior patterns, which enabled the categorization of behavior patterns and the estimation of future physical activity from current physical activity patterns.

Figure 1. A behavior-recording system with a voice input app.

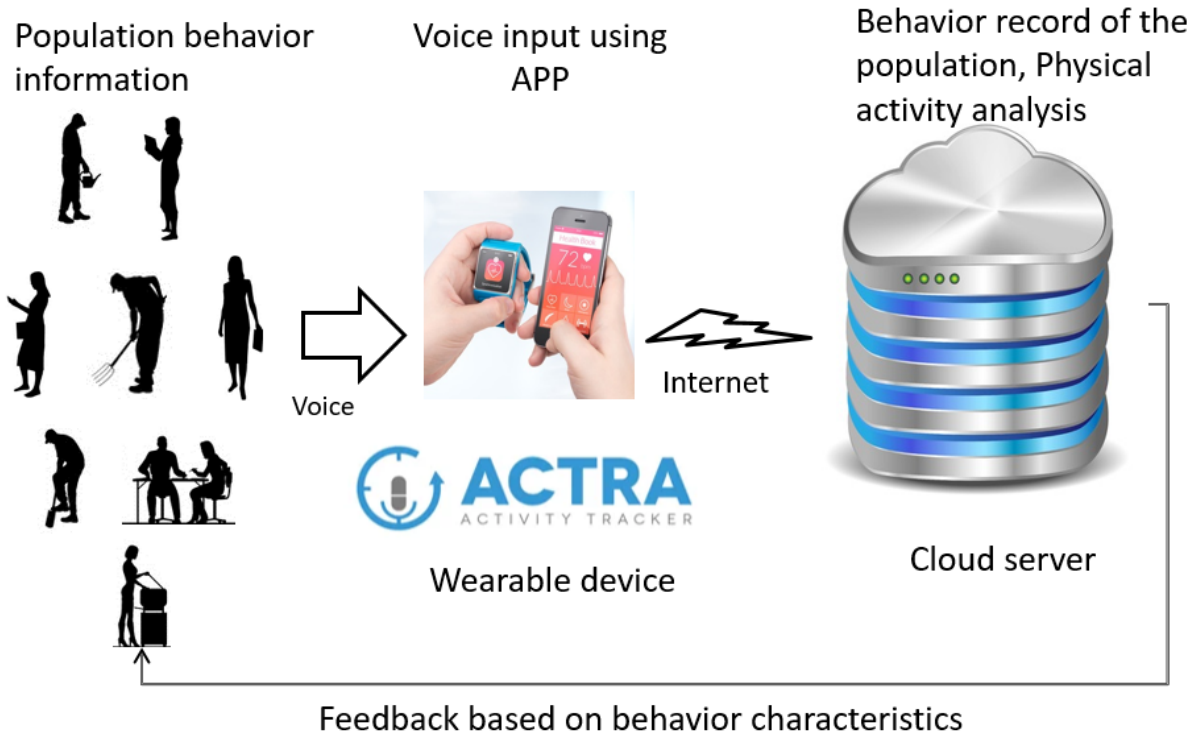
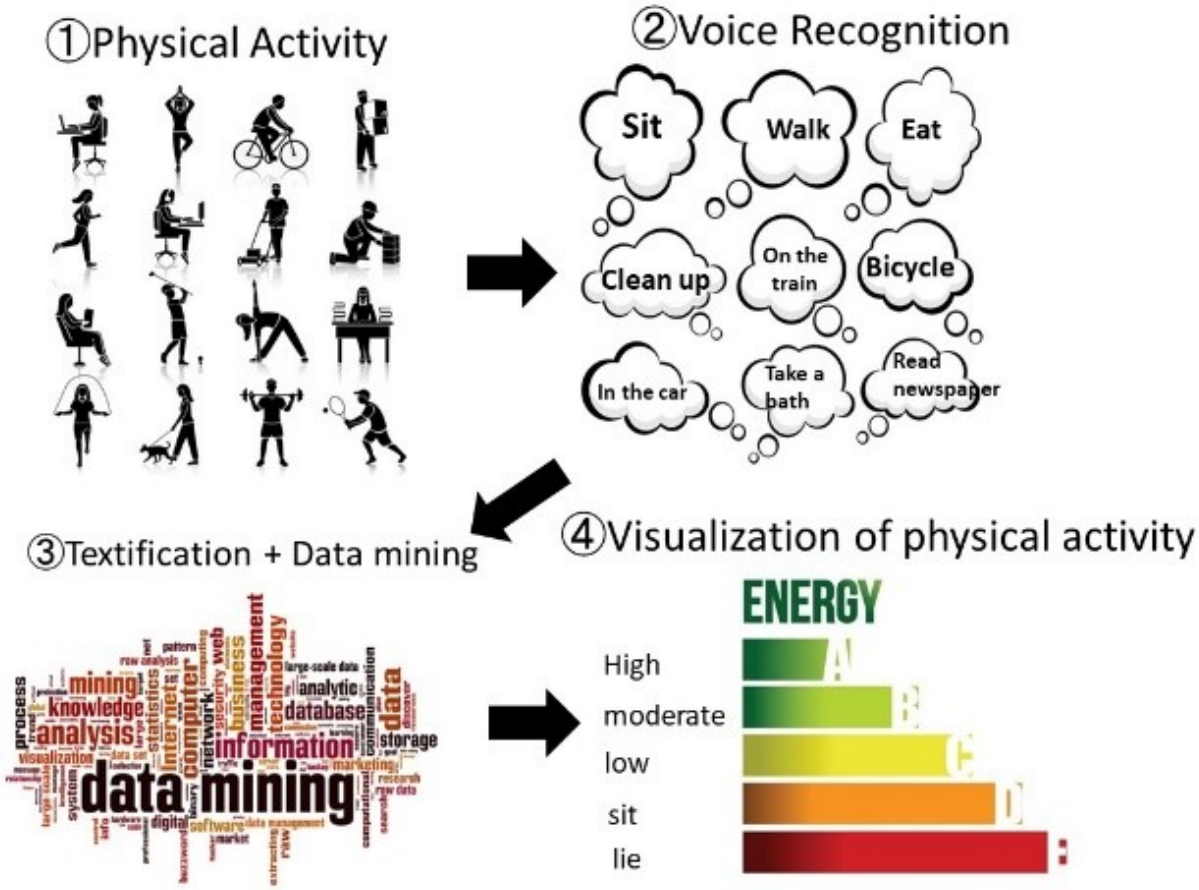


Figure 2. Conceptual view of the visualization of physical activity from the voice input.



We developed an algorithm that can calculate total energy expenditure based on voice data. To our knowledge, this method is the first to evaluate physical activity and energy expenditure based on a behavioral record generated via self-inputted voice data. Figure 3 shows the screen of the developed voice recognition app. When the user presses the microphone button and tells the smartphone what they are or were doing, their voice is recognized and converted into text data (Figure 3, left). The terms "start," "end," "from," "to," and "do" were set as key words to separate words and list the activity in time units with the machine learning function (Figure 3, right). Participants could use the device in real time, such as by saying "start now" when they began an activity. They could also look back at the activities completed that day, and input activities as follows: "Starting at △△○○ hour □□ minute, ●● hour ▼▼ minute end." The behavior database is queried from the word, and the "start" and "end" of the trigger word are extracted. Then, the total time of each behavior is extracted, and the activity intensity from the start to the end of the behavior is automatically recorded. During the study, it was difficult to convert the recorded behavior into text completely automatically, and in some cases human judgment was necessary. The exercise intensity was determined from the text data obtained by the voice recognition program as well as from the correspondence table [20] between the behavior and exercise intensity. For each behavior, the product of exercise intensity and time was obtained, and METs time per day was calculated. The average METs per day was obtained by dividing the total METs value per minute by the total activity time.

Figure 3. Screen of the developed voice recognition app.

Physical Activity Measurement via Accelerometer

Participants wore a 3-axis Active Style Pro accelerometer (HJA-750C, Omron Healthcare, Kyoto) on their lower back for one week as confirmed by the input data of the physical activity measurement system. The accelerometer was considered to have a better correlation with doubly labeled water in the measurement of physical activity energy expenditure than other accelerometers [9]. Participants were instructed to always wear the unit except for during sleep and bathing. The epoch length for processing the data obtained from the accelerometer was set to 10 seconds. The activity intensity (METs) was estimated from the combined acceleration in the vertical, front-back, and left-right directions, and was collected in units of 10 seconds. Activity intensity below the detection threshold, categorized as "a zero count," was a period in which no activity was detected for more than 60 minutes. This time was considered to correlate to periods when the participant was not wearing the device (nonwearing time); the wearing time was determined by subtracting the nonwearing time from 24 hours. Datasets for

days with daily wearing times of 10 hours or more were analyzed. To calculate the 24-hour average METs, 0.9 METs were calculated without measurement data.

Time-Matched Analysis of Both Voice Recognition and Accelerometer Methods

The advantage of using acceleration is that it has high temporal resolution; voice recognition can also be analyzed every minute. Therefore, we performed an analysis in which acceleration data was collated with the behavior recording timeline of voice recognition. As a procedure for time matching, the activity intensity obtained from the accelerometer was matched with multiple behavior records obtained from voice input every minute, and the average METs per day was calculated by each method. The time matching was not automated and was done manually.

Statistical Analysis

Mean METs per day are shown as the mean with the standard deviation. A paired *t* test was performed to compare the average METs values obtained from the voice recognition app and the 3-axis accelerometer. Pearson's correlation coefficient was calculated to examine the correlation between METs obtained from the voice recognition app and the 3-axis accelerometer. The average METs values were calculated for cases where the voice recognition was recorded for 10 hours of activity or more, 14 hours or more, and for time matching analysis. For the values calculated under each analysis condition, the average METs values were compared, and the correlation coefficient was calculated. Using Bland-Altman plots, we related the difference in METs between voice recognition and accelerometer (y axis) to the arithmetic mean of METs for voice recognition and accelerometer (x axis) [21]. Statistical analysis was performed using SPSS version 25 IBM (IBM Corporation, Somers, NY, USA) with an alpha level of less than .05.

Results

In the voice input data of the subject, 601 text-converted data could be confirmed, of which 471 (78.4%) could be automatically converted into behavioral words. For example, phrases with trigger words such as “study from 9:00 to 12:00,” “start meal at 7:00, end meal at 8:00,” and clear words could be automatically converted. However, automatic conversion was not possible for phrases without trigger words such as “get up at 7:00,” or for long sentences containing objects such as “going home by riding a bicycle.” Therefore, the text data that could not be automatically converted was checked and converted by hand.

Table 1 shows the average METs for the voice recognition app and accelerometer. There were 36 days with more than 10 hours of voice recognition data and 16 days with more than 14 hours of voice recognition data. The average daily METs measured by the accelerometer was 1.47 (SD 0.23) and 1.48 (SD 0.24) compared to 1.56 (SD 0.24) and 1.58 (SD 0.28) when the voice recognition data were analyzed for days with 10 or more hours of activity and 14 or more hours, respectively. Under all data extraction conditions, the average METs value determined by voice recognition was significantly higher than that determined by the accelerometer (≥10 hours P=.02, ≥14 hours P=.04). In the time-matched analysis, the voice recognition and accelerometer values were 1.64 (SD 0.20) and 1.63 (SD 0.20), respectively, and there was no significant difference (P=.57).

Table 1. Comparison of average daily METs[a] measured by voice recognition app and accelerometer.

Dataset	Voice recognition, average METs (SD)	Accelerometer, average METs (SD)	*P* value of the *t* test	Number of analyzed days
Days with 10 or more hours of measured activity	1.56 (0.24)	1.47 (0.23)	0.02	36
Days with 14 or more hours of measured activity	1.58 (0.28)	1.48 (0.24)	0.04	16
Time-matched analysis	1.64 (0.20)	1.63 (0.20)	0.57	36

[a]METs: metabolic equivalents.

Figure 4 shows the correlation between the voice recognition data and the average daily METs measured by the accelerometer. Figure 4A shows the data for voice recordings more than 10 hours long, Figure 4B shows the data for recordings that were more than 14 hours long, and Figure 4C shows the date for the time-matched analysis of both voice recognition and accelerometer methods. Under all conditions, the average METs values from voice recognition ranged from 1.2 METs to 2.3 METs. In contrast, the average METs values measured by the accelerometer ranged from 1.0 METs to 2.2 METs. There was a significant correlation between the average METs obtained from voice recognition and that obtained by the accelerometer in the 10 hour dataset, 14 hour dataset, and the time-matched analysis (r=0.545, P=.02; r=0.750, P=.008; and r=0.830, P<.001, respectively).

Figure 4. Correlation of average measured METs with voice recognition app and accelerometer. METs: metabolic equivalents.

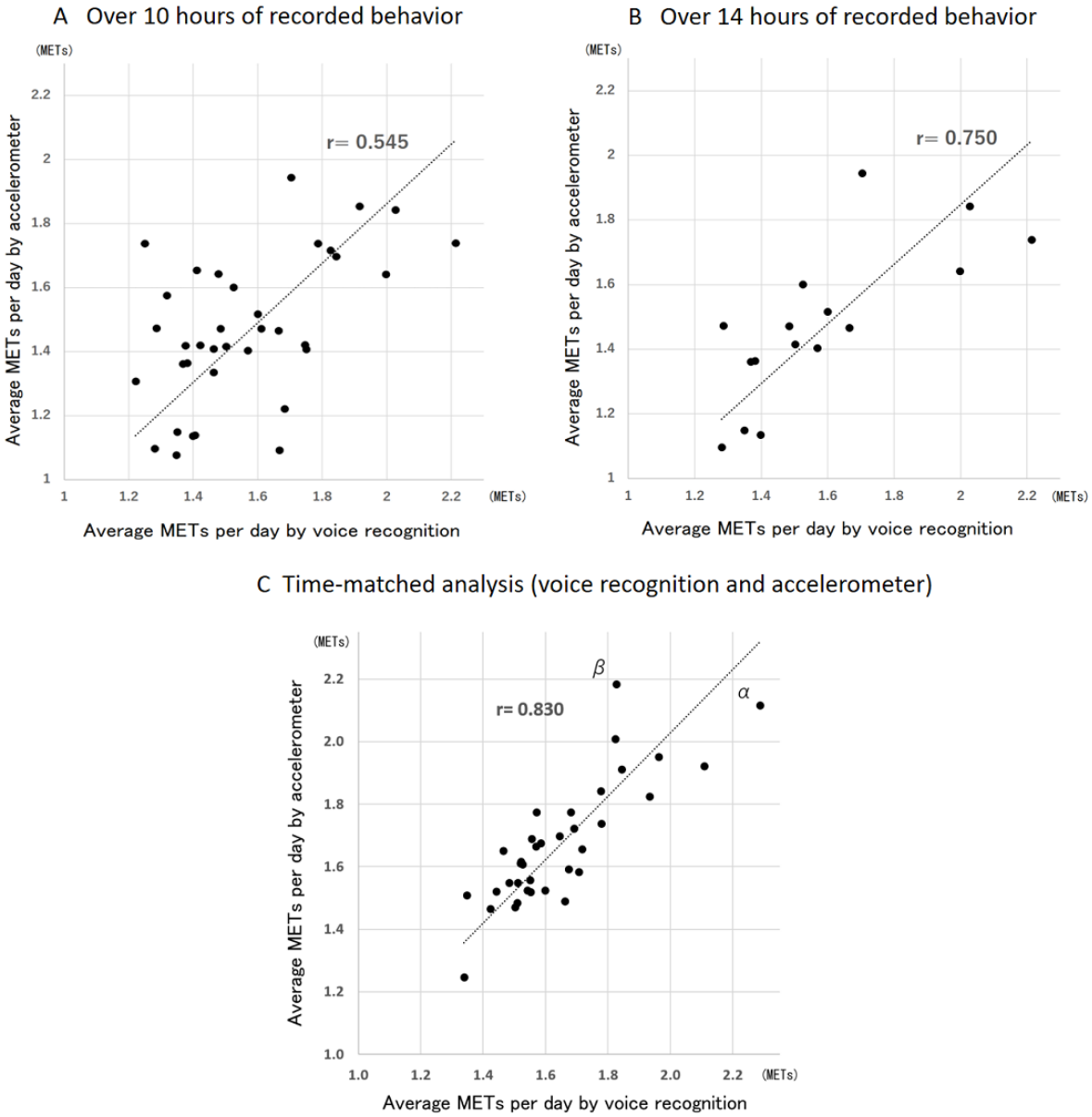


Figure 5 shows the Bland-Altman plot for METs as measured by voice recognition compared with METs as measured by the accelerometer. Figure 5A shows the data for voice recordings more than 10 hours long, Figure 5B shows the data for recordings that were more than 14 hours long, and Figure 5C shows the date for the time-matched analysis of both voice recognition and accelerometer methods. The mean difference (Figure 5A) for the voice recognition and the accelerometer was small (0.09 METs), and the limits of agreement were large at 0.44 METs (SD 1.96). The test for trend was not statistically significant. The regression equation was y = 0.054x + 0.012 (r=.049, P=.78). The mean difference (Figure 5B) between the two methods was small (0.11 METs), and the limits of agreement were large at 0.37 METs (SD 1.96). The test for trend was not statistically significant. The regression equation was y = 0.167x – 0.15 (r=.217, P=.42). In the Bland-Altman plot (Figure 5C) for METs measured by voice recognition compared with METs measured by the accelerometer, the mean difference between the two methods was very small (0.02 METs), and the limits of agreement were 0.24 METs (SD 1.96). The test for trend was not statistically significant. The regression equation was y = –0.0341x + 0.035 (r=–.056, P=.74).

Figure 5. Bland–Altman plot illustrating the difference in average measured METs between the voice recognition app and accelerometer. METs: metabolic equivalents.

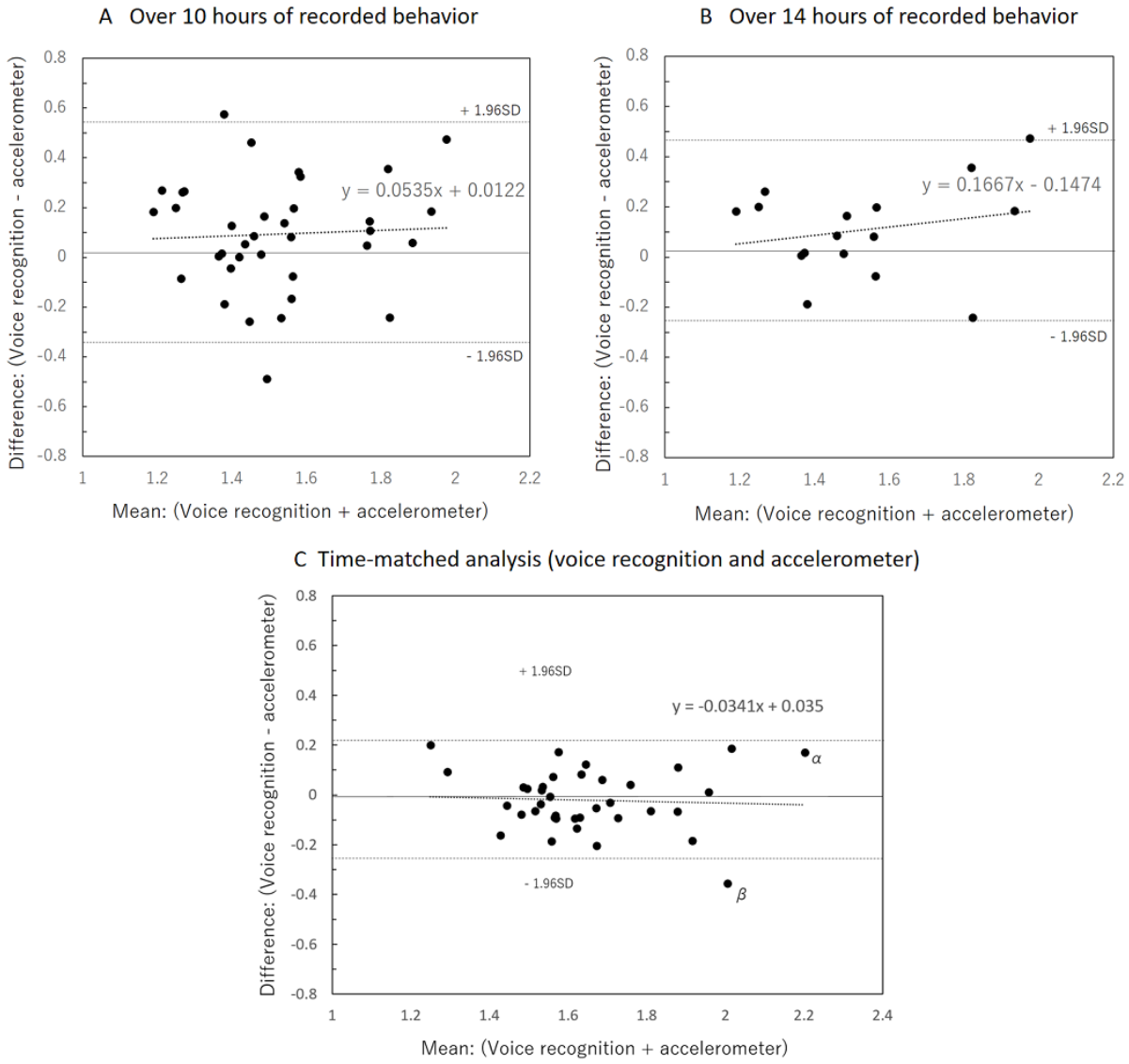
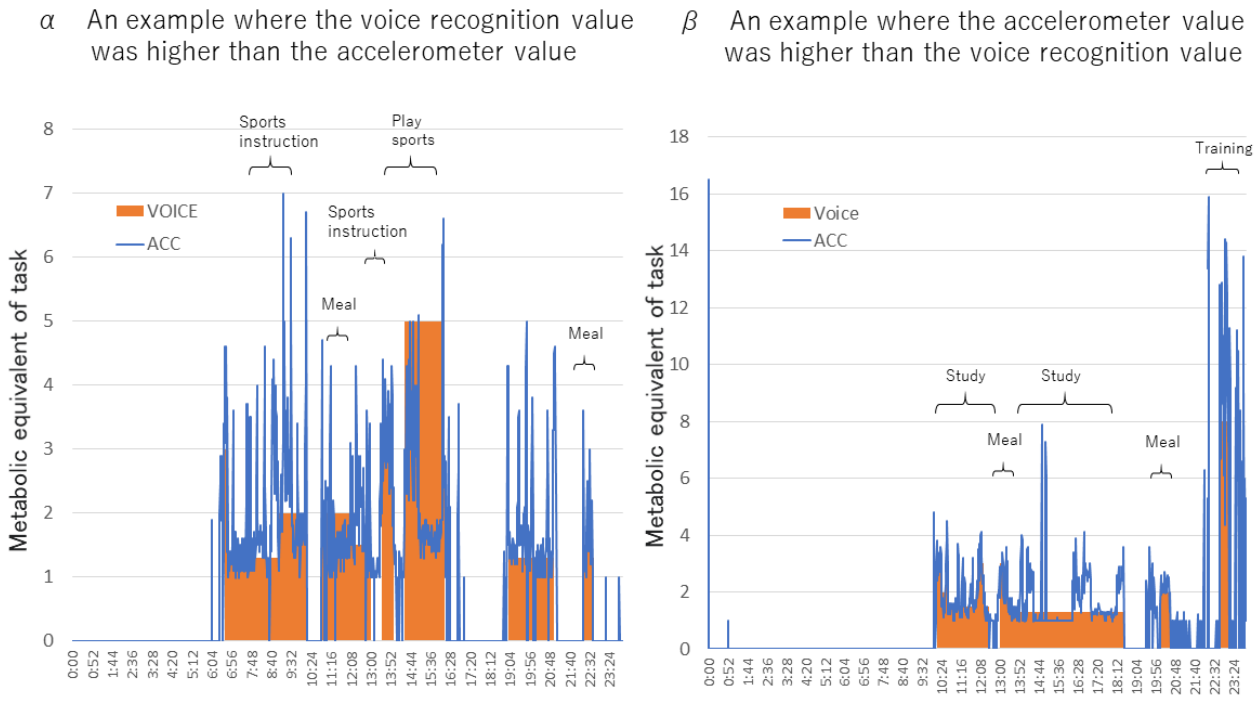


Figure 6 shows two examples in which the error between the METs value measured by voice recognition and the METs value measured value by the accelerometer was large. Figure 6 (α, β) shows the individual data for α and β in Figures 4 and 5. Figure 6α shows that voice recognition app recorded higher METs values than the accelerometer. This was due to the fact that the voice recognition app determined that the game was in progress, while the acceleration data showed a low intensity period. On the other hand, in Figure 6β, the METs values measured by the accelerometer were higher than those measured by the voice recognition app. The cause was judged to be that the voice recognition app recorded the activity as training and set the METs value to 5, whereas, in reality, the activity was accompanied by intense exercise exceeding 10 METs.

Figure 6. 24-hour data for the two high-error examples (α and β in Figure 4C, Figure 5C).



Discussion

Overview

In this study, we developed a behavior-recording app that uses voice recognition technology and verified the validity of the app using an accelerometer. Participants' exercise intensity was estimated from the activity recorded using the voice recognition app, and the average daily METs was calculated [10,11]. When the validity of this data was verified using an accelerometer, we found a high correlation between the time-matched analysis of both app- and accelerometer-measured data, and a moderate correlation when the behavior was recorded for both 10 hours or more and 14 hours or more in a single day. Further, the average daily METs value measured by the voice recognition app in time-matched analysis was not significantly higher than that measured by the accelerometer. However, the average daily METs value measured by the voice recognition app was significantly higher than that measured by the accelerometer for both the 10 hour or more dataset and the 14 hour or more dataset. In a previous study [22], a 24-hour behavior recording strategy using a website produced a significantly higher METs value than an accelerometer. Therefore, when estimating energy expenditure via voice recognition recording, the relevance was moderate if it did not time match the acceleration data, but the average might be overestimated.

Validity of Physical Activity Evaluation Using Voice Recognition

It has been reported that the total energy expenditure reported by an accelerometer underestimates physical activity, even when the gold standard doubly labeled water method is used as a standard [9]. This is because the device is removed during bathing and to charge the device. Therefore, in the present study, the average METs obtained by the voice recognition app might have been slightly higher than that of the accelerometer. In addition, the accelerometer does not count movements that occur in a sitting position or stationary standing position, and acceleration is not measured. Acceleration does not match energy expenditures when riding a bicycle, climbing stairs, or walking on slopes [23]. Also, energy expenditures due to movements of antigravity muscles should be considered [24]. Further, the influence of dietary thermogenesis upon digesting food may be an overlooked factor [7]. Although there is a limit to the physical activity evaluation of the accelerometer, the correlation with physical activity energy expenditure by doubly labeled water has also been shown [9]. In this study, physical activity energy expenditure is not examined because the basal metabolism was not measured, and the average METs was evaluated. Figure 5 (A, B, C) indicates that the error range of the values recorded by voice recognition did not change, regardless of the highest and lowest average METs values. Figure 5 demonstrates no significant correlation between the data in any of the graphs; the error range was shown (0.44 METs (A), 0.37 METs (B), and 0.24 METs (C), 95% confidence interval) regardless of the size of the estimate. In Figure 5B, the regression line of the Brand Altman plot was slightly upward ($y = 0.1667X - 0.1474$), and the data varied. It might be said that the higher the average METs value, the larger the error. In other words, it could be said that there was a day when the acceleration did not actually move much, even if a high-intensity behavior was inputted by voice. On the contrary, the accelerometer moved well, but, in some cases, it didn't move much during voice recording.

Therefore, in consideration of the time resolution, which is the merit of voice recognition and accelerometer use, the time of both methods was matched and analyzed. By using this analysis method, the validity was verified by excluding behaviors that cannot be measured by the accelerometer, which is a weak point

of the accelerometer. For example, data for periods of bathing, periods of daytime sleep, and instances in which participants forgot to put on the accelerometer were excluded. In Figure 5C, the regression line of the Brand Altman plot was flat ($y = -0.0341X - 0.035$) and was within the 95% confidence interval, except for one case. Even so, there was a case where there was a large difference between the voice behavior record and the daily average METs recorded by the accelerometer, so Figure 6 shows the details of the α and β data. The graphs of α and β show cases where voice recognition is overestimated and underestimated compared to acceleration data in sports activities. Other than with sports activities, errors occurred when the activity intensity was difficult to understand for long periods, such as when participants rode bicycles or worked part-time jobs.

Possibility of Behavior Recording by Voice Recognition

In the physical activity measurement system developed previously [10], behavior was reported every 15 minutes. Thus, an activity that was performed for about 10 minutes would have been counted as having taken place for 15 minutes. In this study, we hypothesized that behaviors would be more accurate when recorded via voice recognition because the behavior would be evaluated on a finer timescale. In the time-matched analysis, a high correlation was found between the behavior recorded by voice recognition and the acceleration data, and the mean value also matched, so our hypothesis could be true. However, in the analysis dealing with all the data of one day, there was a difference in the mean METs values taken by the two measurement systems, and the correlation coefficient was moderate. This may be due to participants having experienced difficulty remembering the details of their behaviors, or fatigue related to the task of inputting behaviors sequentially in real time. In recent studies that used the estimation method of physical activity with wearable trackers, data collection compliance and validity were positively correlated, but stricter compliance may have increased the number of excluded data points [25]. It is necessary to correctly calculate the average daily METs, assuming practical application that is useful for dietary guidance and lifestyle improvement.

Limitations and Future Studies

The limitation of this study is that if participant compliance is not high, the measurement accuracy will decrease. In this study, the average daily METs per person were treated as one dataset. If all inputs were perfect, there would have been a total of 140 days of recording, but the final data set we analyzed included 36 days with 10 or more hours of activity and 16 days with 14 or more hours of activity. Increasing compliance when evaluating physical activity using voice recognition is an important factor for future research. There was a technical problem to convert all voice data automatically and mechanically into exercise intensity. However, this study is the first report to validate the evaluation of physical activity using voice input technology. The advantage of the app developed in this research is that it is possible to record behavior by voice input with just a single touch on the device.

This section describes applied research and practice for the future of this research. The method developed in this study could replace the traditional questionnaire in epidemiological studies and could be used to evaluate the physical activity of many people with less burden on the subjects. Time-matched analysis showed that physical activity could be evaluated with extremely high accuracy using voice recognition technology. The developed app enables a simple and low-cost evaluation of physical activity measurement, which may contribute to disease prevention. Future apps that incorporate deep learning using artificial intelligence may be useful for physical activity evaluation research. Although the details of participants' behavior are not known from the acceleration data, the voice recognition method might be useful for analyzing many behavior patterns.

Conclusions

We developed a behavior-recording app using voice recognition and examined its validity using an accelerometer. The system was found to be an effective method for collecting physical activity data and is appropriate for use in epidemiological studies. Although the conversion of voice-to-text data into behavior was not perfect, voice recognition technology is evolving day by day and improvement could be expected. The results of the time-matched analysis showed that physical activity could be measured with high accuracy by voice recognition technology. Participant compliance when using voice input technology is important for ensuring data validity. This research proposes a new method for evaluating physical activity using voice recognition technology.

Acknowledgments

I would like to thank Sydney Koke, MFA, from Edanz Group for editing a draft of this manuscript. This work was supported by JSPS KAKENHI 17K01800. I would also like to thank YONEFU International Group for their help in system development.

Conflicts of Interest

The development of the voice recognition app was outsourced to YONEFU International group.

References

1. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. Med Sci Sports Exerc 2016 Mar 17;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]

2. Pratt M, Sarmiento OL, Montes F, Ogilvie D, Marcus BH, Perez LG, et al. The implications of megatrends in information and communication technology and transportation for changes in global physical activity. The Lancet 2012 Jul 21;380(9838):282-293. [doi: 10.1016/S0140-6736(12)60736-3] [Medline: 22818940]
3. mHealth Solutions Market by Connected Devices (Blood Pressure Monitor, Glucose Meter, Peak Flow Meter) Apps (Weight Loss, Woman Health, Personal Health Record, & Medication) Services (Diagnostic, Remote Monitoring, Consultation) - Global Forecasts to 2022. San Francisco, California: MarketsandMarkets; Aug 2017.
4. Cabinet O. Fifth Science and Technology Basic Plan. Japanese Government. 2016. URL: https://www8.cao.go.jp/cstp/kihonkeikaku/index5.html [accessed 2020-12-29]
5. Ikeda N, Inoue M, Iso H, Ikeda S, Satoh T, Noda M, et al. Adult mortality attributable to preventable risk factors for non-communicable diseases and injuries in Japan: a comparative risk assessment. PLoS Med 2012 Jan;9(1):e1001160 [FREE Full text] [doi: 10.1371/journal.pmed.1001160] [Medline: 22291576]
6. Ministry OH, LaborWelfare. The 2016 National Health and Nutrition Survey. Japanese Government. 2017. URL: https://www.nibiohn.go.jp/eiken/kenkounippon21/download_files/eiyouchousa/2016.pdf [accessed 2020-12-29]
7. Neilson HK, Robson PJ, Friedenreich CM, Csizmadi I. Estimating activity energy expenditure: how valid are physical activity questionnaires? Am J Clin Nutr 2008 Feb;87(2):279-291. [doi: 10.1093/ajcn/87.2.279] [Medline: 18258615]
8. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Jul 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
9. Murakami H, Kawakami R, Nakae S, Yamada Y, Nakata Y, Ohkawara K, et al. Accuracy of 12 Wearable Devices for Estimating Physical Activity Energy Expenditure Using a Metabolic Chamber and the Doubly Labeled Water Method: Validation Study. JMIR Mhealth Uhealth 2019 Aug 02;7(8):e13938 [FREE Full text] [doi: 10.2196/13938] [Medline: 31376273]
10. Namba H, Yamaguchi Y, Yamada Y, Tokushima S, Hatamoto Y, Sagayama H, et al. Validation of Web-based physical activity measurement systems using doubly labeled water. J Med Internet Res 2012 Sep 25;14(5):e123 [FREE Full text] [doi: 10.2196/jmir.2253] [Medline: 23010345]
11. Namba H, Yamada Y, Ishida M, Takase H, Kimura M. Use of a Web-based physical activity record system to analyze behavior in a large population: cross-sectional study. J Med Internet Res 2015 Mar 19;17(3):e74 [FREE Full text] [doi: 10.2196/jmir.3923] [Medline: 25794109]
12. Lee H, Park J, Choi SW, Yoon YE, Park HE, Lee SE, et al. Impact of a Telehealth Program With Voice Recognition Technology in Patients With Chronic Heart Failure: Feasibility Study. JMIR Mhealth Uhealth 2017 Oct 02;5(10):e127 [FREE Full text] [doi: 10.2196/mhealth.7058] [Medline: 28970189]
13. Kim KM, Park KS, Lee HJ, Lee YH, Bae JS, Lee YJ, et al. Efficacy of a New Medical Information system, Ubiquitous Healthcare Service with Voice Inception Technique in Elderly Diabetic Patients. Sci Rep 2015 Dec 11;5:18214 [FREE Full text] [doi: 10.1038/srep18214] [Medline: 26658492]
14. Glanz K, Beck AD, Bundy L, Primo S, Lynn MJ, Cleveland J, et al. Impact of a health communication intervention to improve glaucoma treatment adherence. Results of the interactive study to increase glaucoma adherence to treatment trial. Arch Ophthalmol 2012 Oct;130(10):1252-1258 [FREE Full text] [doi: 10.1001/archophthalmol.2012.1607] [Medline: 22688429]
15. Hezarjaribi N, Mazrouee S, Ghasemzadeh H. Speech2Health: A Mobile Framework for Monitoring Dietary Composition From Spoken Data. IEEE J Biomed Health Inform 2018 Jan;22(1):252-264. [doi: 10.1109/JBHI.2017.2709333] [Medline: 29300701]
16. Koebnick C, Wagner K, Thielecke F, Moeseneder J, Hoehne A, Franke A, et al. Validation of a simplified physical activity record by doubly labeled water technique. Int J Obes (Lond) 2005 Mar;29(3):302-309. [doi: 10.1038/sj.ijo.0802882] [Medline: 15672111]
17. Rohlfing ML, Buckley DP, Piraquive J, Stepp CE, Tracy LF. Hey Siri: How Effective are Common Voice Recognition Systems at Recognizing Dysphonic Voices? Laryngoscope 2020 Sep 19. [doi: 10.1002/lary.29082] [Medline: 32949415]
18. Tan S, Fatehi F. Sweet Talking: Voice Technology and Virtual Assistants in Clinical Diabetes Management. Stud Health Technol Inform 2019 Aug 21;264:1787-1788. [doi: 10.3233/SHTI190648] [Medline: 31438344]
19. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of Conversational Agents to Health and Lifestyle Prompts: Investigation of Appropriateness and Presentation Structures. J Med Internet Res 2020 Feb 09;22(2):e15823 [FREE Full text] [doi: 10.2196/15823] [Medline: 32039810]
20. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 08;1(8476):307-310. [Medline: 2868172]
22. Namba H, Kurosaka Y, Minato K, Yamada Y, Kimura M. Validation of a web-based physical activity measurement system using a tri-axial accelerometer. Research in Exercise Epidemiology. 2015 Mar. URL: http://jaee.umin.jp/REE/J/17_1_19.pdf [accessed 2020-12-29]

23. Yamazaki T, Gen-No H, Kamijo Y, Okazaki K, Masuki S, Nose H. A new device to estimate VO2 during incline walking by accelerometry and barometry. Med Sci Sports Exerc 2009 Dec;41(12):2213-2219. [doi: 10.1249/MSS.0b013e3181a9c452] [Medline: 19920753]
24. Ohkawara K, Ishikawa-Takata K, Park JH, Tabata I, Tanaka S. How much locomotive activity is needed for an active physical activity level: analysis of total step counts. BMC Res Notes 2011 Nov 24;4:512 [FREE Full text] [doi: 10.1186/1756-0500-4-512] [Medline: 22114990]
25. Alharbi M, Straiton N, Smith S, Neubeck L, Gallagher R. Data management and wearables in older adults: A systematic review. Maturitas 2019 Jun;124:100-110. [doi: 10.1016/j.maturitas.2019.03.012] [Medline: 30910279]

Abbreviations

METs: metabolic equivalents

Edited by G Eysenbach; submitted 03.04.20; peer-reviewed by C Cleland, B Woo, M Miyachi; comments to author 20.04.20; revised version received 15.06.20; accepted 09.01.21; published 21.01.21.

Please cite as:
Namba H
Physical Activity Evaluation Using a Voice Recognition App: Development and Validation Study
JMIR Biomed Eng 2021;6(1):e19088
URL: http://biomedeng.jmir.org/2021/1/e19088/
doi:10.2196/19088
PMID:

Original Paper

Subspace Clustering of Physiological Data From Acute Traumatic Brain Injury Patients: Retrospective Analysis Based on the PROTECT III Trial

Sina Ehsani[1], MSc; Chandan K Reddy[2], PhD; Brandon Foreman[3], MD; Jonathan Ratcliff[4], MD, MPH; Vignesh Subbian[1,5], PhD

[1]Department of Systems and Industrial Engineering, College of Engineering, The University of Arizona, Tucson, AZ, United States
[2]Department of Computer Science, Virginia Polytechnic Institute and State University, Arlington, VA, United States
[3]Department of Neurology & Rehabilitation Medicine, University of Cincinnati, Cincinnati, OH, United States
[4]Department of Emergency Medicine, Emory University School of Medicine, Emory University, Atlanta, GA, United States
[5]Department of Biomedical Engineering, College of Engineering, The University of Arizona, Tucson, AZ, United States

Corresponding Author:
Vignesh Subbian, PhD
Department of Biomedical Engineering
College of Engineering, The University of Arizona
1127 E James E Rogers Way
Tucson, AZ, 85721
United States
Phone: 1 5206216559
Email: vsubbian@email.arizona.edu

Abstract

Background: With advances in digital health technologies and proliferation of biomedical data in recent years, applications of machine learning in health care and medicine have gained considerable attention. While inpatient settings are equipped to generate rich clinical data from patients, there is a dearth of actionable information that can be used for pursuing secondary research for specific clinical conditions.

Objective: This study focused on applying unsupervised machine learning techniques for traumatic brain injury (TBI), which is the leading cause of death and disability among children and adults aged less than 44 years. Specifically, we present a case study to demonstrate the feasibility and applicability of subspace clustering techniques for extracting patterns from data collected from TBI patients.

Methods: Data for this study were obtained from the Progesterone for Traumatic Brain Injury, Experimental Clinical Treatment–Phase III (PROTECT III) trial, which included a cohort of 882 TBI patients. We applied subspace-clustering methods (density-based, cell-based, and clustering-oriented methods) to this data set and compared the performance of the different clustering methods.

Results: The analyses showed the following three clusters of laboratory physiological data: (1) international normalized ratio (INR), (2) INR, chloride, and creatinine, and (3) hemoglobin and hematocrit. While all subclustering algorithms had a reasonable accuracy in classifying patients by mortality status, the density-based algorithm had a higher F1 score and coverage.

Conclusions: Clustering approaches serve as an important step for phenotype definition and validation in clinical domains such as TBI, where patient and injury heterogeneity are among the major reasons for failure of clinical trials. The results from this study provide a foundation to develop scalable clustering algorithms for further research and validation.

(JMIR Biomed Eng 2021;6(1):e24698) doi:10.2196/24698

KEYWORDS

cluster analysis; unsupervised machine learning; traumatic brain injury

Introduction

Traumatic brain injury (TBI) is broadly defined as disruption in normal brain function or other evidence of brain pathology as a result of mechanical force directed at the head or a rapid acceleration/deceleration event. TBI is the most common cause of death and disability in children and adults aged less than 44 years [1]. However, there has been little change in TBI-related deaths despite advancements in care delivery [2]. Additionally, a major challenge to both TBI-related clinical research and acute care is reliably identifying candidates for targeted interventions [3]. While there have been substantial advances in technological and computational approaches to TBI phenotyping [4-6], there is still a dearth of actionable information that can be used for pursing secondary clinical research in this domain.

Existing approaches to stratification of patients based on clinical presentation does not adequately address the heterogenous nature of TBI, whereas data mining and machine learning techniques have shown promise in identifying subgroups [5], predicting outcomes [7], and prognosticating among TBI patients [8]. In particular, clustering-based techniques serve as an important step for phenotype definition and have the potential to uncover previously unrecognized relationships between various physiologic variables [9]. For example, in other clinical domains, traditional cluster analyses have been helpful in identifying unique subgroups of patients. These studies include application of k-means cluster analysis for identifying distinct phenotypes of asthma patients [10], as well as using hierarchical clustering to identify both new and known relationships between physiologic variables collected from critically ill patients [9]. In this study, we applied *subspace clustering* (or subclustering) methods on physiologic data collected from TBI patients and compared the performance of different subspace clustering methods (density-based, cell-based, and clustering-oriented methods). The rationale for applying subspace clustering over traditional clustering methods (eg, k-means) is the ability to account for the multiple low-dimensional subspace structure of higher dimensional data [11]. In terms of critical illnesses, such as acute TBI, we hypothesize that the complex latent relationships between various physiologic variables are better represented in subspaces and thus better captured by subclustering methods than traditional methods that are often limited to spatial proximity of data points in individual clusters.

Methods

Data Source

Data for this study were obtained from the Progesterone for Traumatic Brain Injury, Experimental Clinical Treatment–Phase III (PROTECT III) study. The PROTECT III trial included a cohort of 882 TBI patients [12], who were originally recruited for a randomized clinical trial to study the effect of progesterone on patients with acute TBI. Patients were randomly assigned to a treatment group that received progesterone within 4 hours of injury or placebo. While the PROTECT III clinical trial showed that there was no difference in patients between the two study groups, the longitudinal data from the trial were made available for secondary analyses and continued research.

This data set included patient demographics, baseline assessment data, 6-month outcome data, including the Glasgow Outcome Scale Extended scale, and mortality status. The temporal data in this study included laboratory test results for the first 7 days of stay. Other clinical and radiologic data were not included in this analysis. Deidentified data were obtained in collaboration with the PROTECT III investigators and are now available through the Federal Interagency Traumatic Brain Injury Research informatics system. The inclusion criteria for this analysis were as follows: (1) subjects were alive for at least 3 days, (2) subjects were not excluded from the parent study, and (3) their baseline laboratory results were available. The 3-day criterion is used because subjects who do not survive for at least 3 days or 72 hours have likely experienced devastating brain injury or other forms of severe trauma, which often require aggressive interventions [13]. Additionally, the first 72 hours of observation is the time interval used for determining the preliminary effect of the injury and is thus recommended for valid prognostication [14,15].

Subspace Clustering

Subspace or projective clustering is a clustering method that emphasizes on clustering in subspaces of high-dimensional spaces, that is, it tries to find clusters in smaller subspaces and builds up to form larger clusters by using overlapping subspaces [16]. Subspace clustering can be classified into the following three main categories: density-based approaches, cell-based approaches, and clustering-oriented approaches. Density-based approaches define subspaces in dense areas [17]. In cell-based approaches, subspaces are formed by predefining the width of grid cells and the number of objects within each cell [18]. Clustering-oriented approaches define properties of the entire set of clusters, as opposed to definition of the cluster itself, and then assign objects to the cluster with the most relevant properties [19].

Density-Based Approach

One of the commonly used clustering algorithms is density-based spatial clustering of applications with noise (DBSCAN) [20]. The key idea of DBSCAN is that after detecting a cluster using density-based grids, it looks at the neighborhood of each cluster point in a defined radius; any point that exists in this radius is contributed to the cluster.

Every cluster C in a subspace projection is defined by a set of objects O, that is a subset of database DB and a set of relevant dimensions S out of the set of all dimensions D.

A clustering result R is a set of clusters k found in the respective subspace projections as follows:

A density-based subspace cluster (O, S) in a two-dimensional space is defined with respect to parameters *minPoints* and ε–*neighborhood* $N_\varepsilon(p) = \{q \in DB \mid dist^S(p, q) \leq \varepsilon\}$, where $dist^S$ represents a distance function constrained to the dimensions S, as follows [20]:

(1) ε–*neighborhood* of a point: Let p and q be two points of the sample, and the distance equation between these two points is defined by *dist* (*p*, *q*). The distance could be defined with Manhattan distance, Euclidean distance, or other different distance methods. The ε–*neighborhood* of a point is defined as follows:

(2) *Directly density reachable:* A point p is directly density reachable from a point q with respect to ε and *MinPts* if

(3) *Density reachable:* A point p is density reachable from a point q with respect to ε and *MinPts* if all the points in a chain of points (including q and p) are directly density reachable from each another.

(4) *Density connected:* A point p is density connected to a point q if only there is point o, which both p and q are density reachable from.

(5) *Noise:* The sets of points in database DB that are not assigned to any cluster are called noise.

To find a cluster, the DBSCAN algorithm starts with a random point p and finds all density reachable points with respect to ε and *MinPts*. DBSCAN also merges two clusters together if the distance between two sets of points is defined as follows:

Density-connected subspace clustering (SUBCLU) is a greedy algorithm built on an adaption of the DBSCAN algorithm for high-dimensional data. It computes all density-connected sets hidden in subspaces of high-dimensional data. Studies have shown that SUBCLU can outperform other subspace clustering methods based on different measures [18,20,21]. SUBCLU is capable of detecting arbitrarily shaped clusters using the DBSCAN algorithm in subspaces. To use DBSCAN in each subspace, let DB be a d-dimensional feature vectors data set with n objects $DB \subseteq R^d$. Let $A = \{a_1, a_2, \ldots, a_d\}$ be the set of all attributes a of DB. Any subset $S \subseteq A$ is called a subspace. The projection of an object o into a subspace S is denoted by $\pi_s(o)$, and the distance function is denoted by d*ist*. For instance, the ε–*neighborhood* of o in S is the same as DBSCAN, but projected in S subspace as follows:

The core object is defined as follows:

The algorithm begins by generating all one-dimensional clusters using the DBSCAN algorithm. For each detected cluster, it checks whether the cluster also exists in higher dimensions or not. For each k-dimensional subspace $S \in S_k$, the algorithm searches all other k-dimensional subspaces $T \in S_k$ having (k-1) attributes in common and combines them to generate (k + 1)-dimensional candidate subspaces. Based on prior studies [21], we chose the *Midpts* to be in the range from 8 to 128 (with five steps) and the ε–*neighborhood* to be from 0.01 to 0.25 (with nine steps). For this study, the initial *Midpts* value was set to 8 and increased by 30 after each run until it reached 128. The ε–*neighborhood* value was initially set to 0.01 and was increased by 0.03 until a maximum of 0.25.

Cell-Based Approach

Cell-based clustering is centered on cell estimate of the data space. The width of the cells is parametrized by w. A cluster R contains a set of cells, and each cell contains at least τ number of data points. One of the popular cell-based methods is the *MineClus* algorithm, which describes each of these cells as the objects of the cluster by a hypercube with width w. These hypercubes are arbitrarily positioned to define a region with frequent data patterns.

A cell-based subspace cluster (O, S) is defined with respect to the minimum number of objectives τ in cells CS of w width specified by intervals I_i per dimension $\forall i \in S$. Each interval is part of the common domain $I_i = [l_i \ldots u_i] \subseteq [0 \ldots v]$ with lower and upper bounds l_i and u_i. For all irrelevant dimensions $\forall j \in D \backslash S$, the interval is the full domain $I_j = [0 \ldots v]$, and the cluster objects $O = \{o | o \in DB \cap CS\}$ fulfill $|O| \geq \tau$ [21].

Clustering-Oriented Approach

Clustering-oriented approaches focus on the clustering result R by specifying objective functions. PROCLUS [22], one of the first top-down subspace clustering algorithms, forms the clusters first and iteratively improves the clustering model. In the PROCLUS algorithm, the number of clusters and the average dimensionality are used as parameters, and data are partitioned into k clusters with the average dimension being l. A clustering-oriented approach is defined with respect to objective function $f(R)$, which is based on the entire clustering result R, and an optimal value parameter $optF$ is a result set R with $f(R) = optF$.

In this case study, we adapted the aforementioned subspace clustering techniques to analyze the PROTECT III data set. Analyses were performed using OpenSubspace [21,23], an open-source framework that extends the WEKA platform [24,25]. All laboratory values were normalized to a scale between 0 and 10 before applying the algorithms.

Evaluation

Evaluation of unsupervised learning methods, such as cluster analysis, is typically informed by domain expertise. For this work, two clinicians (coauthors of this work [BF and JR]) independently evaluated the results and validated the clusters based on their experiences in the clinical management of TBI as well as clinical research in neurotrauma. The informatician on the team (VS) coordinated the clinician validation process. Mechanistic interpretations for potential markers or associations indicated by clusters were offered based on clinical expertise. To demonstrate alignment of subclustering solutions to a clinical outcome, mortality at 6 months after injury was examined.

Additional evaluation metrics used in this study included F1 score, entropy, coverage, average dimension, and accuracy of classification. The F1 value, a common metric for evaluating

clustering algorithms, is defined as the harmonic mean of precision and recall. Entropy is a metric that accounts for clarity of clustering [26]. Coverage characterizes how clusters cover the input data space. Average dimension is the average of number of dimensions that the clusters cover in each run. Accuracy of classification compares the patterns detected in the model in relation to labeled data, such as outcome. Here, the mortality status of TBI patients was used as the outcome. Finally, the performance of subspace clustering algorithms was compared to traditional k-means clustering, which partitions *n* data points into *k* clusters, placing each observation in one of the clusters with neared mean representation. While k-means rely on distance metrics and proximity of observations within individual clusters, subspace methods group data points based on their lower-dimensional subspaces. Given these distinct algorithmic differences between subspace and k-means clustering in formulation of the clustering problem, a direct comparison of the clusters formed and interpretation of clusters may not be appropriate. Instead, we report performance metrics for comparison purposes.

Results

Subject Characteristics

Of the 882 study subjects in the parent PROTECT III trial, 643 subjects met the inclusion criteria for this study. Table 1 shows the characteristics of these study subjects at baseline. Ten different laboratory results were used in this study, including blood serum chemistry and hematology results at baseline (Table 2). Coagulation tests, such as the international normalized ratio (INR) and activated partial thromboplastin time, were also included.

Table 1. Patient characteristics.

Characteristic	Value (N=643)
Age (years), mean (range)	34 (17-93)
Male sex, n (%)	475 (73.9)
Black people, n (%)	105 (16.3)
Hispanic people, n (%)	97 (15.1)
Cause of injury, n (%)	
Motor vehicle accident	242 (37.7)
Motorcycle or scooter accident	121 (18.8)
Pedestrian struck by a moving vehicle	78 (12.1)
Other	202 (31.4)

Table 2. Laboratory results.

Laboratory parameter	Value, mean (range)
Glucose, mg/dL	151.6 (68-554)
Creatinine, mg/dL	1.015 (0.3-4.2)
Potassium, mmol/L	3.667 (1.5-5.8)
Sodium, mmol/L	139.8 (125-157)
Chloride, mmol/L	105.4 (88-130)
Bicarbonate, mmol/L	22.77 (8.0-34.0)
Hemoglobin, g/dL	13.66 (4.9-18.6)
Hematocrit, %	40.31 (14.6-54.2)
Total white blood cell count, $\times 10^9$/L	14.85 (3.2-41.40)
Platelet count, $\times 10^3/mm^3$	249.7 (51-700)

Application of Subspace Clustering Algorithms to PROTECT III Data

All three types of subspace clustering algorithms (density-based [SUBCLU], cell-based [MineClus], and clustering-oriented [PROCLUS] algorithms) were applied to the PROTECT III data set. The INR, which characterizes the clotting tendency of blood, was identified as one of the distinct clusters. This could represent coagulopathy, a marker of secondary insult in TBI patients [27]. For example, coagulopathy is associated with increased risk of ongoing bleeding and expansion of any intracranial traumatic hemorrhage. One of the clinicians also noted that progressive coagulopathy, which is resistant to correction, is further associated with worse outcomes in TBI patients.

The clustering models also showed a strong relation among INR, chloride, and creatinine. Both clinicians noted and agreed

that elevations in chloride levels are often related to fluid administration for treatment of intracranial hypertension or a shock (hypoperfusion) state. Therefore, elevations in these parameters may also be indicators that the clinical team needed to treat a sicker patient more aggressively. Creatinine may be elevated at baseline in patients with chronic illness or may indicate that secondary kidney failure may impact outcome as a complication of TBI. One of the clinicians noted that there is a further relationship between elevated chloride and subsequent elevation in creatinine, though a wide variety of insults may lead to elevations in creatinine. Finally, in models with higher dimensions, a relationship between the hemoglobin level and hematocrit percentage was noted. This relationship is quite intuitive, given that both measure similar properties. These observations are demonstrated in Figure 1 and Figure 2.

Figure 1.

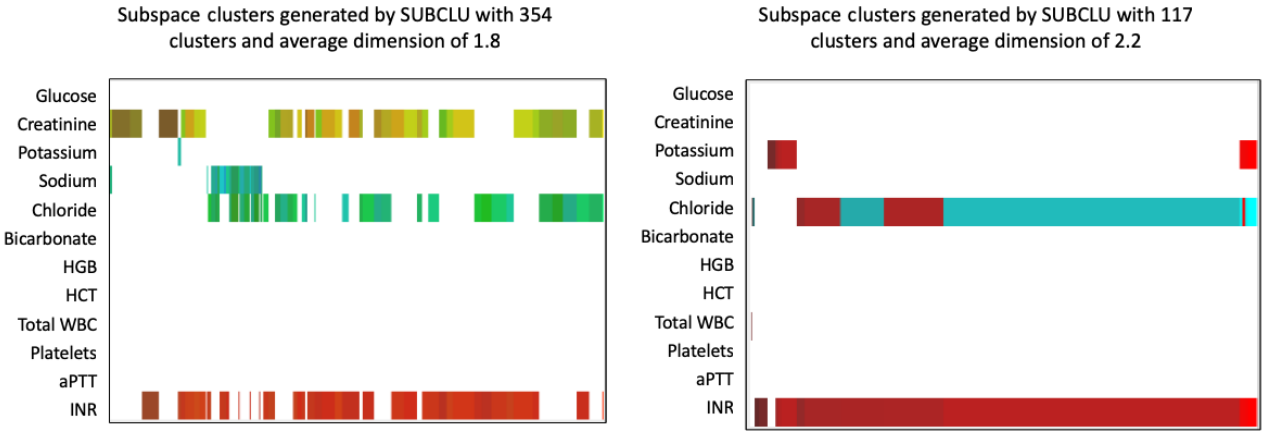
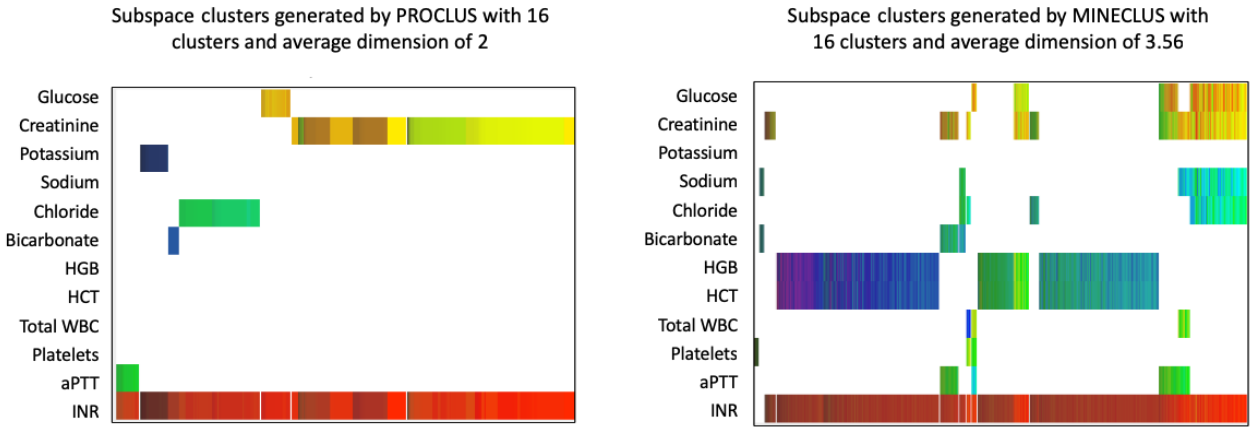


Figure 2.



The performances of different subspace clustering methods as well as the traditional k-means algorithm on the PROTECT III data set were compared using various evaluation metrics and the mortality status as the outcome (Table 3). The density-based algorithm (SUBCLU) had higher F1 and coverage. The cell-based algorithm (MineClus) had good performance on the F1 measure while having lower number of clusters. The clustering-oriented algorithm (PROCLUS) performed reasonably in terms of accuracy and entropy, while it had the lowest F1 compared to other models. K-means, given its simplicity, was the fastest algorithm, but performed worst in all other metrics.

Table 3. Comparison of subspace clustering algorithms.

Evaluation metric	Density-based algorithm (SUBCLU[a]), min-max	Cell-based algorithm (MineClus), min-max	Clustering-oriented algorithm (PROCLUS), min-max	K-means, min-max
F1	0.45-0.69	0.42-0.64	0.36-0.44	0.21-0.30
Entropy	0.45-0.59	0.44-0.55	0.48-0.63	0.51-0.53
Coverage	0.9-1	0.78-0.97	0.43-0.82	0.11-1
Number of clusters	6-1024	6-64	8-32	2-32
Average dimensions	2.3-9	3.2-6.1	2-9	12
Accuracy (%)	81-88	88-88	88-88	51-63
Runtime (s)	367-745,785	58-194	155-402	0.07

[a]SUBCLU: density-connected subspace clustering.

Discussion

Currently, clinical data used to predict outcomes after TBI come from modeling and validation performed across two older clinical studies in TBI encompassing more than 15,000 patients [28,29]. The covariates that were significant in these prior regression models included glucose and hemoglobin, in addition to clinical predictors such as age and clinical examination. However, the area under the curve of these models is suboptimal. Clusters of data may also incorporate clinical knowledge such as the observation that the combination of lactic acidosis, hypothermia, and coagulopathy at presentation after major trauma imparts poor prognosis. Furthermore, many of these patients do not survive the 72 hours required for inclusion in the current analysis.

Lack of access to multiple data sources has limited further external validation of the proposed methods. Nonetheless, clinician validation is important to inform analyses of data from ongoing observational studies and provide valuable insights into the development of clinically relevant tools for TBI management. This case study serves as a demonstration for such applications. As a next step, focus on temporal data and methods for time-series analyses are warranted.

Conclusion

This study explored the application and feasibility of subspace clustering techniques for a specific clinical condition, TBI, using clinical data from a randomized clinical trial. The analyses showed the following three clusters of laboratory physiological data: (1) INR, (2) INR, chloride, and creatinine, and (3) hemoglobin and hematocrit. While all subclustering algorithms had a reasonable accuracy in classifying patients by mortality status, the density-based algorithm had a higher F1 score and coverage. Clustering approaches serve as an important step for phenotype definition and validation in clinical domains, such as TBI, where patient and injury heterogeneity are among the major reasons for failure of clinical trials. Results from this study also provide a foundation to develop scalable clustering algorithms for further research and validation.

Acknowledgments

The authors acknowledge the Progesterone for Traumatic Brain Injury, Experimental Clinical Treatment–Phase III (PROTECT III) principal investigator, Dr David Wright (Emory University), for providing deidentified data for this study. This material is based upon work supported by the National Science Foundation under grants #1838730 and #1838745. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Conflicts of Interest

None declared.

References

1. Taylor CA, Bell JM, Breiding MJ, Xu L. Traumatic Brain Injury-Related Emergency Department Visits, Hospitalizations, and Deaths - United States, 2007 and 2013. MMWR Surveill Summ 2017 Mar 17;66(9):1-16 [FREE Full text] [doi: 10.15585/mmwr.ss6609a1] [Medline: 28301451]
2. Ghajar J. Traumatic brain injury. The Lancet 2000 Sep;356(9233):923-929. [doi: 10.1016/s0140-6736(00)02689-1]
3. Subbian V, Ratcliff JJ, Korfhagen JJ, Hart KW, Meunier JM, Shaw GJ, et al. A Novel Tool for Evaluation of Mild Traumatic Brain Injury Patients in the Emergency Department: Does Robotic Assessment of Neuromotor Performance Following Injury Predict the Presence of Postconcussion Symptoms at Follow-up? Acad Emerg Med 2016 Apr 21;23(4):382-392 [FREE Full text] [doi: 10.1111/acem.12906] [Medline: 26806406]

4. Subbian V, Ratcliff J, Meunier J, Korfhagen J, Beyette F, Shaw G. Integration of New Technology for Research in the Emergency Department: Feasibility of Deploying a Robotic Assessment Tool for Mild Traumatic Brain Injury Evaluation. IEEE J. Transl. Eng. Health Med 2015 Apr 23:1-1. [doi: 10.1109/jtehm.2015.2424224]
5. Nielson JL, Cooper SR, Yue JK, Sorani MD, Inoue T, Yuh EL, TRACK-TBI Investigators. Uncovering precision phenotype-biomarker associations in traumatic brain injury using topological data analysis. PLoS One 2017 Mar 3;12(3):e0169490 [FREE Full text] [doi: 10.1371/journal.pone.0169490] [Medline: 28257413]
6. Subbian V, Meunier J, Korfhagen J, Ratcliff J, Shaw G, Beyette F. Quantitative assessment of post-concussion syndrome following mild traumatic brain injury using robotic technology. 2014 Presented at: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 26-30, 2014; Chicago, IL, USA p. 5353-5356. [doi: 10.1109/EMBC.2014.6944835]
7. de Toledo P, Rios PM, Ledezma A, Sanchis A, Alen JF, Lagares A. Predicting the outcome of patients with subarachnoid hemorrhage using machine learning techniques. IEEE Trans Inf Technol Biomed 2009 Sep;13(5):794-801. [doi: 10.1109/TITB.2009.2020434] [Medline: 19369161]
8. Hemphill JC, Andrews P, De Georgia M. Multimodal monitoring and neurocritical care bioinformatics. Nat Rev Neurol 2011 Jul 12;7(8):451-460. [doi: 10.1038/nrneurol.2011.101] [Medline: 21750522]
9. Cohen MJ, Grossman AD, Morabito D, Knudson MM, Butte AJ, Manley GT. Identification of complex metabolic states in critically injured patients using bioinformatic cluster analysis. Crit Care 2010;14(1):R10 [FREE Full text] [doi: 10.1186/cc8864] [Medline: 20122274]
10. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster Analysis and Clinical Asthma Phenotypes. Am J Respir Crit Care Med 2008 Aug;178(3):218-224. [doi: 10.1164/rccm.200711-1754oc]
11. Elhamifar E, Vidal R. Sparse Subspace Clustering: Algorithm, Theory, and Applications. IEEE Trans. Pattern Anal. Mach. Intell 2013 Nov;35(11):2765-2781. [doi: 10.1109/tpami.2013.57]
12. Wright DW, Yeatts SD, Silbergleit R, Palesch YY, Hertzberg VS, Frankel M, et al. Very Early Administration of Progesterone for Acute Traumatic Brain Injury. N Engl J Med 2014 Dec 25;371(26):2457-2466. [doi: 10.1056/nejmoa1404304]
13. Jochems D, van Wessem KJP, Houwert RM, Brouwers HB, Dankbaar JW, van Es MA, et al. Outcome in Patients with Isolated Moderate to Severe Traumatic Brain Injury. Crit Care Res Pract 2018 Sep 23;2018:3769418-3769417 [FREE Full text] [doi: 10.1155/2018/3769418] [Medline: 30345113]
14. Eriksson EA, Barletta JF, Figueroa BE, Bonnell BW, Sloffer CA, Vanderkolk WE, et al. The first 72 hours of brain tissue oxygenation predicts patient survival with traumatic brain injury. The Journal of Trauma and Acute Care Surgery 2012;72(5):1345-1349. [doi: 10.1097/ta.0b013e318249a0f4]
15. Souter MJ, Blissitt PA, Blosser S, Bonomo J, Greer D, Jichici D, et al. Recommendations for the Critical Care Management of Devastating Brain Injury: Prognostication, Psychosocial, and Ethical Management : A Position Statement for Healthcare Professionals from the Neurocritical Care Society. Neurocrit Care 2015 Aug 18;23(1):4-13. [doi: 10.1007/s12028-015-0137-6] [Medline: 25894452]
16. Theeramunkong T, Kijsirikul B, Cercone N, Ho TB. Advances in Knowledge Discovery and Data Mining. Berlin, Germany: Springer; 2009.
17. Chu Y, Huang J, Chuang K, Yang D, Chen M. Density Conscious Subspace Clustering for High-Dimensional Data. IEEE Trans. Knowl. Data Eng 2010 Jan;22(1):16-30. [doi: 10.1109/TKDE.2008.224]
18. Sembiring R, Zain J. Cluster Evaluation of Density Based Subspace Clustering. J Comput 2010;2(11):2151-9617 [FREE Full text]
19. Aggarwal CC, Wolf JL, Yu PS, Procopiuc C, Park JS. Fast algorithms for projected clustering. SIGMOD Rec 1999 Jun 01;28(2):61-72. [doi: 10.1145/304181.304188]
20. Ester M, Kriegel H, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. 1996 Aug Presented at: Second International Conference on Knowledge Discovery and Data Mining; August 2-4, 1996; Portland, Oregon p. 226-231. [doi: 10.5555/3001460.3001507]
21. Müller E, Günnemann S, Assent I, Seidl T. Evaluating clustering in subspace projections of high dimensional data. Proc. VLDB Endow 2009 Aug 01;2(1):1270-1281. [doi: 10.14778/1687627.1687770]
22. Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data for data mining applications. 1998 Jun Presented at: 1998 ACM SIGMOD International Conference on Management of Data; June 1998; Seattle, WA, USA p. 94-105. [doi: 10.1145/276305.276314]
23. Assent I, Müller E, Krieger R, Jansen T, Seidl T. Pleiades: Subscpace Clustering and Evaluation. In: Daelemans W, Goethals B, Morik K, editors. Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2008. Lecture Notes in Computer Science, vol 5212. Berlin, Heidelberg: Springer; 2008.
24. Witten I, Frank E, Mark H. Data Mining: Practical Machine Learning Tools and Techniques. Burlington, MA: Morgan Kaufmann; 2011.
25. Frank E, Hall M, Witten I. The WEKA Workbench. In: Data Mining (Fourth Edition). Burlington, MA: Morgan Kaufmann; 2016:553-571.
26. Sequeira K, Zaki M. SCHISM: a new approach to interesting subspace mining. IJBIDM 2005;1(2):137. [doi: 10.1504/ijbidm.2005.008360]

27. Stein SC, Smith DH. Coagulopathy in Traumatic Brain Injury. NCC 2004;1(4):479-488. [doi: 10.1385/ncc:1:4:479]
28. Murray GD, Butcher I, McHugh GS, Lu J, Mushkudiani NA, Maas AI, et al. Multivariable prognostic analysis in traumatic brain injury: results from the IMPACT study. J Neurotrauma 2007 Feb;24(2):329-337. [doi: 10.1089/neu.2006.0035] [Medline: 17375997]
29. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med 2008 Aug 05;5(8):e165; discussion e165 [FREE Full text] [doi: 10.1371/journal.pmed.0050165] [Medline: 18684008]

Abbreviations

DBSCAN: density-based spatial clustering of applications with noise
INR: international normalized ratio
PROTECT III: Progesterone for Traumatic Brain Injury, Experimental Clinical Treatment–Phase III
SUBCLU: density-connected subspace clustering
TBI: traumatic brain injury

Edited by G Eysenbach; submitted 01.10.20; peer-reviewed by H Zhang, G Lim; comments to author 23.10.20; revised version received 31.12.20; accepted 16.01.21; published 02.02.21.

Please cite as:
Ehsani S, Reddy CK, Foreman B, Ratcliff J, Subbian V
Subspace Clustering of Physiological Data From Acute Traumatic Brain Injury Patients: Retrospective Analysis Based on the PROTECT III Trial
JMIR Biomed Eng 2021;6(1):e24698
URL: http://biomedeng.jmir.org/2021/1/e24698/
doi:10.2196/24698
PMID:

Original Paper

Point-of-Care Quantification of Serum Alpha-Fetoprotein for Screening Birth Defects in Resource-Limited Settings: Proof-of-Concept Study

Balaji Srinivasan[1], PhD; Julia L Finkelstein[1], MPH, SM, ScD; David Erickson[2], PhD; Saurabh Mehta[1], MBBS, ScD

[1]Division of Nutritional Sciences, Cornell University, Ithaca, NY, United States

[2]Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, United States

Corresponding Author:
Saurabh Mehta, MBBS, ScD
Division of Nutritional Sciences
Cornell University
314 Savage Hall
Ithaca, NY, 14850
United States
Phone: 1 607 255 2640
Email: smehta@cornell.edu

Abstract

Background: Maternal serum alpha-fetoprotein (MSAFP) concentration typically increases during pregnancy and is routinely measured during the second trimester as a part of screening for fetal neural tube defects and Down syndrome. However, most pregnancy screening tests are not available in the settings they are needed the most. A mobile device–enabled technology based on MSAFP for screening birth defects could enable the rapid screening and triage of high-risk pregnancies, especially where maternal serum screening and fetal ultrasound scan facilities are not easily accessible. Shifting the approach from clinic- and laboratory-dependent care to a mobile platform based on our point-of-care approach will enable translation to resource-limited settings and the global health care market.

Objective: The objective of this study is to develop and perform proof-of-concept testing of a lateral flow immunoassay on a mobile platform for rapid, point-of-care quantification of serum alpha-fetoprotein (AFP) levels, from a drop of human serum, within a few minutes.

Methods: The development of the immunoassay involved the selection of commercially available antibodies and optimization of their concentrations by an iterative method to achieve the required detection limits. We compared the performance of our method with that of commercially obtained human serum samples, with known AFP concentrations quantified by the Abbott ARCHITECT chemiluminescent magnetic microparticle immunoassay (CMIA).

Results: We tested commercially obtained serum samples (N=20) with concentrations ranging from 2.2 to 446 ng/mL to compare the results of our point-of-care assay with results from the Abbott ARCHITECT CMIA. A correlation of 0.98 (P<.001) was observed on preliminary testing and comparison with the CMIA. The detection range of our point-of-care assay covers the range of maternal serum AFP levels observed during pregnancy.

Conclusions: The preliminary test results from the AFP test on the mobile platform performed in this study represent a proof of concept that will pave the way for our future work focused on developing a mobile device–enabled quad-screen point-of-care testing with the potential to enable the screening of high-risk pregnancies in various settings. The AFP test on the mobile platform can be applied to enable screening for high-risk pregnancies, within a few minutes, at the point of care even in remote areas where maternal serum tests and fetal ultrasound scans are not easily accessible; assessment of whether clinical follow-up and diagnostic testing may be needed after a positive initial screening evaluation; and development of surveillance tools for birth defects.

(JMIR Biomed Eng 2021;6(1):e23527) doi:10.2196/23527

KEYWORDS

alpha-fetoprotein; point-of-care testing; screening; neural tube defects; mobile phone

Introduction

Background

Neural tube defects (NTDs) are one of the most common and debilitating birth defects documented in the United States and globally. These defects arise when the neural folds fail to fuse entirely during early embryogenesis. Outcomes for an infant born with an NTD vary by severity and the affected region of the neural fold [1,2]. Anencephaly and other open NTDs that affect the infant's brain are incompatible with life, further leading to fetal loss during pregnancy or death soon after birth, whereas those affecting the spine can lead to serious neurological and physical impairment [1,2]. Closed NTDs have a layer of skin covering the defect and are less severe; however, they can require surgery and cause motor and sensory impairments [1,2]. It is estimated that there are over 260,000 NTD cases globally per year, with the burden ranging from 1 to 80 per 10,000 births globally, leading to 70,800 deaths and loss of 6.4 million disability-adjusted life years [3-8]. The total lifetime direct cost of care [9] for a child born with spina bifida in the United States is estimated to be US $791,900.

In cases with elevated maternal serum alpha-fetoprotein (MSAFP) levels, an ultrasound examination is recommended to further determine whether an NTD or another anomaly associated with elevated MSAFP levels is present, in addition to confirming the gestational age, fetal viability, and number of fetuses. If ultrasound findings are ambiguous or show an apparently normal fetus, then genetic counseling and further evaluation through amniocentesis are usually performed. Screening and an early diagnosis of affected pregnancies provides parents with the options for diagnostic and clinical follow-up, interventions during pregnancy, and preparation for the birth of an affected child, including associated medical costs for surgical or nonsurgical treatments. In addition, prenatal NTD detection also informs clinical triage regarding the optimal timing, route, and site of delivery (eg, referral to high-risk or tertiary care hospitals, cesarean delivery). In the United States, the universal screening for NTDs is supported by the American College of Obstetricians and Gynecologists (ACOG) [10] and the American College of Medical Genetics [11], while emphasizing the need to provide adequate counseling and follow-up services. NTD screening approaches, such as biomarker assessment (measurement of MSAFP) and ultrasound examinations (anatomical), also enable the screening and detection of fetal abnormalities other than NTDs and inform clinical care and follow-up [12,13].

MSAFP is one of the biomarkers included in the triple screen test for pregnant women. The triple screen test is a maternal blood screening test that looks for 3 distinct analytes—MSAFP, human chorionic gonadotropin (hCG), and unconjugated estriol—to identify women who are at an increased risk of having a baby with NTD or trisomy syndrome. The triple screening is recommended between 15 and 22 weeks of gestation and is most accurate if performed between 16 and 18 weeks of gestation. Fetal serum alpha-fetoprotein (AFP) concentrations peak at 10-13 weeks' gestation and decline progressively until term, whereas maternal levels peak during the third trimester [14]. Elevated MSAFP levels with a screen positive rate of 5% or less can detect 75%-90% of NTDs and ≥95% of anencephaly [11], which is the most severe type of NTD that is incompatible with life [1]. Abnormally low AFP values [15] (most often a median value of <0.5) are associated with Down syndrome and other chromosomal abnormalities. MSAFP levels may also detect 85% of the ventral wall defects [11].

MSAFP levels are typically quantified using immunoassay-based methods. Conventional immunoassays include enzyme-linked immunosorbent assay [16], radioimmunoassay [17], fluoroimmunoassay [18], electrochemiluminescence [19], and the latex-enhanced immunoturbidimetric method [20]. Several fully automated benchtop instruments, such as the µTasWako i30 (Fujifilm Wako Diagnostics), IMMULITE 2000 Xpi Immunoassay system (Siemens Healthineers), and the ARCHITECT *i*1000SR immunoassay analyzer (Abbott Diagnostics), are also commercially available. However, many settings do not have access to cold chain and centralized laboratories for these laboratory tests. Typically, these conventional immunoassays take a few days, starting from sample collection to a patient finally getting access to test results through a health care provider.

Mobile platforms (smartphones) are positioned to be the hub of the future of medicine [21], with smartphone- and tablet-based medical devices continuing to be integrated into patients' lives in various settings. The increased use of smartphone-based apps and analytical devices has been demonstrated in recent years for numerous apps such as diet tracking apps [22], well-being apps [23], environmental monitoring [24], food toxin screening [25], and medical diagnostics [26-29]. Point-of-care testing (POCT), using smartphones [30], is rapidly emerging as a potential alternative to conventional screening and laboratory-based diagnostic testing, particularly in resource-limited settings.

Objectives

In this study, we present a proof of concept for the lateral flow immunoassay–based rapid screening of serum AFP levels on a mobile platform, from a drop of human serum, within a few minutes. We aim to demonstrate the quantification of AFP in commercial serum calibrators and preliminary results with commercially obtained human serum samples with known AFP concentrations, quantified by the Abbott ARCHITECT CMIA. Preliminary results from this work will pave the way for our future work focused on developing a mobile device–enabled quad-screen test at the point of care in resource-limited settings.

Methods

Overview

The components of the test strip were selected to achieve optimum flow rates as well as the volume of reagents and AFP in the test samples. Development of the immunoassay involved the selection of commercially available antibodies and optimization of their concentrations by an iterative method to achieve the required detection limits. The entire testing process is guided by a mobile app *AFPhone*, which is designed to input

important patient information, provide step-by-step instructions to the user for running the test, and display the AFP concentrations on the device screen. Briefly, the testing process required a drop of the test sample on the test strip to initiate the test. The camera within the portable reader captures the relative intensity changes of the colored bands on the test strip. A custom code is used to batch process the captured images of the test strip and provide the AFP concentrations. The test strip design was optimized, and the calibration curves were experimentally determined using commercially available serum-based AFP calibrators and commercially obtained serum samples with known AFP concentrations. This study does not contain any studies involving human participants; hence, ethical approval was not required. This development process is described in detail in the following sections.

Reagents and Materials

Gold nanoparticle (AuNP) conjugation kit (InnovaCoat 20OD) with 40 nm diameter AuNPs were obtained from Expedeon, Inc. Purified AFP from human fetal cord serum (Cat# 8F8) and monoclonal mouse antihuman-AFP antibodies (Hytest Cat# 4F16-4A3, RRID:AB_2223930, and Hytest Cat# 4F16-5H7, RRID:AB_2223929) were purchased from HyTest Ltd (Finland). Rabbit antimouse immunoglobulin G (IgG) was purchased from Jackson ImmunoResearch Inc. Audit AFP calibrators (Linearity FD Tumor Markers, Cat# K719M-5) were purchased from AUDIT MicroControls, Inc. Amine-free phosphate buffer saline (at 0.01 M) with a pH of 7.4, Tween 20, bovine serum albumin (BSA), borate buffer, and sucrose were acquired from Sigma-Aldrich. A glass fiber conjugate pad with dimensions of 300 mm × 10 mm, Hi-Flow Plus 180 membrane cards, and a cellulose fiber pad for an absorbent pad were acquired from EMD Millipore. The membrane for the sample pad was purchased from mdi Membrane Technologies, Inc.

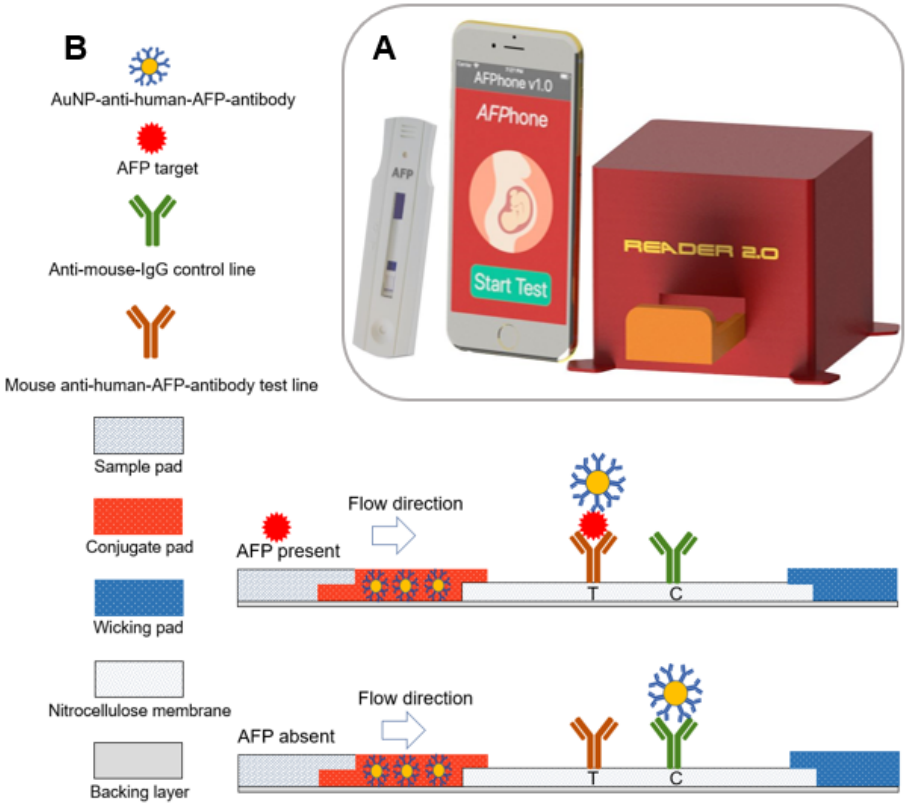
Equipment

The following equipment was used in this study: lateral flow reagent dispenser (Claremont BioSolutions, Upland), Legato 200 Dual Syringe Pump (Claremont BioSolutions LLC), matrix 2360 programmable shear (Kinematic Automation, Inc), and V-1200 Spectrophotometer.

Technology and Components

The technology comprises a custom-made test AFP test strip, cassette for housing the test strip, portable test strip reader, and a mobile app *AFPhone* to guide the user through the various steps of the testing protocol. A custom-developed image processing code was applied to batch process the acquired images to compute the test and control line intensity (TC) ratios for each test strip. Figure 1 shows the technology components of the point-of-care approach described in this study.

Figure 1. (A) Components of the point-of-care testing system. (B) Schematic showing various. components of the test strip with sandwich-type assay for alpha-fetoprotein detection. AFP: alpha-fetoprotein; AuNP: gold nanoparticle; C: control; IgG: immunoglobulin G; T: test.



Test Strip Architecture and Immunoassay Format

The AFP test strip in Figure 1 was based on a sandwich format immunoassay and comprised a whole blood filtration membrane as the sample pad; a conjugate pad for prestoring the AuNP–antihuman–AFP–antibody conjugates in a dry form; a nitrocellulose membrane with antihuman–AFP monoclonal antibodies and secondary antibodies, respectively; and a wicking pad made of cellulose fiber that functions as a waste reservoir. The addition of the test sample and running buffer causes the

AuNP–antihuman–AFP-antibody conjugates to flow freely due to capillary action and react with AFP in the test sample. At high AFP concentrations in the test sample, most of the AuNP–antihuman–AFP–antibody conjugates will bind with the free AFP, eventually binding to the antihuman–AFP–antibody on the test line, resulting in a sandwich complex. All unbound AuNPs–antihuman–AFP were captured at the control line. This relative binding of the AuNP–antihuman–AFP–antibody at the test and control lines increases the test line (T) to control (C) line intensity ratio (TC ratio) in test samples with higher AFP concentration. Similarly, in test samples with lower AFP concentrations, binding of the AuNP–antihuman–AFP–antibody to form a sandwich complex at the test line is reduced, thereby causing an overall decrease in the TC ratio.

AuNP–Antihuman–AFP–Antibody Conjugate Pad Preparation

The antihuman–AFP–antibody was conjugated with AuNPs, following the protocol provided in the gold conjugation kit. The protocol provided by the vendor was used to obtain AuNP–anti-AFP–antibody conjugation. To remove the excess unbound antibodies, a 1:10 dilution of the quencher with water was added up to 10 times the volume of the conjugate mixture and the suspension was centrifuged at 9000 × g for 10 min. The remaining pellet of AuNP–anti-AFP–antibody was resuspended in a solution comprising a 1:10 dilution of quencher with water. The final optical density (OD) was measured using a Spectramax 384 spectrophotometer at 530 nm. The AuNP–anti–AFP conjugate was diluted to 0·35 OD in a conjugate buffer (2 mM borate buffer with 5% sucrose). The conjugate pad was soaked in diluted conjugates for 1 min and oven dried at 37°C for 2 hours, followed by storage at room temperature overnight.

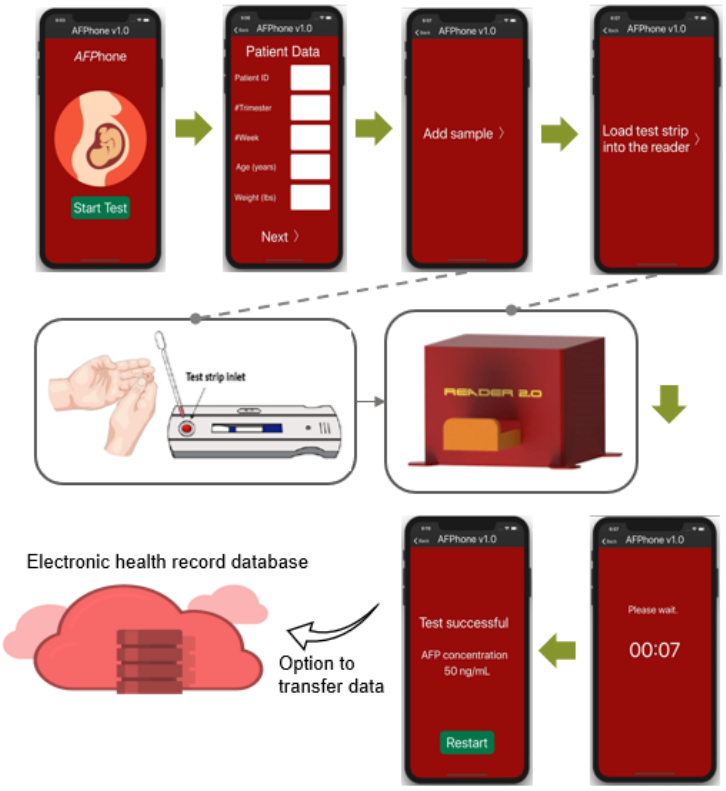
Test Strip Assembly

The membrane card comprised a polyester film backed with a nitrocellulose layer on top. Striping of the test and control line antibodies (1 mm wide and 3 mm spacing), consisting of antihuman AFP–antibody and antimouse–IgG on the nitrocellulose membrane, was performed using the lateral flow antibody dispenser. Membrane cards were then immediately dried for 2 hours at 37°C in a forced convection oven and stored at room temperature in a humidity-controlled box. The conjugate pad, absorbent pad, and sample pad were then assembled with a 2-mm overlap between each pad. The assembled card was cut using an automated shear cutter to obtain test strips of 5 mm width.

Testing Protocol

Figure 2 shows a schematic of the various steps involved in conducting point-of-care AFP testing. The user is guided with step-by-step instructions on the mobile app. Briefly, the user first adds the test sample comprising a mixture of the archived serum or serum-based standards and chase buffer (1× Tris-buffered saline with 1% BSA, 1.5% Tween20, and 0.1% sodium azide) to the test strip to initiate capillary flow within the test strip, which causes the AuNP–antihuman–AFP–antibody conjugates to be released from the conjugate pad. The free AFP in the test sample reacts with the AuNP–antihuman–AFP–antibody and flows downstream to further react with antibodies at the test and control lines. The remaining sample was finally collected in the absorbent pad. The user inserts the test strip into the test strip reader to capture the images of the colorimetric signals with the camera and to analyze via the mobile app to provide the AFP concentrations.

Figure 2. Testing protocol with representative screenshots of the mobile app providing the user with step-by-step instructions.



Preparation of AUDIT AFP Calibrators

The AUDIT AFP calibrators (vials S0-S5) available in a lyophilized form were dissolved in deionized water, based on the instructions provided in the kit to obtain concentrations ranging from 4 to 1034 ng/mL.

Image Processing

The mobile app performs image processing steps on the captured test strip image to improve the accuracy and detection limit. Details of the image processing approach have been previously reported. Briefly, captured images are cropped and converted into grayscale to extract the local minima of pixel intensities and calculate the TC ratios, which correlate with AFP concentrations.

Statistical Analysis

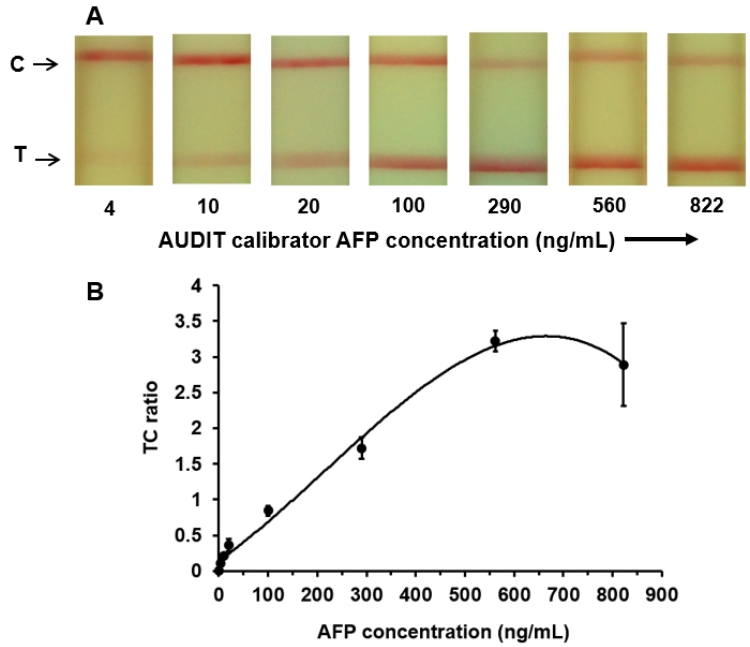
All analyses were performed using Excel (Microsoft) and R software (RStudio version 1.1.456, RStudio Inc). TC ratios were compared with AFP concentrations determined by the reference method with nonparametric bootstrap resampling analysis conducted in R.

Results

Calibration Curve for AUDIT AFP Calibrators

Commercially available AUDIT calibrators were obtained as 5 separate vials labeled A-E, with concentrations ranging from 4 to 1034 ng/mL. Testing was performed simultaneously for each concentration in triplicates. The colorimetric changes in the test and control lines at various known concentrations of the AUDIT calibrators are shown in Figure 3.

Figure 3. (A) Colorimetric variation of the test and control line regions on the alpha-fetoprotein test strip at various known concentrations of alpha-fetoprotein in AUDIT serum-based calibrators. (B) Calibration curve showing the test and control line intensity ratios of the colorimetric signals at various alpha-fetoprotein concentrations in AUDIT serum-based calibrators. AFP: alpha-fetoprotein; C: control; T: test.

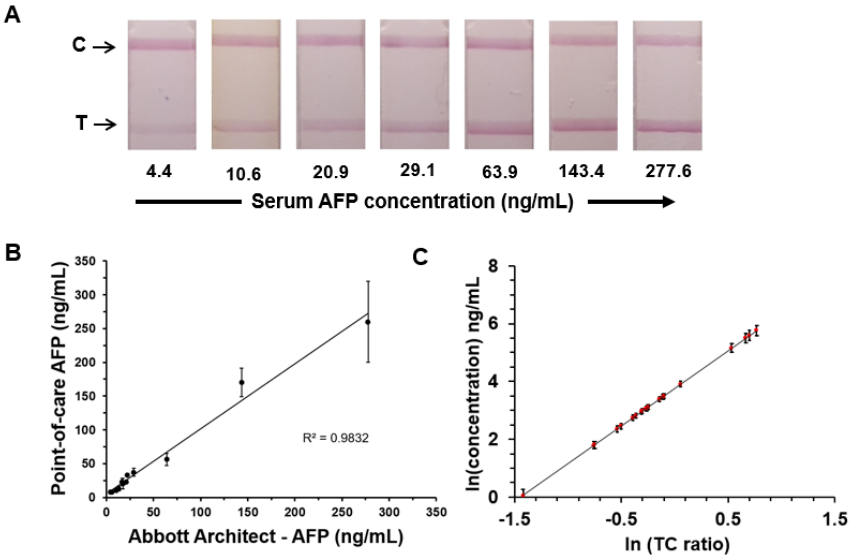


The calibration curve shown in Figure 3 demonstrated that the TC ratios were correlated with the AFP concentration in the calibrators. The TC ratio increased with increasing AFP concentrations until approximately 650 ng/mL and then began to decrease beyond the physiological range because of the hook effect [31,32].

Calibration Curve for Human Serum Samples

The performance of the AFP test strips was further evaluated using archived human serum samples. Serum samples included commercially available serum samples with known AFP concentrations provided by the vendor based on Abbott ARCHITECT CMIA. The colorimetric changes in the test and control lines at various known concentrations of serum samples are presented in Figure 4.

Figure 4. (A) Colorimetric variation of the test and control line regions on the alpha-fetoprotein test strip for serum samples. (B) Correlation plot of serum alpha-fetoprotein concentrations predicted from point-of-care test strip results against the corresponding Abbott ARCHITECT results. (C) Results of bootstrapping to compare bootstrap means of predicted alpha-fetoprotein concentrations and standard errors for the observed test and control line intensity ratio values from a power model using 1000 resampled data sets. AFP: alpha-fetoprotein; TC: test and control line intensity.



We selected TC data for 8 samples and compared our test strip results with the corresponding reference method (Abbott ARCHITECT) results to obtain an initial calibration curve. This calibration curve was then applied to predict the AFP concentrations of the remaining samples tested on the mobile platform. Figure 4 shows a correlation plot comparing the AFP levels predicted by our point-of-care technology against the corresponding levels provided by the vendor, based on Abbott ARCHITECT testing.

Bootstrap resampling analysis was performed using RStudio with serum AFP concentration results of 20 samples to assess the correlations between TC ratios determined on the point-of-care system with Abbott ARCHITECT–determined AFP concentrations. The bootstrapping function was applied to resample 1000 times, and the resulting correlation coefficients were computed. Figure 4 presents the bootstrapping results to compare the bootstrap means of predicted AFP concentrations and SEs for the observed TC ratio values from a power model, using 1000 resampled data sets. A 95% CI for the correlation coefficient between TC ratios and AFP concentrations of the tested serum samples was also obtained (0.846-0.975). Findings from bootstrap analyses provide quantitative evidence that TC ratios from our point-of-care technology and AFP concentrations of the tested serum samples are highly correlated.

Discussion

Principal Findings

In this study, we demonstrated a proof of concept for a sandwich-type immunoassay test strip on our mobile platform for the quantification of AFP concentrations in human serum samples. We determined calibration curves for the AFP assay on a mobile platform with commercially available AUDIT AFP serum-based calibrators and commercially obtained serum samples with known AFP concentrations. The detection range demonstrated in the AUDIT AFP calibrators was 4-650 ng/mL, which covers the range of maternal serum AFP levels observed during pregnancy. The point-of-care AFP assay on the mobile platform was successfully applied to quantify the AFP concentrations in commercially obtained human serum samples with concentrations ranging from 2.2 to 456 ng/mL based on the Abbott ARCHITECT method. On preliminary testing and comparison with Abbott ARCHITECT, a correlation of 0.98 (P<.001) was observed with high sensitivity.

Interpretation of the Findings

On preliminary testing and comparison with Abbott ARCHITECT, a correlation of 0.98 (P<.001) was observed with high sensitivity. The detection range of our point-of-care assay covers the range of maternal serum AFP levels observed during pregnancy. The mobile platform for AFP has the capability to quantify AFP within a few minutes without the need for expensive and time-intensive methods. Obstetricians worldwide face the challenge of screening high-risk pregnancies among the overwhelming number of pregnant women presenting to hospitals. Advances in medical technology often widen health disparities, seldom reaching resource-limited populations. State-of-the-art diagnostic equipment is costly, is bulky, and requires sophisticated training for operation and maintenance. Shifting the approach from clinic- and laboratory-dependent care to a mobile platform based on our point-of-care approach will enable translation to resource-limited settings and the global health care market. Mobile devices, which are increasingly ubiquitous even in resource-limited settings, have the capacity to transform clinical care, health services, research, and surveillance [33,34] across populations. An early confirmation of high-risk pregnancies gives parents more time to process the information and learn about early intervention programs including establishing care in a patient-centered medical home, reviewing eligibility for parental financial and psychological support programs. Clinicians can provide parents with unbiased, comprehensive, and culturally sensitive information about congenital birth defects and available services. An early screening is critical to identify and enroll newborns in

state-specific early intervention programs as soon as possible to improve the short- and long-term outcomes.

Screening for MSAFP concentrations is a standard of prenatal care to identify pregnancies that may increase the risk of NTDs and some trisomy-affected pregnancies, such as trisomy 21 (Down syndrome) and trisomy 18 (Edwards syndrome). These methods aim to screen for the risk of adverse pregnancy outcomes by quantifying biochemical markers in the maternal serum and, in some cases, also incorporate fetal nuchal translucency measurements obtained via ultrasound. This screen is aimed at identifying pregnancies at a higher risk, so that these patients can be offered diagnostic testing and counseling if required. MSAFP results during pregnancy are usually expressed as multiple of the median (MoM) for each gestational week. MoM values are easy to derive, more stable, and allow for an interlaboratory variation. MSAFP levels that are above 2.0 to 2.5 MoMs are considered abnormal. Using the MSAFP cut-off level of 2 to 2.5 MoM for a given gestational age in singleton pregnancies, the detection rates are approximately 95% for anencephaly and between 70% and 85% for other NTDs [35]. In many cases, high-resolution ultrasound is used in conjunction with MSAFP screening, with detection rates depending on gestational age and the type and severity of the NTD.

MSAFP assessment can also be used to screen for other fetal malformations. Studies have shown that elevated second trimester MSAFP can also indicate gestational age, multiple pregnancies, intrauterine death, or other fetal malformations [36]. High MSAFP levels are also observed in certain types of ovarian germ cell tumors (eg, endodermal sinus tumor and embryonal carcinoma), with levels often greater than 1000 ng/mL, especially with pure endodermal sinus (yolk sac) tumors, in which MSAFP levels are greater than 10,000 ng/mL. MSAFP screening has been shown to play a valuable role in the management of twin pregnancy [37], both in the detection of twins and in the prediction of perinatal outcomes in twin pregnancies.

When MSAFP is elevated, targeted ultrasound examination is offered as the initial diagnostic test, in addition to, or in place of, amniocentesis. The quadruple test, which includes AFP, is performed in the early second trimester, optimally at 15-18 weeks of gestation. Errors in the estimation of gestational age [38,39] are the most common reason for a false-positive result. If the true gestational age is earlier than reported, then AFP MoM values will be falsely interpreted as low. However, in resource-limited settings, even with a high number of false positives, a screening test can identify those in need of an ultrasound; furthermore, the number of false positives can be reduced with simple sequencing during clinical examinations, such as matching the last menstrual period with fundal height assessment where relevant and feasible. The sensitivity of serum AFP screening for NTDs has been shown to significantly improve when the gestational age used for the AFP MoM calculation was verified by ultrasound [40]. There is also evidence that in pregnancies resulting in spontaneous early preterm delivery, MSAFP level at 11-13 weeks' gestation is higher, and MSAFP measurement improves the prediction of preterm delivery compared with maternal characteristics and obstetric history alone [41]. The mobile app for this test can be designed to include a data collection module to compile maternal age, weight, ethnicity, gestational age, and other relevant patient history. This additional patient information presented along with the AFP results can enable an obstetrician to interpret the test results and make informed decisions regarding any diagnostic and clinical follow-up. The mobile app can also be designed to wirelessly transmit the test result data to a centralized database of laboratories or public health agencies that can be accessed by obstetricians to interpret the test results and make informed decisions regarding any diagnostic and clinical follow-up.

Serum AFP is a widely accepted serum marker for the detection of hepatocellular carcinoma (HCC). Serum AFP is elevated in tumors, including HCC, hepatoblastoma, and nonseminomatous germ cell tumors of the ovary and testis. Most studies report elevated AFP concentrations in approximately 70% of patients with HCC and in 50% to 70% of patients with nonseminomatous testicular tumors. The 5-year survival rate of primary liver cancer is approximately 15%, and the mortality rate is mainly attributed to late diagnosis in many patients, with a high recurrence rate after curative treatment. There is an urgent need for regular screening for patients at risk for HCC to enable an early detection of this tumor or its recurrence. The second model list of essential in vitro diagnostics list (EDL) was recently released by the World Health Organization for detecting, diagnosing, and monitoring a wide array of disease conditions. There is an urgent need for POCT, especially for the diagnostics listed under the first tier in community and health settings without laboratories. The second model list of EDL includes AFP diagnostic testing for the screening of HCC and staging and disease monitoring of germ cell tumors. HCC accounts for 70%-85% of all primary liver cancers and is the ninth leading cause of cancer-related deaths in the United States.

In areas with the highest burden of NTDs, such as India [5,42-44], there is also limited access to centralized laboratories, cold chain, amniocentesis, and ultrasound technology. Such limited access makes a POCT device for screening NTDs especially relevant for these populations. The use of a drop of capillary blood from a finger prick in our approach is minimally invasive, and sample collection is faster compared with venous blood sampling used in a traditional laboratory approach. A study comparing AFP values of capillary and venous blood in 43 participants concluded that there were no significant differences, with a high correlation (r=0.995) between the sampling methods [45]. AFP concentrations in men and nonpregnant women vary by age and race but are typically in the range of 0-40 ng/mL. MSAFP levels in pregnancy begin to increase beginning around 14 weeks of gestation until approximately 32 weeks. Between weeks 15 and 20, MSAFP levels mostly range between 10 and 150 ng/mL [46]. In adults, serum AFP levels greater than 200 ng/mL in patients with liver cirrhosis are a strong indicator of HCC.

Our preliminary work presented here is a quantitative, point-of-care lateral flow immunoassay–based screening test for the quantification of serum AFP concentrations. This technology will enable a rapid screening of high-risk pregnancies and enable physicians to make informed decisions, especially in resource-limited settings with limited access to

diagnostic laboratories. The ACOG recommends integrated or sequential screening tests with high detection rates earlier in pregnancy, which can provide patients with diagnostic options to consider. The findings suggest that at the current stage of development, this technology can play a significant role as a screening tool for high-risk pregnancies to assess whether further diagnostic testing may be needed.

Limitations

This study has some limitations. There were limited number of test samples used in this study. In addition, there is a need for validation with larger samples in our future work and a more comprehensive evaluation of the diagnostic performance. In future studies, we plan to optimize the performance of the AFP point-of-care assay on a mobile platform using whole blood in human validation studies among a greater number of participants with a broader range of AFP concentrations to improve the calibration curve and conduct an extensive evaluation of diagnostic performance.

Conclusions

In conclusion, we developed and performed a preliminary testing of a point-of-care screening test on our mobile platform for the detection of serum AFP levels from a drop of serum sample within a few minutes. On the basis of preliminary testing results, the AFP screening test on the mobile platform reported in this study will pave the way for our future work focused on developing a point-of-care mobile device–enabled quad-screen test in real time in clinical and field settings. State-of-the-art diagnostic equipment is costly, is bulky, and requires sophisticated training for operation and maintenance. Shifting the approach from clinic- and laboratory-dependent care to a mobile platform based on our point-of-care approach will enable translation to resource-limited settings and the global health care market. Screening for high-risk pregnancies will enable physicians to make informed decisions on whether further diagnostic testing, such as ultrasound and amniocentesis, should be considered. Prenatal NTD detection also informs decisions about the optimal time, route, and site of delivery. Our future work will focus on conducting appropriately powered diagnostic test accuracy studies with maternal serum samples and developing a multiplexed assay to include hCG and unconjugated estriol as a part of the triple screening test. An early confirmation of high-risk pregnancies provides parents more time to process the information and learn about early intervention programs, including establishing care in a patient-centered medical home, reviewing eligibility for parental financial and psychological support programs. An early screening is critical to identify and enroll newborns in state-specific early intervention programs as soon as possible to improve short- and long-term outcomes. Overall, the preliminary results reported in this work serve as a foundation for our future research focused on developing a quad-screen POCT to enable (1) screening for high-risk pregnancies within a few minutes at the point of care even in remote areas; (2) identification of patients who might need continued health care advice and counseling; (3) planning for enrolling newborns in state-specific early intervention programs as soon as possible to improve short- and long-term outcomes; and (4) development of surveillance tools for birth defects.

Conflicts of Interest

DE and SM have an equity interest in VitaScan Technologies, Inc, which is commercializing point of care assays for nutritional status developed in their labs. The remaining authors have no financial relationships relevant to this paper.

References

1. Greene NDE, Copp AJ. Neural tube defects. Annu Rev Neurosci 2014;37:221-242 [FREE Full text] [doi: 10.1146/annurev-neuro-062012-170354] [Medline: 25032496]
2. Copp AJ, Stanier P, Greene NDE. Neural tube defects: recent advances, unsolved questions, and controversies. Lancet Neurol 2013 Aug;12(8):799-810 [FREE Full text] [doi: 10.1016/S1474-4422(13)70110-8] [Medline: 23790957]
3. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet 2012 Dec;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4]
4. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet 2012 Dec;380(9859):2095-2128. [doi: 10.1016/S0140-6736(12)61728-0]
5. Zaganjor I, Sekkarie A, Tsang BL, Williams J, Razzaghi H, Mulinare J, et al. Describing the prevalence of neural tube defects worldwide: a systematic literature review. PLoS One 2016;11(4):e0151586 [FREE Full text] [doi: 10.1371/journal.pone.0151586] [Medline: 27064786]
6. Atta CAM, Fiest KM, Frolkis AD, Jette N, Pringsheim T, St Germaine-Smith C, et al. Global birth prevalence of spina bifida by folic acid fortification status: a systematic review and meta-analysis. Am J Public Health 2016 Jan;106(1):e24-e34. [doi: 10.2105/AJPH.2015.302902] [Medline: 26562127]
7. International Clearinghouse for Birth Defects Surveillance and Research. ICBDSR Birth Defects Annual Report. 2014. URL: http://www.icbdsr.org/resources/annual-report/ [accessed 2021-02-10]
8. Blencowe H, Kancherla V, Moorthie S, Darlison MW, Modell B. Estimates of global and regional prevalence of neural tube defects for 2015: a systematic analysis. Ann N Y Acad Sci 2018 Feb;1414(1):31-46. [doi: 10.1111/nyas.13548] [Medline: 29363759]

9. Grosse SD, Berry RJ, Mick Tilford J, Kucik JE, Waitzman NJ. Retrospective assessment of cost savings from prevention: folic acid fortification and spina bifida in the U.S. Am J Prev Med 2016 May;50(5 Suppl 1):S74-S80 [FREE Full text] [doi: 10.1016/j.amepre.2015.10.012] [Medline: 26790341]
10. Practice bulletin No. 187: neural tube defects. Obstet Gynecol 2017 Dec;130(6):e279-ee90. [doi: 10.1097/AOG.0000000000002412] [Medline: 29189693]
11. Driscoll DA, Gross SJ, Professional Practice Guidelines Committee. Screening for fetal aneuploidy and neural tube defects. Genet Med 2009 Nov;11(11):818-821 [FREE Full text] [doi: 10.1097/GIM.0b013e3181bb267b] [Medline: 19915395]
12. Bartkute K, Balsyte D, Wisser J, Kurmanavicius J. Pregnancy outcomes regarding maternal serum AFP value in second trimester screening. J Perinat Med 2017 Oct 26;45(7):817-820. [doi: 10.1515/jpm-2016-0101] [Medline: 27771626]
13. Gagnon A, Wilson RD, Society of Obstetricians and Gynaecologists of Canada Genetics Committee. Obstetrical complications associated with abnormal maternal serum markers analytes. J Obstet Gynaecol Can 2008 Oct;30(10):918-932. [doi: 10.1016/S1701-2163(16)32973-5] [Medline: 19038077]
14. Graves JC, Miller KE, Sellers AD. Maternal serum triple analyte screening in pregnancy. Am Fam Physician 2002 Mar 01;65(5):915-920 [FREE Full text] [Medline: 11898965]
15. Muller F, Dreux S, Sault C, Galland A, Puissant H, Couplet G, et al. Very low alpha-fetoprotein in Down syndrome maternal serum screening. Prenat Diagn 2003 Jul;23(7):584-587. [doi: 10.1002/pd.646] [Medline: 12868089]
16. Burgett MW, Kwan S. An enzyme-labelled immunoassay for the measurement of α-fetoprotein. Ricerca in clinica e in laboratorio January 01 1982;12(1):299-306. [doi: 10.1007/bf02909336]
17. Kemp HA, Simpson JS, Woodhead JS. Automated two-site immunoradiometric assay of human alpha-fetoprotein in maternal serum. Clin Chem 1981 Aug;27(8):1388-1391. [Medline: 6168413]
18. Ao L, Gao F, Pan B, He R, Cui D. Fluoroimmunoassay for antigen based on fluorescence quenching signal of gold nanoparticles. Anal Chem 2006 Feb 15;78(4):1104-1106. [doi: 10.1021/ac051323m] [Medline: 16478100]
19. Yuan S, Yuan R, Chai Y, Mao L, Yang X, Yuan Y, et al. Sandwich-type electrochemiluminescence immunosensor based on Ru-silica@Au composite nanoparticles labeled anti-AFP. In: Talanta. Amsterdam: Elsevier; Sep 15, 2010:1468-1471.
20. Diao B, Wen K, Chen J, Liu Y, Yuan Z, Han C. Diagnosis of acute respiratory syndrome Coronavirus 2 infection by detection of nucleocapsid protein. medRxiv. 2020.07 2003:2524. [doi: 10.1101/2020.03.07.20032524]
21. Hernández-Neuta I, Neumann F, Brightmeyer J, Ba Tis T, Madaboosi N, Wei Q, et al. Smartphone-based clinical diagnostics: towards democratization of evidence-based health care. J Intern Med 2019 Jan;285(1):19-39 [FREE Full text] [doi: 10.1111/joim.12820] [Medline: 30079527]
22. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]
23. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
24. Finazzi F, Paci L. Quantifying personal exposure to air pollution from smartphone-based location data. Biometrics 2019 Dec;75(4):1356-1366. [doi: 10.1111/biom.13100] [Medline: 31180147]
25. Sergeyeva T, Yarynka D, Piletska E, Linnik R, Zaporozhets O, Brovko O, et al. Development of a smartphone-based biomimetic sensor for aflatoxin B1 detection using molecularly imprinted polymer membranes. Talanta 2019 Aug 15;201:204-210. [doi: 10.1016/j.talanta.2019.04.016] [Medline: 31122412]
26. Guo J. Uric acid monitoring with a smartphone as the electrochemical analyzer. Anal Chem 2016 Dec 20;88(24):11986-11989. [doi: 10.1021/acs.analchem.6b04345] [Medline: 28193075]
27. Leggio M, Villano A, Fusco A. Smartphone-based clinical diagnostics and atrial fibrillation: the need to avoid being out of sight. J Intern Med 2019 Sep;286(3):358-360. [doi: 10.1111/joim.12916] [Medline: 31069873]
28. Petrie T, Samatham R, Goodyear SM, Webster DE, Leachman SA. MoleMapper: an application for crowdsourcing mole images to advance melanoma early-detection research. Semin Cutan Med Surg 2019 Mar 01;38(1):E49-E56. [doi: 10.12788/j.sder.2019.001] [Medline: 31051024]
29. Bifolck E, Fink A, Pedersen D, Gregory T. Smartphone imaging for the ophthalmic examination in primary care. J Ame Acad Phy Assistants 2018 Aug;31(8):34-38. [doi: 10.1097/01.JAA.0000541482.54611.7c] [Medline: 30048353]
30. Liu J, Geng Z, Fan Z, Liu J, Chen H. Point-of-care testing based on smartphone: the current state-of-the-art (2017-2018). Biosens Bioelectron 2019 May 01;132:17-37. [doi: 10.1016/j.bios.2019.01.068] [Medline: 30851493]
31. Tate J, Ward G. Interferences in immunoassay. Clin Biochem Rev 2004 May;25(2):105-120 [FREE Full text] [Medline: 18458713]
32. Butch AW. Dilution protocols for detection of hook effects/prozone phenomenon. Clin chem 2000;46(10):1719. [doi: 10.1093/clinchem/46.10.1719]
33. Reefhuis J, Gilboa SM, Anderka M, Browne ML, Feldkamp ML, Hobbs CA, National Birth Defects Prevention Study. The national birth defects prevention study: a review of the methods. Birth Defects Res A Clin Mol Teratol 2015 Aug 02;103(8):656-669 [FREE Full text] [doi: 10.1002/bdra.23384] [Medline: 26033852]

34. González-Andrade F, López-Pulles R. Congenital malformations in Ecuadorian children: urgent need to create a National Registry of Birth Defects. Appl Clin Genet 2010;3:29-39 [FREE Full text] [doi: 10.2147/tacg.s8794] [Medline: 23776361]
35. Ghi T, Dall'asta A, Pilu G, Contro E, De MF, Frusca T. 41 - Neural tube defects. In: Obstetric Imaging: Fetal Diagnosis and Care (Second Edition). Amsterdam: Elsevier; 2018:213-226.
36. Adams MJ, Windham GC, James LM, Greenberg F, Clayton-Hopkins JA, Reimer CB, et al. Clinical interpretation of maternal serum α-fetoprotein concentrations. Am J Obstet Gynecol 1984 Feb;148(3):241-254 [FREE Full text] [doi: 10.1016/s0002-9378(84)80062-9]
37. Johnson JM, Harman CR, Evans JA, MacDonald K, Manning FA. Maternal serum alpha-fetoprotein in twin pregnancy. Am J Obstet Gynecol 1990 Apr;162(4):1020-1025. [doi: 10.1016/0002-9378(90)91308-y] [Medline: 1691590]
38. Millner SN. Effect of incorrect gestational dating on Down's syndrome and neural tube risk assessment. Ann Clin Biochem 2001 May;38(Pt 3):230-234. [doi: 10.1258/0004563011900650] [Medline: 11392497]
39. Schnittger A, Selbing A. Inaccurate gestational dating and maternal serum alpha-fetoprotein screening. Acta Obstet Gynecol Scand 1984 Jan;63(s119):43-47 [FREE Full text] [doi: 10.3109/00016348409156967]
40. Dashe JS, Twickler DM, Santos-Ramos R, McIntire DD, Ramus RM. Alpha-fetoprotein detection of neural tube defects and the impact of standard ultrasound. Am J Obstet Gynecol 2006 Dec;195(6):1623-1628 [FREE Full text] [doi: 10.1016/j.ajog.2006.03.097] [Medline: 16769022]
41. Beta J, Bredaki FE, Rodriguez Calvo J, Akolekar R, Nicolaides KH. Maternal serum α-fetoprotein at 11-13 weeks' gestation in spontaneous early preterm delivery. Fetal Diagn Ther 2011;30(2):88-93 [FREE Full text] [doi: 10.1159/000324352] [Medline: 21411977]
42. Bhide P, Sagoo GS, Moorthie S, Burton H, Kar A. Systematic review of birth prevalence of neural tube defects in India. Birth Defects Res A Clin Mol Teratol 2013 Jul;97(7):437-443. [doi: 10.1002/bdra.23153] [Medline: 23873811]
43. Allagh KP, Shamanna BR, Murthy GVS, Ness AR, Doyle P, Neogi SB, Wellcome Trust-PHFI Folic Acid project team. Birth prevalence of neural tube defects and orofacial clefts in India: a systematic review and meta-analysis. PLoS One 2015;10(3):e0118961 [FREE Full text] [doi: 10.1371/journal.pone.0118961] [Medline: 25768737]
44. Cherian AG, Jamkhandi D, George K, Bose A, Prasad J, Minz S. Prevalence of congenital anomalies in a secondary care hospital in South India: a cross-sectional study. J Trop Pediatr 2016 Oct;62(5):361-367. [doi: 10.1093/tropej/fmw019] [Medline: 27044502]
45. Vidranski V, Bukovec-Megla Z, Lenicek KJ, Bokulic A. Differences between capillary and venous blood values of alpha-fetoprotein in healthy adults. In: Proceedings of the 3rd EFLM-BD European Conference on Preanalytical Phase. 2015 Presented at: 3rd EFLM-BD European Conference on Preanalytical Phase; 2015; Porto, Portugal p. 2015. [doi: 10.13140/RG.2.1.3318.7366]
46. Adigun OO, Yarrarapu SNS, Khetarpal K. Alpha Fetoprotein. Florida: StatPearls Publishing LLC; 2020.

Abbreviations

ACOG: American College of Obstetricians and Gynecologists
AFP: alpha-fetoprotein
AuNP: gold nanoparticle
BSA: bovine serum albumin
CMIA: chemiluminescent magnetic microparticle immunoassay
EDL: essential in vitro diagnostics list
HCC: hepatocellular carcinoma
hCG: human chorionic gonadotropin
IgG: immunoglobulin G
MoM: multiple of the median
MSAFP: maternal serum alpha-fetoprotein
NTD: neural tube defect
OD: optical density
POCT: point-of-care testing
TC: test and control line intensity

Edited by G Eysenbach; submitted 14.08.20; peer-reviewed by S Cicero, J Rezende Filho, W Leung; comments to author 24.01.21; revised version received 27.01.21; accepted 01.02.21; published 03.03.21.

Please cite as:
Srinivasan B, Finkelstein JL, Erickson D, Mehta S
Point-of-Care Quantification of Serum Alpha-Fetoprotein for Screening Birth Defects in Resource-Limited Settings: Proof-of-Concept Study
JMIR Biomed Eng 2021;6(1):e23527
URL: https://biomedeng.jmir.org/2021/1/e23527
doi:10.2196/23527
PMID:34746648

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org