A Statistical (Process Monitoring) Perspective on Human Performance Modeling in the Age of Cyber-Physical Systems

Fadel M. Megahed[1], L. Allison Jones-Farmer[1], Miao Cai[2], Steven E. Rigdon[2] and Manar Mohamed [1]

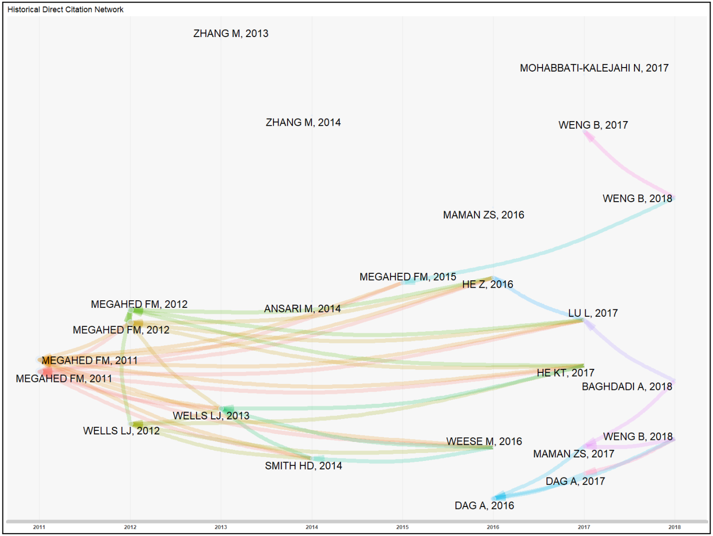
[1]Miami University, OH, USA

[2]Saint Louis University, MO, USA

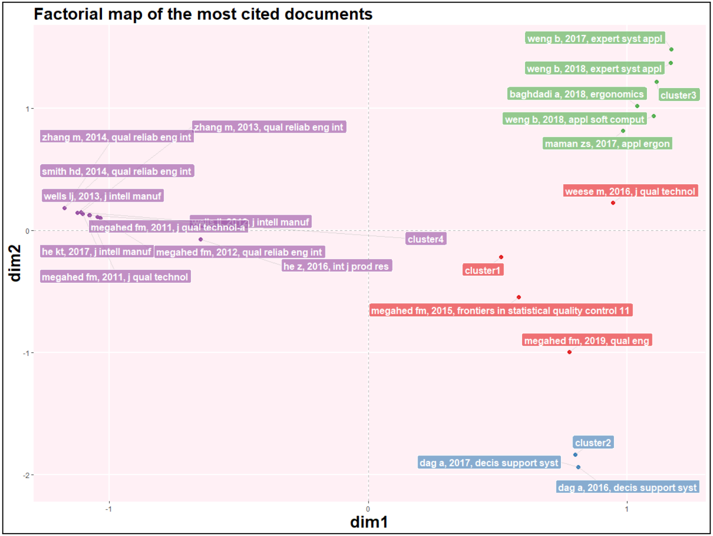
August 14th, 2019

Outline

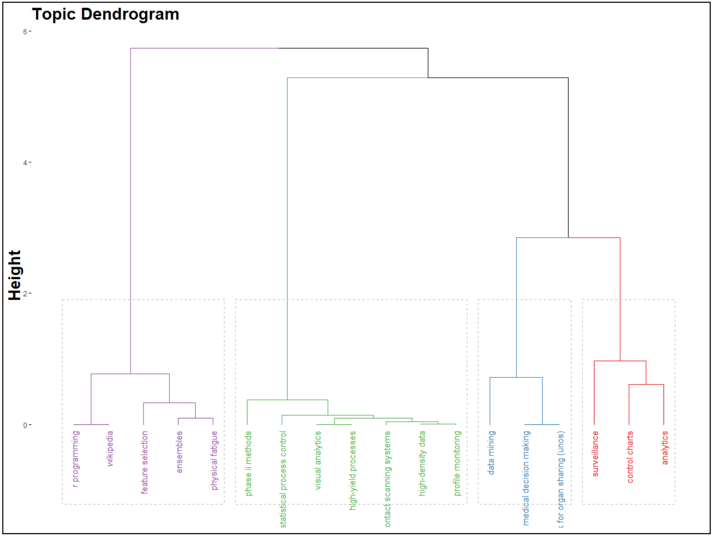
My Research Portfolio's Evolution



My Research Portfolio’s Evolution

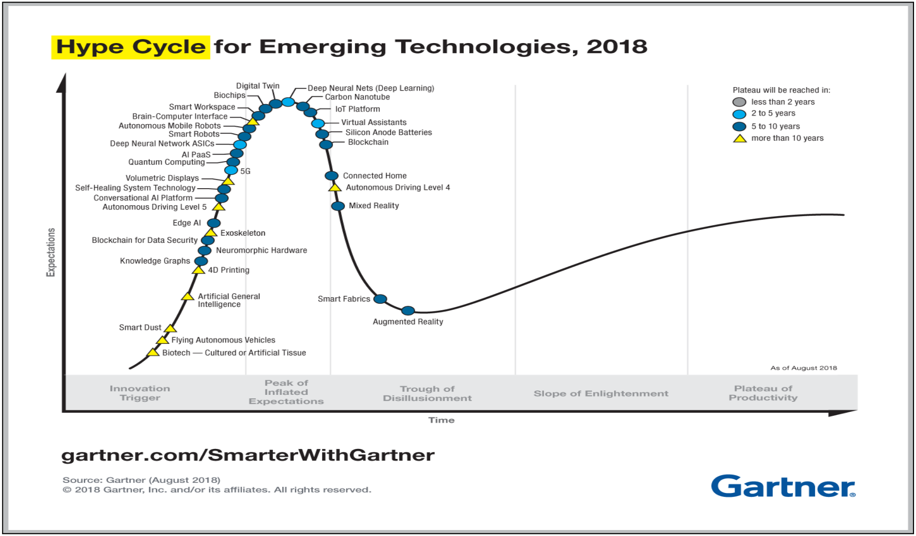


My Research Portfolio's Evolution



Outline

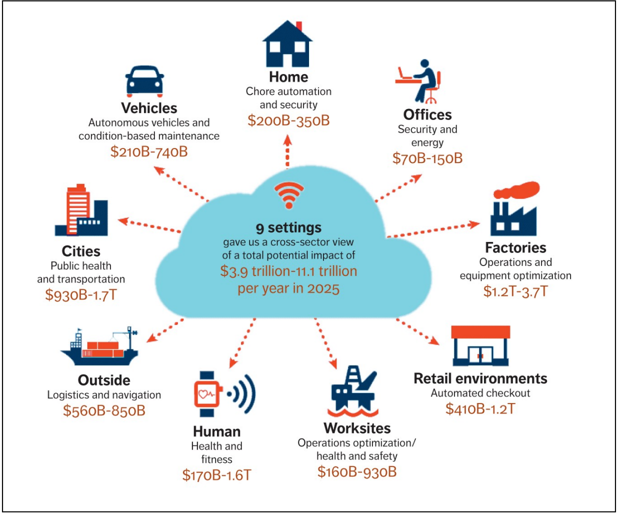
Gartner's Hype Cycle for Emerging Technologies



NSF's Alignment with the Gartner's Perspective

"Cyber-physical systems (CPS) are engineered systems that are built from, and depend upon, the seamless integration of computation and physical components. Advances in CPS will enable capability, adaptability, scalability, resiliency, safety, security, and usability that will expand the horizons of these critical systems. CPS technologies are transforming the way people interact with engineered systems, just as the Internet has transformed the way people interact with information. New, smart CPS drive innovation and competition in a range of application domains including agriculture, aeronautics, building design, civil infrastructure, energy, environmental quality, healthcare and personalized medicine, manufacturing, and transportation. Moreover, the integration of artificial intelligence with CPS creates new research opportunities with major societal implications." [1]

Expected Value for CPS Applications



Source: McKinsey&Company. The McKinsey Global Institute [2].

Humans: A Cornerstone of CPS/IoT Applications

- **From the McKinsey report [2], one can argue that data pertaining to human/workers is important to the following five applications:**
  - **Factories [Market size: $1.2T-3.7T]**
  - Human health and fitness [Market size: $170B-1.6T]
  - Worksites [Market size: $160B-930B]
  - Logistics and Navigation [Market size: $560B-850B]
  - Offices [Market size: $70B-150B]
- At a minimum, the market size for these five applications is estimated to be $2.16T in 2025.

Humans: A Cornerstone of CPS/IoT Applications

- **From the McKinsey report [2], one can argue that data pertaining to human/workers is important to the following five applications:**
  - **Factories [Market size: $1.2T-3.7T]**
  - **Human health and fitness [Market size: $170B-1.6T]**
  - **Worksites [Market size: $160B-930B]**
  - **Logistics and Navigation [Market size: $560B-850B]**
  - **Offices [Market size: $70B-150B]**
- **At a minimum, the market size for these five applications is estimated to be $2.16T in 2025.**

Humans: A Cornerstone of CPS/IoT Applications

- **From the McKinsey report [2], one can argue that data pertaining to human/workers is important to the following five applications:**
  - **Factories [Market size: \$1.2T-3.7T]**
  - **Human health and fitness [Market size: \$170B-1.6T]**
  - **Worksites [Market size: \$160B-930B]**
  - **Logistics and Navigation [Market size: \$560B-850B]**
  - **Offices [Market size: \$70B-150B]**
- **At a minimum, the market size for these five applications is estimated to be \$2.16T in 2025.**

Humans: A Cornerstone of CPS/IoT Applications

- **From the McKinsey report [2], one can argue that data pertaining to human/workers is important to the following five applications:**
  - **Factories [Market size: $1.2T-3.7T]**
  - **Human health and fitness [Market size: $170B-1.6T]**
  - **Worksites [Market size: $160B-930B]**
  - **Logistics and Navigation [Market size: $560B-850B]**
  - **Offices [Market size: $70B-150B]**
- **At a minimum, the market size for these five applications is estimated to be $2.16T in 2025.**

Humans: A Cornerstone of CPS/IoT Applications

- **From the McKinsey report [2], one can argue that data pertaining to human/workers is important to the following five applications:**
  - **Factories [Market size: $1.2T-3.7T]**
  - **Human health and fitness [Market size: $170B-1.6T]**
  - **Worksites [Market size: $160B-930B]**
  - **Logistics and Navigation [Market size: $560B-850B]**
  - **Offices [Market size: $70B-150B]**
- **At a minimum, the market size for these five applications is estimated to be $2.16T in 2025.**

Humans: A Cornerstone of CPS/IoT Applications

- From the McKinsey report [2], one can argue that data pertaining to human/workers is important to the following five applications:
  - Factories [Market size: $1.2T-3.7T]
  - Human health and fitness [Market size: $170B-1.6T]
  - Worksites [Market size: $160B-930B]
  - Logistics and Navigation [Market size: $560B-850B]
  - Offices [Market size: $70B-150B]
- **At a minimum, the market size for these five applications is estimated to be $2.16T in 2025.**

Humans: A Cornerstone of CPS/IoT Applications

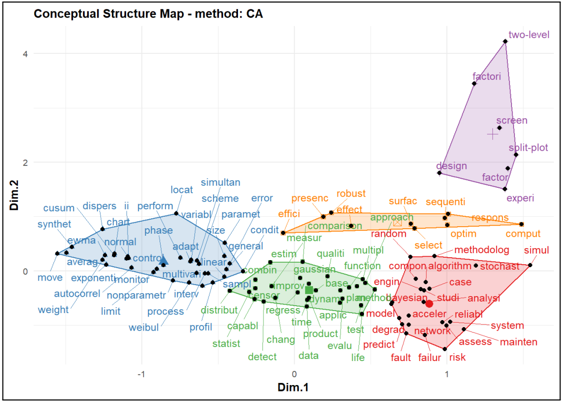
- From the McKinsey report [2], one can argue that data pertaining to human/workers is important to the following five applications:
  - Factories [Market size: $1.2T-3.7T]
  - Human health and fitness [Market size: $170B-1.6T]
  - Worksites [Market size: $160B-930B]
  - Logistics and Navigation [Market size: $560B-850B]
  - Offices [Market size: $70B-150B]
- At a minimum, the market size for these five applications is estimated to be $2.16T in 2025.

Is human performance monitoring an area of emphasis in our SQC research methodologies and/or applications?

SQC and Human Performance Monitoring

The answer is NO!!! Here is how we reached this conclusion:

(1) We "read" all 1576 articles published in: *JQT*, *Technometrics*, *QE*, and *QREI* in 2014 - August 1, 2019. Here is what we have learned:



SQC and Human Performance Monitoring

The answer is NO!!! Here is how we reached this conclusion:

(2) In the Frontiers of Statistical Quality Control 12, we had 20 outstanding contributions. The papers considered important application domains such as: web data, multi-stage processes, and single server queues.

None of the papers explicitly considered monitoring human performance data. However, the discussion in Lazariv and Schmid [3] can be easily thought of in the context of human performance monitoring applications.

SQC and Human Performance Monitoring

The answer is NO!!! Here is how we reached this conclusion:

(3) If we examine JQT as an example, it had three excellent special issues in 2018:

- **"statistical process control on big data streams"**
- **"quality engineering in advanced manufacturing"**
- **"reliability and maintenance modeling with big data"**

None of the 19 special issue papers considered human performance modeling.

SQC and Human Performance Monitoring

The answer is NO!!! Here is how we reached this conclusion:

(4) Note: In ISQC 2019, we have had two excellent talks that discuss the combination of CPS technologies with individualized health monitoring.

- **"Monitoring Two-state Processes Using Indirectly Observed Data" by: O. Hryniewicz, K. Kaczmarek-Majer, K. Opara**
- **"Personalized Health Management for Elderly Care" by: KL Tsui**

So hopefully, these talks lead to increased interests in an important area that affects public interests and health.

SQC and Human Performance Monitoring

The answer is NO!!! Here is how we reached this conclusion:

Why should SQC care about human performance monitoring applications? If you think of advanced manufacturing, which is our most common application area in SQC and industrial statistics:

In addition to the potential of CPS technologies, there is a direct link of human performance to quality outcomes. In a recent review of 73 published empirical studies by [4], the authors clearly stated that:

> ***Quality deficits were associated with undesirable human effects of workload like fatigue and injury-related risk factors. Forty-six percent of the studies reported on efforts to improve HF in the [operations systems] with effect sizes for quality improvements reaching up to 86%.***

Outline

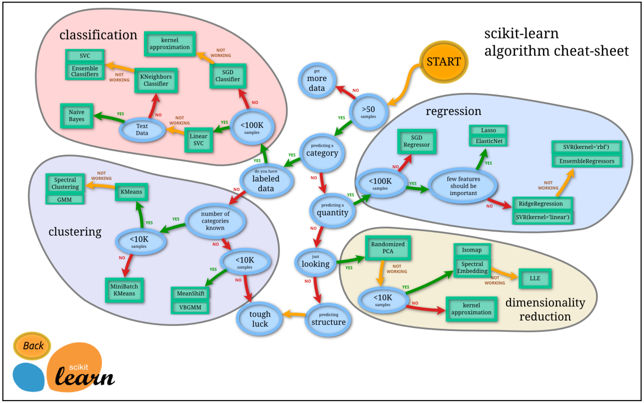
Common Frameworks for AI/ Machine Learning

The KDD Process	**SEMMA**	**CRISP-DM**
Pre KDD	————-	Business understanding
Selection	Sample	Data understanding
Pre processing	Explore	
Transformation	Modify	Data preparation
Data mining	Model	Modeling
Interpretation/Evaluation	Assessment	Evaluation
Post KDD	————-	Deployment

Table: A comparison of the three most commonly used AI/Machine Learning Frameworks. Table adapted from [5].

Common Frameworks for AI/ Machine Learning

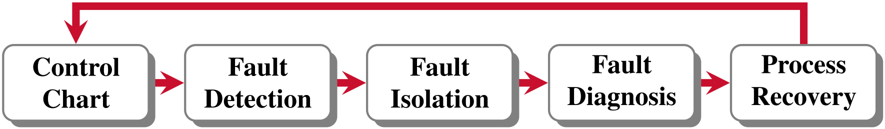
> **These three frameworks encourage iterating, but do not provide sufficient guidance to practitioners/ researchers. Guidance: Typically focuses on the data mining / modeling step.**



Source: SciKit-Learn [6]

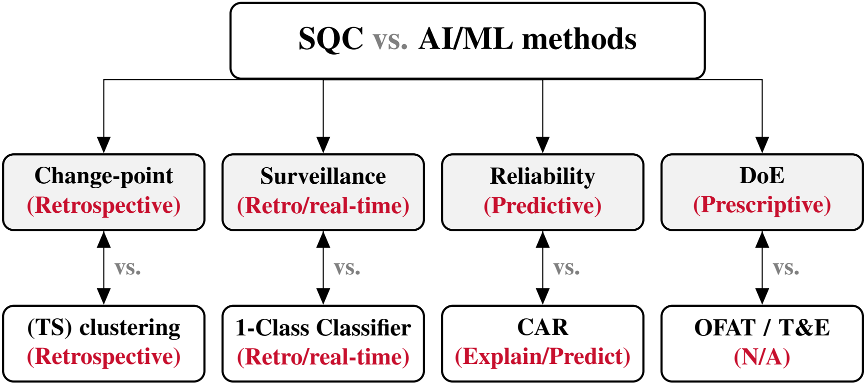
An Engineering Perspective on SQC

In many engineering cases, the practitioner not only cares about a sound monitoring methodology, but also ensuring that the steady/in-control process is recovered as soon as a change occurs.



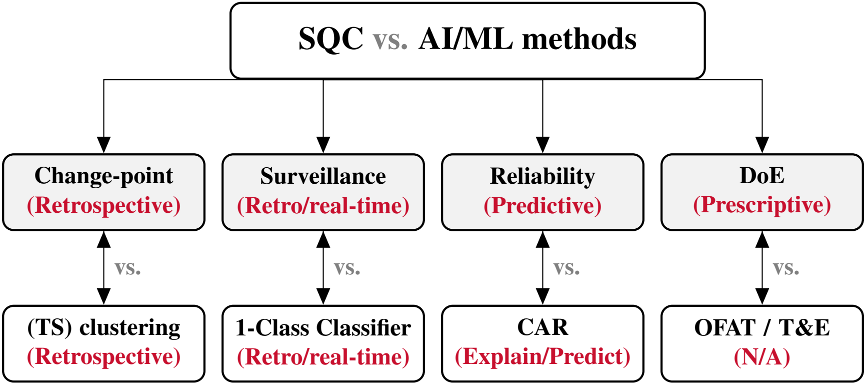
Source: Figure adapted from [7].

Role of SQC in AI/ ML (CPS) Applications



Let us examine three examples (wearables, trucking & cyber security).

Role of SQC in AI/ ML (CPS) Applications



Let us examine three examples (wearables, trucking & cyber security).

Outline

1. Preface
2. The Era of Cyber-Physical Systems (CPS)
3. How SQC can Inform/Augment AI for CPS Applications
4. **Wearable Sensors for Physical Fatigue Management**
5. CPS-Enabled Management Procedures for Unsafe Driving Behaviors
6. An Introduction to biometrics for Cyber/Computer Security
7. Conclusions

Wearables: How I got involved?

BCS Championship - Florida State vs Auburn Box Score, January 6, 2014

Florida State	Auburn
34	**31**
14-0	12-2
« Prev Game	« Prev Game

BCS Championship
Monday Jan 6, 2014
8:30 PM ET
Pasadena, CA

	1	2	3	4	Final
Florida State (1)	3	7	3	21	34
Auburn (2)	7	14	0	10	31

Wearables: How I got involved?

FSU rides GPS technology to title

David M. Hale

Jun 23, 2014

TALLAHASSEE, Fla. -- It was hot and it was muggy and it was a Friday in the middle of summer, all of which should've been enough to strangle any enthusiasm from a group of Florida State's skill-position players running through offseason drills with the team's strength-and-conditioning staff last week. But as freshman tailback Dalvin Cook eased to a stop after an obviously impressive 40-yard sprint, a mad scientist on the sideline with his face buried in a laptop had everyone's attention.

Florida State receiver Rashad Greene has seen the benefits of the technology both on and off the field. AP Photo/Don Juan Moore

The man is Chris Jacobs, an honest-to-goodness rocket scientist tasked with monitoring every movement the Seminoles make in practice and in the weight room. Jacobs had worked as a propulsion engineer with the space program before government cutbacks forced him out of the job, but a timely meeting with a member of Florida State's booster club brought him here.

On the Role of Experimental Design to Collect Data

- **Cross-sectional laboratory study using a one-factor within-subjects design.**
- **Designed factor was the physical level of the task at three levels (low, medium, and high) based on postural, biomechanical, and physiological demand.**

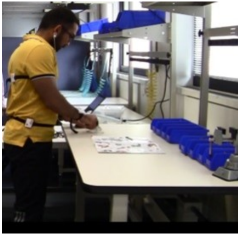
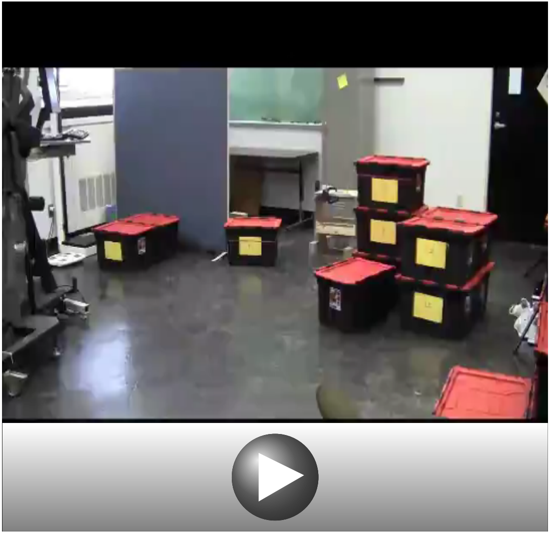


Figure: Part assembly

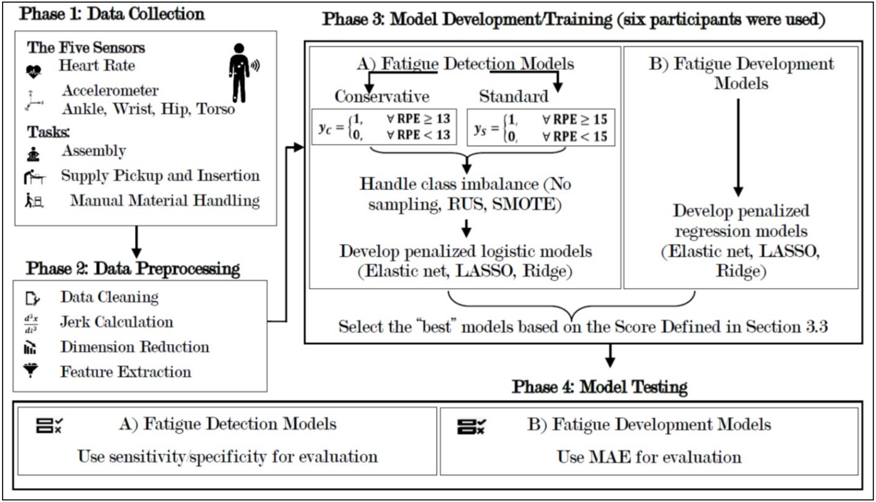
Figure: Supply pickup

Figure: MMH

On the Role of Experimental Design to Collect Data



Preliminary Analysis: Are the sensors a viable option?



For more details, please refer to our published paper in [8].

Preliminary Analysis: Are the sensors a viable option?

Selected features for the logistic LASSO model with RUS sampling (standard scenario).

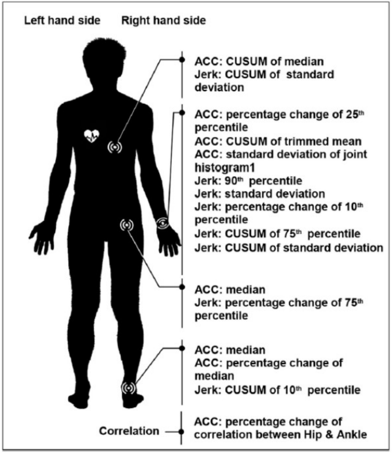
Definition of the Selected Features	Coefficient
Wrist ACC: standard deviation of joint histogram1	0.88
Torso Jerk: CUSUM of standard deviation	0.78
Hip ACC: median	0.39
Wrist Jerk: standard deviation	0.37
Hip Jerk: percentage change of 75th percentile	0.05
Wrist ACC: percentage change of 25th percentile	0.03
Wrist Jerk: 90th percentile	0.001
Ankle Jerk: CUSUM of 10th percentile	−0.03
Wrist Jerk: CUSUM of standard deviation	−0.05
Hip & Ankle ACC: percentage change of correlation	−0.09
Wrist Jerk: percentage change of 10th percentile	−0.19
Ankle ACC: percentage change of median	−0.27
Ankle ACC: median	−0.31
Torso ACC: CUSUM of median	−0.78
Wrist Jerk: CUSUM of 75th percentile	−1.10
Wrist ACC: CUSUM of trimmed mean	−1.20
(Intercept)	−1.28

Training Performance:

 # Features Selected = 16

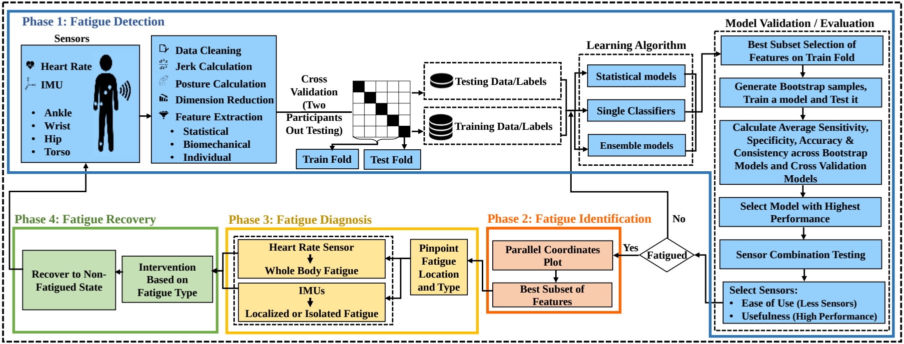
Sensitivity = 0.95

Specificity = 0.89

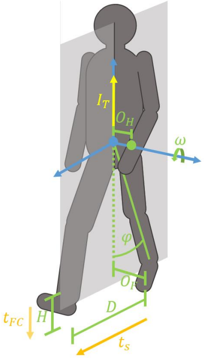


For more details, please refer to our published paper in [8].

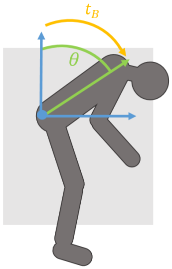
More Data, Better Features ⟶ Less Sensors?



More Data, Better Features $\longrightarrow$ Less Sensors?



I_t: torso vertical impact
O_H: hip oscillation
ω: leg rotational velocity in sagittal plane
φ: leg rotational oscillation
O_F: foot oscillation
D: step length
H: leg raise
t_s: step time
t_{FC}: foot contact time



θ: back bent angle
t_B: time bent

More Data, Better Features $\longrightarrow$ Less Sensors?

Table: Mean performance and the corresponding standard deviation of the classification methods for fatigue detection in MMH task, (the recommended model is **in bold**)

Category	Model	Sensitivity	Specificity	Accuracy	$n_{features}$
BSS	**Bagging**	**0.872 (0.13)**	**0.869 (0.15)**	**0.870 (0.09)**	**5.352**
	Boosting	0.871 (0.13)	0.872 (0.15)	0.870 (0.08)	5.352
	Support Vector Machine	0.811 (0.18)	0.828 (0.17)	0.820 (0.11)	5.352
	Logistic Regression	0.790 (0.17)	0.766 (0.20)	0.778 (0.11)	5.352
LASSO	Penalized Logistic Regression*	0.802 (0.20)	0.916 (0.11)	0.859 (0.11)	18.943
	Penalized Logistic Regression	0.810 (0.13)	0.775 (0.17)	0.793 (0.08)	11.133

* features used in the model are only those generated in our preliminary analysis paper [8]

More Data, Better Features ⟶ Less Sensors?

Table: Mean performance and standard deviation of the Bagging model

# sensors	Sensor Combination					Sensitivity	Specificity	Accuracy
5	Ankle	Hip	Wrist	Torso	HR	0.872 (0.13)	0.869 (0.15)	0.870 (0.09)
4	Ankle	Hip	Wrist	Torso		0.875 (0.13)	0.868 (0.15)	0.871 (0.09)
	Ankle	Hip		Torso	HR	0.850 (0.15)	0.875 (0.13)	0.862 (0.09)
		Hip	Wrist	Torso	HR	0.877 (0.12)	0.863 (0.15)	0.870 (0.08)
	Ankle		Wrist	Torso	HR	0.872 (0.12)	0.864 (0.15)	0.868 (0.08)
	Ankle	Hip	Wrist		HR	-	-	-
3			Wrist	Torso	HR	0.877 (0.12)	0.864 (0.15)	0.870 (0.08)
	Ankle			Torso	HR	0.844 (0.15)	0.874 (0.13)	0.859 (0.09)
	Ankle	Hip		Torso		0.850 (0.15)	0.875 (0.13)	0.862 (0.09)
		Hip	Wrist	Torso		0.877 (0.12)	0.863 (0.15)	0.870 (0.08)
		Hip		Torso	HR	0.859 (0.15)	0.875 (0.14)	0.867 (0.10)
	Ankle		Wrist	Torso		0.873 (0.12)	0.863 (0.15)	0.868 (0.08)
	Ankle	Hip			HR	-	-	-
	Ankle	Hip	Wrist			-	-	-
	Ankle		Wrist		HR	-	-	-
		Hip	Wrist		HR	-	-	-
1				**Torso**		**0.842 (0.15)**	**0.859 (0.14)**	**0.851 (0.10)**
	Ankle					-	-	-
		Hip				-	-	-
			Wrist			-	-	-
					HR	-	-	-

Commentary: On the Use of Statistical and ML Models

- **So far:** SPC has only inspired some of our features and possibly our DoE.
- **Assumptions:** A dichotmous outcome variable (fatigued vs. not fatigued)
- **In practice:** This is not probably how fatigue works.

Commentary: On the Use of Statistical and ML Models

- **So far:** SPC has only inspired some of our features and possibly our DoE.
- **Assumptions:** A dichotmous outcome variable (fatigued vs. not fatigued)
- **In practice:** This is not probably how fatigue works.

Source: Etsy Images

A Change Point Approach: Examining Features

fmegahed.github.io/fatigue_case_jqt.html

Work   MU Mail   MU Calendar   Overleaf - Projects   OSC Dashboard   Machine Learning f...

June 28, 2019

In this paper, we attempt to examine whether the use of change-point analysis techniques is appropriate for detecting fatigue based on data captured from wearable sensors. As such, we perform a **secondary data analysis** to the data generated in: Baghdadi et al., 2018. The reader should note that their raw data was preprocessed using:

A. *Kalman Filter:* Used to process the raw data from sensors to: (i) estimate the spatial orientation of the body with respect to the global reference frame, and (ii) to estimate the kinematics of motion.

B. *Segmentation:* The motion segments where then segmented using an algorithm that assumes the existence of two peaks in the translational acceleration of the gait cycle. This assumption was justified based on the results of Tongen and Wunderlich, 2010 as well as the authors' preliminary analyses.

Since these are not a primary focus of our paper (where detailed in Section 3 of the paper to present the background of how the raw sensor data are transformed to features), we do not detail this analysis in this file. Instead, we refer the reader to the supplementary Matlab files in our GitHub Repository. ** Thus, this R Markdown primarily focuses on the analysis and results described in Sections 4-6 of our paper.**

The reader can **show** any code chunk by clicking on the *code* button. We chose to make the default for the code hidden since we: (a) wanted to improve the readability of this document; and (b) assumed that the readers will not be interested in reading every code chunk.

1 Loading Data & Generating Features

The snippet below documents the list of **R** libraries that were used in this research. For convenience, we used the package since it allows for installing/loading the needed libraries in one step.

Code

1.1 Feature Generation

In the snippet below, we extract the 15 ".mat" files in the GitHub repository (where we loaded the data to allow for the reproduction of our work). Note that these files were originally produced in Baghdadi et al., 2018. Then, we perform several transformation steps: (a) extracting the data for the first three columns in the matlab arrays; and (b) computing three kinematic features from the data corresponding to these columns. Due to the differences between Matlab and R, this requires two nested *for* loops. The outer loop increments over the number of subjects, while the inner loop increments based on the different number of rows of data for each subject. Please see the comments within the code chunk for more details.

Research Opportunity: Low-Hanging SPM Fruit?

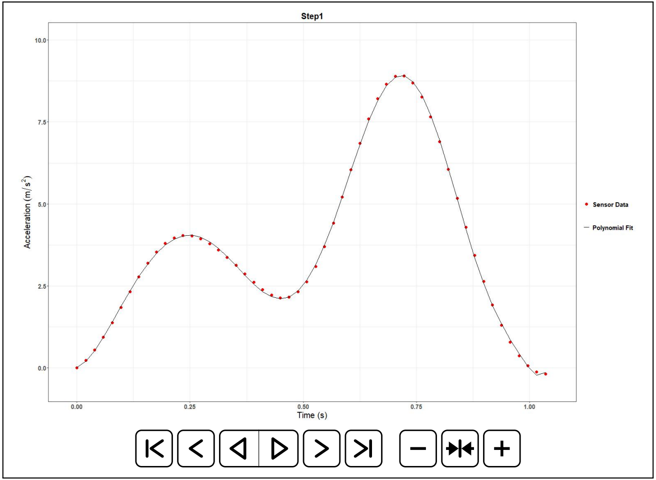


Figure: Ten consecutive gait cycle segments for an arbitrary participant (P5). The x-axis represents time, and the y-axis captures the magnitude of acceleration.

Some Other Research Opportunities [1]

- **Retrospective Analysis:** **Multiple change-point methodology for interdependent multivariate time-series.** (Ideally, allowing for multiple change-points detection)
- **Real-time Surveillance:** **Monitoring the probabilities of the ML/SL models.** (In essence similar to our community's approach to risk-adjustment, e.g., see Steiner et al [9])
- **Predicting Time-to-Fatigue:** **Extending Reliability models for human data seems natural** (see distinction from TTF models and machine learning applications)
  - Fatigue Intervention = Maintenance
  - Fatigue probability $\neq$ Fatigue probability??

Some Other Research Opportunities [1]

- **Retrospective Analysis: Multiple change-point methodology for interdependent multivariate time-series. (Ideally, allowing for multiple change-points detection)**
- **Real-time Surveillance: Monitoring the probabilities of the ML/SL models. (In essence similar to our community's approach to risk-adjustment, e.g., see Steiner et al [9])**
- **Predicting Time-to-Fatigue: Extending Reliability models for human data seems natural (see distinction from TTF models and machine learning applications)**
  - **Fatigue Intervention = Maintenance**
  - **Fatigue probability $\neq$ Fatigue probability??**

Some Other Research Opportunities [1]

- **Retrospective Analysis:** **Multiple change-point methodology for interdependent multivariate time-series.** (Ideally, allowing for multiple change-points detection)
- **Real-time Surveillance:** **Monitoring the probabilities of the ML/SL models.** (In essence similar to our community's approach to risk-adjustment, e.g., see Steiner et al [9])
- **Predicting Time-to-Fatigue:** **Extending Reliability models for human data seems natural** (see distinction from TTF models and machine learning applications)
  - Fatigue Intervention = Maintenance
  - Fatigue probability $\neq$ Fatigue probability??

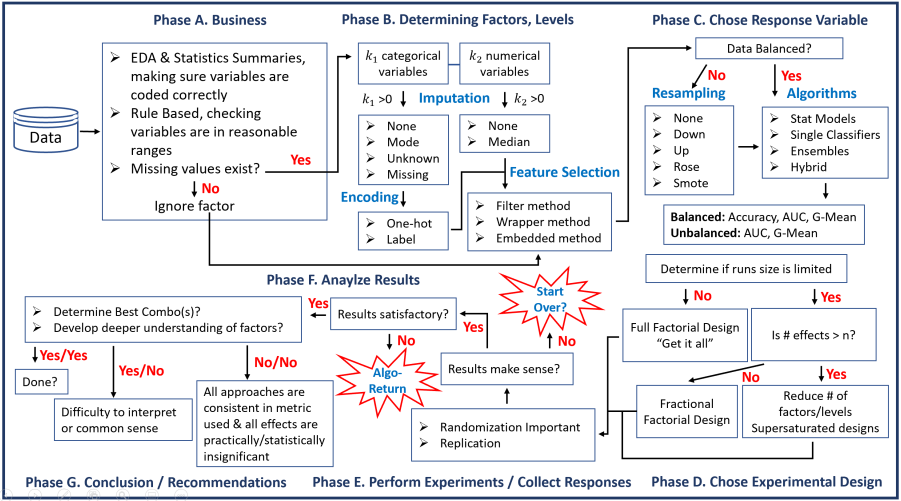
Some Other Research Opportunities [1]

- **Retrospective Analysis:** **Multiple change-point methodology for interdependent multivariate time-series.** (Ideally, allowing for multiple change-points detection)
- **Real-time Surveillance:** **Monitoring the probabilities of the ML/SL models.** (In essence similar to our community's approach to risk-adjustment, e.g., see Steiner et al [9])
- **Predicting Time-to-Fatigue:** **Extending Reliability models for human data seems natural** (see distinction from TTF models and machine learning applications)
  - **Fatigue Intervention = Maintenance**
  - Fatigue probability $\neq$ Fatigue probability??

Some Other Research Opportunities [1]

- **Retrospective Analysis: Multiple change-point methodology for interdependent multivariate time-series.** (Ideally, allowing for multiple change-points detection)
- **Real-time Surveillance: Monitoring the probabilities of the ML/SL models.** (In essence similar to our community's approach to risk-adjustment, e.g., see Steiner et al [9])
- **Predicting Time-to-Fatigue: Extending Reliability models for human data seems natural** (see distinction from TTF models and machine learning applications)
  - Fatigue Intervention = Maintenance
  - **Fatigue probability $\neq$ Fatigue probability??**

Some Other Research Opportunities [2]



Outline

Description of the Data

Data Sources

Real-time pings (lat + long + driver ID + speed)

Weather (visibility + precipitation prob + precipitation intensity + wind speed)

Road geometry (speed limits + number of lanes + type of road)

Driver demographics

SCE (hard brake + headway + collison mitigation + rolling stability)

Description of the Data

Trips: real-time pings stopped at $\geq$ 20 minutes

Shifts: Trips are further divided into different shifts if the driver was not moving for at least eight hours.

Description of the Data

Table 1: ping data

ping_time	speed	latitude	longitude	driver
2015-10-23 08:00:00	0	33.94360	-118.1681	canj1
2015-10-23 08:08:10	0	33.94358	-118.1681	canj1
2015-10-23 08:09:26	5	33.94288	-118.1681	canj1
2015-10-23 08:22:58	4	33.97146	-118.1677	canj1
2015-10-23 08:23:12	8	33.97178	-118.1677	canj1

Table 2: Transformed trips data

driver	trip_id	start_time	end_time	trip_time	distance
canj1	100160724	2015-10-23 08:09:26	2015-10-23 08:37:26	28	4.473
canj1	100160725	2015-10-23 09:04:24	2015-10-23 11:21:24	137	46.721
canj1	100160726	2015-10-23 12:00:36	2015-10-23 15:37:36	217	164.576
canj1	100160727	2015-10-23 16:38:10	2015-10-23 18:37:10	119	52.907
canj1	100160728	2015-10-26 07:49:04	2015-10-26 10:52:04	183	104.085

Table 3: Transformed 30-minute intervals

driver	interval_id	start_time	end_time	interval_time	distance
canj1	197089	2015-10-23 08:09:26	2015-10-23 08:38:00	28	4.538
canj1	197090	2015-10-23 09:04:24	2015-10-23 09:34:24	30	2.645
canj1	197091	2015-10-23 09:34:24	2015-10-23 10:04:24	30	0.984
canj1	197092	2015-10-23 10:04:24	2015-10-23 10:34:24	30	5.928
canj1	197093	2015-10-23 10:34:24	2015-10-23 11:04:24	30	17.348

Table 4: Transformed shifts data

driver	shift_ID	shift_start	shift_end	shift_length
canj1	1	2015-10-23 08:09:26	2015-10-23 18:37:56	628
canj1	2	2015-10-26 07:49:04	2015-10-26 15:06:58	437
canj1	3	2015-10-27 01:59:48	2015-10-27 07:58:56	359
canj1	4	2015-10-28 08:05:08	2015-10-28 20:20:32	735
canj1	6	2015-10-30 09:27:12	2015-10-30 21:18:22	711

Table 5: safety critical events

driver	event_time	event_type
canj1	2015-10-23 14:46:08	HB
canj1	2015-10-26 15:06:03	HB
canj1	2015-10-28 11:58:24	HB
canj1	2015-10-28 17:42:36	HB
canj1	2015-11-02 07:13:56	HB

Table 6: drivers

driver	age
canj1	46
farj7	54
gres0	55
hunt	48
kell0	51
lewr10	27
rice30	34
smiv	49
sunc	37
woow59	24

Table 7: Road geometry from the OpenStreetMap API

driver	latitude	longitude	speed_limit	num_lanes
farj7	30.32650	-89.86389	65	2
farj7	30.34032	-91.73116	65	2
farj7	30.34174	-91.72572	60	2
farj7	30.35075	-91.69085	60	2
farj7	30.35165	-91.68755	60	2

Table 8: weather from the DarkSky API

ping_time	latitude	longitude	precip_intensity	precip_probability	wind_speed	visibility
2015-10-23 08:09:26	33.94288	-118.1681	0	0	0.21	9.82
2015-10-23 08:22:58	33.97146	-118.1677	0	0	0.22	9.81
2015-10-23 08:23:12	33.97178	-118.1677	0	0	0.22	9.81
2015-10-23 08:23:30	33.97233	-118.1678	0	0	0.22	9.81
2015-10-23 08:38:00	34.00708	-118.1798	0	0	0.24	9.81

Description of the Data

All data and code available at https://github.com/caimiao0714/ISQC2019truck

Note on Hierarchical Bayesian Logit & Poisson Models

Bayesian Random-Effects Logistic Regression Model:

$$
\begin{aligned}
Y_i &\sim \text{Bernoulli}(p_i) \\
\log \frac{p_i}{1-p_i} &= \beta_{0,d(i)} + \beta_{1,d(i)} \cdot \text{CT}_i + \sum_{j=1}^{J} x_{ij}\beta_j \\
\beta_{0,d} &\sim \text{i.i.d. } N(\mu_0, \sigma_0^2), \quad d = 1, 2, \cdots, D \\
\beta_{1,d} &\sim \text{i.i.d. } N(\mu_1, \sigma_1^2), \quad d = 1, 2, \cdots, D
\end{aligned}
\tag{1}
$$

- Y_i: whether SCEs occurred during the interval, 0 or 1,
- CT_i: cumulative driving time,
- $\beta_{0,d}$: random intercepts for each driver,
- $\beta_{1,d}$: random slopes for CT_i for each driver,
- x_{ij}: covariates, including age, road geometry, and weather.

Note on Hierarchical Bayesian Logit & Poisson Models

Bayesian Random-Effects Poisson Regression Model:

$$
\begin{aligned}
Y_i &\sim \text{Poisson}(T_i \cdot \lambda_i) \\
\log \lambda_i &= \beta_{0,d(i)} + \beta_{1,d(i)} \cdot \text{CT}_i + \sum_{j=1}^{J} x_{ij}\beta_j \\
\beta_{0,d} &\sim \text{i.i.d. } N(\mu_0, \sigma_0^2), \quad d = 1, 2, \cdots, D \\
\beta_{1,d} &\sim \text{i.i.d. } N(\mu_1, \sigma_1^2), \quad d = 1, 2, \cdots, D
\end{aligned} \tag{2}
$$

- Y_i: the number of SCEs during the interval, a non-negative integer,
- T_i: length of the interval as an offset term,
- CT_i: cumulative driving time,
- $\beta_{0,d}$: random intercepts for each driver,
- $\beta_{1,d}$: random slopes for CT_i for each driver,
- x_{ij}: covariates, including age, road geometry, and weather.

Note on Hierarchical Bayesian Logit & Poisson Models

Weakly informative priors and hyper-priors as recommended by Gelman [10].

$$
\begin{aligned}
\mu_0 &\sim N(0, 5^2) \\
\mu_1 &\sim N(0, 5^2) \\
\sigma_0 &\sim \text{Gamma}(1, 1) \\
\sigma_1 &\sim \text{Gamma}(1, 1) \\
\beta_2, \beta_3, \cdots, \beta_J &\sim N(0, 10^2)
\end{aligned}
\tag{3}
$$

- μ_0, σ_0: hyper-parameters for $\beta_{0,d}$,
- μ_1, σ_1: hyper-parameters for $\beta_{1,d}$,
- $\beta_2, \beta_3, \cdots, \beta_J$: fixed parameters for covariates x_{ij}.

Non-homogeneous Poisson Process: Introduction

- The **intensity function** of a point process is

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(N(t, t + \Delta t] \geq 1)}{\Delta t} \tag{4}$$

- **Nonhomogeneous Poisson Process (NHPP)**: a Poisson process whose intensity function is non-constant.
- **Power law process (PLP)**: the intensity function of a NHPP is

$$\lambda(t) = \frac{\beta}{\theta} \left(\frac{t}{\theta} \right)^{\beta - 1}, \quad \beta > 0, \theta > 0. \tag{5}$$

- β is the shape parameter. θ is the scale parameter.
- $\beta > 1 \rightarrow$ reliability deterioration; $\beta > 1 \rightarrow$ reliability improvement.

Non-homogeneous Poisson process: data

- $T_{d,s,i}$: the time to the d-th driver's s-th shift's i-th critical event,
- d: driver index, $d = 1, 2, \cdots, D$
- s: shift index, $s = 1, 2, \cdots, S_d$,
- i: SCE index, $i = 1, 2, \cdots, n_{d,S_d}$
- $n_{d,s}$: the total number critical events of d-th driver's s-th shift.

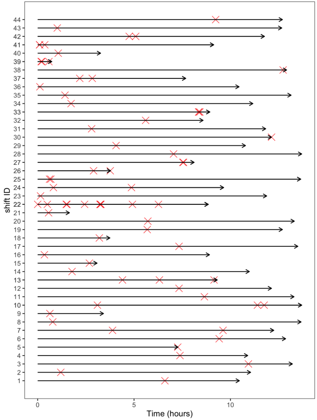


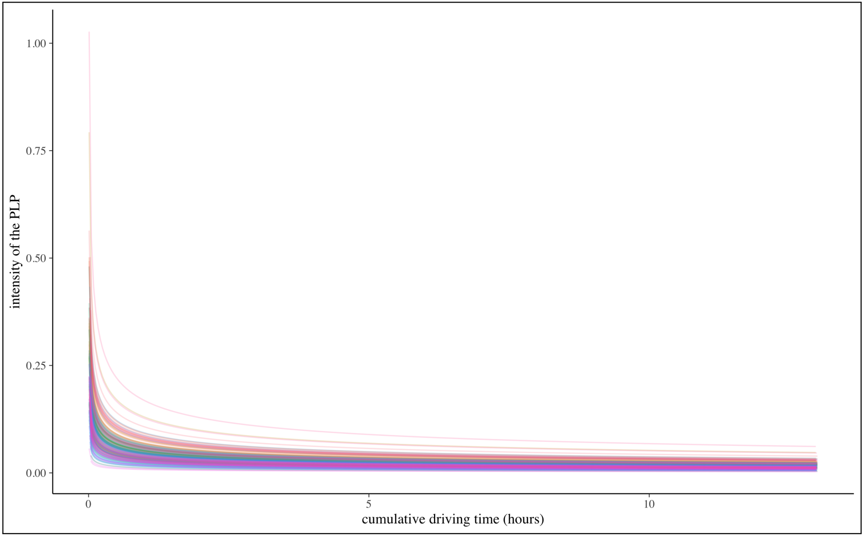
Figure: Arrow plot of time to SCEs in each shift

Bayesian random-intercepts NHPP

$$\begin{aligned}
T_{d,s,1}, T_{d,s,2}, \cdots, T_{d,s,n_{d,s}} &\sim \text{PLP}(\beta, \theta_{d,s}) \\
\beta &\sim \text{Gamma}(1,1) \\
\log \theta_{d,s} &= \gamma_{0d} + \gamma_1 x_{d,s,1} + \gamma_2 x_{d,s,2} + \cdots + \gamma_k x_{d,s,k} \\
\gamma_{01}, \gamma_{02}, \cdots, \gamma_{0D} &\sim \text{i.i.d. } N(\mu_0, \sigma_0^2) \\
\gamma_1, \gamma_2, \cdots, \gamma_k &\sim \text{i.i.d. } N(0, 10^2) \\
\mu_0 &\sim N(0, 5^2) \\
\sigma_0 &\sim \text{Gamma}(1,1)
\end{aligned} \tag{6}$$

- a fixed β across drivers,
- random parameters $\theta_{d,s}$ across drivers
- $\theta_{d,s}$: random intercepts γ_{0d} for scale parameter θ,
- $x_{d,s,k}$: covariates.

Initial Results for the Hierarchical Bayesian NHPP



Opportunities for Future Research

- **Model tuning/optimization: Incorporating the effects of weather, road geometry, etc in the model**
- **Parallelization: Facilitating the computations for the entire dataset of 12,000 drivers.**

Opportunities for Future Research

- **Model tuning/optimization:** Incorporating the effects of weather, road geometry, etc in the model
- **Parallelization:** Facilitating the computations for the entire dataset of 12,000 drivers.

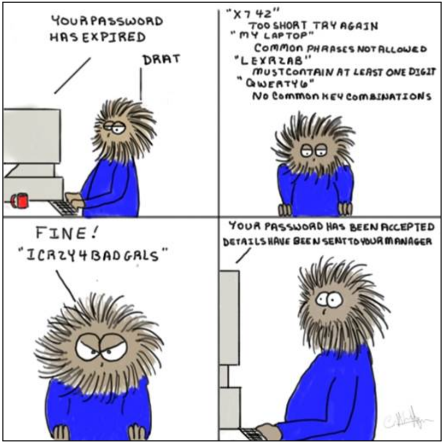
Outline

Human-Human Authentication

- User authentication
  - Differentiate one human user from another
- Prominent authentication approaches
  - Passwords
  - Traditional biometrics

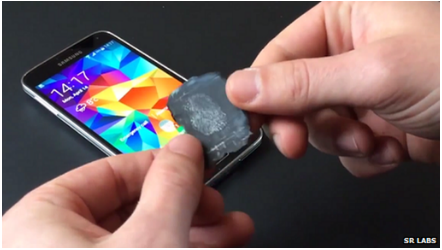
Limitations of Existing User Authentication Solutions

- Passwords
  - Either insecure or unusable
- Traditional biometrics (e.g., fingerprints)
  - Invasive
  - High rejection rates
  - Require additional hardware
  - Susceptible to impersonation or spoofing



Limitations of Existing User Authentication Solutions

- Passwords
  - Either insecure or unusable
- Traditional biometrics (e.g., fingerprints)
  - Invasive
  - High rejection rates
  - Require additional hardware
  - Susceptible to impersonation or spoofing

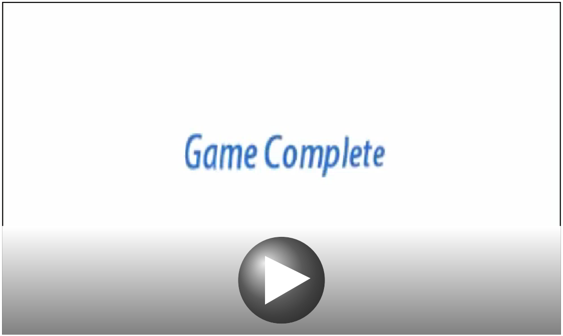


Behavioral Biometrics

- **Keystroke dynamics** [11]
- **Mouse movement patterns** [12]
- **Touch gesture biometrics**
  - **Sliding horizontally and vertically** [13]
  - **Sliding up, down, left, and right and tap** [14]
  - **Horizontal slide and the pattern unlock** [15]

Gametrics: An Example Based on [16]

- **Interactive game-based behavioral biometrics**
- **Why games?**
  - Fully supported by web browsers and touch screen devices
  - Randomized, interactive and cognitive nature
  - Sufficient cues can be extracted in a short period of time



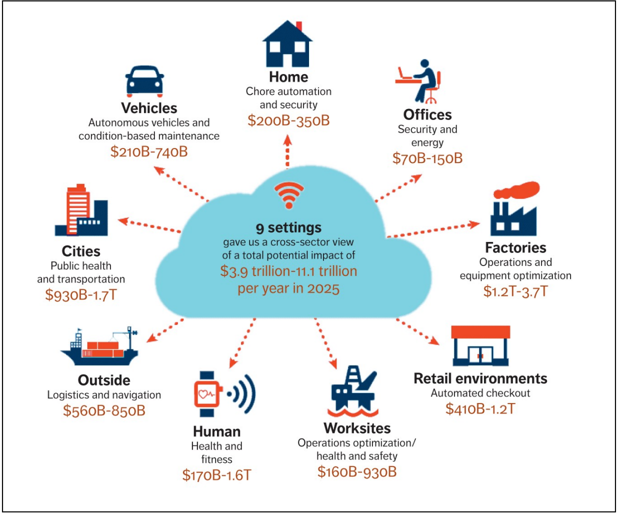
Future Research Opportunities

Using statistical methods to characterize/model wihin-user variations.

- **In my opinion, we should not treat this as a binary classification problem.**

Outline

Summary ("Luck is What Happens When Preparation Meets Opportunity")



Source: McKinsey&Company. The McKinsey Global Institute [2].

Acknowledgments

Key Contributors to the Projects

- Lora Cavuoto, PhD — University at Buffalo
- Tessa Chen, PhD — University of Dayton
- Zahra Sadighi Maman, PhD — Adelphi University
- Mohammad Ali Alamdar Yazdi, PhD — John Hopkins University
- Alex Vinel, PhD — Auburn University

Acknowledgments

This work was supported in part by: the **National Science Foundation** (CMMI-1635927 and CMMI-1634992); the **Ohio Supercomputer Center** (PMIU0138 and PMIU0162); the **American Society of Safety Professionals (ASSP) Foundation**; the **University of Cincinnati Education and Research Center Pilot Research Project Training Program**; the **Transportation Informatics Tier I University Transportation Center (TransInfo)**; a **Google Cloud Platform research grant** for data management; and a **Dark Sky grant** for extended API access. Dr. Megahed's research was also partially supported by the **Neil R. Anderson Endowed Assistant Professorship at Miami University**.

References [1]

[1] Cyber physical systems (CPS) — NSF 19-553.
https://www.nsf.gov/pubs/2019/nsf19553/nsf19553.htm, 2019.
[Online. Last accessed August 4, 2019].

[2] The Internet of Things: Mapping the value beyond the hype.
McKinsey&Company. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/The%20Internet%20of%20Things%20The%20value%20of%20digitizing%20the%20physical%20world/The-Internet-of-things-Mapping-the-value-beyond-the-hype.ashx, 2015.
[Online. Last accessed August 8, 2019].

[3] Taras Lazariv and Wolfgang Schmid.
Challenges in monitoring non-stationary time series.
In *Frontiers in Statistical Quality Control 12*, pages 257–275. Springer, 2018.

[4] Ahmet Kolus, Richard Wells, and Patrick Neumann.
Production quality and human factors engineering: a systematic review and theoretical framework.
Applied Ergonomics, 73:55–89, 2018.

References [2]

[5] Ana Isabel Rojão Lourenço Azevedo and Manuel Filipe Santos.
Kdd, semma and crisp-dm: a parallel overview.
IADS-DM, 2008.

[6] Choosing the right estimator.
https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html, 2019.
[Online. Last accessed August 4, 2019].

[7] Leo H Chiang, Evan L Russell, and Richard D Braatz.
Fault detection and diagnosis in industrial systems.
Springer Science & Business Media, 2000.

[8] Zahra Sedighi Maman, Mohammad Ali Alamdar Yazdi, Lora A Cavuoto, and Fadel M Megahed.
A data-driven approach to modeling physical fatigue in the workplace using wearable sensors.
Applied ergonomics, 65:515–529, 2017.

References [3]

[9] Stefan H Steiner, Richard J Cook, Vern T Farewell, and Tom Treasure.
Monitoring surgical performance using risk-adjusted cumulative sum charts.
Biostatistics, 1(4):441–452, 2000.

[10] Andrew Gelman, Daniel Simpson, and Michael Betancourt.
The prior can often only be understood in the context of the likelihood.
Entropy, 19(10):555, 2017.

[11] Fabian Monrose and Aviel Rubin.
Authentication via keystroke dynamics.
In *Proceedings of the 4th ACM Conference on Computer and Communications Security*, pages 48–56. Citeseer, 1997.

[12] Nan Zheng, Aaron Paloski, and Haining Wang.
An efficient user verification system via mouse movements.
In *Proceedings of the 18th ACM Conference on Computer and Communications Security*, pages 139–150. ACM, 2011.

References [4]

[13] Mario Frank, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song.
Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication.
IEEE transactions on information forensics and security, 8(1):136–148, 2012.

[14] Lingjun Li, Xinxin Zhao, and Guoliang Xue.
Unobservable re-authentication for smartphones.
In *NDSS*, volume 56, pages 57–59, 2013.

[15] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann.
Touch me once and i know it's you!: implicit authentication based on touch screen patterns.
In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 987–996. ACM, 2012.

References [5]

[16] Manar Mohamed and Nitesh Saxena.
Gametrics: towards attack-resilient behavioral authentication with simple cognitive games.
In *Proceedings of the 32nd Annual Conference on Computer Security Applications*, pages 277–288. ACM, 2016.

A Statistical (Process Monitoring) Perspective on Human Performance Modeling in the Age of Cyber-Physical Systems

Fadel M. Megahed[1], L. Allison Jones-Farmer[1], Miao Cai[2], Steven E. Rigdon[2] and Manar Mohamed [1]

[1]Miami University, OH, USA

[2]Saint Louis University, MO, USA

August 14th, 2019