Mobile Sensor Big Data Software Platforms from MD2K

Santosh Kumar, Ph.D.
Director, MD2K Center of Excellence
Moss Chair of Excellence Professor in Computer Science
University of Memphis
MD2K Team
78 faculty, students, and staff spanning 13 funded projects

Advancing biomedical discovery and improving health through mobile sensor big data

Cornell Tech ♦ Georgia Tech ♦ Harvard ♦ U. Memphis ♦ Northwestern ♦ Ohio State ♦ Open mHealth
UCLA ♦ UC San Diego ♦ UC San Francisco ♦ UMass Amherst ♦ U. Michigan ♦ Utah ♦ WVU
MD2K Investigators at the Tech Showcase Today

Polo Chau
Georgia Tech

Emre Ertin
Ohio State

Deborah Estrin
Cornell Tech

Zach Ives
UPenn

Deepak Ganesan
UMASS

Ben Marlin
UMASS

Ida Sim
UCSF
Utility of Mobile Sensor Big Data

Development/Validation of Markers & Digital Biobank

Marker Discovery & Validation

- Lab Studies
- Labeled Data Collection
- Field Studies
- Ground Truth Video

Markers

- Visualizations
- Self reports
- Interventions
- Marker Data

Biomedical Research

Advancing biomedical discovery and improving health through mobile sensor big data

Cornell Tech ✧ Georgia Tech ✧ Harvard ✧ U. Memphis ✧ Northwestern ✧ Ohio State ✧ Open mHealth
UCLA ✧ UC San Diego ✧ UC San Francisco ✧ UMass Amherst ✧ U. Michigan ✧ Utah ✧ WVU
Mobile Sensors to Collect Labeled Data

- **AutoSense sensors**: ECG, respiration, accelerometers
- **EasySense (contactless) sensors**: heart motion, lung motion, lung fluid level
- **Smartphone sensors**: GPS, accelerometers, self-report
- **Microsoft Band**: accelerometers, gyroscopes, HR
- **MotionSense HRV**: accelerometers, gyroscopes, PPG
- **Smart toothbrush**: brushing, Pressure
Mobile Sensor Big Data Software on Smartphone

70 million samples/day
200 gigabytes/day

Ground Truth Video

Sense

Data Quality

MotionSense
AutoSense
Phone Sensors

Markers
Stress
Smoking
Eating

Self-reports
Questionnaires
Interventions

EMA/EMI Scheduler

Act

Micro-randomization
JITAI

Cerebral Cortex

Event Labels and Annotations

AND/OR

Advancing biomedical discovery and improving health through mobile sensor big data

Cornell Tech • Georgia Tech • Harvard • U. Memphis • Northwestern • Ohio State • Open mHealth
UCLA • UC San Diego • UC San Francisco • UMass Amherst • U. Michigan • Utah • WVU
Mobile Sensor Big Data Capabilities in mCerebrum

1. High speed data collection
   - 9,500 samples/s using data exchange architecture

2. High volume storage using append-only model
   - Performance within 92% of optimal

3. Microbatching of data ingestion for efficiency
   - Incurs 8.4 times lower CPU usage vs. AWARE

4. Real-time computation of biomarkers
   - Stress, smoking, driving, activity, etc.

5. Biomarker-triggered notification/intervention
Cerebral Cortex – Use Case Scenarios

Study 1
Study 2
Study N

Raw Sensor Data
Real-time participant monitoring
Marker Development and Validation
Analysis of Marker Data

Cerebral Cortex

Study Coordinators
Data Science Researchers
Health Researchers

Advancing biomedical discovery and improving health through mobile sensor big data

Cornell Tech ♦ Georgia Tech ♦ Harvard ♦ U. Memphis ♦ Northwestern ♦ Ohio State ♦ Open mHealth
UCLA ♦ UC San Diego ♦ UC San Francisco ♦ UMass Amherst ♦ U. Michigan ♦ Utah ♦ WVU
Field Deployments of MD2K Platforms

<table>
<thead>
<tr>
<th>Study</th>
<th>Users</th>
<th>Person-Days</th>
<th>Samples (Billions)</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking/Eating</td>
<td>225</td>
<td>3,150</td>
<td>136</td>
<td>9TB</td>
</tr>
<tr>
<td>Smoking</td>
<td>300</td>
<td>4,200</td>
<td>182</td>
<td>12TB</td>
</tr>
<tr>
<td>Smoking</td>
<td>300</td>
<td>4,200</td>
<td>182</td>
<td>12TB</td>
</tr>
<tr>
<td>Smoking/fMRI</td>
<td>90</td>
<td>1,260</td>
<td>55</td>
<td>3.5TB</td>
</tr>
<tr>
<td>Smoking/Stress</td>
<td>24</td>
<td>336</td>
<td>15</td>
<td>1TB</td>
</tr>
<tr>
<td>Heart Failure</td>
<td>225</td>
<td>6,750</td>
<td>224</td>
<td>15TB</td>
</tr>
<tr>
<td>Oral Health</td>
<td>162</td>
<td>29,160</td>
<td>968</td>
<td>65TB</td>
</tr>
<tr>
<td>Cocaine Use</td>
<td>25</td>
<td>350</td>
<td>18</td>
<td>1.5TB</td>
</tr>
<tr>
<td>Behavior Change</td>
<td>100</td>
<td>1,400</td>
<td>58</td>
<td>4TB</td>
</tr>
<tr>
<td>Job Performance</td>
<td>800</td>
<td>56,000</td>
<td>2,891</td>
<td>185TB</td>
</tr>
<tr>
<td>Totals</td>
<td>2,251</td>
<td>106,806</td>
<td>4,729</td>
<td>300TB</td>
</tr>
</tbody>
</table>

Both mCerebrum and Cerebral Cortex are open-source licensed.
mHealth Biomarker Highlights

- SMOKING
- STRESS
- CRAVING
- GEO-EXPO
- ACTIVITY & POSTURE
- COCAINE
- EATING
- CONVERSATION
- TYPING
- SLEEPING
- HEART FAILURE
- ORAL HEALTH
- FAST FOOD
- VISUAL EXPO
- FATIGUE

Advancing biomedical discovery and improving health through mobile sensor big data

Cornell Tech ♦ Georgia Tech ♦ Harvard ♦ U. Memphis ♦ Northwestern ♦ Ohio State ♦ Open mHealth
UCLA ♦ UC San Diego ♦ UC San Francisco ♦ UMass Amherst ♦ U. Michigan ♦ Utah ♦ WVU
An Example of MD2K Marker Data Streams

Location Data Streams

- $ST_1$, $ET_1$, ‘Home’
- $ST_2$, $ET_2$, ‘Work’
- $ST_3$, $ET_3$, ‘Other’
- ...
- $ST_n$, $ET_n$, ‘Home’

Feature Computation

- Cumulative Staying Time at a given place
- Transition Frequency between two centroids
- The total distance covered
- The maximum distance between two location
- The standard deviation of the displacement
- The maximum distance from ‘home’, ‘work’
- The number of different significant places visited
- The number of different places visited
- The radius of gyration
- The routine Index

GPS Data

Self-report Data

Model to mark POIs trained on labeled data

Places Of Interest marking

- Compute locations (gps co-ordinates) dwelled by participants for 10+ minutes from GPS time series.
- Using google_places_api, mark places of interest (POI) in close vicinity of the computed locations.

Feature Computation

- Cumulative Staying Time at each of the places of interest each day, during the study period
Mobile Sensor Big Data Cloud Capabilities

Real-time Data Processing

Sensor data

Questionnaires

Storage System

Web data

Temporal Annotations

Heterogeneous Data Streams

Concurrent Marker Development

Library of 400+ Marker Data Streams

Sensor data

Questionnaires

Storage System

Web data

Temporal Annotations

Real-time Data Processing

Concurrent Marker Development

Library of 400+ Marker Data Streams

Stress

Smoking

Mobility

Typing

MD2K
Center of Excellence for Mobile Sensor Data-to-Knowledge

Advancing biomedical discovery and improving health through mobile sensor big data

Cornell Tech • Georgia Tech • Harvard • U. Memphis • Northwestern • Ohio State • Open mHealth

UCLA • UC San Diego • UC San Francisco • UMass Amherst • U. Michigan • Utah • WVU
mCerebrum and Cerebral Cortex Personal Edition

mCerebrum Android App
https://md2k.org/mcp

Installation Instructions
https://md2k.org/personal

Software:
https://github.com/MD2Korg/

Feedback/Suggestions:
software@md2k.org

Mailing List:
md2k-software@googlegroups.com

Open source licensed: BSD 2-Clause

Live Demo at Tables 14, 15 and 16
MD2K Software Team

Timothy Hnat, Ph.D.
Chief Software Architect

Nasir Ali, Ph.D.
Research Asst. Prof.

Anand Tirtha, Ph.D.
Data Science Software Engineer

Syed Monowar Hossain, Ph.D.
Lead Software Engineer

Nusrat Nasrin, M.S.
Mobile Software Engineer

Advancing biomedical discovery and improving health through mobile sensor big data

Cornell Tech  Georgia Tech  Harvard  U. Memphis  Northwestern  Ohio State  Open mHealth
UCLA  UC San Diego  UC San Francisco  UMass Amherst  U. Michigan  Utah  WVU