IBM **Research**
Australia

Temporal and Spatial Data Analysis with Applications

Mahsa Salehi
August 2016

Brief Biography

2002

- Amirkabir University of Technology (Tehran Polytechnique)
  - Double B.Sc degrees
    - Software engineering
    - Information Technology

2004

2006

  - M.Sc degree in software engineering

2012

- University of Melbourne
  - PhD in computer science

2014

- Research scientist at Emotiv

2015

- Postdoctoral Researcher at IBM Research
  - Spatio-Temporal Data Analysis in Emergency management

IBM Research Overview

Famous for its science and vital to IBM

3000 Researchers

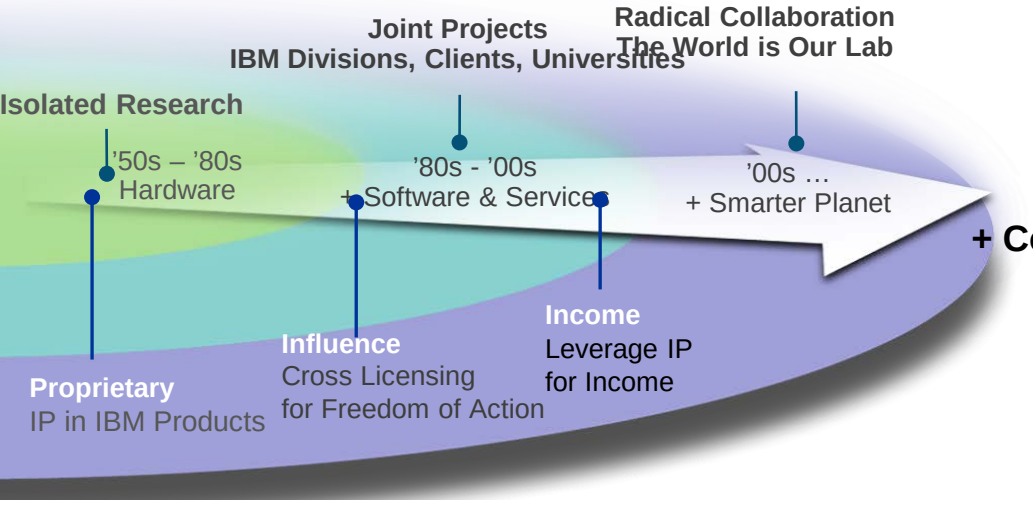
6 Nobel Laureates

20+ years of Patent Leadership

Innovation that Matters

$6B R&D Budget

6 Turing Awards



The IBM Research – Australia Lab

- **IBM Research - Australia** is located
  Level 5, 204 Lygon Street
  Carlton, VIC 3053
- **Internship program:**
  http://www.research.ibm.com/careers/

Outline

- What is Data Mining?
- What is Spatial and Temporal Data Mining?
- Diagnosis of Phenomena
- Driver Distraction Detection
- Disaster Management
  - Bushfire risk prediction
  - Fire Fighting Appliance Pre-Deployment (Demo)
  - Evacuation planner (Demo)
- Summary

What is Data Mining?

- The practice of examining large pre-existing databases in order to generate new information.

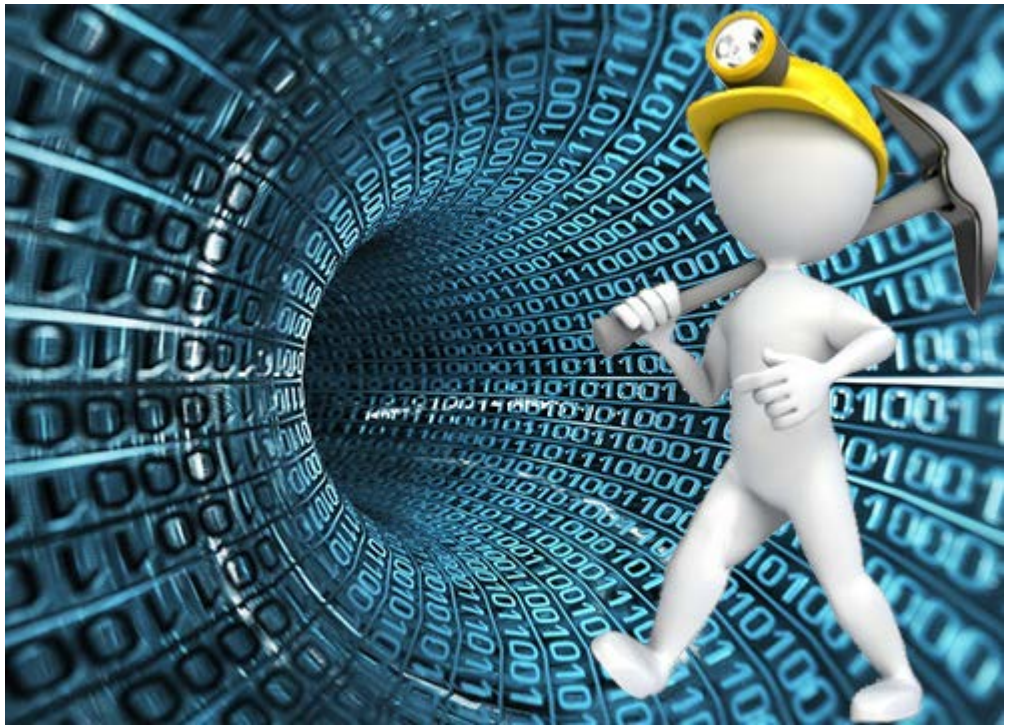


Image invensis.net

- Examples? Features?

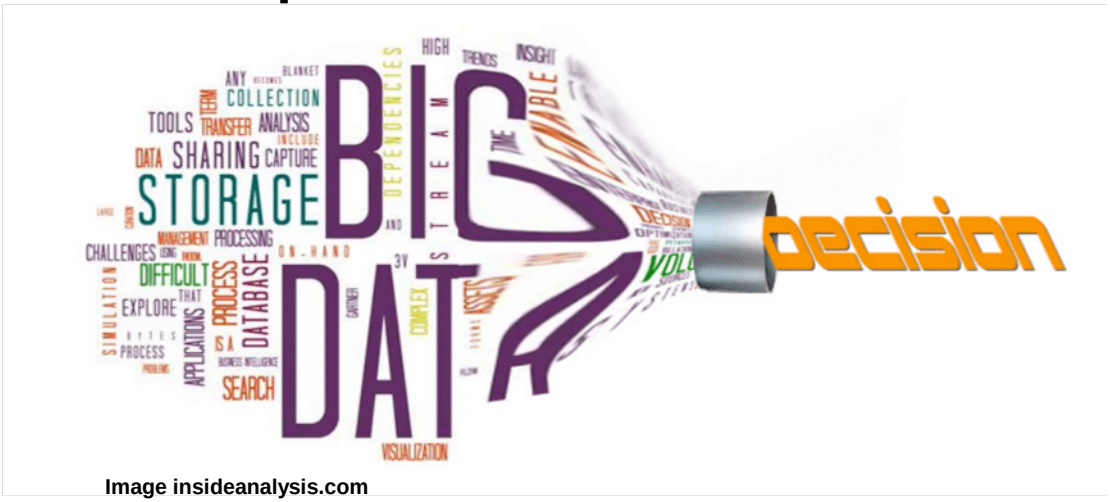


Image insideanalysis.com

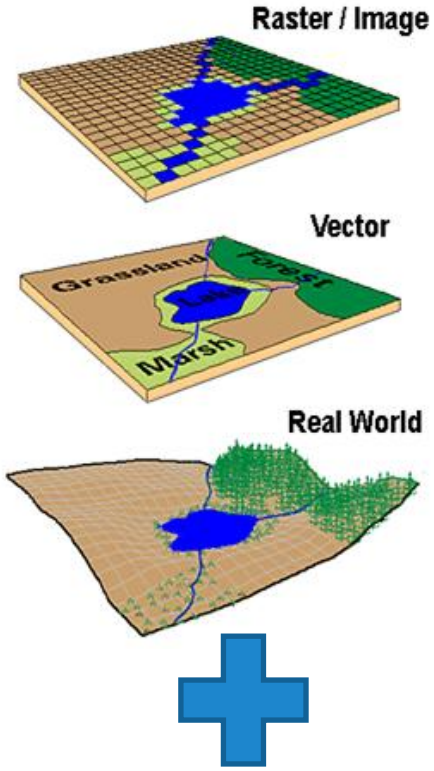
What is the Internet of Things (IoT)?

- Interconnection of uniquely identifiable and interoperable objects (e.g. Sensors, Mobile Phones) with Cloud Computing via the Internet backbone.

Image gadgets.ndtv.com

Different Data Types

- Spatial data (also known as geospatial data):
  - is information about a physical object that can be represented by numerical values in a geographic coordinate system.



- Temporal data:
  - is data that varies over time.

- Spatio-Temporal data:
  - is the integration of space and time. Multidimensional data like points, line segments, regions, polygons, volumes, or other kinds of geometric entities that vary in the course of time.

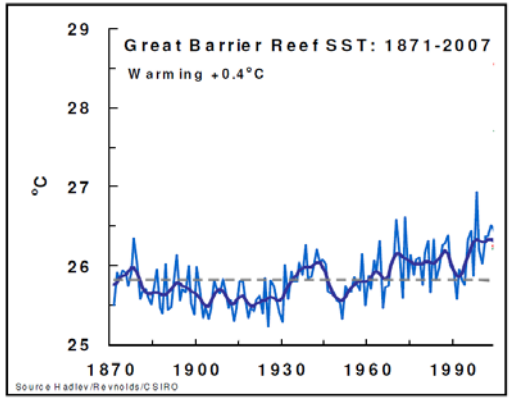


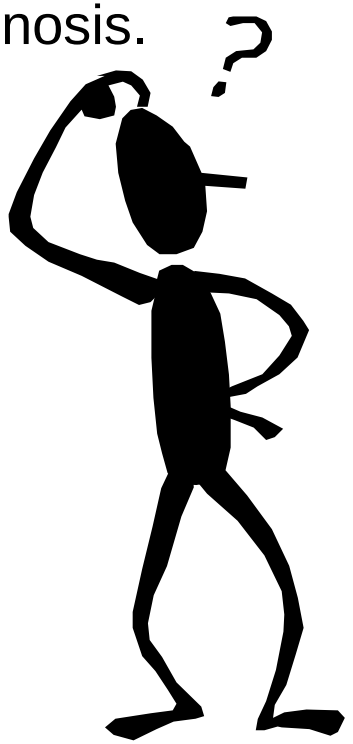
Image planet.botany.uwc.ac.za

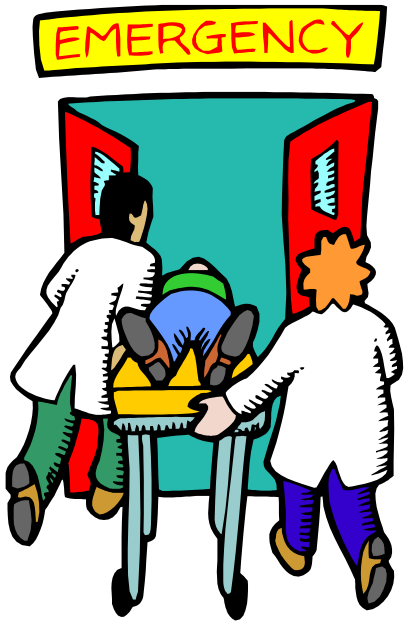
Application 1: Diagnosis of Pneumonia (Microsoft Imagine Cup Project)

We won the first place in Australia, and were ranked among the top 20 in the world!

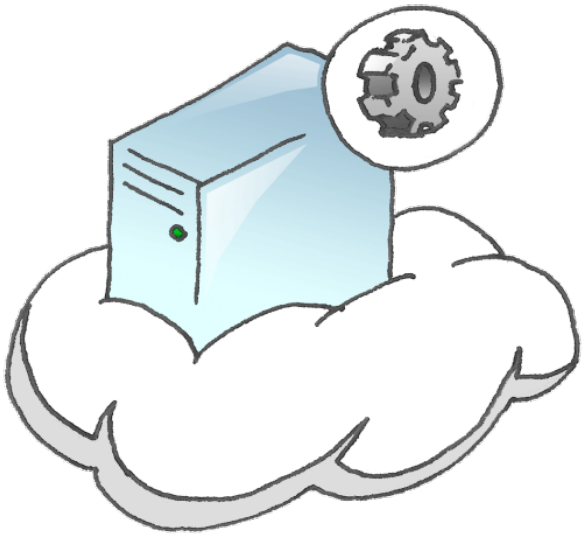
Diagnosing pneumonia is difficult

- Pneumonia is the single biggest killer of children.
- It kills 1.5 million children under 5 per year, mostly in developing countries.
- The key to change this lies in early diagnosis.

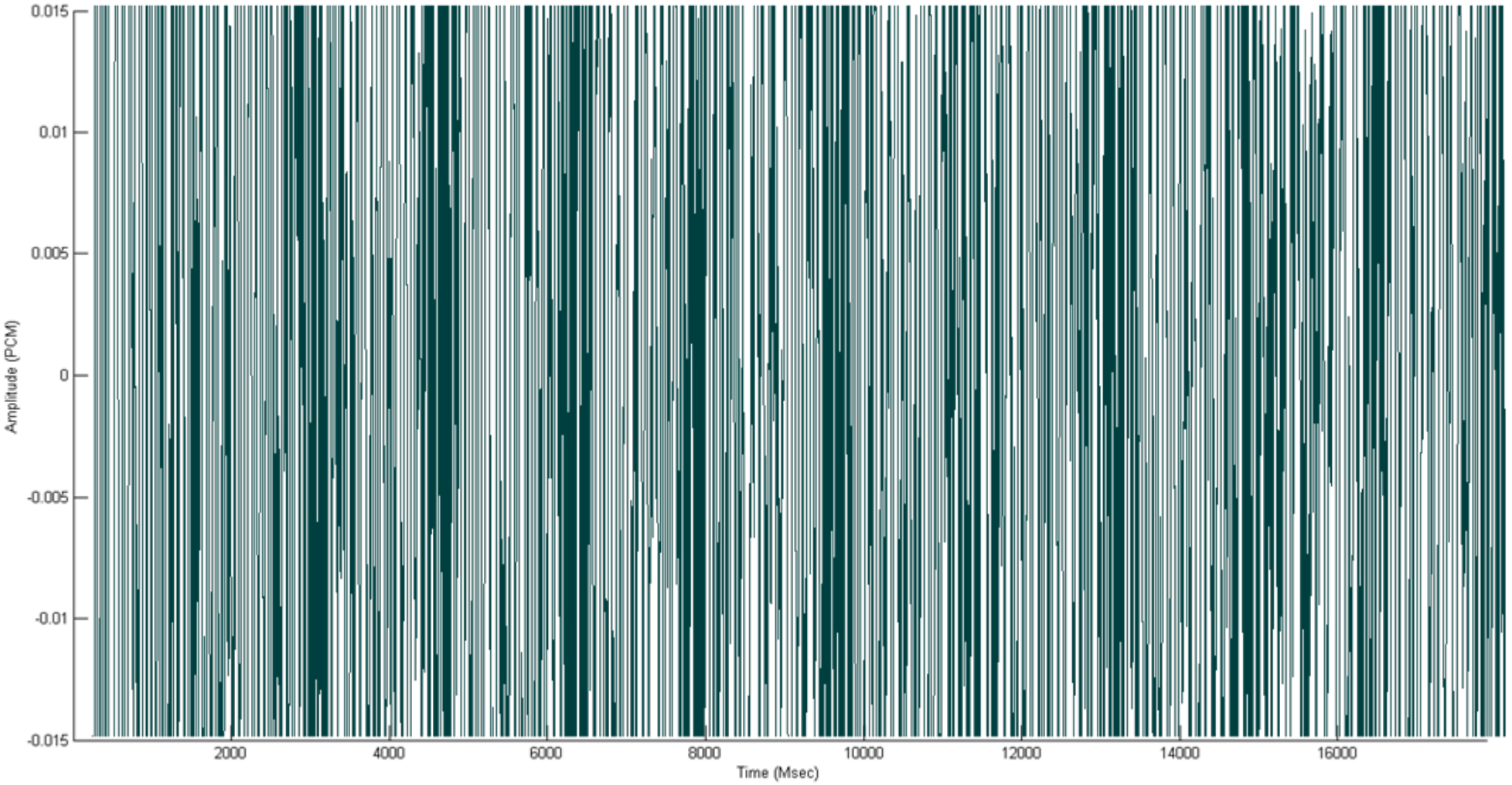


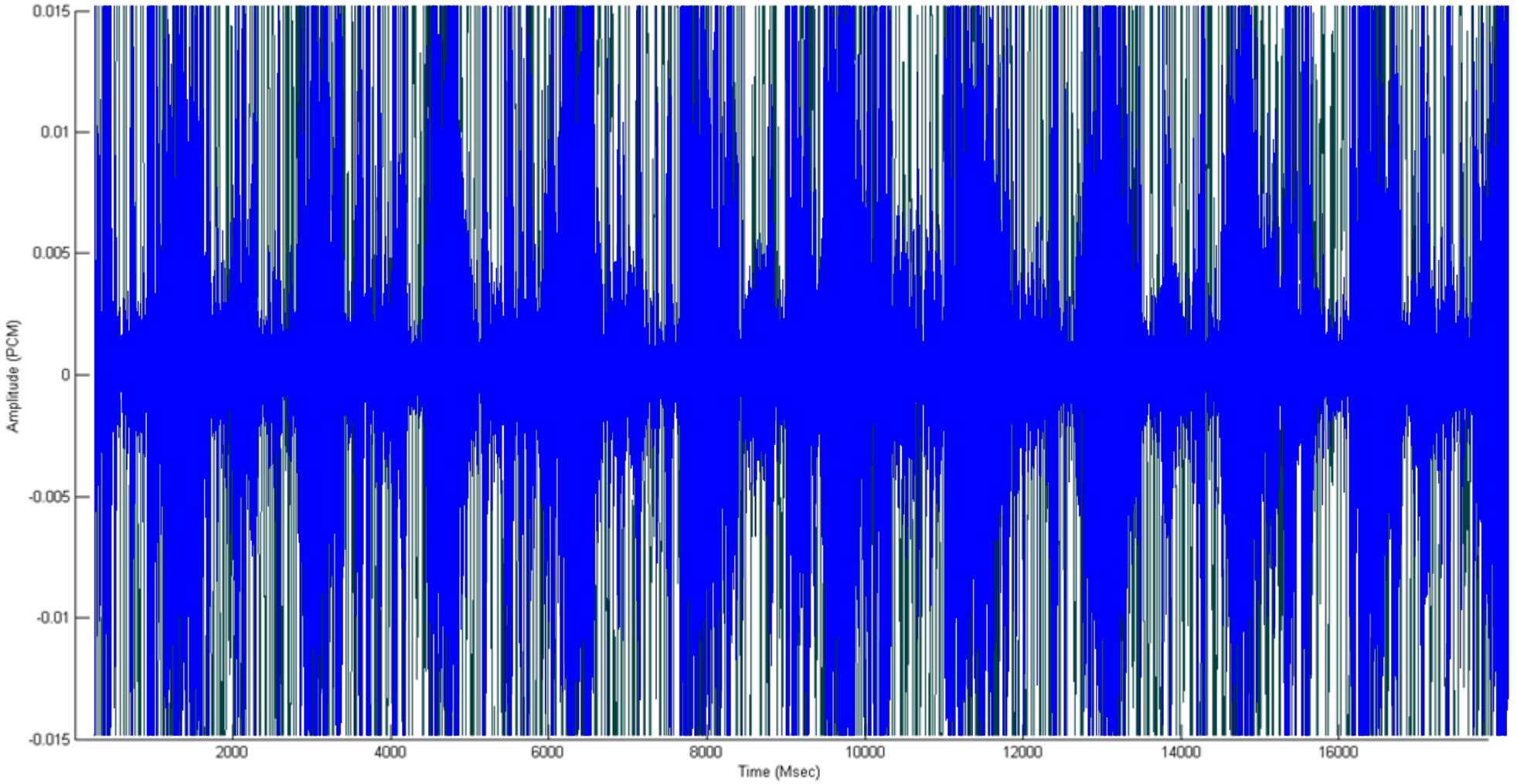


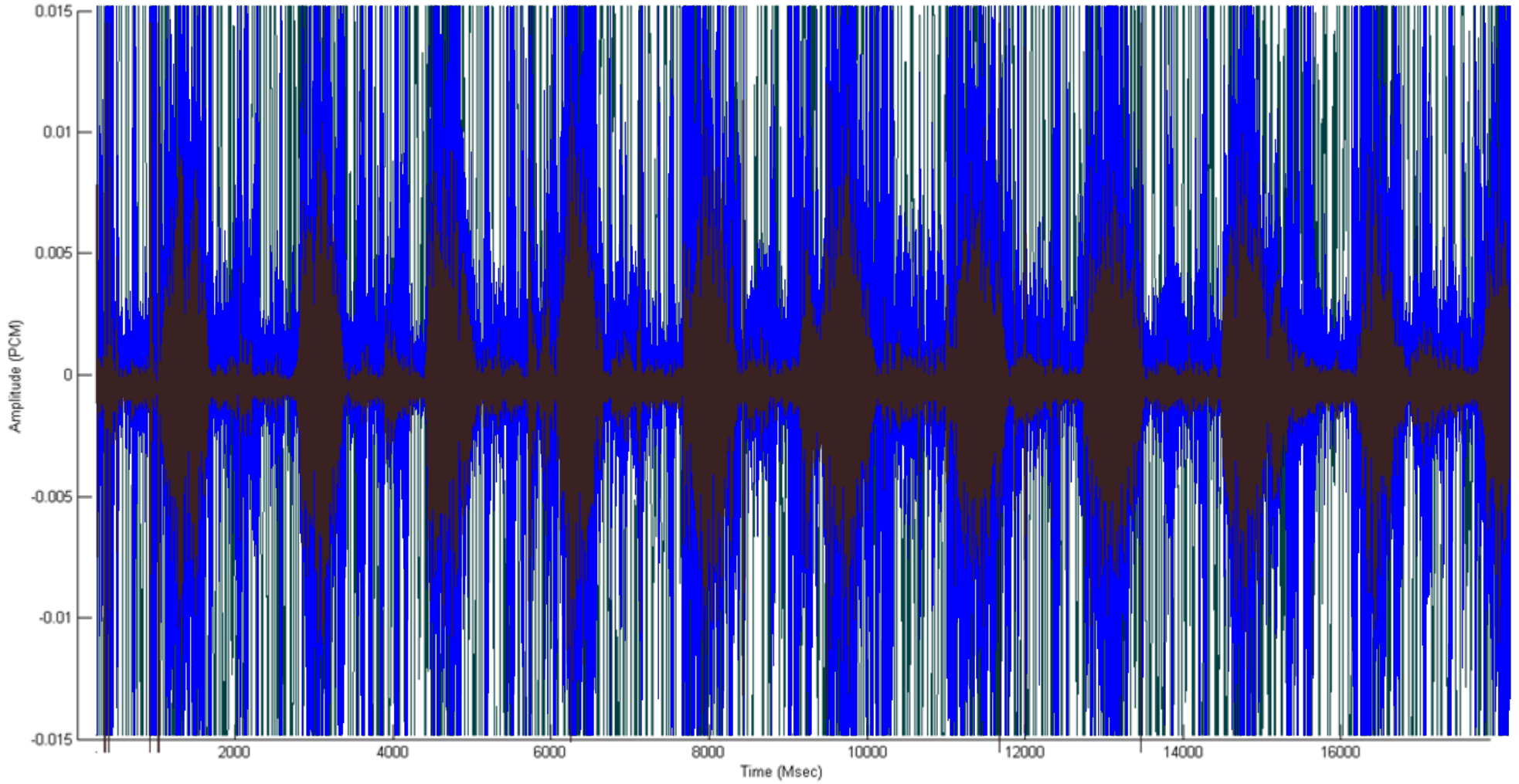
Our IoT based solution: StethoCloud

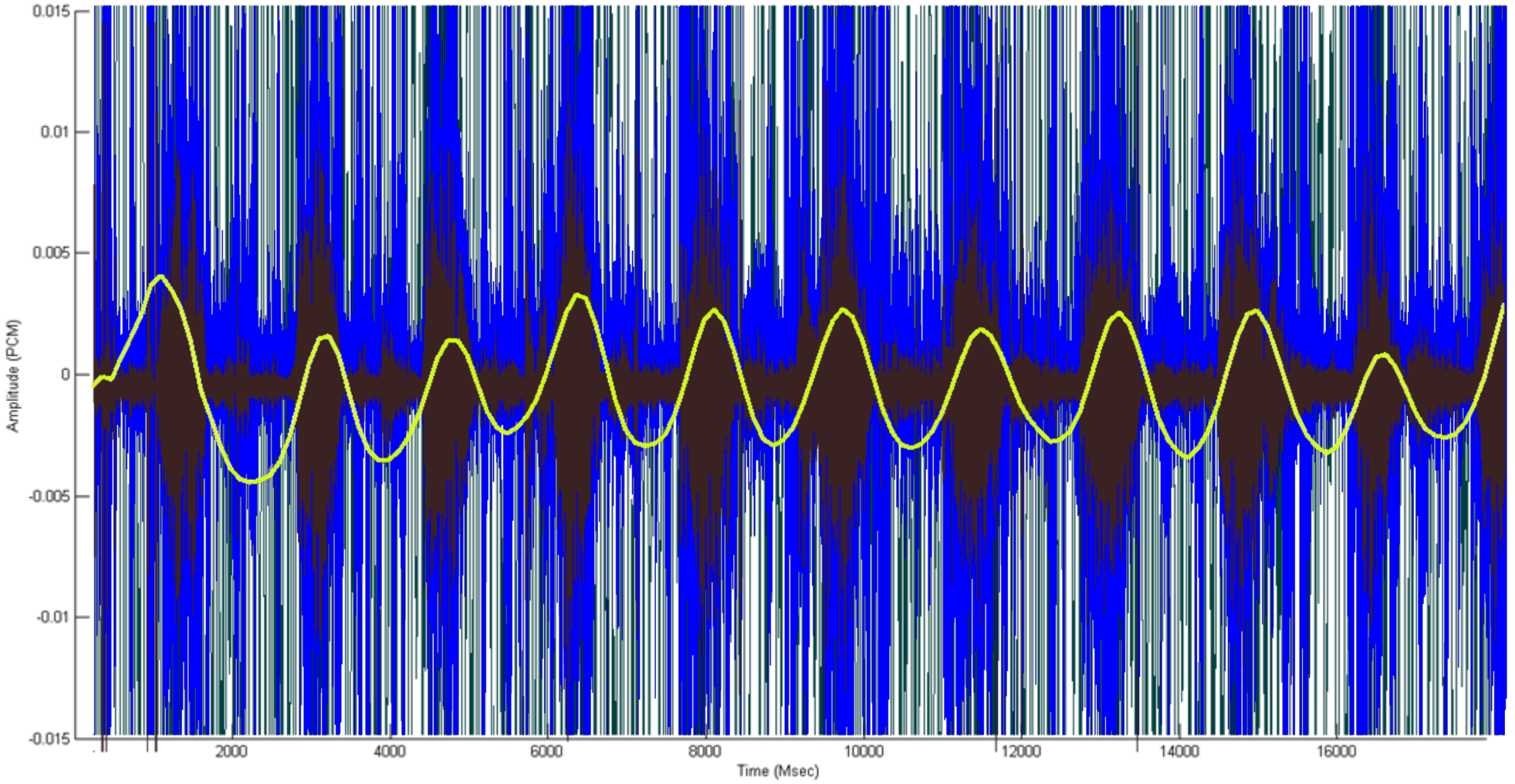


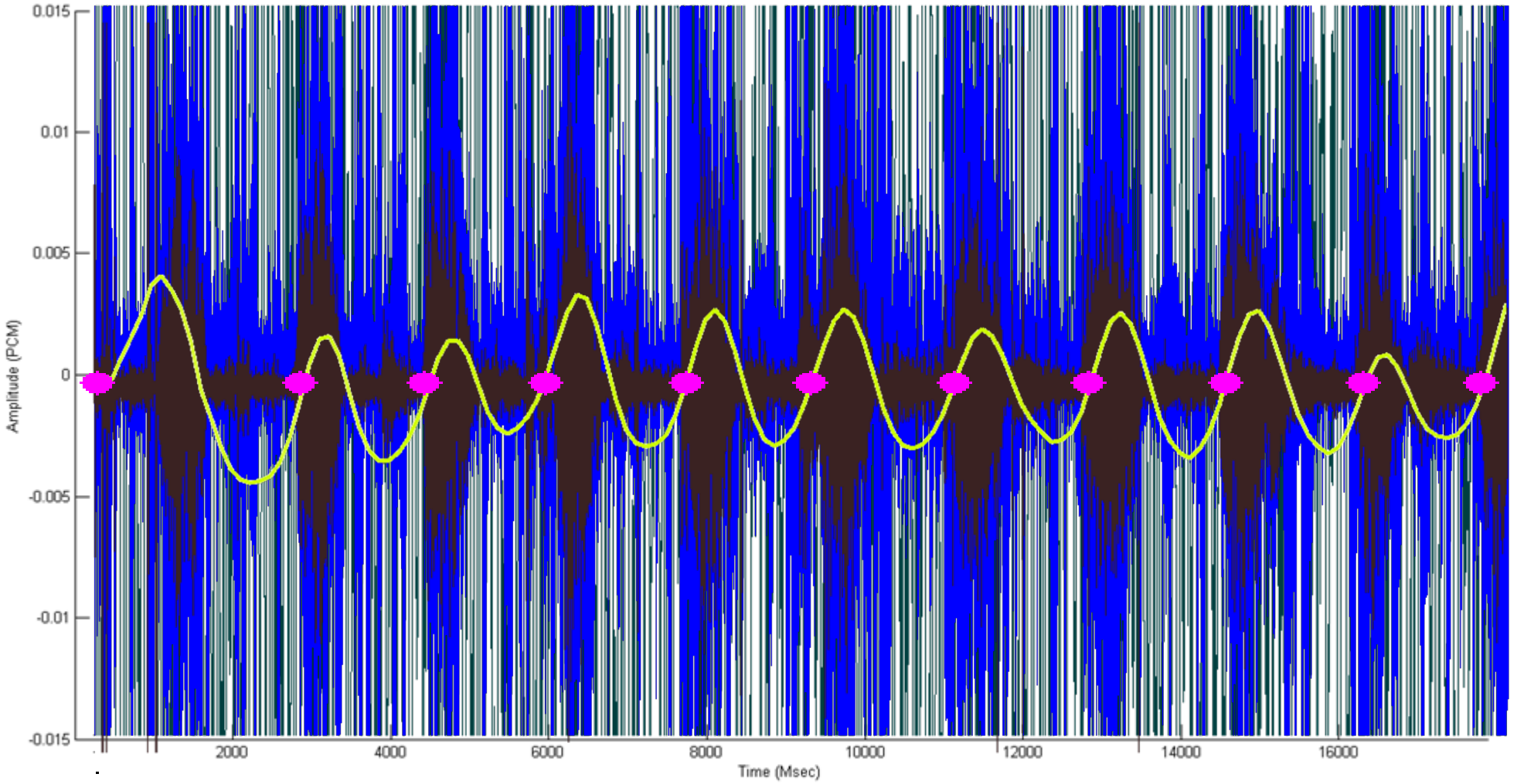
Respiratory Data is collected from six different locations


0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)


0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)


0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)


0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)

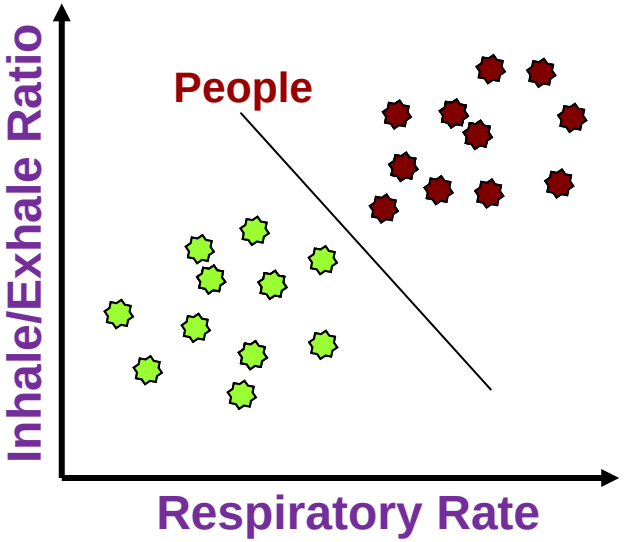

0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)

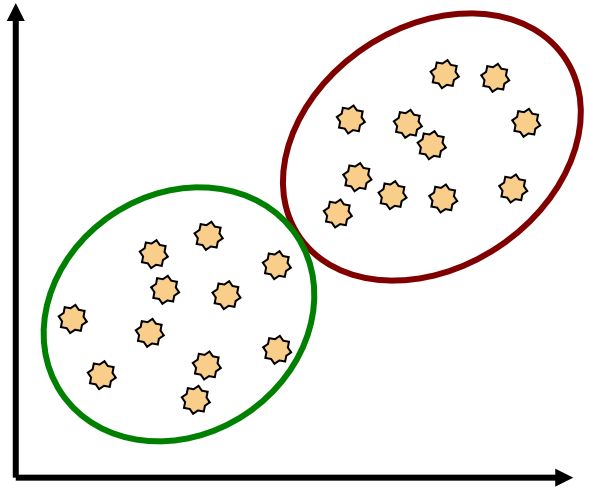
Classification and Clustering

- Objects characterized by one or more features

- Classification
  - Have labels for some points
  - Want a “rule” that will accurately assign labels to new points
  - Supervised learning

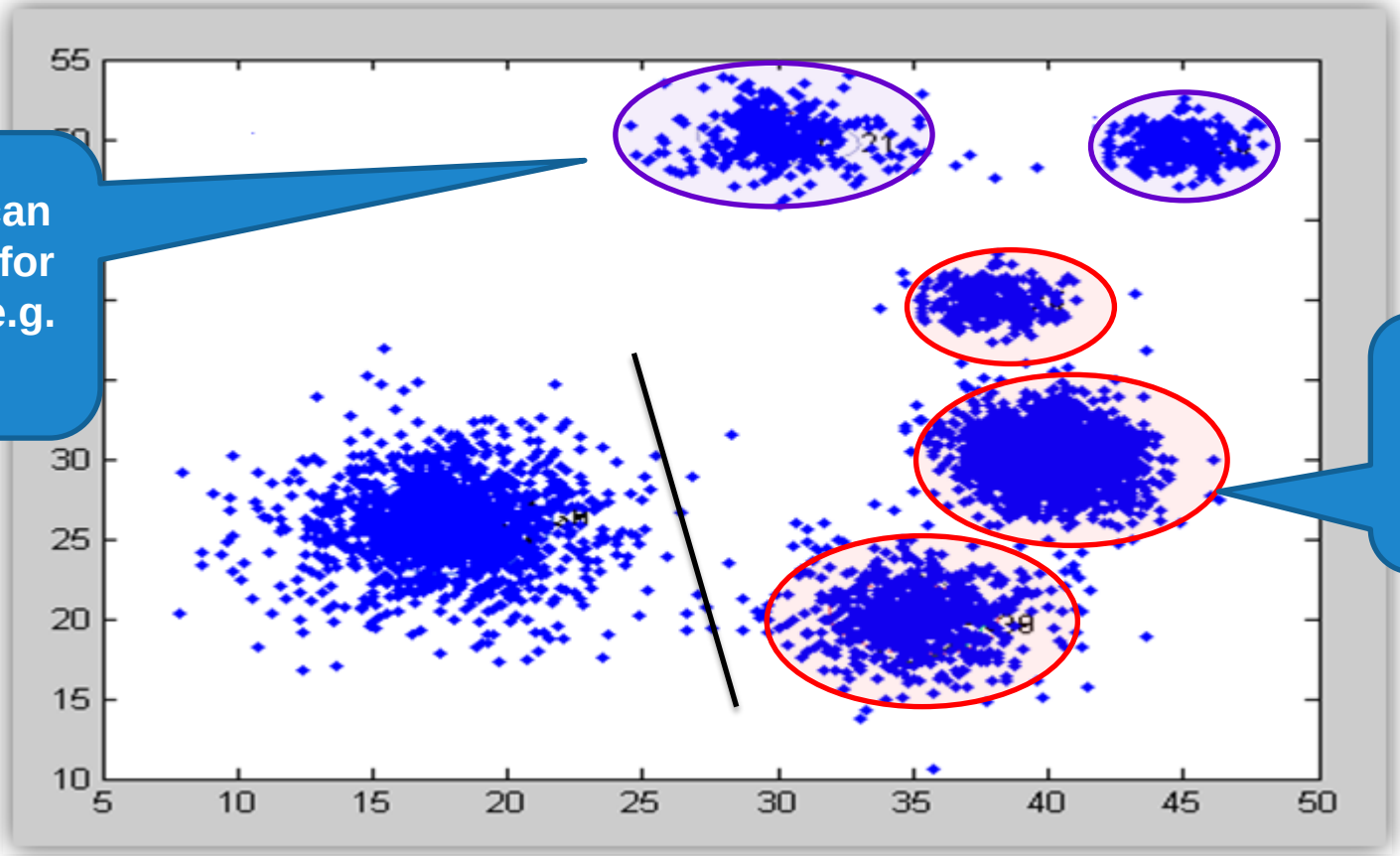
- Clustering
  - No labels
  - Group points into clusters based on how “near” they are to one another
  - Identify structure in data
  - Unsupervised learning





This is what the data might look like

Other Diseases



This technique can directly be used for other diseases, e.g. asthma

Clustering

Possible patterns emerging with more data, e.g., different subtypes

Normal Subject

Pneumonia

Classification

Application 2: Car-Racing Driver Distraction Detection

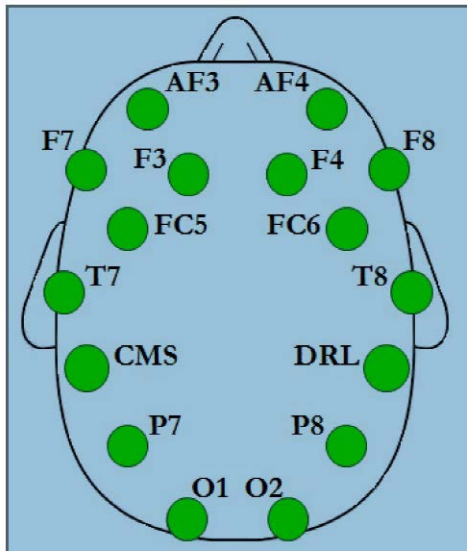
Image emotiv.com

Why is it important?

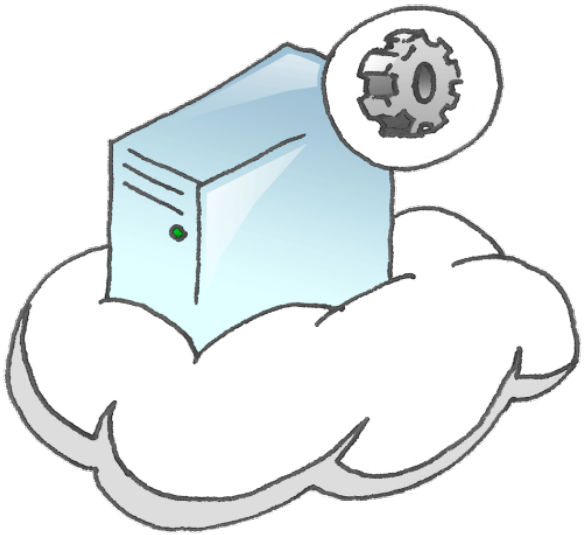
- Monitoring driver attention has a direct effect on decreasing injury/fatality rates, and improving his performance.
- In **car racing** environments reaction times are short, and distraction leads to a reduction in driver's performance during a race.

Our solution: Brain Computer Interface (BCI) Approach

Emotiv EPOC headset



14 Brain EEG Signals
+ Gyroscope



Experimental Setup

- Simulation

- Talking to passenger
- Mobile call, Recording call
- Solving simple mental arithmetic challenges placed on the road, assessed by the driving through the chosen answer from a range of alternatives on screen

EEG data analysis

- Classification
  - Random forest has best results
  - 73.5% detection accuracy with EEG
  - 81% detection accuracy with EEG + Gyroscope

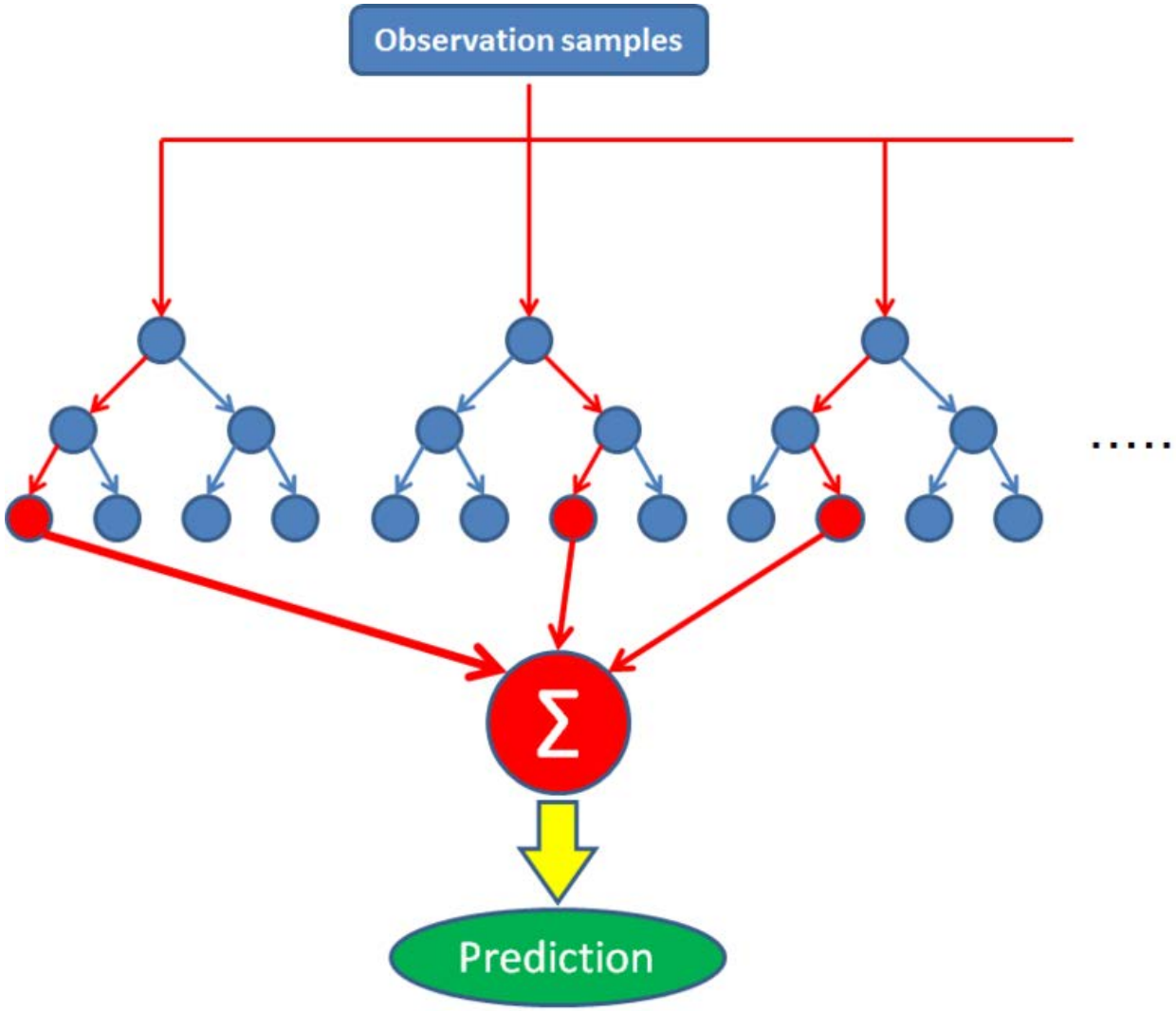


Image cdn-ak.f.st-hatena.com

Application 3: Disaster Management

Image abc.net.au

Why bushfires?

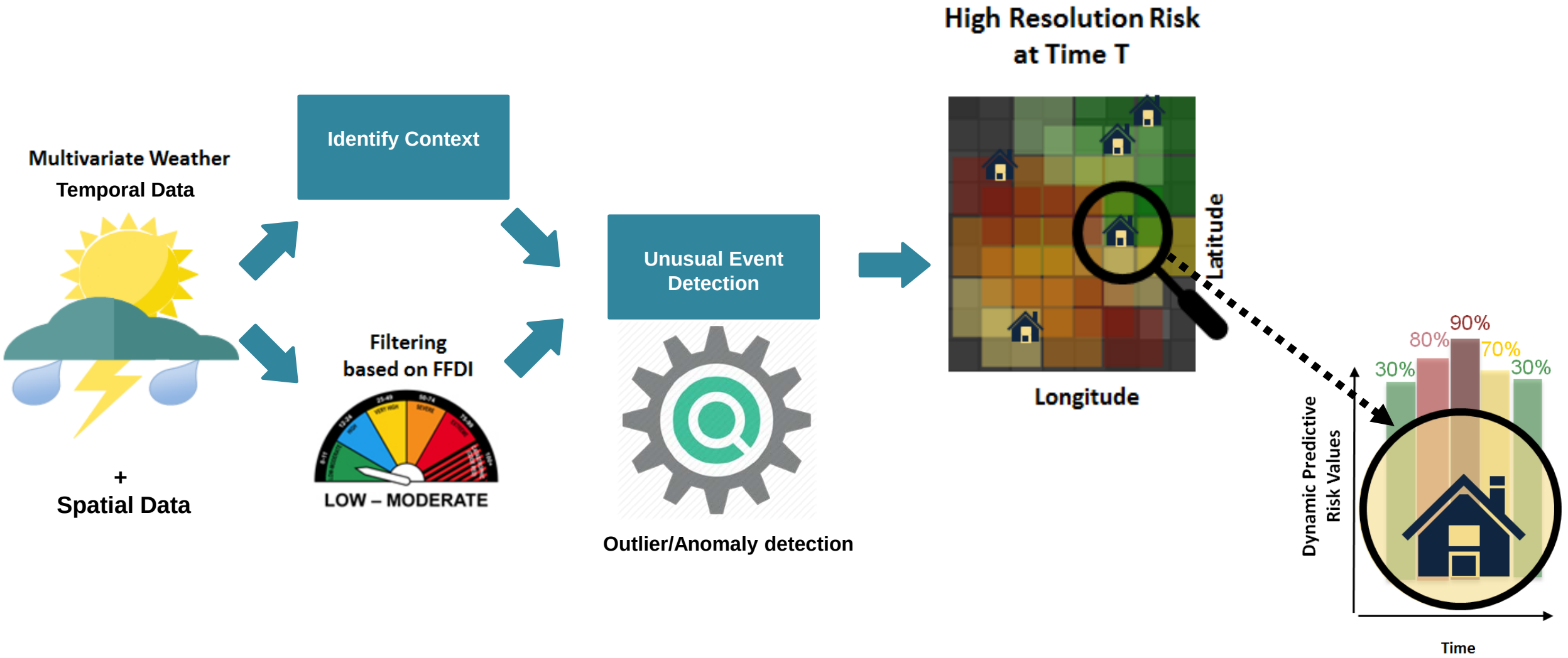
- Bushfires have shaped the Australian landscape for many years. With urban expansion and changing weather patterns the thread to homes and families is increasing.

- Australia experiences bushfires as the most damaging disasters.

- Examples:
  - Ash Wednesday bushfires 1983
  - Black Saturday 2009

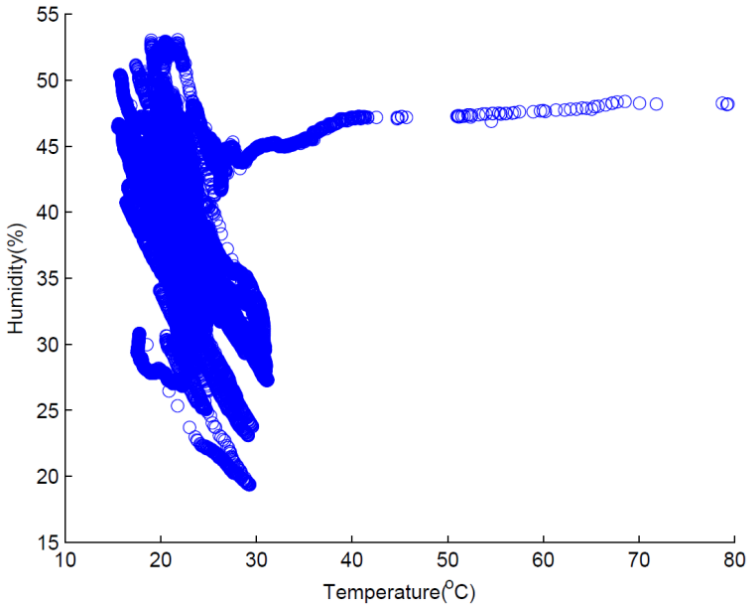
- Ability to ***predict*** the risk of bushfires is crucial in helping emergency services in their decision-making processes, to mitigate and reduce the impact of such events.

Dynamic bushfire risk prediction



Anomaly/Outlier Detection

- Anomaly: data points that are inconsistent to the normal data points
- Anomaly detection: the process of finding anomalous patterns in data sets
- In this application, anomalies are relevant to the episodes of time with high bushfire risk

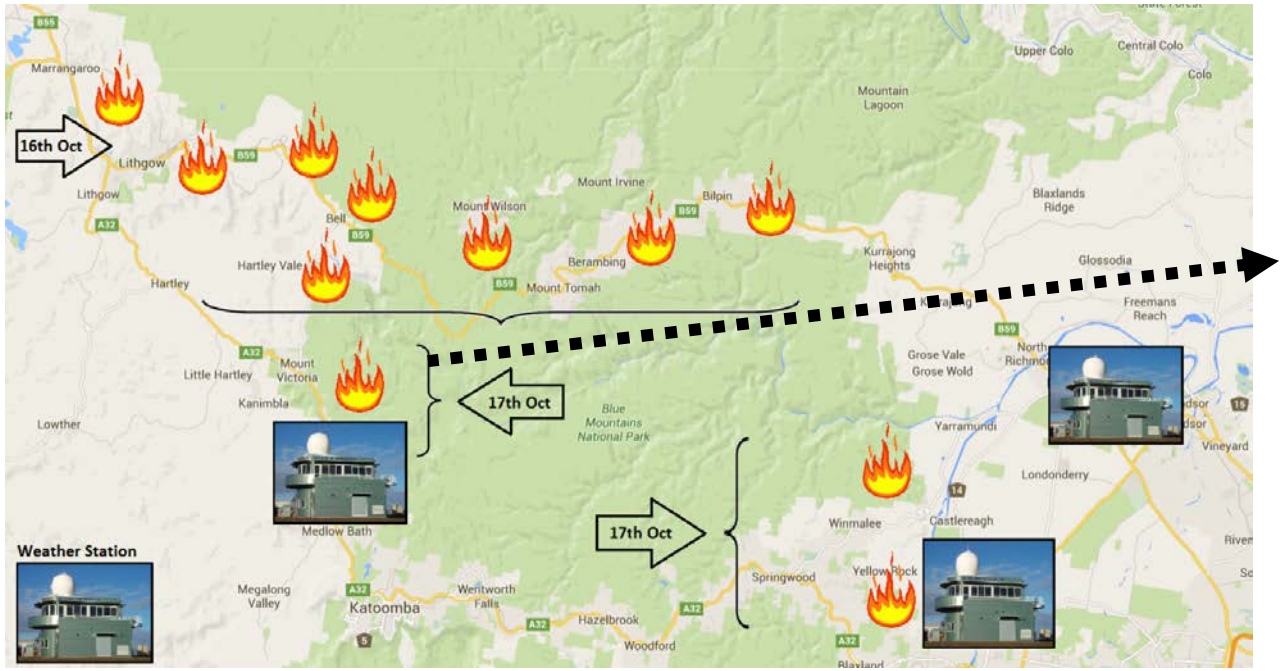


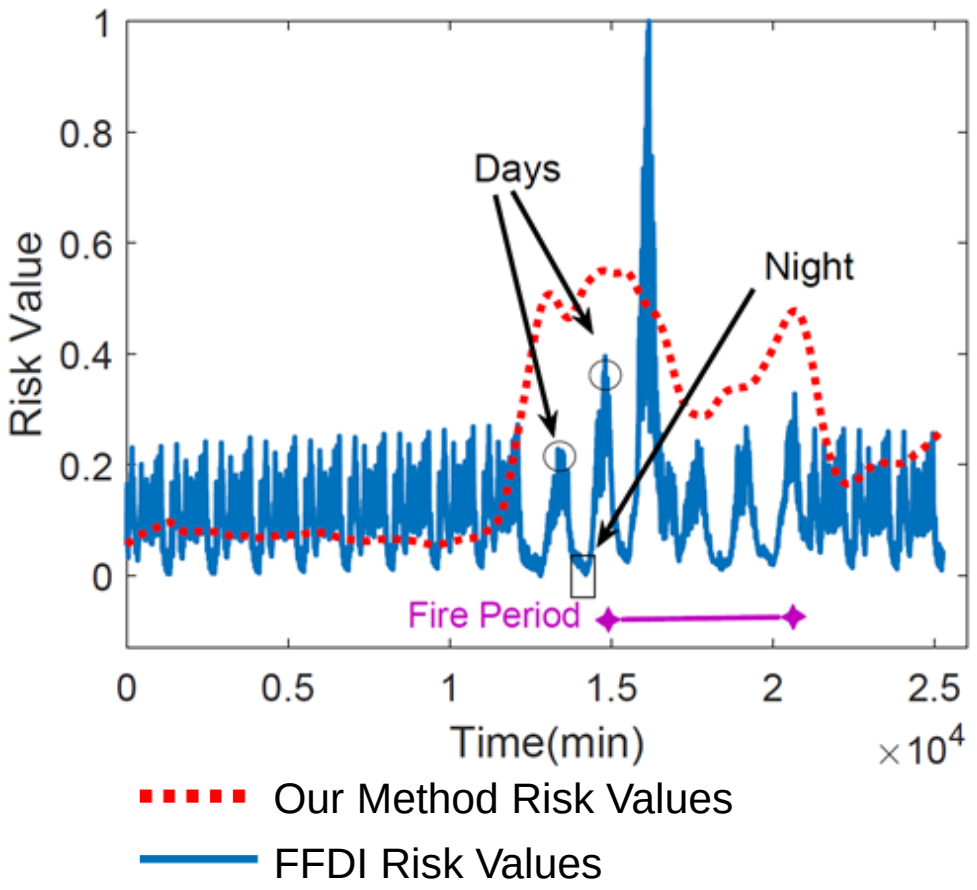
Case study- 2013 Blue Mountains Fire, NSW

- Why Blue Mountains?
  - 11 days of intense bushfires
  - 2 fatalities
  - 248 houses destroyed
  - 183.4M AUD in insurance claims

- Data Available
  - ~ 45 thousands houses
  - historical weather measurement

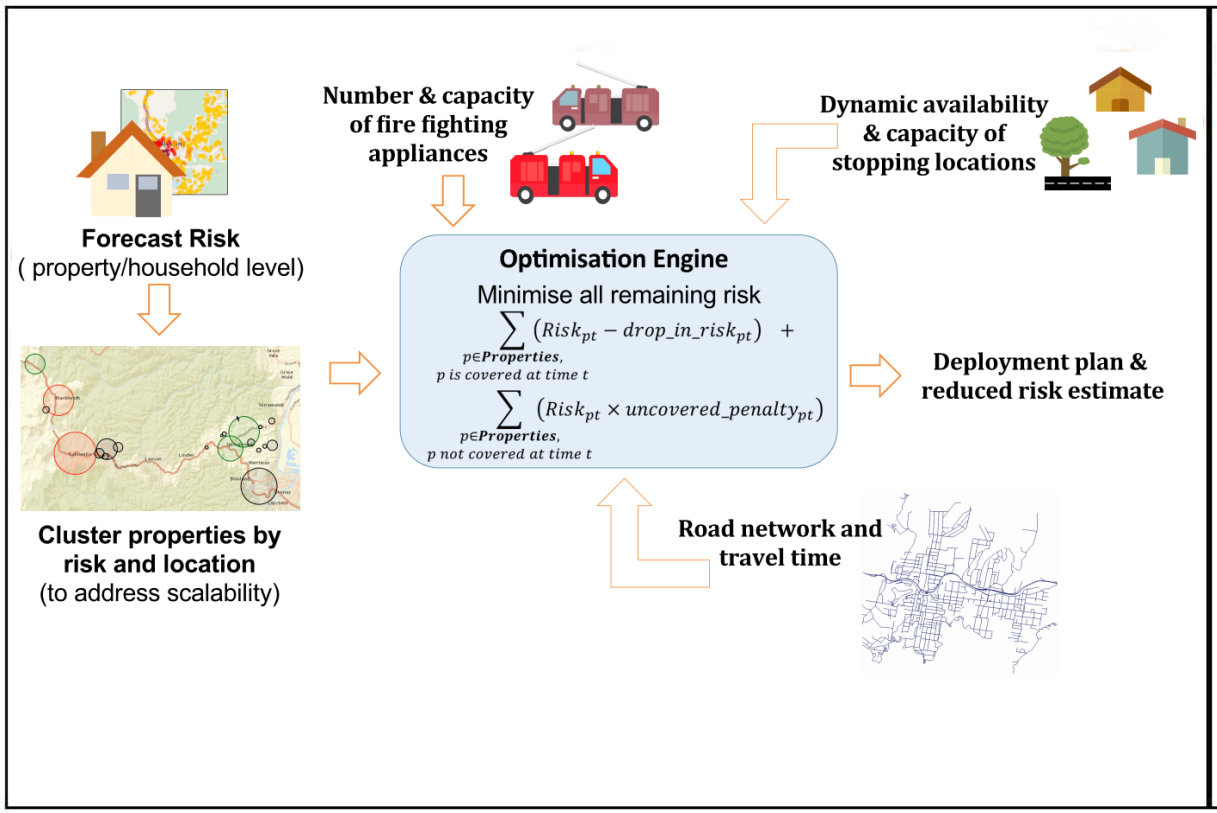
Case study- 2013 Blue Mountains Fire, NSW

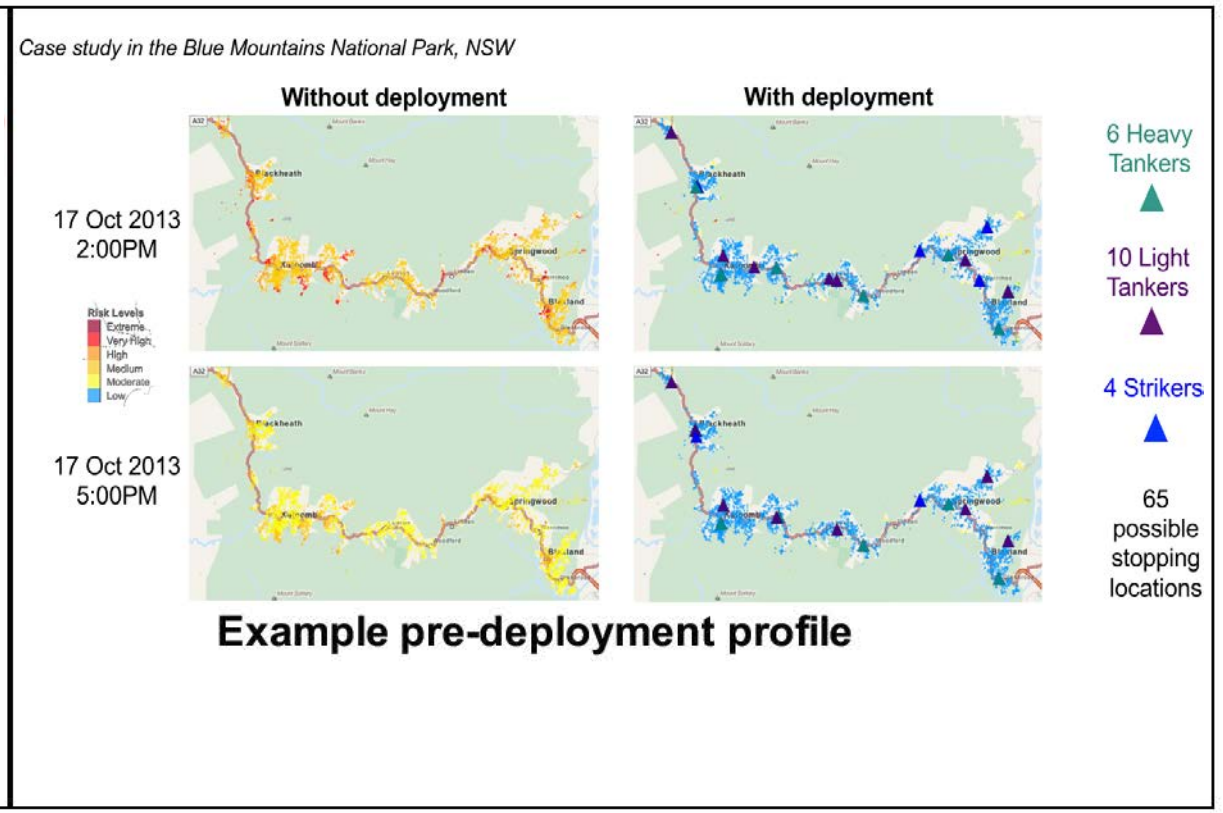




Fire Fighting Appliance Pre-Deployment

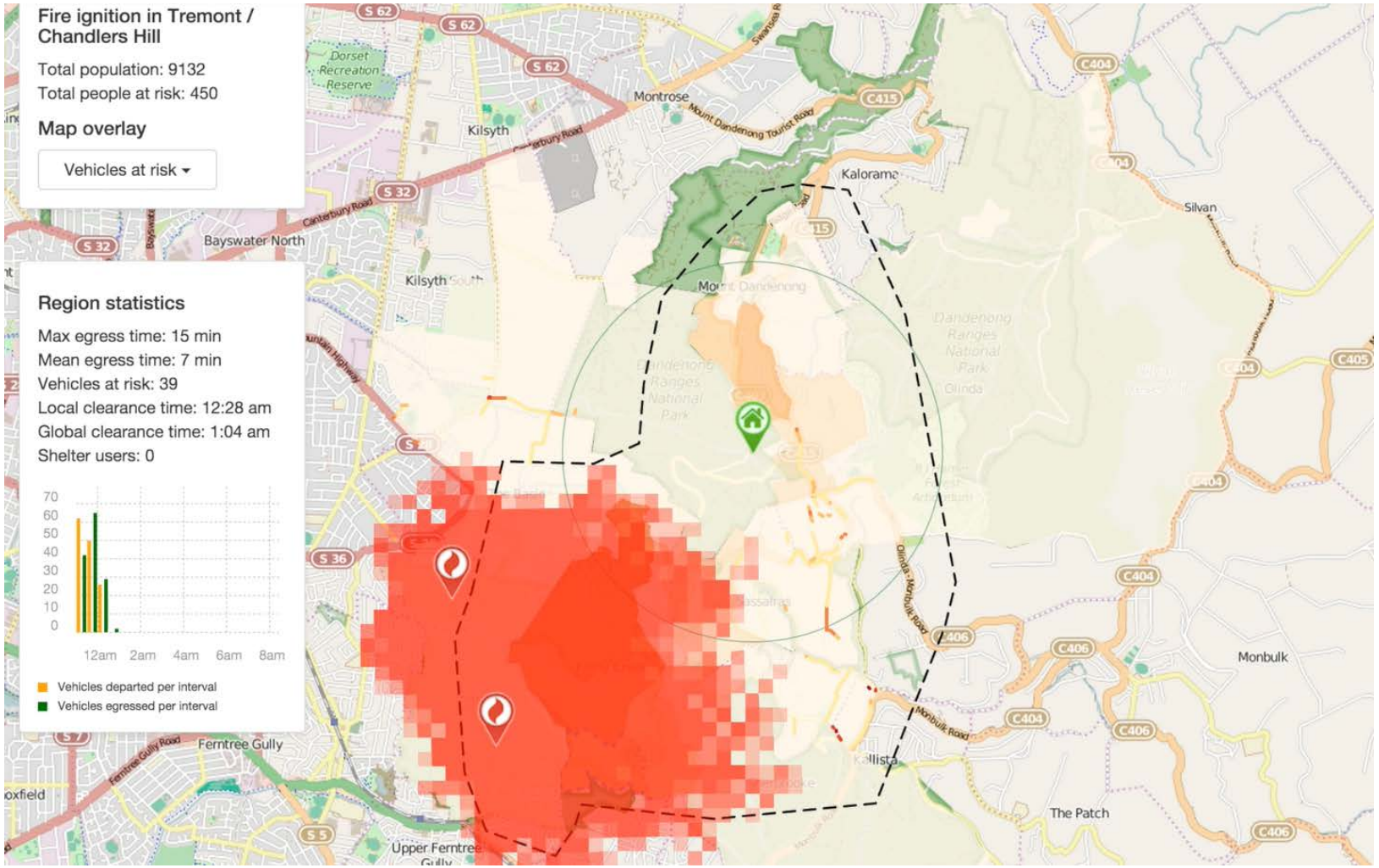
- Optimise locations of available fire fighting resources subject to forecast risk and roadside stopping locations with time dependent availability.



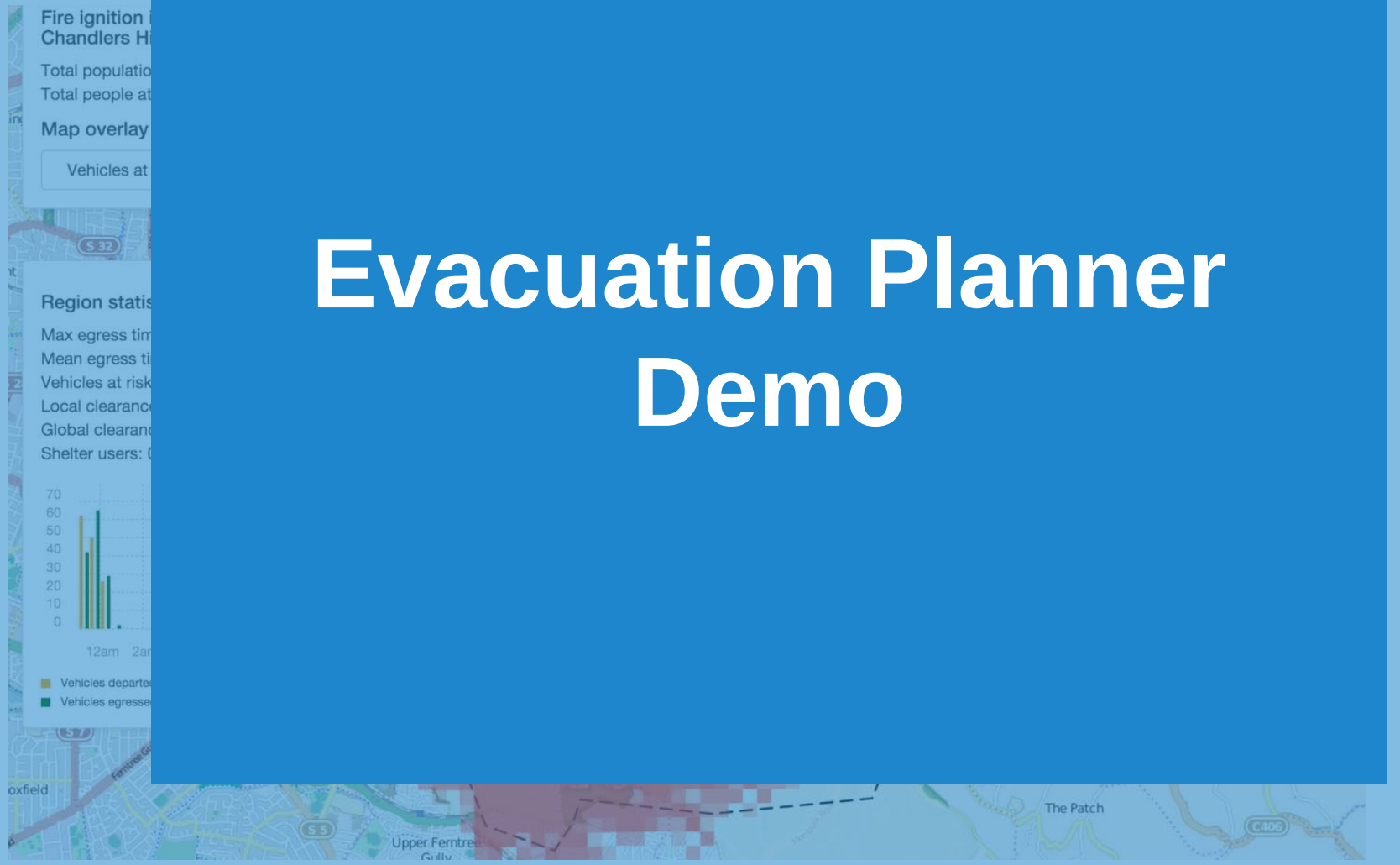


Evacuation Planner

- Spatio-Temporal fire expansion



Evacuation Planner



Evacuation Planner Demo

Summary

- Data types:
  - Spatial data
  - Temporal data
  - Spatio-temporal data
- Data mining techniques
  - Classification
  - Clustering
  - Anomaly/outlier detection
- Applications:
  - Diagnosis of pneumonia
  - Car racing driver distraction detection
  - Bushfire risk prediction, fire-fighting and evacuation planner

Thank you!