SMU
SINGAPORE MANAGEMENT UNIVERSITY

School of
Computing and
Information Systems

IS614 Group Project Proposal
Step Counter

Group 1

31 January 2023

Team Members:

- Clarence San (clarencesan.2021@smu.edu.sg) – Team member
- Spencer Keith Marley (skmarley.2021@smu.edu.sg) – Team member
- Jian Lorenzo Gopez Chan (jian.chan.2021@smu.edu.sg) – Team member
- Sin Hong Tang (shtang.2021@smu.edu.sg) – Team member
- Licheng Yan (licheng.yan.2021@smu.edu.sg) – Team member
- Songhan Wong (songhanwong.2020@smu.edu.sg) – Team member

Faculty Supervisor:

- MA Dong
- Pius LEE

1. Problem Statement

Insurance providers traditionally rely on independently conducted research to quantify mortality risk (Quah 2019). However, the growth in popularity of the wearables market in recent years has sparked significant interest in exploring how the data collected could be used to aid decision-making. At an addressable market of over 250 million yearly volume sales, and a projected compounded annual growth rate of more than 20%, modern wearables are able to track, collect, and transmit a variety of user data, such as daily steps activity, heart rate, and hours spent sleeping, at scale (Chui, Collins & Patel 2021; Musale et al. 2019, p. 37883; Ho, 2016). Of significant interest to health insurance providers are activity markers, which serve as an additional factor when assessing individual health risks and providing competitive risk underwriting (Quah 2019).

On the other hand, incorporating user activity data into pricing insurance premiums results in new, unique challenges. One significant source of premium leakage occurs when consumers artificially inflate their activity to bring down premiums. Such opportunistic, dishonest activity can occur in the form of shaking the device, or using a synthetic step generator. As ensuring the integrity of the collected data is critical to sustain competitive and fair premiums, it is imperative for manufacturers, data and underwriting professionals to collaborate and prevent the proliferation of misuse and abuse by errant users (Quah 2019).

Substantial research has been undertaken to accurately identify users' step activity data for location tracking. For instance, several scholars have proposed (Zeng et al. 2019; Muhsen, Al-Amaydeh & Hamlan 2020) robust step-counting algorithms that rely on a smartphone's built-in accelerometer and gravity sensor. To improve the accuracy of location tracking when indoors, techniques such as Dead Reckoning can also be integrated. Additionally, approaches such as utilising different daily activities, choosing appropriate segregation methods, and fixing thresholds for every experiment have also been suggested to improve algorithm robustness (Muhsen, Al-Amaydeh & Hamlan 2020). Individual gait detection has been explored as a potential use case in biometric authentication (Muaaz & Mayrhofer 2017). For example, Musale et al. (2019, p. 37883) found that spatial data collected through embedded sensors could be aggregated into measures, such as mean, standard deviation, skewness, and correlation, to identify unique users (Musale et al. 2019, p. 37889). Several novel tweaks, such as the use of a pedometer sensor to detect user movement, has also been suggested as a means to minimise unnecessary data collection (Musale et al. 2019, p. 37885). Please refer to Appendix 6.1 for visual examples of collected data from related literature.

Deep learning-based approaches have demonstrated their effectiveness as a robust framework to track human movement (Santos et al. 2023, p. 1). In more recent years, the introduction of Capsule Networks and Generative Adversarial Networks within the field of Computer Vision has attracted attention due to their potential to tackle viewpoint variance and their ability to resist network attacks (Santos et al. 2023, p. 15). However, Computer Vision based solutions require a significantly high investment in infrastructure and hardware, and suffer from unique limitations such as occlusion and higher data requirements (Santos et al. 2023, p. 15). Furthermore, although Deep learning-based approaches might be able to capture more granularity, previous research indicates such granularity is unnecessary. In contrast, wearable sensors will be able to extract the dynamic factors of walking patterns via simpler signals without sacrificing real-time detection (Marsico & Mecca 2019). Critically, IoT sensors and devices are also cost-effective and simple to procure.

We anticipate several challenges with the implementation of our minimum viable product. Firstly, the usage of such multiple connectives on wearable devices can expose a variety of personal information and further increase the risk of security breaches, which necessitates robust security measures (Musale et al. 2019, p. 37883). Secondly, the use of sensors requires the devices to be relatively unobstructive to the user, yet hardy to moisture, humidity, and heat. Finally, the data collection process will require the data to be stored on the device before transmission. Hence, effective data sampling and collection methods will need to be implemented to optimise device storage.

2. Proposed IoT Solution

1. a) Proposed Architecture

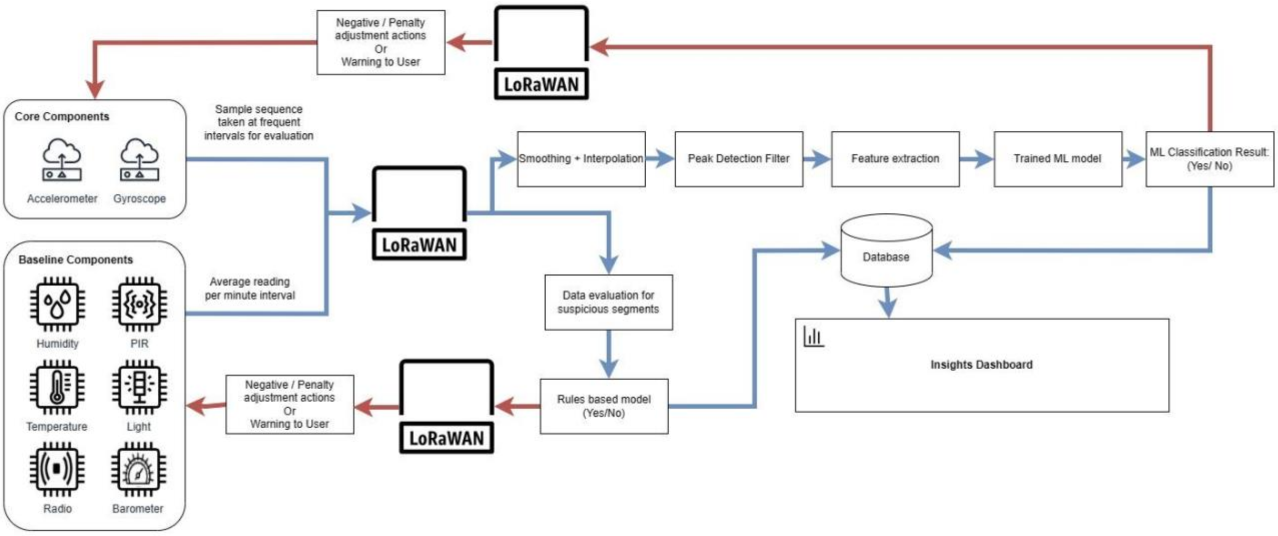


Diagram 1: Proposed IoT architecture

Our proposed architecture incorporates an IoT device designed to detect dishonest movements on a step counter via machine learning inference on the data collected from the device. By leveraging the radio communication feature of micro:bit, we aim to establish a proximity validation mechanism to improve the accuracy of the recorded counts. Please refer to Appendix 6.2 for the group's proposed mockup for the output device.

2. b) Description of architecture

Raw data is collected via both the onboard sensors of micro:bit, as well as the attached sensors that are used to eliminate baseline attempts to apply dishonest activity. After simple processing onboard, raw values are transformed into numerical readings ready for analysis. Due to the limited flash memory of 128kb, the data is uploaded to the cloud during collection. To optimise this process, our team will consider simple preprocessing techniques, such as simple aggregation or reporting a frequency value to minimise data traffic.. A more detailed set of descriptions for these relevant sensors can also be found below (see Appendix 6.3).

The ideal implementation relies on the LoRaWAN network, a low-cost, low-power consumption protocol, to transmit data across longer distances (Tomás 2017). As a working proof of concept, our IoT product relies on the transmitter-receiver model via two micro:bits and a laptop as an intermediary. Once the collected readings are transmitted to the cloud, it will then be fed through a trained Machine Learning model to derive insights. Some initial directions to explore include classifying if the acceleration and angular rotation sequence are indicative of actual human movement, or identifying irregularities in the environment, such as a stagnant room temperature or atmospheric pressure. These insights can be used to make strategic business decisions. Please see Appendix 6.4 for a sample schema for the data to be collected.

To maintain the robustness and reliability of the step counter, specific actions, such as adjusting the recorded number of steps on the step counter, or suspending its usage for a penalty period, can be triggered remotely. A closed loop feedback solution will be implemented to return business decisions back to the step counter via the radio transmitter function.

For further experimentation, our team will also consider edge deployment of the machine learning model on micro:bit itself. This is made possible by specialised ML models that are compact enough to fit onto the micro:bit while still able to perform inference on the go. One such solution is the edge impulse platform which has a working example with micro:bit that our team hopes to explore more on. Appendix 6.5 illustrates a proposed architecture for this approach.

3. c) Justification for sensors and protocols

The following table details the sensors and connectivity components required to implement our proposed IoT architecture. The proposed list of equipment to procure was selected using the following framework:

i. What contextual information is required for our IoT use case?
ii. What sensor modalities can be used to obtain the identified contextual information?

iii. How should the proposed IoT devices be deployed?
iv. Are the sensor specifications of the selected devices sufficient to meet application needs?
v. Are sensor readings periodic or event-driven?
vi. Can the sensor provide valid measurements in the target operating environment?

Sensor	**Motivation**	**Reason for use**
Accelerometer	This device measures the magnitude of acceleration in the X-Y-Z axis. Such information is critical in evaluating if a person is walking or not because changes in acceleration magnitude is indicative of a person swinging the arms or leg which is typical in a walking movement. Irregular and sudden spikes in readings can indicate dishonest intent by the user.	Critical data to analyse user movement.
Gyroscope (Joy-it MPU6050) - **I2C required**	Because accelerometers only measure linear acceleration, dishonest activity cannot be detected if a perpetrator designs a simple shaking device which operates in boundaries accepted within the verification threshold. To enhance the robustness of our step counter, we will take into consideration the angular acceleration a human creates while walking. Thus we now have 6DOF that can be used to verify valid walking patterns.	Critical data to supplement the accelerometer readings. **<u>Not included on micro:bit.</u>**
Light Sensor	Light sensor indicates the raw magnitude of the brightness in the surrounding area. It can be an excellent indication of changes in environment conditions when a person moves from one place to another. This provides a new dimension for our team to evaluate dishonest behaviour.	Included on micro:bit, easy to deploy with minimal effort. The readings are highly responsive.
Temperature and humidity (DHT-11)	This sensor measures ambient temperature and the humidity of the surface / surrounding area. These values can indicate to us a change in environment which is a typical behaviour of movement from one location to another. A constant temperature over a long period of time with large increments in steps can indicate dishonest behaviour.	Easy to deploy with minimal effort **<u>Not included on micro:bit.</u>**
Barometer (GY-63 MS5611) - I2C required	A barometer provides air pressure readings. We can infer the changes in altitude due to movement from one place to another.	**<u>Not included on micro:bit.</u>**
Ultrasound (HC-SR04)	The ultrasound sensor detects distance to the nearest object. Changes in the value can represent movement of the device past walls or around obstacles. In instances of cheating, ultrasound will return stagnant value.	Easy to deploy with minimal effort. **<u>Not included on micro:bit.</u>**
PIR (HC-SR501)	Passive infrared sensors detect IR radiation emitted ideally by living bodies. This is a binary sensor.. By attaching this sensor to our device, our team hopes to detect the presence of the body torso as the wrist swings past the body.	Easy to deploy with minimal effort. **<u>Not included on micro:bit.</u>**

Connectivity	**Motivation**	**Reason for use**
Radio on micro:bit	Radio communication is required to stream the data collected to the laptop intermediary. Moreover, the radio on micro:bit has a range of up to 70m. Further, micro:bit is able to output the strength of the signal which allows us to evaluate and validate the proximity of 2 devices that is critical to our team's design intention.	Included on micro:bit, there are detailed instructions online to set up beacons and communications.

4. d) Detailed sense making action i.e information we expect to mine from IoT data

Our team envisions sense-making via the following 3 key aspects:

i) Sense-making of motion
By leveraging ML, our team employs a classification model to qualify valid walking movements based on accelerometer and gyro readings by sampling on frequent intervals the readings streamed from the device. A possible approach is through peak detection in the waveforms of acceleration value; in this case, the number of peaks in the waveform represents number of steps taken (Zeng et al. 2019). This can then be extended to derive a time based count of negative classification over a period of time when steps are registered. The presence of a cluster of negative classification over a short period of time then signals that suspicious behaviour is indeed occurring. From this we will then perform the business decision by first issuing a warning followed by halting registration of steps on the device.

ii) Sense-making of the environment
Data such as light intensity, temperature, humidity, pressure will be collected on the go and analysed. Over a period where steps are registered, we can look for regions where there are little fluctuations in environmental changes, such areas can flag out potential dishonest activity. Moreover, it allows us to send feedback to the user to encourage them to adjust or change their walking habits. An example could be that a long walking session should be accompanied by variations in environmental data, otherwise it is an indication of cheating using a shaking mechanism.

iii) Sense-making of Interaction
Lastly, using the radio function of micro:bit, we intend to place multiple beacon micro:bits across several areas more than 70m radius. When our device gets in range of micro:bit, it will then measure the strength of the signal as raw data. For instances where the strength is above a certain threshold e.g. 80%, our device will receive a secret key from the beacon. Since a user is only able to receive a secret key when they are close to the beacon, this provides us an additional way to verify physical activity has taken place. Moreover, when analysing the exchange event, suspicious behaviour can come in the form of a user always exchanging data with the same beacon, which shows walking on the spot, a behaviour we discourage.

5. e) List of required hardware

Priority	Main PIC	Description / Justification	Qty	Source	Cost (SGD)
1	Wrist Strap	Gather data via wrist / ankle	1	Purchase link	16
1	Breakout Board	For access to all 16 I/O (elefreaks)	3	Purchase link	48
1	Gyro 3 Axis	Measures angular accel (JoyIT) - Overseas	3	Purchase link	25.5
2	Barometer	Capture altitude change	2	Purchase link	32
3	HR Sensor	Simple pulse sensor - Overseas	2	Purchase link	64
4	Tmp & Humid (DHT-11)	Catch changes in environment	2	Purchase link	3.5
4	Ultrasound (HC-SR04)	Indication of movement past objects	2	Purchase link	5.0
4	PIR	Detects IR from body part arms swing	2	Purchase link	2.36
				Total	196.36

3. Project Plan

For a more detailed task breakdown and timeline overview of the project, please refer to Appendix 6.9.

General Task	Main PIC
Procurement of hardware	Licheng Yan
Hardware assembly (MVP)	Spencer Keith Marley
IoT experimentation (connectivity, sense-making)	Spencer Keith Marley
User feedback / stress testing	Clarence San
Second product iteration	Sin Hong Tang
Feature extraction / preprocessing and ML	Jian Lorenzo Chan
Database and dashboard development	Wong Songhan
Medium article and presentation	All

4. Current Progress/Preliminary Work Done

6. a) Initial exploration with Edge impulse

We connected the micro:bit to our laptop, and coded a basic Python program to output the accelerometer data in dimensions x, y and z to the serial display. Then we captured some sample data by (1) making a walking movement with our hand whilst holding the micro:bit, and (2) making a shaking movement, in order to simulate real and fake walking accelerometer data.

We captured the data with 4 fields: (1) the timestamp, (2) the x dimension acceleration, (3) the y dimension acceleration, and (4) the z dimension acceleration. With this basic simulation of real and fake data, we can see differences in the data collected:

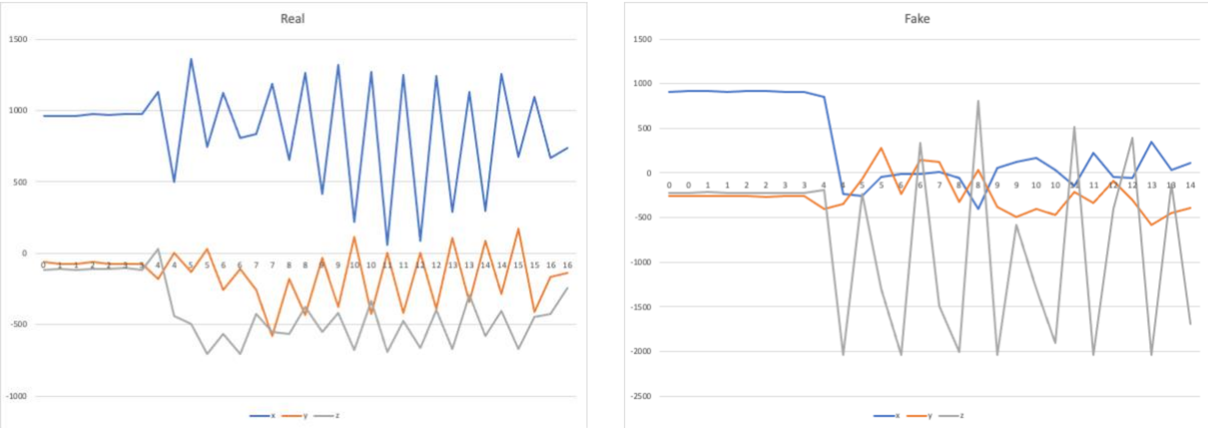


Diagram 2: Basic simulation of dishonest activity

Next we experimented with this data on EdgeImpulse.com, which is the leading development platform for machine learning on edge devices. We proceeded as follows: (1) uploaded the raw data captured by the micro:bit, (2) added a processing block for spectral analysis for analysing repetitive motions, such as data from accelerometers, since it extracts the frequency and power characteristics of a signal over time, (3) added a classification learning block that learns patterns from data and apply these to new data, to categorize movement and hence fit for our task, (4) defined the output features to be “fake” and “real”, (5) trained a basic neural network with 2 hidden layers, and 2 classes in the output layer, (6) collected new testing data, as described above for the data used for training, (7) tested the neural network and found that the real test data was classified correctly with 100% confidence, whilst the fake data was classified as correctly with 81% confidence.

The immediate next steps for us will be to: (1) collect more datasets, including different types of “fake” and “real” data with different people, (2) optimising the hyperparameters and the architecture of the neural network, (3) performing live inference to classify the data being captured, so that we can demonstrate to our peers.

Finally, to prepare for an initial prototype, our team conducted a comprehensive mapping of the user journey, incorporating the Data, Information, Knowledge, Wisdom (DIKW) pyramid as a framework to base sensor procurement decisions on. To further strengthen the comprehensiveness of the MVP and to ensure that most forms of dishonest activity are addressed, the team engaged in ideation exercises with the goal of identifying possible methods of cheating and determining the necessary features to mitigate such risks. The resulting frameworks will be subject to continuous updates based on the results of user testing and feedback. An in depth examination of the frameworks can be found in Appendix 6.6 to 6.8.

5. References

Abadleh, A., Al-Hawari, E., Alkafaween, E., & Al-Sawalqah, H. (2018). Step Detection Algorithm for Accurate Distance Estimation using Dynamic Step Length. arXiv. https://doi.org/10.48550/arXiv.1801.02336.

Chui, M., Mark Collins, M., & Patel, M. (2021). The Internet of Things: Catching up to an accelerating opportunity. *McKinsey and Co*. https://www.mckinsey.com/.

Ho, E. (2016). How Fitness Trackers and the Internet of Things are changing Health Insurance. AIBP. https://iotbusiness-platform.com/insights/fitness-trackers-internet-things-changing-health-insurance/.

Marsico, M., & Mecca, A. (2019). A survey on Gait Recognition via Wearable Sensors. *ACM Computing Surveys*, 52(4), 86:1-86:39. https://doi.org/10.1145/3340293.

Muaaz, M., & Mayrhofer, R. (2017). Smartphone-based gait recognition: From authentication to imitation. IEEE Trans. Mobile Computing, 16(11), 3209–3221.

Muhsen, H., Al-Amaydeh, O., & Al-Hamlan, R. (2020). Algorithm Design for Accurate Steps Counting based on Smartphone Sensors for Indoor Applications. *ASTES Journal*, 5(6), 811-816. Retrieved January 22, 2023, from https://www.astesj.com/v05/i06/p96/.

Musale, P., Baek, D., Werellagama, N., Woo, S., & Choi, B. (2019). You Walk, We Authenticate: Lightweight Seamless Authentication Based on Gait in Wearable IOT Systems. *IEEE Access*, 7, 37883–37895. https://doi.org/10.1109/ACCESS.2019.2906663.

Quah, J. (2023). The future is now: Wearables for insurance risk assessment. Munich Re. https://www.munichre.com/us-life/en/perspectives/wearables/wearables-the-future-is-now-wearables-for-insurance-risk-asses.html

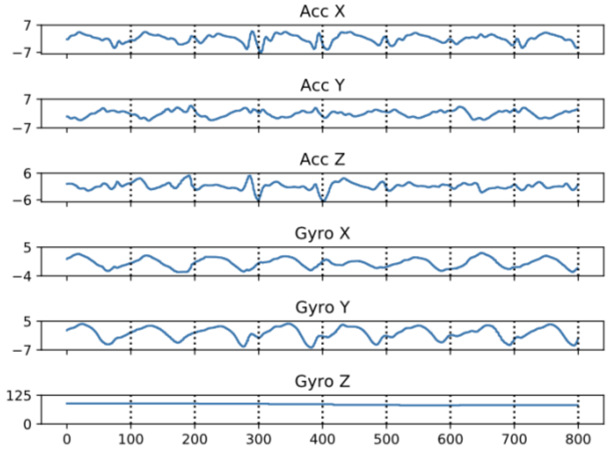
Santos, C., Oliveira, D., Passos, L., Pires, R., Santos, D., Valem, L., Moreira, T., Santana, M., Roder, M., Papa, J., & Colombo, D. (2023). Gait recognition based on Deep Learning: A survey. *ACM Computing Surveys*, 55(2), 1–34. https://doi.org/10.1145/3490235.

Tomás, J. P. (2017, June 12). *What is Lorawan and what are the main benefits of this technology?* RCR Wireless News. Retrieved January 29, 2023, from https://www.rcrwireless.com/20170612/internet-of-things/what-lowrawan-main-benefits-technology-tag23

Zeng, Q., Zhou, B., Jing, C., Kim, N., & Kim, Y. (2015). A Novel Step Counting Algorithm based on Acceleration and Gravity Sensors of a Smartphone. *International Journal of Smart Home*, 9(4), 211–224. https://doi.org/10.14257/ijsh.2015.9.4.22.

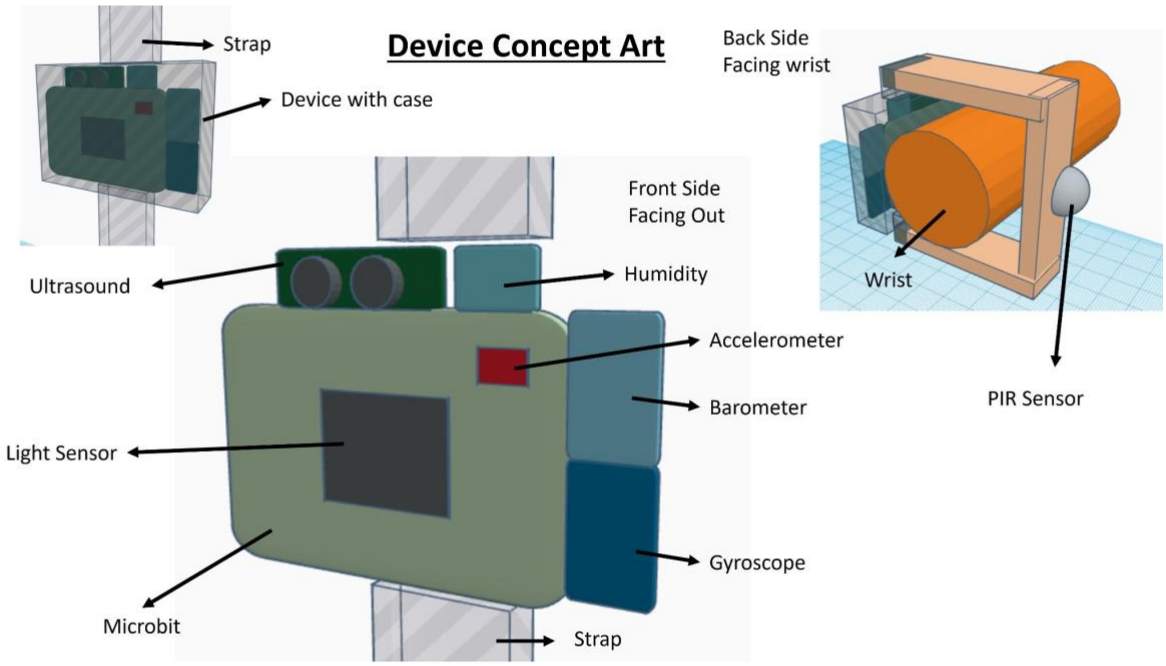
6. Appendix

7. 6.1 Example Accelerometer and Gyroscope Data



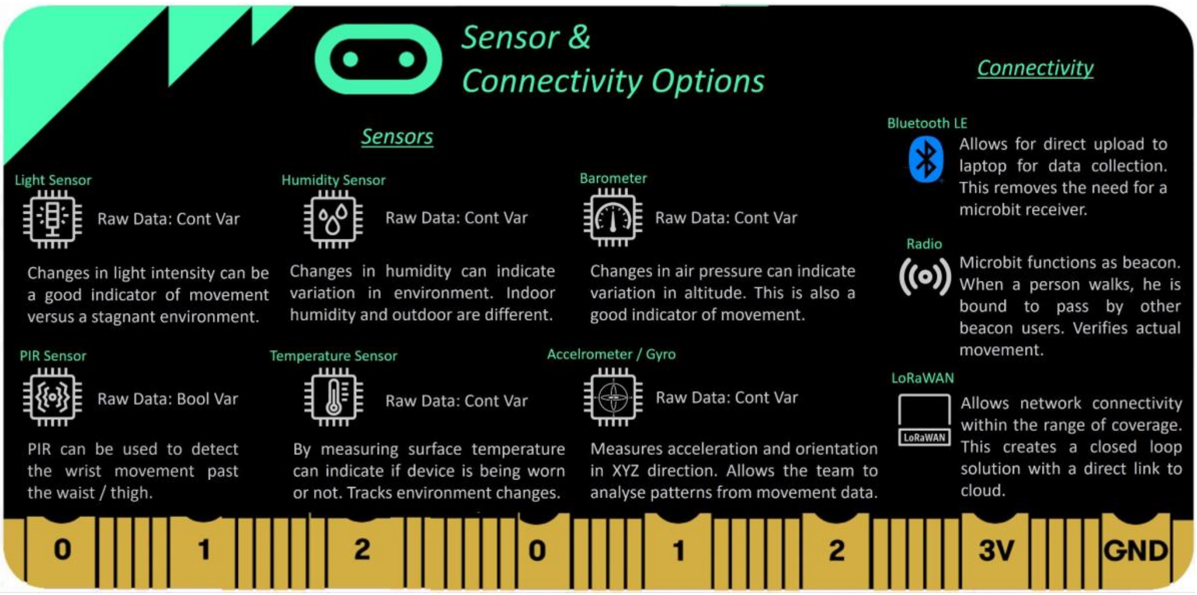
The above diagram shows visuals of collected data from Musale et al (2019).

8. 6.2 Device Concept Mockup



The above concept art showcases how the components are all connected together in the ideal case where our team managed to acquire all components.

9. 6.3 Diagram of Sensors and Connectivity



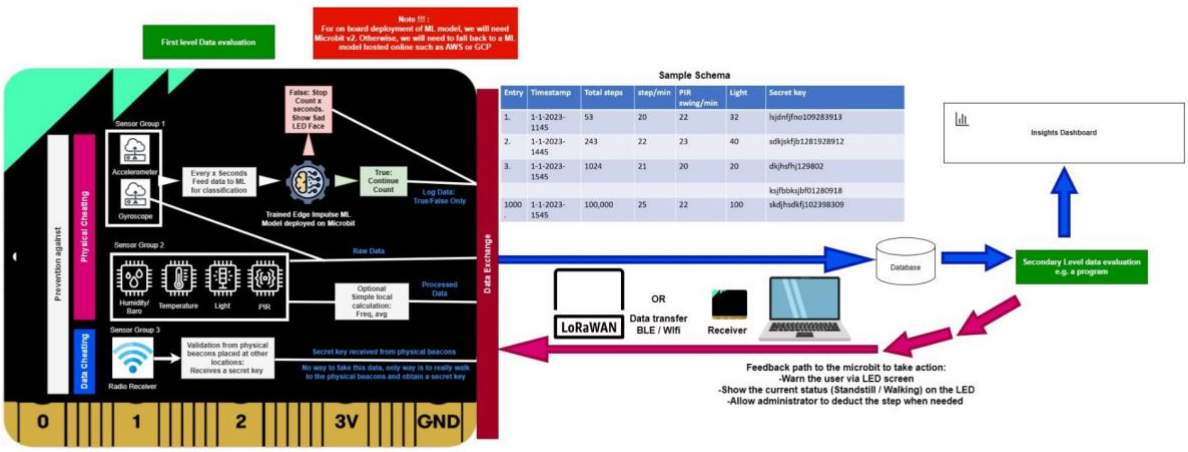
The diagram above describes the various sensors and connectivity options more in-depth.

10. 6.4 Sample Schema of Data

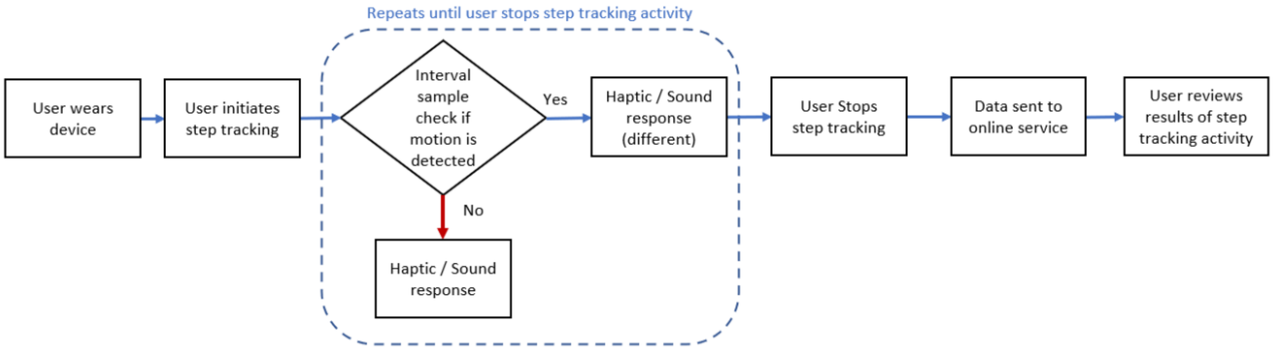
Entry	Time	Total Steps	Steps / Min	PIR swing / min	Light	XYZ - ACC	XYZ - Gyro	Secret Key Time
1.	1-1-2023-1145	53	20	22	32	12-55-12	1-5-0	Lsdnfjfno_0900
2.	1-1-2023-1445	243	22	23	40	12-22-33	1-3-0	-
3.	1-1-2023-1545	1024	21	20	20	45-99-84	9-2-1	Dkjhsfhj12_1045
…	…	…	…	…	…	…	…	…
1000.	1-1-2023-1545	100,000	25	22	100	12-22-33	1-3-0	Skdjhsdkfj1_1456

The above schema illustrates the structure of data that our team intends to collect.

11. 6.5 Architecture for Edge Deployed ML Model on micro:bit V2



12. 6.6 Proposed User Journey



13. 6.7 DIKW Framework

Wisdom

- Classification learning block in Edge impulse to learn patterns and apply to new data, this results in robust step counter
- If we can identify some personal information of user, we can perform demographics analysis for insurance company to recalibrate their incentive programs

Knowledge

- Extracted features to be classified by a trained basic neural network with 2 hidden layers and 2 classes in the output later ("Fake" and "Real")

Information

- Spectral analysis processing block to extract features (time-based count of negative classification) over a period of time where steps are registered
- Amplitude and frequency of the 3D accelerations and the patterns of steps and pauses

Data

- Acceleration in the X-Y-Z axis of user motion
- Rotational velocity in the X-Y-Z axis of user motion
- (additional complement) surface temperature, humidity, pressure and light intensity data of the user environment
- Radio beacon hashed secret key data to determine user location

14. 6.8 Possible Methods of Cheating

Methods of Cheating	Remarks
Attaching the step tracking to a dog / robot vacuum	
Swinging of user arms	
Shaking of user wrist, or using a step counter device	
Wearing another device to impersonate the owner and perform the activity	Project Limitation

15. 6.9 Detailed Task Breakdown and Timeline Overview

Timeline	Deliverable (subtask)	Priority
Week 4 (30 Jan - 5 Feb)	• Project proposal • Finalise IoT system design • Start procurement of hardware	MVP
Week 5 (6 Feb- 12 Feb)	• Arrival of hardware • Assembly of IoT hardware • Experimentation with sensors and connectivity	MVP

Week 6
(13 Feb- 19 Feb)
	• MVP for step counter
	    • Primary sensors ready
	    • Basic detection for steps
	• Preprocessing of signals
		MVP

Week 7 (19 Feb - 26 Feb)	• Begin ML development • Database & dashboard development	P0 (Target)

Week 8
(27 Feb - 5 Mar)
	• ML model deployed and able to make inference
	• End-to-end connectivity ready
	• Database & dashboard ready
	• Full product specs ready
	    • Sense-making and wearable
	    • Connectivity
	    • Data analytics and insights
	    • Fraud detection
		P0 (Target)

Week 9 (6 Mar- 12 Mar)	• Product stress test & user testing • Edge ML development • LoRaWan	P1 (Stretch)

Week 10
(13 Mar - 19 Mar)
	• Edge ML deployed
	• Preparation for final presentation
	    • Demo walkthrough & practice
	    • Slides
		P1 (Stretch)

Week 11 (20 Mar - 26 Mar)	• Final presentation • Preparation for medium article	P0 (Target)
Week 12 (26 Mar- 19 Mar)	• Submission of medium article • Account creation for Medium	P0 (Target)