# Light Based Activity Recognition Using Realistic Data

Jasper Vos
Supervisor(s): M.A. Zuñiga Zamalloa, Miguel Chavez Tapia
EEMCS, Delft University of Technology, The Netherlands

June 19, 2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

**Abstract**

In the field of Visible Light Sensing, light sensors are used to extract information from objects which do not actively communicate any information. Previous research within this field proposed the system called SolAR, and proved the possibility of using a solar cell as both a power source and an activity sensor. A wrist mounted solar cell generates more energy than it uses during operation, while achieving a high classification accuracy for different activities. While the wearer performs different activities, the power output of the solar cell fluctuates. In turn, these fluctuations are used to recognise activities. To extend on the concept of SolAR, this paper introduces a prototype to obtain data from different activities while performing day-to-day tasks. During these activities, ordinary actions are performed to emulate natural circumstances. Analysis of this data initially shows no significant drop in accuracy when compared to SolAR. Further examination shows significant differences in mislabelling rates when comparing to the results of SolAR.

# 1 Introduction

Within the field of Visible Light Communication (VLC), the topic of Visible Light Sensing (VLS) has been advancing over the last years. While VLC relies on the communication between two or more active parties, VLS gathers information from passive objects. Any object can become a passive transmitter by either moving past, or interacting with, a light sensor. This type of sensor observes changes in light intensity or colour. These changes in light can then be used to extract information about the passing object.

Examples of VLS include the recognition of differences in light levels to identify passing vehicles or objects [Wang et al., 2016], by using black and white stripes to form a barcode-like identifier. Customers can be tracked using a unique customer profile, based on the colour of their clothing, without the use of cameras [Zhang et al., 2020], which reduces the risk of privacy infringement. Gestures on smart wearables can be sensed by recognising light blockage patterns, while harvesting power with the same sensors, resulting in self powered gesture recognition [Li et al., 2018]. These examples emphasise the reduced power draw from VLS systems, together with the reduction of privacy concerns, as no cameras are used. Depending on the situation, different approaches regarding sensor placement and sensor types have been used: multiple photodiodes placed on the floor of a room [Li et al., 2016], a single colour sensor placed on the ceiling of a store [Zhang et al., 2020] and a single solar panel placed on a table [Ma et al., 2019].

In a recent study under the name of SolAR [Sandhu et al., 2021], Human Activity Recognition (HAR) has been achieved through the use of a solar panel, worn on the wrist. Conventional HAR systems mostly use accelerometers, sometimes accompanied by other sensors [Lara and Labrador, 2013]. Generally, accelerometers are cheap and they are accurate when recognising activity. However, accelerometers do not have the ability to regenerate energy. By using a solar panel, combined with a machine learning classification algorithm, SolAR proved the possibility of recognising activity while generating more power than the system consumes. To achieve this form of HAR, multiple subjects were asked to perform five different activities while wearing the SolAR system: sitting, standing, walking, running and traversing stairs. Each subject performed the activities over a time span of 3 minutes per activity, both indoors and outdoors. Resulting data is cut into sample windows, from which the most defining features were extracted to represent each window. These windows were then used to train a Machine Learning algorithm to classify the different activities performed by the subjects, resulting in a classification accuracy reaching over 93%.

This paper describes the research that tries to replicate the results of SolAR, with more representable data, gathered from five different activities performed throughout the day. This data consists of the same five predefined activities as in SolAR, but in different lighting conditions, over a longer activity time. During with the activities, ordinary actions are performed during the data gathering process. Like SolAR, this data is then be used to train a Machine Learning algorithm to classify this data as a specific activity.

The general structure of the experiments are explained in section 2. The addition of realistic activity data is discussed in section 3. Section 4 defines the inner details regarding each step. Results are reported in Section 5, a comparison between algorithms is made and differences are reasoned on. Aspects on ethics and reproducibility are explored in section 6, followed by a comparison of the results against the results of SolAR [Sandhu et al., 2021] in section 7. This paper ends with the conclusion and recommendations for future research in section 8.

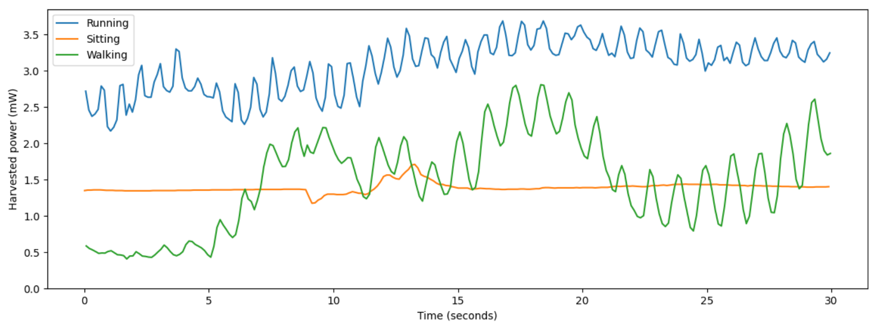



Figure 1: Example of harvested power sample while performing three different activities: running, sitting and walking.

## 2 Methodology

This section explains the structure of the activity recognition system, further details on this implementation are explained in section 4.

When a certain activity is performed, like running, walking or sitting, minor changes in light intensity can be measured using a light sensor worn on the wrist. In this case, a solar panel acts as the sensor, which generates power based on light intensity. Therefore, changes in power output reflect changes in light intensity. The power output of a solar cell during these three activities can be seen in figure 1.

In order to obtain data from activities, a prototype in the form of a small, watch-like, wearable solar panel is used, as can be seen in figure 2. This wearable measures the solar panel output. This output is labelled based on the activity performed.

Ideally, like the system of SolAR [Sandhu et al., 2021], the prototype would be powered by the solar cell in combination with an internal battery. However, to reduce complexity, size and safety of the system, the prototype only measures the output power of the solar cell, while being powered by a power bank. The concept of measuring and harvesting solar energy output has been explored and proven in the research of SolAR.

To measure the output power of the solar panel, a constant load is placed in series with the cell. This way, the output voltage of the cell reflects the output power of the cell, which can be measured easily with an Analogue-to-Digital Converter.

To efficiently classify activities, data is split into short measurement windows. Features which are inherit to the data are extracted to reduce each window from a time-based dataset to a set of defining features. For example, in figure 1 we can observe different oscillations during the activities. A rough estimation on which activity is presented can be made by observing the frequency of the of the corresponding data.

For classification, several Machine Learning algorithms are implemented and compared. The labelled data is combined and split into a train and test set. The train set is used to train multiple Machine Learning algorithms for activity recognition, based on the presented features. After training, the test set is used to evaluate the performance of the classification algorithms against unknown data.

# 3 Introducing Realistic Activity Data

In this section, we analyse the works of SolAR [Sandhu et al., 2021], and present the extension on the concept of SolAR by introducing real-world data.

SolAR has proven the concept of achieving accurate HAR using a solar cell, while generating more power than the system uses. However, data gathered for SolAR was obtained in a controlled environment. Each activity is performed in a specific location, like designated room. In a controlled environment, conditions occurring naturally could be emphasised or suppressed. For example, a designated room can have a unchanged light level, putting an emphasis on this specific light level. By introducing a field study using a prototype similar to SolAR, we can investigate a case close to the true potential for solar cell based HAR.

This paper aims to introduce data from routine activities, under ordinary conditions. Activities are performed indoors and outdoors, on sunny and cloudy days and different times of the day to obtain varying lighting conditions. During these activities, ordinary actions will be performed, like writing, flipping a page or motioning during a conversation.

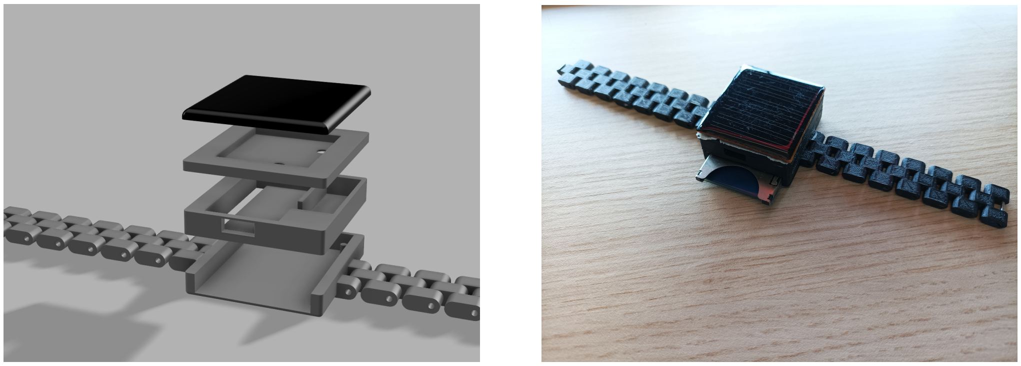
# 4 Experimental Setup

Firstly, this section explains the prototype created for data collection in section 4.1, followed by the means and circumstances of the data collection which are explained in section 4.2. Finally, the methods for data processing are explained in section 4.3.

## 4.1 Prototype

The output power of the solar panel changes based on the light intensity of the light received by the panel. The output power is defined in watts, which itself cannot be measured and is calculated as the product of potential and current ($P = V * I$). Normally, one requires the measurement of these factors. By using a known resistance in series with the solar cell, we can calculate the power output by solely measuring the potential difference of the cell output ($P = \frac{V^2}{R}$).

Accordingly, the microcontroller chosen for this prototype is the Arduino Pro Micro. The small size (35mm x 18mm) of the Pro Micro enables the small area of the prototype,



(a) Digital prototype

(b) Physical prototype, including electronics

Figure 2: Digital and physical prototype, used for collecting and storing the solar panel output data.

giving it the watch-like appearance like in figure 2. The use of an Arduino microcontroller enables one to program the microcontroller with a comprehensive program, interfacing with Arduino compatible modules.

To measure potential difference of the solar cell output, the ADS1115 Analogue-to-Digital Converter (ADC) [Texas Instruments, 2009] is used. Instead of selecting the built-in Arduino ADC, the ADS1115 has been selected for its greater measuring resolution and on board voltage reference, providing precise measurements unaffected by external factors. To obtain the most accurate voltage data from the ADC, an 8 Hz sample speed is used.

In order to save the data output of the Arduino, a generic SD-card reader & writer module is used. This enables the microcontroller to log the data output without the need for an external server, while making the data obtained easily accessible for an external computer.

The combination of these components enables a stand-alone prototype, requiring only a power source for operation. The USB interface of the Arduino can be used to power the prototype, using a power bank which enables mobility when wearing the prototype. The complete wiring diagram for the prototype can be found in figure 3.

To reduce the height of the prototype, the headers of all modules and the microcontroller are desoldered, so modules have wires directly soldered between each other. This enables for vertical stacking of the modules, microcontroller and solar cell while maintaining a relatively small height.

These components together are placed in a three layer container. To prevent unwanted shorting between components, each layer has a thin non-conductive plastic separator. The container design can be seen in figure 2a and is 3D-printed using PLA plastic. As can been seen in the design, the prototype has been designed with a wristband for wrist mounting. The final prototype can be seen in figure 2b.

## 4.2 Data Collection

To collect data, five activities are performed under different, uncontrolled lighting conditions. This results in fuzzy, but realistic data when compared to a lab environment. Like SolAR [Sandhu et al., 2021], the five day-to-day activities are performed during data collection: sitting, standing, walking, running and traversing stairs. Details on different activity

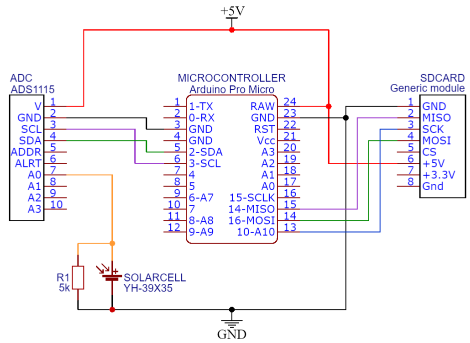



Figure 3: Electrical design of the prototype

conditions can be found in appendix A. Additionally, while performing these activities, ordinary actions are performed. These ordinary actions can be found in appendix B. Although these actions form the majority of actions performed while collecting data. Rarely performed actions, like picking up a dropped pen, are not included from this table.

Most activities are performed under sunlight, at different times during the day in variable weather conditions, resulting in variable light intensity. Another factor in changing light intensity is the surrounding of streets in and near the city of Delft, with apartments, housing and trees significantly influencing the light intensity.

The watch like appearance and feeling of the prototype enables prolonged wearing without being bothersome. It should enable the wearer to perform daily tasks normally and reduce attention given to the wearable, resulting in usage close to a real-life scenario.

The first 200 seconds of each data collection session are disregarded, to account for a short setup time needed to mount the prototype together with the power bank.

For power, an elastane sleeve is used to hold a lightweight power bank on the lower arm. Note that not every power bank is usable for this purpose, as the current draw of the prototype might not exceed the minimum required current draw.

## 4.3 Data Processing

In this subsection, we first start by explaining feature extraction, which is performed to make the data obtained during collection usable for a classification algorithm. After feature extraction, an explanation of the training and testing of multiple classification algorithms is presented.

### 4.3.1 Feature Extraction

The data obtained in section 4.2 presents itself as time series data: at discrete points in time, one single value exists. To use a classification algorithm, the raw data needs to be transformed in such a manner that the new data reflects important aspects of the original data.

To extract data from a meaningful time frame, the complete time series are cut into windows of 8 seconds, each overlapping 50% with the previous window. Each window is reduced to its 21 most descriptive features, previously identified in SolAR [Sandhu et al., 2021] for outdoor HAR. As these features are originally selected for outdoor HAR, where

| Features |
| --- |
| Peak-to-peak value, Coefficient-of-variation, Absolute area, Max. distance between peaks, 1st Quartile, 2nd Quartile, Frequency domain entropy, Median, Spectral peak, Min. value, Mean distance between peaks, Range, Max. value, Min. peaks, Standard deviation, Median absolute deviation, Frequency domain energy, Mean, 3rd Quartile, Max. peak, Autocorrelation. |

Table 1: Selected features, used for feature extraction on the 8 second windows cut from the data.

lighting conditions are naturally varying, they should perform well under different lighting conditions of both indoor and outdoor activities. The selected features can be found in table 1, and are explained in detail in the works of HARKE [Khalifa et al., 2018].

All features are extracted using a combination of functions provided by the Python packages NumPy, SciPy and Pandas [Harris et al., 2020, Virtanen et al., 2020, McKinney et al., 2010].

#### 4.3.2 Classification

Using the data obtained from feature extraction, several Machine Learning (ML) classification algorithms are applied and compared. Although data is collected on activity both indoors and outdoors, no distinction on indoor and outdoor activities is made for the classification process.

To create a reliable classification algorithm, data from different activities is combined into a single dataset. The label of the activities is removed and stored in a separate list. The dataset and list of labels is split into two: a training set and test set.

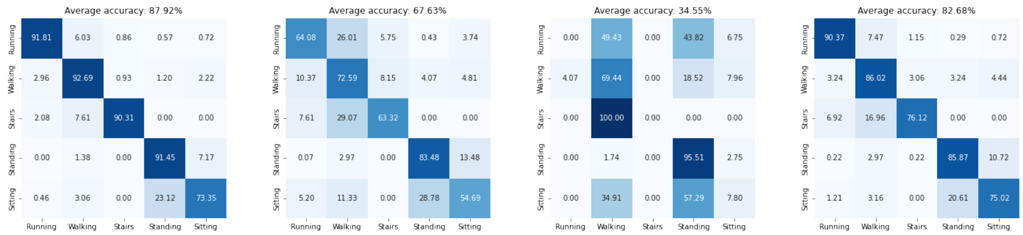
The training set is used to train the ML model to generate reasonable predictions, without interacting with the data selected for the test set.

The model is used to predict the entries from the test set after training. A final evaluation of the model can be made by comparing these predictions against the actual activities of the test set. To compensate for the effect of train test leakage (neighbours of testing samples are too similar and appear in the training set), cross validation is performed by leaving out sequential windows of 6 minutes. Using these sequential windows, almost all time-wise neighbours of the test samples are included in the test set and unseen during the training phase.

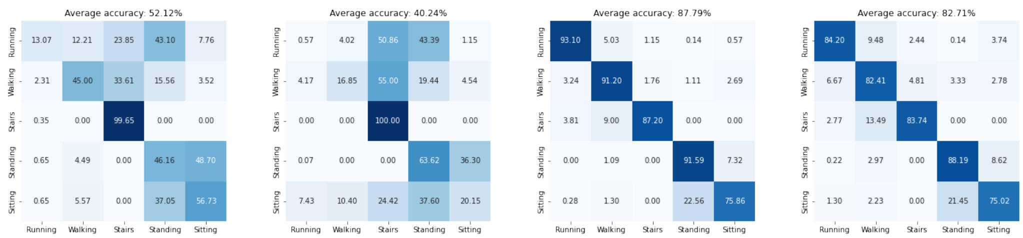
To obtain robust accuracy scores, each 6 minute window is used to cross validate on the training set.

In this paper, the algorithms chosen for prediction are Random Forest (RF), K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), Nearest Centroid (NC), Gradient Boosting (GB) and Multi Layer Perceptron (MLP). All algorithms, except for the MLP, are previously used and compared for SolAR [Sandhu et al., 2021]. The MLP, a subclass of neural networks, is included as it could capture relations which are too complex for other classification algorithms.

For classification, models are split, trained and tested using the Python package scikit-learn [Pedregosa et al., 2011].




(a) Random Forest (b) K-Nearest Neighbor (c) Support Vector Machine (d) Decision Tree




(e) Naive Bayes (f) Nearest Centeroid (g) Gradient Boosting (h) Multi Layer Perceptron

Figure 4: Confusion matrix for different classification algorithms, dark values indicate a high prediction rate for the given actual and predicted label. **Rows**: actual activity, **Columns**: predicted activity.

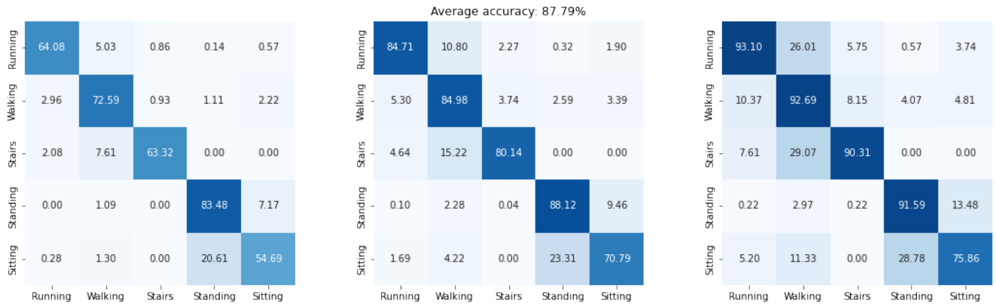
# 5 Results

In SolAR [Sandhu et al., 2021], a clear distinction between indoor and outdoor measurements were made due to significant differences in lighting conditions. In this paper, no distinction between indoor and outdoor circumstances are made. This choice enables us to emulate a real-life scenario, where one might enter or exit a building during a walk or perform similar tasks both indoors and outdoors.

Figure 4 depicts confusion matrices for each of the machine learning algorithms. The diagonal line of squares from top left to bottom right indicates accurate classification. We can make several observations when viewing the matrices of each algorithm.

Three of the classifiers perform relatively bad when compared to the others, reaching a line beyond which they can be considered unusable. Specifically, these classifiers are the SVM, NB and NC (average accuracy of 34%, 52% and 40%, respectively). As these algorithms perform notably worse, we can assume several properties of the activity data based on assumptions made by these algorithms:

1. The activities are not linearly separable (assumed by the State Vector Machine).
2. The features are not distributed normally (assumed by the Naive Bayes classifier).
3. The activities are not clustered around centre points (assumed by the Nearest Centroid classifier).




(a) Minimum accuracy values (b) Average accuracy values (c) Maximum accuracy values

Figure 5: Aggregate confusion matrices of the results for the five more accurate algorithms. Note that only the average accuracy scores add up to 100%. **Rows**: actual activity, **Columns**: predicted activity.

From here on, we will adjust focus towards the other five, more accurate, classification algorithms: RF, KNN, DT, GB and MLP. These achieve an accuracy range of 67 to 87%. Interestingly, one pair and one triplet of labels are most prominently confused by these five classifiers.

To better understand the accuracy scores of the five classification algorithms, data is aggregated and presented in figure 5, which shows the minimum, average and maximum accuracy scores for each pair.

When comparing the pair of sitting and standing, we can observe standing is labelled as sitting in 7% of the cases and sitting is labeled as standing in 20% of the cases. As depicted in figure 4, each of the five more accurate classifiers confuse this pair. Presumably, data obtained by sitting and standing is similar to such a degree that it cannot be properly distinguished in a significant amount of cases.

Changing the focus to the triplet of walking, running and traversing stairs, we can observe mislabelling with a preference for walking over running and traversing stairs for each of the five accurate classifiers. Interestingly, this preference differs greatly depending on each classifier, especially the KNN and MLP classifiers show a tendency to mislabel running and traversing stairs. However, no direct pattern can be observed when combining the accuracy of the classifiers, for this specific triplet, indicating these three different activities are separable using a classification algorithm for most of the cases.

# 6 Responsible Research

This section explains the ethical aspects of the research performed, followed by the steps taken to enable reproducibility of the project.

## 6.1 Ethics

The research done for this paper includes several ethical aspects: data collection and storage, data trimming, and publication of results.

Data collected throughout the project presents itself in the form of CSV (Comma Separated Values) files, each entry in the files is defined as a measurement value at a given timestamp. After a data collection session, the resulting file is renamed for labelling and copied locally to a computer for redundancy. After project completion, the data on both the SD card and computer is deleted.

No personal information is gathered throughout the process, furthermore, any data stored and labelled contains no further (personal) information. It could however, contain unexpected underlying data about the participant. It is therefore necessary to delete the gathered data. After project completion, data on both the SD card and computer is deleted.

To simulate a realistic scenario, data from all collection sessions is removed from the start of the session. This trimming is performed on the first 200 seconds of collection, to prevent the inclusion of data generated when setting up the prototype.

Published results during this project are limited to the outcomes of the applied Machine Learning algorithm, with the exception of the preview in figure 1. The outcomes of such an algorithm are based on several unpublished internal variables, meaning the results are not traceable to the original data.

## 6.2 Reproducibility

The steps taken to obtain results can be split up into four parts: prototype design, data collection, feature extraction and classification.

The prototype design is explained in section 4.1, where we can find the exact electronics used to measure the solar cell output values. The explanation together with the electrical diagram presented in figure 3 enables a reader to create a similar measurement setup to reproduce measurements results.

Data collection details can be found in section 4.2, where different circumstances are identified when data is collected. These variables can be recreated to collect similar data. However, the data collection process for this project is partly performed under natural weather circumstances. Sunny and cloudy days can heavily differ based on location and season. The reproduced data could be vastly different than data obtained during this project.

Operations for feature extraction are explained in section 4.3.1, first explaining how data is transformed from raw data into usable and meaningful data. The features extracted are properly identified and a reference for further explanation is included. The functional parts used to extract data are noted for future use. Extracted data is not saved for long term storage, but saved in memory for use of classification only.

Classification of processed data is explained in section 4.3.2. To achieve reproducibility for this part, tools and algorithms used for classification are explained. Although Machine Learning classification algorithms can be based on selection of random variables, the variables tend to converge towards an optimal point. Resulting in similar outcomes when employing these algorithms.

# 7 Discussion

When comparing results obtained for this paper to the results achieved by SolAR [Sandhu et al., 2021], we can observe a similar accuracy. Generally, the combined accuracy achieved by SolAR exceeds 88%, while results obtained in figure 4 reached a maximum of 87.9%. At fist, this appears unremarkable, but when we compare the confusion matrices of figure 5b to the results of SolAR, some notable difference appear.

The mislabelling rate for sitting and standing are notably higher when normal activities are introduced, with the maximum labelling accuracy for standing and sitting being at 91% and 75%, respectively, as observable in figure 5c. In SolAR, a much higher accuracy was achieved when labelling both sitting and standing: at least 99 % for standing and 91% for sitting.

The mislabelling rate for walking and running are notably lower when normal activities are introduced, where the average confusion rate of running and walking is 10% and 5%, the average confusion in SolAR is 10% and 12%. Interestingly, SolAR showed a bigger confusion for running and walking indoor than outdoor, which could in turn be an effect of the classification algorithms fitted to the lighting conditions of the controlled environment (on a treadmill for example).

Further inspecting the confusion matrices in figure 5b, we can observe significant confusion of the pair {running, walking} for {standing, sitting}, and vice versa with a minimum 1 to 2 percent in some cases. Such a confusion was not present in SolAR, which achieved a mislabelling rate of 0 percent.

To understand this difference, we will take a look at the circumstances surrounding data collection. Activities performed as they would be performed in day-to-day circumstances, include small accompanying actions, like writing along, flipping a page or motioning during a conversation. These actions can influence extracted features in an unpredictable manner, similar to the fluctuation occurring during sitting in figure 1. Such a fluctuation can influence the features to appear similar to other classes, which can in turn cause the sample to be mislabelled.

# 8 Conclusions and Future Work

This paper looks at the previous research of SolAR [Sandhu et al., 2021], which uses a hand-mounted solar panel to achieve energy positive Human Activity Recognition, and introduces data generated in a realistic scenario.

A prototype is created to wrist mount a solar cell, which mimics a watch. Solar cell output is used to classify activities performed while wearing the prototype. Five different activities are performed in varying circumstances. Like SolAR, the data is transformed and used to train and test several classification algorithms.

When comparing the results with the results achieved by SolAR, we observe that at first sight, no significant impact on accuracy occurs. However when we further examine the results, we find that the mislabelling rate is distributed more towards the pair of sitting and standing, instead of walking and running. A significant increase in mislabelling rate occurs, presumably caused by the introduction of ordinary actions during activities.

With the focus of obtaining realistic activity data, several options remain unexplored. Firstly, creating a small stand-alone prototype without the need for external power would result in unhindered usage for the most realistic scenario. Secondly, introducing a large representable pool of testers would result in representable activity data. Additionally, one can further raise the accuracy by introducing an outlier detection algorithm to remove outliers from the predicted data.

# A Activity Scenarios

| Activity | Indoor/ Outdoor | Location | Scenario | Time[1] (minutes) |
|---|---|---|---|---|
| Sitting | Indoor | at desk | working at pc | 72 |
| | | on couch | reading | |
| | Outdoor | at table | working at laptop | |
| | | in chair | reading | |
| Standing | Indoor | at desk | working at pc | 77 |
| | | in kitchen | cooking | |
| | Outdoor | - | in conversation | |
| Walking | Outdoor | in Delft | routine walk | 71 |
| | Both | on campus | walking through campus | |
| Running | Outdoor | in Delft | routine jogging | 50 |
| | Indoor | on treadmill | routine jogging | |
| Traversing stairs | Indoor | in stairwells | walking up & down stairs | 50 |
| | Outdoor | in Delft | | |

# B Actions

| Activity | Action |
|---|---|
| Sitting | writing along, drinking coffee/tea/water, flipping page, fidgeting, opening/using phone, moving chair, adjusting headphones |
| Standing | writing along, drinking coffee/tea/water, flipping page, fidgeting, opening/using phone, motioning in conversation, using kitchen utensils, adjusting headphones |
| Walking | opening/using phone, pressing button to cross street, adjusting headphones |
| Running | watching time/pace, drinking water, adjusting pace (treadmill), adjusting headphones |
| Traversing stairs | opening/using phone, adjusting headphones |

# References


[Harris et al., 2020] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585:357–362.


[1] Excluding 200 seconds grace period to set up

[Khalifa et al., 2018] Khalifa, S., Lan, G., Hassan, M., Seneviratne, A., and Das, S. K. (2018). Harke: Human activity recognition from kinetic energy harvesting data in wearable devices. *IEEE Transactions on Mobile Computing*, 17(6):1353–1368.

[Lara and Labrador, 2013] Lara, O. D. and Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials*, 15(3):1192–1209.

[Li et al., 2016] Li, T., Liu, Q., and Zhou, X. (2016). Practical human sensing in the light. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '16, page 71–84, New York, NY, USA. Association for Computing Machinery.

[Li et al., 2018] Li, Y., Li, T., Patel, R. A., Yang, X.-D., and Zhou, X. (2018). Self-powered gesture recognition with ambient light. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, page 595–608, New York, NY, USA. Association for Computing Machinery.

[Ma et al., 2019] Ma, D., Lan, G., Hassan, M., Hu, W., Upama, M. B., Uddin, A., and Youssef, M. (2019). Solargest: Ubiquitous and battery-free gesture recognition using solar cells. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, New York, NY, USA. Association for Computing Machinery.

[McKinney et al., 2010] McKinney, W. et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

[Sandhu et al., 2021] Sandhu, M. M., Khalifa, S., Geissdoerfer, K., Jurdak, R., and Portmann, M. (2021). Solar: Energy positive human activity recognition using solar cells. In *2021 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10.

[Texas Instruments, 2009] Texas Instruments (2009). Ads111x ultra-small, low-power, 2c-compatible, 860-sps, 16-bit adcs with internal reference, oscillator, and programmable comparator.

[Virtanen et al., 2020] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

[Wang et al., 2016] Wang, Q., Zuniga, M., and Giustiniano, D. (2016). Passive communication with ambient light. In *Proceedings of the 12th International on Conference on*

*Emerging Networking EXperiments and Technologies*, CoNEXT '16, page 97–104, New York, NY, USA. Association for Computing Machinery.

[Zhang et al., 2020] Zhang, R., Zamalloa, M. A. Z., Jelicic, V., and Siegel, M. (2020). Exploiting color sensors to provide optimal lighting and anonymous tracking in stores. In *2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 1–9.