An Energy Harvesting Portable and Rollable Large Area Gestural Interface Using Ambient Light

George Quine, Computer Science, Swansea University, 954390@swansea.ac.uk

ABSTRACT

In this work, we explored the viability of using ambient light and a large photovoltaic sheet for energy harvesting and gesture recognition. Our prototype consists of a large, portable, and rollable gestural interface which can uniquely distinguish and classify distinct hand gestures performed by a user. The system works under the principle that the amount of power harvested by the photodiodes decreases when a near-field object blocks the surrounding light. By monitoring these fluctuations, we recognised that different shadow patterns produce a distinct signature in the amplitude of the harvested voltage. Delegating the detection responsibilities to machine learning, it was possible to capture the hidden meaning within the hand gestures to perform an action in real time. We focused on two classifiers, one utilising a machine learning technique, Random Forest (RF), and the other a deep learning classifier, a Convolutional Neural Network (CNN). To further improve the robustness of the system, we applied two pre-processing techniques known as Normalisation and Principal Component Analysis to reduce inherent noise caused by inevitable environmental and human factors. We evaluated the proposed system under a variety of lighting conditions, as well as assessing the significance of the two pre-processing techniques. We trained our models with 1,050 incidents of 5 unique gestures. The CNN demonstrated the highest overall accuracy in all lighting conditions, with 95% accuracy in 1K lux. The RF performed similarly well, obtaining 93% accuracy in 1K lux. Using a designed Graphical User Interface (GUI), both models are capable of recognising an unseen gesture in 0.05 seconds

KEYWORDS

Gesture Recognition, Photodiodes, Solar Energy Harvesting, Machine Learning

1 INTRODUCITON

Today, smart watches, portable tablets and smart home appliances have changed the way we collect, use, and share data [1][2][3]. As society and technology continue to change, there is a need for a new approach to communicate with these devices. Hand gesture recognition presents a new way for computers to begin to understand human body language. This provides a more efficient pathway for human machine interfacing (HMI) without the need of physical touch. It is believed that gesture-based interfaces can reduce the complexity of communication [4]. Imagine the safety benefits of interacting with your car's functions without taking your eyes off the road [5].

The term gesture recognition refers to the whole process of tracking the performed human gesture, converting the gesture to numerical data, and interpreting the data to represent the meaning of the given command, a non-verbal communication. Many studies have been conducted to track body movements, the most common method being the use of cameras and computer vision algorithms to interpret gestures. The Microsoft Kinect [6] and the Leap Motion Controller [7] are two such examples, both use infrared cameras to gather a detailed analytical view of the user. Although this technique has proven to be very effective, to recognise these gestures applications must invade the user's privacy and possess the image input from a camera to gain information of the user's intensions. Other methods studied include contact-based devices which identify gestures by analysing the physical interaction of the user from sensors like accelerometers and gyroscopes. The Nintendo Wii [8] remote and CyberGlove III [9][10] can obtain relatively accurate information due to the direct attachment with the users, however, a contact-based device can be inconvenient to wear and may require the user to stand in front of a receiving sensor. As the demand for remote and disposable devices increase, there is growing interest in battery free systems [11].

Gesture technology offers multiple advantages over other forms of HMI across multiple applications. In the gaming industry, using gesture recognition in computer games would make the experience more interactive for consumers, especially when combined with existing Virtual Reality systems [12]. Developers have put a lot of energy into the graphics, UX and sound to build a sense of reality. Although this sense of reality is removed when the user must hold a controller in either hand. This reminds the user that their surroundings are an illusion. Gesture recognition will allow the user to have the freedom of movement without the burden of holding a controller. The technology also presents applications in industrial settings for use with human robot collaboration. Some factory environments make other HMI system challenging to implement, as noise levels may be too high for voice recognition systems and using touch screen interfaces usually requires a worker to stop his task. Gesture recognition systems would enable workers to control smart robotics and other IoT systems from a distance while still at their station.

1.2 PROPOSED METHOD

In this paper, we propose a battery-free system that can perform gesture detection and recognition using the surrounding ambient light. Our prototype will consist of a large, portable, and rollable gestural interface which can uniquely distinguish distinct hand gestures from a user. By featuring a rollable photovoltaic sheet, its display size and form factor can be dynamically changed to the user's preference. This enables the system to be highly adaptable with many environments and provides the user a free range of interactions with its portability. The photovoltaic sheet can be rolled out quickly, offering a perfect solution for portable use. The system can be deployed over many large surfaces such as a wall, table, window blind, and cabinet. Additionally, the flexible material enables the system to be deployed over curved or undulating surfaces such as a sofa arm rest. This advancement should be advantageous for consumers where space, size and surface characteristics vary.

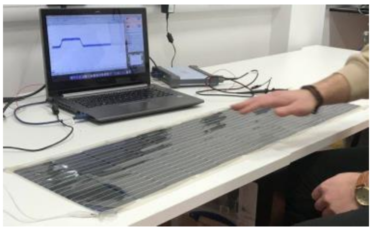

Figure 2 - The Photovoltaic Sheet and Recognition Signal

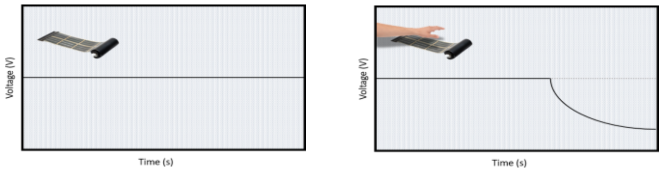

Figure 1 – Time-series of photodiodes harvested power as no hand is placed above the surface interface (left) or hand is placed above surface interface (right).

Relying heavily on batteries significantly limits an applications scope. Replacing batteries or recharging is inconvenient and can be problematic if there is no nearby power resource. In recent decades, solar energy has been one of the most widely investigated source for energy due to the increasing affordability of the technologies as well as increased political pressure to shift towards green energy sources [13]. Sunlight can help overcome the energy limitations of batteries. The photovoltaic sheet has self- powering capabilities by harvesting energy from the surrounding light. Furthermore, as awareness of the global climate emergency increases, solar energy provides a viable alternative to burning fossil fuels and a pathway to reduce global warming [14]. The most important element of the system, the photodiodes, can directly convert sunlight into electricity by a phenomenon known as the photovoltaic effect. The maximum power generated by the photodiodes increases with the increase in light intensity. This will offer advantages for low-cost installation, portability, and sustainability. The final system will consist of a large rollable photovoltaic (PV) sheet [15], an energy harvesting circuit board (InfinityPV OPV3W60V [16]) and a microcontroller board for gesture recognition. As it can be seen in Figure 1, the length and width of the PV sheet is 110 cm and 23 cm, respectively, weighing no more than 300g. The photovoltaic sheet can generate up to 12 W at 1 sun or 0.03 W at 300 lux for indoor lighting. We receive a voltage signal which appears as an AC modulation that is filtered by a 1 μF series capacitor. This signal will be used for gesture detection.

The system will work under the principle that the amount of power harvested by the photodiodes decreases when a near-field object blocks the surrounding light. As demonstrated in Figure 2, moving a hand above the surface interface in close range will cut off specific light arrays from reaching the photodiodes, causing a sharp dip in the receiving voltage signal. By monitoring these fluctuations, we recognise that different gestures leave distinguishable signatures. Delegating the detection responsibilities to machine learning, we can capture the hidden meaning in the hand gestures to perform an action in real time. As long as there is light, the system will work anywhere. This is particularly useful for remote regions with no access to any other source of electricity.

2 RELATED WORK

Y.li et al, presented two prototypes under the domain of gesture detection, applying the concept to a smart watch and smart pair of glasses [17]. Their approach relied on an array of visible small, low-cost photodiodes surrounding the object to harvest energy from the surrounding ambient light and to recognise specific finger gestures in proximity. No machine learning technique was used in this project. Recording the change in light intensity across every individual photodiode, Y.li et al could determine the direction of the users' fingers movement by noting the index of the first and last blocked photodiode. Both the watch and glasses achieved 99% precision and 98% recall. Y.li et al conducted a study to measure the amount of energy the two prototypes can harvest in four indoor lighting conditions (200 lux to 2K lux) and three outdoor lighting conditions (4K lux to 110K lux). In the indoor conditions, the devices were capable of harvesting power in ranges from 23 μW to 124 μW. Whereas, in the outdoor conditions, the amount of power harvested by the two prototypes were significantly higher, ranging from 1.3 mW to 46.5 mW. Their studies show that the harvested energy is sufficient to power the gesture recognition module unless the user is sat in a dark room. This issue can be alleviated by using super-capacitors, where 1 – 3 seconds under direct sunlight can provide the system sufficient enough of power to last for one hour in a dark room.

Yet, D. Ma, et al. evaluated the performance of multiple machine learning classifiers [18]. This included a Random Forest (RF) and three other classifiers: Support-Vector Machine (SVM), K-nearest neighbours (KNN) and a Decision Tree (DT). D. Ma, et al. explored the use of three different solar cells, a 10x5cm silicon-based solar opaque solar cell (S1) and two 1x1cm transparent solar cells (T1 and T2) with different transparencies (T1 20.2% and T2 35.3%). Under different light conditions, six user friendly hand gestures were explored, these included: "*Up, Down, UpDown, DownUp, LeftRight and FlipPalm*". All three solar cells used in this project were smaller than the average user's hand which limited their scope for more gestures. Because the system lacks a sense of direction, it would struggle with complex movements such as clockwise movement or swiping from left to right. Due to the small size of the gesture interface, the system had difficulty distinguishing between the basic motions, FlipPalm and UpDown. Although removing the *FlipPalm* gesture, the authors received a maximum accuracy of 97.1%, 96.5% and 96.1% for S1, T1, and T2, respectively using KNN at a super high light intensity of 2600 lux.

3 PROPOSED METHOD

In this section we will be exploiting the effect of different hand movements above the photodiode surface. Here, we introduce the chosen hand gesture set, followed by our recognition frameworks, and the proposed pre-processing techniques used to remove noise from the signal waveform.

3.1 GESTURE SET

As shown in Figure 3, we explored 5 different hand gestures, these included *a circular swipe clockwise, a circular swipe anticlockwise, opening and closing of two fists, a swipe to the left, a swipe to the right*. These gestures were chosen for the prototype with useful functional controls in mind. A circular swipe clockwise could turn the volume up, whereas the circular swipe anticlockwise could turn the volume down. Open and closing of the fist could select the item of choice. Swiping to the left and right could be used to navigate through a horizontal catalogue.

Figure 3 – (1) Circular clockwise, (2) Circular anticlockwise, (3) Opening and closing of two fists, (4) Swipe to the left, (5) Swipe to the right.

Typically, the photocurrent module would produce a stable voltage reading with little noise. Although as explained, moving a hand above the surface interface in close range will cut off specific light rays from reaching the photodiodes, causing a substantial dip in the receiving voltage signal and therefore changing the shape of the signal's waveform. By observing different hand gestures, it was clear to us that these 5 gestures selected above left distinguishable unique signatures within the receiving voltage, as seen in Figure 4. When the signals corresponding to the circular clockwise and circular anticlockwise gestures are compared, they are inverted in time. This may also be seen when comparing the left and right swipe gestures. A double hand gesture, on the other hand, causes the voltage signal to drop and rise sharply with a sustained period of low voltage in between.

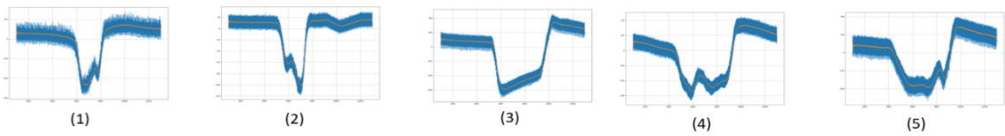

Figure 4 – (1) Circular clockwise, (2) Circular anticlockwise, (3) Opening and closing of two fists, (4) Swipe to the left, (5) Swipe to the right.

3.3 DATA COLLECTION

During data collection, we connected the photodiode interface to an oscillator (PicoScope 5000 series [19]) as shown in Figure 1. Here we carried out the series of gestures listed above in section 4.1. We recorded 15006 photocurrent voltage values over a 5-second time period using the PicoScope and its accompanying data logging software. The data was obtained at a maximum sampling rate of 20k samples per division, with a 500-millisecond collection period and a photocurrent input range of ±500 mV. For a single gesture, the photovoltaic module's resultant readings are saved as a CSV file, each containing data separated under two columns, Time (s) and Channel A (mV). The total number of rows in each file is 15009, the top 3 being header rows. The gesture incidents were then saved into folders, each folder named after the gesture it contains. The Data was collected all by one user using their right-hand. The user performed the task sitting at a desk 2 meters below a ceiling flush light. Data collection was carried in a bright laboratory which measured 750 lux using a digital light meter. Prior to the start of the study, the user was given several minutes to practice the hand gesture before the analysis began. Once the user was confident, the user carried out the 5 hand gestures, repeating each gesture 30 times. In total, we collected 1,050 gesture instances for training and testing of the machine learning models (7 Sessions x 5 gestures x 30 repetitions).

3.4 RECOGNITION FRAMEWORK

When the user performs a hand gesture above the photodiode surface, the system captures a time-series of photocurrents and transfers the data to a gesture recognition system in a numerical form. The main efficiency metric for our system is the gesture recognition accuracy. In this paper we are proposing a machine learning based gesture detection system to achieve superior feature extraction from the time series of photocurrents received from the photovoltaic sheet as a 1-dimensional voltage waveform. Before we can apply our machine learning models to the raw data, it is extremely important that we pre-process our data. Every time a user performs a gesture, the time series data produced by the photodiodes will differ to some degree. Differences in the resulting data may be caused by a variety of cases. 1) Variations in user parameters such as hand size, hand angle, speed of the hand motion, and the proximity of the user's hand to the solar interface; 2) Differences in environmental parameters such as light intensities, for example, turning on/off a light whilst performing an action, or temporary blockage of sunlight due to a moving cloud. Therefore, for the machine algorithm to achieve better results, we must identify and discard the inherent noise in order to minimize the variations before being applied to the classifier.

3.5 PRE-PROCESSING OF RAW DATA

As it can be witnessed in Figure 5, It was evident that for every individual incident of a gesture the graph presented a stable voltage reading before and after the gesture had been performed. This was a result of the user's hand not being present above the solar panel and the gesture had yet to be performed. I realised I could remove extraneous datapoints by reducing the time range from both the start and end of a gesture. Because all gestures must contain the same number of features, it was critical that the new time range was broad enough to encompass the start and end of all gesture signatures while being small enough to remove superfluous datapoints. In addition, I extracted the estimated weighted mean with a span of 200 to significantly reduce noise whilst preserving the gestures direction. As a result, after optimising the time range and extracting the estimated weighted mean, we were able to delete 3,006 superfluous datapoints from each unique gesture. This left each gesture with 12,000 voltage datapoints from its original 15,006, resulting in a significant reduction in memory and computational time.

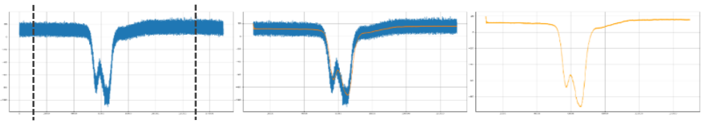

Figure 5 – The removal of extraneous data. First, the data was cropped. Second, the Estimated Weighted Mean was Calculated. Third, noisy data was removed

3.5.1 Principal Component Analysis (PCA)

High dimensional data can pose problems for machine learning as predictive model may run into the risk of overfitting. Dimensions are nothing but features that represent the data. Feeding our machine learning models with this many datapoints will lead to a very slow performance and poor accuracy. PCA is the most used and most popular statistical method for transforming attributes of a dataset into a new set of uncorrelated attributes known as principal components (PCs) [20]. PCA's main goal is dimensionality reduction and feature selection, taking datasets with many features and simplifying that dataset by capturing the principal components that hold the most variance to summarize the data using less properties. PCA is performed by eliminating insignificant features from a high-dimensional space and projecting the most important features into a lower-dimensional subspace, improving classification accuracy and reducing computational time [21]. In this paper, we wanted to apply the minimum number of principal components such that 99% of the variance was retained. A sample output of PCA is shown in Figure 6, with cumulative explained variance sketched as a function of the number of principal components. We discovered that 15 principal components are the bare minimum for retaining 99.007% of the variance in the dataset. In other words, using PCA we have reduced 12,000 datapoints for each gesture to 15 PCs without compromising the information extracted from the data

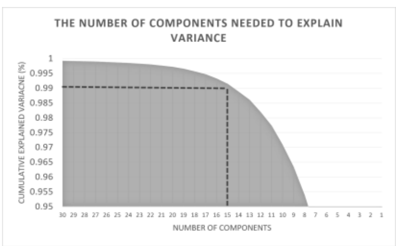

Figure 6 - The number of components needed to explain variance

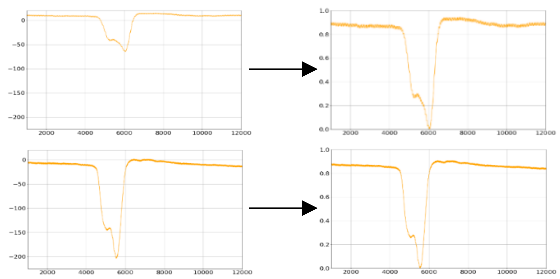

Figure 7- (Top) The application of Normalisation to a gesture performed at 350 lux. (Bottom) The application of Normalisation applied to a gesture performed at 750 lux

3.5.2 Magnitude Normalisation

The magnitude of the voltage produced by the photodiodes will most likely change between gestures as a result of a combination of human and environmental factors such as the distance between the user's hand and the photovoltaic interface or gestures performed under varying light intensities. Therefore, for every gesture, we applied a MinMax scaling technique to the voltage magnitude and map the features between the range of 0 and 1, where 1 denotes the maximum magnitude value and 0 denotes the minimum magnitude value.

In Figure 7, the above two graphs were captured at a light intensity of 350 lux, whereas the bottom two graphs were captured at a light intensity of 750 lux. The results prior to normalisation are on the left, and the results after normalisation are to the right. The magnitude of the two light intensities differs dramatically, with 350 lux only capable of producing a photodiode input range of 20 to -65mV. Whereas, at 750 lux, the signal range is much larger, ranging from 20 to -205mV. Applying normalisation will guarantee that all features are in the exact same voltage magnitude scale, allowing the classifier to easily compare different gestures performed under different intensities.

4 Classifiers

For this project we will be comparing one commonly used classifier, Random Forest, against a deep learning machine learning method, Convolutional Neural Network, for gesture recognition in order to determine which model is best suited for my application. Using the 1,050 gesture instances obtained throughout the 7 sessions, we will evaluate the two models by comparing their performance results.

4.1 Random Forest

A Random Forest is a simple yet powerful technique that works on the assumption that a collective opinion of a crowd is smarter than a single individual. The Random Forest, as the name suggests, is an algorithm that aggregates a large group of individually trained decision trees to create a 'forest' that will act as a committee to make a more accurate prediction. Each individual decision tree is trained in parallel using various subsets of the training dataset, resulting in a large diversity of trees which generally results in improved performance. For the final decision, the RF classifier aggregates the decision of each individual tree and selects the prediction result with the most votes as the final prediction. A Random Forest is invariant to the scaling of the data and requires little effort in the fine-tuning of hyperparameters, yet it maintains a high reproducibility. A RF can almost work out of the box and that is one reason why they are very popular [22][23].

4.2 Convolutional Neural Network

A CNN is a deep learning machine learning method that is based on the artificial neural network (ANN). The most commonly used CNN is the 2 dimensional-CNN which is best suited for pattern recognition and image classification. Inherited from the 2D-CNN is a newly emerged 1 dimensional-CNN which has been proven to be very effective in extracting features from 1D sequential time-series data. The primary difference between the two models is the structure of the input data and how the convolutional filter traverses through the data to extract features. Prior to the introduction of 1D-CNNs, 1D signals would have had to be explicitly converted into suitable 2D image formats before being fed into a conventional 2D-CNN. This conversion of 1D signals into image results in a loss of information, which makes 1D CNNs superior in cases of 1D signals. Due to the one-dimensional nature of our data, the proposed model will be a 1D-CNN, where the input layer will receive the raw 1D signal before passing it forward towards the two convolutional layers (), a fully connected layer, and a SoftMax activation function to determine the final output of the node. The output for the model will be a 5-element vector containing the probability of a given window belonging to each of the 5 gestures. [24] [25]

5 SYSTEM TESTING

In this section, we will assess the performance of our final two systems in relation to a variety of environmental conditions. In order to demonstrate the significance of both pre-processing techniques, we shall also evaluate both models with and without PCA and Normalisation. All evaluations will be performed on Google Collaboratory's NVIDIA Tesla T4 GPU [26].

CNN				
Evaluated Dataset (lux)	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Validation Dataset (750)	96	96	96	95
Near Window (1K)	95	96	95	95
Bright Lab (750)	94	94	94	94
Normal Office (650)	83	83	83	83
Kitchen (450)	71	67	71	67
Living Room (350)	47	48	47	44
Dark Room (250)	31	26	31	28

Table 2 - Evaluation Table for CNN

RF				
Evaluated Dataset (lux)	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Validation Dataset (750)	90	90	90	89
Near Window (1K)	93	94	93	93
Bright Lab (750)	86	88	86	86
Normal Office (650)	80	81	80	80
Kitchen (450)	67	62	67	63
Living Room (350)	40	45	40	40
Dark Room (250)	26	23	26	22

Table 1 - Evaluation Table for RF

EXPERIMENT 1: THE IMPACT LIGHT INTENSITY HAS ON OUR FINAL MODELS

Both our models rely on ambient light to identify the gesture being performed above. Here we will evaluate both models' robustness across 6 unseen different light intensities, including a near window (1K lux), a bright lab (750 lux), a normal office (650 lux), a kitchen (450 lux), a living room (350 lux), and a dark room (250 lux). In each of the tested light conditions, we have collected 150 gesture instances (5 gestures x 30 repetitions). These light intensities were measured using a digital light meter directly above the surface of the photovoltaic sheet. All samples were taken by one user. Table 1 and 2 shows the results of all tested conditions with their accompanying evaluation metrics.

Both models demonstrated excellent precision and recall in unseen light levels surpassing 750 lux. This came to no surprise given that the dataset used to train both models were collected at 750 lux. When placed near a window at 1K lux, both models outperformed the validation dataset. At high light intensities, the photovoltaic sheet could compute clean clear signal waveforms when a gesture was being performed. This makes it easier for the model to distinguish between gestures, which is useful for gesture recognition. According to our confusion matrices for 650 lux in Figure 8 and 9, a right swipe is the first of the gestures to become difficult to classify and is commonly misinterpreted as a double hand gesture. As we continue to decrease the light intensity to 450 lux, the system now anticipates all right swipe actions as a double hand gesture. Left swipe also becomes harder to distinguish and is commonly mistaken as a right swipe gesture, whilst for all other classes, both models continue to perform flawlessly. More than half of all gestures classified below 350 lux are now inaccurate which is unacceptable for a gesture detection system. In low light conditions the light intensity is not sufficient enough to create substantial signatures in the receiving voltage.

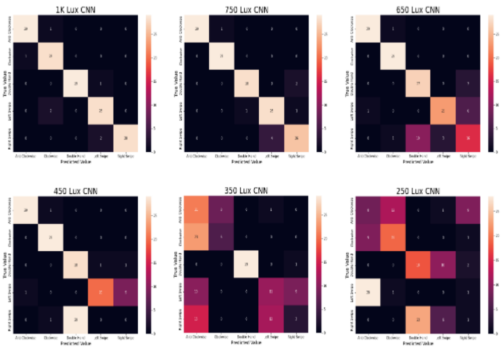

Figure 8 - Confusion Matrix for CNN in various light intensities

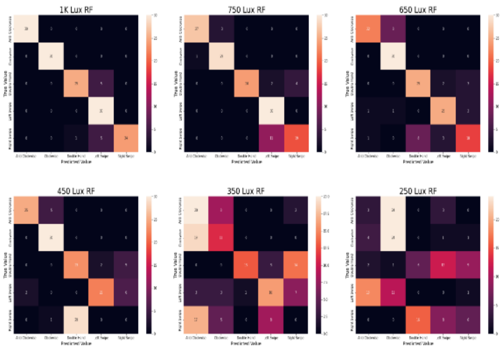

Figure 9 - Confusion Matrix for RF in various light intensities

EXPERIMENT 2: THE FLICKER EFFECT

By executing the gestures beneath a flickering light, we can now explore the influence of sudden and abrupt changes in ambient light. To begin with, we have changed the light above the photovoltaic sheet to a flickering light, flashing at approximately 10Hz, oscillating between 750 lux to 550 lux at the photovoltaic sheet. The user performed 14 instances of each gesture below the flickering light (5 gestures x 14 repetitions = 70 total instances). Results show that both models are incapable of detecting the gestures when the light intensity is oscillating (31% accuracy for both RF and CNN). Changing the light intensities at such a rapid rate will degrade the voltage signal captured by the photovoltaic sheet. Although this scenario would be uncommon in any environment since the flashing light is noticeable to the naked eye.

EXPERIMENT 3: THE IMPORTANCE OF PRINCIPAL COMPONENT ANALYSIS

Figure 10 shows the comparison between both models with and without the incorporation of Principal Component Analysis. A Convolutional Neural network with no PCA application shows no ability to learn. In fact, the CNN model was so inept that it identified all gestures as anti-clockwise for all light intensities. This could be owing to the capacity and complexity of our model. Our chosen CNN model was built for a small number of inputs, with only two convolutional layers: the first with the same number of neurons as the number of inputs, and the second with 400 neurons. When the CNN model is introduced with 12,000 raw data inputs instead of 15 principal components, the model tries to fit to the noise of the input data and so it becomes excessively specific, making it difficult to generalise to new training data. When the model attempts to condense the feature representation from 12,000 features into 400 neurons found at the second convolutional layer, in a bottleneck fashion, it fails to sufficiently learn and preserve the most important features prior to classification.

In Random Forest, however, we discover that the performance ability of a model with and without PCA is not significantly different. Although PCA increases the model's performance, particularly in low-light conditions, the Random Forest model can still learn just as well when fed vast amounts of raw data. Unlike our CNN model, Random forests are immune to feature magnitude, they are invariant to monotonic transformations of individual features. What does enhance the model's ability to perform better is that PCA removes irrelevant, noisy, and redundant features from the feature vector, hence feeding the model with clean improved data will enhance the RF model's performance. Based on the test results of the two models with and without PCA, it can be concluded that the success of PCA can be used to improve the accuracy performance of both Convolutional Neural Network and Random Forest.

EXPERIMENT 4: THE IMPORTANCE OF NORMALISATION

Now that we've established the significance of PCA. In this experiment, we will evaluate the effectiveness of normalisation on the feature input. Figure 11 depicts the accuracy of both models when normalisation has been used and hasn't been used. The experimental results show that when normalisation is used, the accuracy is always higher than when normalisation has not been utilised. As we can see, this pre-processing technique is capable of minimising the impact of light intensity on the model's performance to the greatest extent possible. Both models show a sharper decline in performance without normalisation than when normalisation has been applied, particularly after 650 lux. Based on the test results of the two models with and without normalisation, we can conclude that normalisation is essential for our final model due to its success in reducing the impact of both models' performances with light intensity.

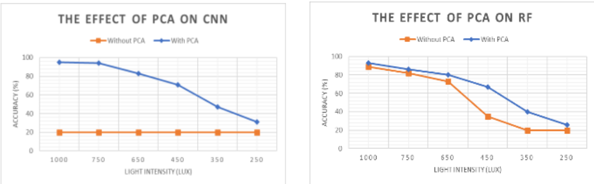

Figure 10 - The effect of PCA on both CNN and RF

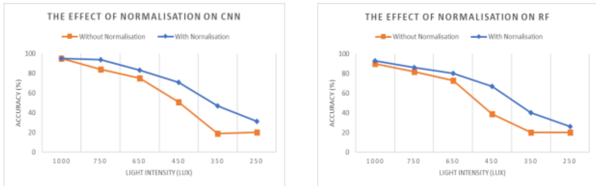

Figure 11 - The effect of Normalisation on both CNN and RF

6 DISCUSSION

The proposed device requires bright light to function reliably. This device is particularly suitable in environments such as a bright lab or near a window during the day. Insufficient light poses as a limitation to the device, for example a living room at 350 lux. With conditions below 350 lux, the photovoltaic sheet fails to produce substantial signatures in the voltage signal which makes it challenging for the classifiers to distinguish the true meaning behind a command. Inconsistent light can also contribute to an unreliable performance of the proposed device. Currently, drastic variations in light, such as turning on and off a light will result in the voltage signal to be too noisy for the device to function reliably. In order to function, the proposed device requires sufficient light. In bright light, such as near a window at 1K, using CNN and RF, the system will perform flawlessly, achieving 95% and 93% accuracy, respectively. The gesture class used to train the proposed system was effective in bright light conditions. Each gesture class left a unique distinguishable signature in the amplitude of the voltage. Although the signatures of a gesture started to deter as the light intensity was reduced. The double hand gesture was the last gesture to become problematic, which could be due to the fact that it is the only gesture that does not require a direction. Both clockwise/anticlockwise and left/right swipe gestures are inverted in time. With our current proposed system, we only have 1 dimension of voltage over a time series variable. This requires the user to always perform the clockwise/anticlockwise and left/right swipe gesture with the same hand that the model was trained with. To address this issue, we could propose that we divide the photovoltaic sheet into 4 electrical arrangements. One section to the right, one section to the left, another section at the top, and a final section at the bottom of the PV sheet. Doing so, having four independent voltage readings will provide the system with a sense of left/right and up/down direction when the user performs a gesture. This will minimise the confusion between inverted gestures and enable the system to recognise the same gesture regardless of which hand the user decides to perform it with.

All gestures used to train and test the proposed system were recorded by one user. The system is yet to be evaluated on other users. The system must be trained and evaluated using as many users as possible to make the system more robust. If the system doesn't work well on a large number of users, we could train and calibrate the system to a single user before use. Although, training of the system has been proven to be computationally expensive, this solution will require a decentralized approach. Because the proposed system is intended to be performed on a micro-controller, to address this problem, a single user could connect the microcontroller to a standard computer via a serial connection during the training process.

It is critical that our prediction models are both computationally inexpensive and demand little energy to function. In all lighting conditions, the CNN classifier outperformed the RF classifier in all metrics. However, we cannot dismiss the RF model just yet. The benefit of using a RF over the CNN is its simplicity to employ. Despite our efforts to keep the CNN architecture as minimal as possible, serialising the CNN model to transfer it to a microcontroller will result in a 35MB file size. A serialised trained RF, on the other hand, will be as little as 2.5 KB.

Using the designed GUI, it can be noted that both models are capable of recognising a gesture in 0.05 seconds. As a result, I can confidently state that the system can operate in real time utilising these classifiers.

7 CONCLUSION

In this paper, we proposed a self-powered rollable gestural interface using ambient light for both energy harvesting and gesture recognition. We developed a model that can detect a hand gesture conducted above the photodiodes in bright light conditions in less than 0.05 seconds.

There are numerous applications in this discipline, including virtual reality, sign language, and smart robotic control. When compared to other gesture recognition approaches, employing ambient light to identify gestures have various advantages. For starters, It is virtually impossible to maliciously create anything from the data of a silhouette that our proposed approach receives. The energy harvesting capabilities of our proposed system also offer a perfect solution for portable use, eliminating the inconvenience of replacing or recharging batteries as well as the difficulty of finding a nearby power resource. Additionally, the flexible rollable material enables the system to be deployed over undulating surfaces where its display size and form factor can be dynamically changed. This makes our system useful for consumers where space, size and surface characteristics vary.

We focused on two classifiers, one utilising a machine learning technique, Random Forest, and the other a deep learning classifier, 1D-Convolutional Neural Network. The CNN showed the best overall accuracy in all lighting conditions, achieving a maximum of 95% accuracy in 1K lux. The RF performed equally as well, achieving 93% accuracy in 1K lux. It is apparent that when the light intensity is reduced, the photovoltaic sheet struggled to produce substantial signatures in the voltage signal which made it challenging for the classifiers to distinguish the true meaning behind a given command.

With this work and the anticipated future work in mind, we successfully demonstrated the viability of using a large photovoltaic sheet for both energy harvesting and gesture recognition using bright light. In today's society, the emphasis must be on sustainability and environmental protection. This self-powered rollable gesture interface will be beneficial in situations where resources are restricted but light is available (sun). Currently, the results obtained from this project serve as a foundation for future development.

8 ACKNOWLEDGMENTS

Dr Deepak Sahoo, my supervisor, for his advice, encouragement, and enthusiasm towards my dissertation

9 REFRENCES

[1] Fitbit Inc. Fitbit. Available at: https://www. fitbit.com/.

[2] Apple Inc. Apple iPad. Available at: http://www. apple.com/ipad/

[3] June Life Inc. June oven. Available at: https:// juneoven.com/

[4] Joshua r. New, Erion Hasanbelliu, Mario Aguilar. Facilitating User Interaction with Complex Systems via Hand Gesture Recognition. 2003. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.132.49

[5] Faheem Khan, Seong Kyu Leem, Sung Ho Cho. Hand-Based Gesture Recognition for Vehicular Applications Using IR-UWB Radar. 2017. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5422194/

[6] Microsoft Corp. Xbox. Kinect. Available at: http: //www.xbox.com/en-US/xbox-one/ accessories/kinect-for-xbox-one

[7] Leap Motion. Available at: https://www. leapmotion.com/

[8] Nintendo Wii. Nintendo. Available at: https://www.nintendo.co.uk/Wii/Accessories/Accessories-Wii Nintendo-UK- 626430.html

[9] CyberGlove III. Cyber Glove Systems. Available at: http://www.cyberglovesystems.com/cyberglove-iii

[10] G. D. Kessler, L. F. Hodges, and N. Walker. Evaluation of the cyberglove as a whole-hand input device. 1995. Retrieved from: https://www.researchgate.net/publication/27521573_Evaluation_of_the_CyberGloveTM_as_a_Whole_Hand_Input_Device

[11] Vamsi Talla, Bryce Kellogg, Shyam Gollakota, Joshua R Smith. Battery-Free Cellphone. 2017. Retrieved from: https://dl.acm.org/doi/10.1145/3090090

[12] Oculus. Virtual Reality. Available from: https://www.oculus.com/

[13] S.P. Beeby, Z. Cao, A. Almussallam. Kinetic, thermoelectric and solar energy harvesting technologies for smart textiles. 2013. Retrieved from: https://www.sciencedirect.com/science/article/pii/B978085709342450 0110

[14] Felix Cruetzig, Peter Agoston, Jan Christoph Goldschmidt. The underestimated potential of solar energy to mitigate climate change. 2017. Retrieved from: https://www.researchgate.net/publication/319396014_The_underestimated_potential_of_solar_energy_to_mitigate_clim ate_change

[15] infinityPV Solar Tape. OPV. Available from: https://infinitypv.com/products/opv/solar-tape

[16] OPV3W60V – DCDC Converter – OPV. Available from: https://infinitypv.com/products/electronics/opv3w60v

[17] Yichen Li, Tianxing Li, Ruchir A. Patel, Xing-Dong Yang, Xia Zhou. Self-Powered Gesture Recognition with Ambient Light. 2018. Retrieved from: http://dx.doi.org/10.1145/3242587.324263

[18] Dong Ma, Guohao Lan, Mahbub Hassan, Wen Hu, Mushfika B. Upama, Ashrad Uddin, Mostafa Youssef. SolarGest: Ubiquitous and Battery-free Gesture Recognition using Solar Cells. 2019. Retrieved from: https://doi.org/10.1145/3300061.3300129

[19] PicoScope. PicoScope 5000 Series. Available from: https://www.picotech.com/oscilloscope/5000/flexible-resolution- oscilloscope

[20] Sartorius. What Is Principal Component Analysis (PCA) and How It Is Used? 2021. Available from: https://www.sartorius.com/en/knowledge/science-snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186

[21] Towards Data Science. A One-Stop Shop for Principal Component Analysis. 2021. Available from: https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c

[22] Tony You. Towards Data Science. Understanding Random Forest. 2019. Available from: https://towardsdatascience.com/understanding-random-forest-58381e0602d2

[23] Niklas Donges. BuiltIn. A complete Guide to the Random Forest Algorithm. 2020. Available from: https://builtin.com/data-science/random-forest-algorithm

[24] Nils. Good Audience. Introduction to 1D Convolutional Neural Networks in Keras for Time Sequences. 2018. Available from: https://blog.goodaudience.com/introduction-to-1d-convolutional-neural-networks-in-keras-for-time-sequences-3a7ff801a2cf

[25] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, Daniel J. Inman. 1D convolutional neural networks and applications: A survey. 2021. Available From: https://www.sciencedirect.com/science/article/pii/S0888327020307846

[26] Google Colab. Available from: https://colab.research.google.com/?utm_source=scs-index