ARTICLE IN PRESS

The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express xxx (xxxx) xxx

www.elsevier.com/locate/icte

Human activity recognition based on wrist PPG via the ensemble method

Omair Rashed Abdulwareth Almanifi[a], Ismail Mohd Khairuddin[a], Mohd Azraai Mohd Razman[a], Rabiu Muazu Musa[b], Anwar P.P. Abdul Majeed[a,c,e,f,d,*]

[a] *Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang, Pahang Darul Makmur, Malaysia*

[b] *Centre for Fundamental and Continuing Education, Department of Credited Co-curriculum, Universiti Malaysia Terengganu, Terengganu, Malaysia*

[c] *School of Robotics, XJTLU Entrepreneur College (Taicang) Xi'an Jiaotong-Liverpool University, Suzhou. 215123, P. R. China*

[d] *EUREKA Robotics Centre, Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, UK*

[e] *Centre for Software Development and Integrated Computing, Universiti Malaysia Pahang, Gambang, Pahang Darul Makmur, Malaysia*

[f] *Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur Campus, Kuala Lumpur, Malaysia*

Received 28 May 2021; received in revised form 11 November 2021; accepted 16 March 2022

Available online xxxx

Abstract

Human activity recognition via Electrocardiography (ECG) and Photoplethysmography (PPG) is extensively researched. While ECG requires less filtering and is less prone to disturbance and artifacts, nonetheless, PPG is cheaper and widely available in smart devices, making it a desired alternative. In this study, we explore the employment of the ensemble method with several pre-trained machine learning models namely Resnet50V2, MobileNetV2, and Xception for the classification of wrist PPG data of human activity, in comparison to its ECG counterpart. The study produced promising results with a test classification accuracy of 88.91% and 94.28% for PPG and ECG, respectively.

Keywords: HAR; Exercise; PPG; ECG; Classification; Ensemble; Machine learning; Transfer learning

1. Introduction

There is undeniable evidence that regular exercise and physical activity provide numerous benefits; promoting general well-being, as well as preventing several chronic diseases including cancers and cardiovascular disease. However, these health benefits can differ depending on many factors such as the frequency, the volume, and the type of activity performed. Automated recognition of exercise and physical activity has been a topic of research for many years. This interest springs from the many possible applications for Human activity recognition (HAR) which include its use in surveillance systems, the monitorization of the elderly, rehabilitation processes of patients, and many uses that pertain to the Internet of Things (IoT) [1].

Predominantly, the most used method for HAR is the usage of camera-produced images, a resource exhaustive method that is still plagues with many obstacles. Nonetheless, several other approaches to HAR were explored over the years, One of which is the use of heart rate measurements, where the rhythmic patterns can be indicative of the type of physical activity. Electrocardiography is a common method of measuring heart rates by detecting electrical changes resulting from cardiac cycles. The heart rate rhythm is often graphed in terms of voltage and time in an electrocardiogram (ECG). ECG is a widely employed method, due to its high accuracy and clear depiction of heart rhythm signals [2]. A significant proportion of research went into the utilization of ECG signals to train machine learning models capable of automated detection. For instance, Jia et al. of the University of Science and Technology of China conducted a study on the recognition of human daily activity, using multi-lead ECG data, infused with an accelerometer for performance enhancement [3]. The dataset used was prepared using a wearable battery-powered ECG sensor with a sampling range of 200 Hz and a resolution of 8 bits, as for the acceleration, sensor B-MA180 chip was used. The features were extracted from ECG data using principal

* Corresponding author at: School of Robotics, XJTLU Entrepreneur College (Taicang) Xi'an Jiaotong-Liverpool University, Suzhou. 215123, P. R. China.

E-mail address: anwarmajeed1983@gmail.com (A.P.P. Abdul Majeed).

Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2022.03.006

2405-9595/

component analysis and Hermite polynomial expansion, before dimensional reduction via Linear Discriminant Analysis. Using Relevance Vector Machine (RVM) as the classification algorithm in the study, The model produced 97.78% accuracy when all of the ECG leads were used and a 99.67% accuracy with accelerometer infusion.

ECG signals show extremely impressive potential in HAR; nonetheless, it requires expensive machines that are normally large in size, rendering it quite notably unproductive for such a task as HAR. Photoplethysmography (PPG) is a viable alternative for the measurement of heart rate and cardiovascular rhythms, through the use of optical sensors that detect the change in light absorbance rate of vascular tissue as blood volume changes with every cardia cycle [4]. This method, contrary to Electrocardiography, does not require any expensive machines or tools, as it requires a very simple oximeter sensor that could be found in modern smartphones and smartwatches.

Some attempts at human activity recognition by means of machine learning using PPG data were seen in the last few years. One of which was conducted by Mahmud et al. in 2020, in a research paper that tested the use of multi-stage long short-term memory (LSTM) architecture [5]. In the study, the authors used LSTM as a method of temporal feature extraction, where the temporal features were extracted from PPG data as well as Inertial Measurements taken from a gyroscope, and different accelerometer sensors. The classification of said features was done by utilizing a Softmax-activated neural network of LSTM and fully connected layers yielding an impressive average accuracy of 83.2% using that combination of sensorial data. However, it is worthy of noting that the classification accuracy when PPG data were used alone was only 72.3%.

Another research work whose primary focus was the use of PPG data for HAR purposes was published by Brophy et al. [6]. The study involved the use of a machine-vision methodology to HAR via a pre-trained CNN model, known as Google's Inception-v3. The data used was from the Wrist PPG During Exercise public dataset, and was processed by plotting every 8 s of the signal into traditional signal graphs using MATLAB software. After being trained for 10000 steps, the model achieved a classification accuracy of 75.8% a relatively higher accuracy compared to other papers that used PGG data alone, as noted by the authors in the study.

Whereas PPG is more practical as its data could be acquired from any of the widely available wearable devices in the market, further facilitating the accessibility of HAR applications to regular consumers, it is often plagued by noise and artifacts compared to ECG signals that are often cleaner, and more accurate in their sampling. The aim of this study is to overcome this limitation by using a computer-vision-based transfer learning ensembled convolutional neural network model, a novel technique that makes use of the full strength of different pre-trained deep learning models to offer a great boost in classification performance while keeping the training period to a minimum.

2. Methodology

2.1. Data acquisition and preparation

The dataset used in this study was the publicly available dataset Wrist PPG During Exercise prepared by Casson et al. in 2017 [7]. As explained in the accompanying paper, the data was recorded from a sample of 8 individuals, 5 of whom were female while the rest were male, with an average age of 26. PPG records were taken via a sensor placed at the wrist of the subjects, which as well recorded Inertial measurements (gyroscope, a low noise accelerometer, a wide range accelerometer, and a magnetometer). Chest ECG signals were also recorded for comparison's sake simultaneously. The exercises performed by the subjects consisted of four types of exercise in a controlled setting using a treadmill and an exercise bike: running, walking, and high/low resistance cycling. The data was sampled at a frequency of 256 Hz for up to 10 min. There was minimal filtering conducted on the dataset; only the cycling recordings were filtered using a low-pass filter with a cut-off frequency of 15, as they exhibited noises unobserved in the treadmill signals.

Since the length of each signal is quite large, the signals were used in parts, where they were cut into smaller signals with an interval of 6 s, segmented every second and a half. The intention was quick detection, as anything beyond 6 s might cause consumers to become weary of the process, not to mention that any increase in the interval might cause information loss, considering the processed image is 224*224 in size (the recommended size for the used transfer learning models). As segmentation was performed, the total number of images became 4461 images, each class consisting of 846, 1166, 980, and 1469 images for high, low resistance cycling, running, and walking, respectively. To create a fair comparison to the literature reviewed at the beginning of this paper, as well as not to divert focus from the aim of this study in exploring the enhanced performance that pre-trained ensembled CNNs provide, the data was left as is with no attempts to resolving the imbalance in the data classes.

It could be noted that the dataset involves only four classes, hence excluding many of the variety of activities performed daily by most individuals. Nevertheless, the dataset seemed to offer a great challenge for classification research due to the amount of noise in the PPG data, which could be observed in the literature reviewed in the introduction of this paper. Not to mention, its utilization of ECG as a method for comparison, and since the purpose of this study is the illustration of the advantages that the ensemble method could provide when dealing with PPG HAR data, it was deemed suitable regardless of the number of classes it provides.

2.2. Signal representation approach

To use computer vision pre-trained convolutional neural network models, the signals needed to be presented visually. Plenty of techniques exist capable of fulfilling such a task in the field of signal processing, one of which is continuous

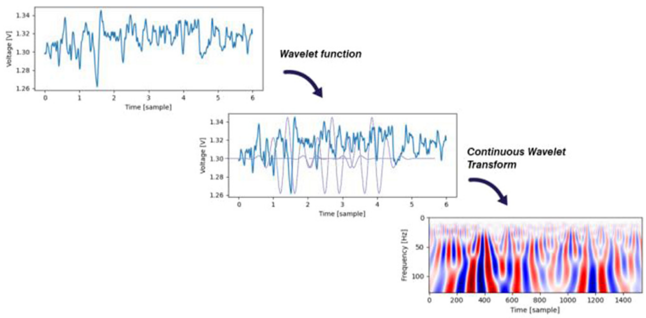

Fig. 1. Process of CWT.

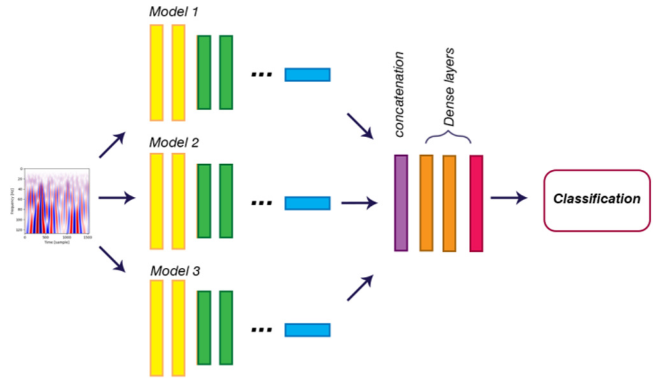

Fig. 2. Illustration of the structure of the ensemble model.

wavelet transform (CWT). CWT is a mathematical tool that generates a spectral representation of a signal by means of decomposing it into wavelets that vary in scale and translation. The translation and scale of the transformation are often expressed in the form of a scale factor that either dilates or compresses the signal, by applying high or low scale factors respectively [8]. Information extraction occurs when a specific wavelet function is multiplied by the signal as the wavelet shifts and stretches across the signal, allowing for a multi-scaled depiction of the signal. For the sake of this study, CWT was used to represent the signals optically, as it is a method that is not affected immensely by noise. The wavelet function used was Morlet, and the widths transform was 128. Below in Fig. 1 is an example of CWT representation of a PPG signal, placed for illustration purposes.

2.3. Pretrained convolutional neural networks and the ensemble method

Classification is a heavily researched field in machine learning, with many algorithms and new architectures appearing almost every day. Recently, Convolution neural networks (CNNs) saw a rapid spike in popularity, in the field of computer vision and image classification in specific, due to their robust architecture that allowed for impressively high classification accuracy with little to no data pre-processing [9]. This was only possible owing to the recent advancements in processing technologies, as CNNs are resource-exhaustive compared to older algorithms such as Random forests. This flaw in CNNs and deep learning, in general, spawned the need for a new approach to deep learning that shortens the time of training while keeping the high performative nature of CNN-based deep learning. One of these approaches is transfer learning, a technique used to transfer knowledge from pre-trained models to newer models with the aim of quickening the process of training, while achieving better accuracy [10]. Over the years, many publicly available pre-trained models found their way to existence, thanks to the efforts of multiple developers. Resnet50V2, MobileNetV2, and Xception are some of these models and the ones chosen for this study, all trained on the Imagenet dataset, which contains more than 14 million images. The weights of the models are transferred upon the process of training, before being re-trained on the dataset in use, this accelerates the process of training greatly, as it removes the need for complete rebuild from scratch [11]. All three models exhibit great performances, and are used worldwide for all types of applications in image recognition.

Just like transfer learning, there exist many other techniques that can even further improve the predictability of models. The ensemble method is one of these techniques, where multiple models are used at the same time to make inferences. This is achieved by using the features with the highest probabilities out of the combined models, using a layer often referred to as a Concatenate layer, hence giving a higher performance than that of an individual model [12]. The ensemble model can be used with any machine learning algorithm, but in this study, it was used to combine the three pre-trained CNN models mentioned above. Resnet50V2, MobileNetV2, and Xception were used together to make the model, using the aforementioned concatenate layer, which takes multiple sensors as input and produces a singular tensor as output. The output of this layer is either passed to more layers or is used directly for classification as shown in Fig. 2.

2.4. Experiment setup

The experiment conducted in this study was set up using a computer with an Nvidia GeForce® GTX 1650 Ti with 4 GB dedicated RAM, and an AMD Ryzen™ 7 4800H microprocessor. The experiment code was written in the programming language Python, a popular language for machine learning software development. The signals were processed mainly using the open-source python package PyWavelets, which was developed to ease the use of wavelet algorithms. After producing the CWT images, they were resized into a tensor of 224 × 224 × 3 and shuffled before being divided into a training and testing set with a ratio of 4:1. The images were then trained on the models which were connected to multiple dense layers with 512, 256, 128, 64 neurons activated with a ReLu activation function, and a dropout layer of 20%, all used to prevent overfitting and boost the performance of the algorithm [13]. The classifier used was a Softmax regression classifier. The same layers were used for all the compared algorithms for a fair comparison. The model was finally compiled using the optimizing algorithm Adam, with a learning

Table 1
Classification Accuracy, Loss and AUC results of ECG data.

		Resnet50V2	MobileNetV2	Xception	Ensemble
Training	Accuracy	95.19%	93.30%	95.22%	95.31%
	Loss	0.7861	0.8161	0.7864	0.1762
	AUC	0.9919	0.9833	0.9883	0.9914
Testing	Accuracy	91.25%	90.58%	93.16%	94.28%
	Loss	0.8315	0.8367	0.8093	0.2265
	AUC	0.9742	0.9666	0.9763	0.9848

Table 2
Classification Accuracy, Loss and AUC results of PPG data.

		Resnet50V2	MobileNetV2	Xception	Ensemble
Training	Accuracy	96.46%	87.53%	92.60%	94.51%
	Loss	0.7870	0.8766	0.8286	0.2157
	AUC	0.9945	0.9688	0.9825	0.9918
Testing	Accuracy	82.53%	83.76%	86.11%	88.91%
	Loss	0.9164	0.9022	0.8851	0.3502
	AUC	0.9945	0.9548	0.9597	0.9770

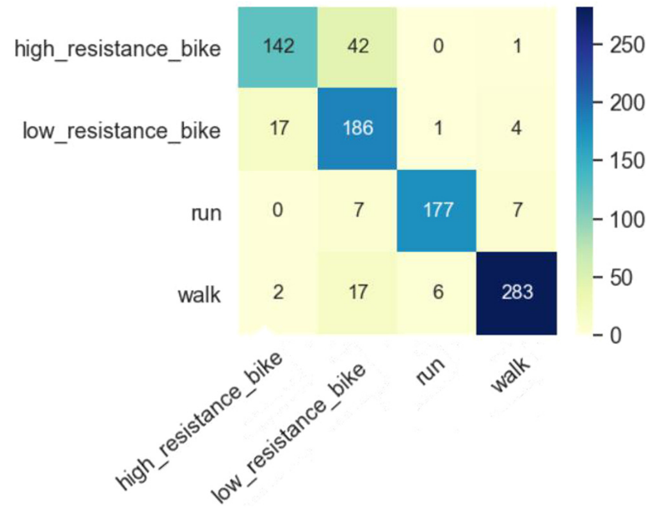

Fig. 3. Confusion matrix of the ensemble model.

rate of 0.00001. Training commenced with a batch size of 16 and continued for as many training epochs before being stopped with early stopping with patience of 3 epochs. The program to used to train the models was developed with the aid of multiple Python packages, namely: Tensorflow, Keras, SkLearn, Numpy, Matplotlib, and OpenCV [14–16].

3. Results and discussion

It is a justifiable tradition in machine learning to utilize a set of performance metrics in order to adequately observe and judge the performance of the created models. In this study, in particular, several performance metrics were used; the accuracy of prediction indicative of the number of correct predictions made by the models in question, the Cross-entropy loss which decreases as the prediction probability is closer to the actual label, and lastly the Area Under the Receiver Operating Characteristics (ROC) curve, often referred to as the Area Under the Curve (AUC) which measures the capability of the model to distinguish between classes. In Tables 1 and 2 the performance of all the models trained on both ECG and PPG data is depicted, respectively, in terms of the aforementioned performance metrics.

The experiment results as illustrated in the two tables demonstrate the improved performance that the use of the ensemble method provides. As can be noted in Table 2, when PPG was used to train the ensemble model there was a 2.8% improvement in classification accuracy(CA), compared to the best performing model (Xception), and a 4.8% improvement compared to the average accuracy of all three models. It is important to note that Resnet50V2 performs better during training; nevertheless, this paper is mostly concerned with the testing accuracy due to the fact that it is a better indictive of the model's ability to perform in practical applications, as it classifies data that was not presented to it before. The same trend of enhanced classification performance can be seen reflected on the AUC and Loss (cost function), with a decrease of 0.5249 and an increase of 0.0173, respectively, compared to the best performing individual model. These results in comparison to prior work done on the same dataset, seen in Brophy et al.'s [6] and Mahmud et al.'s [5] studies, showed much improved performance, as both studies produced CA of 75.8% and 72.3% respectively. This shows the great potential of ensembled transferred models at enhancing the process of HAR with PPG data.

In Table 1, the results of the performance of the ECG data are depicted, showing a difference of about 5.4% classification accuracy between the ensemble models. This can be explained by considering the accuracy that ECG machines can provide in comparison to PPG sensors which are widely available at a much lower cost. Despite this expected difference in performance, a difference of 5.4% with an accuracy of 88.91% is relatively low, especially in comparison to previous literature. PPG data after all can be acquired easily using smartphones and smart wristbands, which could open the door for applications that are hard to implement with ECG sensors, considering the size and price.

While the previously depicted metrics of evaluation provide adequate evaluation of the models performance, confusion matrices can provide even further information into the model's capability to accurately classify data into their relative classes. Fig. 3 shows the confusion matrix of the ensemble model when used with PPG data, the focus of this study, Normally, the diagonal values are indicative of the accuracy of classification, and it can be seen that the values of the ensemble model are concentrated diagonally which indicates a good classification accuracy.

It is worth noting that the data sampled for this study was small, with a sample of only 8 subjects. We suspect that an increase in the size of the dataset, can improve the performance of the models even further. The use of different classifiers and larger more complex pre-trained models, can also prove further enhancement if used in future research.

4. Conclusion

While ECG data is commonly used for human activity recognition, PPG can be an inexpensive alternative. This study aims at using a computer-vision approach at detecting human activities using PPG signals, with the aid of the ensemble method used to combine 3 pre-trained models; Resnet50V2, MobileNetV2, and Xception. The ensemble model under controlled conditions was compared to each of its composing models, trained on both ECG and PPG data of heart rate recorded during different activities. Although the models when trained on ECG data performed better, The performance of PPG gave a relatively high accuracy at 88.91% indicative of the potential of using wrist PPG data with the ensemble method in HAR for accurate detection. The field of machine learning is very vast and there is spacious room for further studies on HAR via PPG data; the use of the method suggested in this research on different datasets with more human activities is suggested as it poses an interesting test of the method's full capacity. Furthermore, future research could compare different preprocessing methods, and hyperparameter tuning techniques as they seem to have the potential to improve the performance of the models even further.

CRediT authorship contribution statement

Omair Rashed Abdulwareth Almanifi: Conceptualization, Writing – original draft, Methodology, Software, Investigation. **Ismail Mohd Khairuddin:** Data curation, Software. **Mohd Azraai Mohd Razman:** Methodology, Validation. **Rabiu Muazu Musa:** Visualization, Formal analysis. **Anwar P.P. Abdul Majeed:** Conceptualization, Supervision, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to give a special thanks to the providers of the dataset; Wrist PPG During Exercise [7].

References

[1] J.K. Aggarwal, M.S. Ryoo, Human activity analysis: A review, ACM Comput. Surv. 43 (2011) 1–43, http://dx.doi.org/10.1145/1922649.1922653.

[2] D.B. Geselowitz, On the theory of the electrocardiogram, Proc. IEEE. 77 (1989) 857–876, http://dx.doi.org/10.1109/5.29327.

[3] R. Jia, B. Liu, Human daily activity recognition by fusing accelerometer and multi-lead ECG data, in: 2013 IEEE Int. Conf. Signal Process. Commun. Comput., ICSPCC 2013, 2013, http://dx.doi.org/10.1109/ICSPCC.2013.6664056.

[4] A.A. Alian, K.H. Shelley, Photoplethysmography: Analysis of the pulse oximeter waveform, in: Monit. Technol. Acute Care Environ., Springer, New York, 2014, pp. 165–178, http://dx.doi.org/10.1007/978-1-4614-8557-5_19.

[5] T. Mahmud, S.S. Akash, S.A. Fattah, W.P. Zhu, M.O. Ahmad, Human activity recognition from multi-modal wearable sensor data using deep multi-stage LSTM architecture based on temporal feature aggregation, in: Midwest Symp. Circuits Syst, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 249–252, http://dx.doi.org/10.1109/MWSCAS48704.2020.9184666.

[6] E. Brophy, J.J. Dominguez, Z. Wang, T.E. Ward, A machine vision approach to human activity recognition using photoplethysmograph sensor data, in: 29th Irish Signals Syst. Conf., ISSC 2018, Institute of Electrical and Electronics Engineers Inc., 2018, http://dx.doi.org/10.1109/ISSC.2018.8585372.

[7] D. Jarchi, A.J. Casson, Description of a database containing wrist PPG signals recorded during physical exercise with both accelerometer and gyroscope measures of motion, Data 2 (2017) http://dx.doi.org/10.3390/data2010001.

[8] Stéphane Mallat, A Wavelet Tour of Signal Processing, second ed., Elsevier, 1999, http://dx.doi.org/10.1016/b978-0-12-466606-1.x5000-4.

[9] S. Albawi, T.A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network, in: Proc. 2017 Int. Conf. Eng. Technol., ICET 2017, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–6, http://dx.doi.org/10.1109/ICEngTechnol.2017.8308186.

[10] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345–1359, http://dx.doi.org/10.1109/TKDE.2009.191.

[11] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345–1359, http://dx.doi.org/10.1109/TKDE.2009.191.

[12] O. Sagi, L. Rokach, Ensemble learning: A survey, wiley interdiscip, Rev. Data Min. Knowl. Discov. 8 (2018) e1249, http://dx.doi.org/10.1002/widm.1249.

[13] A. Pauls, Determining optimum drop-out rate for neural networks, Midwest Instr. Comput. Symp. (2018) 0–11.

[14] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, TensorFlow: A system for large-scale machine learning, in: Proc. 12th USENIX Symp. Oper. Syst. Des. Implementation, OSDI 2016, 2016.

[15] F. Chollet, Keras documentation, keras.io, 2015.

[16] Numpy, numpy — Numpy, numpy website, 2017, http://dx.doi.org/10.1017/CBO9781107415324.004.