Physical Activity Recognition using Deep Transfer Learning with Convolutional Neural Networks

Berke Ataseven*
Mechanical Engineering
Koc University
Istanbul, Turkey
bataseven15@ku.edu.tr

Alireza Madani*
Mechanical Engineering
Koc University
Istanbul, Turkey
amadani20@ku.edu.tr

Beren Semiz
Electrical and Electronics Engineering
Koc University
Istanbul, Turkey
besemiz@ku.edu.tr

M. Emre Gursoy
Computer Engineering
Koc University
Istanbul, Turkey
emregursoy@ku.edu.tr

Abstract—Current wearable devices are capable of monitoring various health indicators as well as fitness and/or physical activity types. However, even on the latest models of many wearable devices, users need to manually enter the type of work-out or physical activity they are performing. In order to automate real-time physical activity recognition, in this study, we develop a deep transfer learning-based physical activity recognition framework using acceleration data acquired through inertial measurement units (IMUs). Towards this goal, we modify a pre-trained version of the GoogLeNet convolutional neural network and fine-tune it with data from IMUs. To make IMU data compatible with GoogLeNet, we propose three novel data transform approaches based on continuous wavelet transform: Horizontal Concatenation (HC), Acceleration-Magnitude (AM), and Pixelwise Axes-Averaging (PA). We evaluate the performance of our approaches using the real-world PAMAP2 dataset. The three approaches result in 0.93, 0.95 and 0.98 validation accuracy and 0.75, 0.85 and 0.91 test accuracy, respectively. The PA approach yields the highest weighted F1 score (0.91) and activity-specific true positive ratios. Overall, our methods and results show that accurate real-time physical activity recognition can be achieved using transfer learning and convolutional neural networks.

Index Terms—Convolutional Neural Networks, Transfer Learning, Wavelet Analysis, Physical Activity Recognition

I. INTRODUCTION

Advancements in wearable sensor technology have made it feasible to integrate many sensor modalities into electronic devices, which enable unobtrusive and continuous acquisition of various physiological signals and parameters [1]. Once analyzed through appropriate algorithms, these signals can be converted into actionable metrics for achieving non-invasive health monitoring [2]. Indeed, there has been a leap forward towards adding health monitoring functionalities into wearable devices, such as smart bracelets and smart watches. With current smart devices available in the market, users can now assess physiological parameters such as heart rate, irregular rhythm, respiration rate and blood oxygen saturation [3]–[5].

In addition to monitoring clinical data such as pulmonary and cardiovascular indicators, another eminent purpose of wearable devices is to track fitness, physical activity, and energy consumption. Nevertheless, one of the present challenges

*The first two authors contributed equally.

is that even on the latest models of these devices, users need to manually enter the type of work-out or physical activity they are performing. Based on the user's entry, the device can then use the associated pre-trained models to assess fitness performance and energy/calorie expenditure [6]. Although there have been several studies focusing on physical activity assessment using single and/or multiple accelerometers attached on wrist, hip, ankle and foot [7]–[12], these frameworks are not suitable for automatic and real-time activity monitoring. Therefore, achieving automatic and real-time physical activity recognition remains as an open research question.

In recent years, deep learning methods, in particular Convolutional Neural Networks (CNNs), have achieved remarkable success in many machine learning problems [13], [14]. However, it is oftentimes difficult to train large CNNs from scratch since they are data-hungry (i.e., they require a large number of labeled training samples) and resource-hungry (i.e., training large CNNs can take multiple days even on expensive hardware). Therefore, *transfer learning* has become an effective strategy to leverage the power of pretrained CNN models. Transfer learning enables the transfer of knowledge gained in solving one problem towards solving another problem. For example, consider a supervised machine learning problem A, labeled training dataset $\mathcal{D}_A$, and CNN model $\mathcal{M}_A$ that was trained to solve A using $\mathcal{D}_A$. Now consider a different but related problem B such that the size of the dataset $\mathcal{D}_B$ is limited. Instead of training a CNN model $\mathcal{M}_B$ from scratch, using transfer learning, model $\mathcal{M}_A$ is taken and fine-tuned using $\mathcal{D}_B$ to obtain an accurate model to solve B.

In recent literature, deep transfer learning was used in several contexts related to health and medical systems, such as motor imagery electroencephalogram classification [15], multichannel sleep stage classification [16], driver fatigue detection [17], pulmonary diagnostics and infection detection [18]. Inspired by these applications, in this study our goal is to develop a transfer learning-based real-time physical activity recognition framework, which can recognize subjects' physical activity in real-time using acceleration data acquired through inertial measurement units (IMUs). Towards this goal, we leveraged a pre-trained version of the GoogLeNet architecture [19], and modified and fine-tuned it with data originating from IMUs. To make IMU data compatible with GoogLeNet, we

developed three novel continuous wavelet transform (CWT)-based data transform approaches: Horizontal Concatenation (HC), Acceleration-Magnitude (AM), and Pixelwise Axes-Averaging (PA). HC concatenates data from two IMUs (wrist and chest). AM enables the usage of all three IMUs (wrist, chest, ankle) by combining the magnitudes of X, Y, Z acceleration axes. Finally, PA enables the usage of all three IMUs by preserving the frequency information through axes averaging rather than computing magnitudes. Experiment results with real-world data show that, in terms of accuracy, PA > AM > HC. In particular, PA achieves a validation accuracy of 98% and test accuracy of 91%.

The rest of this paper is organized as follows: Section II-A describes our dataset and physical activity recognition problem. Section II-B discusses the preliminaries of CWT. Sections II-C, II-D, and II-E introduce our novel CWT-based data transform approaches. Section II-F presents the details of our transfer learning and fine-tuning steps. Section III contains our experimental evaluation and discussion. Finally, Section IV concludes the paper.

II. METHODS

A. Dataset and Problem Description

In this study, we worked with the publicly available PAMAP2 dataset, which includes data from 9 healthy subjects [20], [21]. Measurements were acquired at 100 Hz with Colibri wireless inertial measurement units (IMUs) mounted on the chest, the wrist of the dominant arm, and the ankle of the dominant leg. We used the tri-axial acceleration data recorded from these three IMUs. NaN values originating from wireless data drop were removed using the intra-extrapolator provided by John D'Errico [22]. A representative tri-axial acceleration data from the wrist IMU is presented in Figure 1a.

Throughout the experimental protocol, participants were asked to perform the following physical activities:

- **Lying:** Lying while doing nothing (small movements were allowed).
- **Walking:** Walking outside with with a speed of 4-6km/h, depending on the subject's comfort.
- **Running:** Jogging outside with a speed suitable for the subject.
- **Cycling:** Cycling outside with a speed suitable for the subject as if the subject is riding to work or cycle for pleasure (but not as a sport activity).
- **Nordic Walking:** Walking outside on asphaltic terrain, using walking poles having asphalt pads.
- **Ascending Stairs:** Ascending a distance of at least five floors in a building between the ground and the top floors.
- **Descending Stairs:** Descending a distance of at least five floors in a building between the top and the ground floors.
- **Ironing:** Ironing one or two shirts or t-shirts.
- **Rope Jumping:** Performing basic jump or alternating foot jump, whichever was more suitable for the subject.

Given acceleration data recorded from the aforementioned IMUs, our task is to automatically recognize which physical activity is being performed by the subject in real-time. In order to simulate real-time data streaming, we work with sliding time windows rather than using the time-series signals as a whole. Information about window length and stride size used in each approach is detailed in Sections II-C, II-D and II-E.

B. Continuous Wavelet Transform (CWT)

The first step of our framework is to use Continuous Wavelet Transform (CWT) towards decomposing the IMU acceleration data and obtaining a 2D coefficient matrix. CWT decomposes a signal into a set of basis functions using a finite-length function called the *mother wavelet* (Ψ) [23]. Depending on the application, different mother wavelets can be chosen. The width and central frequency of Ψ can be changed by moving it across the signal. These scaled and shifted versions of Ψ are called *daughter wavelets*, and can be combined to represent the original signal, as shown in Equation 1.

$$F(\tau, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} f(t) \cdot \Psi^*(\frac{t - \tau}{s}) dt \qquad (1)$$

Here, s is the scaling factor (compression or dilation), τ is the translation factor (time shift), $f(t)$ is the signal being analyzed, and Ψ^* is the shifted and scaled mother wavelet. The output $F(\tau, s)$ denotes the corresponding wavelet coefficients.

In our study, we first applied CWT on each signal window and obtained a complex-valued 2D coefficient matrix. We then computed the magnitude of this matrix by multiplying it with its complex conjugate. The resulting matrix can be treated as one channel of an image, i.e., each coefficient in the matrix represents one pixel in the image, and the value of the coefficient determines the color intensity of the corresponding pixel. In case of a single channel, the resulting image is gray-scale (low vs high intensity corresponds to blackness vs whiteness of the pixel). In case of colored (RGB) images, there exist 3 channels: Red (R), Green (G), Blue (B); each of which can be populated by a different coefficient matrix. From this point onwards, we will refer to the consecutive processes of computing the 2D coefficient matrix and calculating its magnitude as CWT.

C. Horizontal Concatenation (HC) Approach

We developed three approaches for physical activity recognition, all of which leverage CWT and transfer learning. The first approach is called Horizontal Concatenation (HC). In this approach, tri-axial (X, Y and Z) acceleration signals from only the chest and wrist IMUs are used; the ankle IMU is not used. The window length was determined as 112. This is equal to half of the required input image width for GoogLeNet (224), which is the CNN we used for transfer learning. The stride size was determined as 1/3 of the window length. The subsequent processing steps performed by the HC approach are as follows:

1) Each of the X, Y, and Z axes windows were converted into 2D coefficient matrices using CWT.
2) Resulting X, Y, and Z axes matrices were treated as the red, green, and blue channels of an image respectively.
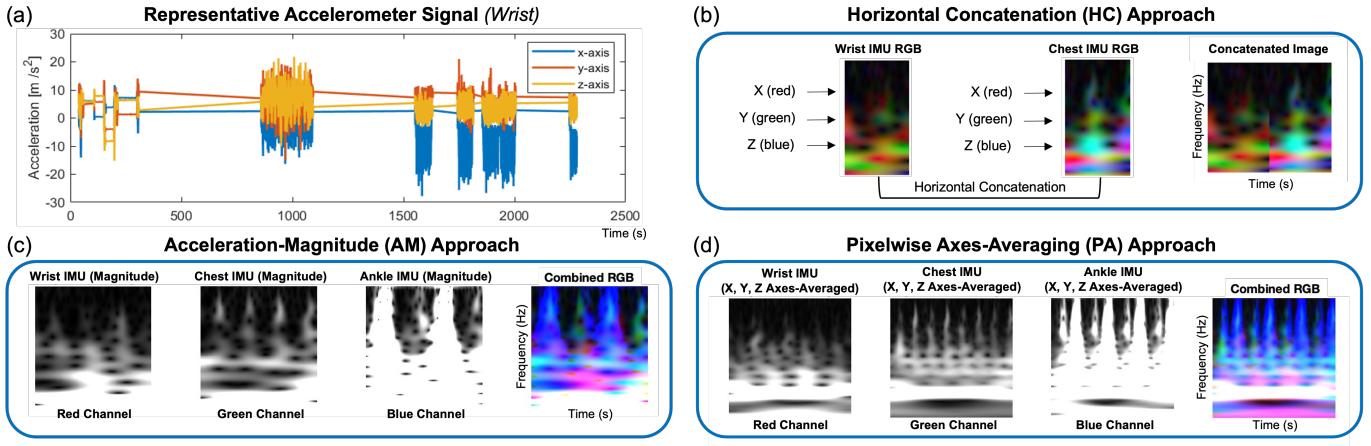
Fig. 1. (a) A representative tri-axial acceleration data from the wrist IMU. (b) Horizontal concatenation approach using the wrist and chest IMUs. (c) Acceleration-magnitude approach using all three IMUs. (d) Pixelwise axes-averaging approach using all three IMUs.

The three channels were combined to construct an RGB image per window, as shown in Figure 1b.

3) Per each window, the RGB image from the wrist IMU and the RGB image from the chest IMU were concatenated horizontally to obtain a 224-by-224 image.
4) Resulting 224-by-224 RGB images were used to train, validate and test the CNN.

D. Acceleration-Magnitude (AM) Approach

Our second approach is called Acceleration-Magnitude (AM). In this approach, tri-axial (X, Y and Z) acceleration signals from all three IMUs (chest, wrist and ankle) are used. The window length was determined as 224, which is directly equal to the input image width required by GoogLeNet. Similar to the HC approach, stride size was determined as 1/3 of the chosen window length. The subsequent processing steps performed by the AM approach are as follows:

1) For each IMU sensor, the acceleration magnitude of the current window was computed in L_2 norm using Equation 2. Let X_i, Y_i and Z_i denote the X, Y, Z signals from the i'th window. The acceleration magnitude of the i'th window, denoted by $\|m\|_i$, is:

$$\|m\|_i = \sqrt{X_i^2 + Y_i^2 + Z_i^2} \qquad (2)$$

2) CWT was applied separately on the magnitude signals $\|m\|_i$ obtained from the chest, wrist, and ankle IMUs.
3) Resulting wavelet coefficient matrices were obtained from the three IMUs. Treating the wrist IMU's matrix as the red channel, chest IMU's matrix as the green channel, and ankle IMU's matrix as the blue channel, an RGB image was constructed (224-by-224). A sample image is illustrated in Figure 1c.
4) Resulting 224-by-224 RGB images were used to train, validate and test the CNN.

E. Pixelwise Axes-Averaging (PA) Approach

Our third approach is called Pixelwise Axes-Averaging (PA). In this approach, similar to the AM approach, tri-axial

(X, Y and Z) acceleration signals from all three IMUs (chest, wrist and ankle) are used. The window length was determined as 450 (equal to 4.5 seconds), which was shown to be appropriate and effective for activity recognition [24]. Similar to the HC and AM approaches, stride size was determined as 1/3 of the chosen window length. The subsequent processing steps performed by the PA approach are as follows:

1) Each of the X, Y and Z axes windows were converted into 2D coefficient matrices using CWT.
2) Let $M_X[j]$, $M_Y[j]$, and $M_Z[j]$ denote the j'th pixel in the coefficient matrix of the X, Y and Z axis respectively. The values of each pixel were averaged to construct a matrix M:

$$M[j] = \frac{M_X[j] + M_Y[j] + M_Z[j]}{3} \qquad (3)$$

3) The above step was repeated separately for all three IMUs, which resulted in: one matrix for wrist IMU, one matrix for chest IMU, one matrix for ankle IMU.
4) Matrices were resized to become 224-by-224, which is the required input size for the CNN model.
5) Resulting matrices were used to construct an RGB image per window, by treating the wrist IMU's matrix as the red channel, chest IMU's matrix as the green channel, and ankle IMU's matrix as the blue channel. A sample image is illustrated in Figure 1d.
6) Resulting 224-by-224 RGB images were used to train, validate and test the CNN.

F. Transfer Learning with CNNs

In all three approaches (HC, AM, PA), the output datasets consist of RGB images. These datasets can be used to train deep learning models for physical activity recognition. However, to avoid training a deep CNN from scratch with limited data and to leverage the power of pretrained models, we used transfer learning. In particular, we applied transfer learning with GoogLeNet as our base (starting) model, which is a popular, 22 layers deep CNN [19]. We used a pretrained

	Validation Accuracy	Test Accuracy
Horizontal Concatenation	0.93	0.75
Acceleration-Magnitude	0.95	0.85
Pixelwise Axes-Averaging	0.98	0.91

TABLE I
VALIDATION AND TEST ACCURACY RESULTS

version of GoogLeNet that was trained using the ImageNet image classification dataset [25]. This CNN originally takes as input images of size 224-by-224 and classifies them into 1000 object categories.

In order to transfer this CNN to the physical activity recognition problem, we performed the following. We discarded the last 3 layers of the GoogLeNet architecture, which correspond to a fully connected layer, a softmax layer, and an output layer with 1000 output classes. We replaced them by a fresh fully connected layer (with output size = 9), a softmax layer, and an output layer with 9 classes and cross entropy loss function. We then fine-tuned the whole CNN by updating all layers' parameters using the RGB image dataset constructed by our proposed approaches (HC, AM or PA). Following the intuition that earlier layers of a CNN contain more generic features whereas the later layers become progressively more specialized towards the task in hand, we used a higher learning rate in the layers that have been freshly added. More specifically, in order to learn the parameters of the newly added fully connected layer faster than the rest of the CNN (which has been pretrained), we set the learning rate factor for this layer to be equal to 5 times the learning rate of the rest of the CNN. This applied to both the weight and bias parameters.

III. EXPERIMENT RESULTS AND DISCUSSION

In this section, we experimentally evaluate each of the three proposed approaches (HC, AM, PA) using the PAMAP2 dataset and discuss the evaluation results.

A. Experiment Setup

The training, validation and test datasets for the three approaches (HC, AM, PA) were constructed separately, i.e., RGB images from two different approaches do not co-exist in any dataset. Recall that the PAMAP2 dataset contains data from 9 healthy subjects. In each approach, data from 2 subjects were left out for testing. Data from the remaining 7 subjects were used for constructing the training and validation sets. We randomly shuffled the data from these 7 subjects and used 80% for training and 20% for validation.

Hyperparameters for fine-tuning the GoogLeNet CNN were set as follows. The Adam solver was used as the optimization algorithm. Initial learning rate was set to 0.0001, validation frequency was set to 50, number of epochs was set to 10, and minibatch size was set to 64. Models were trained on an Intel i7 CPU. Cross entropy was used as the loss function. The final CNN was chosen as the network that had the best validation loss throughout the iterations of fine-tuning.

B. Model Performance Results

After CNN models are trained, their accuracy on the validation and test sets are measured. As reported in Table I, we

obtained validation accuracy values of 0.93, 0.95, and 0.98 for the Horizontal Concatenation (HC), Acceleration-Magnitude (AM) and Pixelwise Axes-Averaging (PA) approaches, respectively. In contrast, test accuracy values for the three approaches were 0.75, 0.85, and 0.91 respectively. It is expected that the test accuracy values are lower than validation accuracy, since the training and validation sets included data from 7 subjects whereas 2 subjects were left out. Since data from those 2 subjects are included in the test dataset, test accuracy suffers from inter-subject variability. Nevertheless, accuracy values remain competitive in general, considering that the accuracy of random guessing is only 11%. In the rest of this section, we evaluate and discuss each approach in more detail.

1) Horizontal Concatenation (HC) Approach: Compared to the AM and PA approaches, the HC approach has lower accuracy since it uses data only from two IMUs (ankle IMU is not used by the HC approach). Window length being equal to 112 samples (1.12 seconds) was also a factor that contributed the lower performance of HC, considering that 1.12 seconds of data from chest and wrist IMUs may sometimes not be sufficient to capture unique information needed to distinguish between different physical activities.

In Table II, we report the precision, recall and F1 score of the HC approach for each different physical activity. Across all activities, the weighted precision, recall and F1 scores were found to be 0.80, 0.75 and 0.75, respectively. We observe that the model performs well for many activities (F1 score ≥ 0.8), but the F1 scores are relatively low for three activities in particular: walking, ascending stairs, and descending stairs. The F1 scores for these three activities are 0.58, 0.48 and 0.34. The reason behind this was further investigated using the confusion matrix reported in Figure 2. The confusion matrix shows that the data from ascending stairs, descending stairs, Nordic walking and walking activities are typically the ones that are confused with one another by the model. Notably, these activities are all inherently walking-type activities, which result in similar chest and wrist acceleration values. Therefore, distinguishing between them was difficult using only the chest and wrist IMUs. Since the ankle IMU carries information regarding the walking style of the subject (upward, downward, with walking pole, etc.) and/or frequency, AM and PA approaches that leverage the ankle IMU perform much better than HC in terms of the aforementioned activities.

2) Acceleration-Magnitude (AM) Approach: We report the precision, recall and F1 score of the AM approach for each different physical activity in Table III. Overall, using the AM approach, we obtained a weighted precision, recall and F1 score of 0.87, 0.85 and 0.85, respectively. The AM approach performs much better than the HC approach in terms of the walking activity, considering that the F1 score for the walking activity is similar to the other activities. Results for the ascending and descending stairs activities are also substantially better in AM compared to HC (F1 scores of 0.76 and 0.63 compared to 0.48 and 0.34); yet, these activities could not reach the F1 scores of the other activities. We also observed a small improvement in the recognition of the Nordic walking

	Precision	Recall	F1 Score
Lying	0.85	0.90	0.88
Walking	0.83	0.44	0.58
Running	0.93	0.93	0.93
Cycling	0.96	0.94	0.95
Nordic Walking	0.94	0.66	0.78
Ascending Stairs	0.34	0.82	0.48
Descending Stairs	0.52	0.25	0.34
Ironing	0.74	0.92	0.82
Rope Jumping	0.99	0.80	0.88
Unweighted Average	0.79	0.79	0.74
Weighted Average	0.80	0.75	0.75

TABLE II
RESULTS FOR THE HORIZONTAL CONCATENATION APPROACH

	Precision	Recall	F1 Score
Lying	0.92	0.87	0.89
Walking	0.91	0.85	0.88
Running	0.87	0.95	0.90
Cycling	0.99	0.94	0.96
Nordic Walking	0.99	0.74	0.85
Ascending Stairs	0.68	0.85	0.76
Descending Stairs	0.57	0.69	0.63
Ironing	0.78	0.95	0.86
Rope Jumping	0.97	0.72	0.82
Unweighted Average	0.85	0.84	0.84
Weighted Average	0.87	0.85	0.85

TABLE III
RESULTS FOR THE ACCELERATION-MAGNITUDE APPROACH
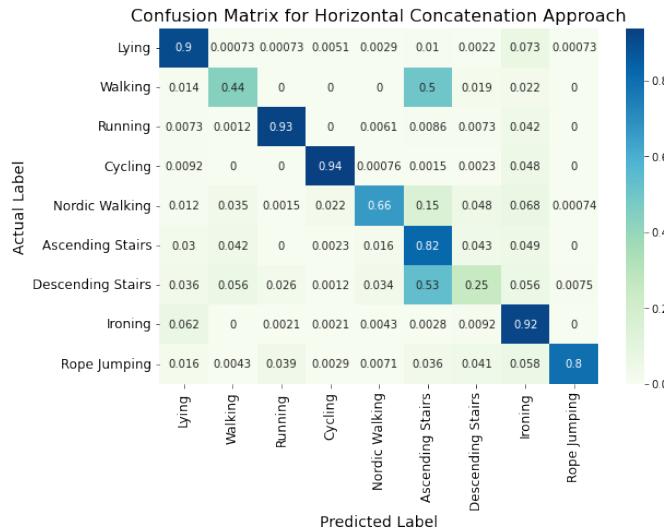
Fig. 2. Row-wise normalized confusion matrix for the HC approach.
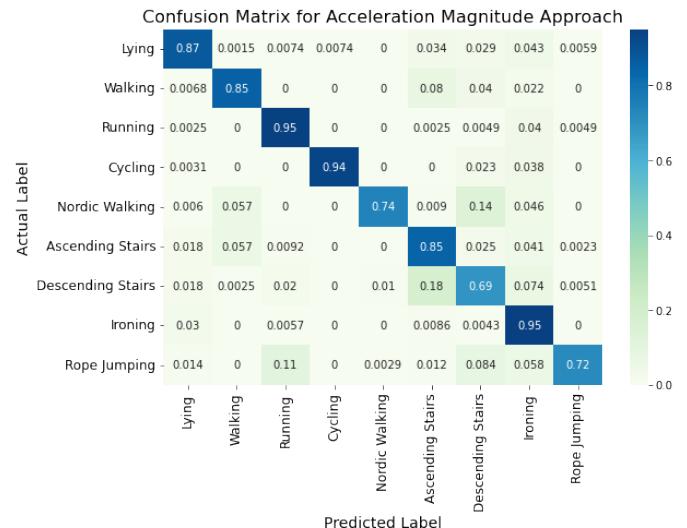
Fig. 3. Row-wise normalized confusion matrix for the AM approach.

activity relative to the HC approach. F1 score increased from 0.78 to 0.85, and normalized true positive ratio increased from 0.66 to 0.74 as shown in Figure 3. Overall, the most significant improvement was in the walking activity: F1 score increased from 0.58 to 0.88, and normalized true positive ratio increased from 0.44 to 0.85.

The above improvements achieved by the AM approach can be attributed to larger window length (224 samples, therefore 2.24 seconds) and the added value of data from the ankle IMU. However, since finding the magnitude involves squaring the signals with frequency contents, the AM approach still remains incapable of handling differences related to frequency. Thus, accuracy can be further improved by the PA approach.

3) Pixelwise Axes-Averaging (PA) Approach: We report the precision, recall and F1 score of the PA approach in Table IV. Overall, we obtained a weighted precision, recall and F1 score of 0.91, 0.91 and 0.91. These results show that the PA approach is indeed the best performing one compared to HC and AM. Notably, there are increases in the F1 scores and normalized true positive ratios of all activities in the PA approach compared to AM (also see Figure 4).

The PA approach performs exceptionally well for the running and ironing activities, in which normalized true positive ratios of 1.00 and 0.99 are achieved, meaning that the model performed almost perfectly in distinguishing these activities from the remaining ones. Activities that have relatively lower true positive ratios are Nordic walking, ascending stairs and descending stairs (0.77, 0.87 and 0.77, respectively). Indeed, all approaches (HC, AM, PA) have difficulty in recognizing these three activities. Nevertheless, we note that the F1 scores of ascending stairs and descending stairs activities are still higher in PA (0.79 and 0.78) compared to AM (0.76 and 0.63) and HC (0.48 and 0.34).

IV. CONCLUSION

In this paper, we developed three approaches for real-time activity recognition from IMU data using transfer learning. We leveraged a modified version of the well-known GoogLeNet architecture as our base model for transfer learning. As the original GoogLeNet CNN takes images of size 224-by-224 as input, we proposed three different CWT-based approaches to convert IMU time series into 2D matrices: Horizontal Concatenation (HC), Acceleration-Magnitude (AM), and Pixelwise Axes-Averaging (PA). We experimentally evaluated all three approaches using the real-world PAMAP2 dataset. Overall, the best performing approach was PA. AM performed slightly worse than PA since its computation of magnitude causes the loss of unique frequency information that PA benefits from.

	Precision	Recall	F1 Score
Lying	0.96	0.93	0.95
Walking	0.91	0.92	0.92
Running	0.97	1.00	0.98
Cycling	1.00	0.96	0.98
Nordic Walking	0.99	0.77	0.87
Ascending Stairs	0.72	0.87	0.79
Descending Stairs	0.78	0.77	0.78
Ironing	0.92	0.99	0.95
Rope Jumping	1.00	0.93	0.96
Unweighted Average	0.92	0.90	0.91
Weighted Average	0.91	0.91	0.91

TABLE IV
RESULTS FOR THE PIXELWISE AVERAGING APPROACH
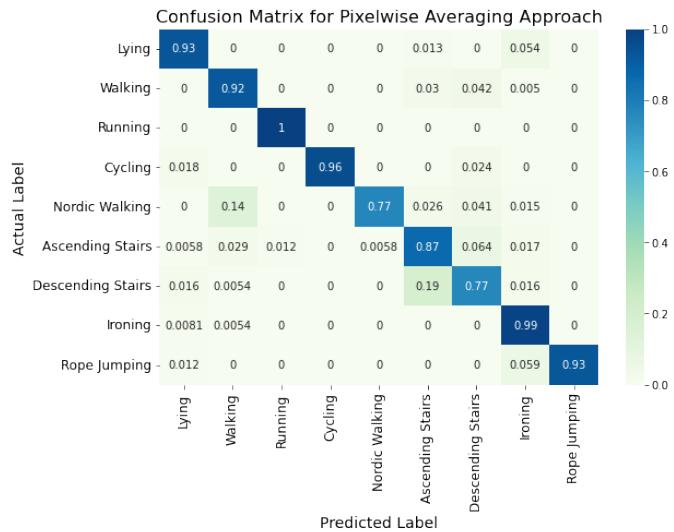
Fig. 4. Row-wise normalized confusion matrix for the PA approach.

HC performed worse than PA and AM since it only uses two IMUs and it has a smaller window length.

Among all three approaches, two common observations were that (i) it was challenging to distinguish between two activities in particular: ascending stairs and descending stairs, and (ii) inter-subject variability decreases the activity recognition accuracy of all three approaches, which can be observed by the difference between the validation accuracy and test accuracy values. Future work will include investigating additional pre-processing approaches to overcome these challenges. Additionally, we will test our approaches on larger datasets in order to better validate their generalizability.

REFERENCES

[1] B. Semiz, "Digital biomarker discovery for non-invasive health monitoring with acoustic and vibration signals," Ph.D. dissertation, Georgia Institute of Technology, 2020.

[2] T. Wang, T. Azad, and R. Rajan, "The emerging influence of digital biomarkers on healthcare," *Rock Health*, 2016.

[3] Apple, "Apple watch. helping your patients identify early warning signs." *https://www.apple.com/healthcare/apple-watch/*, 2022.

[4] Fitbit, "Tune in to your body with fitbit health metrics," *https://www.fitbit.com/global/us/technology/health-metrics*, 2022.

[5] Garmin, "Health science," *https://www.garmin.com.my/minisite/garmin-technology/health-science/*, 2022.

[6] Apple, "Use the workout app on your apple watch," *https://support.apple.com/en-us/HT204523*, 2022.

[7] F. Marin, K. Lepetit, L. Fradet, C. Hansen, and K. B. Mansour, "Using accelerations of single inertial measurement units to determine the intensity level of light-moderate-vigorous physical activities: Technical and mathematical considerations," *Journal of biomechanics*, vol. 107, p. 109834, 2020.

[8] K. Lu, L. Yang, F. Seoane, F. Abtahi, M. Forsman, and K. Lindecrantz, "Fusion of heart rate, respiration and motion measurements from a wearable sensor system to enhance energy expenditure estimation," *Sensors*, vol. 18, no. 9, p. 3092, 2018.

[9] B. Cvetković, R. Szeklicki, V. Janko, P. Lutomski, and M. Luštrek, "Real-time activity monitoring with a wristband and a smartphone," *Information Fusion*, vol. 43, pp. 77–93, 2018.

[10] B. Dong, S. Biswas, A. Montoye, and K. Pfeiffer, "Comparing metabolic energy expenditure estimation using wearable multi-sensor network and single accelerometer," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 2866–2869.

[11] F. Fotouhi-Ghazvini and S. Abbaspour, "Wearable wireless sensors for measuring calorie consumption," *Journal of Medical Signals and Sensors*, vol. 10, no. 1, p. 19, 2020.

[12] J. H. Zhang, D. J. Macfarlane, and T. Sobko, "Feasibility of a chest-worn accelerometer for physical activity measurement," *Journal of science and medicine in sport*, vol. 19, no. 12, pp. 1015–1019, 2016.

[13] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[14] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.

[15] M. Wei, R. Yang, and M. Huang, "Motor imagery eeg signal classification based on deep transfer learning," in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2021, pp. 85–90.

[16] F. Andreotti, H. Phan, N. Cooray, C. Lo, M. T. Hu, and M. De Vos, "Multichannel sleep stage classification and transfer learning using convolutional neural networks," in *2018 40th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 171–174.

[17] W. M. Shalash, "Driver fatigue detection with single eeg channel using transfer learning," in *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2019, pp. 1–6.

[18] S. Byun, B. G. B. Bueno, Y. Gupta, N. Dhadge, S. Pawar, R. Kodgule, and R. R. Fletcher, "The use of thermal imaging and deep learning for pulmonary diagnostics and infection detection," in *2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2021, pp. 1–4.

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[20] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *Proceedings of the 2012 16th Annual International Symposium on Wearable Computers (ISWC)*. USA: IEEE Computer Society, 2012, p. 108–109. [Online]. Available: https://doi.org/10.1109/ISWC.2012.13

[21] ——, "Creating and benchmarking a new dataset for physical activity monitoring," in *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*. New York, NY, USA: Association for Computing Machinery, 2012. [Online]. Available: https://doi.org/10.1145/2413097.2413148

[22] J. DErrico, "inpaint nans," *MATLAB Central File Exchange, https://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint_nans*, 2022.

[23] R. Polikar, "The wavelet tutorial second edition part i," *Fundamental Concepts & An Overview of The Wavelet Theory*, 1996.

[24] G. Wang, Q. Li, L. Wang, W. Wang, M. Wu, and T. Liu, "Impact of sliding window length in indoor human motion modes and pose pattern recognition based on smartphone sensors," *Sensors*, vol. 18, no. 6, p. 1965, 2018.

[25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.