International Journal of Medical Informatics 172 (2023) 105004

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

AccNet24: A deep learning framework for classifying 24-hour activity behaviours from wrist-worn accelerometer data under free-living environments

Vahid Farrahi [a,b,*], Usman Muhammad [b], Mehrdad Rostami [b], Mourad Oussalah [a,b]

[a] Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
[b] Center of Machine Vision and Signal Analysis, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland

ARTICLE INFO

Keywords:
Activity recognition
Transfer learning
Raw acceleration
Machine learning
Activity classification

ABSTRACT

Objective: Although machine learning techniques have been repeatedly used for activity prediction from wearable devices, accurate classification of 24-hour activity behaviour categories from accelerometry data remains a challenge. We developed and validated a deep learning-based framework for classifying 24-hour activity behaviours from wrist-worn accelerometers.
Methods: Using an openly available dataset with free-living wrist-based raw accelerometry data from 151 participants (aged 18–91 years), we developed a deep learning framework named AccNet24 to classify 24-hour activity behaviours. First, the acceleration signal (x, y, and z-axes) was segmented into 30-second nonoverlapping windows, and signal-to-image conversion was performed for each segment. Deep features were automatically extracted from the signal images using transfer learning and transformed into a lower-dimensional feature space. These transformed features were then employed to classify the activity behaviours as sleep, sedentary behaviour, and light-intensity (LPA) and moderate-to-vigorous physical activity (MVPA) using a bidirectional long short-term memory (BiLSTM) recurrent neural network. AccNet24 was trained and validated with data from 101 and 25 randomly selected participants and tested with the remaining unseen 25 participants. We also extracted 112 hand-crafted time and frequency domain features from 30-second windows and used them as inputs to five commonly used machine learning classifiers, including random forest, support vector machines, artificial neural networks, decision tree, and naïve Bayes to classify the 24-hour activity behaviour categories.
Results: Using the same training, validation, and test data and window size, the classification accuracy of AccNet24 outperformed the accuracy of the other five machine learning classification algorithms by 16%–30% on unseen data.
Conclusion: AccNet24, relying on signal-to-image conversion, deep feature extraction, and BiLSTM achieved consistently high accuracy (>95 %) in classifying the 24-hour activity behaviour categories as sleep, sedentary, LPA, and MVPA. The next generation accelerometry analytics may rely on deep learning techniques for activity prediction.

1. Introduction

Historically, accelerometers were used to monitor activities during waking hours [1,2], followed by the development of different age-specific sets of cut-points for translating acceleration into activity types and intensity categories [3–5]. Although such cut-points remain widely used [6], their accuracy in estimating activity intensities remains limited [7]. Recent studies suggest that all activity behaviours within the 24-hour cycle, including sleep, sedentary behaviours, and light-intensity (LPA) and moderate-to-vigorous physical activity (MVPA), may be codependently related to health indicators [8–10].

In recent years, a paradigm shift, from threshold-based approaches to using machine learning methodologies and raw accelerometry, is noticeable [11–14]. Existing studies have demonstrated that conventional machine learning algorithms such as random forest (RF) [15,16], artificial neural networks (ANN) [16,17], support vector machines (SVM) [16], and decision tree (DT) [16,18] can provide accurate predictions of activity intensities and categories. However, existing

* Corresponding author at: Research Unit of Health Sciences and Technology, P.O. 5000, FI-90014 University of Oulu, Finland.
E-mail address: Vahid.farrahi@oulu.fi (V. Farrahi).

https://doi.org/10.1016/j.ijmedinf.2023.105004
Received 26 October 2022; Received in revised form 9 December 2022; Accepted 20 January 2023
Available online 25 January 2023

literature reviews reveal that machine learning methods have achieved variable success in relatively small samples, and their generalization to free-living data remains limited [11,13,19].

Conventional machine learning techniques have limited ability to process data in their raw form and therefore require hand-designed feature extractors [20]. In practice, an extensive list of time and frequency features can be manually extracted from raw acceleration data [11,21,22], but only some of them appear to contain relevant information for discriminating activity intensities and categories [22]. An alternative method that does not rely on hand-designed feature extractors is deep learning approach [20,23]. Compared to conventional machine learning algorithms, deep learning approaches identify and learn representations from data in an automated manner [20]. Although potentially better than conventional machine learning methodologies in several ways [20], recent reviews of literature indicate that existing studies have continued to use conventional machine learning techniques for activity prediction with no study using deep learning techniques for predicting 24-hour activity behaviours [11,13,21,23]. Following the emergence of the 24-hour activity cycle paradigm [8,9], this study developed and validated a deep learning framework, dubbed AccNet24, to classify 24-hour activity behaviour into sleep, sedentary behaviours, LPA, and MVPA from free-living wrist-based raw accelerometry data and compared its classification performance with five commonly employed machine learning classification algorithms [11,13,21].

2. Methods

2.1. Dataset

We used data from the Capture-24 dataset [24,25], an open-access accelerometer validation study of 151 adults aged 18–91. Participants were asked to wear an Axivity AX3 wrist-worn accelerometer for 24 h. They were also asked to wear a Vicon Autographer wearable camera on their chest while awake and keep records of their sleeping periods in a time-use diary. The accelerometer was set to record the raw acceleration data at 100 Hz, and the wearable camera was set to take an image every 30 s. A detailed description of the data and recruitment process is outlined elsewhere [24,25].

2.2. Annotations and 24-hour activity behaviour categories

Trained researchers annotated the accelerometer data based on camera images with hierarchical labels from the Compendium of Physical Activities [26], as detailed elsewhere [24,25]. Sleeping periods were marked based on the diary. The activity types were mapped to sleep, sedentary behaviour, LPA, and MVPA according to the body posture and activity energy expenditure values listed in the Compendium to form the 24-hour activity behaviour categories. Table 1 shows the definitions of the activity behaviour categories, together with some examples of camera image annotations.

Table 1
Examples of camera image annotations on the basis of the Compendium of Physical Activities [26], and how those were mapped into 24-hour activity behavior categories. A full list of the activities in the dataset is presented elsewhere [25].

Class	Definition	Example of hierarchical annotations: Compendium category and code
Sleep	Self-reported bedtime	• Sleep: 7030 sleeping
Sedentary	Waking behaviour at < 1.5 METs in a sitting, lying or reclining posture	• Home activity, miscellaneous, sitting: 9060 sitting/lying reading or without observable/ identifiable activities • Transportation; private transportation:16010 driving automobile or light truck (not a semi) • home activity; miscellaneous; sitting:11580 office/computer work general
Light-intensity physical activities (LPA)	Waking behaviour at < 3 METs not meeting the sedentary behaviour definition.	• Home activity; household chores; preparing meals/cooking/ washing dishes: 5035 kitchen activity general cooking/ washing/dishes/cleaning up • Leisure; miscellaneous; walking: 21,070 (generic) walking and occasional standing (no more than two consecutive images) • Leisure; miscellaneous: 5060 shopping miscellaneous • Occupation; miscellaneous: 11,475 (generic) manual labour
Moderate-to-vigorous-physical activities (MVPA)	Waking behaviour at ≥ 3 METs	• Transportation; private transportation:1010 bicycling • Occupation; interruption: 17,133 walking upstairs • Leisure; sports; water activities:18070 canoeing/rowing • Home activity; household chores; house cleaning; furniture: 5020 cleaning heavy such as car/ windows/garage

METs = metabolic equivalents.

2.3. Deep learning framework for predicting 24-hour activity behaviours (AccNet24)

Fig. 1 shows the architecture of our AccNet24 model designed for predicting 24-hour activity behaviours from accelerometry data. The AccNet24 framework was designed to operate with two-dimensional (2D) images rather than one-dimensional raw acceleration data. Based on the recent success in medical image analysis [27], this aimed to transform the signal processing problem into a machine vision problem, enabling the use of advanced deep learning techniques originally proposed for image data.

Briefly, the main steps of AccNet24 are as follows. First, the acceleration signal was segmented, and signal-to-image conversion was performed for each segment. The signal images were used to automatically extract deep features using transfer learning. These features were then transformed into an uncorrelated, lower-dimensional feature space, and the transformed features were used as inputs for a bidirectional long short-term memory (BiLSTM) recurrent neural network (RNN) to classify the 24-hour activity behaviour categories as sleep, sedentary behaviour, LPA, and MVPA.

2.4. Time series-to-image encoding

The acceleration data in each axis were segmented into nonoverlapping 30-second windows, and a signal image was created for each window in the ×, y, and z- directions. Signal-to-image conversion was performed with the Gramian angular field (GAF) technique. An image created with the GAF technique represents the temporal correlation between each pair of values in the time series [28]. GAF images are created by representing the normalized signal in the polar coordinate system and computing the trigonometric function [28]. This procedure produces a square matrix representing a pairwise temporal correlation between the elements that can be plotted as an image in two-dimensional space, as illustrated in Fig. 2.

2.5. Deep feature extraction using transfer learning

Automatic deep feature extraction was performed with transfer learning. In theory, the underlying assumption behind transfer learning is that when a prediction model is trained well to extract important features for prediction, the knowledge can be transformed to another

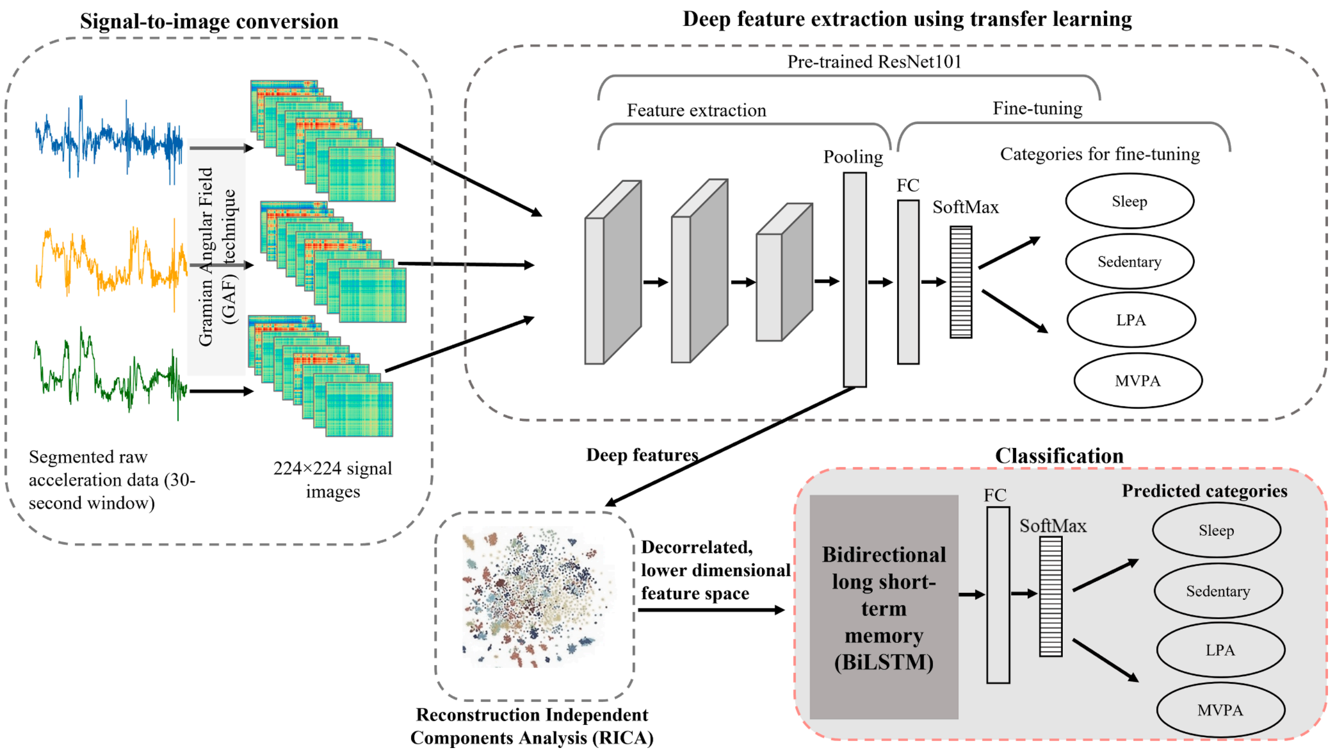

Fig. 1. Overall deep learning-based framework of AccNet24. FC = Fully connected layer; LPA = Light-intensity physical activity. MVPA = Moderate-to-vigorous-intensity physical activity.

domain after fine-tuning the model to avoid the need to train a model from scratch [29–31].

In practice, the state-of-the-art convolutional neural network (CNN)-based architectures trained on the ImageNet dataset have been the basis for performing transfer learning in machine vision applications [32]. This is partly because the ImageNet dataset is a massive dataset with millions of labelled natural images [33], and different CNN-based architectures with automated feature extraction strategies have achieved excellent performances with it [34]. To extract high-quality features for predicting the 24-hour activity behaviours from the signal images, we fine-tuned pretrained ResNet101 architecture on the ImageNet dataset and changed the final fully connected layer to predict sleep, sedentary behaviours, LPA, and MVPA. ResNet101 is a 101-layer-deep CNN-based architecture known for its stability and performance [35].

2.6. Lower-dimensional feature space

Data representation (or features), generated either automatically or manually, can directly influence the performance of deep learning and conventional machine learning algorithms [36]. Among many existing methods for data representation [36], independent component analysis (ICA) is a widely used technique for representing the features in an uncorrelated, lower-dimensional space [37]. Briefly, ICA aims to separate the input space into maximally independent components and has been repeatedly used with deep learning algorithms to decorrelate the feature space and improve prediction performance [37].

However, predicting the 24-hour activities is an imbalanced classification task. In the 24-hour context, MVPA is a sparse category constituting only 3 %–5% of people's time [38]. Therefore, we used a special form of ICA for transforming the feature space into a decorrelated, lower-dimensional feature space called reconstruction ICA (RICA). Compared to the original version of ICA, RICA has a better ability to work with sparse data, also providing a better chance for the prediction algorithm to perform well for sparse categories [39]. The formulation and assumptions behind RICA are presented in the online supplemental methods.

2.7. Classification with BiLSTM

RNNs are a special class of deep neural networks that can use sequential information in the network [40]. This property is essential in time-series data analysis applications, where the embedded structure in the data sequence conveys useful knowledge [40]. However, although movement and nonmovement behaviours may contain sequential information (e.g., walking and running), activity prediction is generally performed with segmented signals (typically 10–30-second windows) [11,13,21], limiting the amount of sequential information available to the network for learning the patterns.

BiLSTM is a class of RNNs designed to address some of the inherent problems of the original RNN architecture (e.g., vanishing gradient problem) and has obtained outstanding performances in many applications with sequential data [40]. Compared to other types of RNNs, an important characteristic of BiLSTM is that it increases the amount of information available to the network for a better prediction by taking input in both backwards and forwards directions [41]. Our motivation for using BiLSTM is to take advantage of the temporal pattern of the signals across signal images. The deep features that were decorrelated with RICA were used as input to BiLSTM to classify the 24-hour activity behaviours as sleep, sedentary behaviours, LPA, or MVPA.

2.8. Conventional machine learning methods

To compare the performance of AccNet24 with other existing methodologies, we also developed and validated five machine learning classifiers for predicting the 24-hour activity behaviours. We trained a balanced RF with random under-sampling in each boostrap to balance the classes, as well as four other classification algorithms including ANN, SVM, DT, and naïve Bayes (NB) [11,13,21].

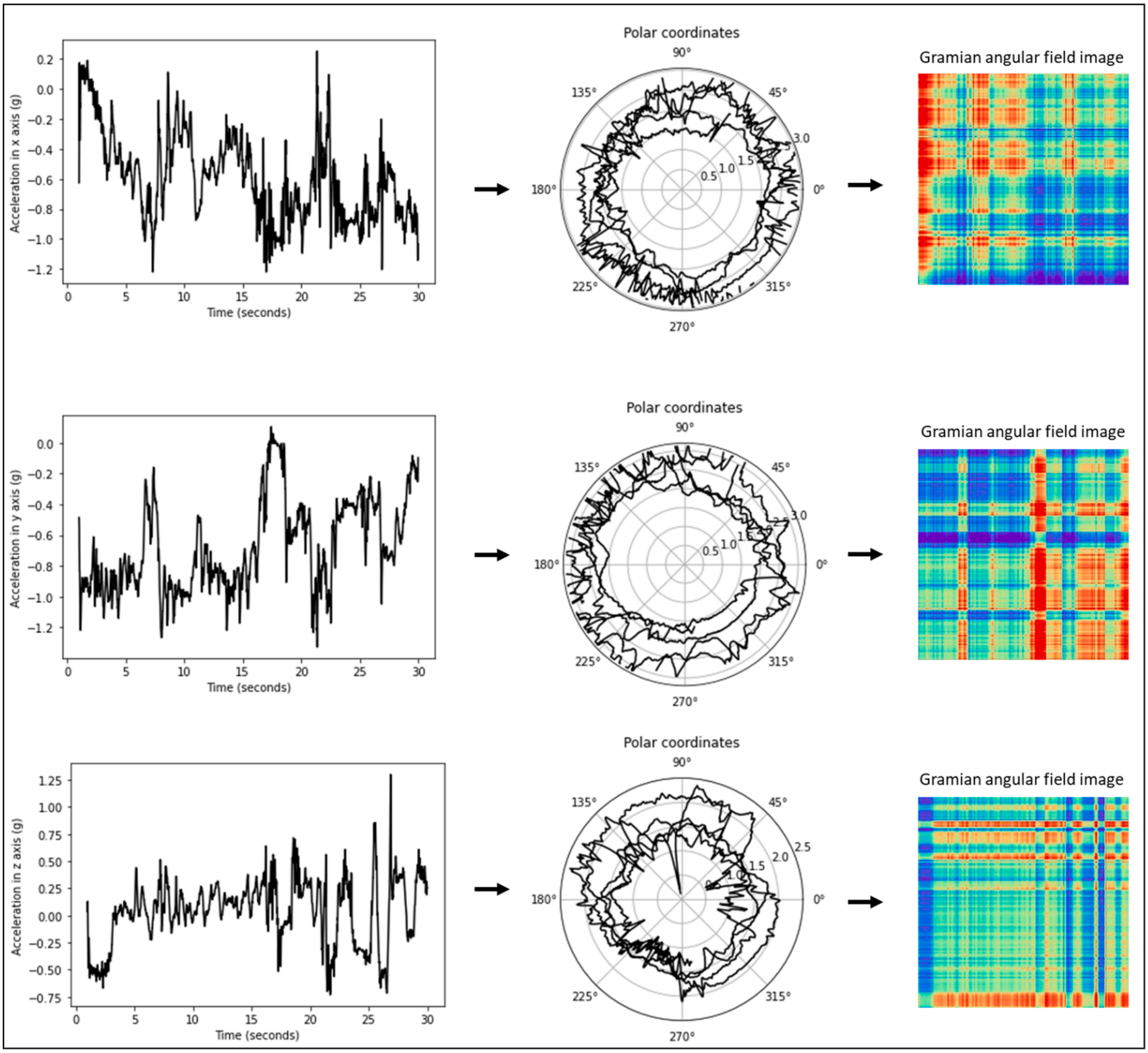

Fig. 2. An example demonstrating how signal-to-image conversion is performed using Gramian Angular Technique (GAF). A 30-second signal segment is from one the participants performing a light-intensity activity.

2.9. Feature generation and extraction

Following the existing studies [11,22], the input features were directly extracted from the three acceleration axes (x, y, and z), vector magnitude, or both. Similarly to AccNet24, the acceleration data in each axis were segmented into nonoverlapping 30-second windows. For each window, we extracted a 112-dimensional feature vector. These features were selected from an extensive list of time and frequency domain features described in previous studies [18,24,25] and are listed in online supplemental Table 1. The extracted features were used as inputs for balanced RF, ANN, SVM, DT, and NB to classify sleep, sedentary behaviour, LPA, and MVPA.

2.10. Training, validation, and testing

We randomly split the whole dataset into training, testing, and validation sets, comprising data from 101 (67 %), 25 (16.5 %), and 25 (16.5 %) of the participants, respectively. Splitting the dataset was participant-based: each participant's data were used either in the training, validation, or testing phase, resembling the commonly used leave-one-participant-out procedure [42] to provide a realistic understanding of the generalizability of the results. All the methods were trained with the training set and validated on the validation set to fine-tune the hyperparameters. The trained model was then tested on unseen data (test set).

For AccNet24, one single image was created for each measurement axis (x, y, and z), resulting in three images per 30-second window. All these images were used in the training phase to increase the amount of data for training. However, this resulted in three predictions per window for the validation and test sets. To obtain a single prediction per 30-second window, we defined the final prediction based on voting among the labels predicted for the three images. To enable direct comparison among the models, we used the same participant's data when training, validating, and testing using AccNet24 and the machine learning classification models.

2.11. Parameters and implementation details

Signal images were created using the "pyts" package in Python [43], and transfer learning and BiLSTM were implemented in MATLAB. All the code is available online (see online supplemental material). All the signal images were resized to 224 × 224 pixels based on the original input size for ResNet101. Both high- and low-dimensional feature space may limit the ability of prediction [36]. To build the lower-dimensional feature space, RICA received the feature vector from the output of the last pooling layer of ResNet101 with a size of 2048 and was set to transform the deep features into a 100-dimensional feature space. Our empirical tests with Capture-24 dataset indicated that 100 features are appropriate for predicting 24-hour activity behaviours (see online supplementary material). We selected appropriate parameters for balanced RF, ANN, SVM, DT, and NB based on previously published studies [18,22,25,44] and fine-tuned them to better fit this dataset. Table 2 provides the full list of parameters.

2.12. Performance and sensitivity analyses

Confusion matrices were used to evaluate each method's classification and misclassification rates, as well as overall classification accuracies on both the validation and test sets. To address the problem of overfitting [36], we performed a sensitivity analysis to ensure that the AccNet24 framework was minimally dependent on the training set. We repeated the training, validation, and testing with the same parameters while exchanging the participants' data, which were initially assigned to validation and test sets (25 and 25 participants) with the data of 50 randomly selected participants from the training set.

3. Results

3.1. Classification performance of AccNet24

Learning curves and confusion matrix for transfer learning are shown in the online supplementary material. Fig. 3 shows confusion matrices illustrating the performances of the AccNet24 framework with the validation and test sets. The overall accuracy of AccNet24 was 96.9 % for the validation data. When tested with unseen data (test set), AccNet24 obtained 98.2 % accuracy, and the prediction accuracy across all categories consistently exceeded 90 %. Comparable results were obtained across all categories when retraining AccNet24 with different participants' data in training, testing, and validation sets, indicating the robustness of AccNet24 irrespective of the training data (see online supplemental results).

Table 2
Hyperparameters used for training AccNet24 and machine learning classifiers.

Method	Hyperparameters
AccNet24	• ResNet101 was fine-tuned using stochastic gradient descent (SGD) with a learning rate of $3e^{-4}$ with mini-batch size of 16 and epochs of 5. Sample shuffling was done between every epoch. • BiLSTM included a layer with 60 hidden units, a fully connected layer, a SoftMax layer, and the number of epochs were fixed to 150. • BiLSTM was trained using an Adam optimizer, setting the learning rate up to 10^{-4}, and the recurrent weights were set with He initializer.
Balanced random forest (RF)	• Each random forest contained 500 trees, and the size of the random subset of features at each split was the square root of the total number of features. Sampling with replacement was done for all classes
Artificial Neural Networks (ANN)	• Each network contained a single hidden layer of 15 nodes and was trained for a maximum of 200 iterations.
Support vector machines (SVM)	• The kernel type was radial basis function (rbf), and the regularization parameter was set to 200.
Decision tree (DT)	• The minimum number of samples required to split an internal node was set to 100, and the minimum number of samples required to be at a leaf node was set to 50

3.2. Comparison of AccNet24 with balanced RF, ANN, SVM, DT, and NB classification algorithms

Fig. 4 presents confusion matrices showing the performances of balanced RF, ANN, SVM, DT, and NB. Among these classifiers, the classification accuracy of balanced RF and ANN was higher on the validation (81.6 % and 80.5 %) and unseen (82.0 % and 82.3 %) data, followed by SVM, DT, and NB. Overall, across all the machine learning classifiers, the accuracy for predicting sleep and sedentary categories was higher than for LPA and MVPA. Compared to the machine learning classifiers, the accuracy of AccNet24 across all categories was highest, leading to 15.3 %–28.7 % and 15.9 %–30.6 % higher classification accuracy on validation and unseen data, respectively.

4. Discussion

The present study developed and validated a deep learning framework for classifying 24-hour activity behaviour categories from free-living wrist-based raw acceleration data. The overall accuracy of the AccNet24 framework in classifying sleep, sedentary behaviour, LPA, and MVPA exceeded 95 %, outperforming the accuracy achieved with the five commonly employed machine learning algorithms, RF, ANN, SVM, DT, and NB, by approximately 16 %–30 %.

Our results are directly comparable with those reported in two published studies that used the same dataset for developing and validating a hybrid machine learning model using RF and the hidden Markov model to classify 24-hour activity categories [24,25]. These studies both reported a lower accuracy than the AccNet24 with a similar window size (88 % [25] and 87 % [24] vs 98 %), particularly due to having a lower classification accuracy for the LPA and MVPA categories. Although not directly comparable due to different datasets and window sizes, the classification accuracy of AccNet24 appears to exceed the achieved accuracy values in most of the previous studies developing and validating machine learning models with free-living data [11,13,21].

Similar to AccNet24, a few studies have used deep learning-based methodologies for predicting activity types and intensities [45–48]. Contrary to expectations, the achieved accuracy values in these studies, if not lower [45], were only slightly higher than those reported for conventional machine learning approaches [11,13,21]. This may be partially because training deep learning models from scratch typically requires mass data [29,30], which is typically expensive and difficult to acquire. These studies have had limited training data [45–47], which appears to hinder the potential ability of deep learning approaches to automatically identify the most appropriate features to perform better prediction than other methodologies. This also appears to be related to the persistent tendency to use conventional machine learning algorithms for predicting activity intensities and categories [11,13,21], compared to other fields of research involving time-series data that have already witnessed the deployment of deep learning techniques [49].

AccNet24 has two distinct differences compared to other existing methodologies [11,13,24,25,45,47]. First, AccNet24 converts raw acceleration signals into 2D images, providing the opportunity to use pretrained deep learning models that have already learned how to identify high-quality features on millions of natural images. Without massive training data, AccNet24 identified deep features for predicting 24-hour activity behaviour categories using transfer learning, which may be a reason for its superior performance. Second, AccNet24 uses BiLSTM for classification, which assumes that the underlying data have useful sequential information [40]. This assumption appears to match well with the problem of predicting daily activities, considering that human movement and nonmovement behaviours may be periodic (e.g., walking and running).

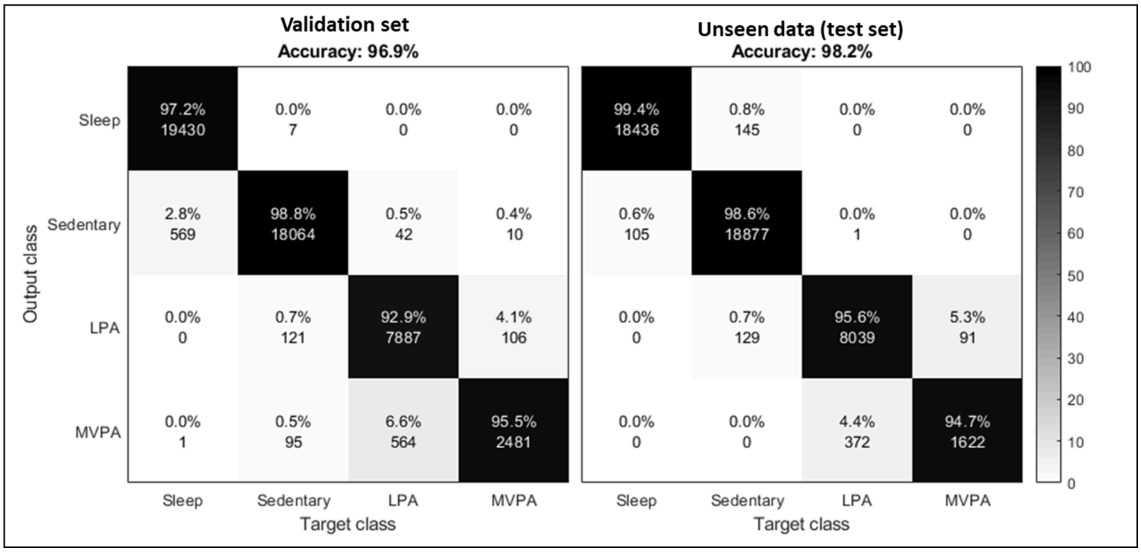

Fig. 3. Confusion matrices showing the performance of AccNet24 on validation and unseen (test set) data. AccNet24 was trained with data from 101 randomly selected participants, and validation and test sets each included data from 25 participants.

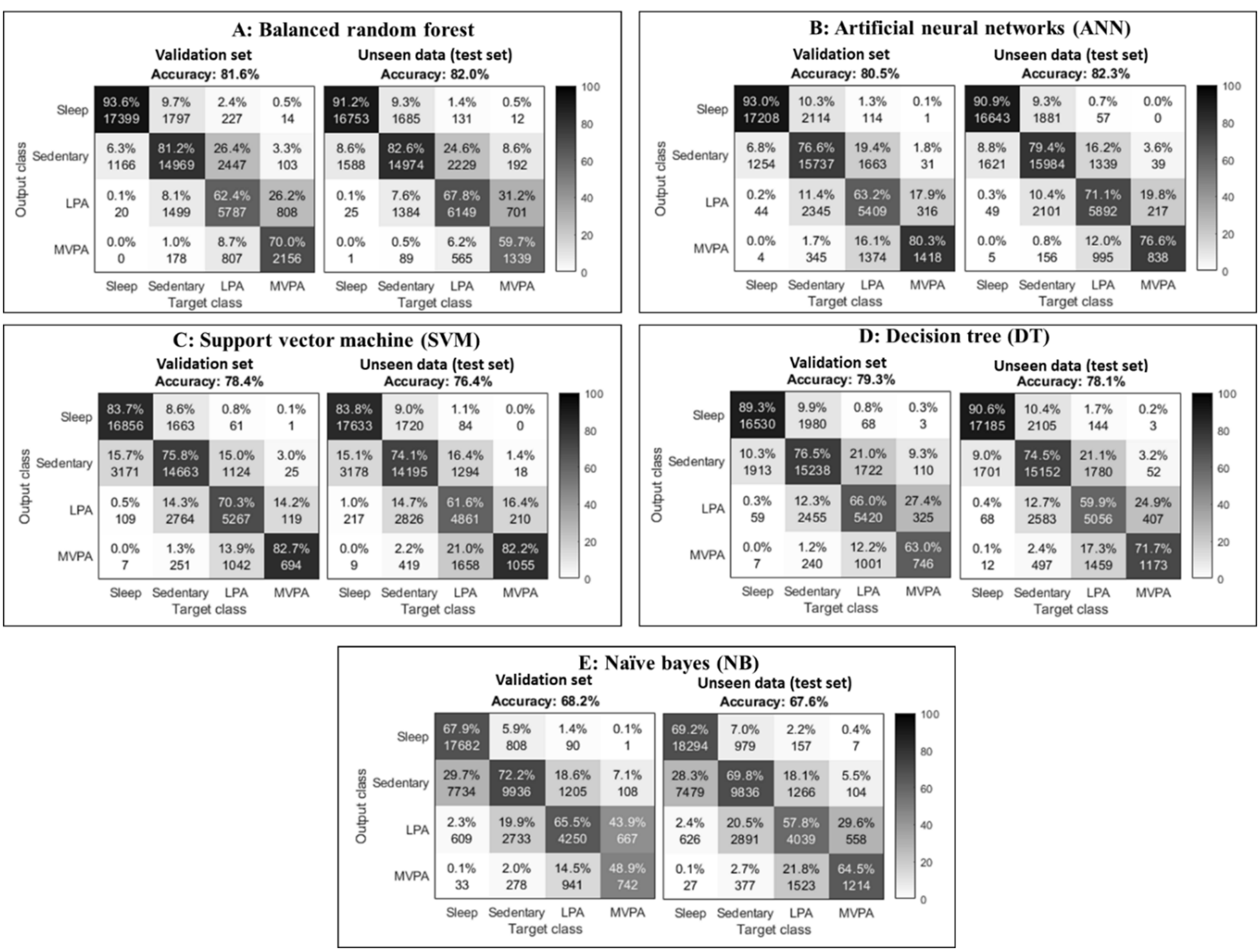

Fig. 4. Confusion matrices showing the performance of (A) balanced random forest, (B) artificial neural networks, (C) support vector machines, (D) decision tree, and (E) Naïve bayes on validation and unseen (test set) data. The models were trained with 101 randomly selected participants, and validation and test set each included data from 25 participants. To enable direct comparison, the participants assigned to training, validation and test sets were the same as those used for training, validating, and testing AccNet24 framework. Note that we trained, validated, and tested a balanced random forest, which randomly under-samples each bootstrap to balance the sample. No strategies for addressing the imbalance classification problem were used when training, validating, and testing, ANN, SVM, DT, and NB.

4.1. Strengths

The main strength of AccNet24 was its consistently high accuracy across all the 24-hour activity behaviour categories. Overall, previous studies have typically obtained varying accuracies across different activity categories [13,19], often performing better in differentiating between sleep and sedentary time but achieving poorer performances in distinguishing between LPA and MVPA [24,25]. AccNet24 was developed and tested with free-living acceleration data. This is a strength because most existing studies developing methodologies for predicting activity behaviours have typically used laboratory-collected data [11,19,21]. Another strength is using a relatively large open-access dataset [23–25], enabling future studies to repeat and compare their results with AccNet24. We also performed a sensitivity analysis to assess the robustness of AccNet24. No overfitting problem was evident, considering that similar results were obtained when retraining the model with different training, testing, and validation data.

4.2. Limitations

The participants providing data for the Capture-24 dataset spanned a wide age range. Due to a lack of measured energy expenditure, we used similar energy expenditure thresholds for defining activity intensity categories for all participants. This resulted in assigning all the activities of the same type to the same intensity category without considering the variability between the participants in energy expenditure. The complexity of the model and its implementation is an inherent problem of deep learning architecture [20], and AccNet24 is no exception. Execution time for AccNet24 was approximately 26 h (on a machine with 2 graphical processing units, a 3.5 GHz CPU, and a 32 GB RAM). The codes for implementing and using pretrained AccNet24 are made available online to facilitate using and implementing AccNet24.

5. Conclusions and future works

AccNet24, relying on signal-to-image conversion, deep feature extraction, and BiLSTM for classification, achieved consistently high accuracy in classifying the 24-hour activity behaviour categories as sleep, sedentary, LPA, and MVPA from raw wrist-based accelerometry data. The overall accuracy of AccNet24 outperformed five commonly used machine learning classification algorithms for activity prediction by 16 %–30 %. With the recent paradigm shift towards 24-hour activity cycle paradigms [8,9], AccNet24 appears to be a step towards precise assessment of 24-hour activity behaviour categories. Future studies will consider such deep learning frameworks for prediction of activity types.

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.

Acknowledgements

The present study is connected to the DigiHealth-project, a strategic profiling project at the University of Oulu. The project is supported by the Academy of Finland (project number 326291) and the University of Oulu. This study has also received funding from the Ministry of Education and Culture in Finland [grant number OKM/20/626/2022]. The funders played no role in designing the study, or collecting, analyzing, and interpreting the data, or writing the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2023.105004.

References

[1] Z. Xie, A. Jo, Y.-R. Hong, Electronic wearable device and physical activity among US adults: an analysis of 2019 HINTS data, Int. J. Med. Inf. 144 (2020), 104297.
[2] I.-M. Lee, E.J. Shiroma, Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges, Br. J. Sports Med. 48 (2014) 197–201, https://doi.org/10.1136/bjsports-2013-093154.
[3] S.G. Trost, P.D. Loprinzi, R. Moore, K.A. Pfeiffer, Comparison of accelerometer cut-points for predicting activity intensity in youth, Med. Sci. Sport. Exerc. 43 (2011) 1360–1368.
[4] S.E. Crouter, D.R. Bassett, A new 2-regression model for the Actical accelerometer, Br. J. Sports Med. 42 (2008) 217–224.
[5] E.A.H. Winkler, P.A. Gardiner, B.K. Clark, C.E. Matthews, N. Owen, G.N. Healy, Identifying sedentary time using automated estimates of accelerometer wear time, Br. J. Sports Med. 46 (2012) 436–442.
[6] J.H. Migueles, C. Cadenas-Sanchez, U. Ekelund, C.D. Nyström, J. Mora-Gonzalez, M. Löf, I. Labayen, J.R. Ruiz, F.B. Ortega, Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations, Sport. Med. 47 (2017) 1821–1845, https://doi.org/10.1007/s40279-017-0716-0.
[7] Y. Kim, M.W. Beets, G.J. Welk, Everything you wanted to know about selecting the "right" Actigraph accelerometer cut-points for youth, but…: a systematic review, J. Sci. Med. Sport 15 (2012) 311–321, https://doi.org/12.001.
[8] J.H. Migueles, E. Aadland, L.B. Andersen, J.C. Brønd, S.F. Chastin, B.H. Hansen, K. Konstabel, O.M. Kvalheim, D.E. McGregor, A.V. Rowlands, S. Sabia, V.T. van Hees, R. Walmsley, F.B. Ortega, GRANADA consensus on analytical approaches to assess associations with accelerometer-determined physical behaviours (physical activity, sedentary behaviour and sleep) in epidemiological studies, Br. J. Sports Med. (2022) 376–384.
[9] M.E. Rosenberger, J.E. Fulton, M.P. Buman, R.P. Troiano, M.A. Grandner, D. M. Buchner, W.L. Haskell, The 24-hour activity cycle: a new paradigm for physical activity, Med. Sci. Sports Exerc. 51 (2019) 454–464.
[10] V. Farrahi, M. Kangas, R. Walmsley, M. Niemelä, A. Kiviniemi, K. Puukka, P. J. Collings, R. Korpelainen, T. Jämsä, Compositional associations of sleep and activities within the 24-h cycle with cardiometabolic health markers in adults, Med. Sci. Sports Exerc. 53 (2021) 324–332, https://doi.org/10.1249/MSS.0000000000002481.
[11] V. Farrahi, M. Niemelä, M. Kangas, R. Korpelainen, T. Jämsä, Calibration and validation of accelerometer-based activity monitors: a systematic review of machine-learning approaches, Gait Posture 68 (2019) 285–299, https://doi.org/10.1016/j.gaitpost.2018.12.003.
[12] D.R. Bassett Jr, A.V. Rowlands, S.G. Trost, Calibration and validation of wearable monitors, Med. Sci. Sports Exerc. 44 (2012) S32–S38.
[13] M. de Almeida Mendes, I.C.M. da Silva, V.V. Ramires, F.F. Reichert, R.C. Martins, E. Tomasi, Calibration of raw accelerometer data to measure physical activity: a systematic review, Gait Posture 61 (2018) 98–110.
[14] R.P. Troiano, J.J. McClain, R.J. Brychta, K.Y. Chen, Evolution of accelerometer methods for physical activity research, Br. J. Sports Med. 48 (2014) 1019–1023.
[15] J. Kerr, R.E. Patterson, K. Ellis, S. Godbole, E. Johnson, G. Lanckriet, J. Staudenmayer, Objective assessment of physical activity: classifiers for public health, Med. Sci. Sports Exerc. 48 (2016) 951–957.
[16] A.H.K. Montoye, B.S. Westgate, M.R. Fonley, K.A. Pfeiffer, Cross-validation and out-of-sample testing of physical activity intensity predictions using a wrist-worn accelerometer, J. Appl. Physiol. 124 (2018) 1284–1293.
[17] V. Farrahi, M. Niemelä, P. Tjurin, M. Kangas, R. Korpelainen, T. Jämsä, Evaluating and enhancing the generalization performance of machine learning models for physical activity intensity prediction from raw acceleration data, IEEE J. Biomed. Heal. Informatics. 24 (2020) 27–38, https://doi.org/10.1109/JBHI.2019.2917565.
[18] S. Zhang, A.V. Rowlands, P. Murray, T.L. Hurst, Physical activity classification using the GENEA wrist-worn accelerometer, Med. Sci. Sport. Exerc. 44 (2012) 742–748.
[19] C.C.T. Clark, C.M. Barnes, G. Stratton, M.A. McNarry, K.A. Mackintosh, H. D. Summers, A review of emerging analytical techniques for objective physical activity measurement in humans, Sport. Med. 47 (2017) 439–447.
[20] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[21] A. Narayanan, F. Desai, T. Stewart, S. Duncan, L. Mackay, Application of raw accelerometer data and machine-learning techniques to characterize human movement behavior: a systematic scoping review, J. Phys. Act. Heal. 17 (2020) 360–383.
[22] J. Chong, P. Tjurin, M. Niemelä, T. Jämsä, V. Farrahi, Machine-learning models for activity class prediction: A comparative study of feature selection and classification algorithms, Gait Posture 89 (2021) 45–53, https://doi.org/10.1016/j.gaitpost.2021.06.017.
[23] J. Wang, Y. Chen, S. Hao, X. Peng, L. Hu, Deep learning for sensor-based activity recognition: a survey, Pattern Recogn. Lett. 119 (2019) 3–11, https://doi.org/10.1016/j.patrec.2018.02.010.
[24] M. Willetts, S. Hollowell, L. Aslett, C. Holmes, A. Doherty, Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants, Sci. Rep. 8 (2018).
[25] R. Walmsley, S. Chan, K. Smith-Byrne, R. Ramakrishnan, M. Woodward, K. Rahimi, T. Dwyer, D. Bennett, A. Doherty, Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease, Br. J. Sports Med. 56 (2021) 1008–1017.
[26] B.E. Ainsworth, W.L. Haskell, S.D. Herrmann, N. Meckes, D.R. Bassett, C. Tudor-Locke, J.L. Greer, J. Vezina, M.C. Whitt-Glover, A.S. Leon, Compendium of Physical Activities, Med. Sci. Sport. Exerc. 43 (2011) (2011) 1575–1581.

[27] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[28] Z. Wang, T. Oates, Imaging time-series to improve classification and imputation, in: Twenty-Fourth Int. Jt. Conf. Artif. Intell., 2015: pp. 3939–3945.
[29] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2009) 1345–1359.
[30] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data. 3 (2016).
[31] U. Muhammad, M.Z. Hoque, M. Oussalah, A. Keskinarkaus, T. Seppänen, P. Sarder, SAM: Self-augmentation mechanism for COVID-19 detection using chest X-ray images, Knowledge-Based Syst. 241 (2022), 108207.
[32] M.A. Morid, A. Borjali, G. Del Fiol, A scoping review of transfer learning research on medical image analysis using ImageNet, Comput. Biol. Med. 128 (2021), 104115.
[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conf. Comput. Vis. Pattern Recognit., 2009: pp. 248–255.
[34] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84–90.
[35] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016: pp. 770–778.
[36] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 1798–1828.
[37] M. Längkvist, L. Karlsson, A. Loutfi, A review of unsupervised feature learning and deep learning for time-series modeling, Pattern Recogn. Lett. 42 (2014) 11–24.
[38] J.-P. Chaput, V. Carson, C. Gray, M. Tremblay, Importance of all movement behaviors in a 24 hour period for overall health, Int. J. Environ. Res. Public Health 11 (2014) 12575–12581.
[39] Q. Le, A. Karpenko, J. Ngiam, A. Ng, ICA with reconstruction cost for efficient overcomplete feature learning, Adv. Neural Inf. Process. Syst. 24 (2011) 1017–1025.
[40] I.H. Sarker, Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions, SN Comput. Sci. 2 (2021) 1–20.
[41] G. Van Houdt, C. Mosquera, G. Nápoles, A review on the long short-term memory model, Artif. Intell. Rev. 53 (2020) 5929–5955.
[42] J. Staudenmayer, W. Zhu, D.J. Catellier, Statistical considerations in the analysis of accelerometry-based activity monitor data, Med. Sci. Sports Exerc. 44 (2012) S61–S67.
[43] J. Faouzi, H. Janati, pyts: A Python Package for Time Series Classification, J. Mach. Learn. Res. 21 (2020) 41–46.
[44] A. Dalton, G., OLaighin, Comparing supervised learning techniques on the task of physical activity recognition, IEEE J. Biomed. Heal. Informatics. 17 (2012) 46–52.
[45] A. Logacjov, K. Bach, A. Kongsvold, H.B. Bårdstu, P.J. Mork, HARTH: A Human Activity Recognition Dataset for Machine Learning, Sensors 21 (2021) 7853.
[46] M. Hagenbuchner, D.P. Cliff, S.G. Trost, N. Van Tuc, G.E. Peoples, Prediction of activity type in preschool children using machine learning techniques, J. Sci. Med. Sport 18 (2015) 426–431.
[47] R. Nawaratne, D. Alahakoon, D. De Silva, P.D. O'Halloran, A.H.K. Montoye, K. Staley, M. Nicholson, M.I.C. Kingsley, Deep learning to predict energy expenditure and activity intensity in free living conditions using wrist-specific accelerometry, J. Sports Sci. 39 (2021) 683–690.
[48] H. Xu, J. Li, H. Yuan, Q. Liu, S. Fan, T. Li, X. Sun, Human activity recognition based on Gramian angular field and deep convolutional neural network, IEEE Access 8 (2020) 199393–199405.
[49] A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, J. Neural Eng. 16 (2019) 31001.