



Framework for Human Activity Recognition on Smartphones and Smartwatches

Blagoj Mitrevski¹, Viktor Petreski¹, Martin Gjoreski²,
and Biljana Risteska Stojkoska¹✉

¹ Faculty of Computer Science and Engineering,
Ss. Cyril and Methodius University, Rugjer Boshkovik 16,
PO Box 393, Skopje, Macedonia
{mitrevski.blagoj,
petreski.viktor.l}@students.fnki.ukim.mk,
biljana.stojkoska@fnki.ukim.mk

² Department of Intelligent Systems, Jozef Stefan Institute,
Jamova 39, Ljubljana, Slovenia
martin.gjoreski@ijs.si

Abstract. As activity recognition becomes an integral part of many mobile applications, its requirement for lightweight and accurate techniques leads to development of new tools and algorithms. This paper has three main contributions: (1) to design an architecture for automatic data collection, thus reducing the time and cost and making the process of developing new activity recognition techniques convenient for software developers as well as for the end users; (2) to develop new algorithm for activity recognition based on Long Short Term Memory networks, which is able to learn features from raw accelerometer data, completely bypassing the process of generating hand-crafted features; and (3) to investigate which combinations of smartphone and smartwatch sensors gives the best results for the activity recognition problem, i.e. to analyze if the accuracy benefits of those combinations are greater than the additional costs for combining those sensors.

Keywords: Smartphone · Smartwatch · Activity recognition

1 Introduction

In the era of pervasive and ubiquity computing, various applications of activity recognition are evident in many real-life, human-centric problems such as eldercare, healthcare, sports etc. [1]. In most developed countries, demographical trends tend toward more and more elderly people, usually left to their own means in receiving healthcare and other services. The effects of these trends are dramatic on public and private healthcare, as well as on the individuals themselves. Therefore, it is economically and socially beneficial to enhance prevention, by shifting from a centralized, expert-driven model to one that permeates home environments, with focus on advanced homecare services dedicated to personal medical assistance [2].

Regular daily exercises reduce the risk and progression of chronic diseases and improve functional abilities, cardiorespiratory fitness and metabolic health in patients with frequent diseases, such as cardiovascular, lung, and neurodegenerative diseases. With the stimulation of physical activity, the risks for these conditions can be greatly reduced. In particular, [3, 4] showed the effect of physical activity on coronary heart diseases and on the risk of hypertension known as high blood pressure. Additionally, in [5] it is confirmed the effect of the physical activity on diabetes, while [6] prescribes exercise therapy against diabetes. The benefits from regular physical activities for healthy population include reductions in body weight and fat, resting heart rate, increased high density lipoprotein cholesterol and an improved maximal oxygen uptake.

Health benefits associated with physical activity depend on the activity duration, intensity and frequency, therefore it is important to monitor and distinguish the physical activities. Activity recognition (AR) attempts to recognize the actions of an agent in environment from sequence of observations. AR has the potential to address the emerging health conditions such as obesity, heart conditions, diabetes, etc., since physical inactivity is the main factors for those conditions or at least strongly coupled with them. Additionally, human activity recognition can help to develop patient recovery trainings or even provide early detection of diseases, strokes, falls, etc.

Therefore, we can identify three main advantages of AR: (i) early detection of falls and other abnormalities in elderly; (ii) help in the process of recovery after an accident/stroke and (iii) prevention of diseases.

The in-depth data analyses can lead to a broader range of societal challenges. However, most of the studies that investigate the benefits from different physical activities are expensive and require complex process of monitoring. In most epidemiological studies, participants are equipped with special sensors, therefore, the cohort size is limited, and the conclusions cannot be simply mapped to broader population [7]. It would be more convenient more participants to be included in the studies, but without financial implications by means of hardware requirements and manual data analyses. In such scenarios, a common framework is needed that is capable not only to gather data from many participants, but also to provide automatic activity recognition.

This paper has three main goals: (1) to design a system architecture and organization for activity recognition using smartphones and smartwatches as sensor devices which gather data and recognize activities, along with a remote cloud system which is in charge of training and improving the models for recognition; (2) to develop a new lightweight algorithm for activity recognition that is easily implementable for smartphone and smartwatch applications; and (3) to evaluate the accuracy of the algorithm on smartphone and smartwatch sensors and to test which sensor combination gives the best results for the activity recognition problem, i.e. if the accuracy benefits of those combinations are greater than the additional costs of combining those sensors.

Our algorithm is based on neural network, i.e. Long Short Term Memory networks. Although this algorithm has been previously used for activity recognition on wearables [8] and smartphones [9], to the best of our knowledge, this is the first research that evaluates the algorithm on data collected from smartwatch sensors.

The remaining sections of this paper are organized as follows. The second section describes and illustrates the architecture for development of AR methods in details. In the third section we describe and test a neural network which can be used as activity

recognition model, as well as the best combination of sensors that outperforms all the others. The paper is concluded in the fourth section.

2 System Architecture for Efficient AR Tools Development

The traditional approach for development new AR tools and techniques includes collecting sensors data to be further used for training and testing. The process of data collection can be made in a special laboratory conditions, or under field conditions. In both cases, all data are stored in a central server and the process of developing AR techniques is made offline.

In the first case, participants equipped with sensors are guided to perform particular activities for a relatively short period. Therefore, most of the collected datasets are with small number of participant, which is eventually leading to development of inconvenient models.

In the second case, participants are wearing the sensors during their normal daily life, reporting the activities in an activity diary. Typical sensors are accelerometers attached to an elastic belt and placed at the hip or at the dominant ankle. The main drawback of this data collection scenario is the lack of accuracy with respect to time spent performing particular activity. Additionally, wearing the sensors for a long time can even reduce participants' compliance to take part in the study [7]. Still, the approaches from the literature that investigate the AR accuracy on data collected under field condition are satisfying, performing almost equally good as on data collected under field conditions. However, this is expensive, labor and time consuming task for the application developers.

In this section we will describe a system architecture that will overcome both problems, i.e. the development of AR tools would be made online on large datasets collected under field condition. The system we purpose consists of three main parts: server (cloud), normal users and users-contributors.

- The server (cloud) is a central part in the architecture. Its function is to store data, build AR models, update AR models and distribute them to the users.
- Normal users get the AR models from the server (cloud) on their wearable (smartphone). They can contribute only by sending their experience back to the server (cloud).
- Users contributors sent labeled data to the server (cloud), so they contribute to the process of data collection. Additionally, they can act as normal users.

The contributors provide labeled data to be used for creating new AR models or improving the existing ones. Before a specific activity, the contributor can turn her device to “contribute mode” to record labeled data to be sent to the server-side for further processing. Data transferring to the server (cloud) should be done in real-time or near real-time. To save energy for data transmission, different techniques can be used for data reduction, like delta compression (for near real-time transfer) or data prediction based on dual prediction scheme [10, 11]. Additionally, some application specific heuristics can be investigated for this purpose, like sending data only when the phone is connected to the charger, or if the available battery is above a predefined threshold (Fig. 1).

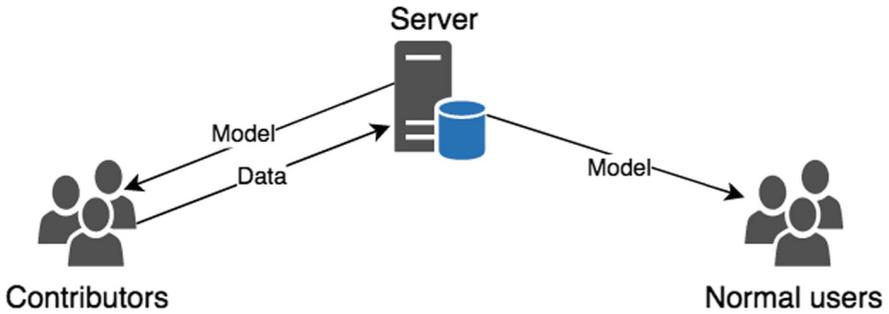


Fig. 1. General framework of the system for online data collection and processing

The architecture we propose is not only suitable for applications based exclusively for smartphone sensors, but can be extended to systems that use different wearables. In this case, the sensors from the wearables send data measurements to the smartphone, which acts as a gateway or hub to retransmit the data to the server (cloud). In this case, it is important for both devices (smartphone and wearable) to be able to communicate using common protocols with low energy requirement, like Bluetooth Low Energy (BLE), etc. Although there are many different protocols for smart devices to transfer data (ZigBee, Z-Wave, Insteon, etc.), the state-of-the-art smartphones are not supporting them [12], therefore this scenario is usually not feasible.

The need for such architecture is not new, but previously it was not feasible since most of the tools for AR require in-depth data analyses usually performed by a team that includes both data scientists and experts in this field to be jointly involved in the process of generating hand-crafted features. Recently, there have been new tools that include automated feature engineering techniques to extract features from the raw accelerometer readings and to select a subset of the most significant features.

3 Designing Accurate and Lightweight Algorithms for AR

The activity recognition at the present-day mainly is sensor-based, implemented with the help of smartphones and wearable devices acting like wearable sensors and computational (recognition) devices. In [13], artificial intelligence (AI) techniques are used to develop daily activity reminders for elders with memory impairments. Moreover, in [14] the authors built abnormal human-activity detection models which can be used to detect and notify of abnormal behavior and early detection of dementia.

Body-worn sensors can be used in sports and physical activities in order to assess and improve the overall sport performance and fitness. In [15] the authors achieved to learn the daily activity pattern of the users and assess the daily energy expenditures in order to help users improve their lifestyle. There are also commercial devices for monitoring sport activities such as Nike + [16] sensor which is placed inside a shoe to keep track (duration) of running and jogging exercises. If connected to a smartphone application, it enables the user to set training goals or to challenge friends.

For the process of activity recognition, different approaches have been proposed in the literature, ranging from simple models to complex neural networks. For instance, [17] examines techniques such as Hidden Markov model, the conditional random field (CRF), the skip-chain CRF, etc., for building activity models. The authors in [18] collected multimodal sensor data, extracted features from their dataset and then employed Support Vector Machine (SVM) on the features. Furthermore, [19, 20] proposed neural networks as classifiers using generated features from the data. As improvement, [21] used Convolutional neural networks (CNN) where the model itself automatically extracts features from the raw sensor data without the need of human expert with prior knowledge of the field to generate handcrafted features. Finally, in [22] the authors suggested a recurrent neural model superior to other neural network models. This model is able to capture temporal correlations from the sensor data, therefore it is applicable to wide range of problems while providing acceptable accuracy. In [7], automated feature engineering technique is performed to extract features from the raw accelerometer readings from epidemiological studies, and four machine learning algorithms are used for classification, showing that only one accelerometer is sufficient for accurate activity recognition.

Apart from the accuracy, other major problems in deploying AR models and applications is the computational cost and time complexity of the algorithm, since it should operate in real-time on devices with limited energy [23–25]. Even in the modern smartphones with performances comparable to those of the computers, the power remains a challenging problem, since battery technology has not kept pace with information and communication technologies. Other issues regarding energy consumption is sampling frequency, as it is an important parameter for the accuracy of the algorithm.

In this section we describe the dataset used in our study, the model we developed for activity recognition, as well as the results of our analyses.

3.1 The Dataset

We use the AR dataset [26, 27] which consists of around 9 million entries, 4 million accelerometer data entries and 4 million gyroscope entries recorded in laboratory conditions. There are recordings of 9 users performing the following activities: ‘Biking’, ‘Sitting’, ‘Standing’, ‘Walking’, ‘Stair Up’ and ‘Stair down’ while data is recorded via two embedded sensors: accelerometer and gyroscope. Four smartwatches (two LG and two Samsung Galaxy Gear) and eight smartphones (two Samsung S3 mini, two Samsung S3, two LG Nexus 4 and two Samsung S+) were used. The data was split into four subsets: smartphone accelerometer data, smartwatch accelerometer data, smartphone gyroscope data and smartwatch gyroscope data.

3.2 Long Short Term Memory Neural Network

Generally, AR techniques first segment the time series data with sliding windows, then apply signal processing and statistical methods for feature extraction from the raw accelerometer measurements and then train machine learning algorithms for classification of different activities.

Among many techniques for activity recognition from the literature, we investigated a neural network based approach known as Long Short Term Memory (LSTM) networks. The main advantages of LSTM can be summarized as: (i) easily implementable on mobile applications; (ii) outperforms other approaches from the literature by means of accuracy; and (iii) robust enough to perform almost equally good on data collected under field conditions as on data collected in a controlled environment [8, 9].

LSTM network as a deep learning system appropriate for temporal modeling was initially proposed by Hochreiter [28] and later improved in 2000 by Gers [29]. It has shown improvements over Deep Neural Networks for speech recognition problem [30]. Since 2016, LSTM became integral part of many applications and services delivered by Google, Microsoft and Apple, including personalized speech recognition on the smartphone [31] and gesture typing decoding [32].

LSTM networks are a special type of neural networks that remember information from further back in the past. Given a sequence of inputs $X = \{x_1, x_2, \dots, x_n\}$, LSTM associates each time step with an input gate, memory gate and output gate, denoted respectively as i_t, f_t and o_t . The information from the past is remembered using the state vector c_{t-1} . The forget gate decides how much of the previous information are going to be forgotten. The input gate decides how to update the state vector using the information from the current input. The l_t vector consists of the information from the current input added to the state. Finally, the output gate decides what information to output at the current time step. This process is formalized as in (1),

$$\begin{aligned}
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t]) \\
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t]) \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t]) \\
 l_t &= \tanh(W_l \cdot [h_{t-1}, x_t]) \\
 c_t &= f_t \cdot c_{t-1} + i_t \cdot l_t \\
 h_t &= o_t \cdot \tanh(c_t)
 \end{aligned} \tag{1}$$

where W_i , W_f , W_o and W_l have dimensions $D \times 2N$, D is the number of memory cells and N is the dimension of the input vector. These matrices represent the parameters of the network. LSTM is local in space and time since its computational complexity per time step and weight is $O(1)$ [28].

As an algorithm for learning models for the task of activity recognition we use LSTM Network implemented in TensorFlow [33]. For every task we tuned the parameters (learning rate, hidden layers, structure of the network) to obtain optimal results. Our basic model consists of two fully connected and two LSTM layers with 64 units each and we use L2 regularization to avoid overfitting. We train the model for 70 epochs.

3.3 Experiments

First the data was split into training and testing sets. Because our dataset contains entries of 9 users, we separated the data from two users and we used it for testing while the remaining data was used for training. With this kind of splitting we avoid

overfitting on user specific data and the results are unbiased. Before feeding the data to the models, we do sliding window of size 200 and step of 50.

The experimental results are presented in Fig. 2. The bars represent the sensor combination used as input to the models. It can be seen that in general the accuracy is higher when the smartphone sensors’ data is used. The highest accuracy of 94% is achieved for the combination accelerometer-gyroscope form the smartphones’ sensors.

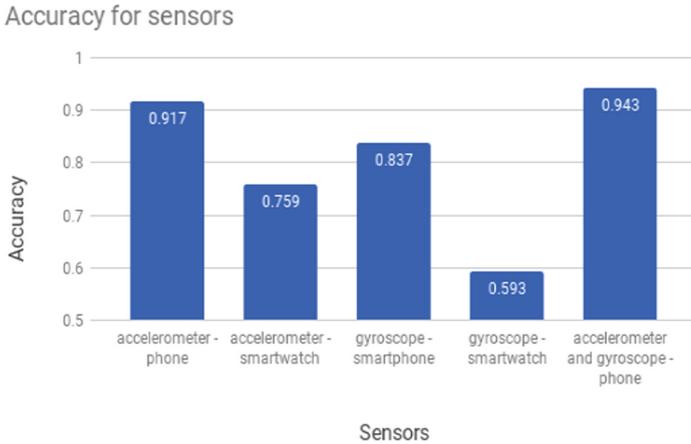


Fig. 2. Accuracy for each sensor combination

Figures 3 and 4 present the confusion matrices for the accelerometer and for the gyroscope data. The rows represent the true class and the columns represent the predicated class. From the confusion matrices it can be seen that misclassifications from the gyroscope phone data and the misclassifications from the accelerometer phone data are contradictory. For example, from Fig. 3 can be seen that the models that use only acceleration data, mostly mix the classes Sitting and Standing. On the other hand, it can be seen in Fig. 4 that the models that use only gyroscope data mix the classes

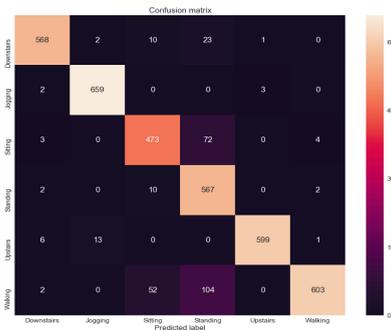


Fig. 3. Confusion matrix for phone accelerometer data

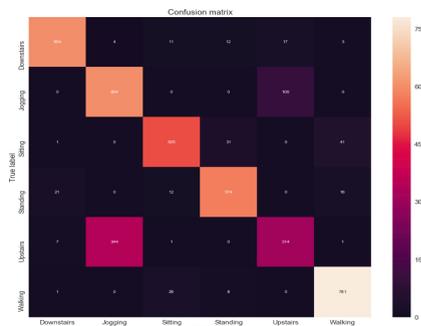


Fig. 4. Confusion matrix for phone gyroscope data

Jogging and Upstairs. To exploit the model variability, we used the accelerometer and the gyroscope data for building the final models.

4 Conclusion

Activity recognition is integral part of many wearable devices, therefore development new tools and algorithms will remain a challenging problem for the research community in the next few years. The process of data collection is usually expensive labor task that is time consuming for both developers of wearable applications and participants that assist in the process of data collection. In this paper we propose a system architecture that can decrease the time to develop new and more accurate methods and tools for activity recognition.

We develop a LSTM technique that performs accurate AR using smartphone and smartwatch sensors. From our experiments, it has been proved that it is better to combine the accelerometer and the gyroscope sensors from the smartphone in order to increase the accuracy of the model, since models that use smartphone acceleration data make different misclassification errors compared to the models that use smartphone gyroscope data.

Acknowledgment. This work was partially financed by the Faculty of Computer Science and Engineering, University “Ss. Cyril and Methodius”, Skopje, Macedonia.

References

1. Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., Havinga, P.: Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: a survey. In 2010 23rd International Conference on Architecture of Computing Systems (ARCS), pp. 1–10. VDE (2010)
2. Risteska Stojkoska, B., Trivodaliev, K., Davcev, D.: Internet of things framework for home care systems. *Wireless Commun. Mobile Comput.* **2017** (2017)
3. Weinstein, A.R., et al.: The joint effects of physical activity and body mass index on coronary heart disease risk in women. *Arch. Intern. Med.* **168**(8), 884–890 (2008)
4. Hu, G., Barengo, N.C., Tuomilehto, J., Lakka, T.A., Nissinen, A., Jousilahti, P.: Relationship of physical activity and body mass index to the risk of hypertension: a prospective study in Finland. *Hypertension* **43**(1), 25–30 (2004)
5. Haapanen, N., Miilunpalo, S., Vuori, I., Oja, P., Pasanen, M.: Association of leisure time physical activity with the risk of coronary heart disease, hypertension and diabetes in middle-aged men and women. *Int. J. Epidemiol.* **26**(4), 739–747 (1997)
6. Jia, Y.: Diabetic and exercise therapy against diabetes mellitus. In: ICINIS 2009. Second International Conference on Intelligent Networks and Intelligent Systems, 2009, pp. 693–696. IEEE (2009)
7. Zdravevski, E., Stojkoska, B.R., Standl, M., Schulz, H.: Automatic machine-learning based identification of jogging periods from accelerometer measurements of adolescents under field conditions. *PLoS ONE* **12**(9), e0184216 (2017)
8. Ordóñez, F.J., Roggen, D.: Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* **16**(1), 115 (2016)

9. Milenkoski, M., Trivodaliev, K., Kalajdziski, S., Jovanov, M., Stojkoska, B.R.: Real time human activity recognition on smartphones using LSTM Networks. In: MIPRO (2018)
10. Stojkoska, B.R., Nikolovski, Z.: Data compression for energy efficient IoT solutions. In: 2017 25th Telecommunication Forum (TELFOR), pp. 1–4. IEEE (2017)
11. Stojkoska, B.R., Trivodaliev, K.: Enabling internet of things for smart homes through fog computing. In: 2017 25th Telecommunication Forum (TELFOR), pp. 1–4. IEEE (2017)
12. Stojkoska, B.L.R., Trivodaliev, K.V.: A review of internet of things for smart home: challenges and solutions. *J. Cleaner Prod.* **140**, 1454–1464 (2017)
13. Pollack, M.E., et al.: Autominder: an intelligent cognitive orthotic system for people with memory impairment. *Robot. Auton. Syst.* **44**(3), 273–282 (2003)
14. Yin, J., Yang, Q., Pan, J.J.: Sensor-based abnormal human-activity detection. *IEEE Trans. Knowl. Data Eng.* **20**(8), 1082–1090 (2008)
15. Long, X., Yin, B., Aarts, R.M.: Single-accelerometer-based daily physical activity classification. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, EMBC 2009, pp. 6107–6110. IEEE (2009)
16. Official website for Nike + . <http://www.nikeplus.com/>
17. Kim, E., Helal, S., Cook, D.: Human activity recognition and pattern discovery. *IEEE Pervasive Comput.* **9**(1), 48 (2010)
18. Liu, X., Liu, L., Simske, S. J., Liu, J.: Human daily activity recognition for healthcare using wearable and visual sensing data. In: IEEE International Conference on Healthcare Informatics (ICHI), 2016, pp. 24–31. IEEE (2016)
19. Yang, J.Y., Wang, J.S., Chen, Y.P.: Using acceleration measurements for activity recognition: an effective learning algorithm for constructing neural classifiers. *Pattern Recognit. Lett.* **29**(16), 2213–2220 (2008)
20. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. *ACM SigKDD Explor. Newslett.* **12**(2), 74–82 (2011)
21. Zeng, M., et al.: Convolutional neural networks for human activity recognition using mobile sensors. In: 2014 6th International Conference on Mobile Computing, Applications and Services (MobiCASE), pp. 197–205. IEEE (2014)
22. Murad, A., Pyun, J.Y.: Deep recurrent neural networks for human activity recognition. *Sensors* **17**(11), 2556 (2017)
23. Anguita, D., Ghio, A., Oneto, L., Llanas Parra, F.X., Reyes Ortiz, J.L.: Energy efficient smartphone-based activity recognition using fixed-point arithmetic. *J. Univ. Comput. Sci.* **19**(9), 1295–1314 (2013)
24. Gordon, D., Czerny, J., Miyaki, T., Beigl, M.: Energy-efficient activity recognition using prediction. In: 2012 16th International Symposium on Wearable Computers (ISWC), pp. 29–36. IEEE, June 2012
25. Oneto, L., Ortiz, J.L., Anguita, D.: Constraint-aware data analysis on mobile devices: an application to human activity recognition on smartphones. In: Adaptive Mobile Computing, pp. 127–149 (2017)
26. Stisen, A., et al.: Smart devices are different: assessing and mitigating mobile sensing heterogeneities for activity recognition. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems (SenSys 2015), Seoul, Korea (2015)
27. UCL link to the dataset. <https://archive.ics.uci.edu/ml/datasets/Heterogeneity+Activity+Recognition>
28. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
29. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to forget: continual prediction with LSTM. *Neural Comput.* **12**(10), 2451–2471 (2000)

30. Sainath, T.N., Vinyals, O., Senior, A., Sak, H.: Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (Icassp), pp. 4580–4584. IEEE (2015)
31. McGraw, I., et al.: Personalized speech recognition on mobile devices. In: 2016 IEEE International Conference on Acoustics, Speech And Signal Processing (ICASSP), pp. 5955–5959. IEEE (2016)
32. Alsharif, O., Ouyang, T., Beaufays, F., Zhai, S., Breuel, T., Schalkwyk, J.: Long short term memory neural network for keyboard gesture decoding. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2076–2080. IEEE (2015)
33. TensorFlow Homepage. <https://www.tensorflow.org/>