



# Deep learning—Accelerating Next Generation Performance Analysis Systems? †

Heike Brock

Honda Research Institute Japan Co., Ltd., 8-1 Honcho Wako-shi, Saitama 351-0188, Japan;  
h.brock@jp.honda-ri.com; Tel.: +81-48-462-5219

† Presented at the 12th Conference of the International Sports Engineering Association, Brisbane, Queensland, Australia, 26–29 March 2018.

Published: 23 February 2018

**Abstract:** Deep neural network architectures show superior performance in recognition and prediction tasks of the image, speech and natural language domains. The success of such multi-layered networks encourages their implementation in further application scenarios as the retrieval of relevant motion information for performance enhancement in sports. However, to date deep learning is only seldom applied to activity recognition problems of the human motion domain. Therefore, its use for sports data analysis might remain abstract to many practitioners. This paper provides a survey on recent works in the field of high-performance motion data and examines relevant technologies for subsequent deployment in real training systems. In particular, it discusses aspects of data acquisition, processing and network modeling. Analysis suggests the advantage of deep neural networks under difficult and noisy data conditions. However, further research is necessary to confirm the benefit of deep learning for next generation performance analysis systems.

**Keywords:** machine learning; deep learning; specialized activity recognition; motion performance analysis; wearable sensor data; data processing

## 1. Introduction

Insights into aerodynamics and biomechanics largely supported the improvement of human performance and motor skills during the last century [1]. However, decades of extensive exploration and frequent deployment exhausted the possibilities of further kinematic performance improvement. This focuses attention on the implementation of novel technologies as for example smart equipment and surfaces [2] or augmented intelligent coaching software [3]. Especially the latter possesses large potential due to the availability and low cost of wearable devices that facilitate the acquisition and accessibility of in-field movement data. Here, the most common sensors are Inertial Measurement Units (IMUs). They were used in various applications like swimming [4], ski jumping [5], cross-country skiing [6], trampolining [7] or table-tennis [8]. Within the last years, increasing attention is furthermore paid to less obtrusive bio-sensors woven into garment [9] or even directly implanted in the skin [10]. Common to all of those devices is their very traditional strategy of deployment: data is collected and transmitted to a main processing system that retrieves relevant information on the basis of feature extraction and similarity or variance measurements or basic machine learning. To implement meaningful performance analysis systems, one has to pose the question whether these computations can ensure the utilization of all significant and relevant motion information.

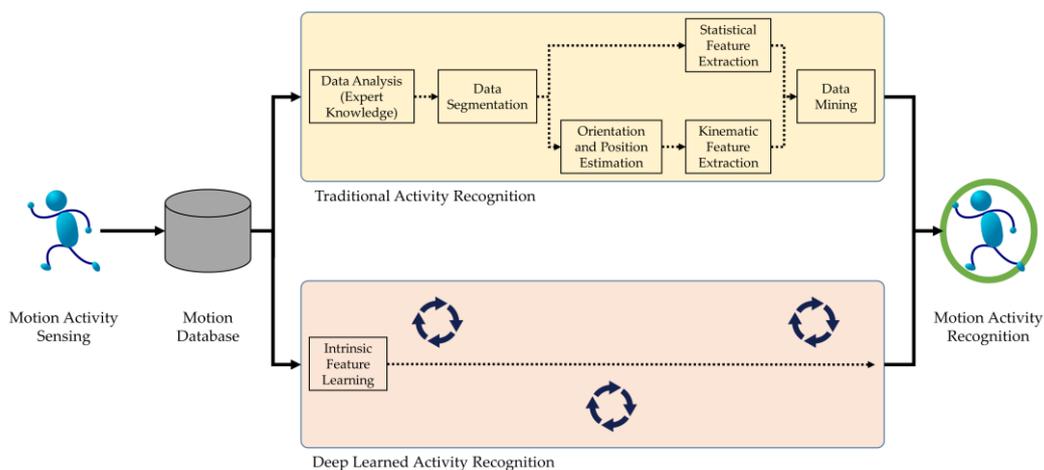
Different computing domains suggest that sensor data analysis software has by far not reached its maximum level of quality yet: for example in image recognition, the introduction of deep neural network architectures improved system error rate by more than 10% [11], and was constantly decreased since then. Similar results are reported for natural language translation and speech recognition. This draws the conclusion that deep learned machine intelligence could also play a major

role for the successful implementation of reliable and intelligent automated training systems. But what is necessary to apply deep learning to high-performance motion data? And how exactly can deep learning be applied in respective trainings systems? Based on reported works utilizing biological time-series data, answers to the previous questions and strategies for the implementation of deep learning motion performance analysis systems shall be given in this survey.

## 2. The Benefit of Deep Learning

Traditional activity recognition systems are based on defined processing steps that aim to control and restrict the multi-variate (and typically high-dimensional) wearable sensor measurement data [12]. The common flow of those steps can be summarized as (1) basic data analysis, (2) data segmentation into activity segments, (3) extraction of meaningful feature representations per data segment and (4) the retrieval of relevant motion information using similarity or variance measures and machine learning. Implementation of those major processing steps is subject to manual, hand-crafted algorithms and data transformations that depend on the specific characteristics of the given application. Consequently, it can be cumbersome to determine those data properties that work best for a given task, and often specific domain knowledge is necessary to ensure meaningful data processing. Common feature extractors range from simple statistical or spectral descriptions as obtained using Wavelet and Fourier transformations to kinematic-induced features such as body pose and body joint position [5]. Using IMUs or other wearable measurement instrumentation, the latter can generally not be determined directly and need to be estimated by further processing functions such as the Kalman Filter. Although kinematic features are close to the biomechanic specifications, most practitioners therefore prefer simple statistical feature extractors of fast implementation and computation time.

As one can see, traditional activity recognition is a very heuristic procedure that largely relies on the idea that the chosen feature extractors are able to display variance between data streams for subsequent data mining. To prevent information loss caused by such work flow, it appears reasonable to utilize deep neural networks: given a sufficient amount of training data, all relevant characteristics of the underlying data are learned intrinsically by stacking layers of transformed data representations [13]. The stacked (deep) layers remain hidden within the neural network architecture and commonly reduce dimensionality of the fundamental data structure to a more compact and discriminant data representation. This procedure guarantees independence from further data augmentation or feature extraction steps (Figure 1). Every hidden layer is learned individually based on the input training data, whereas often only the higher level layers need to be retrained for different training input of similar data structure [14].



**Figure 1.** The work flow of a traditional human activity recognition system can be reduced considerably when using deep learning. Specific expert or domain knowledge is not necessary to learn subtle connections within the data.

Numerous variations of deep network architectures have been developed over the last decade, with most of them being variations of either Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs). RNNs are commonly employed with sequential data such as in machine translation and base on probabilistic calculus. CNNs on the other hand utilize two- or three-dimensional data and were initially developed for image recognition. Both network architectures were already employed in activity recognition tasks of human motion data and also appear useful for future sport performance analysis.

### *Deep Learning in Human Activity Recognition*

Deep neural network architectures are utilized for the recognition of human motion data from wearable sensor measurements since 2012. Multiple studies demonstrated the advantage of deep learning over conventional feature-engineered systems under simple data sets of low-performance motion sequences like jogging, running and jumping with both RNNs and CNNs. For example in [15], a RNN architecture with long short term memory achieved 95% of recognition accuracy on a set of accelerometer data collected with a mobile phone. In a more extensive comparison of multiple network models, RNNs and CNNs were both shown to constantly achieve better F1 (meaning precision and recall) scores than shallow networks [16] on multiple public activity recognition data sets. A CNN architecture using one-dimensional convolution along the time domain was furthermore reported to consistently outperform the best shallow baseline networks by more than 5% [17]. Most recently, even higher accuracy was reported by the introduction of a deep recurrent convolutional network, which could outperform any of the previous networks by 4% on average [18].

## **3. Deep Learning High-Performance Sports Data**

Various research is reported that deployed deep learning network architectures for recognition or prediction of actions in sport scenarios. However, most of these works developed network models for the use of sport video recordings. To date, four performance analysis systems are known that utilized non-image data to learn deep neural network architectures. In concrete, these are one system for the prediction of basketball trajectories [19], one system for stroke recognition in beach volleyball [20], one system for the automatic judging of ski jumping [21] and one system for the classification of cross-country skiing gears [22]. Since the latter three systems are based on wearable sensor data, only their system specifications shall be discussed in more detail in the following, whereas the beach volleyball recognition system shall be referred to as BV, the ski jumping judging system as SJ and the cross-country classification system as XCS.

### *3.1. System and Data Requirements*

As discussed in Section 2, deep neural network architectures are able to learn subtle connections within a set of training data without domain knowledge and information loss. However, this data-independence requires a large amount of data with ideally up to then thousands of training samples. The collection of such masses of data can be a very difficult task under common constraints of experimental field measurements like restricted economic resources, access to participating athletes or sporting venues. For the design and training of a suitable network architecture, it is furthermore beneficial to have access to good computing hardware respectively high power graphical processing units: the most accurate network models are seldom found immediately. Instead, basic network architectures are retrained multiple times with varying properties to determine the parameters best suited for the given problem. Once a meaningful network is trained on the other hand, retrieval of the desired information is fast and relevant data provided within seconds.

Difficulties in system implementation caused by the need of large data sets for network training reflect within the present studies. In SJ, only 88 ski jumps were available for network learning and evaluation. In BV, performances of a much larger data set of approximately 4300 motion actions were classified. Although large from a sensing and measurement perspective, this data collection is still small as compared to the data sets used in image or text classification tasks. The authors of XCS claim

to have used 416.737 data recordings, which appears to be a huge collection of skiing data. However, one can suspect the total number of recordings to constitute the absolute number of frame counts here, and the total number of data captures should be expected to be much smaller accordingly.

### 3.2. Deployed Architectures

Research suggests that different network models should be applied to different motion types: recurrent networks were shown to outperform convolutional networks on short and temporally ordered motion sequences, whereas CNNs worked better under long and repetitive actions [16]. For future training software, this would mean that RNNs should be favored for acyclic movements like jumping and throwing and CNNs for cyclic movements like rowing, skiing or swimming. The BV and SJ studies do not follow this recommendation and evaluate networks from a more general perspective. As a pilot study within the sports activity recognition domain, BV uses one CNN architecture and compares its performance to a variety of shallow baseline networks. To recognize errors without contextual dependence, the authors of SJ designed three different types of CNNs and compare their error classification results with two shallow baseline networks. In XCS, both RNNs and CNNs of different model design and parameters per layer were learned and evaluated, providing a more comprehensive evaluation of the different network types with sport performance data (Table 1).

**Table 1.** Deep network architectures and the number of their different model parameters evaluated for use with high-performance sports data.

	# Tested CNNs	# Conv. Filter Variations	# Tested RNNs	# Memory Cell Variations	# Baseline Networks
BV	1	none	none	n.a.	6
SJ	3	none	none	n.a.	2
XCS	2	3	3	3	1

### Performance Results

The performance and quality of deep learning systems largely depends on the choice of their network model and its internal parameters. As previously mentioned, it is therefore common to train a large number of multiple network variations until a best solution is found to obtain highly accurate performance evaluation systems. BV does not discuss any model fine-tuning and only evaluates the chosen CNN with a fixed number of convolution and pooling filter. Within the (most likely) optimized network design, significant improvement in classification accuracy of up to 16% could be achieved as compared to the baseline networks, indicating the potential of deep performance analysis systems. To exclude any bias caused by network optimization processes, SJ evaluates the deep networks without prior model fine-tuning. Instead, it only evaluates three CNN architectures under one basic design. As a result, accuracy of the convolutional networks was not significantly improved for errors that could already be classified sufficiently well with the shallow baseline networks. However, results appear promising for those errors that could not be reliably classified by the shallow networks beforehand. Here, recognition accuracy could be improved by approximately 10%. As stated by the authors, especially those errors are subject to large variations in execution as well as bias in the ground truth annotation, suggesting that deep networks might be able to learn more distinct features under noisy and erroneous data. However due to the small number of sample data, this finding cannot be generalized without further investigation.

In contrast to the more general studies, the purpose of XCS was to determine the optimal network model within a number of network variations for the given classification problem. Therefore, the best working network models for both RNN and CNN were of high accuracy (2.4% and 1.6% of error rate) and perform considerably better than the baseline method (14.6% of error rate). In contrary to previous research, RNNs achieve higher accuracy than CNNs for the cycling skiing application. However, this might be a result of the chosen study and system design and should

not be generalized yet: for example, only two different types of motion (respectively skiing gears) had to be classified, which makes it difficult to evaluate the general relevance of the study for future system implementations.

#### 4. Impacting Performance Analysis Systems?

To date, present studies that utilize wearable sensor data cannot be considered sufficient to allow for an extensive evaluation of the benefit of deep learning. This is mainly due to the following two reasons. Firstly, only activity recognition tasks have been investigated. To enable a valid conclusion, further tasks that might be important for the training of motor skills should be evaluated (e.g., the prediction of movements to intervene erroneous or even dangerous movements or the learning of regression models to obtain numeric values for data display and evaluation). Secondly, current systems only classify a small number of different action categories and movement patterns. Network models learned on general activity recognition data sets seldom contain more than ten different motion classes. Similarly, for the three reported sport performance systems the number of motion classes were ten different volleyball strokes (BV), nine different types of ski jump motion errors (SJ) and two different types of cross-country skiing turns (XCS). Since semantic differences between differing motion actions of sports performances are usually less distinct than in general human activity recognition, it is currently hard to predict how reliable and accurate data-driven performance analysis systems might become in future.

## 5. Conclusions

This work discusses the application of deep learning technologies in future sport performance analysis systems. To date, several investigations are reported that utilize deep neural network architectures on wearable sensor data, whereas the majority of all designed network models are evaluated on general human activity data of simple, everyday movements. Here, results clearly show that deep learning models are capable of improving recognition accuracy as compared to traditional data mining methods of heuristic approach. This suggests that relevant and discriminative characteristics of a motion could be learned that otherwise get lost in the process of feature engineering. As a consequence, one can expect the inclusion of deep neural network architectures to have a significant positive effect on the accuracy and reliability of intelligent training software.

Studies that learned deep networks on sports motion data seem to confirm the previous results. However to date, only a very small number of different network model designs and parameter variations were evaluated. To predict the long-term effect of deep learning, network performance needs to be further evaluated with a higher number of semantically similar motion patterns or a higher number of movement variations as found within motor executions of one single sport performance. Furthermore, it should be emphasized that results of field studies are commonly not as significant or explicit as results of studies that employed data captured under laboratory conditions. This is mostly due to a smaller amount of available training data paired with higher noise and less discriminative data specifications. One of the main challenges in the implementation of deep learned training systems is consequently the provision of numerous and meaningful sensor data: to enable sufficient training samples for the learning of deep neural network architectures, it is necessary to overcome common constraints of data collection such as restricted time and economic resources. Then and only then, deep learning has the chance to bring sport performance analysis systems to a new level of reliability and accuracy to support and enhance every person exercising sports in the future.

**Acknowledgments:** No additional funding has been received in support of this research.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. McGinnis, P. *Biomechanics of Sport and Exercise*; Human Kinetics: Champaign, IL, USA, 2013.
2. Stefanyshyn, D.J.; Wannop, J.W. Biomechanics research and sport equipment development. *Sports Eng.* **2015**, *18*, 191–202.
3. Sigrist, R.; Rauter, G.; Riener, R.; Wolf, P. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychon. Bull. Rev.* **2013**, *20*, 21–53.
4. Magalhaes, F.A.D.; Vannozzi, G.; Gatta, G.; Fantozzi, S. Wearable inertial sensors in swimming motion analysis: A systematic review. *J. Sports Sci.* **2015**, *33*, 732–745.
5. Brock, H.; Ohgi, Y. Assessing Motion Style Errors in Ski Jumping Using Inertial Sensor Devices. *IEEE Sens. J.* **2017**, *17*, 3794–3804.
6. Fasel, B.; Favre, J.; Chardonnes, J.; Gremion, G.; Aminian, K.; An inertial sensor-based system for spatio-temporal analysis in classic cross-country skiing diagonal technique. *J. Biomech.* **2015**, *48*, 3199–3205.
7. Helten, T.; Brock, H.; Müller, M.; Seidel, H.P. Classification of trampoline jumps using inertial sensors. *Sports Eng.* **2011**, *14*, 155–164.
8. Blank, P.; Groh, B.H.; Eskofier, B.M. September. Ball speed and spin estimation in table tennis using a racket-mounted inertial sensor. In Proceedings of the 2017 ACM International Symposium on Wearable Computers, Maui, HI, USA, 11–15 September 2017; pp. 2–9.
9. Yang, G.; Xie, L.; Mäntysalo, M.; Zhou, X.; Pang, Z.; Da Xu, L.; Kao-Walter, S.; Chen, Q.; Zheng, L.R. A health-iot platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box. *IEEE Trans. Ind. Inform.* **2014**, *10*, 2180–2191.
10. Olivo, J.; Carrara, S.; De Micheli, G. Energy harvesting and remote powering for implantable biosensors. *IEEE Sens. J.* **2011**, *11*, 1573–1586.
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2012; pp. 1097–1105.
12. Bulling, A.; Blanke, U.; Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.* **2014**, *46*, 436–444.
13. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
14. Cook, D.; Feuz, K.D.; Krishnan, N.C. Transfer learning for activity recognition: A survey. *Knowl. Inf. Syst.* **2013**, *36*, 537–556.
15. Chen, Y.; Zhong, K.; Zhang, J.; Sun, Q.; Zhao, X. LSTM Networks for Mobile Human Activity Recognition. In Proceedings of the 2016 International Conference on Artificial Intelligence: Technologies and Applications, Bangkok, Thailand, 24–25 January 2016.
16. Hammerla, N.Y.; Halloran, S.; Ploetz, T. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv* **2016**, arXiv:1604.08880.
17. Yang, J.; Nguyen, M.N.; San, P.P.; Li, X.; Krishnaswamy, S. Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition. In Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015; pp. 3995–4001.
18. Ordóñez, F.J.; Roggen, D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* **2016**, *16*, 115.
19. Shah, R.; Romijnders, R. Applying deep learning to basketball trajectories. *arXiv* **2016**, arXiv:1608.03793.
20. Kautz, T.; Groh, B.H.; Hannink, J.; Jensen, U.; Strubberg, H.; Eskofier, B.M. Activity recognition in beach volleyball using a Deep Convolutional Neural Network. *Data Min. Knowl. Discov.* **2017**, *31*, 1678.
21. Brock, H.; Ohgi, Y.; Lee, J. Learning to judge like a human: Convolutional networks for classification of ski jumping errors. In Proceedings of the 2017 ACM International Symposium on Wearable Computers, Maui, HI, USA, 11–15 September 2017; pp. 106–113.
22. Rassem, A.; El-Beltagy, M.; Saleh, M. Cross-Country Skiing Gears Classification using Deep Learning. *arXiv* **2017**, arXiv:1706.08924.

