# Sensor-Based Gait Parameter Extraction With Deep Convolutional Neural Networks

Julius Hannink, Thomas Kautz, Cristian F. Pasluosta, Karl-Günter Gaßmann, Jochen Klucken,
and Bjoern M. Eskofier, *Member, IEEE*

*Abstract*—Measurement of stride-related, biomechanical parameters is the common rationale for objective gait impairment scoring. State-of-the-art double-integration approaches to extract these parameters from inertial sensor data are, however, limited in their clinical applicability due to the underlying assumptions. To overcome this, we present a method to translate the abstract information provided by wearable sensors to context-related expert features based on deep convolutional neural networks. Regarding mobile gait analysis, this enables integration-free and data-driven extraction of a set of eight spatio-temporal stride parameters. To this end, two modeling approaches are compared: a combined network estimating all parameters of interest and an ensemble approach that spawns less complex networks for each parameter individually. The ensemble approach is outperforming the combined modeling in the current application. On a clinically relevant and publicly available benchmark dataset, we estimate stride length, width and medio-lateral change in foot angle up to $-0.15 \pm 6.09$ cm, $-0.09 \pm 4.22$ cm and $0.13 \pm 3.78°$ respectively. Stride, swing and stance time as well as heel and toe contact times are estimated up to $\pm 0.07$, $\pm 0.05$, $\pm 0.07$, $\pm 0.07$ and $\pm 0.12$ s respectively. This is comparable to and in parts outperforming or defining state of the art. Our results further indicate that the proposed change in the methodology could substitute assumption-driven double-integration methods and enable mobile assessment of spatio-temporal stride parameters in clinically critical situations as, e.g., in the case of spastic gait impairments.

*Index Terms*—Convolutional neural networks (CNNs), deep learning, mobile gait analysis, regression, spatio-temporal gait parameters.

## I. INTRODUCTION

A VARIETY of neurological and musculoskeletal diseases affects human gait quality and manifest in specific stride

characteristics. Parkinson's disease (PD), for example, is associated with a reduced stride length, shuffling steps or impaired gait initiation. As reduced gait quality can lead to severe reductions in patient mobility and quality of life [1], it is important to quantify, detect, and treat gait impairments as early as possible.

Objective quantification of gait impairment is based on stride-specific characteristics such as stride length or stride time. These parameters are commonly extracted with the help of several electronic measurement systems, including computerized pressure mats ([2], [3]), optical motion-capture systems [4] or mobile, sensor-based solutions ([5]–[11]). While the first two require a laboratory environment and are limited in availability, the latter is mobile and inexpensive. This renders mobile, sensor-based solutions the primary choice for unobtrusive gait analysis systems.

Choosing this modality though introduces a conflict between the abstract variables of measurement and the readout parameters requested by the users and is as such entangled with the physical constraints in wearable sensing: For instrumented medical healthcare applications, one might be able to measure accelerations and angular rates at a patient's foot by using state-of-the-art inertial sensors. However, the treating physician is not interested in interpreting acceleration signatures for a given stride but rather wants to monitor variables directly related with the situation as for example stride length or heel-strike (HS) angle. The efficient translation of abstract data into context-related knowledge thus is the underlying challenge in all applications of wearable sensors and mobile healthcare technologies.

In the extraction of stride parameters, this challenge is addressed from several perspectives. The majority of methods are based on physical and geometric reasoning to extract spatial gait parameters by using double integration of inertial sensor signals ([6]–[9], [12]). The main limitation regarding this type of approach is the dependence on a zero-velocity phase within each stride that is needed to reinitialize the integration process. In clinical practice, however, this assumption is easily violated [13]. Other approaches aim at driving biomechanical models of the lower extremity with sensor data [10], [11] or apply machine learning in order to extract the parameters of interest [13].

The underlying problem of efficient data to knowledge translation is recently being addressed very successfully in the field of image understanding. Here, data from images is identified to belong to a certain object class [14], translated to captions that describe the image content [15] or used to identify persons based on face recognition [16]. All these applications represent ground-breaking advances in their respective field in terms of
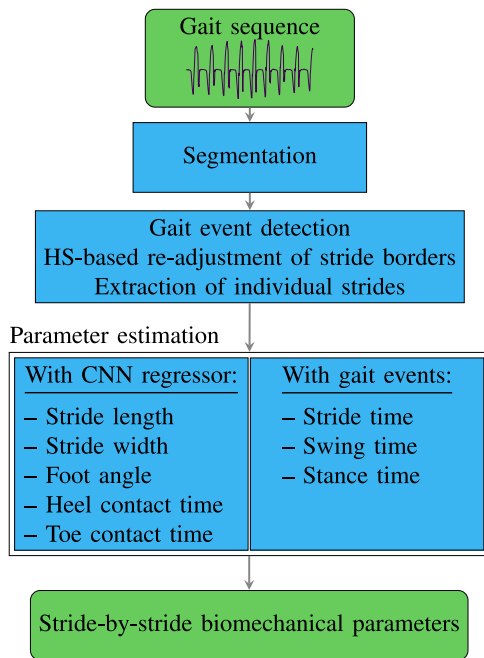
Fig. 1. Conceptual flowchart for applying the proposed system to mobile gait analysis: an in-comming gait sequence is segmented and gait events are identified within each segment. In the second step, individual strides are defined from HS to HS and fed to the biomechanical feature extraction routines. These either compute the timings directly related with the identified gait events or estimate spatio-temporal parameters with a CNN.

recognition rates. The common underlying methodology that allows these achievements is a branch of machine learning called deep learning.

Due to its success in other domains, deep learning is starting to appear in the context of wearable sensing and computing to extract meaningful information from sensor data ([13], [17]–[20]). This particular branch of machine learning is said to have large potential in mobile sensing and computing regarding inference accuracy, robustness or class scaling, which are partly missing from state-of-the-art [21]. Applications of deep learning in wearable sensing and computing are, however, largely focusing on activity recognition ([17]–[20]). To the authors' best knowledge, this and their prior work [13] are the first applications regarding other topics in the field.

In this paper, we present a framework based on deep CNNs and aim at translating the abstract information provided by wearable sensors into context-related expert features requested by the users. The system is trained on a regression task between the sensor data and a set of reference output parameters. Thereby, it extends the authors' prior work [13] that only addresses a single output parameter. A prerequisite for this is a knowledge base, i.e. a collection of wearable sensor data captured in a controlled environment and annotated with the help of a reference system that can directly measure the expert features of interest.

We apply the proposed framework in the context of mobile gait analysis as illustrated in Fig. 1. In doing so, we focus specifically on the extraction of biomechanical stride parameters



Fig. 2. Placement of the inertial sensor and axes definition.

with CNN regression while potential benefits from deep learning approaches to other parts of the pipeline (e.g. segmentation) might be addressed in future work. A total of eight exemplary and stride-specific characteristics are extracted that are clinically relevant as they define gait quality. To this end, two different modeling approaches are compared: a combined model that uses one network architecture to estimate all expert features of interest and an ensemble approach where one neural network is spawned for each output parameter individually. Both models are trained and evaluated on a publicly available and clinically relevant benchmark dataset.

In summary, our main contributions are: (1) A generalisable method for data-driven and integration-free extraction of spatio-temporal gait characteristics, and (2) Technical validation of the proposed method on a clinically relevant and publicly available dataset.

## II. METHODS

### A. Data Collection and Setup

We use a benchmark dataset collected by Rampp *et al.* [7] that is publicly available at https://www5.cs.fau.de/activitynet/benchmark-datasets/digital-biobank/ and briefly described here.

The inertial sensor platform Shimmer2R [22] consisting of a 3D-accelerometer (range $\pm 6$ g) and a 3D-gyroscope (range $\pm 500\,°/s$) was used for data collection. It was attached laterally below each ankle joint (see Fig. 2). In order to avoid gait changes due to different shoe characteristics [23], the same shoe model (adidas Duramo 3) was used by all subjects. Data was captured at 102.4 Hz at a resolution of 12 bit. Simultaneously, validation data was acquired with the well-established pressure mat GAITRite with a spatial resolution of $\pm 1.27$ cm [3].

In total, 116 geriatric inpatients were assessed at the Geriatrics Centre Erlangen (Waldkrankenhaus St. Marien, Erlangen, Germany). Written informed consent was obtained prior to the gait assessment in accordance with the ethical committee of the medical faculty at Friedrich-Alexander University Erlangen-Nürnberg (Re.-No. 4208).

For our study, the annotation on the dataset was extended with additional parameters compared to the annotation reported on by Rampp *et al.* [7]. This was based on positions and timings of the patients' heel and toe as measured by the GAITRite reference system. The spatial parameter set was enlarged to cover not
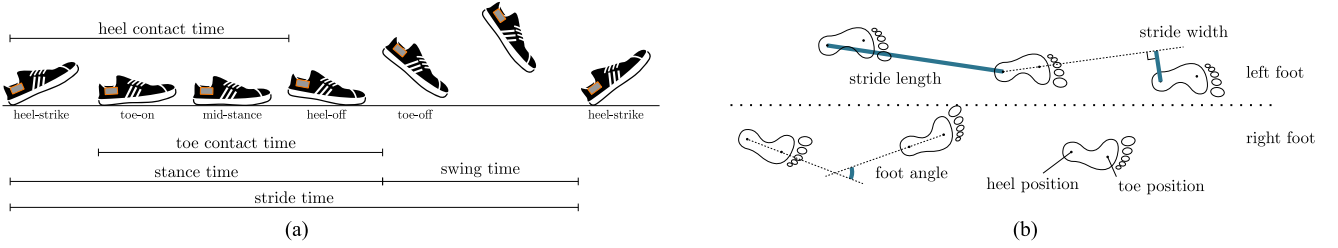
Fig. 3. Definition of temporal (a) and spatial (b) gait parameters based on heel and toe positions and timings measured with the GAITRite reference.

TABLE I
EXTENDED SET OF ANNOTATION PARAMETERS ON THE DATASET
AND THEIR MEAN VALUES, STANDARD DEVIATIONS,
AS WELL AS MINIMAL/MAXIMAL VALUES

| Output Parameter | Unit | Mean ± Std. | [Min, Max] |
|---|---|---|---|
| Stride length | cm | 80.63 ± 23.23 | [20.01, 129.81] |
| Stride width | cm | −1.44 ± 13.29 | [−37.52, 33.03] |
| Foot angle | ° | 0.07 ± 3.49 | [−11.93, 15.86] |
| Stride time | s | 1.23 ± 0.19 | [0.74, 2.06] |
| Swing time | s | 0.37 ± 0.08 | [0.01, 1.05] |
| Stance time | s | 0.85 ± 0.16 | [0.48, 1.65] |
| Heel contact time | s | 0.64 ± 0.14 | [0.16, 1.52] |
| Toe contact time | s | 0.69 ± 0.17 | [0.25, 1.57] |



Fig. 4. Exemplary input signal for a stride defined from HS→HS after preprocessing.

only stride length, but also stride width and change in medio-lateral foot angle. Additionally, heel and toe contact times were added to the list of temporal parameters stride, stance and swing time. Fig. 3 gives an overview on the definitions of temporal and spatial parameters. Stride width was defined as shown in Fig. 3(b) and positive values were measured towards the lateral side of the shoe.

Patients performed an extensive geriatric assessment, described by Rampp *et al.* [7] in detail. For the scope of this paper, we focused on the free walking test over the GAITRite mat at comfortable walking speed instrumented with the inertial sensors. After excluding datasets from 8 patients due to medical reasons (i.e. patients could not complete the measurement protocol), 2 due to inertial sensor malfunction and additional 7 due to measurement errors with the GAITRite system, 99 patients were left for training and evaluation of the proposed method. Compared with Rampp *et al.* [7], reference values for heel and toe contact times could not be computed for two patients.

Gait disorders or fall proneness were diagnosed in 54% of the study population. The other top three diagnoses were heart rhythm disorder (70%), arterial hypertension (69%), and coronary artery disease (41%), which are also associated with gait and balance disorders [24]. In summary, this dataset constitutes a clinically relevant study population both in terms of the number of patients and the presence of unpredictable gait alterations.

Table I gives an overview on the extended set of reference parameters on the dataset and their mean value, standard deviation, as well as their minimal/maximal values.
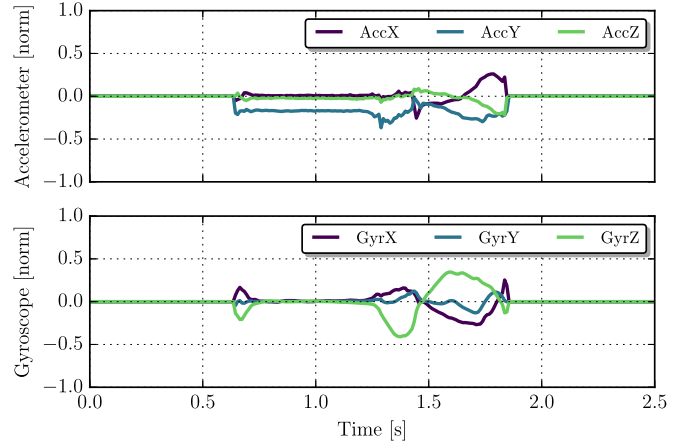
## B. Preprocessing

Before the inertial sensor data is fed to the CNN, we perform a series of preprocessing steps. The segmentation step mentioned in the overview Fig. 1 is already provided by the dataset in our case. Preprocessing therefore includes extraction of annotated strides from the continuous recordings, calibration from raw sensor readings to physical units, coordinate system transformations to align sensor axes on left and right feet, normalization w.r.t. sensor ranges and padding to fixed length of 256 samples per stride to ensure fixed size input to the network.

The system is trained on data segments from HS→HS. This choice of stride definition is beneficial since it does not assume a zero-velocity phase that state-of-the-art double-integration approaches need to reinitialize the integration process. In clinical practice, this assumption is easily violated as, e.g., in the case of spastic gait impairments. However, initial ground contact can still be detected for, e.g., spastic gait patterns and provides a valid segmentation of the signal into strides. This is the type of scenario that we intend to address with stride segments defined from HS→HS.

In order to provide this kind of input data, we need to detect HS and toe-off (TO) events in the stride segmentation provided by the dataset and adjust the stride borders accordingly. Detection of HS and TO events within the sensor data is done according to Rampp *et al.* [7]. An exemplary input signal for one stride defined from HS→HS is shown in Fig. 4.

Furthermore, we can directly compute three expert features from the list given in Table I based on the HS and TO events in the datastream. Stride time is defined as the HS→HS distance and given a TO event, we can subdivide each stride into its two phases and compute stance and swing time (see Fig. 3(a) or Rampp *et al.* [7]). This leaves a set of five output parameters to be estimated by the deep CNNs.

## C. Network Architectures

The network architectures used here are based on three elementary building blocks: convolutional, max-pooling, and densely connected layers.

A convolutional connection between layers is defined by a set of $N_i$ kernels $\boldsymbol{\psi}_1 \dots \boldsymbol{\psi}_{N_i}$ of length $L_i$ and biases $b_1 \dots b_{N_i}$. The index $i$ thereby represents a label for the layer at hand. Given a multichanneled input vector $\boldsymbol{x}_j$ with $j = 1 \dots N_{i-1}$, the activation/output of the convolutional connection is computed as

$$\boldsymbol{a}_k = \text{ReLU} \left( \sum_j \boldsymbol{\psi}_{k,j} * \boldsymbol{x}_j + b_k \right) \quad (1)$$

with $k = 1 \dots N_i$. We used a rectifying linear unit (ReLU) as the activation function for this type of connection as this nonlinearity is said to model biological neurons better compared with tangent or sigmoid functions [25].

Convolutional layers are often followed by max-pooling layers to increase robustness of the extracted features [26]. This type of connection downsamples the feature maps obtained by the convolutional connection by taking the maximum in temporal windows of length $r$. The downsampling factor is thus $1/r$.

The third type of connection used here is the densely connected layer. This type of connection is defined by a set of weight vectors $\boldsymbol{W}_1 \dots \boldsymbol{W}_{N_i}$ and biases $b_1 \dots b_{N_i}$. Given a single-channel input vector $\boldsymbol{x}$, the activation of the densely connected layer is computed by matrix multiplication as

$$a_j = \text{ReLU} \left( \sum_i W_{i,j}\, x_i + b_j \right) \quad (2)$$

with $j = 1 \dots N_i$. In case the output of the previous layer is a multichanneled vector, the single-channel input vector $\boldsymbol{x}$ is constructed by concatenation of the individual channels, i.e. flattening. Again, we use a ReLU for activation.

Finally, a readout layer compresses the last dense layer to the number of output variables for the task at hand. The readout layer is identical to a densely connected layer with the identity instead of the ReLU as an activation function. The number of output variables is then encoded into the number of weight vectors for this layer.

Based on these elementary blocks, two models are built:
- Model A: Estimating the complete set of output variables with a combined model (see Fig. 5, left).
- Model B: Estimating each output variable individually with an ensemble of networks (see Fig. 5, right).

Consequently, the individual network architectures in model B can be less complex compared with model A in order to achieve comparable model complexities.

Regarding the application of data to knowledge translation in the context of mobile gait analysis, we decide for a network architecture built from three convolutional layers with max-pooling followed by three densely connected layers and a readout layer for model A. In the convolutional layers, we train $N_1 = 32$, $N_2 = 64$ and $N_3 = 128$ kernels of size 30, 15 and 7 samples respectively, as well as the corresponding number of bias terms. Max-pooling is done in nonoverlapping, temporal windows of size $r = 2$ samples. Given the sampling frequency, the kernel size corresponds to approximately 0.29 s on all three layers. The three densely connected layers are trained with $N_4 = 2048$, $N_5 = 1024$ and $N_6 = 512$ weight vectors and bias terms respectively. The readout layer has $N_{\text{output}} = 5$ nodes for model A.

For model B, the individual network architectures are built from two convolutional layers with $N_1 = 16$ and $N_2 = 32$ filters of size 30 and 15 samples respectively and one densely connected layer with $N_3 = 1024$ nodes. The max-pooling layers are identical to model A with a downsampling factor of $\frac{1}{2}$. However, the readout layer has only $N_{\text{output}} = 1$ node as each individual architecture is responsible for one of the output parameters. Fig. 5 gives an overview on the network architectures used in both models.

The theoretical motivation for this choice is to address the most crucial question in network design of global versus individual modeling with two representative cases. In model A, we only distinguish between different kinds of output parameters at the last level of the network. The features extracted in this architecture therefore have to be general enough to capture information about all of the output parameters. In model B, however, each output parameter has its own feature extraction path that can be optimized to the parameter at hand.

## D. Training

Training of neural networks is viewed as an optimization problem regarding an error function (implicitly) depending on the network parameters. This error defines a discrepancy measure between predicted output and a ground truth reference on the training dataset or subsets thereof. Weights and biases on all layers are then changed using backpropagation and with the aim to minimize the error. In practice, however, only random subsets of the training dataset (mini-batches), are shown to the optimizer in one iteration of the training loop to speed up the learning phase (stochastic learning) [27].

Because of the different numeric ranges and physical units in the output parameters (see Table I), the network is trained to estimate normalized and dimensionless output variables $\hat{y}_i$. Therefore, each reference $y_{i,\text{ref}}$ is scaled to the range $[0, 1]$ by using the minimum/maximum value attained on the entire training set $\mathcal{S}_{\text{train}}$:

$$\hat{y}_{i,\text{ref}} = \frac{y_{i,\text{ref}} - \min_{\mathcal{S}_{\text{train}}} y_{i,\text{ref}}}{\max_{\mathcal{S}_{\text{train}}} y_{i,\text{ref}} - \min_{\mathcal{S}_{\text{train}}} y_{i,\text{ref}}}. \quad (3)$$
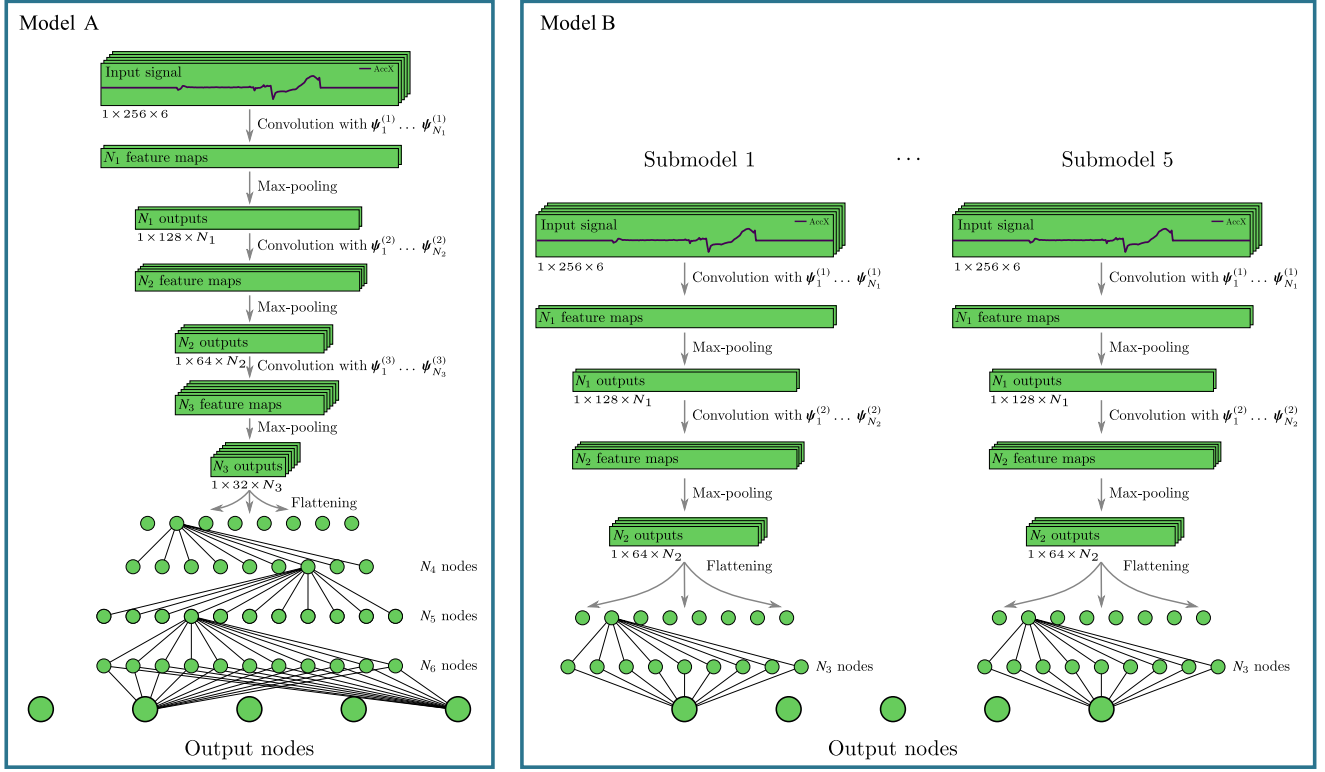
Fig. 5. The two network architectures for data to knowledge translation in the context of mobile gait analysis: Model A consists of three convolutional layers with max-pooling followed by three densely connected layers. Model B, however, spawns smaller networks consisting of two convolutional layers with max-pooling and one densely connected layer for each of the output variables. Additionally, dimensionalities of important layers are indicated.

Predictions on the test set are later rescaled to their physical dimensions using the scaling parameters from the training set.

Given predictions $\hat{y}_i$ on a mini-batch of size $N_{\text{batch}}$ for each output variable $i = 1 \ldots N_{\text{output}}$, we define the error as the sum of the individual root-mean-square errors on the mini-batch:

$$E = \sum_i \text{rmsq} \left( \hat{y}_i - \hat{y}_{i,\text{ref}} \right). \tag{4}$$

For optimization, we use Adam [28], a state-of-the-art optimization method for stochastic learning. On benchmark datasets, it shows faster convergence than other stochastic optimization routines and we use default settings of $\alpha = 1\mathrm{e}^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1\mathrm{e}^{-8}$ (for details see [28]). All weights are initialized by sampling a truncated normal distribution with standard deviation 0.01 and biases are initially set to 0.01. We train for a fixed number of 4000 iterations with a mini-batch size of $N_{\text{batch}} = 100$ strides.

To prevent overfitting, dropout is used on the densely connected layers. This technique effectively samples a large number of thinned architectures on the hidden layers by randomly dropping nodes during training. With this, overfitting could be significantly reduced in many use cases and was superior to weight-regularization methods [29]. We use fixed dropout probabilities of $p^{(4)} = 0.75$, $p^{(5)} = 0.5$ and $p^{(6)} = 0.0$ for model A. For the individual architectures in model B, a dropout probability of $p^{(3)} = 0.5$ is used. Every connection thus has a 50%

chance of being inactive. During testing, however, the full architectures are used and no connections are dropped.

The networks are implemented and trained using Google's TensorFlow library [30].

### E. Evaluation Scheme

Evaluation of the two modeling approaches is based on a 10-fold cross-validation scheme. The stride-specific sensor data from 99 patients on the dataset are sorted into training and test partitions depending on the patient identifier to ensure distinct splits of the dataset. For each of the two models, we iterate over the complete dataset in this fashion and estimate the output variables on the test set in each fold. The estimates from individual folds are then pooled to arrive at average statistics for each output variable and model. As an evaluation statistic, we use average accuracy $\pm$ precision which correspond to the mean and standard deviation of the signed error distribution. The two models are compared based on this statistic and a Levene test of equal variances between the respective distributions to check whether precisions differ significantly. Because the error distributions for the parameter heel contact time are slightly non-Gaussian (checked by visual inspection of $q$–$q$ plots), a Levene test is preferred over e.g. a Bartlett test that is less robust against non-normality.

In order to assess the learning speed and performance of models A and B, we compute the training error for each of
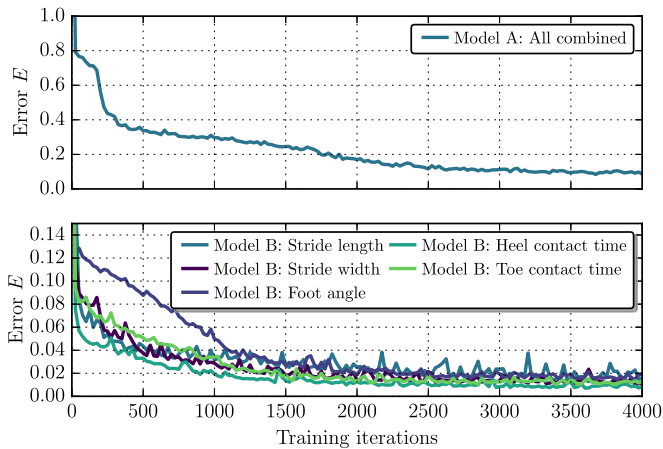
Fig. 6. Error $E$ evaluated on the entire training set as a function of the training iterations for model A, all submodels that constitute model B and an exemplary 90/10% split of the dataset.

the models over the training iterations for an exemplary and patient-wise 90/10% train/test split of the dataset.

## III. RESULTS

### A. Training

Fig. 6 shows the error evaluated on the entire training set over the iterations for an exemplary 90/10% train/test split of the dataset. The error is evaluated for model A and for each of the submodels that constitute model B. In all cases, the fixed number of 4000 iterations is sufficient to reach a stable regime of the error on the entire training dataset and hence we stop the training. Furthermore, the adaptation of the two models to the training data is comparable w.r.t. the selected error function as $E_{1...5} \approx 0.02$ in model B corresponds to a total/summed error of $E \approx 0.1$ in model A (see Fig. 6).

### B. Stride Parameter Estimation on Unseen Data

Table II lists average accuracy and precision on the unseen test data achieved by the two models w.r.t. the pooled estimates from each cross-validation fold. The ensemble approach B that spawns one CNN for each output variable reaches significantly better precision regarding stride length and width, while the corresponding mean accuracies also exceed those achieved by model A. On the remaining three parameters foot angle, heel and toe contact time, both models perform similarly. Therefore, we consider the ensemble approach B to be superior in the context of mobile gait analysis.

Detailed results for all output parameters on the dataset and the superior model B can be found in Fig. 7 and Table III. Fig. 7 includes Bland–Altman plots for each of the output variables as estimated by model B as well as the achieved mean accuracy and precision. Table III lists the error statistics for each output parameter as well as state-of-the-art results.

## IV. DISCUSSION

We present a method for data to knowledge translation in the context of sensor-based gait analysis to extract a total of eight spatio-temporal stride characteristics. Within the proposed framework, we compare two different approaches in the case of vectorial knowledge: A) A combined modelling approach estimating the complete set of output parameters and B) An ensemble approach where individual, less complex models are spawned for each output parameter. The resulting model complexities are thereby designed to be of similar magnitude. With the superior ensemble approach B, the spatial parameters stride length, stride width and foot angle are estimated with mean accuracy and precision of $-0.15 \pm 6.09$ cm, $-0.09 \pm 4.22$ cm and $0.13 \pm 3.78°$ respectively. The temporal stride characteristics stride, swing and stance time are predicted with average precision of $\pm 0.07$, $\pm 0.05$ and $\pm 0.07$ s respectively. Additionally, the heel and toe contact times are determined up to $\pm 0.07$ and $\pm 0.12$ s respectively. Thereby, we provide technical validation on a clinically relevant dataset containing 1185 individual strides from 99 geriatric patients.

The estimation of stride length is outperforming the double-integration result by Rampp *et al.* [7] by 2.3 cm (27%, statistically significant) in precision. Compared with the deep learning approach presented by Hannink *et al.* [13] that estimates this parameter up to $-0.27 \pm 5.43$ cm based on HS→HS strides, our results are slightly worse. This might be due to reduced network complexity in the model presented here (Hannink *et al.* [13] use twice as much kernels and biases on each of the two convolutional layers) or the fact that the underlying datasets are not completely identical (101 vs. 99 patients here due to technical reasons). Regarding results on other datasets, Ferrari *et al.* [8] report a measurement error of $-0.16 \pm 7.02$ cm for the parameter stride length based on double integration and a dataset of 1314 strides captured from 10 elderly PD patients. Trojaniello *et al.* even reach an average accuracy and precision of $0.1 \pm 1.9$ cm and thereby almost resolve their reference precision of $\pm 1.3$ cm [9]. However, this result is evaluated only on a small dataset of 532 strides from 10 elderly PD patients and points out the limits of this comparison: The results achieved have to be seen as a function of the variability across subjects captured in the evaluation dataset. A different evaluation dataset does not necessarily ensure a fair comparison of methods and it shows the need for a unified evaluation of stride length estimation methods presented here and in the literature ([7]–[9], [13]) on the same, large cohort study as, e.g., the publicly available dataset described by Rampp *et al.* [7].

Regarding stride width, there is little related work. Horak and Mancini [31] even consider it "difficult to obtain with body-worn sensors". Nevertheless, Rebula *et al.* [12] report an intraclass correlation coefficient (ICC) of 0.88 between a motion-capture reference and a sensor-based estimation of stride width. Our result corresponds to an ICC of 0.95 and is thus outperforming state-of-the-art double-integration methods w.r.t. stride width.

TABLE II
COMPARISON OF MODEL A AND B REGARDING AVERAGE ACCURACY AND PRECISION REACHED ON UNSEEN TEST DATA

|  | Stride length | Stride width | Foot angle | Heel contact time | Toe contact time |
|---|---|---|---|---|---|
| Model A | $-0.34 \pm 8.10$ cm | $0.41 \pm 7.79$ cm | $-0.05 \pm 3.59°$ | $-0.00 \pm 0.08$ s | $-0.00 \pm 0.12$ s |
| Model B | $\mathbf{-0.15 \pm 6.09}$ **cm** | $\mathbf{-0.09 \pm 4.22}$ **cm** | $0.13 \pm 3.78°$ | $0.00 \pm 0.07$ s | $0.00 \pm 0.12$ s |
| Levene test | sign. | sign. | n.s. | n.s. | n.s. |

To compare precisions, a Levene test was performed on the respective error distributions at the 0.01 significance level.

Mariani *et al.* [6] determine the foot or turning angle up to $0.12 \pm 3.59°$ on data from 10 elderly subjects and a study protocol that included a $180°$ turn as opposed to our straight walking data. In this respect, our results are comparable.

For the parameters stride, stance, and swing time, our results are identical to Rampp *et al.* [7] due to identical methods. Estimation of heel and toe contact times has not been reported in the literature to the best of the authors' knowledge. Sabatini [32] proposes the detection of heel-off and toe-on/foot-flat events by thresholding the angular velocity in the sagittal plane. However, this approach is rather heuristic and the precision regarding these events was not evaluated. Thus, our result constitutes the state-of-the-art regarding sensor-based estimation of heel and toe contact times.

Based on these contact times and HS/TO events that are detectable with state-of-the-art methods, this work enables the detection of heel-off and toe-on events (see Fig. 3) that do not manifest that clearly in the sensor signals. Thereby, each gait cycle can be further subdivided into loading response, mid-stance phase, terminal stance and preswing as defined by Perry *et al.* [33] and their dependence on disease state or speed could be evaluated in future work. The latter would extend the work by Hebenstreit *et al.* [34] that was based on motion capture data to a mobile setting.

The entire processing pipeline presented in this work is based on stride segments defined from HS to HS and is therefore independent of the zero-velocity assumption. In everyday clinical practice and the presence of impaired gait, this assumption is easily violated and limits applicability of state-of-the-art double-integration approaches [13]. However, there are no theoretical considerations that prohibit the application of the proposed method to populations experiencing severe gait disturbances as in the case of spasticity. Thereby, the proposed system is a suitable substitute for the assumption-governed double-integration techniques and could enable mobile gait analysis in these clinically critical cases.

The main limitation of the proposed method is that the resulting data to knowledge translation is as good as the knowledge base. This is because the knowledge base is used to learn the nonlinear relationship between the input sensor data and the output parameters and this mapping implicitly depends on the samples collected in the knowledge base. We thus strongly stress the importance of sharing benchmark datasets within the community and create larger, community-maintained knowledge bases.

The implementation of the framework is generalizable and flexible. Information from other wearable sensing modalities (e.g. barometric pressure) can be introduced by adding addi-
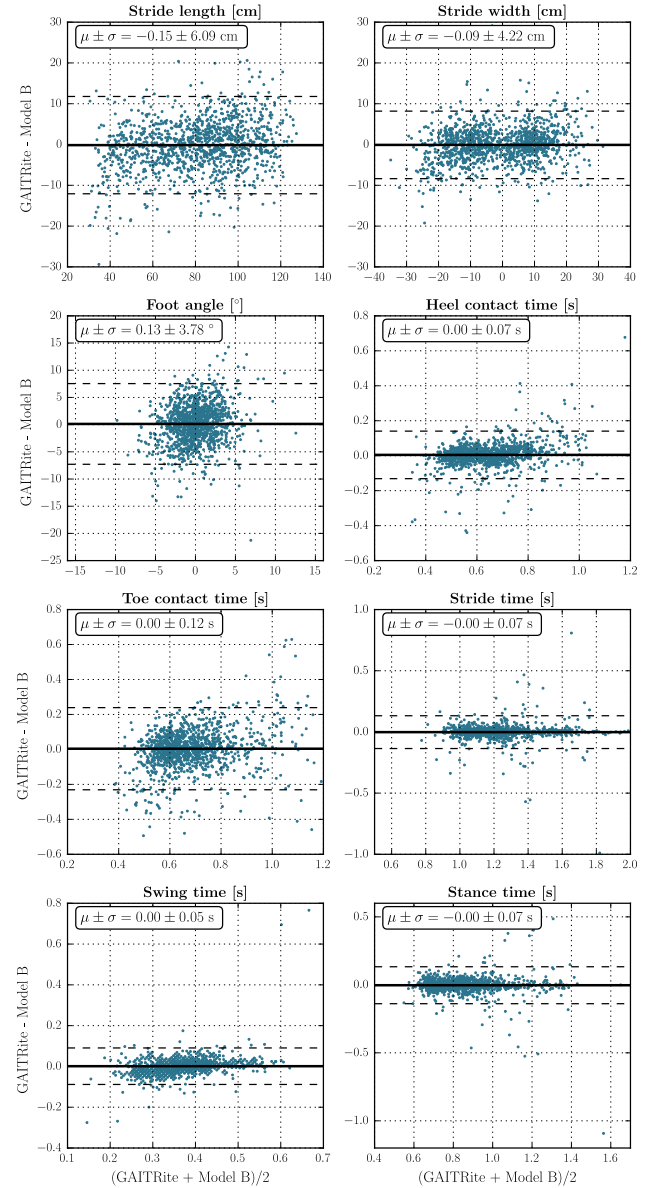


Fig. 7. Bland–Altman plots for each of the output variables estimated by model B on unseen test data. Additionally, the mean accuracy (solid line) and $\pm 1.96\sigma$ bounds (dashed lines) are shown.

tional channels to the input signal. Application to other data to knowledge translation problems in this field can be done by exchanging the knowledge base. TensorFlow [30] generically supports model quantization and lower level arithmetics that are needed for inference with deep CNNs on mobile devices. Although this is not a necessity in mobile gait analysis, where

TABLE III
AVERAGE ESTIMATES AS ACHIEVED BY MODEL B ON UNSEEN TEST DATA AND THE REFERENCE SYSTEM GAITRITE. THE ERROR OF MEASUREMENT IS
REPORTED AS MEAN ACCURACY AND PRECISION. ADDITIONALLY, RESULTS FROM STATE-OF-THE-ART DOUBLE-INTEGRATION APPROACHES ARE LISTED.

| Output Parameter | Unit | Average estimates | | Mean acc. $\pm$ prec. | |
|---|---|---|---|---|---|
| | | Model B | GAITRite | Model B | State-of-the-art |
| Stride length | cm | $80.78 \pm 21.82$ | $80.63 \pm 23.23$ | $-0.15 \pm 6.09$ | $-0.26 \pm 8.37$[7] |
| | | | | | $-0.16 \pm 7.02$[8], * |
| | | | | | $0.10 \pm 1.90$[9], * |
| Stride width | cm | $-1.34 \pm 12.49$ | $-1.44 \pm 13.29$ | $-0.09 \pm 4.22$ | —(see text) |
| Foot angle | ° | $-0.06 \pm 2.90$ | $0.07 \pm 3.49$ | $0.13 \pm 3.78$ | $0.12 \pm 3.59$[6], * |
| Heel contact time | s | $0.64 \pm 0.12$ | $0.64 \pm 0.14$ | $0.00 \pm 0.07$ | – |
| Toe contact time | s | $0.68 \pm 0.14$ | $0.69 \pm 0.17$ | $0.00 \pm 0.12$ | – |
| Stride time | s | $1.23 \pm 0.19$ | $1.23 \pm 0.19$ | $-0.00 \pm 0.07$ | $0.00 \pm 0.07$[7] |
| Swing time | s | $0.37 \pm 0.07$ | $0.37 \pm 0.08$ | $0.00 \pm 0.05$ | $-0.00 \pm 0.05$[7] |
| Stance time | s | $0.86 \pm 0.16$ | $0.85 \pm 0.16$ | $-0.00 \pm 0.07$ | $0.00 \pm 0.07$[7] |

Model B: $n = 99$ geriatric patients and 1185 individual strides; *different evaluation dataset.

the emphasis lies on the mobility of the sensing technology, it might be needed in future work.

Future work includes the application of the proposed framework to other data to knowledge translation problems and thereby the establishment of a generally applicable system. In this respect, especially the end of training needs to be addressed in a data-adaptive manner. As this work did not include a rigorous exploration of the parameter space (number and dimensionality of kernels, etc.), this part is left for future work. In the context of mobile gait analysis, individualization or domain adaptations aspects as well as sequential modeling approaches that account for across-stride context will be investigated. Additionally, the aforementioned benchmark evaluation of several biomechanical parameter estimation methods for a fair comparison of methods will be covered in future work.

## REFERENCES

[1] T. Ellis, J. T. Cavanaugh, G. M. Earhart, M. P. Ford, K. B. Foreman, and L. E. Dibble, "Which measures of physical function and motor impairment best predict quality of life in parkinson's disease?" *Parkinsonism and Related Disorders*, vol. 17, no. 9, pp. 693–697, 2011.

[2] U. Givon, G. Zeilig, and A. Achiron, "Gait analysis in multiple sclerosis: Characterization of temporal-spatial parameters using GAITRite functional ambulation system," *Gait & Posture*, vol. 29, no. 1, pp. 138–142, 2009.

[3] K. E. Webster, J. E. Wittwer, and J. A. Feller, "Validity of the GAITRite walkway system for the measurement of averaged and individual step parameters of gait," *Gait & Posture*, vol. 22, no. 4, pp. 317–321, 2005.

[4] R. W. Kressig *et al.*, "Temporal and spatial features of gait in older adults transitioning to frailty," *Gait & Posture*, vol. 20, no. 1, pp. 30–35, 2004.

[5] J. Klucken *et al.*, "Unbiased and mobile gait analysis detects motor impairment in parkinson's disease," *PLoS ONE*, vol. 8, no. 2, 2013, Art. no. e56956.

[6] B. Mariani, M. C. Jiménez, F. J. G. Vingerhoets, and K. Aminian, "On-shoe wearable sensors for gait and turning assessment of patients with parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 155–158, Jan. 2013.

[7] A. Rampp, J. Barth, S. Schülein, K.-G. Gaßmann, J. Klucken, and B. M. Eskofier, "Inertial sensor based stride parameter calculation from gait sequences in geriatric patients," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1089–1097, Apr. 2014.

[8] A. Ferrari, P. Ginis, M. Hardegger, F. Casamassima, L. Rocchi, and L. Chiari, "A mobile kalman-filter based solution for the real-time estimation of spatio-temporal gait parameters," *IEEE Trans. Neural Syst. Rehabil. Eng.*, no. 99, pp. 764–773, Jul. 2015.

[9] D. Trojaniello *et al.*, "Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: Application to elderly, hemiparetic, parkinsonian and choreic gait," *J. Neuroeng. Rehabil.*, vol. 11, no. 152, 2014, Art. no. 152.

[10] K. Aminian, B. Najafi, C. Büla, P. F. Leyvraz, and P. Robert, "Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes," *J. Biomech.*, vol. 35, no. 5, pp. 689–699, 2002.

[11] A. Salarian, P. R. Burkhard, B. M. Jolles, and K. Aminian, "A novel approach to reducing number of sensing units for wearable gait analysis systems," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 72–77, Jan. 2013.

[12] J. R. Rebula, L. V. Ojeda, P. G. Adamczyk, and A. D. Kuo, "Measurement of foot placement and its variability with inertial sensors," *Gait & Posture*, vol. 38, no. 4, pp. 974–980, 2013.

[13] J. Hannink *et al.*, "Stride length estimation with deep learning." 2016. [Online]. Available: https://arxiv.org/abs/1609.03321

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition." 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[15] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3156–3164.

[16] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1701–1708.

[17] N. Y. Hammerla, J. M. Fisher, P. Andras, L. Rochester, R. Walker, and T. Ploetz, "PD disease state assessment in naturalistic environments using deep learning," in *Proc. Conf. Innovative Appl. Artif. Intell.*, pp. 3207-3215, 2015.

[18] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, convolutional, and recurrent models for human activity recognition using wearables." 2016. [Online]. Available: http://arxiv.org/abs/1604.08880

[19] F. J. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, 2016, Art. no. 115.

[20] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, "Time series classification using multi-channels deep convolutional neural networks," in *Web-Age Information Management* (Lecture Notes in Computer Science series), vol. 8485. Berlin, Germany: Springer-Verlag2014, pp. 298–310.

[21] N. D. Lane and P. Georgiev, "Can deep learning revolutionize mobile sensing?" in *Proc. Int. Workshop Mobile Comput. Syst. Appl., HotMobile'15*, 2015, pp. 117–122.

[22] A. Burns *et al.*, "SHIMMER—A wireless sensor platform for noninvasive biomedical research," *Sensors*, vol. 10, no. 9, pp. 1527–1534, 2010.

[23] J. C. Menant, J. R. Steele, H. B. Menz, B. J. Munro, and S. R. Lord, "Effects of walking surfaces and footwear on temporo-spatial gait parameters in young and older people," *Gait & Posture*, vol. 29, no. 3, pp. 392–397, 2009.

[24] B. Salzman, "Gait and balance disorders in older adults," *Amer. Family Physician*, vol. 82, no. 1, pp. 61–68, 2010.
[25] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, vol. 15, pp. 315–323.
[26] Y. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. Int. Conf. Mach. Learni.*, 2010, pp. 111–118.
[27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org
[28] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–13.
[29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
[30] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems." 2015. Online. Available: https://tensorflow.org
[31] F. B. Horak and M. Mancini, "Objective biomarkers of balance and gait for parkinson's disease using body-worn sensors," *Movement Disorders*, vol. 28, no. 11, pp. 1544–1551, 2013.
[32] A. M. Sabatini, "Quaternion-based strap-down integration method for applications of inertial sensing to gait analysis," *Med. Biol. Eng. Comput.*, vol. 43, no. 1, pp. 94–101, 2005.
[33] J. Perry, J. M. Burnfield, and L. M. Cabico, *Gait Analysis: Normal and Pathological Function*, vol. 12. Thorofare, NJ, USA: Slack, 1992.
[34] F. Hebenstreit, A. Leibold, S. Krinner, G. Welsch, M. Lochmann, B. M. Eskofier, "Effect of walking speed on gait sub phase durations," *Human Movement Sci.*, vol. 43, pp. 118–124, 2015.

**Karl-Günter Gaßmann** received the degree and M.D. from the School of Medicine, Ruperto-Carola-University, Heidelberg, Germany, in 1986. He currently holds a professorship at the Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany, and is the Director of the Geriatrics Center Erlangen, Waldkrankenhaus St. Marien, Erlangen, Germany. His research interests include fall prevention, gait analysis, and frailty assessment in older age.

**Jochen Klucken** received the Graduate degree in medicine from the University of Regensburg, Regensburg, Germany, in 1996. He received the Habilitation degree in neuroscience in 2009, working on cell biology aspects of neurodegeneration in Parkinson's disease.

He is the Vice Head of the Department of Molecular Neurology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany. He is cosupervising the Movement Disorder Unit of the hospital and performing several translational research studies in wearable sensor technologies for patients with movement disorders. His main focus is Parkinson's disease and he is also studying the underlying cell biology aspects of Parkinson's disease, as well as novel diagnostic devices serving as objective biomarkers for the disease. He worked for four years on his experimental medical thesis in the field of atherosclerosis in the Department of Clinical Chemistry, University of Regensburg.

In 2004, he also started the translational research projects applying sensor-based motion detection in movement disorders. From 2008 he was a Principle Investigator in the Department of Molecular Neurology (Head Prof. Dr. Juergen Winkler) focusing on cell biology aspects of neurodegeneration and mobile healthcare technologies for integrated care concepts in Parkinson's disease.

**Julius Hannink** received the B.Sc. and M.Sc. degrees in physics in 2011 and 2014 from the University of Göttingen, Göttingen, Germany. He is currently working towards the Ph.D. degree in the Digital Sports Group, Pattern Recognition Laboratory, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany.

His work focuses on individualized, sensor-based therapy monitoring and mobility impairment analysis in neurodegenerative diseases. His research interests include wearable computing, biomedical signal analysis and machine learning for healthcare applications.

**Thomas Kautz** received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from the Ilmenau University of Technology, Ilmenau, Germany, in 2012 and 2013, respectively. He is currently pursuing a Ph.D. degree at the Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany.

His research interests include pattern recognition, probabilistic filtering and biomedical signal processing.

**Bjoern M. Eskofier** (SM'16) received the Graduate degree in electrical engineering from the Friedrich-Alexander University (FAU) Erlangen-Nürnberg, Erlangen, Germany, in 2006, and the Ph.D. degree in biomechanics from the University of Calgary, Calgary, Canada, in 2010.

Since February 2011, he has been the Assistant Professor of computer science in sports (endowed professorship of the Adidas AG) and the Head of the Digital Sports Group, Pattern Recognition Laboratory, FAU. He has authored or coauthored more than 100 peer-reviewed articles and submitted 5 patent applications. His current research interests include machine learning and signal analysis for wearable computing systems in health care and sports.

**Cristian F. Pasluosta** received the degree in electrical engineering from the National University of Cordoba, Cordoba, Argentina, in 2005, the M.S. degree in engineering, and the Ph.D. degree in biomedical engineering from Louisiana Tech University, Ruston, LA, USA, both in 2010. From 2011 to 2014, he was a Postdoctoral Research Fellow at the Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA. Since 2014, he is a Postdoctoral Researcher in the Digital Sports Group, Pattern Recognition Laboratory, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany. His research interests include applying machine learning techniques for biomedical applications, wearable technologies, motion analysis, mechatronic design and control systems.