A Comparative Study of End-To-End Discriminative Deep Learning Models for Knee Joint Kinematic Time Series Classification

M. Abid[1,2], A. Mitiche[3], Y. Ouakrim[1,2], P. A. Vendittoli[4], A. Fuentes[5], N. Hagemeister[2] and N. Mezghani[1,2]

1. Centre de Recherche LICEF, TELUQ university, Montreal, Quebec, Canada
2. Laboratoire LIO, Centre de Recherche du CHUM, Montreal, Quebec, Canada
3. INRS, Centre Énergie, Matériaux et Télécommunications, Montreal, Canada
4. Centre de recherche Hôpital Maisonneuve-Rosemont, Montreal, Quebec, Canada
5. EMOVI, Laval, Quebec, Canada
{mariem.abid, nicola.hagemeister}@etsmtl.ca, mitiche@emt.inrs.ca,
{oua.youssef, alexandre.fuentes}@gmail.com, pa.vendittoli@me.com, neila.mezghani@teluq.ca

Abstract—One of the main motivations for classifying knee kinematic signals, namely the variation during a locomotion gait cycle of the angles the knee makes with respect to the three-dimensional (3D) planes of flexion/extension, abduction/adduction, and internal/external rotation, is to assist diagnosis of knee pathologies. These signals are informative but high dimensional, and highly variable, which has posed difficulties that have been addressed by machine learning algorithms. The purpose of this study is to investigate classification of knee kinematic signals through the entire gait using deep neural networks. The signals are first pre-processed to identify representative patterns, which are then used for deep learning of discriminative classifiers. This paper describes an efficient means of distinguishing between knee osteoarthrisis patients and asymptomatic participants, and our methods and experiments which validate it.

I. INTRODUCTION

The knee is a complex joint that requires perfectly coupled three-dimensional (3D) motions for proper function. As a result, reliable diagnosis of knee-joint pathologies is a difficult task, requiring in many cases a combination of imaging-based examinations, such as magnetic resonance imaging and computed tomography, and clinical tests. Such methods do not provide direct objective information on the functional aspects of the knee-joint, and are not typically performed during knee movement. For this reason, biomechanical gait analysis has become essential in knee-joint pathology diagnosis: it provides quantitative information about the structure and motion of the knee-joint to complement the common evaluation methods for more accurate diagnosis [1]. 3D knee kinematic signals, measuring knee flexion/extension, abduction/adduction, and internal/external rotation, are now commonly used in gait analysis to assist knee-joint pathology diagnosis. Knee kinematic gait signals can be acquired in a normal clinical setting, using a commercially available treadmill and a simple noninvasive knee attachment system [2]. Recorded as functions of time, they can be viewed as time series. High-dimensionality, the significant within-pathology class variability, and the low between class separation, usually make kinematic signal interpretation quite challenging [3], [4]. Several studies have used knee kinematic signals to distinguish between knee osteoarthritis (OA) and asymptomatic (AS) subjects, and to further classify OA populations according to severity levels. A common approach to knee kinematic signal classification has been to extract and select knee kinematic features, such as flexion angle peak values from the kinematic time series representations, and then apply simple statistical methods such as Student t-test to know whether there were statistically significant differences between normal and pathological groups [5], [6]. Recently, general methods of feature extraction and machine learning have been applied to knee kinematic signals classification. However, general feature extraction faces the high dimensionality and variability that characterize knee kinematic data signals. Recent investigations have obtained good results but have been tested on small datasets, which limits the generality of their conclusions. For instance, 40 knee OA subjects and 40 healthy subjects participated in a study in which 3D knee kinematics data (flexion/extension, abduction/adduction and internal/external rotation) were recorded [7]. The authors could only examine a set of 70 kinematic features to determine the most discriminant features to use. Regression tree representation gave 85% accuracy in discriminating knee OA subjects from healthy subjects. In this study, we use deep neural networks [8] for the automatic classification of asymptomatic and knee OA kinematic data using the entire signal as the initial features on a relatively large dataset compared to previous studies of biomechanical data classification [9]. The motivation for investigating deep neural networks in knee kinematic data classification is their success in classification of time series at large on UCR and/or MTS archive datasets [10], [11], [12], [13] from different domains such as human activity recognition and sleep stage identification. The advantage of deep neural networks is their ability for automatic feature extraction from raw, complex, and high-dimensional data. In this work, pre-processed knee kinematic data signals are the inputs to deep neural networks to learn kinematic features capable of discriminating knee OA patients from asymptomatic participants. Our investigation provides a comparison of the effectiveness of various deep neural networks in knee kinematic time series classification problem. To the best of our knowledge, this is the first study on classifying raw knee kinematic time series using end-to-end discriminative deep learning classifiers via analyzing

the entire kinematic signal.

II. METHODS

A. Ethics statement

Ethical approval for this study was given by the Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM) and the École de Technologie Supérieure (ÉTS), Hôpital Maisonneuve-Rosemont (HMR), local ethics committee. All subjects provided an informed consent before participation.

B. Data collection: Subjects and procedures

We used 3D knee kinematics of 226 subjects collected in different centers. The first group included 81 asymptomatic (AS) subjects (37 males and 44 females). The AS subjects ranged in age from 20 to 58, had a mean age of 32.7 ± 10 years, and a mean body mass index (BMI) 24.5 ± 4.1 kg/m^2 (normal weight). The second group included 145 knee osteoarthritis (OA) patients (59 males and 86 females). The OA subjects ranged in age from 40 to 80, had a mean age of 62.8 ± 10.1 years, and a mean body mass index (BMI) 31.7 ± 7.4 kg/m^2 (obese). 3D knee kinematics are acquired with the $KneeKG^{TM}$ system (Emovi Inc. Canada) during gait on a treadmill (45 s duration). KneeKG is a non-invasive system consisting of a harness that is placed on the participant's knee, an infrared camera, and a computer equipped with the $KneeKG^{TM}$ software [14]. The accuracy [15], reproducibility [16], repeatability [17], and reliability [18] of the system have been studied. In particular, accuracy of the KneeKG system was assessed in studies which evaluated the mean repeatability of measures ranging from $0.4°$ to $o.8°$ for knee angles. Each participant underwent a series of successive gait trials during a given session. In each trial, the motion trajectories in the sagittal, frontal, and transverse planes of the knee reference system are recorded. These data are filtered using a non-parametric time series analysis called Singular Spectrum Analysis (SSA) [19], and transformed into 3D knee-joint angles [17]. A database is created for each participant, containing the 3D knee kinematics in the sagittal, frontal, and transverse planes, i.e, abduction-adduction, flexion-extension, and internal-external rotation, respectively, representing the angle time series, i.e., the time-varying angle values.

C. Data analysis

Knee kinematics classification systems typically have three parts: data-preprocessing, feature extraction and selection, and classification. In this work, pre-processed knee kinematic data curves (entire gait cycle) are fed directly into deep neural networks as inputs.

1) Kinematic data pre-processing: Once data collection is complete, the raw kinematic data are pre-processed in order to find robust representative patterns for each participant, via steps of missing data interpolation, gait events detection, normalization, outlier detection, cycles' selection, and averaging. The data curve is interpolated by cubic spline interpolation to fill gaps that may be present between data point measurements. The curves of each

participant are then divided into distinct gait cycles. Our approach centers on the location of local maxima values in the sagittal plane signal, since the data from the sagittal plane are more reproducible than those from the other two planes [20]. Heel strike (HS) points are the first local minima after the local maxima and are specified as the start of each gait cycle. This is followed by normalization to 100% of the gait cycle, giving 100 measurement points for each participant in each plane. Thus, for each plane, the superposed normalized cycles constitute the observations to describe the representative patterns characterizing the given participant. These observations correspond to a family of curves due to within-subject variability from stride-to-stride, and are possibly affected by outliers. Representative gait patterns of a given subject are then determined by within-subject evaluation for outliers removal and reliable curves selection. The within-subject variability is estimated using Boxplot, which allows the efficient elimination of outlier curves, which occur outside upper and lower limits. This is followed by cross-validation, which consists of keeping only about 15 of the most repeatable curves, i.e., those having the lowest Root Mean Square Error (RMSE). Finally, the gait of each subject is characterized by a single curve, which is the mean of these [21]. Figure 1 shows the time-series kinematic signals in the sagittal plane, frontal plane, and transverse plane for the whole dataset.

2) Deep neural networks for knee kinematic time series classification task: The purpose is to classify knee kinematic time series using deep neural networks. We tested various end-to-end discriminative deep learning models designed specifically for time series classification, in order to find which model works best for knee kinematic time series classification. In contrast to feature engineering, end-to-end discriminative deep learning directly learns the mapping between the raw input of a time series and outputs a class probability distribution. This is important in our study because it avoids the bias due to handcrafted features. A deep neural network has an input layer, an output layer, and more than two hidden layers. In the present case, deep neural networks are applied for each plane separately, namely to the flexion/extension angle, with respect to the sagittal plane, the abduction/adduction angle, with respect to the frontal plane, and the internal/external angle, with respect to the transverse plane, to determine the contribution of each plane in discriminating patients with OA and AS participants. For each plane, the input layer inputs a labeled knee kinematic time series dataset $D = [(Y_1, X_1), (Y_2, X_2), ..., (Y_N, X_N)]^T$, N denotes the dataset size, X_i is a univariate time series and Y_i its corresponding two-class label (AS and OA). Target values are 0 for the AS class and 1 for the OA class. $X_i = [x_{i1}, x_{i2}, ..., x_{iT}]$, $T = 100$ denotes the time series length (corresponding to the gait cycle percentage). The elements x_{ij} of the time series are real numbers corresponding to the knee kinematic angles. The task of kinematic time series classification consists therefore of training a classifier on a dataset D in order to map the input set X_i into its class label Y_i.
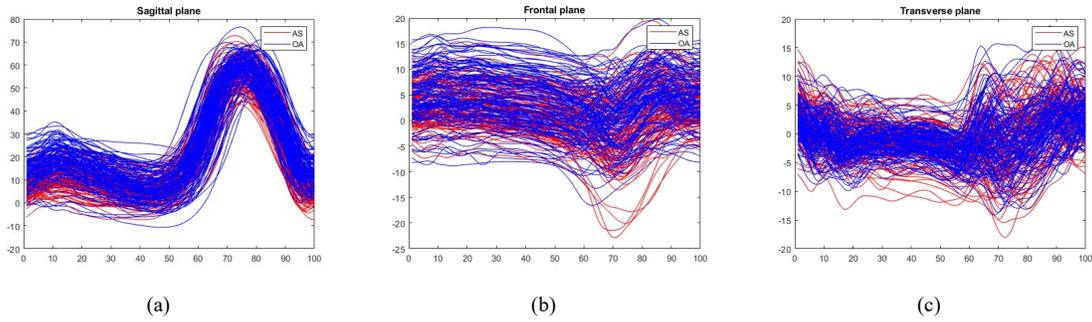
Figure 1. Plots of knee kinematic time series for all subjects in all three planes.

Hidden layers of a deep network are designed to learn hierarchical feature representations of the data. During training, a set of hyper-parameters is optimized, and the weights are initialized randomly [22]. Using gradient descent [22], [23], the weights are updated using the back propagation algorithm, in a way that minimizes the cost function on the training set. The choice of the model, the architecture, and the cost function are crucial for obtaining a network that generalizes well, and are generally problem and data dependent. We trained seven deep learning models, which have convolutional neural network (CNN)-based architecture [24], and designed specifically for time series classification. The different hyper-parameters of CNN are the optimization algorithm (momentum), the number of epochs, the number of layers, the number of filters, the filter size, the activation function, the cost function, the batch size, and the weight initialization. In the remainder of this paper, a convolutional block is denoted $Block_k$ with the number of filters c_k, and the filter size z_k. Here following are the deep learning models that we investigated for their ability to discriminate between knee kinematic signals of patients with OA and AS participants.

Time convolutional neural network (time-CNN) [25]: Figure 2a summarizes the architecture of the time-CNN model. There are 3 layers in this network including 2 convolution blocks, and 1 fully-connected layer, with a Sigmoid activation function. The convolutional block consists of a convolution layer, a sigmoid layer, and a max pooling layer with pooling size 3. The number of filters c_i={6, 7}, and the filter size z_i={12, 7}.

Multiscale convolutional neural network (MCNN) [26]: Figure 2b summarizes the architecture of the MCNN model. There are 4 layers in this network including 2 convolution blocks, 1 fully-connected layer with 256 neurons and Sigmoid activation function, and a softmax layer. The convolutional block consists of a convolution layer, a sigmoid layer, and a max pooling layer. The search space for the pooling factor is {2, 3, 5}, which denotes the number of outputs of max pooling. The number of filters c_i={256, 256}, and the filter size z_i={{5, 10, 20}, {5, 10, 20}}. The MCNN comprises a multi-scale transformation stage that apply various transformations on the input time series, including identity mapping, down-sampling, and smoothing, each of which is an input to an independent CNN. A deep concatenation technique have been used after the first convolutional block to concatenate all the resulting feature maps at each scale. Window slicing have been used to increasing the size of the training size. The length of slices is set to be 0.9T.

Multi Channel Deep Convolutional Neural Network (MCDCNN) [27]: Figure 2c summarizes the architecture of the MCDCNN model. There are 4 layers in this network including 2 convolution blocks, 1 fully-connected layer with 732 neurons and a ReLU activation function, and finally a softmax layer. The convolutional block consists of a convolution layer, a ReLU layer, and a max pooling layer with pooling size 2. The number of filters c_i={8, 8}, and the filter size z_i={5, 5}. The MCDCNN inputs individual univariate time series (channels), and then concatenates the resulting feature maps of the second convolutional block.

Time Le-Net (t-LeNet) [28]: Figure 2d summarizes the architecture of the t-LeNet model. There are 4 layers in this network including 2 convolution blocks, 1 fully-connected layer with 500 neurons and a ReLU activation function, and finally a softmax layer. The convolutional block consists of a convolution layer, a ReLU layer, and a max pooling layer with pooling size 2 and 4 on the first and second block respectively. The number of filters c_i={5, 20}, and the filter size z_i={5, 5}. Two data augmentation techniques have been proposed namely window slicing and window warping.

Fully Convolutional Neural Network (FCN) [12]: Figure 2e summarizes the architecture of the FCN model. There are 5 layers in this network including 3 convolution blocks, 1 global average pooling layer, and finally a softmax layer. The convolutional block consists of a convolution layer, a batch normalization layer, and a ReLU activation layer. The number of filters c_i={128, 256, 128}, and the filter size z_i={8, 5, 3}.

Encoder [29]: Figure 2f summarizes the architecture of the encoder model. There are 5 layers in this network including 3 convolution blocks, 1 global average polling layer, and finally a softmax layer. The convolutional block consists of a convolution layer, an instance normalization layer, a PReLU activation layer, a max pooling layer with pooling size 2, and a dropout of 0.2. The number of filters c_i={128,

256, 512}, and the filter size z_i={5, 11, 21}.

Residual Network (ResNet) [12]: Figure 2g summarizes the architecture of the ResNet model. There are 11 layers in this network including 9 convolution blocks, 1 global average polling layer, and finally a softmax layer. The convolutional block consists of a convolution layer, a batch normalization layer, and a ReLU activation layer. The number of filters c_i={64, 64, 64, 128, 128, 128, 128, 128, 128}, and the filter size z_i={8, 5, 3, 8, 5, 3, 8, 5, 3}.

All deep learning models used in our study have an output layer with 2 neurons, corresponding to the binary classification in this application.

We should note that all deep learning models were initialized randomly using the Glorot's uniform initialization method [30].

Table I summarizes the optimized hyper-parameters configuration for each model.

III. RESULTS AND DISCUSSION

We trained the deep learning models presented above with 10 different runs each. Following Fawaz *et al.* [13] and for the sake of fair comparison, each run uses the same train/test split of the knee kinamtic time series dataset in the same way as the UCR archive, but with a different random weight initialization, which enables us to take the mean accuracy over the 10 runs. That is, the training and testing set have approximately the same amount of data. Metrics of accuracy, precision and recall were used over the test set for model selection, i.e. the ability of the model to discriminate between AS and OA participants. These metrics are defined in Equation 1-3. In these equations, TP stands for true positives, i.e., the number of OA participants correctly classified as OA participants. TN stands for true negatives, i.e., the number of AS participants correctly classified as AS participants. FP stands for false positives, i.e., the number of AS participants misclassified as OA participants, and FN stands for false negatives, i.e., the number of OA participants misclassified as AS participants.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

In this experiment, the compared deep learning models are available in an open source deep learning framework which is implemented using the open source deep learning library Keras with the Tensorflow back-end. Table II summarizes the accuracy, precision, recall and training time (s) of the compared deep learning models: time-CNN, MCNN, MCDCNN, t-leNet, FCN, encoder, and ResNet, applied for each plane separately, namely the frontal (Abduction/adduction), sagittal (Internal/external rotation), and transverse (flexion/extension) planes.

The results show that ResNet reached competitive accuracy compared to the other deep learning approaches on the knee kinematic time series dataset. This finding concur with the deep learning for time series review where deeper neural networks performed better in related fields [13]. The approaches that involve pre-processing techniques such as data segmentation (MCNN, MCDCNN and t-LeNet) achieved the worst accuracies. In our problem settings, the gait cycle event-based segmentation techniques would be more adequate than the sliding window segmentation technique. The former would split the knee kinematic signal based on the gait cycle events (Heel Strike, toe-off, etc), whereas the latter split the signal into windows of a fixed size. We have noticed that the Adam optimization is also adopted for time series classification as for computer vision. We should note that the ResNet architecture exclude the pooling operation, which means that the length of a time series is kept unchanged throughout the convolution, so as to prevent overfitting. We should note also that the batch size must be more than or equal to one and less than or equal to the number of samples in the training dataset. That is why we replaced the batch size value in models where it is set greater to the training dataset at hand. One of the non-successful strategies we tried the batch mode (where the batch size is equal to the total dataset) and the stochastic mode (where the batch size is equal to one). From a biomechanical perspective, the focused analysis using each plane data separately corroborates that the abduction/adduction patterns are the most discriminative patterns that are able to distinguish OA and AS participants. This result is consistent with studies related to the pathogenesis of knee OA [31], [32], [33], and machine learning-based studies [34].

IV. CONCLUSION AND FUTURE WORK

To the best of our knowledge, the present study is the first to investigate the application of deep learning to differentiate between gait patterns of patients with osteoarthritis and asymptomatic participants using knee kinematic data. The data was collected from different sites to have a larger number of OA patients and AS participants compared to previous studies, which gives better generalization capabilities. Descriptive statistics such as peak angles are commonly extracted from the gait signal. In this study, the entire signal is employed as the initial features. We started from the most successful existing deep learning models applied in various time series domains in order to answer the question of selecting the most appropriate and best-performing model for the knee kinematic time series classification problem to distinguish knee OA and AS participants. Even though we found promising results for knee kinematics time series classification using end-to-end deep learning models, the problem remains challenging. An early task is to fine-tune the ResNet model with much larger datasets, and conduct more extensive experiments on knee kinematic time series. In a future work, we intend to consider a multivariate knee kinematic time series dataset, i.e, each participant in the dataset is represented by a vector of 15 cycles and not their mean. Features could

Table I. hyperparameters'Optimization for knee kinematic time series dataset.

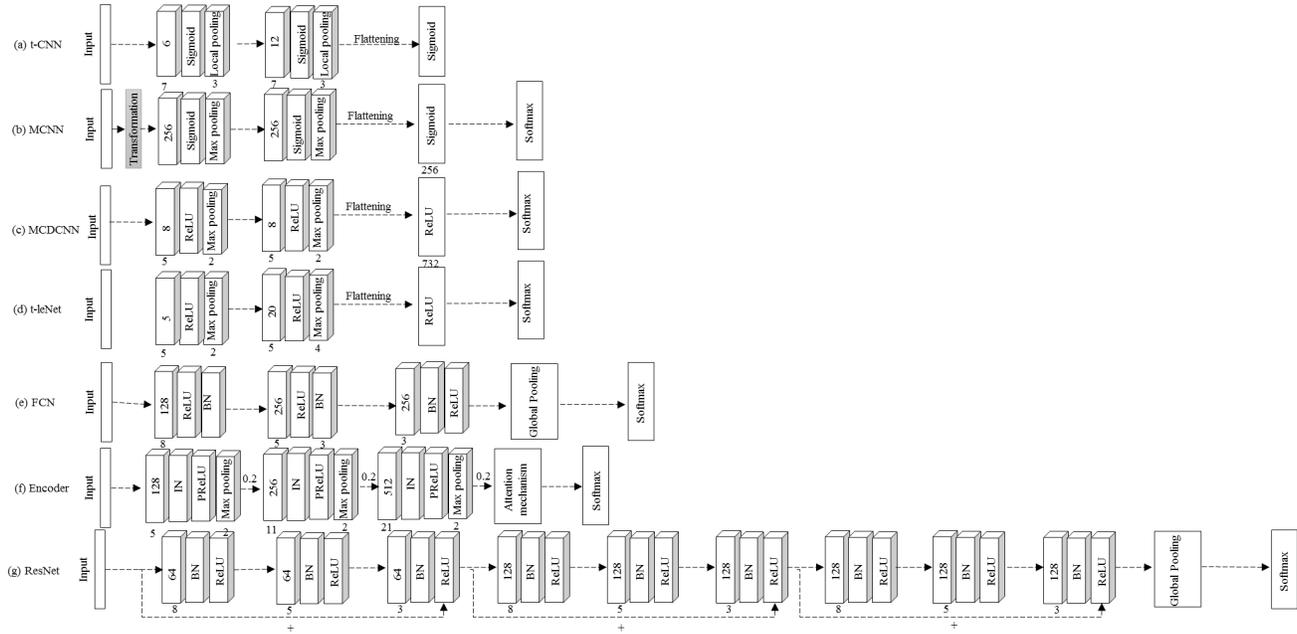| Model	Epochs	Cost function	learning rate	Batch size	Optimizer	Activation function
Time-CNN	2000	MSE	0.001	16	Adam	Sigmoid
MCNN	200	Entropy	0.1	16	Adam	Sigmoid
MCDCNN	120	Entropy	0.01	16	SGD	ReLU
t-leNet	1000	Entropy	0.01	16	Adam	ReLU
FCN	2000	Entropy	0.001	16	Adam	ReLU
Encoder	100	Entropy	0.00001	12	Adam	PReLU
ResNet	1500	Entropy	0.001	16	Adam	ReLU

Figure 2. The architecture of the 7 tested end-to-end deep learning models.

be learned independently on each plane, then the learned features would be concatenated and fed into the classifier. We intend also to compare the ResNet model toward traditional machine learning algorithms applied to a set 70 handcrafted features from the knee kinematic time series [7]. We also intend to understand the learned features by ResNet by applying Class Activation Map (CAM) [35], after improving accuracy to reach 100%, and compare them to the previously cited handcrafted features. This study has some limitations. Indeed, OA patients were older and heavier than those in the AS group which can effect the gait measures and the classification results. However, in our experiments, we used the same data set for all the classifiers. Our aim is to compare several deep learning models for knee joint kinematic time series data. Moreover, to measure the performance of the deep learning models on the test knee kinematic data, we adopted the general evaluation measure: accuracy, precision, and recall (similarly to Fawaz *et al.* [13]). In this study, our primary concern is accuracy. In medical tests, sensitivity and specificity are more readily grasped by clinicians. Therefore, we shall adopt these measures in the future work.

REFERENCES

[1] V. Medved, *Measurement of Human Locomotion*. CRC Press, 2000.
[2] J. D. Guise, N. Mezghani, R. Aissaoui, and N. Hagemeister, "New comprehensive methods for the biomechanical analysis of knee osteoarthritis," in *Understanding Osteoarthritis from Bench to Bedside*, June 2011, pp. 85–102.
[3] N. Mezghani, I. Mechmeche, A. Mitiche, Y. Ouakrim, and J. de Guise, "An analysis of 3d knee kinematic data complexity in knee osteoarthritis and asymptomatic controls," *PLoS ONE*, vol. 13, p. e0202348, 2018.
[4] A. Phinyomark, G. Petri, E. Ibáñez-Marcelo, S. T. Osis, and R. Ferber, "Analysis of big data in gait biomechanics: Current trends and future directions," *Journal of Medical and Biological Engineering*, vol. 28, no. 2, pp. 244–260, 2018.
[5] Y. Nagano, K. Naito, Y. Saho, S. Torii, T. Ogata, K. Nakazawa, M. Akai, and T. Fukubayashi, "Association between in vivo knee kinematics during gait and the severity of knee osteoarthritis," *The Knee*, vol. 19, pp. 628–32, 2012.
[6] D. Bytyqi, B. Shabani, S. Lustig, L. Cheze, N. Karahoda, and G. P. Neyret, "Gait knee kinematic alterations in medial osteoarthritis: three dimensional assessment," *International Orthopaedics*, vol. 8, p. 1191–1198, 2014.
[7] N. Mezghani, Y. Ouakrim, A. Fuentes, A. Mitiche, N. Hagemeister, P. A., and J. de Guise, "Mechanical biomarkers of medial compartment knee osteoarthritis diagnosis and severity grading: Discovery phase," *Journal of biomechanics*, vol. 52, pp. 106–112, 2018.
[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 512, p. 436–444, 2015.

Table II. The results of applying 7 end-to-end deep learning models on each of the plane knee kinematic data separately. Note that the ResNet model is showing the highest accuracy, which is highlighted with bold font.

Model	Precision			Accuracy			Recall			Training time (s)		
	Flex.	Abd.	Rot.	Flex.	Abd.	Rot.	Flex.	Abd.	Rot.	Flex.	Abd.	Rot.
Time-CNN	75.82	77.18	55.66	75.21	77.11	61.53	75.14	77.11	61.55	69	53.05	159.39
MCNN	24.78	25	24.78	49.57	50	4957	50	50	50	509.43	288.33	418.09
MCDCNN	25.21	25	48.64	50.42	50	48.71	50	50	48.90	2.52	1.88	2.22
t-leNet	25.21	25	24.78	50.42	50	49.57	50	50	50	680.72	654.54	704.97
FCN	70.37	75.85	72.88	69.23	72.88	53.84	69.12	72.88	53.76	1253.16	765.03	776.71
Encoder	72.37	71.71	48.71	69.23	66.10	48.62	69.06	66.10	48.62	165.25	169.27	171.54
ResNet	77.31	81.60	62.43	**76.92**	**80.5**	**62.39**	76.87	80.50	62.36	1956	1204.56	1087.103

[9] M. Abid, N. Mezghani, and A. Mitiche, "Knee joint biomechanical gait data classification for knee pathology assessment: A literature review," *Applied Bionics and Biomechanics*, vol. 2019, p. 14, 2019.

[10] Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, "The ucr time series classification archive," July 2015, www.cs.ucr.edu/~eamonn/time_series_data/.

[11] M. G. Baydogan, "Multivariate time series classification datasets," 2015, http://www.mustafabaydogan.com.

[12] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 1578–1585.

[13] I. H. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P. A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, Mar 2019.

[14] S. Lustig, R. A. Magnussen, L. Cheze, and P. Neyret, "The kneekg system: a review of the literature," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 20, no. 4, pp. 633–638, 2012.

[15] M. Sati, J. de Guise, S. Larouche, and G. Drouin, "Improving in vivo knee kinematic measurements: application to prosthetic ligament analysis," *The Knee*, vol. 3, no. 4, pp. 179–190, 1996.

[16] N. Hagemeister, L. Yahia, N. Duval, and J. de Guise, "In vivo reproducibility of a new non-invasive diagnostic tool for three-dimensional knee evaluation," *The Knee*, vol. 6, no. 3, pp. 175–181, 1999.

[17] N. Hagemeister, G. Parent, M. V. de Putte, N. St-Onge, N. Duval, and J. de Guise, "A reproducible method for studying three-dimensional knee kinematics," *Journal of Biomechanics*, vol. 38, no. 9, pp. 1926–1931, 2005.

[18] D. R. Labbe, N. Hagemeister, M. Tremblay, and J. de Guise, "Reliability of a method for analyzing three-dimensional knee kinematics during gait," *Gait & Posture*, vol. 28, no. 1, pp. 170–174, 2008.

[19] R. Aissaoui, S. Husse, H. Mecheri, G. Parent, and J. a. D. Guise, "Automatic filtering techniques for three-dimensional kinematics data using 3d motion capture system," in *2006 IEEE International Symposium on Industrial Electronics*, vol. 1, 2006, pp. 614–619.

[20] B. Yu, T. Kienbacher, E. S. Growney, M. E. Johnson, and K.-N. An, "Reproducibility of the kinematics and kinetics of the lower extremity during normal stair-climbing," *Journal of Orthopaedic Research*, vol. 15, no. 3, pp. 348–352, 1997.

[21] M. Abid, N. Mezghani, and Y. Ouakrim, "Representative knee kinematic patterns identification using within-subject variability analysis," in press.

[22] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, *Efficient BackProp*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–48.

[23] Y. Bengio, *Practical Recommendations for Gradient-Based Training of Deep Architectures*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 437–478.

[24] Y. LeCun and Y. Bengio, "The handbook of brain theory and neural networks," M. A. Arbib, Ed. Cambridge, MA, USA: MIT Press, 1998, ch. Convolutional Networks for Images, Speech, and Time Series, pp. 255–258.

[25] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162–169, 2017.

[26] Z. Cui, W. Chen, and Y. Chen, "Multi-scale convolutional neural networks for time series classification," *CoRR*, vol. abs/1603.06995, 2016.

[27] 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*, F. Li, G. Li, S.-w. Hwang, B. Yao, and Z. Zhang, Eds. Cham: Springer International Publishing, 2014, pp. 298–310.

[28] A. Le Guennec, S. Malinowski, and R. Tavenard, "Data Augmentation for Time Series Classification using Convolutional Neural Networks," in *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data*, Riva Del Garda, Italy, 2016.

[29] J. Serrà, S. Pascual, and A. Karatzoglou, "Towards a universal neural network encoder for time series," in *CCIA*, 2018.

[30] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10). Society for Artificial Intelligence and Statistics*, 2010.

[31] T. D. V. Cooke, E. A. Sled, and R. A. Scudamore, "Frontal plane knee alignment: a call for standardized measurement." *The Journal of Rheumatology*, vol. 34, no. 9, pp. 1796–1801, 2007.

[32] R. Cerejo, D. D. Dunlop, S. Cahue, D. Channin, J. Song, and L. Sharma, "The influence of alignment on risk of knee osteoarthritis progression according to baseline stage of disease," *Arthritis & Rheumatism*, vol. 46, no. 10, pp. 2632–2636, 2002.

[33] L. Sharma, J. Song, D. T. Felson, S. Cahue, E. Shamiyeh, and D. D. Dunlop, "The Role of Knee Alignment in Disease Progression and Functional Decline in Knee Osteoarthritis," *JAMA*, vol. 286, no. 2, pp. 188–195, 2001.

[34] N. Mezghani, N. Gaudreault, A. Mitiche, L. Ayoubian, Y. Ouakrim, N. Hagemeister, and J. A. de Guise, "Kinematic gait analysis of workers exposed to knee straining postures by bayes decision rule," *Artificial Intelligence Research*, vol. 4, no. 2, pp. 106–111, 2015.

[35] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2921–2929.