2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST) | 978-1-6654-1847-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/MOCAST52088.2021.9493378

# Fatigue Detection Using Deep Long Short-Term Memory Autoencoders

Konstantinos Balaskas, Kostas Siozios
Aristotle University of Thessaloniki, Department of Physics, Thessaloniki, Greece
Email: kompalas@physics.auth.gr, ksiop@auth.gr

***Abstract*—Efficient time series data mining techniques are an essential part of real world measurement systems and can yield meaningful results from unlabeled data by taking advantage of feature extraction principles. In this paper, we perform kinematic analysis on time series data from IMU sensors for fatigue detection on runners, using several unsupervised machine learning techniques. We propose a robust feature extraction scheme composed of an LSTM Autoencoder, to exploit the advantages of recurrent neural networks and the data compression capabilities of an Autoencoder. The proposed model combines the advantages of several clustering algorithms for accurate fatigue detection in real time, making it suitable for implementation in an embedded device. Experimental evaluation of the feature extraction algorithms showcased their capabilities to produce meaningful features, overcoming the obstacle of extremely limited training data. The inference procedure yielded successful detection in 43% of our representative sample, indicating the efficiency of our model in extracting robust features from unseen kinematic data.**

***Index Terms*—Machine Learning, Fatigue Detection, LSTM, Autoencoders, Clustering, Feature Extraction**

## I. Introduction

Measurement systems and sensor devices require efficient mining of time series data to facilitate real-time applications on the edge. The biggest challenge for data scientists is the extraction of meaningful features from the abundance of raw unlabeled data obtained from such systems, especially when real-time decision-making is needed. Wearable devices on athletes allow for exact monitoring of their movement in real-time and can provide insights on their performance improvement. Fatigue detection is the core indicator of performance limitation and its study reveals crucial information about running gait patterns. However, discovering running features better correlated with fatigue requires prolonged analysis and human expertise. Carefully designed machine learning algorithms alleviate the need for hand-crafted feature extraction mechanisms, saving considerable human labor and provide the necessary performance during inference. While supervised learning is limited by the need of a ground truth target, which is very rare in real-world datasets, unsupervised feature learning techniques operate directly on the unlabeled data and can produce robust features by leveraging data structure.

In this paper, we propose a fully unsupervised framework, composed of an LSTM Autoencoder for flexible feature extraction and clustering algorithms for fatigue detection. The choice of LSTM cells is supported by their ubiquitous employment in time series tasks (e.g. segmentation), whereas the Autoencoder format contributes to data compression/expansion my modifying the network architecture and not the LSTM cells. Experimental results validate our choices by successfully extracting meaningful features and detecting fatigue on a number of unseen kinematic datasets.

## II. Related Work

With the limelight of big data computing on machine learning algorithms, many automated tools for extracting features from sequences have been introduced. Malhotra et al. [11] used stacked LSTM networks for anomaly detection in time series, by using training examples free from any indication of anomalies. Similarly, Marchi et al. [4] exploited the reconstruction error of an LSTM Denoising Autoencoder as activation signal to detect novel acoustic signals. In a more unsupervised approach, Shin et al. [5] aimed to identify organs in magnetic resonance medical images. Single-layer and stacked sparse autoencoders extracted temporal and visual features, thus allowing for a weakly supervised training for the image classifier. In [13], the features from the latent representation of a stacked autoencoder on segmented time series data were compared to each other by a distance measure to detect breakpoints.

Research on running gait patterns and the biomechanical properties that influence them has benefited from the increased availability of wearable measuring devices. Reenalda et al. [7] used IMU's to objectify changes in mechanics over the course of a marathon, using differential analysis on specific kinematic parameters. In a related paper, researchers in [9] attempted to identify changes in biomechanical running gait patterns by classifying inclination condition. Tackling the issue of excessive fatigue during running, De Beeck et al. [10] explored whether machine learning is able to predict the rating of perceived exertion (RPE), constructing a diverse feature space with statistical, sport-science and symmetry features. Hu et al. [2] demonstrated the ability of Recurrent Neural Networks to detect surface conditions and age-group status from an individual's walking behavior. In [3], researchers studied the potential of PCA in combination with pre-trained Self Organizing Maps to classify gait patterns in different time periods in association to complete leg exhaustion, as a result of isokinetic leg exercises.

Most papers focus on the effect of specific environmental and biomechanical parameters to running gait patterns, and do not apply a holistic approach to feature extraction. Machine learning seems to be applied only in the classification part of the workflow, while its capabilities can extend to extracting robust and meaningful features in an abstract manner. Furthermore, in many publications supervised classification models are used to distinguish between different running gait patterns, thus requiring labeled data sets and a universal ground truth, which is rarely available in most databases. In this paper, we overcome these obstacles by operating in a completely unsupervised manner, using machine learning not only as a classifying mechanism, but for feature extraction as well.

## III. Proposed Framework

Our framework is built on the basis that fatigue physically impacts the human body and alters running gait patterns. Under that scope, fatigue detection becomes detached from respiratory reactions of the human body and its kinematic influence can be studied separately. As a guideline, feedback in regards to respiratory behavior of runners was provided only as a sanity check (our model is purely unsupervised). It also proved useful for comparison between respiratory fatigue and "kinematic" fatigue.

### A. Data acquisition

Kinematic data from 14 runners were obtained through a 20-minute evaluation test of cardiorespiratory fitness. IMU sensors were attached to their lower back and cervical area (lower neck) with sampling rate of 200 Hz. The wearable part of the measuring system was composed of a 3-axis accelerometer (ADXL345), a 3-axis gyroscope (ITG3200) and a microcontroller (MSP430FR5969) to control communication protocols and data acquisition from the sensors. Each IMU provided three triaxial biomechanical variables: acceleration, angular velocity and angular displacement. Raw data were filtered using a $2^{nd}$ order Butterworth filter, with cut-off frequency set at 10 Hz. In total, a dataset would include 18 variables.

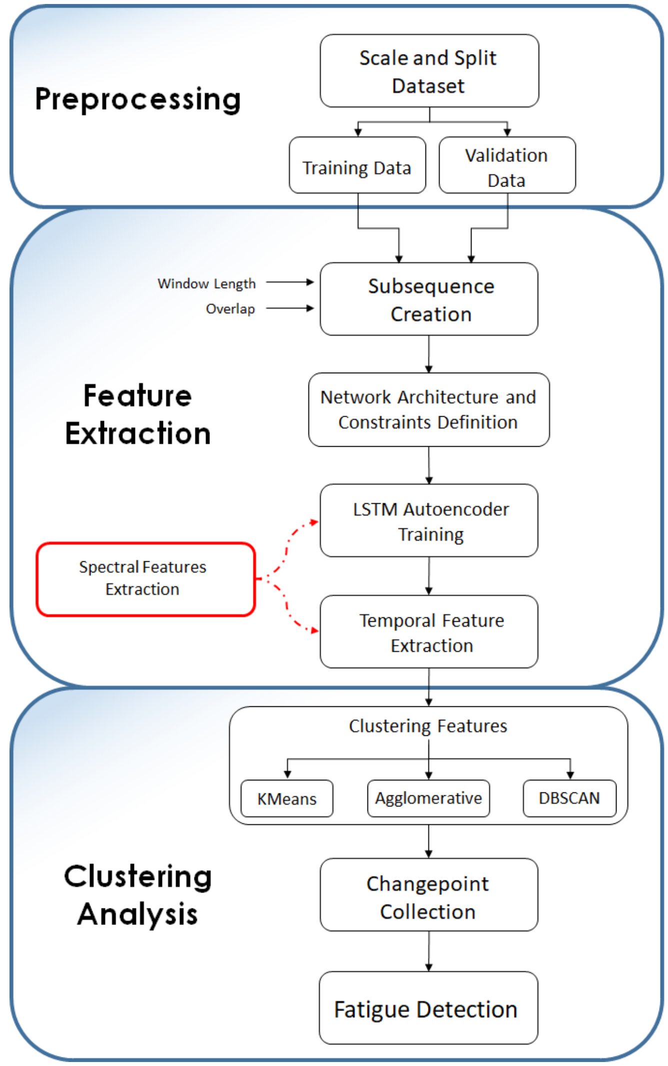



Fig. 1: Algorithmic flow of the training procedure.

### B. Fatigue Detection Algorithmic Flow

Our proposed framework employs different models for the training and inference procedure. The algorithmic flow for the training procedure is shown in Fig. 1. It can be separated in three distinct processes:

*Preprocessing:* Data were scaled and split into the training and validation set, by separating the first 20% of all time series, which roughly consisted of the warming up phase, thus maintaining the temporal relationship between data points. A moving window technique was implemented to create subsequences from the time series, using two user-defined variables called "Window Length" and "Overlap", which determined the subsequence length and step respectively. The effect of both user-defined parameters on the clustering results was evaluated.

*Feature Extraction:* As mentioned, above the feature extractor was an LSTM neural network, constructed in the form of an Autoencoder. Two architectural approaches to the autoencoder format were studied: the standard and well-documented undercomplete approach and an overcomplete approach, whose feature extraction capabilities have not been explored in the bibliography. The latter produces the necessary features at the code layer by applying dimensionality expansion (instead of reduction) to the code layers, making for a larger feature space. To avoid the risk of overfitting, constraints were applied in the training procedure. First, sparsity was introduced, as suggested in [1]. Sparsity is extremely important for effective feature learning, since it discards representation solutions that contain unimportant information (or non-highly activated nodes) [14]. Also, a contractive term was added to the reconstruction training function, so that representations would be locally invariant in many directions of change of the input sequence. Spectral information was provided to the model by applying Fast Fourier Transform (FFT) to the input sequences. Their contribution as either additional training examples or finalized spectral features was studied. We constrained the training of the feature extractor to a single dataset, in order to avoid statistical studies within the representative sample to include all possible variants of running gait patterns. This also helped to showcase our model's invariance to per-individual traits during the inference procedure.

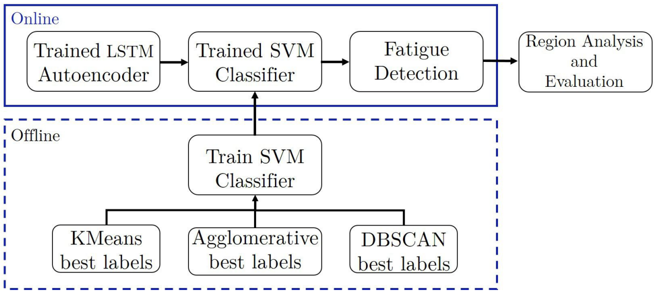
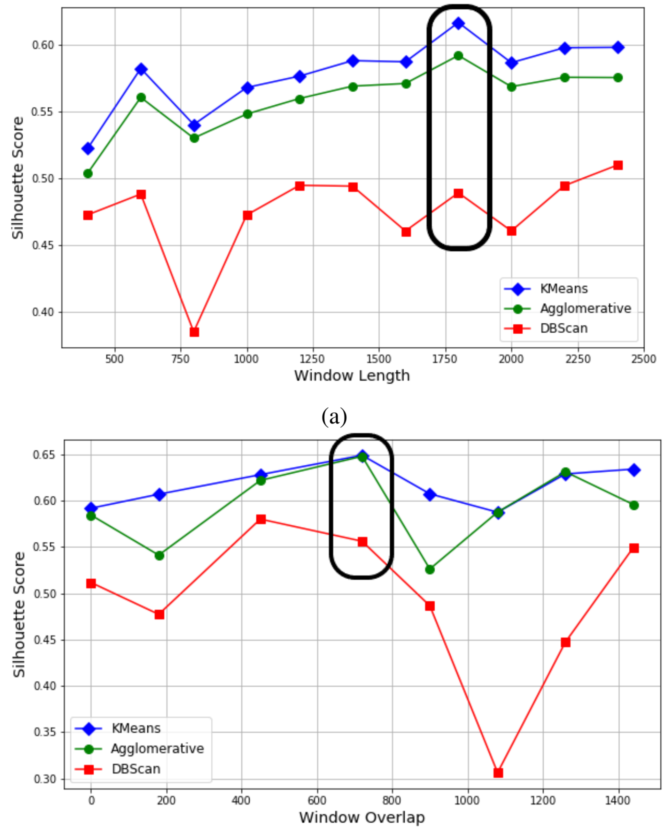



Fig. 2: Algorithmic flow of the inference procedure.




(a)

(b)

Fig. 3: Impact of the window length (a) and the overlap (b) parameter to the clustering results.

*Clustering Analysis:* Time series segmentation can be viewed as a constrained clustering problem: the data points should be grouped together based on their similarity, but with an additional constraint so that data within a group correspond to consecutive time events. Segments that are clustered together correspond to similar running gait patterns and are expected to appear in consecutive time periods. Thus, fatigue is detected at the time interval where a segment of predefined length displays behavior dissimilar to its temporal predecessors, based on a clustering distance metric of our choice (euclidean distance in our case). Binary clustering was the desired outcome, corresponding to separation of subsequences into a pre-fatigue and a post-fatigue states in time. Three clustering algorithms (KMeans, Agglomerative, DBSCAN) were used to cluster the unlabeled features provided by the neural network. Change-point analysis led to the fatigue detection decision. The converging clustering results from all three were evaluated with a weighted version of the Silhouette Score, called Quality of Result, that accounted for intra-cluster coherence, inter-cluster dissimilarity and separation between the pre-fatigue and post-fatigue states.

The inference model depicted in Fig. 2 was applied to unseen kinematic data, using an offline mechanism to support fatigue detection in real time. Using the labels from all three clustering algorithms, an SVM classifier was trained to unify the decision making process into a single algorithm. The modification aimed to increase efficiency and decrease storage complexity.

## IV. Experimental Evaluation

Experimental results from the training procedure concern the architecture and hyperparameter definition of the LSTM Autoencoder and also the shape of the feature space. Spectral features were deemed unnecessary as they dominated the final decision over kinematic features and steered the fatigue detection towards the first indications of respiratory exhaustion. This interesting discovery could be of use to physical educators as a way to study respiratory exhaustion without extensive statistical analyses of respiratory data, but is out of scope of this paper.

At the algorithmic level, the overcomplete LSTM Autoencoder produced features with the slightly higher score on the Quality of Result metric, whilst providing an expanded feature space to the clustering algorithms than its undercomplete counterpart. At the architecture level, the subsequence length and window overlap were set to optimal values of 9 seconds (1800 samples) and 3.6 seconds (720 samples) respectively, through the design space exploration shown in Fig. 3. Higher values of the subsequence length of the original time series provided richer features at the expense of exponentially larger execution time, due to the increased amount of learnable parameters. The overlap between subsequences seemed to only impact the resolution on the final fatigue detecting decision. Hyperparameters concerning the clustering algorithms were set to their optimal values.

The temporal evolution of clustered features can be seen in Fig. 4. Values in the x-axis represent time in the form of subsequence index (or sample number) and y-values correspond to the formed clusters: 1 for the pre-fatigue state and 0 for the post-fatigue. Binary classification was achieved and the clustered features (in color) maintain strong temporal coherence, as the two clusters are clearly separated in specific time intervals. Fatigue is detected at the red dotted vertical line and was located within 6.63-6.73 minutes of the evaluation test of cardiorespiratory fitness. Blue vertical lines represent the end of the warming up phase (left) and the onset of respiratory exhaustion (right). Kinematic fatigue can be detected at a much earlier time, providing physical educators with the necessary information about when running gait patterns change drastically, before fatigue affects the respiratory system. The change point analysis ceases to search for a fatigued state at

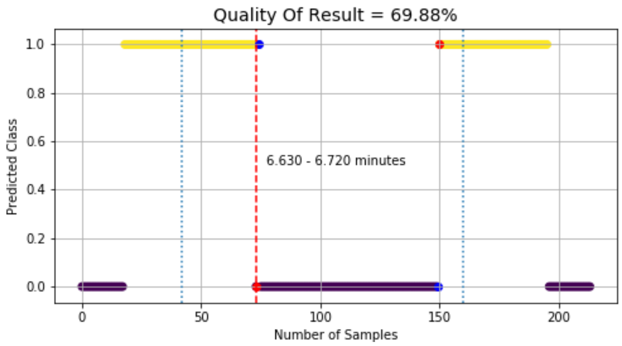



Fig. 4: Temporal evolution of clustered features.

that point, saving crucial computational time during the real time application. Regressing back to running patterns similar to those of the pre-fatigue state is a frequent phenomenon in running gait analysis, but its appearance is out of the scope of this paper.

The proposed inference model of Fig. 2 predicted the temporal occurrence of fatigue in 13 additional datasets, unseen during the training procedure. The Quality of Result metric was used to evaluated the allocation of features into clusters and the correctness of the final decision (i.e. fatigue detection). The collective amount of accurately predicted datasets per step of the Quality of Result metric are presented in Fig. 5. Essentially, Fig. 5 reports accuracy measurements per different runners. The red dotted line represents the threshold above which any detection can be considered valid. This level of confidence was empirically set at 50%. At that level, fatigue was accurately detected at 43% of our representative sample (6 runners), a result that proves our model's ability to confidently generalize its findings to unseen running mechanisms and patterns, overcoming the constraint imposed by the limited training examples. An underlying trade-off between Quality of Result and the range of confident decisions our model can make exists in Fig. 5: by lowering the threshold, fatigue can be predicted in more athletes but less accurately. This can provide useful information to the user when only an approximate but successful detection is needed.

## V. Conclusions

In this paper, a purely unsupervised framework for detecting fatigue in kinematic data was introduced. A novel feature extraction network in the form of an overcomplete LSTM Autoencoder provided robust and meaningful features, as proven by high quality clustering results. The model was able to accurately allocate temporal features to the pre-fatigue or post-fatigue state by recognizing change in running gait patterns, with optimal architecture and decision resolution. An inference model for detecting fatigue in real time achieved 43% accurate detection in our representative sample, despite the limited training of the feature extractor with a single dataset of running kinematic data.

**Acknowledgment:** This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-01681).

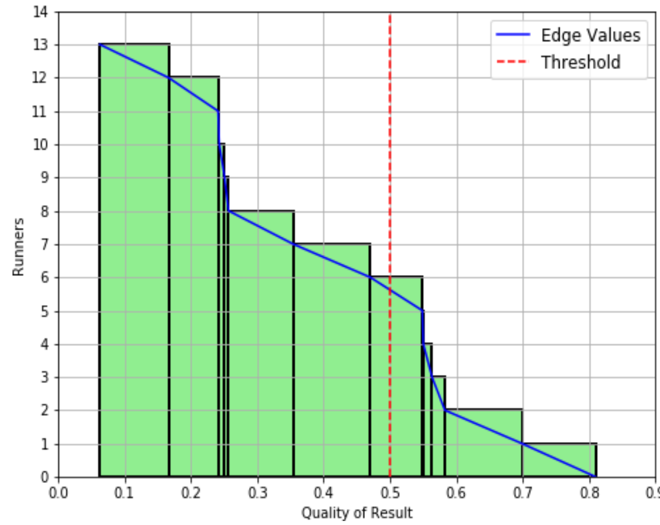



Fig. 5: Collective amount of detection attempts that supersede a threshold on the Quality of Result metric.

## References

[1] A. Makhzani et al. K-sparse autoencoders. *arXiv preprint arXiv:1312.5663*, 2013.

[2] B. Hu et al. Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface-and age-related differences in walking. *Journal of biomechanics*, 71:37–42, 2018.

[3] D. Janssen et al. Diagnosing fatigue in gait patterns by support vector machines and self-organizing maps. *Human movement science*, 30(5):966–975, 2011.

[4] E. Marchi et al. Non-linear prediction with lstm recurrent neural networks for acoustic novelty detection. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2015.

[5] H.C. Shin et al. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1930–1943, 2012.

[6] J. Nogas et al. Fall detection from thermal camera using convolutional lstm autoencoder. In *Proceedings of the 2nd workshop on Aging, Rehabilitation and Independent Assisted Living, IJCAI Workshop*, 2018.

[7] J. Reenalda et al. Continuous three dimensional analysis of running mechanics during a marathon by means of inertial magnetic measurement units to objectify changes in running mechanics. *Journal of biomechanics*, 49(14):3362–3367, 2016.

[8] M. Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[9] N. Ahamed et al. Subject-specific and group-based running pattern classification using a single wearable sensor. *Journal of biomechanics*, 84:227–233, 2019.

[10] O. De Beéck et al. Fatigue prediction in outdoor runners via machine learning and sensor fusion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 606–615, 2018.

[11] P. Malhotra et al. Long short term memory networks for anomaly detection in time series. In *Proceedings*, volume 89. Presses universitaires de Louvain, 2015.

[12] W. H. Press et al. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, USA, 3 edition, 2007.

[13] W.H. Lee et al. Time series segmentation through automatic feature learning. *arXiv preprint arXiv:1801.05394*, 2018.

[14] Y. Netzer et al. Reading digits in natural images with unsupervised feature learning. *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.