# Developing a Start-to-Finish Pipeline for Accelerometer Activity Recognition Using Long Short-Term Memory Recurrent Neural Networks


**Christian McDaniel[1], Shannon Quinn[2]**

[1]University of Georgia, Institute for Artificial Intelligence, Athens, GA, USA

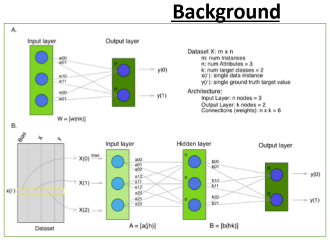
[2]University of Georgia, Department of Computer Science, Athens, GA, USA



## Abstract

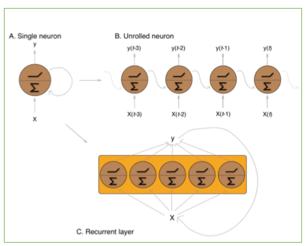
Increased prevalence of smartphones and wearable devices has facilitated the collection of triaxial accelerometer data for numerous Human Activity Recognition (HAR) endeavors. Concurrently, advances in the theory and implementation of Long Short-Term Memory (LSTM) recurrent neural networks (RNNs) has made it possible to process this data in its raw form, enabling on-device online analysis. In this two-part experiment, we have first amassed the results from thirty studies and reported their methods and key findings in a meta-analysis style review. We then used these findings to guide our development of a start-to-finish data analysis pipeline, which we implemented on a commonly used open-source dataset in a proof of concept fashion. The pipeline addresses the large disparities in model hyperparameter settings and ensures the avoidance of potential sources of data leakage that were identified in the literature. Our pipeline uses a heuristic-based algorithm to tune a baseline LSTM model over an expansive hyperparameter search space and trains the resulting model on standardized windowed accelerometer signals alone. We find that we outperform benchmark results from other baseline LSTMs trained on the same model.


## Background

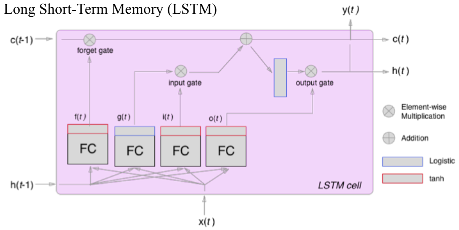


Artificial Neural Networks (ANNs): Data-centric machine learning models which fine-tune layered weight embeddings to learn non-linear data representations.

Recurrent Neural Networks (RNNs): Capture temporal dependencies in data by processing data along the temporal dimension, retaining information from all previous time steps as the input is analyzed.



Long Short-Term Memory (LSTM)



LSTMs enhance the perceptive abilities of RNNs, particularly for long-term dependencies.

- Implementing these networks involves making numerous hyperparameter (HP) selections, yielding an enormous HP search space.
- The tree-structured Parzen (**TPE**) expected improvement (EI) algorithm iteratively suggests, tests, and reupdates "best" HP combinations.
- Human Activity Recognition (**HAR**) research aims to categorize a person's movement through accelerometer time series data.

## Part I: Meta-Analysis Findings

We conducted **a meta-analysis style overview** of the current state of LSTM use for HAR tasks across thirty studies.

Findings:

- General lack of consensus regarding hyperparameter (HP) choices
- HP choices are rarely explained or justified.
- Many reports lacked architecture detail, making reproduction impossible.
- Identified potential sources of data leakage (exposing information about the testing data to the model during training)
- Very little evidence of cross validation or repeated trials

## Part II: The Pipeline

From our observations during Part I, we

1) designed the data analysis pipeline and
2) determined the HP ranges/ values to test during model optimization.

*Data Analysis Pipeline*:

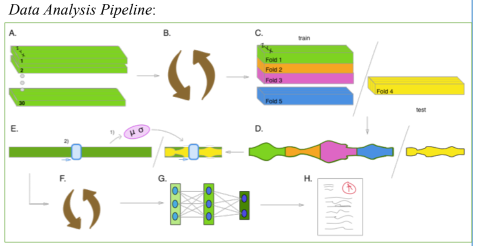


**Figure 4.** Separate data by participant so test-train split doesn't include data from same person (A); use cross-validation when practical (C); standardize after test-train split and use mean and standard deviation from training set only for both training set and test set (E1); shuffle windows to smooth the learning process (F).

*Hyperparameter Ranges (to be Tuned by TPE Algorithm):*

| Category | Hyperparameter | Range |
|---|---|---|
| Data Processing | Window Size | 24, 48, 64, 128, 192, 256 |
| | Stride | 25%, 50%, 75% |
| | Batch Size | 32, 64, 128, ..., 480 |
| Archi-tecture | Units | 2, 22, 42, 62, ..., 522 |
| | Layers | 1, 2, 3 |
| Forward Processing | Activation Function (unit, state) | softmax, tanh, sigmoid, ReLU, linear |
| | Bias | True, False |
| | Weight Initialization (cell, state) | zeros, ones, random uniform dist., random normal dist., constant (0.1), orthogonal, Lecun normal, Glorot uniform |
| Regular-ization | Regularization (cell, state, bias, activation) | None, L2 Norm, L1 Norm |
| | Weight Dropout (unit, state) | uniform distribution (0, 1) |
| | Batch normalization | True, False |
| Learning | Optimizers | SGD, RMSProp, Adagrad, Adadelta, Nadam, Adam |
| | Learning Rate | $10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$ |

**Table 1.** Values/ranges included for Model Hyperparameter Optimization prior to final cross validation on data. "Unit" and "cell" refer to the LSTM cells' outputs, "State" refers to the recurrent weight matrices.

## Materials

**Dataset**: Human Activity Recognition Using Smartphones from Anguita, et. al. available on the University of California at Irvine's Machine Learning Repository (UCI HAR Dataset)

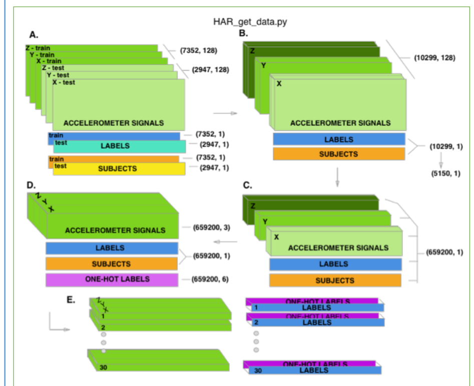



**Figure 5.** Data provided as axis-wise windowed signals pre-split into training and testing sets. Before passing through our pipeline, we processed the data so as to undo as much of this preprocessing as possible, ending with a single set of triaxial signals split by record.

**Languages and Libraries:**

- TPE HP optimization: Hyperas from Hyperopt
- LSTM: Keras Keras with TensorFlow backend
- Grid search, Standardization, Cross Validation: Scikit Learn
- Data Formatting, etc.: NumPy, Pandas

## Results

Most models found in literature:

- High-dimensional data (hand crafted features, multiple sensors and modalities
- Highly processed data (remove gravity component, etc.)
- Complex architecture

**Our models:** *only* triaxial **total accelerometer signals**;

- **least preprocessing** and **lowest number of features used**
- **Baseline LSTM** (simple architecture to save computational expense)

| Model Name | Performance | Features |
|---|---|---|
| Baseline LSTM1 | 90.77% | 9 (T, B, G) |
| Baseline LSTM2 | 85.35% | 3 – 9 (?) |
| **Pipeline Part 1 (Best)** | **94.96%** | **3 (T)** |
| **Pipeline Part 2 (CV) Acc** | **90.97%** | **3 (T)** |
| **Pipeline Part 2 (CV) F1** | **0.91*** | **3 (T)** |
| **Pipeline Part 2 (Best) Acc** | **95.25%** | **3 (T)** |
| **Pipeline Part 2 (Best) F1** | **0.96*** | **3 (T)** |

**Table 2.** The results from the benchmark models using the UCI HAR dataset. Pipeline Part 1 = best test score from hyperparameter optimization; Pipeline Part 2 = final cross validation on data; CV = cross validation; "Best" = single best training trial.

## Conclusions

We demonstrate the ability for a **baseline LSTM** model trained solely on **raw triaxial accelerometer data** to outperform other baseline models trained on this data and perform competitively with classical models trained on hundreds of hand-crafted features as well as other more complex LSTM models trained on higher dimensional sensor data.

We demonstrate the ability to optimize a data-centric model over an **expansive hyperparameter search space** and train it end-to-end within a scientifically rigorous and deliberate **Data Analysis Pipeline.**

Going forward, we would like to repeat this experiment to average performances from different models returned by the TPE algorithm; we would also like to repeat this experiment on other HAR datasets. Further exploration should be done to analyze why the algorithm's selections are indeed superior, how different data affect these choices, and how the LSTM cells within the models themselves are representing this type of data as has been done with LSTMs in other domains.

We hope that this Pipeline will serve useful in producing explicit and reproducible experiment results and in pushing the field forward in a methodical way.

## References


Dataset:

- Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz. Energy efficient smartphone-based activity recognition using fixed-point arithmetic. *Journal of Universal Computer Science, Special Issue in Ambient Assisted Living - Home Care*, 2013.
- Downloaded from UCI Machine Learning Repository

TPE:

- James Bergstra, Remi Bardenet, Yoshua Bengio, and Balazs Kegl. Algorithms for hyper-parameter optimization. *NIPS*, 2011.

ANN/RNN/LSTM and Data Analysis Background:

- ANN: WS McCulloch and W Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 5(4):115–133, 1943.
- RNN: Ronald J Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1989.
- LSTM: Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. *Neural Computation*, 1997.
- Aurelien Geron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. OReilly Media, Inc., Sebastopol, CA, 2017.
- John A Miller. Introduction to data science using scalation. Technical report, University of Georgia Department of Com- puting Science, 2018.

Full list of citations for literature review and other aspects of experiment in paper with the same name and on GitHub at the link listed below.


## Acknowledgements & Contacts


Christian McDaniel, clm121@uga.edu; Shannon Quinn

Quinn Research Group



Thanks to the UGA IAI for providing the computing resources needed to carry out this project. Thanks to SciPy 2018.


Code available on GitHub: https://github.com/xtianmcd/accelstm