II-10

Case Study on Small-Scale Dynamic Neural Networks Explainability

Martynas DUMPIS[1], Dalius NAVAKAUSKAS[1]

E-mail: martynas.dumpis@vilniustech.lt, [1]Department of Electronic Systems, Vilnius Gediminas Technical University

Introduction

Dynamic artificial neural networks (DNN) are increasingly used for time-varying data processing, yet their explainability remains a challenge. This study investigates small-scale DNN, particularly Long Short-Term Memory (LSTM) and Finite Impulse Response Neural Networks (FIRNN), focusing on their explainability in human activity classification tasks. By combining Layer-wise Relevance Propagation (LRP) and Fast Fourier Transform (FFT) techniques, this research provides complementary insights into DNN decision-making mechanisms.

Aims

The goal of this study is to enhance the understanding of decision-making processes in small-scale dynamic neural networks through explainability techniques.

The main objectives:

a) evaluate the classification accuracy and complexity of FIRNN and LSTM used for a selected human activity classification task;

b) employ the LRP technique to investigate the contribution of time-varying signal samples to FIRNN and LSTM decisions;

c) employing the FFT technique to gain insights into input signal spectral characteristics and their possible connection with FIRNN filter order or LSTM hidden cells' number.

DNN Architectures and Training

The accelerometer magnitude data (1,710 s captured human activities – walking and running), was used to train and evaluate FIRNNs and LSTMs with three-layer 1-(m)-1-1 architectures, here $m \in [1, 100]$ – filter order or cells' number that was varied. The 100 training attempts (stopping when validation error increases) per configuration resulted in 10,000 trained DNNs.

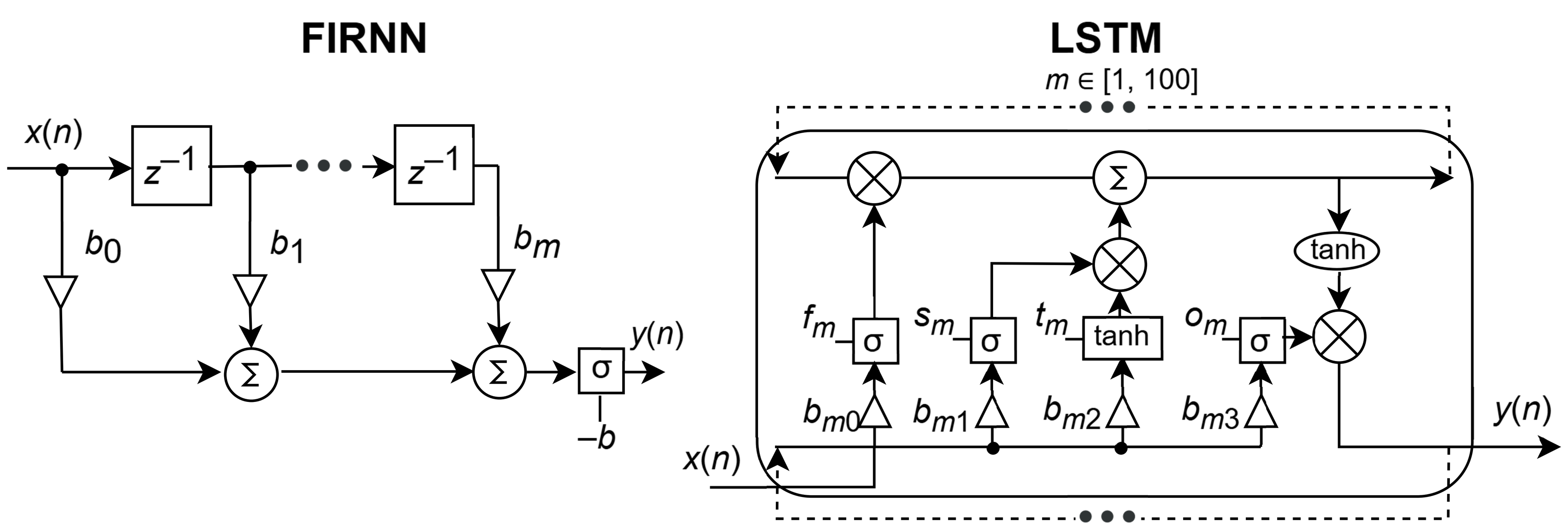
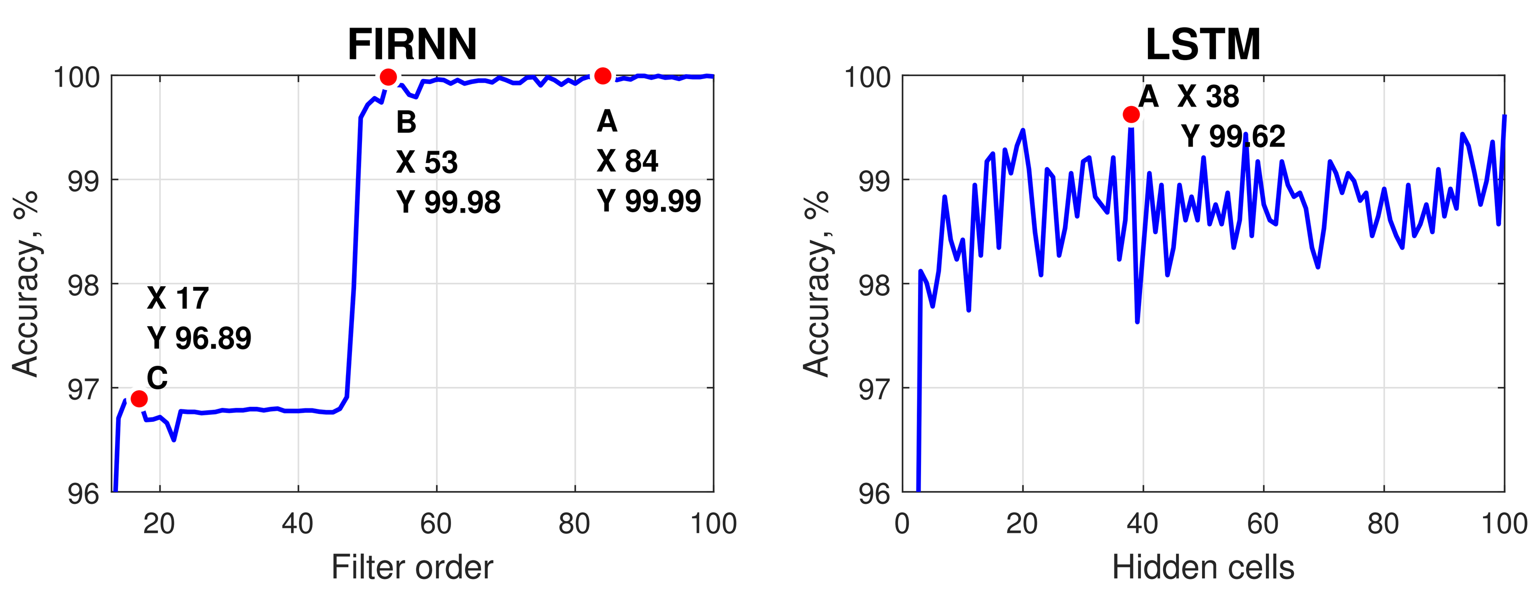


Illustration of synapses in the hidden layer of DNNs

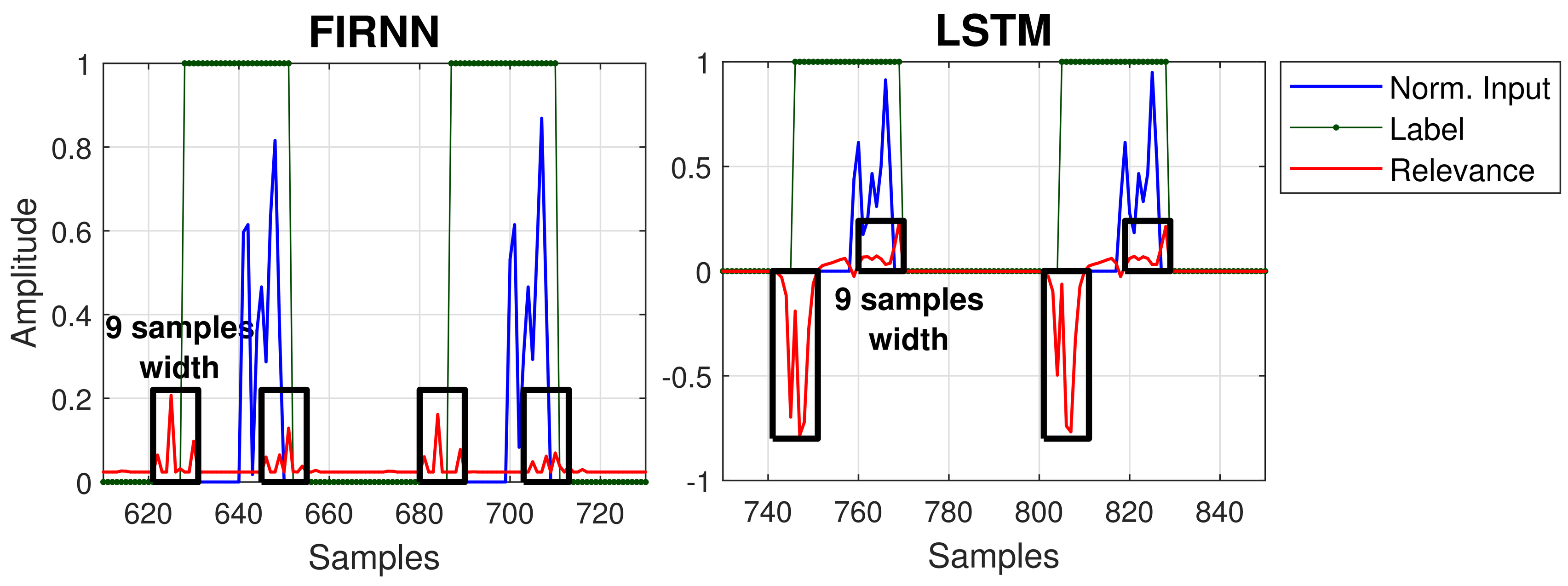
The highest accuracy reaching DNNs per configuration were considered.



The highest accuracy of DNNs varying filter order or hidden cells' number

LRP Analysis

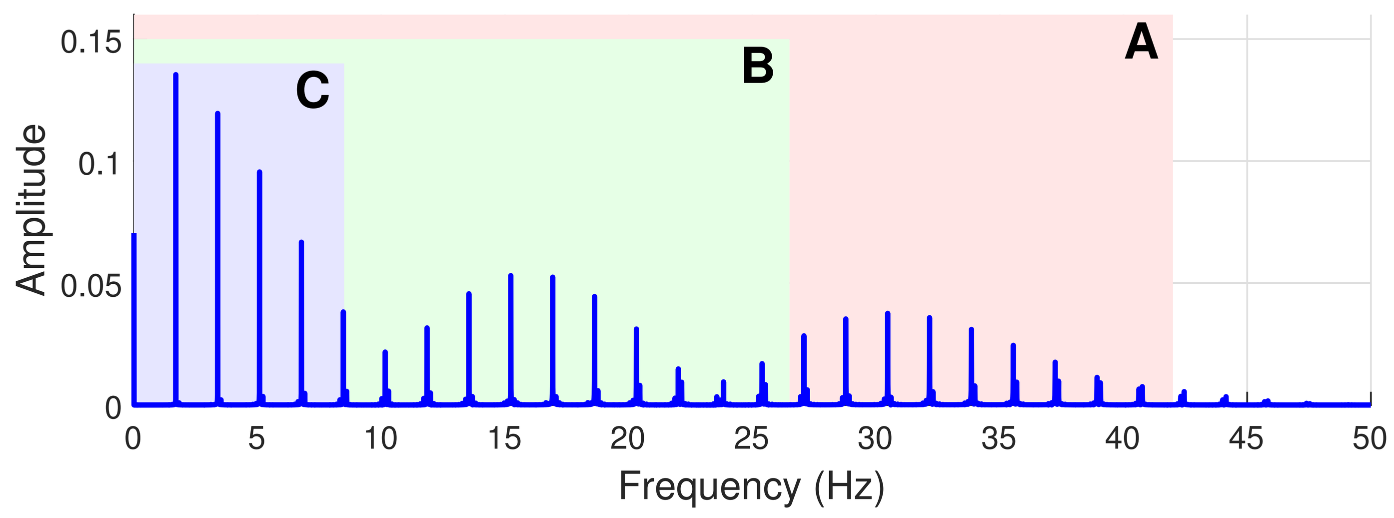
The highest accuracy reaching FIRNN and LSTM configurations ("A") were used for LRP analysis. For each localized peak or valley, the number of samples associated with these specific features was calculated to determine how much input data contributes to the DNN's decision-making process.



Example of the relevance for the highest accuracy reaching DNNs superimposed on a normalized input signal and classification labels

Frequency Analysis

The FFT was applied to the input signal, sampled at 100 Hz, to analyze its frequency composition. Key frequency ranges were identified, corresponding to the periodic components most relevant to the DNN's performance.



The magnitude of the input signal with outlined three key frequency ranges (A–C)

Results

- **FIRNN:** 99.99% accuracy, 85 binary multiplications, 85 binary additions, 2 activation functions.
- **LSTM:** 99.62% accuracy, 6,080 binary multiplications, 6,004 binary additions, 191 activation functions.
- **LRP analysis:** 9 samples are the most relevant for class separation.
- **Frequency analysis:** key freq. ranges 8.5 Hz, 26.5 Hz, and 42 Hz.

Conclusions

1. FIRNN achieves highest accuracy with significantly lower complexity, requiring 71 times less additions/multiplications and 95 times less activation functions compared to LSTM.

2. LRP analysis shows that both DNNs rely on the similar width sample size for class transitions, but their use (relevance) differs depending on the DNNs type.

3. Frequency analysis confirms that only in FIRNN case filter order can be inferred from the input signal magnitude, avoiding the need for an extensive full search of the highest accuracy reaching structure and enabling granuality (A–C key frequency ranges) in the selection of the FIRNN complexity.