Sensor Sequential Data-Stream Classification Using Deep Gated Hybrid Architecture

Nelly Elsayed[†] Anthony S. Maida[†] Magdy Bayoumi[⋆]
[†]*School of Computing and Informatics*
[⋆]*Electrical and Computer Engineering Department*
[†⋆]*University of Louisiana at Lafayette, USA*

***Abstract*—Sensors are the main components to supply information for resource management in a smart city. This paper studies the sensor data-stream classification problem using different time series state-of-the-art classification models. In this study, we found that the hybrid architecture of gated recurrent units and temporal fully convolutional neural network (GRU-FCN) model outperforms the existing state-of-the-art classification techniques in most of the benchmark sensor-obtained datasets. Moreover, the GRU-FCN model is simpler than the other existing gate-based recurrent classification architectures. Thus, it is an appropriate model to be implemented on small or portable hardware devices.**

***Index Terms*—Deep learning, GRU, GRU-FCN, convolution neural network, sensors, time series, smart city.**

I. Introduction

A smart city focuses on its most demanding needs and on the greatest opportunities to improve lives. The goal of creating a smart city is to give a good quality of life to its citizens, and achieve a sustainable clean environment [1]. A smart city could be defined as an urban area that uses different types of digital electronic data collector devices such as sensors, cameras to supply information which is used to manage resources and assets efficiently. This data information is processed and analyzed to monitor and manage law enforcement, traffic and transportation systems, parking control, water supply networks, electricity supply network, waste management, smart marketing, hospitals, schools, colleges, parks, power plants, information systems, libraries, and other community services [2]–[4].

A sensor is a device, subsystem, or module whose purpose is to detect any changes or events in its surrounding environment. Then, the sensor sends the obtained information to other electronic devices such as computers or cell phones. There are various types of sensors. The most common sensors that used in daily applications are: light sensor, color sensor, alcohol sensor, touch sensor, proximity sensor, accelerometer, infrared sensor (IR sensor), ultrasonic sensor, flow sensor, level sensor, temperature sensor, pressure sensor, smoke sensor, and gas sensor [5]. Sensors are the main components to build a smart city because they are easy to use and to apply. However, processing and analyzing the data obtained by sensors is a difficult process due to a large amount of data, especially foe sensor data-streams.

A sensor data-stream is a type of time series data. There are several proposed methods for univariate and multivariate time series classification. However, there is a little literature that focus only on the sensor data-stream classification problem. In this paper, we examine several time series state-of-the-art classification methods in the sensor data-streams classification task. We found that the deep gated recurrent and convolutional network hybrid model (GRU-FCN) which was proposed in [6] has a significant improvement in classification accuracy on most of the UCR benchmark sensor-obtained datasets [7] compared to the state-of-the-art classification methods. The GRU-FCN is inspired by the proposed model in [8] which combines the long short-term memory (LSTM) with a temporal fully convolutional neural network (FCN) for time series classification. However, the GRU-FCN combines the gated recurrent unit (GRU) and a temporal fully convolutional neural network. The GRU-FCN model does not have any additional supporting algorithms such as attention algorithm or masking to increase the performance comparing to the LSTM based classification architectures [6]. Furthermore, the GRU has a smaller number of parameters and smaller architecture than the LSTM. Therefore, the GRU is preferable to be implemented on small or portable devices. Moreover, due to the temporal convolution architecture, GRU-FCN does not require any data preprocessing in prior to the training process [9]. Therefore, it is preferable to be used in sensors data-related problems. We tested the GRU-FCN model and the state-of-the-art classification methods on different UCR [7] benchmark Sensor-source obtained datasets. We found that the GRU-FCN model outperforms the other state-of-the-art classification techniques in the classification accuracy.

II. GRU-FCN Model Architecture

The GRU-FCN model proposed in [6] based on the framework introduced in [8], [9]. The GRU-FCN model architecture is shown in Figure 1. The GRU-FCN architecture has two parallel parts: a GRU and a temporal FCN. The GRU-FCN model uses three layers of temporal FCN architecture proposed in [9]. The FCN layers learn to extract feature representations from the data without prior data reprocessing or feature engineering requirements. Therefore, the temporal FCN is responsible for feature extraction from the sensor-source datasets [9]. The GRU enables the model to recognize temporal dependencies within these sequential data-streams [10]. Therefore, the GRU-FCN model can learn both the features and temporal dependencies to predict the correct class for each sensor data

978-1-7281-1457-6/19/$31.00 ©2019 IEEE

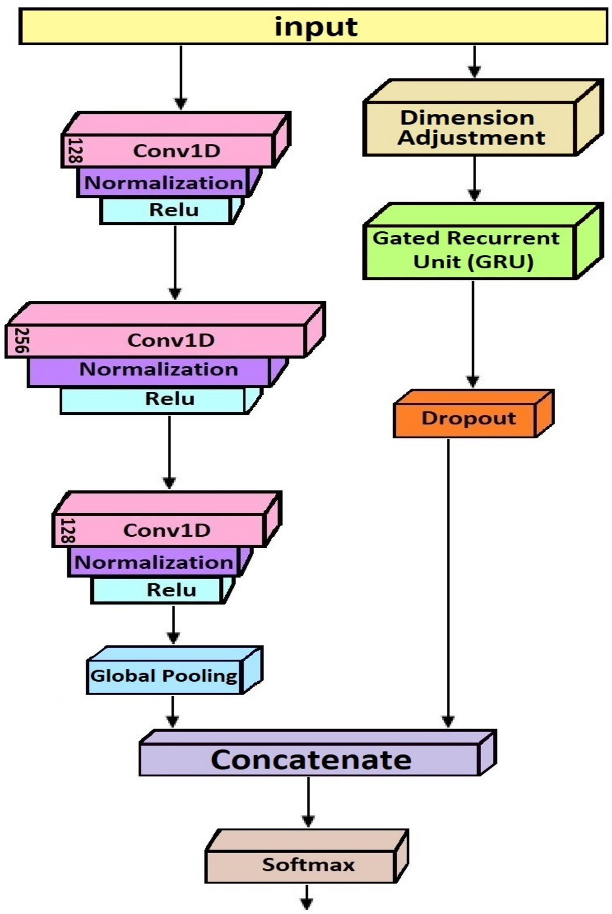

Fig. 1: The GRU-FCN model architecture proposed in [6] which was inspired by architectures in [8], [9].

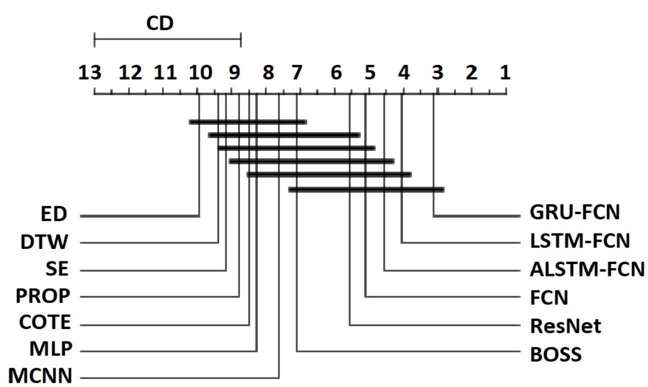

Fig. 2: Critical difference diagram based on arithmetic mean of model ranks.

training. Moreover, the GRU has smaller architecture than the LSTM. Therefore, the number of parameters, memory requirements and training time of a GRU are smaller than an LSTM [6], [10].

III. Experiments and Results

In our experiment, we examine several state-of-the-art classification models. We selected FCN [9] as a temporal convolutional network based model, LSTM-FCN [8], and ALSTM-FCN [8] as long short-term memory and fully convolutional networks based models, ResNet [9] which based on convolutional residual networks, MCNN [11] that is based on convolutional networks, and MLP [9] which based on multilayer perceptrons. In addition, we selected the COTE [12], DTW [13], PROP [14], BOSS [15], and SE [12] as the most widely used state-of-the-art classification models.

A. Datasets

We conducted our experiments on 18 different sensor-source obtained datasets from the UCR benchmark (2015) [7]. Each dataset was divided into training and testing data. The description of number of classes, training data size, testing data size, data length, number of epochs, and train and test batch sizes of each dataset used in our experiments are shown in Table I. The size of training and testing datasets was chosen according to the UCR benchmark [7] description.

B. Results

We regenerated the GRU-FCN [6], FCN [9], LSTM-FCN [8], ALSTM-FCN [8], and MLP [9] models to obtain their classification results. We used the same internalizations that have been used by the original literature proposed by each these classification models. For ResNet [9], MCNN [11], COTE [12], DTW [13], PROP [14], BOSS [15], SE [12], and ED [7] we obtained their results from their proposed literature. The results of each classification model over the sensor-obtained UCR benchmark 2015 [7] are shown in Table II where the symbol / means that for the certain dataset, the corresponding classification model does not have a published result and we were not able to regenerate the model to obtain this certain result due to hardware resources limitations and time limited consumptions. The GRU-FCN shows the lowest classification error for 10 out of 18 datasets. The GRU-FCN also shows the overall smallest classification error, mean per-class classification error (MPCE), and arithmetic average rank compared to the other state-of-the-art models as shown in Table II.

Figure 2 shows the critical difference diagram [16] of the GRU-FCN and the state-of-the-art models. The critical difference diagram based on the rank arithmetic mean of each classification model on the UCR benchmark sensor-obtained datasets using Bonferroni-Dunn or Nemenyi test [17] with $\alpha = 0.05$. This graph shows that the GRU-FCN has significant classification accuracy improvement compared to the other state-of-the-art models.

Figures 3, 5, and 4 show the loss value through the training and validation processes using the recurrent-based classification models (GRU-FCN, LSTM-FCN, and ALSTM-FCN) over the Car, ChlorineCon, and CincECGTorso datasets. These figures show that the average difference between both training and validation processes is the lowest for GRU-FCN model over the Car, ChlorineCon, and CincECGTorso datasets.

TABLE I: UCR 18 Sensor-source obtained datasets descriptions [7].

Dataset	***# Classes***	***Train size***	***Test size***	***Length***	***# epochs***	***Train Batch***	***Test Batch***
Car	4	132	172	577	2000	128	128
ChlorineConc	3	467	3840	166	2000	128	128
CinCECGTorso	4	40	1380	1639	500	128	128
Earthquakes	2	322	139	512	2000	128	128
FordA	2	3601	1320	500	2000	128	128
FordB	2	3636	810	500	1600	128	128
InsectWingbeatSound	11	220	1980	256	1000	128	128
ItalyPower	2	67	1029	24	2000	64	128
Lightning2	2	60	61	637	4000	128	128
Lightning7	7	70	73	319	3000	32	32
MoteStrain	2	20	1252	84	2000	128	128
Phoneme	39	214	1896	1024	2000	64	128
Plane	7	105	105	144	200	16	16
SonyAIBORobotI	2	20	601	70	2000	64	128
SonyAIBORobotII	2	27	953	65	2000	64	128
StarLightCurves	3	1000	8236	1024	2000	64	64
Trace	4	100	100	275	1000	64	128
Wafer	2	1000	6164	152	1500	64	64

TABLE II: Classification testing error for 18 Sensor-source obtained datasets from the UCR [7] benchmark.

Dataset	**Classification Method and Testing Error**												
	GRU-FCN	***FCN***	***LSTMFCN***	***ALSTMFCN***	***ResNet***	***MCNN***	***MLP***	***COTE***	***DTW***	***PROP***	***BOSS***	***SE***	***ED***
Car	**0.016**	0.050	0.174	0.174	0.067	/	0.117	/	0.267	/	/	/	0.267
ChlorineCon	**0.002**	0.157	0.191	0.193	0.172	0.203	0.125	0.314	0.352	0.360	0.334	0.312	0.350
CinCECGTorso	0.124	0.187	0.191	0.193	0.172	0.058	0.158	0.064	0.349	0.062	0.125	**0.021**	0.103
Earthquakes	**0.171**	0.199	0.177	0.173	0.214	/	0.208	/	0.281	0.281	0.186	/	0.288
FordA	0.074	0.094	**0.072**	0.073	**0.072**	/	0.231	/	0.444	0.182	0.083	/	0.335
FordB	0.083	0.117	0.088	**0.081**	0.100	/	0.371	/	0.38	0.265	0.109	/	0.394
InsectWingbeatSound	0.446	0.598	0.342	**0.329**	0.469	/	0.369	/	0.643	/	0.479	/	0.438
ItalyPower	**0.027**	0.030	0.037	0.040	0.040	0.030	0.034	0.036	0.050	0.039	0.053	0.053	0.045
Lightning2	0.246	0.197	0.197	0.213	0.246	0.164	0.279	0.164	0.131	0.115	0.148	**0.098**	0.246
Lightning7	**0.137**	**0.137**	0.164	0.178	0.164	0.219	0.356	0.247	0.274	0.233	0.342	0.274	0.425
MoteStrain	0.076	**0.050**	0.061	0.064	0.105	0.079	0.131	0.085	0.165	0.114	0.073	0.113	0.121
Phoneme	0.644	0.655	0.650	**0.640**	0.676	/	0.902	/	0.772	/	0.733	/	0.891
Plane	**0**	**0**	**0**	**0**	**0**	/	0.019	/	**0**	/	/	/	0.038
SonyAIBORI	**0.017**	0.032	0.018	0.030	0.015	0.230	0.273	0.146	0.275	0.293	0.321	0.238	0.305
SonyAIBORII	**0.018**	0.038	0.022	0.025	0.038	0.070	0.161	0.076	0.169	0.124	0.098	0.066	0.141
StarLightCurves	0.025	0.033	0.024	0.023	0.029	0.023	0.043	0.031	0.093	0.079	**0.021**	0.093	0.151
Trace	**0**	**0**	**0**	**0**	**0**	**0**	0.180	0.010	**0**	0.010	**0**	0.05	0.240
Wafer	**0.001**	0.003	**0.001**	0.002	0.003	0.002	0.004	0.001	0.020	0.003	**0.001**	0.002	0.005
best	**10**	4	4	5	3	1	0	0	2	0	3	2	0
MPCE	**0.025**	0.033	0.032	0.033	0.035	0.134	0.063	0.135	0.080	0.079	0.066	0.138	0.077
AVG Rank	**3.111**	5.139	3.861	4.556	5.583	7.611	8.333	8.444	9.389	8.778	7.083	9.139	9.944

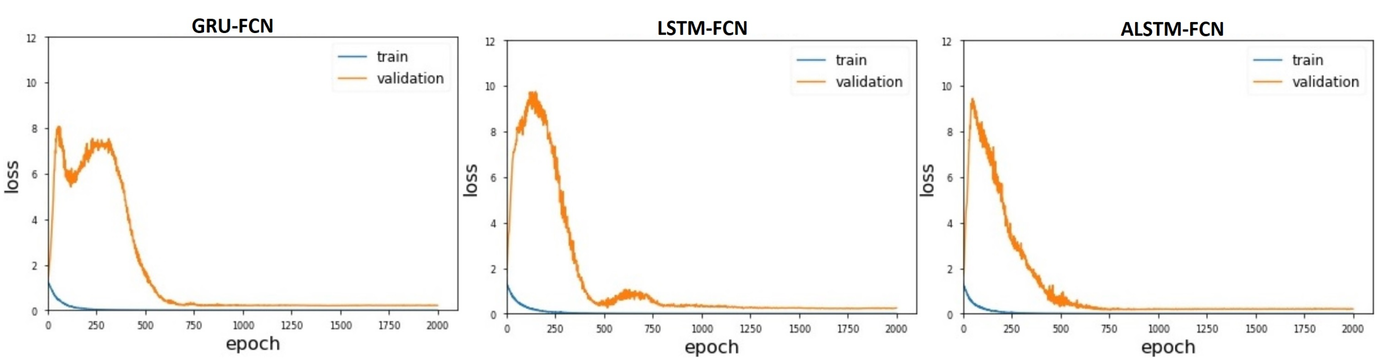

Fig. 3: The loss value of GRU-FCN, LSTM-FCN, and ALSTM-FCN models over the Car dataset training and validation processes.

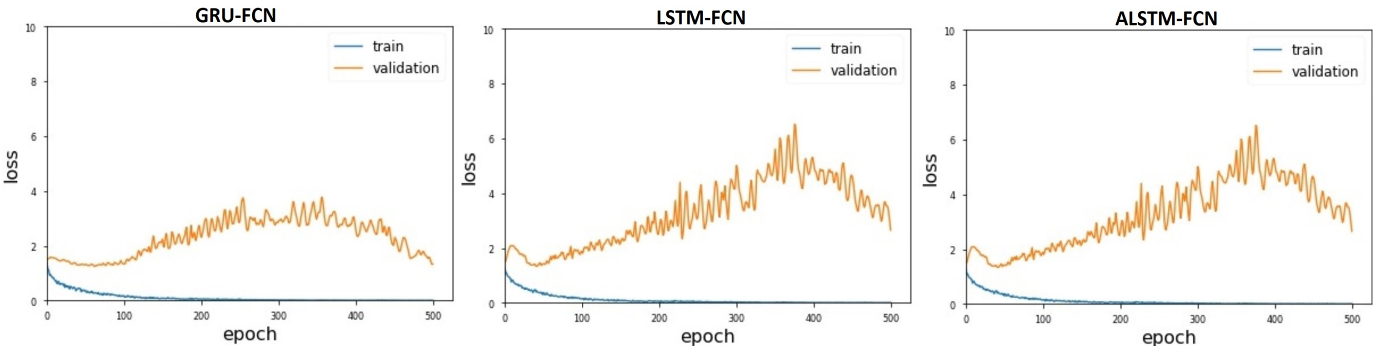

Fig. 4: The loss value of GRU-FCN, LSTM-FCN, and ALSTM-FCN models over the CinECGTorso dataset training and validation processes.

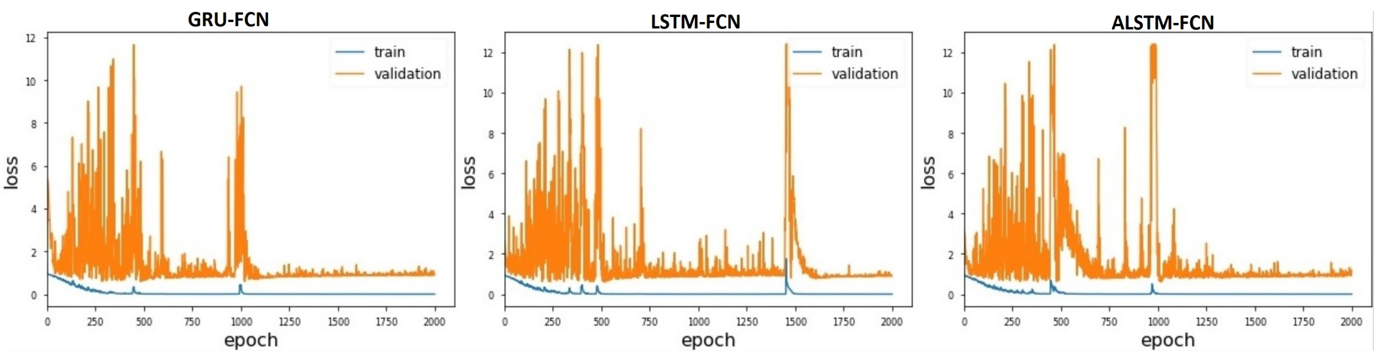

Fig. 5: The loss value of GRU-FCN, LSTM-FCN, and ALSTM-FCN models over the ChlorineCon dataset training and validation processes.

IV. Conclusion

Processing sensor-obtained data is an essential component of building an efficient smart city. There are several successful classification models for different sensor-obtained times series classification tasks. However, we empirically found that the GRU-FCN classification model enhances the classification accuracy without requiring extra algorithmic support compared to the existing recurrent-based classification models. Moreover, the GRU-FCN classification model achieves the performance of state-of-the-art models and has the highest average arithmetic ranking and the lowest mean-per-class error (MPCE) through the UCR benchmark sensor-source obtained datasets classification compared to the state-of-the-art models.

References

[1] R. P. Dameri and C. Rosenthal-Sabroux, "Smart and digital city : A systematic literature review," 2017.

[2] M. Angelidou, "Smart city policies: A spatial approach," *Cities*, vol. 41, pp. S3–S11, 2014.

[3] V. Albino, U. Berardi, and R. M. Dangelico, "Smart cities: Definitions, dimensions, performance, and initiatives," *Journal of Urban Technology*, vol. 22, no. 1, pp. 3–21, 2015.

[4] A. Vanolo, "Smartmentality: The smart city as disciplinary strategy," *Urban Studies*, vol. 51, no. 5, pp. 883–898, 2014.

[5] M. Kretschmar, S. Welsby, and L. Precision, "Capacitive and inductive displacement sensors," in *Sensor Technology Handbook*. Elsevier, 2005, pp. 193–222.

[6] N. Elsayed, A. S. Maida, and M. Bayoumi, "Deep gated recurrent and convolutional network hybrid model for univariate time series classification," *arXiv preprint arXiv:1812.07683*, 2018.

[7] 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/.

[8] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

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

[10] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[11] Z. Cui, W. Chen, and Y. Chen, "Multi-scale convolutional neural networks for time series classification," *arXiv preprint arXiv:1603.06995*, 2016.

[12] A. Bagnall, J. Lines, J. Hills, and A. Bostrom, "Time-series classification with COTE: the collective of transformation-based ensembles," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 9, pp. 2522–2535, 2015.

[13] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu, "Weighted dynamic time warping for time series classification," *Pattern Recognition*, vol. 44, no. 9, pp. 2231–2240, 2011.

[14] J. Lines and A. Bagnall, "Time series classification with ensembles of elastic distance measures," *Data Mining and Knowledge Discovery*, vol. 29, no. 3, pp. 565–592, 2015.

[15] P. Schäfer, "The BOSS is concerned with time series classification in the presence of noise," *Data Mining and Knowledge Discovery*, vol. 29, no. 6, pp. 1505–1530, 2015.

[16] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, no. Jan, pp. 1–30, 2006.

[17] T. Pohlert, "The pairwise multiple comparison of mean ranks package (PMCMR)," *R Package*, vol. 27, 2014.