B.TECH PROJECT REPORT

On

Light Weight Dilated CNN for Activity Classification and Intent Recognition

Pranav Khanna

ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
DECEMBER, 2019

Light Weight Dilated CNN for Activity Classification and Intent Recognition

A PROJECT REPORT

Submitted in partial fulfillment of the requirments for the award of the degrees

of
BACHELOR OF TECHNOLOGY
in

ELECTRICAL ENGINEERING

Submitted by :
Pranav Khanna

Guided by :
Dr. Vivek Kanhangad and Dr Apurva Narayan(External Supervisor)

INDIAN INSTITUTE OF TECHNOLOGY INDORE
DECEMBER, 2019

Candidate's Declaration

I hereby declare that the project entitled **Light Weight Dilated CNN for Activity Classification and Intent Recognition** submitted in partial fulfillment for the award of the degree of Bachelor of Technology in Electrical Engineering completed under the supervision of **Dr. Vivek Kanhangad and Dr Apurva Narayan(External Supervisor)** IIT Indore is an authentic work.

Further, I/we declare that I/we have not submitted this work for the award of any other degree elsewhere.

Pranav Khanna
160002023
B.Tech, Electrical Engineering
IIT Indore, India

Certificate by BTP Guide

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

Dr. Vivek Kanhangad
Associate Professor
Discipline of Electrical Engineering
IIT Indore

Preface

This report on "Light Weight Dilated CNN for Activity Classification and Intent Recognition" is prepared under the guidance of Dr Apurva Narayan and Dr Vivek Kanhangad. In this report, we highlight the benefits of dilated convolutional networks and introduce the concept of Human Intent Recognition. The report highlights the computational efficiency of dilated CNNs and compares its performance with Recurrent Neural Networks like LSTM and GRU. Through this report we have also described the procedure for creating the Intent Recognition dataset and presented our results on this new dataset.

Pranav Khanna
160002023
B.Tech, Electrical Engineering
IIT Indore, India

Acknowledgement

I wish to thank Dr. Vivek Kanhangad and Dr Apurva Narayan(External Supervisor) for their kind support and valuable guidance. It is their help and support, due to which I was able to complete the model and technical report. I have to express my appreciation to the guides for sharing their pearls of wisdom with me during the course of this research. Without their support this project would have been a distant reality. I am thankful to my family who were always there and supported me unconditionally and also to our friends who were there to cheer me up anytime we needed them. I would also like to thank the Institute and the Department of Electrical Engineering for providing me with the opportunity to work on this project.

Pranav Khanna
160002023
B.Tech, Electrical Engineering
IIT Indore, India

Abstract

Time series data is available from a diverse set of sensors in real life. It is of prime importance in the domain of machine learning and artificial intelligence to analyze such data and identify outliers or anomalies, characteristic of the underlying activities and predict the future. Traditionally, time-series analysis involves identifying features using exploratory data analysis and using statistical approaches for classification and prediction. However, with the advent of convolutional neural networks (CNN), our ability to extract features automatically has substantially improved. In this project, we propose a novel light-weight deep learning architecture of dilated CNN for classification and predicting time series data sets. We evaluate our model on a real-world *human activity recognition* time series data set and a synthetically crafted pseudo-realistic dataset for *human intent recognition*. Our model outperforms the state-of-the-art models and is light-weight.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Today sensors are ubiquitous and can be found in almost any technical device ranging from washing machines to pacemakers and wearable devices. These sensors produce large quantities of data in the form of time series. Time series analysis finds application in a wide range of fields such as health care [1], climate [2], robotics [3], and finance [4]. It is of high value to consumers and manufacturers to leverage the large quantity of data in real-time to make useful inferences from it. In the domain of machine learning and artificial intelligence, we want to use these time-series signals and either perform classification or regression tasks that help us identify system behaviour, makes them safe, secure and resilient. There is substantive evidence of machine learning research in the domain of time series classification [5]. However, either these techniques cannot handle big data or are not accurate for the task, and the models are computationally expensive that cannot execute on low power devices.

In this project, we propose a novel light weight deep learning architecture of dilated CNN (dCNN) for time series classification. Our work adapts the dCNN architecture from [6] where authors introduce the concept of dilations and its effects on the convolution operation. It highlights the benefits of dCNN for performing semantic segmentation. We evaluate the performance of our model on popular activity recognition datasets Mobifall [7] and WISDM [8]. Moreover, we also highlight the computational efficiency of dCNN when compared to the state of the art models consisting of LSTMs and GRUs. We believe that this lightweight architecture could pave the way for running highly accurate models on embedded systems for various applications and would reduce the need for expensive systems.

The second major contribution of this project is the introduction of the novel dataset for Human Intent Recognition(HIR) or Activity

Prediction. HIR is an upcoming field of research which aims to recognize the intent of a user based on a sequence of actions. HIR extends Human Activity Recogntion (HAR) by predicting the final action of a user that might happen in the future, unlike HAR, which recognizes an activity after seeing the data. We introduce a novel pseudo-realistic dataset created from an existing HAR dataset [8] for HIR. The dataset has been carefully developed such that it reflects practical situations. Each sample in the dataset combines activities from the HAR dataset in different combinations which would help predict the final intent of the user. We benchmark the performance of our dCNN model, LSTM and GRU on this new dataset. These models are evaluated based on computational complexity and predictive accuracy.

Chapter 2

Related Work

Analyzing data from sensors is same as performing time series analysis. It is a well-studied problem and still remains the central component in diverse areas of research. For instance, in [9] they combined the statistical machine learning approach of auto-regressive and moving average with artificial intelligence (AI) techniques to provide accurate time-series predictions. In AI, recurrent neural networks are traditionally considered a de-facto for modelling time series data. In [5] used a recurrent neural network for time series forecasting whereas as [10] use the modern variant of recurrent networks, the long-short term memory networks for forecasting. Time series forecasting has been an active area of research for deep learning researchers as well as where [11] and [12] employed deep recurrent and deep convolutional neural networks for time series classification. There is also extensive research in computational intelligence techniques such as fuzzy logic [13] and Bayesian models [14] for time series forecasting.

Deep neural networks have found profound success in the domain of computer vision and natural language processing with advances in convolutional and recurrent neural networks. However, recently, researchers are investigating the use of various variants and state-of-the-art deep learning models in the domain of time series analysis. In [15] they use the WaveNet model for multivariate time series prediction where conditional dependence of each variate on the other is given special consideration. In another work [16] designed a multi-channel CNN where each channel handles a variable in a multivariate time series. Time series classification plays a key role in numerous industrial applications such as in [17] use a CNN for anomaly detection and diagnosis in semiconductor manufacturing. In [18] presented a novel architecture of time series classification with each time series transformed to a 3-D tensor and subsequently processed through

the network. A comprehensive survey on deep learning techniques for time series forecasting is presented in [19], and a robust benchmarking of algorithms is presented in [20]. Time series from the domain of HAR has been studied in [21] where CNN is applied on the frequency features of a time series.

Past research has explored the domain of HIR and is sometimes referred to as plan recognition and goal recognition. In [22] uses goal recognition to recognize a player's high-level intentions using a computational model trained on a player behaviour corpus. The project highlights the advantages of deep LSTM based goal recognition models over single-layer LSTMs, n-gram encoded feedforward neural networks and Markov logic networks. In [23] authors have applied plan recognition on the popular game Starcraft navigation and Monroe Plan Corpus. The project presents a recursive neural network model that learns such a decision model automatically. In [24] recursive Bayesian approach has been used for HIR in the context of shared control robotics. The recursive Bayesian filtering approach is fused with multiple non-verbal observations to probabilistically reason about the intended goal of the user. The project concludes that its approach outperforms existing solutions and demonstrates that the probabilistic fusion of multiple observations improves intent inference and performance for shared control operation. In [25] developed HIR model in the domain of Natural Language Processing (NLP) to retrieve relevant information from a sentence for a voice-based human-machine interface for modern intelligent vehicles. In another work [26] propose a new model to predict intention of neighbouring vehicles from raw sensor data for self-driving cars.

Chapter 3

Data Set and Experiment

To evaluate the performance of our proposed model, we apply the model to two publicly available datasets on human activity recognition containing accelerometer data. We subsequently propose a pseudo-synthetic dataset for human intent recognition adapted from the Wisdm HAR dataset. The results are compared against the published results on these datasets.

3.1 Human Activity Recognition datasets

3.1.1 Mobifall Dataset

The Mobifall dataset [7] contains data for three sensors. A accelerometer, a gyroscope and an orientation sensor (a software-based sensor and derives its data from the accelerometer and the geomagnetic field sensor) of a smartphone. A Samsung Galaxy S3 device with the LSM330DLC inertial module (3D accelerometer and gyroscope) is used to capture the motion data. The accelerometer sensor gives acceleration force along the `x`, `y` and `z` axes (including gravity). The gyroscope sensor gives the rate of rotation around the `x`, `y` and `z` axes (angular velocity) and finally, the orientation sensor gives angle around the x, y and z axes.

The dataset consists of activities related to daily living and falls. The daily activities include standing, walking, jogging, jumping, moving up the stairs, moving down the stairs, sitting on the chair, stepping in the car and stepping out of the car. The Fall activities include

forward-lying, front-knees-lying, back-sitting-chair, sideward-lying.

3.1.2 Wisdm Dataset

The Wisdm dataset [27] contains data for the sensor, accelerometer. The dataset was collected from users using an Android smartphone who carried the phone in their trousers and asked to walk, jog, ascend stairs, descend stairs, sit, and stand for specific periods of time. The accelerometer data was collected at a rate of 20 samples per second. The accelerometer sensor in this case also gives acceleration force along the `x`, `y` and `z` axes (including gravity). The acceleration recorded includes gravitational acceleration toward the centre of the Earth. The dataset includes the activities walking, jogging, upstairs, downstairs, sitting, standing.

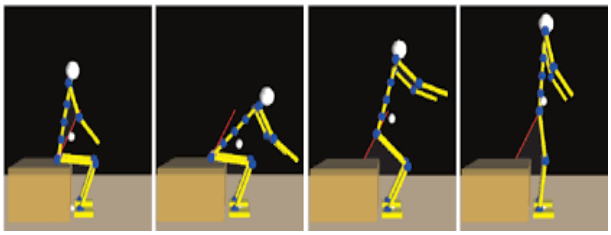

Figure 3.1: Positions under consideration for the Intent Sitting to Standing. The second and third positions correspond to the Transition positions for the Intent Sitting to Standing

3.2 Human Intent Recognition Dataset

The HIR dataset includes the intents from sitting to standing, standing to walking, and walking to jogging. The Wisdm HIR dataset is split into a training dataset containing data from 11 users and a test dataset containing data from 4 users.

3.2.1 Transition Positions

For every intent, there are three positions taken into consideration in this project, two primary positions and one transition position. These transition positions help us in identifying the final intent of the person. The movement from one primary position to another

$$A = \textit{Sensor values for Primary Position} : A \quad (1)$$
$$B = \textit{Sensor values for Primary Position} : B \quad (2)$$
$$T = \textit{Sensor values for Transition Position} \quad (3)$$
$$S = \textit{Sample of the HIR dataset} \quad (4)$$
$$T = A \cap B \quad (5)$$
$$K1 = \textit{Total Number of timestamps for which data for position A is recorded in a sample.} \quad (6)$$
$$K2 = \textit{Total Number of timestamps for which data for Transition position is recorded in a sample.} \quad (7)$$
$$K3 = \textit{Total Number of timestamps for which data for position B is recorded in a sample.} \quad (8)$$
$$K1 = [1,8],\ K2 = [1,5],\ K3 = [0,8] \quad (9)$$
$$S = \{A_{n=0}, A_{n=1}, A_{n=2}, \ldots\ A_{n=k1-1}, A_{n=k1}, T_{n=k1+1}, T_{n=k1+2}, T_{n=k1+3}, \ldots\ T_{n=k1+k2-1}, T_{n=k1+k2}, \quad (10)$$
$$B_{n=k1+k2+1}, B_{n=k1+k2+2}, B_{n=k1+k2+3}, \ldots\ B_{n=k1+k2++k3-1}, B_{n=k1+k2+k3}\}$$

primary position is a smooth continuous movement. The intermediate positions between the two primary positions are collectively grouped as transition positions for that intent. The overlapping sensor values of the two primary positions are considered to be the values for transition positions. Consider the positions in the Figure 3.1 which illustrate the different positions taken into consideration for the intent *Sitting to Standing*. The intermediate positions between the two primary positions are collectively defined as transition positions for that intent.

3.2.2 Procedure

Equations 1 - 10 give an insight about the process of creating the HIR dataset.

1. Equation 1, 2 and 3 represent the sensor values for different positions. Equation 5 illustrates the method of obtaining the sensor values for the transition position. The overlapping values of the two primary positions in each dimension `x`, `y`, and `z` are considered as the sensor values for the transition position.

2. Equation 6, 7, 8 describe the features of the HIR dataset sample. Consider a sample with $\{K1 = 3,\ K2 = 5\ ,\ K3 = 4\}$. The first three timestamps of the sample contains the data for position A , the next five timestamps for position T and the last four timestamps for position B. The main feature of the HIR is its ability to predict the intent irrespective of the number of events happening before and after the transition and different values of K1, K2 and K3 help reflect practical situations in the HIR dataset samples.

3. A special case arises when *K3 = 0*, which highlights the main difference between HAR and HIR. Thus when K3 = 0, the sample does not contain the sensor data for the destination primary position. This highlights that HIR is able to predict the final position of the user on the basis of its starting and transition position, which is however not the case in HAR which is only able to recognise the position after it receives the sensor data for that position.

Chapter 4

Dilations in CNN

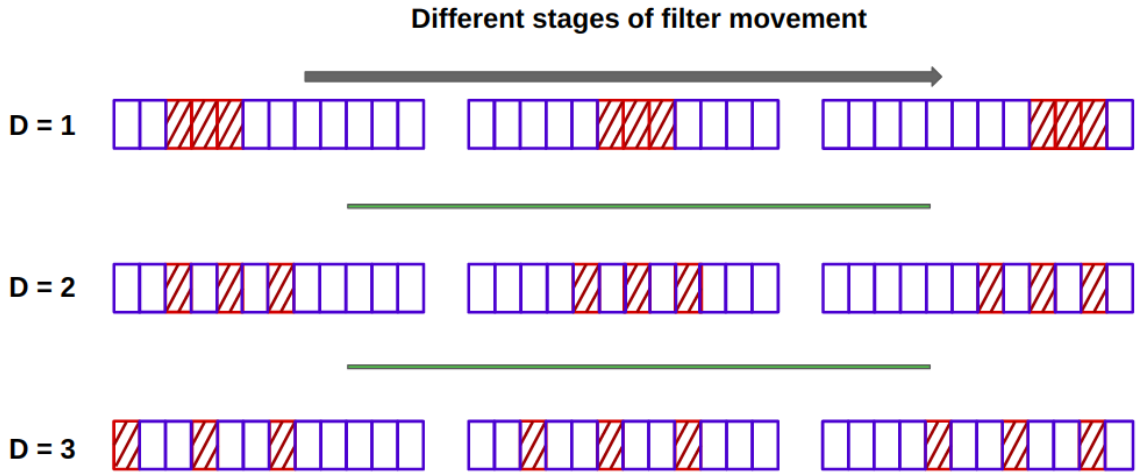

Figure 4.1: Different stages of filter movement across the input for different dilation rates in 1d convolutions. The filter in this image is of size three.

Till recently, Recurrent neural networks (RNNs) have been considered the most optimum neural network architectures for time series analysis however recent studies show that convolutional neural networks can perform comparably or even better. Dilated convolutions help one-dimensional convolutional neural networks to effectively learn time series dependencies. CNNs have many advantages over RNN based solutions. They are much more computationally efficient than RNNs and are much more easier to train. Further hidden layers in CNNs can be visualised which is not the case in RNNs and this helps in the explaining the predictions of the model. Figure 4.1 illustrates the effect of change in dilation rate on filter movement across the input. The filter used for performing convolutions is of size three.

Chapter 5

dCNN Module

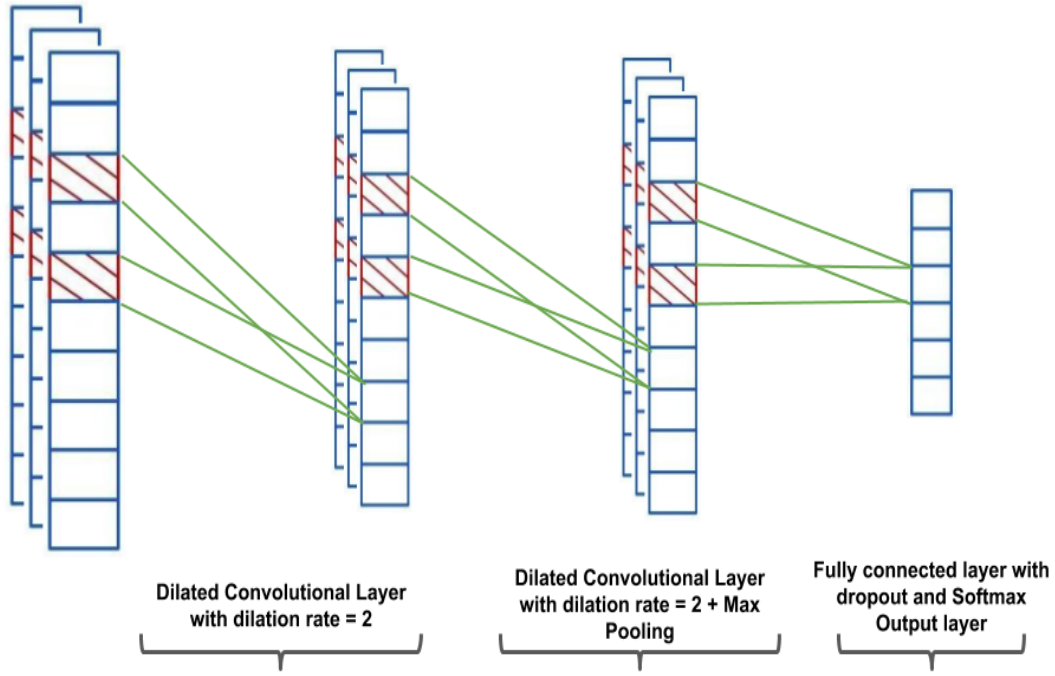

Figure 5.1: Block Diagram of the proposed Neural Network Architecture. The red shaded boxes illustrate filter operation on the layer input.

Our proposed approach uses the dCNN Module. The dilated convolutional neural network (dCNN) module is composed of a sequence of layers of dilated convolutional layers with ReLU (Rectified Linear Unit) activations. We apply a 1-dimensional dilated convolution to extract features from the segments of the time series. In the dCNN architecture presented in Figure 5.1, the first layer in the dCNN module is the dilated convolutional layer; filters in this layer are of $size = 2$

and slide over the sensor time series with $stride = 1$ and $dilation\ rate = 2$ in the vertical direction. Moreover, employing dilations in the convolution layer allows one to learn features between the observations that are far off.

The next layer in the dCNN module similar to the first layer that performs dilated convolutions on the feature maps obtained from the previous layers. It allows to extract features between much far off sensor values. Subsequently, we have a max pooling layers that is used to abstract the high-level features that are learnt by our dCNN filters. Finally, we have the last fully-connected layer that is applied to the outputs of the dCNN layers. The softmax layer computes the probability values for the predicted class.

5.1 dCNN Architecture Analysis

The network designed for the task of human activity recognition and prediction has four layers including: Dilated Convolution Layer 1 (No. of filters = 8, Filter Size = 2, Dilation Rate = 2) $\Longrightarrow$ Dilated Convolution Layer 2 (No. of filters = 8, Filter Size = 2, Dilation Rate = 4) $\Longrightarrow$ Max Pooling Layer and followed by a fully connected layer. To train the network, we used batch size of 2000 on Google Cloud Instance with 1 NVIDIA Tesla T4 GPU. We used the $cross-entropy$ loss function with the ADAM optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and the learning rate was set to 10^{-5}. We used drop out at a rate of 0.5.

5.2 LSTM and GRU Architecture Analysis

We develop two architectures of single layer network of LSTMs and GRUs to compare the effectiveness of our dCNN model. We use a batch size of 2000 to train each network and use ADAM optimizer with hyperparameter $\beta_1 = 0.9$ and $\beta_2 = 0.999$. A softmax layer is the terminating layer in each of the architecture for classification.

Chapter 6

Results

In this Section we present our results on the three datasets. Two for activity recognition and one for pseudo-synthetic dataset for activity prediction or intent recogntion. We compare state of the art methods with our approach and demonstrate superiority in performance along with the model being light weight enabling deployment on real-world low power mobile devices.

We present our results on activity recognition. In Table 6.1 we present our results for the Wisdm dataset on the recurrent neural networks based on LSTM, GRU, and our dCNN model. We compare the performance of our model on accuracy and size of the model (the number of parameters in the model).

Model Name	**Accuracy**	**# of Parameters**
LSTM	96.73%	94,214
GRU	96.12%	89,862
dCNN	**94.95%**	**9,606**

Table 6.1: Performance evaluation of the dCNN model on the WISDM data set for activity recogniton

Similar to the results on the Wisdm dataset, we performed similar experiments on another publicly available dataset, the Mobifall, to evaluate the generalizability of our model. It turns out that results are promising and reveal interesting insights about the process. In Table 6.2 we present our results for the Mobifall dataset and compare the results obtained from state-of-the-art models.

In both Table 6.1 and Table 6.2 the results follow a similar trend where the dCNN model has better or nearly as good as state-of-the-art performance however, the model has significantly low number of

Model Name	**Accuracy**	**# of Parameters**
LSTM	81.05 %	132,617
GRU	80.98%	128,265
dCNN	**78.16%**	**7,257**

Table 6.2: Performance evaluation of the dCNN model on the Mobifall data set for activity recognition

parameters. The low number of parameters make the model modular and portable in terms of devices with low memory and power.

We briefly introduced the notion of intent recognition as a problem similar to predicting the activity instead of classifying it. We created a pseudo-synthetic data set for researchers interested in the problem of intent recognition.In Table 6.3, we present results and comparison of the dCNN model for HIR. Our results are state-of-the-art at the moment. Our dataset is opensource and we welcome researchers to test their models and present results in the domain of human intent recognition.

Model Name	**Accuracy**	**# of Parameters**
LSTM	98.65 %	22,788
GRU	98.74%	18,436
dCNN	**96.259%**	**452**

Table 6.3: Performance evaluation of the dCNN model on the intent recognition dataset

It may be noted from the results in Table 6.3 that although the model's performance is approximately 2% lower however there is a 40-fold reduction in the model size. It is a very useful result and paves way forward for real-time intent recognition algorithms that can be use in health care and automotive industry.

Chapter 7

Summary

We have proposed a new convolutional neural network architecture for sensor data classification and prediction. We present a novel dilated CNN model for human activity recognition and human intent recognition. The dCNN architecture uses stacks of dilated convolutions. The extracted features for classification task are obtained by considering the inter and intra relation between variates. Our experiments show that the proposed model is as effective as previous models working with hand-crafted features such as spectrogram and statistical features. In this project we also introduce a novel dataset of Human Intent Recognition using Human Activity Recognition. We wish to pursue further work in HIR which has the ability to expand the scope of HAR to many other diverse applications and domains.

Bibliography

[1] Spencer S Jones, R Scott Evans, Todd L Allen, Alun Thomas, Peter J Haug, Shari J Welch, and Gregory L Snow. A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of biomedical informatics*, 42(1):123–139, 2009.

[2] Zhihua Zhang, Zhang, and Khelifi. *Multivariate Time Series Analysis in Climate and Environmental Research*. Springer, 2018.

[3] Claudia Pérez-D'Arpino and Julie A Shah. Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 6175–6182. IEEE, 2015.

[4] Nijolė Maknickienė, Aleksandras Vytautas Rutkauskas, and Algirdas Maknickas. Investigation of financial market prediction by recurrent neural network. *Innovative Technologies for Science, Business and Education*, 2(11):3–8, 2011.

[5] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):6085, 2018.

[6] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *CoRR*, abs/1511.07122, 2015.

[7] G. Vavoulas, M. Pediaditis, E. G. Spanakis, and M. Tsiknakis. The mobifall dataset: An initial evaluation of fall detection algorithms using smartphones. In *13th IEEE International Conference on BioInformatics and BioEngineering*, 2013.

[8] Gary M. Weiss Jennifer R. Kwapisz and Samuel A. Moore. Activity recognition using cell phone accelerometers. In *Proceedings*

of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10), Washington DC, 2010.

[9] G Peter Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.

[10] Kyongmin Yeo. Model-free prediction of noisy chaotic time series by deep learning. *arXiv preprint arXiv:1710.01693*, 2017.

[11] Bernardo Pérez Orozco, Gabriele Abbati, and Stephen Roberts. Mordred: Memory-based ordinal regression deep neural networks for time series forecasting. *arXiv preprint arXiv:1803.09704*, 2018.

[12] Omolbanin Yazdanbakhsh and Scott Dick. Multivariate time series classification using dilated convolutional neural network. *arXiv preprint arXiv:1905.01697*, 2019.

[13] Omolbanin Yazdanbakhsh and Scott Dick. Forecasting of multivariate time series via complex fuzzy logic. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(8):2160–2171, 2017.

[14] Feras A. Saad and Vikash K. Mansinghka. Temporally-reweighted Chinese restaurant process mixtures for clustering, imputing, and forecasting multivariate time series. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS 2018)*, volume 84 of *Proceedings of Machine Learning Research*, pages 755–764, Playa Blanca, Lanzarote, Canary Islands, 2018. PMLR.

[15] Anastasia Borovykh, Sander Bohte, and Cornelis W Oosterlee. Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691*, 2017.

[16] Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J Leon Zhao. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science*, 10(1):96–112, 2016.

[17] Ki Bum Lee, Sejune Cheon, and Chang Ouk Kim. A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, 30(2):135–142, 2017.

[18] Chien-Liang Liu, Wen-Hoar Hsaio, and Yao-Chung Tu. Time series classification with multivariate convolutional neural network. *IEEE Transactions on Industrial Electronics*, 66(6):4788–4797, 2018.

[19] John Cristian Borges Gamboa. Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887*, 2017.

[20] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.

[21] Daniele Ravi, Charence Wong, Benny Lo, and Guang-Zhong Yang. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE journal of biomedical and health informatics*, 21(1):56–64, 2016.

[22] Wookhee Min, Bradford W. Mott, Jonathan P. Rowe, and James C. Lester. Deep lstm-based goal recognition models for open-world digital games. In *AAAI Workshops*, 2017.

[23] Francis Bisson, Hugo Larochelle, and Froduald Kabanza. Using a recursive neural network to learn an agent's decision model for plan recognition. In *IJCAI*, 2015.

[24] S. Jain and B. Argall. Recursive bayesian human intent recognition in shared-control robotics. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3905–3912, Oct 2018.

[25] Yudi Zheng, Yongkang Liu, and John H. L. Hansen. Intent detection and semantic parsing for navigation dialogue language processing. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6, 2017.

[26] Sergio Casas, Wenjie Luo, and Raquel Urtasun. Intentnet: Learning to predict intention from raw sensor data. In Aude Billard, Anca Dragan, Jan Peters, and Jun Morimoto, editors, *Proceedings of The 2nd Conference on Robot Learning*, volume 87 of *Proceedings of Machine Learning Research*, pages 947–956. PMLR, 29–31 Oct 2018.

[27] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.