Temperature Forecasting via NeuralNetworks Based on Time-Series Data

Rawnaq Naili[1], *Rania Chettab*[1], *Mohamed AbdouBouteldja*[2], Mohamed Badeche[2*].

[1]Software and Information Systems Technologies (TLSI) Department, University of Constantine 2 - AbdelhamidMehri, Algeria.

[2]LIRE Laboratory, University of Constantine 2 - AbdelhamidMehri, Algeria.

[*] Corresponding address: mohamed.badeche@univ-constantine2.dz

Abstract—in this paper, we presentdeep neural network architectures and use themintime series weather forecasting. We started with a MLP (Multi-Layer Perceptron) architecture, then we used CNN (Convolutional Neural Network) architecture, and moving on to LSTM (Long-Short Term Memory), to finally making a hybrid architecture of CNN and LSTM. Eight years (2009-2016) of hourly meteorological data were used to train each model and the results show that neural network could be a good way to predict future temperatures.

Keywords—Artificial neural networks, Temperature forecasting, Time series, MLP, CNN, LSTM, Time series forecasting.

I. INTRODUCTION:

Weather forecasting is the prediction of the state of the atmosphere for a given location using the application of science and technology. It is nowadays an incredibly complex technical feat, and only the most powerful computers can handle the ever-increasing amount of information collected. [6] And when talking about weather forecasting, temperature prediction is one of the major concerns in the domain of meteorology. [5]

The current wideavailability of massive weather observation datamotivated researchers to explore pattern hidden in thelarge dataset and make weather prediction. The adventof information and computer technology in the lastdecade also support the research of weather forecastingto get more accurate results.[13]. Moreover, the best and most used technique that could handle the huge amount of collected data is time series forecasting.

Time series forecasting is the use of certain model to forecast future values based onpast observed values, and thus can be understood as a method for predicting future values by understanding past values [9].

Numerical weather forecasts use atmospheric models to predict future weather conditions based on current weatherconditions [10, 11, 12]. Unlike numerical weather prediction, time series forecasting uses a model to predictfuture values based on past values.

The prediction of time series data in meteorology can help in decision-making processes carried out by organizations responsible for the prevention of disasters. [4]

II. DATA COLLECTION:

1. Raw data :

The dataset contains weather conditions recorded at the Weather Station at the Max Planck Institute for Biogeochemistry in Jena, Germany, over several years. In this dataset, 14 different quantities (including air temperature, atmospheric pressure and humidity) were recorded every 10 minutes (see Table 1). [7]

FEAUTURE	UNIT
PRESSURE	Mbar
TEMPERATURE	C°
T POT	K
T DEW	C°
RH	%
VP MAX	Mbar
VP ACT	Mbar
VP DEF	Mbar
SH	g/kg
H2OC	Mmol/mol
RHO	g/m**3
WV	m/s
MAX WV	m/s
WD	Deg

Table 1: Features recorded in the dataset

The original data goes back to 2003, but in our case we used a subset, from 2009 to 2016 (both inclusive). In addition, we used a further reduced version in which recordings are kept for every hour.

The graph below represents the variation of the temperature in the chosen weather data:

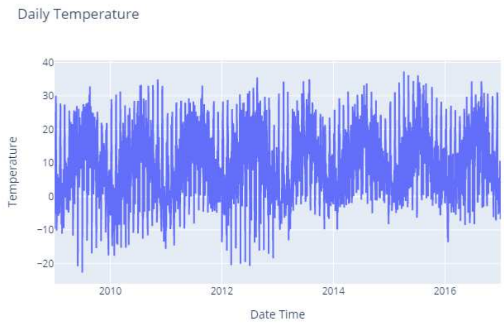

Figure 1: Temperature variation (2009 - 2016)

We divided the chosen weather data into two successively selected groups, the training group, corresponding to the first 75% of the dataset, and the test group, corresponding to 25% of the rest of the dataset so that the generalization capacity of network could be checked after training phase.

2. Data pre-processing:

As a rule of thumb, whenever we use a neural network, we should normalize or scale our data. For it helps in fastertraining of models.

The goal of normalization is to change the values ofnumeric columns in the dataset to a common scale,without affecting differences in the ranges of values orlosing any information.

The preprocessing that we carried out here isnormalizing data using MinMaxScaler. Which does exactly what we mentioned before.

For each value in a feature, MinMaxScalersubtracts the minimum value in the feature and then divides by the range. The range is the difference between the original maximum and original minimum.

The default range for the feature returned by MinMaxScaler is 0 to 1.

MinMaxScaler preserves the shape of the original distribution and does not change the information embedded in the original data.

For measuring the error made by the neural network, we used the Mean Absolute Error (MAE) :

$$mae = \frac{1}{n}\sum_{j=1}^{n} | y_j - \hat{y}_i |$$

Where:

- y_j is the true value.
- $\hat{y}$ is the predicted value.
- n is the total number of test sample.

III. RECURRENT NEURAL NETWORKS:

Neural networks have seen big interest lately, and they are being successfully appliedin various domains, especially when it comes to solve prediction, classification, or control problems.

They are defined as an interconnected of simpleprocessing element whose functionality is based on thebiological neuron which is a unique piece ofequipment that carries information or a bit of knowledge andtransfers to other neuron in the chain of networks. Artificialneuron imitates these functions and their unique process oflearning [3].

In our case, we used three different architectures of artificial neural networks.

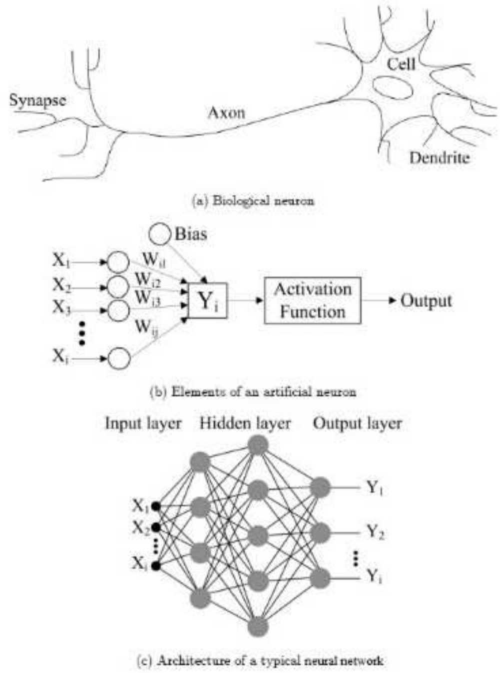

Figure 2: Neural network architecture [8]

1. MULTI-LAYER PERCEPTRON (MLP):

The multi-layer perceptron is the first kind ever of artificial neural networks, it has many layers, the first layeris the input layer, the last one is the output layer, themiddle layers are called hidden layers, and each layer includes several neurons. The neurons in MLP have thecharacteristics of inter-layer full connection and intra-layerconnectionless. (Figure 2)

Following this architecture, we developed a sequential model of one input layer, successively followed by a dense hidden layers with a ReLu activation function (used to decide whether a neuron can be activated or not), and one last dense layer of one neuron that represents the output of this model.

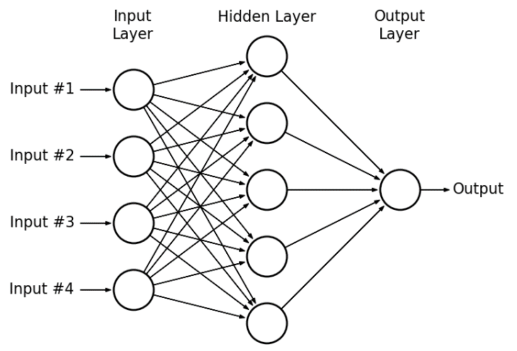

Figure 3: Multi-Layer Perceptron Network

2. CONVOLUTIONAL NEURAL NETWORK (CNN):

Convolutional neural networks are generally associated with computer vision applications. Their architecture is specifically adapted to carrying out complex visual analyzes. It is defined by a three-dimensional arrangement of neurons, instead of the standard two-dimensional network. The first layer of these neural networks is called "Convolutional Layer". Each neuron in this layer only processes information from a small part of the visual field. The convolutional layers are followed by rectified or ReLU layer units, which allows CNN to process complicated information.

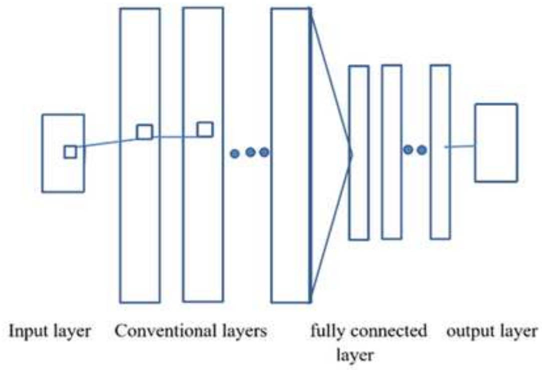

Figure 4: CNN architecture

Although traditionally developed for two-dimensional image data, CNNs can be used to model univariate time series forecasting problems, in our case it is temperature prediction. For that, we developed sequential model composed of one input layer, one convolutional hidden layer with a ReLu activation function. This layer is used to extract the relationship between the values of the 14 different features fed as input and to see if they have a useful impact to precise predictions.And then a pooling layer that is used to distill the output of the convolutional layer to the most salient elements.

The convolutional and pooling layers are followed by a fully connecteddense layer that interprets the features extracted by the convolutional part of the model.

A flatten layer is used between the convolutional layers and the dense layer to reduce the feature maps to a single one-dimensional vector.

3. LONG-SHORT TERM MEMORY (LSTM):

The developed RNN model is based on one of the neural network architecture named LSTM (Long short-Term Memory).

LSTM is a network that is composed of cells (LSTM blocks) linked to each other. Each LSTM block contains three types of gate: Input gate, Output gate, and Forget gate, respectively, which implement the functions of writing, reading, and resetting on the cell memory. [1]

The presence of these gates, allows LSTM cells to remember information for an indefinite time; in fact, if the following Input gate is the activation threshold, the cell will retain the previous state, and if the current state is enabled, it will be combined with the input value. As the name suggests, the Forget gate resets the current state of the cell, and the Output gate decides whether the value in the cell must be carried out or not. [1]

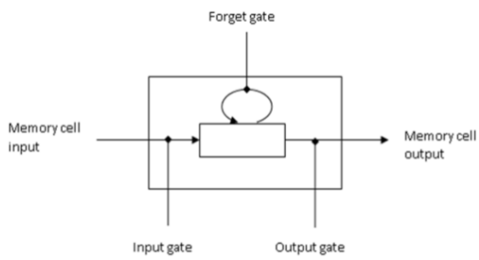

Figure 5: Block diagram of an LSTM cell [1]

The reason behind using it is that it is free from the problem of vanishing gradients, and offers excellent results and performance;also, it is ideal for prediction and classification of temporal sequences.[1]

Following this background we developed a sequential modelcomposed of multiple layers: one input layer and one single LSTM layer with ReLu activation function, and to make it more robust; it was followed by a dense layer with only one neuron that predicts the future temperature value.

4. CNN-LSTM:

In this section, we developed a hybrid architecture of the two previous models, started with the convolutional and the pooling layers of the developed CNN; and ended by the LSTM layer with a ReLu activation function of the

LSTM model, which replaced the fully connected dense layer in this model.

In addition, the flatten layer is used between the convolutional layers and the LSTM layer as before.

In each of the four architectures, we trained the model using all the 14 features, in order to forecast only the temperature.

When compiling the model, and to optimize the algorithm, we used the Adam optimizer, which realizes the benefits of both AdaGrad and RMSProp. Adaptive Gradient Algorithm maintains a per-parameter learning rate that improves performance on problems with sparse gradients. Root Mean Square Propagation that also maintains per-parameter learning rates that are adapted based on the average of recent magnitudes of the gradients for the weight.[2]

IV. TOOLS USED:

- *Pandas* - Open source data structuring and analysis tool.
- *Keras* - An open source neural network library in Python.
- *NumPy* - Used for scientific computing in Python.
- *Matplotlib* - Python library for visualization of data.
- *TensorFlow* - Python library used for dataflow programming.

V. RESULT AND DISCUSSION:

The optimal structures for the developed model in every part for obtaining minimum prediction error are shownbelow:

MLP	Number of 1st denselayer's neurons	100
	Batch size	32
	Number of epochs	10
	MAE	**0.64**
CNN	Number of dense layer's neurons	50
	Batch size	32
	Number of epochs	10
	MAE	**1.03**
LSTM	Number of dense layer's neurons	50
	Batch size	32
	Number of epochs	10
	MAE	**0.60**
CNN-LSTM	Number of dense layer's neurons	50
	Batch size	32
	Number of epochs	10
	MAE	**0.61**

Table 2 :Optimal structures' configurations

The graph below represents the results of the MLP's optimal structure's configuration:

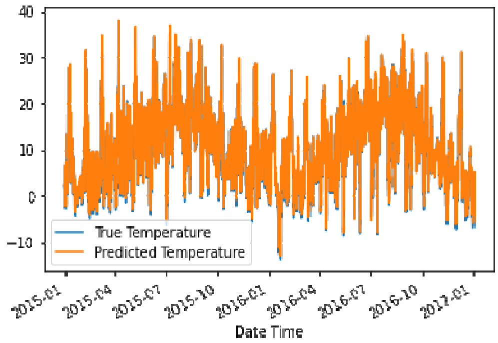

Figure 6: MLP predictions

For the CNN, these are the results:

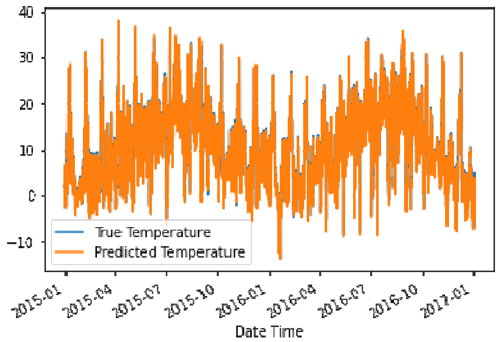

Figure 4: CNN predictions

And these are the results of the LSTM model:

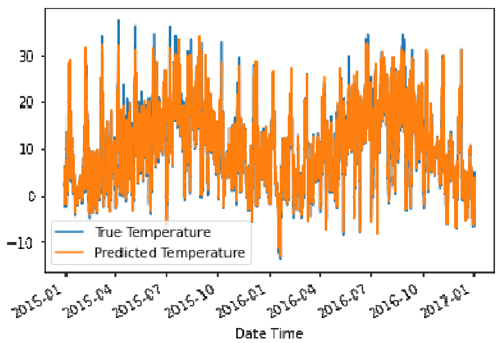

Figure 7: LSTM predictions

Lastly,the results of the CNN-LSTM model are:

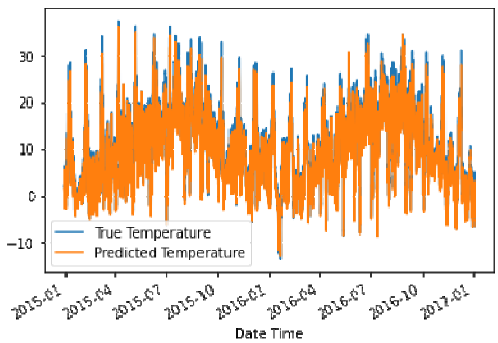

Figure 8:LSTM model predictions

VI. CONCLUSION:

The results of this paper show that the MAE values obtained were as small as possible, which suggest that neural networks generally and LSTM specifically could be a good way to relate on to predict temperature of general weather conditions. In addition, numerical weather forecastscan be used with these models to make such reasonable and accurate predictions and help the weather decision making systems in case of alerts or warning.

REFERENCES:

[1] Deep Learning with TensorFlowby Giancarlo Zaccone, Md. Rezaul Karim, Ahmed Menshawy, April 2017.

[2] Neural networks and deep learningby Aurélien Géron, March 2018.

[3] L. Fausett, Fundamental of Neural Networks, New York , Prentice Hall.A well-written book, with very detailed worked examples to explainhow the algorithm function, 1994.

[4] Modise Wiston and Mphale KM (2018) Weather Forecasting: From the Early Weather Wizards to Modern-day Weather Predictions. JClimatol Weather Forecasting 6: 229. doi:10.4172/2332-2594.1000229.

[5] METEOROLOGICAL APPLICATIONS.Meteorol. Appl. 15: 431–445 (2008).Published online 9 July 2008 in Wiley InterScience(www.interscience.wiley.com) DOI: 10.1002/met.83

[6] Weather forecasting article by Encyclopaedia Britannica, Inc17 January 2019. Encyclopaedia Britannica. https://www.britannica.com/science/weather-forecasting.

[7] Weather time series dataset recorded by the Max Planck Institute for Biogeochemistry

[8]GORDAN, Meisam; RAZAK, Hashim Abdul; ISMAIL, Zubaidah and GHAEDI, Khaled. Recent Developments in Damage Identification of Structures Using Data Mining. Lat. Am. j. solids struct. [online]. 2017, vol.14, n.13 [cited 2020-05-30], pp.2373-2401. Available from: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1679-78252017001302373&lng=en&nrm=iso>. ISSN 1679-7825. https://doi.org/10.1590/1679-78254378.

[9]Raicharoen, T., Lursinsap, C., &Sanguanbhokai, P. (2003). Application of critical support vectormachine to time series prediction. International Symposium on Circuits and Systems (Vol.5,pp.V-741-V-744 vol.5). IEEE

[10] Barker, D., Huang, X. Y., Liu, Z. Q., Auligné, T., Zhang, X., &Rugg, S., et al. (2012). Theweather research and forecasting model's community variational/ensemble data assimilationsystem: wrfda. Bulletin of the American Meteorological Society, 93 (6), 831 - 843.

[11] Shen, F., Min, J., &Xu, D. (2016). Assimilation of radar radial velocity data with the wrf hybridetkf–3dvar system for the prediction of hurricane ike (2008). Atmospheric Research, 169, 127-138.P.G. Clem, M. Rodriguez, J.A. Voigt and C.S. Ashley, U.S. Patent 6,231,666. (2001).

[12] Xu, D., Min, J., Shen, F., Ban, J., & Chen, P. (2016). Assimilation of mwhs radiance data fromthe fy□3b satellite with the wrf hybrid□3dvar system for the forecasting of binary typhoons.Journal of Advances in Modeling Earth Systems, 8 (2).

[13] AfanGalih Salman *et al.* / Journal of Computer Science 2018, 14 (7): 930.938 DOI: 10.3844/jcssp.2018.930.938