

Meteonowcasting using Deep Learning Architecture

Sanam Narejo and Eros Pasero

Department of Electronics and Telecommunication
Politecnico Di Torino
Torino, Italy

Abstract—The area of deep learning has enjoyed a resurgence on its peak, in almost every field of interest. Weather forecasting is a complicated and one of the most challenging tasks that includes observing and processing huge amount of data. The present paper proposes an effort to apply deep learning approach for the prediction of weather parameters such as temperature, pressure and humidity of a particular site. The implemented predictive models are based on Deep Belief Network (DBN) and Restricted Boltzmann Machine (RBM). Initially, each model is trained layer by layer in an unsupervised manner to learn the non-linear hierarchical features from the input distribution of dataset. Subsequently, each model is re-trained globally in supervised manner with an output layer to predict the appropriate output. The obtained results are encouraging. It is found that the feature based forecasting model can make predictions with high degree of accuracy. This implies that the model can be suitably adapted for making longer forecasts over larger geographical areas.

Keywords—Deep learning architectures; deep belief network; time series prediction; weather nowcasting

I. INTRODUCTION

In the few last decades, the use of machine learning has spread rapidly beyond the limitations of computer science field. Machine learning is extensive and so pervasive today that one probably uses it dozens of times a day without knowing it. Deep learning is a category of machine learning models. Recently, the area of Deep learning has enjoyed a resurgence on its peak, in almost every field of interest. Deep learning architectures include several models such as Deep Neural Networks (DNNs), Convolutional Neural networks (CNNs), Recurrent Neural Network (RNN), Deep Belief Network (DBN), Recursive NN and more [1]. Present literature suggests that these architectures are being applied widely and it has produced state of the art results on various problems in major fields like computer vision, automatic speech recognition, natural language processing, audio recognition and bioinformatics.

Weather forecasting is a complex time series forecasting problem. This is due to its dynamic and non-linear chaotic behaviour [2], [3]. It is an arduous skill that involves observing and processing vast amounts of data. In literature, several approaches have been proposed in order to deal with accurate time series forecasting problem. Artificial Neural Network (ANN) is known to be one of the successfully developed models widely used in solving many time series forecasting and prediction problem in diversity of applications [4]-[6]. ANN is general, flexible, non-linear tool capable of approximating any arbitrary function [7]. ANN is based on a

collection of connected units called artificial neurons. Neurons receive input, change their internal state (activation) according to that input, and produce output depending on the input and activation. The network forms by connecting the output of certain neurons to the input of other neurons forming a directed, weighted graph. The weights as well as the functions that compute the activation can be modified by a process called learning.

Deep learning is an application of ANN to learning tasks that contain more than one hidden layer. Deep ANNs contain numerous levels of non-linearities depending on the depth of hidden layers. The deep hierarchical architecture allows them to efficiently represent highly nonlinear patterns and highly varying functional abstractions. Although, it was not clear how to train such deep networks, as the random initialization of network parameters appears to often get stuck in poor solutions [8].

Nowcasting is defined as the prediction of the present, the very near future. The term is a contraction for now and forecasting. This term has been used for a long time in meteorology. In other words, nowcasting is a strategy to perform very short range forecasting. This procedure maps the current weather and then uses an estimate of its speed and direction of movement to forecast the weather a short period ahead. A critical aspect regarding the time series prediction problem is to capture the temporal relationship [9] and underlying structure residing in given input series data [10].

This research work is inspired by the recent advances in the realm of deep learning methods. In this work, a predictive ANN model based on the deep learning is obtained by firstly training the layers of Restricted Boltzmann Machine (RBM) in an unsupervised fashion. Subsequently, stacking those trained RBMs to create Deep Belief Network (DBN). Afterwards, the DBM is finally trained in supervised manner to predict the parameters of weather, i.e., temperature, pressure and humidity. For each parameter, a separate predictive model is implemented and trained. The accuracy of predictions confirms the promising performance of Deep learning algorithm specifically DBN. The data used in this study is sampled every 15 minutes by means of a traditional metrological station. The approach proposed here is basically a local level and it is restricted to a particular geographical area. However, it can be further extended and possible to apply on global level.

A literature review is presented in Section II. Section III deals with the illustration of adopted research methodology and experimental setup. Section IV presents the obtained results and discussion in detail. The paper ends with conclusions and

suggestions for possible future research as specified in Section V.

II. LITERATURE REVIEW

Keeping our focus particularly in the realm of time series forecasting and prediction, we shed some light by presenting some research related with time series forecasting with deep learning methods. As aforementioned in previous section, various different architectures come under the umbrella of “Deep Learning models”.

The initial research advocates that DBN models are efficient at the classification and prediction tasks but, it actually lacks the efficacy to model temporal sequences. The authors in [11] have reported that recurrent neural networks performed drastically better on energy load forecasting using dataset from kaggle competition. They further argued that the greedy layerwise trained feed forward neural networks with stacked AutoEncoders obtained discouraging results with no significant performance gain but added complexity. As it is well known fact that feed forward networks are deficient in capturing temporal dynamics. Thus, these models are unable to access the past terms while modelling the underlying structure.

On the contrary, the Stacked AutoEncoders are also deployed in [12] in order to learn feature representations for weather forecasting. The empirical evaluations are further compared between models using raw features and models using learned representations as features. The obtained results in the above mentioned study prove that Deep Neural Network (DNN) is capable enough to provide better feature space for highly varying and non-stationary data like weather data series of temperature, pressure and wind speed. Related study has been provided in [13] for short term wind prediction and in [14] for load forecasts.

A predictive model has been proposed for time series data in [15] by using a DBN with RBM. Additionally, the performance of the proposed model is evaluated on data of CATS benchmark [16], [17] and chaotic time series. According to the experimental outcomes, it was confirmed that the proposed prediction model, DBN with RBM using pre-training and fine-tuning learning algorithm and PSO structure decision, performed better than traditional models although it is unable to beat the best of IJCNN 2004 competition model. The authors in [10] proposed a novel hybrid approach for multi-step ahead time series forecasting by using deep learning and Nonlinear Autoregressive Neural Network.

The research in [18] presents another novel hybrid model with discriminative and generative components for spatio-temporal inference about weather. Furthermore, a data driven kernel is implemented that forms the predictions according to physical laws. A detailed review of unsupervised feature learning and deep learning for time series modelling has been conferred in [19]. The article presents the detailed review on time series analysis and temporal sequence modelling using deep architectures. However, according to our observation, the study was found to be deficit, as far as the overview regarding time series forecasting with deeper models is concerned. The CRBM was introduced in the family of Deep Learning by applying it to capture the activity related to human motion [20].

Similarly, the other variants of RBM, for example, Temporal RBM [21], [22] and Gated RBM [23], [24] have also been introduced. With the exception to these models, another deep architecture producing outstanding state-of-the-art research is Convolutional Neural network (CNN). These models are of high interest specifically for image data or high dimensional time series data. Apart from being stand alone, convolution has also been applied as Convolutional RBM [25], [26] and Convolutional AutoEncoders [27]-[29].

III. MATERIALS AND METHODS

A. Meteorological Nowcasting

The activity conducted in this current work is related with our earlier work done in [30]. In above mentioned research, a statistical neural system was used to “nowcast” meteorological data measured by a weather station deployed at Neuronica laboratory, Politecnico Di Torino, see Fig. 1. For further details please refer [30]-[32]. By utilizing the same resources of meteorological data, i.e., “NEMEFO”, we have performed weather parameter prediction by using deep learning algorithm.

In our previous work [33], we presented predictive models for internet traffic prediction by using DBN. We explored the useful strategies for topological architecture for deeper networks and we also did validation on standard benchmark time series. Keeping all those aspects in mind, which we earned for successful training of deep models, we were motivated to attempt some more case studies for real time data sets.



Fig. 1. METEO weather station.

Weather forecasting has been one of the most challenging problems around the world for more than a half century. However, nowcasting is weather forecasting on a very short

term. It makes difficult for traditional mathematical or statistical models to adapt irregular patterns of data which cannot be written in form of function, or deduced from a formula. In response to this, we developed and trained some more DBN models for nowcasting the air temperature, relative humidity, and air pressure for the next future value. The pictorial view of our contributed activity is presented in Fig. 2.

The standard training of deeper models through gradient back propagation appears to be difficult until Hinton gave a breakthrough in 2006. The standard training strategies attempts to allocate the parameters in the region of parameter space that generalize poorly. This has been shown practically in number of studies [34].

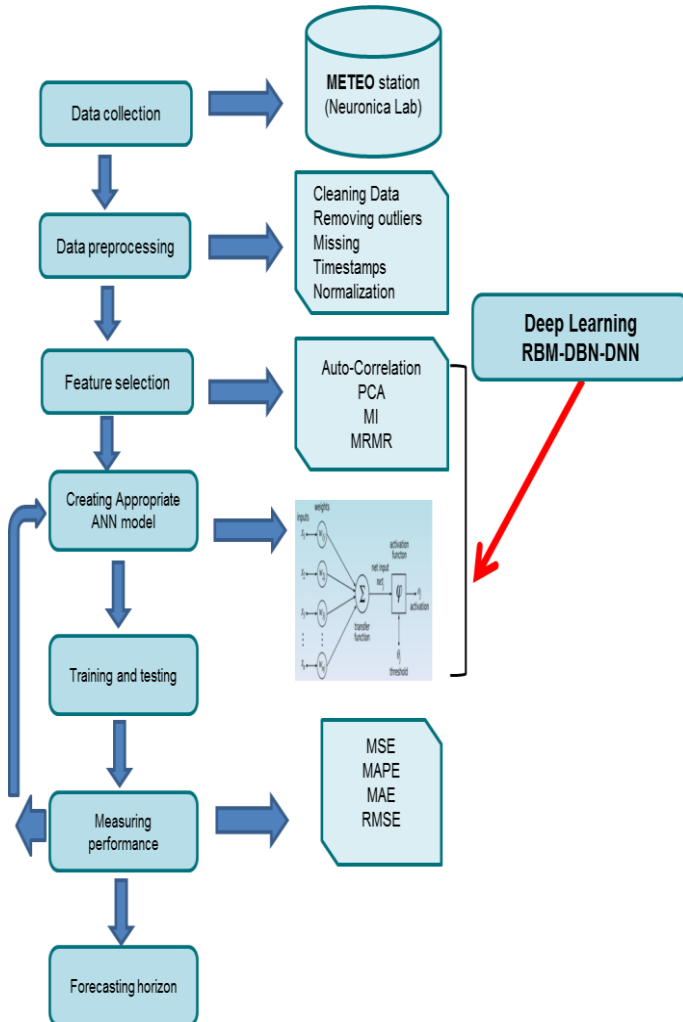


Fig. 2. Illustration of contributed activity for weather nowcasting.

B. METEO Weather Station and NEMEF0

The Weather forecasting is a complicated and one of the most challenging tasks that includes observing and processing huge amount of data. NEMEF0 stands for NEural METeological FOrecasts. It is basically a software tool connected to Meteo weather station at Neuronica laboratory, which samples meteo data after every 15 minutes. Meteo station contains following recorded weather data.

- Air temperature
- Relative Humidity
- Air Pressure
- Solar radiation
- Wind velocity
- Precipitation
- Wind Direction
- Corresponding Date and Time.

The sensors at Meteo station provide a new recording after every fifteen minutes. The dataset was downloaded from weather station. It contains the records from 4 October 2010 to 3 September 2015. However, the predictive models are only implemented for nowcasting of Air temperature, Relative humidity, and Air pressure as mentioned previously. The data recorded through sensors may have noise, some of missing samples and unwanted frequency fluctuations. In order to detect the outliers and to remove sensor noise, some of the pre-processing in the form of filtering has been done on the data prior to considering it as an input set. Subsequently, features are extracted individually for each case to be predicted for the next sample.

C. Air Temperature Prediction

Temperature is one of the most common parameters for an accurate weather forecast. The unit of recorded temperature at Meteo station is Celsius. It is one of the known facts that temperatures gets effected by season. For example, in extreme summer we face scorching heat by sun and in winter we experience freezing cold temperature. Consequently, the recorded temperature has maximum and minimum values. Apart from this the second effecting parameter could be the particular hours in a day; at that time the air temperature can possibly vary, i.e., the day time hours and the night time hours. Since temperature is clearly dependent on the season and hour, these two attributes have been taken into account in order to reach a right nowcasting. Month and Hour have been computed using the date of the record and they have been used as predictors. They have been preprocessed in order to transform them as sinusoidal features as shown in Fig. 3 and 4, respectively.

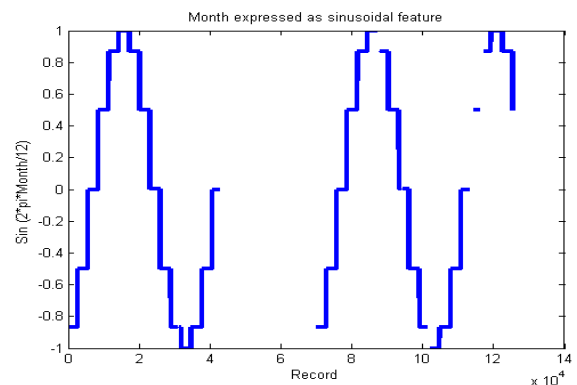


Fig. 3. Recorded months converted into sine waveform.

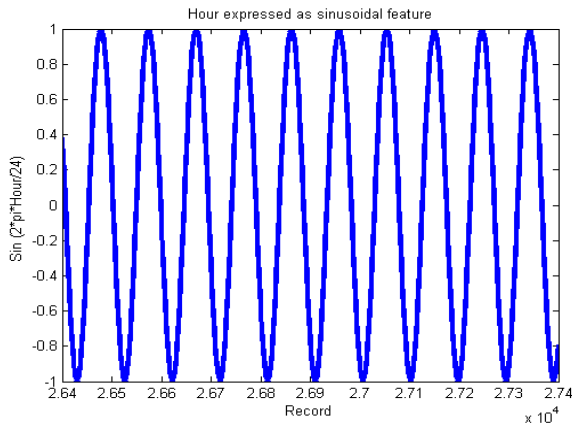


Fig. 4. Recorded hours converted into sine waveform.

In order to predict temperature at time $t+1$, the final input feature set contains particular values of month, hour, temperature at interval (t) and temperature at ($t-1$). Although, before taking the temperature as attribute, we have done pre-processing to reduce the noisy fluctuations from raw sensor data which includes Butterworth lowpass filter with order 2 and 0.11 Cutoff frequency in mHz. The difference between actual and filtered data can be seen in Fig. 5. Moreover, the identified outliers in series were replaced by NAN. Additionally, the interpolation method was applied to cover the missing samples where sensor was unable to record the samples.

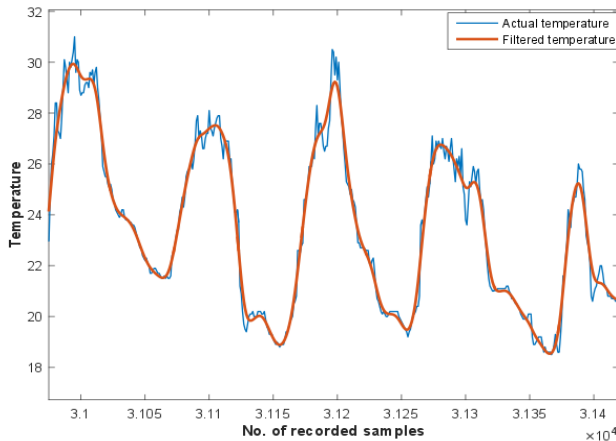


Fig. 5. Actual and filtered temperature series.

The most noticeable step is that the input data as well as labels were normalized in the range of (0,1). Apparently because, we have used RBM which deals with binary hidden and visible units. The detail explanation related to this has already been demonstrated [34]. The training data was selected from October 2010 to March 2014. As aforementioned, the input data set consist of five attributes. According to this, the input layer was based on five nodes, whereas, the output layer with one output neuron. In order to select the number of hidden layers and the size of hidden units in each layer, we preferred random search method. In response to this, we developed and trained several architectures. The selection of hyper-parameter for this model and the next upcoming models

presented in Section III-D and III-E was based on our earlier hypothesis which provided great support to select better model. The best predictive model for temperature prediction, which was initially pretrained layer by layer with total four hidden layers was with the dimension (500-200-100-10). After training each layer separately the model was trained globally by adding an output layer with temperature labels. The architecture of model for temperature prediction is illustrated in Fig. 6. The results are further discussed in Section IV.

D. Relative Humidity Prediction

Humidity is a quantity representing the amount of water vapour in the atmosphere. However, relative humidity depends on the temperature and the pressure of the system of interest. The variation of the temperature, which has a larger variability, depends on the hour and season. Apart from considering the above mentioned attributes, we applied Mutual Information Criteria (MIC) to find the best correlations in between of weather parameters. This further confirmed the attributes selection as mentioned below. The correlations between features computed via MIC are presented in Table 1.

To further explain the feature selection procedures through MIC assume a target class labelled as c . For selecting the features with the highest relevance of attributes to the target class c is crucial. Relevance is usually characterized in terms of correlation or mutual information, of which the latter is one of the widely used measures to define dependency of variables.

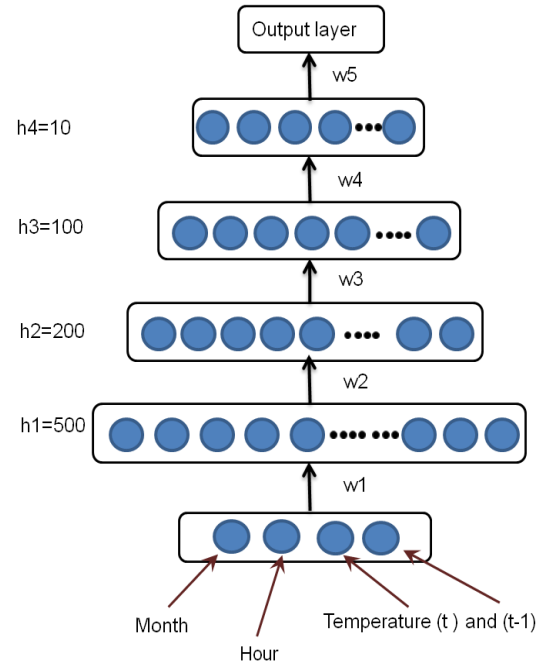


Fig. 6. DBN-RBM for temperature prediction.

Given two random variables x and y , their mutual information is defined in terms of their probabilistic density functions $p(x)$, $p(y)$ and $p(x,y)$:

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad 1$$

The selected features x_i are required, individually, to have the largest mutual information, i.e., $I(x_i; c)$ with the target class

c, reflecting the largest dependency on the target class. In terms of sequential search, the m best individual features, i.e., the top m features in the descent ordering of $I(x_i;c)$, are often selected as the first m features [35].

Hence, features used as inputs for the training are corresponding temperature, previous pressure, previous humidity, corresponding Month and Hour. This feature set and labels were further normalized in the range of (0,1) prior to training. The humidity data was filtered with Butterworth filter corresponding same order of 2 and cutoff frequency at 0.11 mHz. The Actual and filtered humidity data is shown in Fig.7.

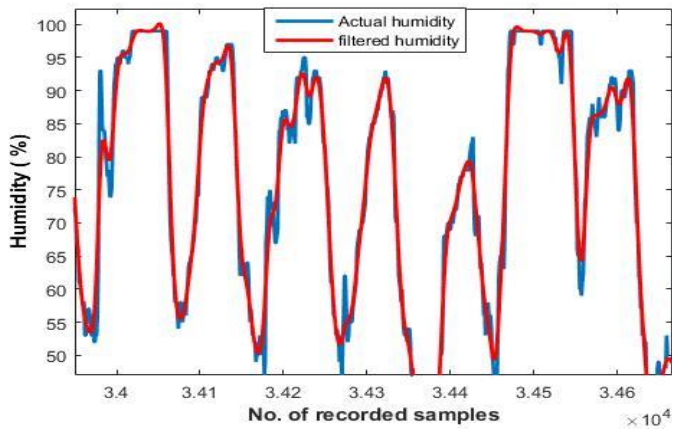


Fig. 7. Actual and filtered humidity series.

In order to construct and train a predictive model for humidity prediction, a DBN was developed with four hidden layers, one input layer consist of Six nodes and one output layer based on one output neuron. The hidden layers were RBM of size (300-200-100-10), which was trained one by one in a layer wise greedy way with contrastive divergence. The model is illustrated in Fig. 8. Initially, all weights and biases were assigned the value zero. The model was trained with total 120k samples and was further tested with rest of the data. The pretraining of RBM was performed using minibatches of size 10, with maximum one iteration for each layer pretraining. After training each layer separately the model was trained globally by adding an output layer with normalized humidity labels. However, fine tuning was performed with Maximum 800 iterations. The results are further discussed in Section IV.

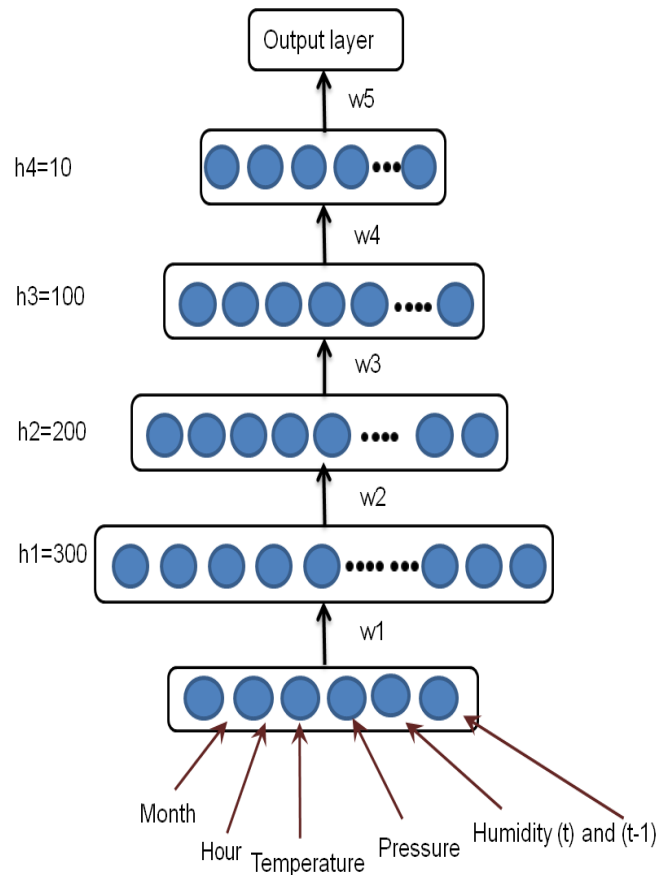


Fig. 8. DBN-RBM for humidity prediction.

E. Pressure Prediction

In general, pressure is a force applied perpendicular to the surface of an object per unit area over which that force is distributed. However, atmospheric pressure or air pressure, sometimes also called barometric pressure, is the pressure exerted by the weight of air in the atmosphere of Earth. The pressure data was smoothed with Butterworth low pass filter in same way as air temperature and relative humidity. However, a high pass filter was also applied on the data to detrend the linearly decreasing trend observed in recorded air pressure series. Fig. 9 presents the graph of actual and filtered pressure samples.

TABLE I. FEATURE SELECTION FOR HUMIDITY USING MUTUAL INFORMATION METHOD

Attributes	Time	Temperature	Pressure	Rain	Humidity	Wind direction	Wind velocity
Time	2.0748	9.06e-4	0.1261	0.0081	0.0246	0.0186	0.0281
Temperature	0.3748	1.933	0.0092	0.0189	0.0189	0.0421	0.0671
Pressure	0.1261	0.0920	1.9787	0.0223	0.0661	0.0227	0.0343
Rain	0.0081	0.0189	0.0223	0.308	0.05	0.01	0.005
Humidity	0.0246	0.0189	0.7256	0.0534	1.8134	0.0415	0.0920
Wind direction	0.0186	0.0421	0.0534	0.0154	0.0415	1.6601	0.0252
Wind velocity	0.0281	0.0671	0.0343	0.0058	0.0920	0.0252	1.7766

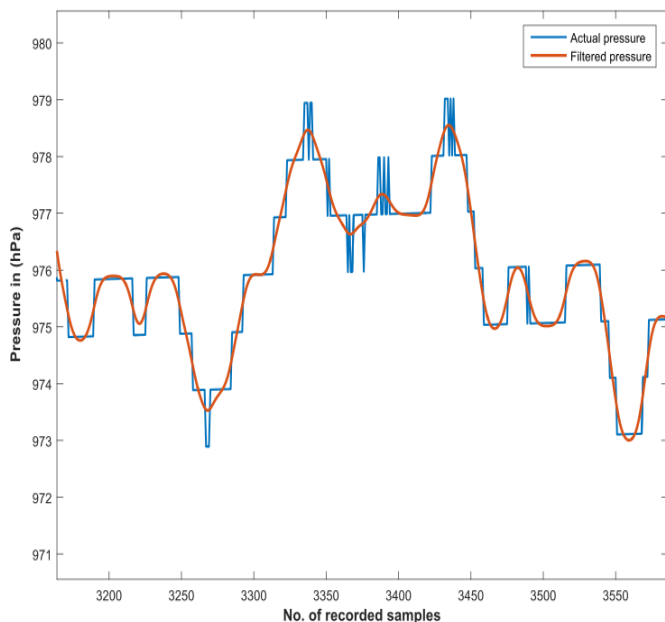


Fig. 9. Actual and filtered humidity series.

In order to extract valuable features for pressure prediction we explored some more aspects. The main factor that affects the air pressure at a given location is the altitude (or height above sea level) of that location. In order to select the meaningful features for air pressure prediction, we did little research. We came to know that the pressure depends on the density or mass of the air. Moreover, the density of air depends on its temperature and from our meteorological dataset the temperature depends on Season (categorized in months) and hour of the day. Thus we took the following parameters as input attributes for predicting the next pressure in series, the month, an hour, corresponding temperature, and pressure at (t) and (t-1).

In order to construct and train a predictive model for pressure prediction, a DBN was developed with three hidden layers, one input layer and one output layer. The hidden layers were RBM of size (300-200-5), which was trained one by one in a layer wise greedy way with contrastive divergence. The model is illustrated in Fig. 10.

The model was trained with total 120k samples and was further tested with the rest of the data. The pretraining of RBM was performed using minibatches of size 10, with maximum one iteration for each layer pretraining. After training each layer separately the model was trained globally by adding an output layer with normalized pressure labels. However, fine tuning was performed with maximum 800 iterations.

Initially, all weights and biases were assigned with the value zero in the training of each predictive model case. However, after pretraining the weights were found to be in form of normally distributed data. This identifies that weights are not randomly initialized in order to find the suitable solution. The weights were further transformed during the fine tuning phase.

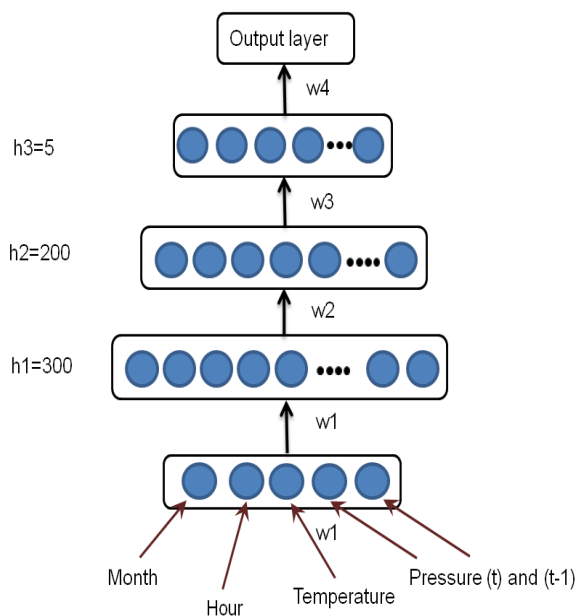


Fig. 10. DBN-RBM for humidity prediction.

Weights in the hidden layers of an each predictive model attempts to explore the nonlinear representations or features from the data. In this regard, the computed weights are also termed as feature detectors or receptive fields. These can be considered as a good way of visualizing which kind of features the hidden units have learned. There is a possibility that less meaningful or insignificant detectors may also be present. The results are further discussed in next section.

IV. RESULTS AND DISCUSSION

In this section, we describe and further discuss in detail the evaluation of trained DBN models for METEO nowcasting. According to our objective, in this work we attempted for the next sample prediction of non-stationary time series through DBN. As aforementioned, we were successful in deploying and training accurate models METEO nowcasting, weather parameter prediction. In our dataset, each weather parameter owns distinct trend and diverse behavior in its time series. For each parameter, we trained separate model and feature selection was performed accordingly. The performance for each predictive model is measured through three different performance metrics, i.e., Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Regression parameter R on Training and Test sets for prediction of Metrological Parameters. The statistics for measuring the performance of each predictive model is presented in the following Table 2.

TABLE II. MSE ON TRAINING AND TEST SETS FOR PREDICTION OF METROLOGICAL PARAMETERS

DBN Models	RMSE		MSE		R
	Training	Test	Training	Test	
Temperature prediction	8.57e-4	9.06e-4	7.35e-7	8.20e-7	0.99
Humidity prediction	1.3e-3	1.5e-3	1.58e-6	2.27e-6	0.99
Pressure prediction	9.07e-4	7.73e-4	8.24e-7	5.98e-7	0.99

The R parameter is linear regression, which relates targets to outputs estimated by network. If this number is equal to 1, then there is perfect correlation between targets and outputs. It is clearly obvious from the measures presented in Table 2 that the number for each model is very close to 1, which indicates a good fit. The MSE is a measure of the quality of predictive model. It is always non-negative, and values closer to zero are better. The MSE is computed as presented in (2), where \hat{y} is a vector of n predictions computed by model, and y is the vector of observed values. Taking the square root of MSE yields the RMSE.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad 2$$

In Fig. 11, we present actual and predicted temperature samples taken from test set. It is visible from the figure that predicted samples are highly replicating the original temperature data.

It is clearly depicted from the Fig. 12 that, predicted humidity samples of test set are very close to the original recorded humidity. In the same way, strong correlations of predicted and recorded pressure can be seen from Fig. 13.

The results obtained from models are robust and shows considerably good predictions. Since, the models perform the forecasting task for only next one sample, however for meteo nowcasting our concerned objective is to predict for next three hours. This is considered as our next future target.

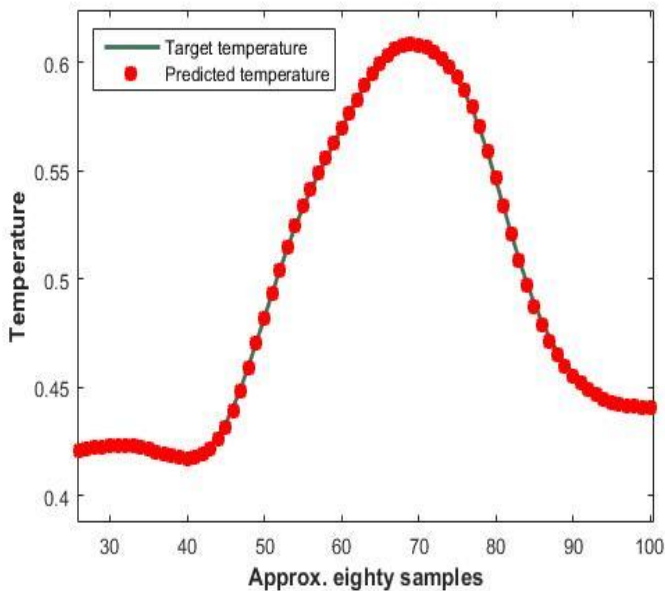


Fig. 11. Close view of target and predicted temperature with eighty samples from test dataset.

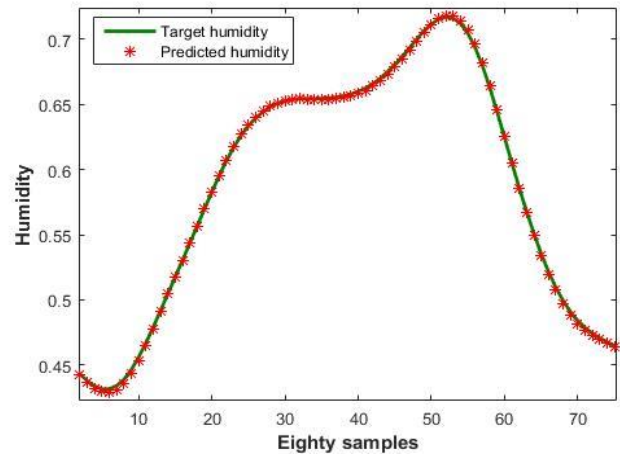


Fig. 12. Close view of target and predicted humidity with eighty samples from test dataset.

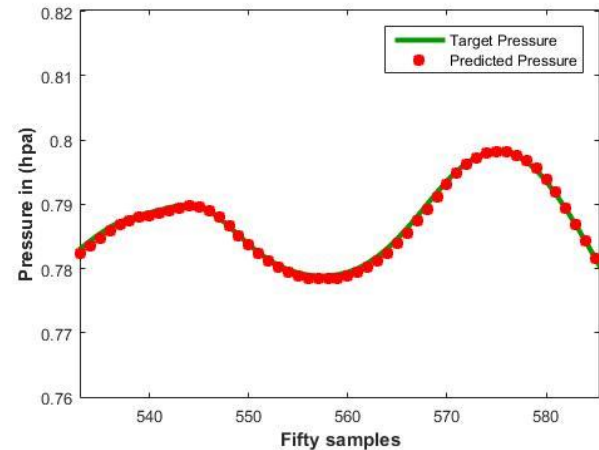


Fig. 13. Close view of target and predicted pressure with fifty samples from test dataset.

V. CONCLUSION AND FUTURE WORK

This research introduces a predictive ANN model based on deep learning hierarchical architecture, for the prediction of weather parameters such as temperature, pressure and humidity. The results shows outstanding performance of implemented DBN models while producing the accurate estimations.

The current research is limited to the implementation of DBN models, with the exception of providing any comparative evaluation with existing traditional neural network models. However, this can be taken as the direction for the future work.

REFERENCES

- [1] QT. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat and A. Rehman, "Sentiment Analysis Using Deep Learning Techniques: A Review," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 8(6), pp. 424-433, 2017.
- [2] S. Narejo and E. Pasero, "Time Series Forecasting for Outdoor Temperature using Nonlinear Autoregressive Neural Network Models," *Journal of Theoretical and Applied Information Technology*, vol. 94(2) pp. 451-463, 2016.
- [3] M. Tektaş, "Weather forecasting using ANFIS and ARIMA models." *Environmental Research, Engineering and Management*. Vol. 51(1), pp. 5-10, 2010.
- [4] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Systems with applications*, vol. 37(1), pp. 479-489, 2010.
- [5] R. E. Abdel-Aal, "Hourly temperature forecasting using abductive networks," *Engineering Applications of Artificial Intelligence*, vol. 17(5), pp. 543-556, 2004.
- [6] K. Abhishek, M. P. Singh, P., S. Ghosh and A. Anand, "Weather forecasting model using artificial neural network," *Procedia Technology*, vol. 4, pp. 311-318, 2012.
- [7] JJ Montaña Moreno, "Artificial neural networks applied to forecasting time series. *Psicothema*, vol. 23(2). 2011.
- [8] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, "Greedy layer-wise training of deep networks," *Advances in Neural Information Processing Systems*, vol. 19, pp. 153-160, 2007.
- [9] F. S. Albuquerque Filho, F. Madeiro, S.M. Fernandes, P.S. de Mattos Neto, T.A. Ferreira, "Time-series forecasting of pollutant concentration levels using particle swarm optimization and artificial neural networks," *Química Nova*, vol. 36(6), pp. 783-789, 2013.
- [10] S. Narejo and E. Pasero. "A Hybrid Approach for Time Series Forecasting Using Deep Learning and Nonlinear Autoregressive Neural Networks.," pp. 69-75, 2016
- [11] E. Busseti, I Osband, and S Wong, "Deep learning for time series modeling," *Technical report, Stanford University*, 2012.
- [12] J.N.K Liu, Y. Hu, J.J. You, and P.W. Chan, "Deep neural network based feature representation for weather forecasting." In *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, p. 1. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2014.
- [13] M. Dalto, J. Matuško, and M. Vašak. "Deep neural networks for ultra-short-term wind forecasting." In *Industrial Technology (ICIT), 2015 IEEE International Conference on*, pp. 1657-1663. IEEE, 2015.
- [14] W. He, "Deep neural network based load forecast." *Comput. Model. New Technol*, vol. 18(3), pp. 258-262, 2014.
- [15] T. Kuremoto, S. Kimura, K. Kobayashi, and M. Obayashi. "Time series forecasting using a deep belief network with restricted Boltzmann machines." *Neurocomputing*, vol. 137, pp. 47-56, 2014.
- [16] Amaury Lendasse Erkki, Erkki Oja, and Olli Simula. *Time series prediction competition: the cats benchmark*. 2004.
- [17] Amaury Lendasse, Erkki Oja, Olli Simula, and Michel Verleysen. *Time series prediction competition: The cats benchmark*, 2007.
- [18] A. Grover, A. Kapoor, and E. Horvitz. "A deep hybrid model for weather forecasting." In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 379-386. ACM, 2015.
- [19] M. Längkvist, L. Karlsson, and A. Loutfi. "A review of unsupervised feature learning and deep learning for time-series modeling." *Pattern Recognition Letters* vol. 42, pp. 11-24, 2014.
- [20] G.W. Taylor, G.E. Hinton, and S.T. Roweis. "Modeling human motion using binary latent variables." In *Advances in neural information processing systems*, pp. 1345-1352, 2007.
- [21] I. Sutskever and G.E Hinton. "Learning multilevel distributed representations for high-dimensional sequences." In *Artificial Intelligence and Statistics*, pp. 548-555, 2007.
- [22] N. Garg, and J. Henderson. "Temporal restricted boltzmann machines for dependency parsing." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers Vol. 2*, pp. 11-17. Association for Computational Linguistics, 2011.
- [23] I. Sorokin, "Classification Factored Gated Restricted Boltzmann Machine." In *AALTD@ PKDD/ECML*. 2015.
- [24] R. Memisevic, and G Hinton. "Unsupervised learning of image transformations." In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1-8. IEEE, 2007.
- [25] H. Lee, P. Pham, Y. Largman, A.Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," *Advances in neural information processing systems*, pp. 1096-1104, 2009.
- [26] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations." In *Proceedings of the 26th annual international conference on machine learning*, pp. 609-616. ACM, 2009.
- [27] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber. "Stacked convolutional auto-encoders for hierarchical feature extraction." *Artificial Neural Networks and Machine Learning-ICANN*, pp. 52-59, 2011.
- [28] K. Chen, M. Seuret, M. Liwicki, J. Hennebert, and R. Ingold. "Page segmentation of historical document images with convolutional autoencoders," In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*, pp. 1011-1015. IEEE, 2015.
- [29] A. Makhzani and B. Frey. "A winner-take-all method for training sparse convolutional autoencoders." In *NIPS Deep Learning Workshop*. 2014.
- [30] E. Pasero and W. Moniaci. "Artificial neural networks for meteorological nowcast," In *Computational Intelligence for Measurement Systems and Applications, CIMSA. 2004 IEEE International Conference*, pp. 36-39. IEEE, 2004.
- [31] M. Costa, W. Moniaci, and E. Pasero. "INFO: an artificial neural system to forecast ice formation on the road." In *Computational Intelligence for Measurement Systems and Applications, 2003. CIMSA'03. 2003 IEEE International Symposium*, pp. 216-221. IEEE, 2003.
- [32] E. Pasero, and W. Moniaci. "Learning and data driver methods for short term meteo forecast." *Lecture Notes in Computer Science 3931*, pp 105, 2006.
- [33] S. Narejo and E. Pasero, "An Application of Internet Traffic Prediction with Deep Neural Network", *Multidisciplinary Approaches to Neural Computing*. Springer , Cham, vol. 69, pp. 139-149, 2018.
- [34] D. Erhan, Y. Bengio, A. Courville, P.A. Manzagol, P. Vincent, and S. Bengio. "Why does unsupervised pre-training help deep learning?." *Journal of Machine Learning Research*, vol. 11(Feb), pp. 625-660, 2010.
- [35] H. Peng, F. Long, and C. Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27(8), pp. 1226-1238, 2005.