A Deep Learning Based Approach for Long-Term Drought Prediction

Norbert A Agana
Department of Electrical and
Computer Engineering
North Carolina A&T State University
Greensboro, NC
Email: naagana@aggies.ncat.edu

Abdollah Homaifar
Department of Electrical and
Computer Engineering
North Carolina A&T State University
Greensboro, NC
Email: homaifar@ncat.edu

***Abstract*—Drought is a natural disaster that comes with high hazardous impacts on the society. Its effects are mostly manifested as hydrological drought. Identifying past droughts and predicting future ones is very vital in limiting their effects. However, the random and nonlinear nature of drought variables makes accurate drought prediction remain a challenging scientific problem. Neural networks have shown great promise over the last two decades in modeling nonlinear time series. But the issue of nonconvex optimization ensues when two or more hidden layers are required for highly complex phenomena. This research looks into the drought prediction problem using deep learning algorithms. We propose a Deep Belief Network consisting of two Restricted Boltzmann Machines for long-term drought prediction using lagged values of Standardized Streamflow Index (SSI) as inputs. The proposed model is applied to predict different time scale drought indices across the Gunnison River Basin located in the Upper Colorado River Basin. The study compares the efficiency of the proposed model to that of traditional approaches such as Multilayer Perceptron (MLP) and Support Vector Regression (SVR) for predicting the different time scale drought conditions. The proposed model shows an edge in performance over the traditional methods using Root Mean Square Error and Mean Absolute Error as metrics.**

***Index Terms*—Deep Belief Network, unsupervised pre-training, Restricted Boltzmann Machine, Drought Prediction.**

I. Introduction

Drought is a natural disaster that occurs when there is a significant deficit in precipitation compared to the long-term average [1]. It impacts society in many ways and has consequences on all components of water-related resources, producing negative environmental, economic and social repercussions [2]. Most effects of drought are manifested as hydrological drought [3]. Due to its slow development, it usually takes a significant the amount of time on its inception for its impact to be perceived by the economic sector. Among all extreme climate events, drought is considered the most complex phenomenon affecting more people [4]. For instance, the United States witnessed a significant drought increase regarding number and severity over the last two decades [5],[4]. As contained in the National Climatic Data Center, USA (2002) database, the United States experienced either severe or extreme during the last century with nearly 10% of the total land area affected [6]. Of the total 58 weather-related disasters recorded within the period, 10 were as a result of droughts and other related heat waves [6]. The ability to design models that can make reliable future predictions constitutes a significant progress in the drought management process. However, the random and nonlinear nature of drought variables makes accurate drought prediction remain a challenging scientific problem.

Several studies conducted in recent years have proposed methods to improve drought forecasting [7],[8],[9],[10][11]. In [8], a seasonal drought prediction model based on a Bayesian framework, was used to characterize hydrologic drought across the Gunnison River Basin. The authors used standardized streamflow index (SSI) for their analysis. In [12], a wavelet-linear genetic programming (WLGP) model was used for long lead-time drought forecasting (with 3, 6, and 12-month lead times) in the state of Texas. They demonstrated that the standard linear genetic programming model is unable to learn the non-linear structure of drought in lead times more than three months. An autoregressive integrated moving average (ARIMA), recursive multistep neural network (RMSNN) as well as a direct multistep neural network (DMSNN) were also used in [13] to forecast drought in the Kansabati River Basin, India. Three different machine learning methods consisting of artificial neural networks, support vector regression, and coupled wavelet-artificial neural networks in [14] were used for long-term drought forecasting at the Awash River Basin in Ethiopia. The forecast results of their study showed that the wavelet-ANN model was most accurate for long-term drought prediction. There are several other methods in the literature for drought forecasting, and the reader can refer to [15] for a more comprehensive review.

All these studies have shown promising results regarding drought forecasts. However, the effects of climate change on climate extreme events, especially in recent decades, highlights the need for more complex methods in predicting these events [16],[17],[18]. Artificial neural networks (ANN) have proven to be suitable for complex time series forecasting [19]. ANN are universal functional approximators that can be used to approximate any complex function, provided a sufficient number of hidden layers and number of units are specified. But the problem of nonconvex optimization ensues when more than

978-1-5386-1539-3/17/$31.00

two hidden layers are required for highly complex systems. This made training of deep neural networks unattractive. Hinton et. al., however rekindled interest in deep learning by proposing an unsupervised greedy layer-wise training for deep networks [20]. Deep learning algorithms trained using this approach have shown empirically to avoid getting stuck in bad local solutions. Deep learning architectures including Deep Belief Networks (DBN) have emerged as promising in applications such as image classification and speech recognition problems [20], [21], [22]. However, their use in time series prediction problems is relatively new and is attracting much attention. Some applications of DBN for time series modeling can be found in [23], [24],[25]. In [23], a Deep Belief Network was used to forecast foreign exchange rates and better performance was found with the DBN model than other standard approaches. In [24], a Deep Belief Network was proposed for short-term drought prediction of the Huaihe River Basin in China. The performance of the DBN-based the model was found to be superior to standard BP neural network. Also in [25], a deep belief network optimized by particle swarm optimization was proposed for time series prediction. We proposed a predictive Deep Belief Network for prediction of long-term (with 6 and 12 months lead times) drought conditions using standardized streamflow index (SSI) as the drought index. Similar to Standardized Precipitation Index (SPI), SSI can be used to monitor drought conditions on a variety of timescales. Different time scales such as 1, 3, 6, 9, 12, 24 months, etc. are commonly computed and used in drought analysis [26]. The various time scales allow these indices to be employed in both short-term and long-term drought applications. The 12-month SSI (SSI 12) for instance, compares the streamflow for 12 consecutive months with all the previous years similar 12 consecutive months streamflow of all available data. SSI 12, SSI 24 and so on are usually used to represent long-term drought patterns. For example, a 12-month SSI at the end of March 2005 compares the streamflow totals for the April 2004 - March 2005 period with similar periods in available history record. The study is carried out in the Gunnison River Basin, located in the Upper Colorado River Basin, shown in Figure 3.

II. Method

In this section, we present the general structure of an artificial neural network, followed by a description of the Restricted Boltzmann Machines which forms the building blocks of the Deep Belief Network. Then we describe how the Deep Belief Network training algorithm is integrated into the standard neural network model for drought forecasting.

A. Artificial Neural Networks

An Artificial Neural Network (ANN) is a machine learning model which operates similarly to that of the operation of biological neurons. Just like biological neural networks, ANN consists of simple neurons and connections that process information to determine a relationship between inputs and outputs [27]. An ANN operates by providing a stimulus to the neuronal model, observing the output, and adjusting the weights until the desired output is achieved. An entry is submitted to the ANN with a corresponding desired target, a defined response for the output. They focus on the variables that are very significant for the output and ignore the information that has little impact on the output.

The multi-layer perceptron (MLP) is the most common ANN architecture used for forecasting [19]. The MLP is a feedforward neural network that consists of at least three layers of neurons. An MLP has an input layer, one or more hidden layers and an output layer with a directed acyclic graph representation network. For instance, let x, and y represent the input and output vectors respectively. The input layer of the network receives the models data vector x, while the output layer gives the models output vector y. An activation function is applied to activate the neurons in the hidden layer. For a three layer network system, the non-linear mapping between the input x and the output y is given by the following equation [26]:

$$y = f_2 \left[\sum_{j=0}^{h} \left[w_j f_1 \left(\sum_{i=0}^{n} w_{ji} x_i \right) \right] \right] \tag{1}$$

where n represents the number of input variables, h, the number of hidden neurons and the y is the output of the network. The parameters w_j and w_{ji} represent the weights and biases connecting the neurons of the input and the hidden layers and between the hidden layers and the output respectively. f_1 and f_2 are the activation functions for the hidden layer and the output layer respectively. An example of a transfer function is the sigmoidal-type transfer function, given by:

$sigm(z) = \frac{1}{1+e^{-z}}$

An ANN is usually learned by adjusting the weights and biases in order to minimize the cost function using the error back-propagation algorithm. The mean-squared error is considered as a cost function in this work and is given by:

$$E_d = \frac{1}{2} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{2}$$

y_i is the actual data and $\hat{y}_i$ is the predicted data.

B. Restricted Boltzmann Machines

Restricted Boltzmann Machines (RBMs) are a type of generative energy based models that are defined in terms of energies of configurations of the visible and hidden units[30]. The configuration simply defines the state of each unit. An RBM consists of two layers, a pair of hidden and visible units [28]. The layer of hidden units are not connected to each other and have undirected, symmetrical connections to a layer of visible units [28]. They only allow connections between a hidden unit and a visible unit but no connections between two visible units or between two hidden units. The restriction is

that their units must form a bipartite graph as shown in Figure 2. The energy of the joint configuration, $E(v,h)$, between visible and hidden units of an RBM is defined by the following equation [29], [30]:

$$E(v,h) = -\sum_{i=visible} a_i v_i - \sum_{j=hidden} b_j h_j - \sum_{ij} v_i h_j w_{ij} \quad (3)$$

where v_i and h_j are the binary states of visible unit i and hidden unit j respectively. a_i and b_j are the biases and w_{ij} is the weight between them. A network with a lower energy indicates that it is in a more desirable state. That is, it has a higher probability of occurring. This energy function is used to calculate the probability that is assigned to every possible pair of a visible and a hidden unit. The probability of a configuration is determined by the energy of the configuration and is given by [30]:

$$P(v,h) = \frac{1}{Z} e^{-E(v,h)} \quad (4)$$

where Z is a partition function (normalization constant) which is a sum of all the possible energy configurations of the visible and hidden units:

$$Z = \sum_{v,h} e^{-E(v,h)} \quad (5)$$

The Contrastive Divergence algorithm is used to train an RBM [28] by presenting a training sample to the visible units and alternatively sampling the hidden units, $p(h|v)$, and visible units, $p(v|h)$. The hidden unit activations are mutually independent given the visible unit activations and vice versa and hence, the conditional probabilities with binary values are calculated using the equations:

$$P(h|v) = \prod_{j=1}^{n} P(h_j|v) \quad (6)$$

$$P(v|h) = \prod_{i=1}^{m} P(v_i|h) \quad (7)$$

where n and m are the number of hidden and visible units respectively. For a single binary hidden and visible unit, the conditional probabilities are determined using:

$$P(h_j = 1|v) = \sigma\left(c_j + \sum w_{ij} v_i\right) \quad (8)$$

$$P(v_i = 1|h) = \sigma\left(b_i + \sum w_{ij} h_j\right) \quad (9)$$

where σ is the activation transfer function. c_j and b_i are the biases, v_i and h_j are the states of the visible and hidden units, and w_{ij} are the weights connecting the units i and j

An RBM is trained such that optimal parameters b, c, and w can be obtained. The gradient of the log probability of a training vector on a weight is of the form [30]:

$$\frac{\delta log p(v)}{\delta w_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model} \quad (10)$$

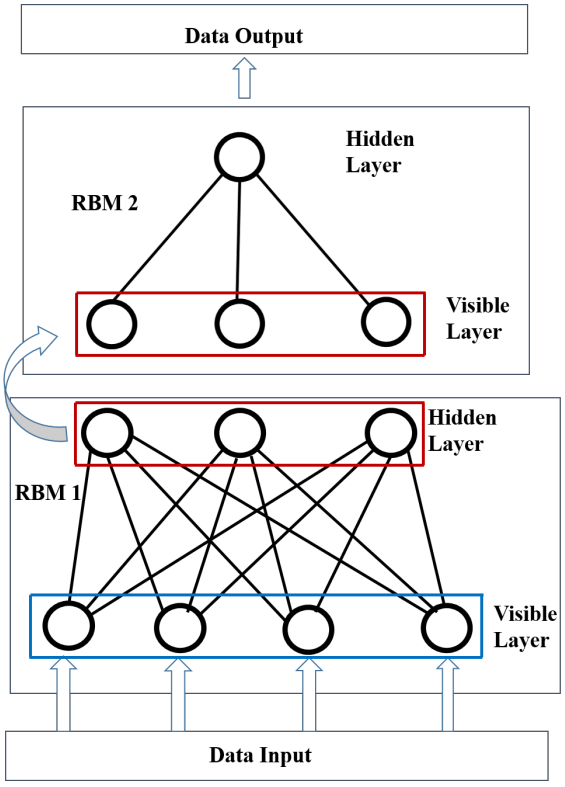

Fig. 1. An Architecture of a Deep Belief Network with two RBMs

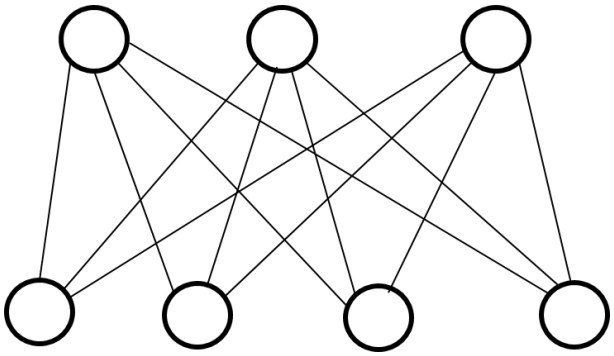

Fig. 2. An Example of a Restricted Boltzmann Machine

where $\langle v_i h_j \rangle_{data}$ expresses the distribution of raw data input to the RBM, and $\langle v_i h_j \rangle_{model}$ is the distribution of data after the model has been reconstructed.

The weights and biases for the RBM can also be updated using the Contrastive Divergence algorithm as follows:

$$\triangle w_{ij} = \alpha\left(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}\right) \quad (11)$$

$$\triangle c_j = \alpha\left(\langle h_j \rangle_{data} - \langle h_j \rangle_{model}\right) \quad (12)$$

$$\triangle b_i = \alpha\left(\langle v_i \rangle_{data} - \langle v_i \rangle_{model}\right) \quad (13)$$

where α is a learning rate.

C. Deep Belief Networks

Although artificial neural networks, trained by back propagation algorithm have proven to be suitable for complex time

series forecasting, the initial values of the weights can affect the learning process. Usually, the parameters are randomly initialized, and this may lead to slower convergence, or the training process may get stuck at local optima solutions, especially networks with several hidden layers [25]. Also, selection of initial weights may lead to different forecast results for each training process. The Deep Belief networks (DBN) by Hinton et al. was proposed to solve these problems [28]. The DBN uses a greedy layer-wise unsupervised learning to pre-train initial weights of the networks and then fine-tune the whole network using standard supervised methods such as a feed-forward neural network with backpropagation algorithm. A Deep Belief Network is a probabilistic generative model with many layers of hidden causal variables. It is constructed by stacking multiple Restricted Boltzmann machines (RBMs) [29] [28]. An example of a DBN consisting of two RBMs is depicted in Figure 1. These RBMs form the building block networks for a DBN. The multilayer neural network can be trained efficiently by using the feature activations of one layer as the training data for the next layer. Better initial values of weights of all layers can be obtained by the layer-wise unsupervised training as compared to random initialization [29].

The training of a DBN consists of two major steps. An unsupervised layer-wise pre-training is performed to obtain optimal parameter values of the network and the entire network is fine-tune using the back-propagation algorithm in a supervised manner. The DBN architecture with an adequate number of hidden layers has a more representative power ability and can approximate highly complex phenomena more effectively than shallow nets.

D. The Proposed Model

We propose a Deep Belief Network with two Restricted Boltzmann Machines to predict long-term drought conditions using lagged values of Standardized Streamflow Index (SSI) as inputs. The structure of the model is shown in Figure 1 with two stacked RBMs. The following steps are followed to implement the prediction model:

1) Obtain the different time scale SSI (SSI 12 and SSI 24 in this case)
2) Normalize the data and split into training and testing data
3) Initialize the parameters of the DBN
4) Pre-train the DBN using unsupervised learning
5) Finetune the entire network using the back-propagation algorithm.
6) Use the test data to test the model: Return to step 3 and retrain the model if unsatisfactory results are obtained.

In a typical RBM, binary logistic units are used for visible nodes We modified the binary nodes to a continuous case to handle the SSI continuous-valued input data. Following the method used in [29], we rescaled the continuous-valued input data to the (0,1) interval and considered each continuous input value as the probability for a binary random variable to take the value 1. The transformation is given by:

$$X_{scaled} = X_{std} * (max - min) + min \tag{14}$$

where $X_{std} = \frac{X_{obs} - X_{min}}{X_{max} - X_{min}}$

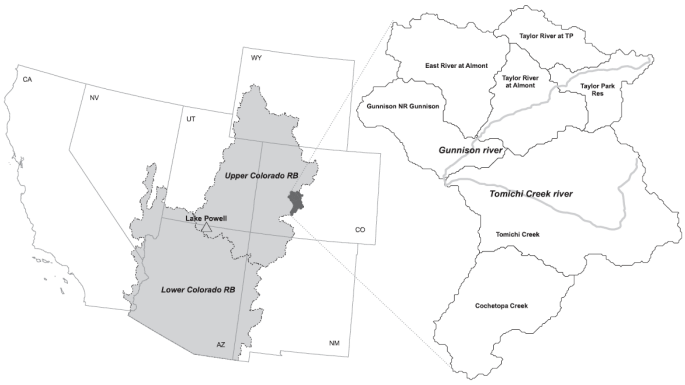

Fig. 3. Location of the Gunnison River Basin [8]

Two performance metrics are used in this work to assess the accuracy of the model. The Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{T}(y_i - \hat{y_i})^2}{T}} \tag{15}$$

and the mean absolute error (MAE):

$$MAE = \frac{\sum_{i=1}^{T}|y_i - \hat{y_i}|}{T} \tag{16}$$

where y_i and $\hat{y}$ are the observed and predicted values respectively. T is the total number of predictions. Both MLP and SVR were implemented using scikits-learn[31].

III. Study Area and Data

The case study is carried out in the Gunnison River Basin, located in the Upper Colorado River Basin shown in Figure 3. The Gunnison River Basin is located in the Southwestern United States with a total drainage area of $5400km^2$ and forms a part of the headwater subbasins of the Colorado River basins [8]. The Colorado River Basin consists of the upper and lower portions. It is made up of seven states: Wyoming, Colorado, Utah, Nevada, California, New Mexico, and Arizona with an approximate drainage area of $640000km^2$ [8]. Similar to the dataset used in [8], streamflow observations at basin outlet station USGS 09114500 (Gunnison River) were used in this paper. We used data from 1912 to 1992 for training and the rest from 1993 to 2013 for testing.

A. Standardized Drought Indices

The Standardized Drought Indices is one group of drought indices. They represent anomalies from a normal condition in a standardized way and is the most commonly used standardized meteorological drought index [3]. We apply the concept employed by McKee et al. [32] for the Standardized Precipitation

TABLE I
CLASSIFICATION OF DROUGHT SEVERITY USED IN THE DROUGHT MONITOR [8]

Category	Severity	SI Value
D0	Abnormally Dry	-0.5 to -0.7
D1	Moderate Drought	-0.8 to -1.2
D2	Severe Drought	-1.3 to -1.5
D3	Extreme Drought	-1.6 to -1.9
D4	Exceptional Drought	-2 or less

Index (SPI) in defining a Standardized Streamflow Index (SSI). The SSI has a calculation procedure similar to SPI. That is, fitting a distribution to the data and transforming it into a normal distribution [3]. Generally, drought indicators defined similar to SPI are called standardized indices (SI). McKee et al used the Gamma distribution for fitting monthly precipitation data series and suggested that the procedure could be applied to other variables relevant to drought other than precipitation such as streamflow, snowpack, soil moisture etc [32]. Droughts are categorized by their level of severity. The severity of droughts characterized by SI is summarized in Table I [8]. These indices are usually identified by the U.S. Drought Monitor classification system. They capture the anomalies from the average moisture status of a region with regards to the drought variables [8]. A similar drought index scheme was used by Madadgar et.al. [8] to characterize the hydrological droughts of the Gunnison River basin. We follow the same procedure by Cacciamani et al [33] to calculate the SSI index. First, model a probability density function on the distribution frequency of the total streamflow cumulated over different time scale of interest (3 months, 6 months, 12 months, 24 months, etc). Then the probability density function is transformed into a normal standardized distribution ensuring zero mean and unit standard deviation [33] [34]. Since the SSI are calculated over different streamflow accumulation periods and scales, they can be used to estimate different potential impacts of a hydrological drought. The 12-month SSI, for instance, shows a comparison of the streamflow for 12 consecutive months with the same 12 consecutive months during the previous years of all available data. Our focus in this study is to forecast the SSI 12 and SSI 24 series which are usually used for long-term drought applications. Longer term drought forecasts can serve as useful information about drought conditions that affect streamflow, groundwater or other hydrological systems.

IV. RESULTS AND DISCUSSION

The proposed drought forecasting model is applied to monthly streamflow data observed from 1912 to 2013, for the Gunnison River Basin, located in the Upper Colorado River Basins. Forecast models for SSI 12 and SSI 24 are presented with lead times of 6 and 12 months. The SSI 12 forecast with 6 months lead time can serve as a 6-month warning time for imminent drought. Likewise, a 12-month lead time forecast can act as a 12-month warning time. The root mean square and mean absolute error performance metrics were used to compare the three models: DBN, MLP, and SVR.

TABLE II
SIX (6) MONTH-LEAD TIME PREDICTION ERRORS FOR SSI 12

Model	R-Squared	MSE	RMSE	MAE
DBN	0.93438	0.00211	0.04572	0.02852
MLP	0.86878	0.00422	0.06468	0.04211
SVR	0.92767	0.00233	0.04821	0.03638

The predictive errors of all three methods were compared and shown in Tables II through V. All the prediction results presented are based on the test data sets. Table II shows the performance results for the SSI 12 forecasts (6 months lead time). From the results, it can clearly be seen that the DBN model had the highest coefficient of determination and recorded the lowest RMSE and MAE errors. Table III shows the same SSI 12 forecast errors but with 12-months lead time. Again, the DBN has the highest prediction accuracy. As expected, the prediction errors for the 12-month lead times are higher than that of the 6-month lead times. Hence, the 6-month lead times forecast captures events with more accuracy than the 12-month lead times prediction. The performance results for the SSI 24 forecast are also presented in Tables IV (6-month lead times) and Table V (12-month lead times). Similar to the SSI 12 forecasts, the DBN shows a comparatively better predictive skill, recording lower prediction errors than that of MLP and SVR. The lower errors resulting from DBN shows that the DBN model can be a better model for drought prediction in the Gunnison River Basin. Also, with the change in the time scale of SSI from 12 to 24, the RMSE and MAE become smaller, implying that the prediction accuracy of SSI 24 is better than that of SSI 12. For a more visual comparison, the RMSE and MAE results of Table II are plotted in Figures 4 and 5 respectively. The DBN model outperforms the MLP model with the SVR model showing a competitive performance as can be observed from the MAE box plot. Figures 6 through 11 shows the plots of the SSI 12 and SSI 24 forecasts using the three models (DBN, MLP, and SVR). The DBN predictions result for SSI 12, and SSI 24 are shown in Figures 6 and 9, respectively. It can be observed that the SSI 24 forecast show fewer deviations and are therefore more accurate than that of the SSI 12 forecast. Similar observations can be made about the prediction results of MLP and SVR. For a more comparative analysis, the three model predictions results for SSI 24 are shown in Figure 12. Also, we show the training and prediction results of the proposed model for the SSI 24 in Figure 13 (with a lead time of 6 months) and Figure 14 (with lead time of 12 months). As can be seen, the prediction results of SSI 24 with 6 months lead time is more accurate than the 12 months lead time prediction.

V. CONCLUSION

This paper looked into the drought prediction problem using a deep learning based approach. We proposed a deep belief network (DBN) for long-term drought prediction and compared its performance to that of standard MLP and SVR models. The DBN model was found to provide better pre-

TABLE III
TWELVE (12) MONTH-LEAD TIME PREDICTION ERRORS FOR SSI 12

Model	R-Squared	MSE	RMSE	MAE
DBN	0.91825	0.00268	0.05177	0.03429
MLP	0.81267	0.00615	0.07844	0.05128
SVR	0.89951	0.00330	0.05745	0.04402

TABLE IV
SIX (6) MONTH-LEAD TIME PREDICTION ERRORS FOR SSI 24

Model	R-Squared	MSE	RMSE	MAE
DBN	0.96415	0.00125	0.03535	0.01876
MLP	0.91302	0.00303	0.05507	0.03969
SVR	0.95481	0.00157	0.03968	0.02325

TABLE V
TWELVE (12) MONTH-LEAD TIME PREDICTION ERRORS FOR SSI 24

Model	R-Squared	MSE	RMSE	MAE
DBN	0.949130	0.001805	0.042485	0.027259
MLP	0.863105	0.004858	0.069699	0.046303
SVR	0.946658	0.001893	0.04350	0.04350

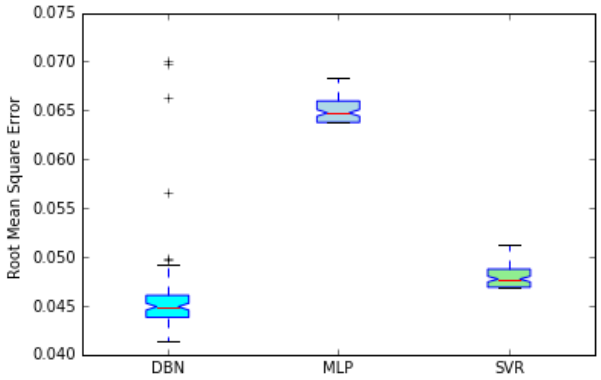

Fig. 4. Comparison of the RMSE of the three models (SSI 12: 6 months lead time)

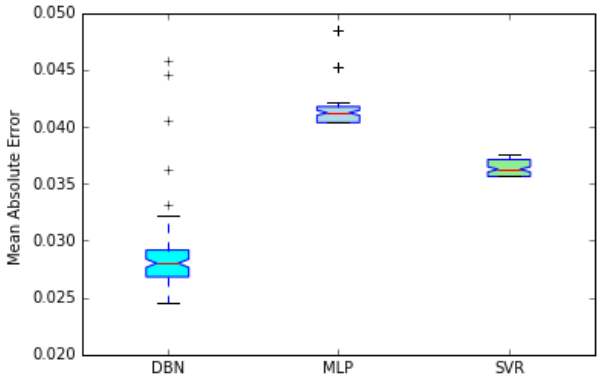

Fig. 5. Comparison of the MAE of the three models (SSI 12: 6 months lead time)

diction results, recording lower prediction errors compared to MLP and therefore can be more reliable and efficient

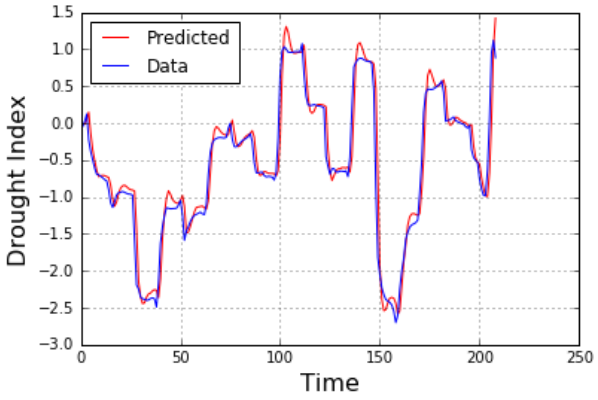

Fig. 6. Observed and predicted SSI 12 using the DBN model (6 months lead time)

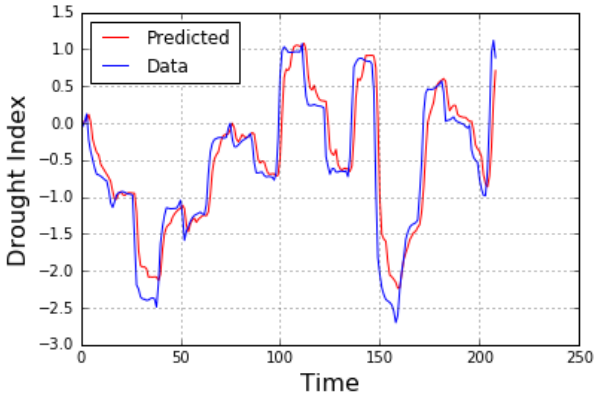

Fig. 7. Observed and predicted SSI 12 using the MLP model(6 months lead time)

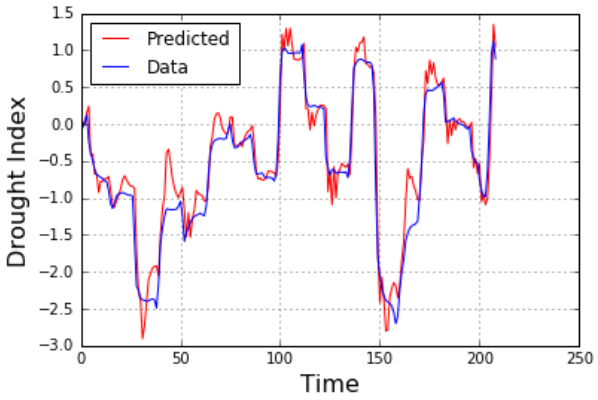

Fig. 8. Observed and predicted SSI 12 using the SVR model(6 months lead time)

for long-term drought prediction. It's performance over the SVR was however less significant. This is probably due to the unavailability of large sample size to fully utilize the capabilities of the deep architecture in the DBN model. The use of only the standardized streamflow index may also be a limitation of this study. Future work will involve evaluating

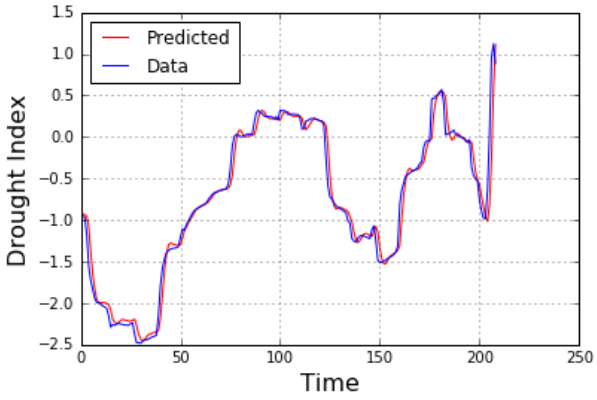

Fig. 9. Observed and predicted SSI 24 using the DBN model(6 months lead time)

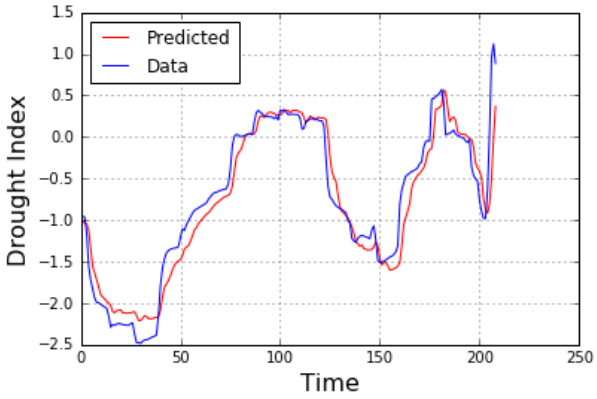

Fig. 10. Observed and predicted SSI 24 using the MLP model(6 months lead time)

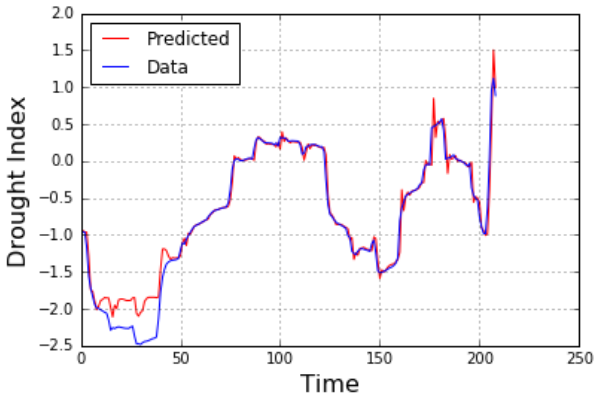

Fig. 11. Observed and predicted SSI 24 using the SVR model(6 months lead time)

the performance of the DBN model for drought prediction by consideration global climate indices in addition to the standardized streamflow index. Since the SVR model showed competitive performance with the DBN model, we may also consider using the DBN for pre-training and the SVR for the final prediction in our future work.

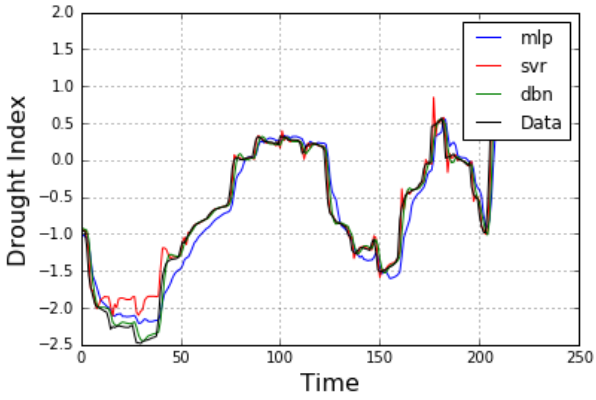

Fig. 12. Observed and predicted SSI 24 for all three models(6 months lead time)

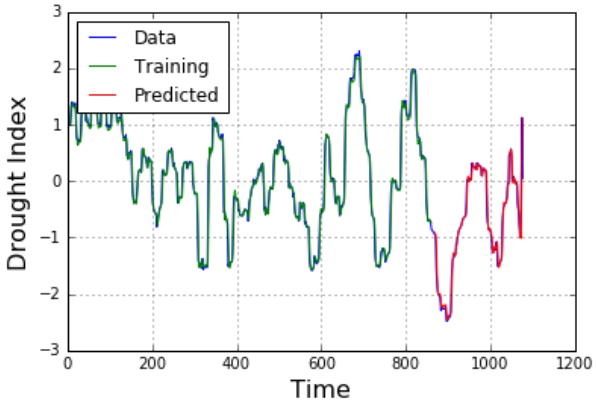

Fig. 13. Training and prediction of SSI 24 using the DBN model(6-Months lead time)

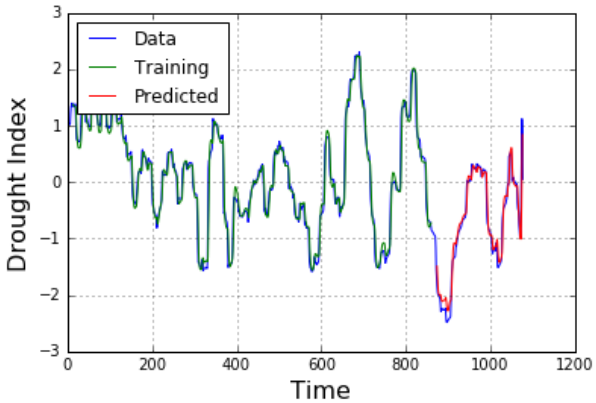

Fig. 14. Training and prediction of SSI 24 using the DBN model(12-Months lead time)

ACKNOWLEDGMENT

This work is partially supported by the Expeditions in Computing by the National Science Foundation under Award Number: CCF-1029731.

REFERENCES

[1] A Belayneh, J Adamowski, B Khalil, and J Quilty. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmospheric Research*, 172:37–47, 2016.

[2] Antonino Cancelliere, G Di Mauro, B Bonaccorso, and G Rossi. Drought forecasting using the standardized precipitation index. *Water resources management*, 21(5):801–819, 2007.

[3] Anne F Van Loon. Hydrological drought explained. *Wiley Interdisciplinary Reviews: Water*, 2(4):359–392, 2015.

[4] Donald A Wilhite and Michael J Hayes. Drought planning in the united states: status and future directions. In *The arid frontier*, pages 33–54. Springer, 1998.

[5] Stanley A Changnon, Roger A Pielke Jr, David Changnon, Richard T Sylves, and Roger Pulwarty. Human factors explain the increased losses from weather and climate extremes. *Bulletin of the American Meteorological Society*, 81(3):437, 2000.

[6] Tom Ross and Neal Lott. *A climatology of 1980-2003 extreme weather and climate events*. US Department of Commerece, National Ocanic and Atmospheric Administration, National Environmental Satellite Data and Information Service, National Climatic Data Center, 2003.

[7] A Ramachandra Rao and G Padmanabhan. Analysis and modeling of palmer's drought index series. *Journal of hydrology*, 68(1):211–229, 1984.

[8] Shahrbanou Madadgar and Hamid Moradkhani. A bayesian framework for probabilistic seasonal drought forecasting. *Journal of Hydrometeorology*, 14(6):1685–1705, 2013.

[9] AK Mishra and VR Desai. Drought forecasting using stochastic models. *Stochastic Environmental Research and Risk Assessment*, 19(5):326–339, 2005.

[10] Cristina Fernández, José A Vega, Teresa Fonturbel, and Enrique Jiménez. Streamflow drought time series forecasting: a case study in a small watershed in north west spain. *Stochastic Environmental Research and Risk Assessment*, 23(8):1063–1070, 2009.

[11] Mehmet Özger, Ashok K Mishra, and Vijay P Singh. Long lead time drought forecasting using a wavelet and fuzzy logic combination model: a case study in texas. *Journal of Hydrometeorology*, 13(1):284–297, 2012.

[12] Ali Danandeh Mehr, Ercan Kahya, and Mehmet Özger. A gene–wavelet model for long lead time drought forecasting. *Journal of Hydrology*, 517:691–699, 2014.

[13] AK Mishra and VR Desai. Drought forecasting using feed-forward recursive neural network. *Ecological Modelling*, 198(1):127–138, 2006.

[14] A Belayneh, J Adamowski, B Khalil, and B Ozga-Zielinski. Long-term spi drought forecasting in the awash river basin in ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508:418–429, 2014.

[15] Ashok K Mishra and Vijay P Singh. Drought modeling–a review. *Journal of Hydrology*, 403(1):157–175, 2011.

[16] Ashok K Mishra and Vijay P Singh. A review of drought concepts. *Journal of Hydrology*, 391(1):202–216, 2010.

[17] Hamid Moradkhani and Matthew Meier. Long-lead water supply forecast using large-scale climate predictors and independent component analysis. *Journal of Hydrologic Engineering*, 15(10):744–762, 2010.

[18] Shahrbanou Madadgar and Hamid Moradkhani. Drought analysis under climate change using copula. *Journal of Hydrologic Engineering*, 18(7):746–759, 2011.

[19] Guoqiang Zhang, B Eddy Patuwo, and Michael Y Hu. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62, 1998.

[20] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[21] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning*, pages 609–616. ACM, 2009.

[22] Tao Liu. A novel text classification approach based on deep belief network. In *International Conference on Neural Information Processing*, pages 314–321. Springer, 2010.

[23] Ren Zhang, Furao Shen, and Jinxi Zhao. A model with fuzzy granulation and deep belief networks for exchange rate forecasting. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 366–373. IEEE, 2014.

[24] Junfei Chen, Qiongji Jin, and Jing Chao. Design of deep belief networks for short-term prediction of drought index using data in the huaihe river basin. *Mathematical Problems in Engineering*, 2012, 2012.

[25] Takashi Kuremoto, Shinsuke Kimura, Kunikazu Kobayashi, and Masanao Obayashi. Time series forecasting using a deep belief network with restricted boltzmann machines. *Neurocomputing*, 137:47–56, 2014.

[26] Seyed Mohammad Hosseini-Moghari and Shahab Araghinejad. Monthly and seasonal drought forecasting using statistical neural networks. *Environmental Earth Sciences*, 74(1):397–412, 2015.

[27] Shishutosh Barua, AWM Ng, and BJC Perera. Artificial neural network–based drought forecasting using a nonlinear aggregated drought index. *Journal of Hydrologic Engineering*, 17(12):1408–1413, 2012.

[28] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[29] Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007.

[30] Geoffrey Hinton. A practical guide to training restricted boltzmann machines. *Momentum*, 9(1):926, 2010.

[31] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[32] Thomas B McKee, Nolan J Doesken, John Kleist, et al. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. American Meteorological Society Boston, MA, 1993.

[33] C Cacciamani, A Morgillo, S Marchesi, and V Pavan. Monitoring and forecasting drought on a regional scale: Emilia-romagna region. In *Methods and tools for drought analysis and management*, pages 29–48. Springer, 2007.

[34] A Cancelliere, G Di Mauro, B Bonaccorso, and G Rossi. Stochastic forecasting of drought indices. In *Methods and Tools For Drought Analysis and Management*, pages 83–100. Springer, 2007.