

Ensemble Learning with CNN and BMO for Drought Prediction

Vandana Sardar, Shilpa Chaudhari, Aniketh Anchalia, Anirudh Kakati, Ankit Paudel, and Bhavana B N
Dept. of Computer Science and Engineering

M S Ramaiah Institute of Technology, Bengaluru, India (Affiliated to VTU)

*Corresponding author: vandana.s@msrit.edu, shilpasc29@msrit.edu, animketh@gmail.com, anirudhkakati@gmail.com, ankitpd12000@gmail.com, bhavanabn07@gmail.com

Abstract— Natural and complex climate disasters like drought have a number of underlying causes that are observed over timescales ranging from months to years. Sustaining natural resources for farming necessitate drought management plans wherein drought prediction is becoming powerful and flexible with intelligent techniques. It has been proved that machine learning and deep learning techniques are successful for drought prediction. Usage of ensemble hybrid intelligent learning algorithm is available for groundwater and gully erosion modeling but rarely emphasized for drought prediction in the literature. This paper discusses ensemble of Convolutional Neural Network (CNN) and Barnacles Mating Optimizer (BMO) to enhance the efficiency of a CNN model for drought prediction. The input for the proposed ensemble learning model includes the indices Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Atmospherically Resistant Vegetation Index (ARVI) and Enhanced Vegetation Index (EVI), which are calculated from satellite data taken for Kolar regions of Karnataka. predicted drought is classified into low drought, moderate drought and severe drought, based on the NDVI value. Improved results are observed.

Keywords—Drought indices, Drought prediction, Convolutional Neural Network and Barnacles Mating Optimizer

I. INTRODUCTION

An intense and persistent lack of precipitation in a specific region causes a weather-related natural disaster called drought, which is propagated through the entire hydrological cycle as shown in figure 1. The land in that

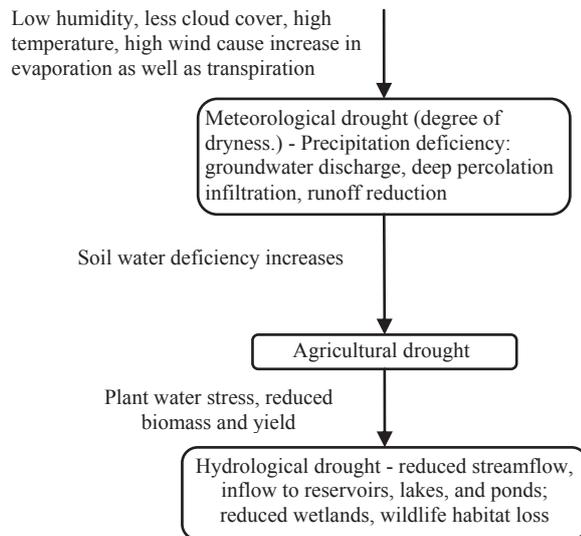


Fig. 1. States of drought in a region

region loses its fertility, thus impacting the natural and socio-economic environments. The events shown on arrows in Figure 1 change the flow/state of the drought.

More than 100 indices can be computed based on various factors for checking severity, location, duration of drought for early warning to the operators to devise contingency plans. This paper uses satellite data as input. The indices that can be computed from satellite data include Enhanced Vegetation Index (EVI), Evaporative Stress Index (ESI), Normalized Difference Vegetation Index (NDVI), Temperature Condition Index (TCI), Vegetation Condition Index (VCI), Vegetation Drought Response Index (VegDRI), Vegetation Health Index (VHI), Water Requirement Satisfaction Index (WRSI), Land Surface Water Index (LSWI), Normalized Difference Water Index (NDWI), Soil Adjusted Vegetation Index (SAVI), Standardized Anomaly Index (SAI).

NDVI monitors the health of vegetation with high spatial coverage. Data continuity, sensor variations, community acceptability, and data maintenance are some of the challenges in using NDVI. TCI determine temperature-related vegetation stress. It measures local and wider region drought. The weakness of TCI is calculated on a weekly timescale which cannot account for the day of the year and/or time of day. Along with NDVI and TCI, VCI is used to measure agricultural drought. VCI predict onset, duration, and severity. It does not indicate cloud contamination and shortage of historical data.

Accurate and timely drought prediction can help to take decision. This paper discusses the ensemble learning used for drought prediction based on NDVI, SAVI, ARVI, and EVI. Computation of various indices uses spatial data and CNN techniques. The proposed learning method performance and its ability to achieve drought prediction accuracy is measured.

The current scenario involves the Drought Management Group that constitutes to coordinate the efforts to deal with drought in various states. National Agricultural Drought Assessment and Monitoring System (NADAMS) use rainfall data for near real-time information on drought prevalence, its severity level and its perseverance. ISRO has developed drought assessment system based on remote sensing-based techniques especially for crop yield. This system uses rainfall data, dry spells, and crop indices. Mahalanobis National Crop Forecast Center (MNCFC) uses it regularly. In addition to many other indicators, comparison of these conditions is made with respect to

normal and previous years as one of the assessment parameters.

Drought assessment includes drought pattern analysis which helps to monitor drought for early warning or future drought prediction. This helps to control or mitigate the drought causes in a particular sector/ecosystem, including water management, agriculture, and dam utilization.

The organization of this paper is as follows. Section II discusses the literature review on the techniques used for drought prediction. Section III explains the proposed methodology wherein the ensemble model design is elaborated. Section IV discusses the experimental result based on the developed hybrid model. Finally, Section V concludes the proposed hybrid model.

II. LITERATURE SURVEY

For the first phase of the literature survey for this work, we focused on papers, texts, journals and articles that clearly defined agricultural drought, the causes and effects of such a calamity. These papers gave us a concise enough understanding of how drought affects every aspect of human life, livestock and the economy of our country. Interpolation of climatological data was used to predict drought, but the advent of machine learning and the advancements in data science facilitate drought prediction with greater precision based on classification and regression methods. The availability of long-range data for drought prediction enhances the capability over interpolation of climatological data. In this section, research articles on agricultural drought using satellite image processing are discussed to understand the patterns of drought and the methods used to predict drought using satellite data. Table I shows the various techniques used for drought assessment. A few of the techniques used include Convolutional Neural Network (CNN), CNN variants like AlexNet, VGGNet, Random Forest (RF), Random Forest Regression (RFR), Support Vector Machine (SVM), Multilayer Perceptron (MLP), MLP neural network (MLP-NN), Deep Belief Network (DBN), Support Vector Regression (SVR), Empirical Mode Decomposition (EMD), Deep forward neural networks (DFNN), Distributed Random Forest (DRF), Gradient Boosting Machine (GBM), Artificial Neural Networks (ANN), Decision Tree (DT), Classification and regression tree (CART), Remote Sensing (RS), Geographic Information System (GIS), Least Squares SVR (LS-SVR), Boosted Regression Trees (BRT), Multivariate Adaptive Regression Splines (MARS), Flexible Discriminant Analysis (FDA).

It is observed that intelligent drought prediction model involves five processes - data collection, feature selection from dataset, prediction of drought, ensembling, performance evaluation and tuning the performance parameters.

Quality and quantity of collected and prepared data impact the drought accuracy. This prepared data using data preprocessing is given as input for selecting features of drought. This paper used satellite images-based dataset. Satellite images are easily available from ISRO satellites. Weather parameters-based drought prediction requires

supercomputers. To collect weather data from earth, which is vast, requires a complex or high-resolution system. Satellite image-based prediction does not require supercomputers. Machine learning models require massive historical data for accurate drought prediction. High resolution satellite images without cloud contamination remains a challenge.

The selected feature helps for drought indices computation. Classification accuracy can be improved by feature selection, which also reduces processing time. The selected features are used in drought prediction models, where intelligent techniques such as ANN, CNN variants, fuzzy, RF etc. are used. It is observed in other areas of research that usage of more than one machine learning algorithm improves accuracy. The ensembling step ensembles machine learning model with other intelligent models.

TABLE I. TECHNIQUES USED FOR DROUGHT ASSESSMENT

Paper	Technique used	Result
[1]	CNN	96% - needs multiclass classification
[2]	CNN, AlexNet, VGGNet	very minimum loss
[3]	Meteorological Drought Identification, Soil moisture assessment, Drought Intensity Classification	Satellite image-based indices useful for agrometeorological drought assessment
[4]	bias-corrected RF, SVM, and MLP-NN	The bias-corrected RF model outperformed other two models.
[5]	DBN, MLP and SVR	Highest coefficient of determination and lowest RMSE and MAE
[6]	EMD, DBN	Improved accuracy
[7]	ILWIS 3.6	Strong relationship between the recorded RF data and the NDVI values
[8]	DFNN, DRF and GBM	DFNN model outperformed both of the other models
[9]	ANN, SVR, DT, RF	Higher accuracy than the other three algorithms.
[10]	CART	Vegetation stress prediction in drought years.
[11]	GIS, ArcScene tool of Arc-GIS	NDVI and rainfall - highly correlated.
[12]	SVR, LS-SVR, RFR	SPI- dependable measure of drought
[13]	CART, BRT, RF, MARS, FDA, and SVM	RF model - highest accuracy
[14]	Drought Index	EDI network model outperforms the SPI model

III. METHODOLOGY

The proposed ensemble learning model using CNN and BMO considers four indices computed from preprocessed satellite images for drought prediction in the Kolar region of Karnataka. The study area is part of Karnataka state in India, specifically, Kolar district, which is located between north latitude 13° 02' 03" and 13° 19' 11" and east longitude 77° 56' 02" and 78° 13' 02". Satellite image data, AwiFs Landsat 8 from 2007 to 2018 is obtained from National Remote Sensing Centre (NRSC) in the areas of Kolar, which are in the form of bands.

The satellite images are uploaded to the database, from where they are acquired and preprocessed. The processed images are used to calculate the indices, which are then saved back into the database. The calculated indices are fed to the CNN model along with BMO to predict the output. The CNN model is built using the weights from the bio-inspired algorithms and satellite images.

The proposed methodology is illustrated in Figure 2 with 5-steps flow. The steps are as follows:

1) *Data collection*: Each image has 4-bands of same row and columns in raster format, which is converted as database object (png format) using Rasterio's open() function from python library as shown in Figure 3.

2) *Drought indices computation*: Green vegetation in near infrared (NIR) and red vegetation in the spectrum (RED) are used. Atmospheric scattering effect with respect to red reflectance spectrum is considered in blue wavelengths (BLUE) for computation of ARVI. EVI considers atmospheric influences as well as soil background signals.

3) *Dataset preparation*: Total 200 images were collected for which indices are computed and stored in database. This indices database is divided into two parts training (70%) and testing (30%).

4) *Ensemble model*: The computed indices are taken as the initial population for the bio-inspired algorithms -BMO. The best fit population is generated and that population is fed as the input to the CNN model. The output is predicted for each image and drought is classified into low drought, moderate drought and severe drought, based on the NDVI value calculated for that image.

5) *Performance Analysis*: The performance of the ensemble model is analyzed in terms of accuracy, convergence time and loss.

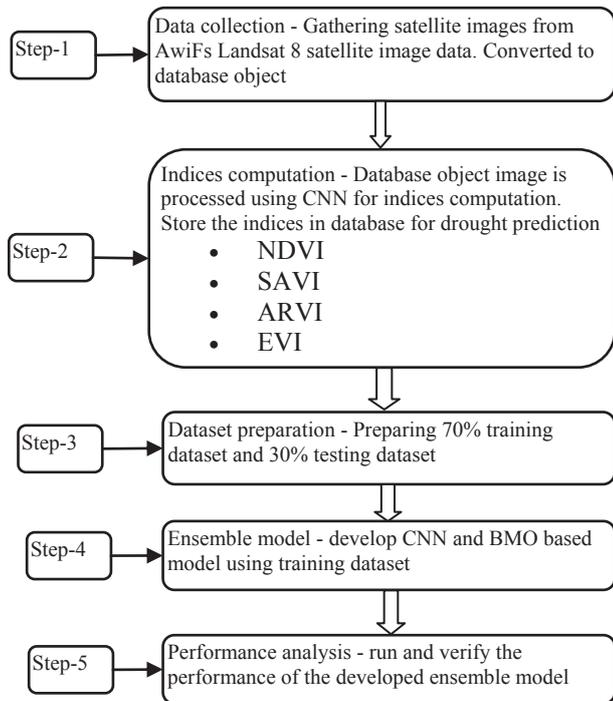


Fig. 2. Flow of event in the proposed ensemble model

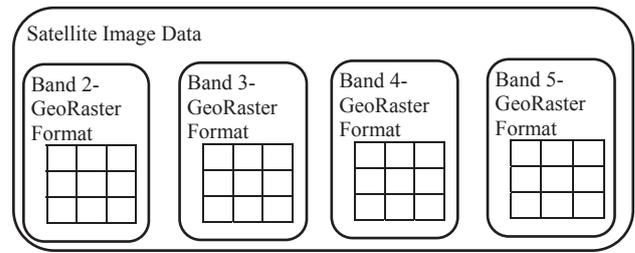


Fig. 3. Satellite Image Bands - rows and columns

The timeline and activation of events for Drought Prediction Model is shown in Figure 4. To execute these functionality, various data is passed from one function to another. The flow of this data is shown in Figure 5.

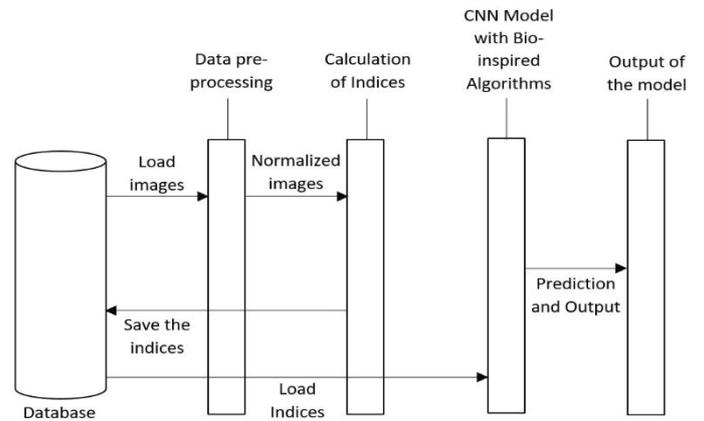


Fig. 4. Timeline of Drought Prediction Model

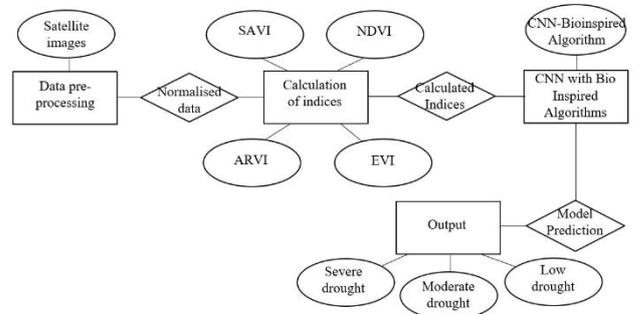


Fig. 5. Various data used for Drought assessment

The classes designed for drought assessment are given in Figure 6.

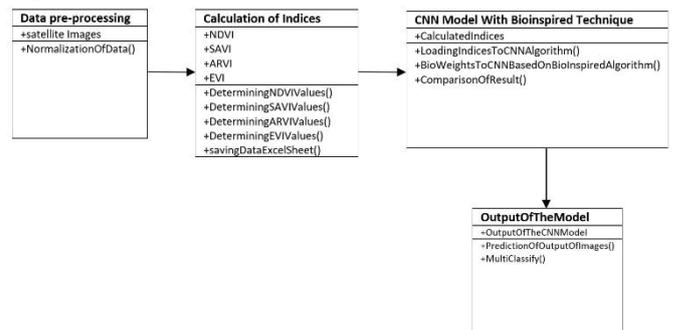


Fig. 6. Classes for Drought assessment

The indices calculation procedure is depicted in Figure 7.

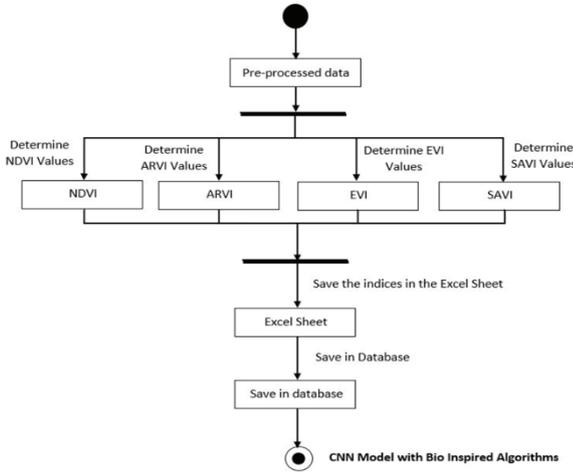


Fig. 7. State Diagram for Drought indices computation

The image's raw pixel data is given as input to CNN model to extract image features automatically. The image features are further classified for further analysis. AS CNN reduces the developer's effort towards feature extraction, it is used in the proposed drought prediction as the intelligent technique. The initial weight is randomly allocated in CNN, which may affect performance.

Bio-inspired techniques help to achieve near-optimal solutions to large-scale optimization problems, which is an emerging approach based on biological evolution of nature. These are stochastic search techniques which are developed to achieve near-optimal solutions to large-scale optimization problems. The Barnacles Mating Optimizer algorithm is used to optimize the weights for the CNN model.

We load the calculated indices from the database and initialize the population for BMO with the indices as shown in (1). We use the standard F3 fitness function to calculate the fitness values.

$$X = \begin{bmatrix} \text{NDVI} & \text{SAVI} & \text{EVI} & \text{ARVI} \\ x_1^1 & x_1^2 & x_1^3 & x_1^4 \\ \vdots & \vdots & \vdots & \vdots \\ x_n^1 & x_n^2 & x_n^3 & x_n^4 \end{bmatrix} \quad (1)$$

The BMO algorithm is implemented as follows:

Set $n=24$

Set $\text{max_iterations}=100$

Initialize the population of barnacle X_i

Calculate the fitness of each barnacle

Sorting to locate the best result at the top of the Population (T =the best solution)

Take pl as the length of the barnacles' penises

q is the normal distributed random number and $q = 1-p$.

while ($l < \text{max_iterations}$)

Set the value of $pl=7$

Using the following equations Selection is done:

$\text{barnacle_dad} = \text{randperm}(n)$

$\text{barnacle_mum} = \text{randperm}(n)$

if selection of dad and mum $\leq pl$

for each variable

Using given equations generate the offspring:

$$x_i^{N_new} = px^N_{\text{barnacle_dad}} + qx^N_{\text{barnacle_mum}} \text{ for } k \leq pl$$

end for

else if selection of dad and mum $> pl$

for each variable

Using given equations generate the offspring:

$$x_i^{N_new} = \text{rand}() \times x^N_{\text{barnacle_mum}} \text{ for } k > pl$$

end for

end if

Bring the current barnacle back if it goes outside the boundaries

Calculate the fitness of each barnacle

Sort and update T if there is a better solution

$l=l+1$

end while

We classify the drought based on the NDVI values calculated and the predicted output of the image from the previous model, as depicted below:

Begin

Take the predicted output of the image from the CNN model.

Classify the severity of drought based on the NDVI values.

ForEach ndvi value i **begin**

if $i > 0.9$ **then**

$\text{class}[i] = \text{'Low'}$

else if $i < 0.9$ and $i > 0.5$ **then**

$\text{class}[i] = \text{'Moderate'}$

else

$\text{class}[i] = \text{'Severe'}$

Display the results

End

The calculated indices extracted from the satellite data have been fed to the bio-inspired algorithms and the best fit population calculated by them has been taken as the initial weights for the CNN algorithm. The output for the images has been predicted and classified into low drought, moderate drought and severe drought.

IV. EXPERIMENTAL RESULTS

For the CNN model, the accuracy was below 90% but during testing, the accuracy of the model was found to be 91%. The loss was found to be more during training the model. We can see that the loss of the model decreased to less than 0.1 during testing. The two graphs are shown in Figure 8 and Figure 9.

The accuracy of the ensemble model during training was close to 92%. During testing of the model, the accuracy increased to 94%. The model loss was high during training close to 0.30 and was close to 0.20 during the testing phase. The accuracy using F3 was found to be the most (94%). The two graphs are shown in Figure 10 and Figure 11.

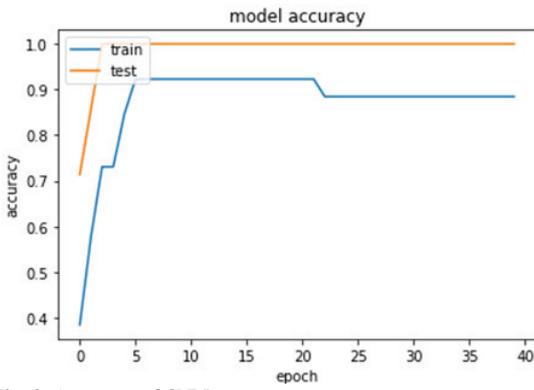


Fig. 8. Accuracy of CNN

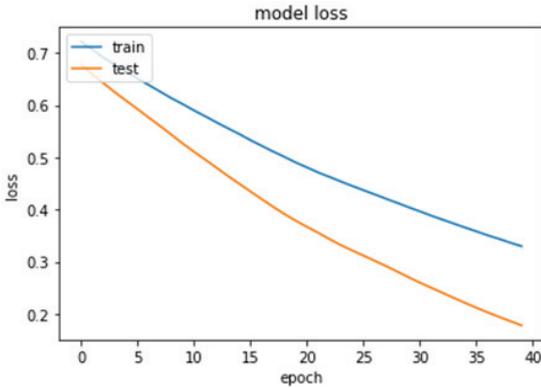


Fig. 9. Loss of CNN

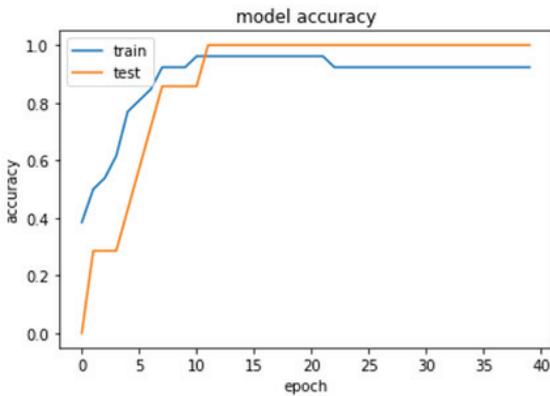


Fig. 10. Accuracy of CNN with BMO

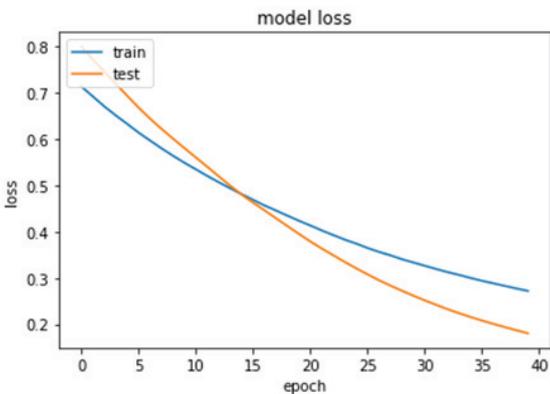


Fig. 11. Loss of CNN with BMO

The multi-classification of drought was done into the categories- Low drought, Moderate drought, and Severe drought. The values of NDVI used for this classification are as follows. (1) Values above 0.9 are classified as chances of low drought. (2) Values between 0.9 and 0.5 are classified as chances of moderate drought. (3) Values below 0.5 are classified as chances of severe drought. It is depicted in Figure 12.

	Image Name	NDVI	SAVI	EVI	ARVI	Classification
0	drought.0.png	0.40	20.31	0.00	1.8600	Severe
1	drought.1.png	0.46	17.12	0.00	1.5700	Severe
2	drought.2.png	0.75	26.78	23.39	25.0000	Moderate
3	drought.3.png	1.09	19.14	92.14	51.0000	Low
4	drought.4.png	0.74	24.83	9.63	37.4832	Moderate
5	drought.5.png	1.09	21.64	50.38	21.0000	Low
6	drought.6.png	0.40	26.78	0.00	6.3300	Severe
7	drought.7.png	1.14	16.26	635.00	51.0000	Low
8	drought.8.png	0.94	16.26	0.00	125.0000	Low
9	drought.9.png	1.00	21.64	11.82	14.0000	Low
10	drought.10.png	0.88	21.64	50.38	25.0000	Moderate
11	drought.11.png	0.88	14.08	4.67	3.0600	Moderate
12	drought.12.png	1.06	20.31	16.25	29.0000	Low
13	nodrought.0.png	1.10	21.64	18.82	37.4832	Low
14	nodrought.1.png	1.00	24.83	0.00	30.3300	Low
15	nodrought.2.png	1.14	16.26	26.00	99.0000	Low
16	nodrought.3.png	1.00	18.08	92.14	49.0000	Low

Fig. 12. Drought classification using NDVI

The results were analyzed considering all the four indices (NDVI, SAVI, EVI, ARVI). The accuracy of the model was found to increase, and the time taken to run the model was found to decrease when BMO was integrated into it. They are tabulated in Table II.

TABLE II. COMPARISON OF ALGORITHMS

Model	Accuracy	Time Taken (Epoch1)		Time Taken (Epoch2)	
CNN	0.91	16s	3s/step	2s	2s/step
CNN with BMO	0.94	3s	2s/step	0s	46ms/step

V. CONCLUSION AND FUTUREWORK

With this study, an ensemble model of BMO and CNN for drought assessment was presented. Different types of drought indices and their relationship were used. Drought prediction using the proposed ensemble approach is not new but unique for drought prediction using satellite images. The developed hybrid model based on deep learning and bio-inspired BMO algorithm improved the accuracy and time taken to run the model is found to decrease. The output was predicted for each image and drought was classified into low drought, moderate drought and severe drought, based on the NDVI value. For the standalone CNN model, the accuracy was 91% while for the CNN with BMO model it was found to be 94%.

For future research, other machine learning techniques could be employed, or a hybrid mode based on different bio-inspired algorithms could be implemented with the goal of increasing the efficiency of the model even further.

REFERENCES

- [1] Sardar, Vandana Sudhakar, K. M. Yindumathi, Shilpa Shashikant Chaudhari, and Prosenjith Ghosh. "Convolution Neural Network-based Agriculture Drought Prediction using Satellite Images." In 2021 IEEE Mysore Sub Section International Conference (MysuruCon), pp. 601-607. IEEE, 2021
- [2] Chaudhari, Shilpa, Sardar, Vandana, D. S. Rahul, M. Chandan, Mahantesh Shivan Shivakale, and K. R. Harini. "Performance analysis of CNN, AlexNet and VGGNet models for drought prediction using satellite images." In 2021 Asian Conference on Innovation in Technology (ASIANCON), pp. 1-6. IEEE, 2021.
- [3] S.Sreelesh, Navneetkaur, Sreeramanaik S. R, "Agricultural drought and soil moisture analysis using satellite, image-based indices", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, volume xlii-3/w6, 2019 ISPRS-GEOGLAM-ISRS Joint Int. Workshop on "Earth Observations for Agricultural Monitoring", 18–20 February 2019, New Delhi, India.
- [4] Feng, Puyu, Bin Wang, De Li Liu, and Qiang Yu. "Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia." *Agricultural Systems* 173 (2019): 303-316.
- [5] Norbert A Agana, Abdollah Homaifar "Deep learning-based approach for long-term drought prediction", Department of Electrical and Computer Engineering North Carolina A&T State University Greensboro, NC.
- [6] Agana, Norbert A., and Abdollah Homaifar. "A hybrid deep belief network for long-term drought prediction." *Proceedings of the Workshop on Mining Big Data in Climate and Environment (MBDCE 2017)*, 17th SIAM International Conference on Data Mining (SDM 2017), Houston, TX, USA, pp. 27-29. 2017.
- [7] Berhan G, Hill S, Tadesse T, Atnafu S. "Using satellite images for drought monitoring: a knowledge discovery approach". *Journal of Strategic Innovation and Sustainability*. 2011 Jul 1;7(1):135-53.
- [8] Prodhan, Foyez Ahmed, Jiahua Zhang, Fengmei Yao, Lamei Shi, Til Prasad Pangali Sharma, Da Zhang, Dan Cao, Minxuan Zheng, Naveed Ahmed, and Hasiba Pervin Mohana. "Deep learning for monitoring agricultural drought in South Asia using remote sensing data." *Remote Sensing* 13, no. 9 (2021): 1715
- [9] Mokhtari, Ramin, and Mehdi Akhoondzadeh. "Data fusion and machine learning algorithms for drought forecasting using satellite data." *فضا و زمین فیزیک*, no. 4 (2021): 231-246
- [10] Tadesse T, Demisse GB, Zaitchik B, Dinku T. "Satellite-based hybrid drought monitoring tool for prediction of vegetation condition in Eastern Africa: A case study for Ethiopia". *Water Resources Research*. 2014 Mar; 50(3):2176-90
- [11] Himanshu SK, Singh G, Kharola N. "Monitoring of Drought Using Satellite Data". *International Research Journal of Earth Sciences*. 2015;3(1):66-72
- [12] Kruseman G, Ornella L, Crossa J, "Satellite Data and Supervised Learning to Prevent the Impact of Drought on Crop Production: Meteorological Drought", DOI: <http://dx.doi.org/10.5772/intechopen.85471>
- [13] Rahmati, Omid, Fatemeh Falah, Kavina Shaanu Dayal, Ravinesh C. Deo, Farnoush Mohammadi, Trent Biggs, Davoud Davoudi Moghaddam, Seyed Amir Naghibi, and Dieu Tien Bui. "Machine Learning Approaches for Spatial Modeling of Agricultural Droughts in the South-east Region of Queensland Australia.", *Science of The Total Environment* 699 (2020): 134230
- [14] Morid S, Smakhtin V, Bagherzadeh K., "Drought forecasting using artificial neural networks and time series of drought indices", *International Journal of Climatology, Int. J. Climatol.* 27: 2103–2111 (2007) Published online 25 April 2007 in Wiley InterScience