PREDICTION OF INDIAN MONSOON USING MACHINE LEARNING

YAJNASENI DASH

CENTRE FOR ATMOSPHERIC SCIENCES

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2022

PREDICTION OF INDIAN MONSOON USING MACHINE LEARNING

by

YAJNASENI DASH

Centre for Atmospheric Sciences

Submitted

in fulfilment of the requirements of the degree of doctor of philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2022

ॐ पूर्णमदः पूर्णमिदं पूर्णात्पूर्णमुदच्यते ।
पूर्णस्य पूर्णमादाय पूर्णमेवावशिष्यते ॥

Om Puurnnam-Adah Puurnnam-Idam Puurnnaat-Puurnnam-Udacyate |
Puurnnasya Puurnnam-Aadaaya Puurnnam-Eva-Avashissyate ‖

Dedicated to Sri Sri Thakur,
My Teachers and Family

Certificate

This is to certify that the thesis entitled "***Prediction of Indian Monsoon Using Machine Learning*" is** being submitted by **Yajnaseni Dash** to the Indian Institute of Technology Delhi for the award of the degree of **DOCTOR OF PHILOSOPHY** is a record of original bonafide research carried out by her. Yajnaseni Dash has worked under our joint guidance and supervision and has fulfilled the requirements for the submission of this thesis. The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree or diploma.

Prof. SAROJ KANTA MISHRA

YES Foundation Chair for Climate Modeling
Principal Investigator, DST Centre of Excellence in Climate Modeling
Associate Professor, Centre for Atmospheric Sciences
Indian Institute of Technology Delhi

Prof. BIJAYA KETAN PANIGRAHI

Institute Chair Professor
Head of the Centre for Automotive Research and Tribology
Professor, Department of Electrical Engineering
Indian Institute of Technology Delhi

Acknowledgments

The thesis on "***Prediction of Indian Monsoon Using Machine Learning***" was a result of the encouragement of many people whose help shaped it. Their provided feedback, direction, and valuable support are noteworthy. It is with hearty gratitude that I acknowledge their contribution to my thesis.

First of all, I would like to thank my guides Prof. Saroj Kanta Mishra and Prof. Bijaya Ketan Panigrahi for their enormous support, constructive suggestions, and advice during the whole duration of my Ph.D. work and for their guidance in my professional career. Successful completion of the present research endeavor could not be possible without their support and guidance. They magnificently guided me through the whole process and helped me to deliver my best efforts all the time. I express my sincere gratitude for the supreme guidance and competent advice received from them. I am especially thankful to my guide Prof. Saroj Kanta Mishra for believing in me and giving me an opportunity to do such interdisciplinary research in atmospheric science.

Besides my advisors, I would like to thank Prof. Sandeep Sahany for his advice and constructive suggestions. I would also like to thank the SRC members Prof. Manju Mohan, Prof. Maithili Saran, and Prof. Abhijit R. Abhyankar for their kind cooperation and constructive inputs. I am extremely thankful to the faculty and staff of the Centre for Atmospheric Sciences for their cooperation during my Ph.D. work.

I would like to thank my seniors, colleagues, and friends in the research group at the DST Centre of Excellence in Climate Modeling and the Centre for Atmospheric Sciences at the Indian Institute of Technology Delhi for their cooperation during my Ph.D. work.

I would like to thank the Ministry of Human Resources and Development and the Indian Institute of Technology Delhi for providing me with financial assistance and computational resources.

This work would not have been possible without the support of my family. I thank my parents, in-laws, elder sister and brother, sister-in-law and brother-in-law, and my two daughters and husband for their unconditional love, moral support, understanding, and tolerance during my long and crazy schedules; without their love and motivation over the years, I would not be at this point today.

Above all, I pay my reverence to the almighty Sri Sri Thakur for His blessings.

Place: New Delhi

Date: 10-10-2022

Yajnaseni Dash

Yajnaseni Dash

Abstract

Accurate prediction of Indian monsoon rainfall is highly desirable; however, it is one of the toughest challenges the scientific community is facing today. Among the various scientific approaches, statistical, numerical, and empirical modeling are popular and widely used in recent times for monsoon prediction, whereas machine learning has not gained any such attention, which although a potential approach, may be able to solve this problem. Rainfall is a product of several atmospheric and oceanic processes and its prediction remains a daunting task. While physically based numerical modeling has progressed considerably over the last half-century, due to the inherent complexities of the underlying physical processes, alternative approaches are explored to make the predictions accurate. Although the dynamical models are capable of providing temporal and spatial distribution of rainfall using cause-effect association among several atmospheric processes, the prediction accuracy by the climate models is not yet at the desired stage and hence, statistical models are still preferred for the rainfall prediction. The statistical models are based on the association of the Indian monsoon with various climate predictors, but, the method of finding a good predictor is quite complex as the monsoon rainfall depends on several weather-related phenomena happening around a year or beyond several years. The other approach besides the aforesaid numerical and statistical based approaches is the empirical approach which uses the past rainfall values to predict the future rainfall values. The conventional models of empirical modeling were based on simple regression based optimization methods; thus failed to produce a higher degree of accuracy in the prediction of an unknown parameter.

In the empirical approach, the rainfall time series is supposed to carry the imprint of all causes and weather phenomena in itself. Thus, it is necessary to analyze the rainfall time series and extract important patterns hidden inside the time series data, and for this,

the explorative data analysis method (Empirical Mode Decomposition- Detrended Fluctuation Analysis (EMD-DFA) was applied in this study. The Indian monsoon rainfall datasets were decomposed into a finite number of empirical modes called Intrinsic Mode Functions (IMF). The obtained IMFs play an important role in rainfall prediction and are also providing a physical basis to relate monsoon rainfall with different meteorological parameters, which show a similar period of occurrences as a particular IMF. The EMD-DFA approach helps us to find out that the first few empirical modes are nonlinear and the rest are linear. Among the different modes of IMFs, the last mode represents the residual, which shows the long-term climatic average and always produces constant values for each corresponding year. According to the current study, the non-linear modes can explain more than 50% of inter-annual variability. The first few IMFs exhibit high-frequency modes. The linear IMFs are slower as compared to the non-linear IMFs and are with less randomization. There is a requirement for analyzing both linear and non-linear frameworks for modeling and predicting the Indian monsoon rainfall. Thus, in this study, both parts were analyzed, and the impact of combining both parts has also experimented.

Presently, machine learning techniques are gaining popularity in other contexts and their lack of application from a climate prediction perspective is a need of concern. The primary goal of inherent non-linear dynamics in meteorology is accurate predictions that need careful model selection, tuning, and validation. The present study is an attempt in this endeavour wherein the applications of machine learning techniques for Indian monsoon rainfall prediction are focused and based on the research outcome and learning experience, attempts were made to improve it. In this study, we have used Artificial Neural Network (ANN), Extreme Learning Machine (ELM) and for the first time applied Regularized Online Sequential Random Vector Function Link (ROS-

RVFL), EMD-DFA and deep Long Short-Term Memory (LSTM) neural network (EMD-LSTM) techniques for the rainfall prediction purpose and subsequently examined their relative performances. Seasonal prediction of rainfall is made a year in advance using the previous year's data. It is found that the deep EMD-LSTM neural network model is producing more accurate prediction outcomes as compared to other techniques used in this study. Based on the computed Central Periods of IMFs, the possible association of the Indian monsoon with the influencing climate predictors has been suggested. The method of decomposing the monsoon rainfall into IMFs and later application of DFA and deep learning techniques is a new approach to Indian monsoon rainfall prediction. The proposed method is efficacious in predicting rainfall statistically as verification is done on an independent test dataset. This study is novel, unique, and expected to give a new direction for the Indian monsoon rainfall prediction.

सार

भारतीय मानसूनी वर्षा की सटीक भविष्यवाणी अत्यधिक वांछनीय है; हालाँकि, यह आज वैज्ञानिक समुदाय के सामने सबसे कठिन चुनौतियों में से एक है। विभिन्न वैज्ञानिक दृष्टिकोणों में, सांख्यिकीय, संख्यात्मक और अनुभवजन्य मॉडलिंग लोकप्रिय हैं और हाल के दिनों में मानसून की भविष्यवाणी के लिए व्यापक रूप से उपयोग किए जाते हैं। मशीन लर्निंग पर ऐसा कोई ध्यान नहीं गया है, जो हालांकि एक संभावित दृष्टिकोण है, इस समस्या को हल करने में सक्षम हो सकता है । वर्षा कई वायुमंडलीय और समुद्री प्रक्रियाओं का एक उत्पाद है और इसकी भविष्यवाणी एक कठिन काम है। जबकि पिछली आधी सदी में भौतिक रूप से आधारित संख्यात्मक मॉडलिंग में काफी प्रगति हुई है, अंतर्निहित भौतिक प्रक्रियाओं की अंतर्निहित जटिलताओं के कारण, भविष्यवाणियों को सटीक बनाने के लिए वैकल्पिक तरीकों का पता लगाया जाता है। हालांकि गतिशील मॉडल कई वायुमंडलीय प्रक्रियाओं के बीच कारण-प्रभाव संघ का उपयोग करके वर्षा का अस्थायी और स्थानिक वितरण प्रदान करने में सक्षम हैं, जलवायु मॉडल द्वारा भविष्यवाणी सटीकता अभी तक वांछित चरण में नहीं है और इसलिए, वर्षा भविष्यवाणी के लिए सांख्यिकीय मॉडल अभी भी पसंद किए जाते हैं। सांख्यिकीय मॉडल विभिन्न जलवायु भविष्यवक्ताओं के साथ भारतीय मानसून के जुड़ाव पर आधारित होते हैं, लेकिन, एक अच्छा भविष्यवक्ता खोजने की विधि काफी जटिल है क्योंकि मानसून की वर्षा एक वर्ष के आसपास या कई वर्षों के बाद होने वाली कई मौसम संबंधी घटनाओं पर निर्भर करती है। उपरोक्त संख्यात्मक और सांख्यिकीय आधारित दृष्टिकोणों के अलावा अन्य एक दृष्टिकोण अनुभवजन्य दृष्टिकोण है जो भविष्य के वर्षा मूल्यों की भविष्यवाणी करने के लिए पिछले वर्षा मूल्यों का उपयोग करता है। पारंपरिक अनुभवजन्य मॉडलिंग सरल प्रतिगमन आधारित अनुकूलन विधियों पर आधारित थे; यह इस प्रकार अज्ञात पैरामीटर की भविष्यवाणी में उच्च स्तर की सटीकता उत्पन्न करने में विफल रहा।

अनुभवजन्य दृष्टिकोण में, वर्षा समय श्रृंखला को सभी कारणों और मौसम की घटनाओं की छाप अपने आप में ले जाना माना जाता है। इस प्रकार, वर्षा समय श्रृंखला का विश्लेषण करना और समय श्रृंखला डेटा के अंदर छिपे महत्वपूर्ण पैटर्न को निकालना आवश्यक है, और इसके लिए, इस अध्ययन में खोजपूर्ण डेटा विश्लेषण विधि (अनुभवजन्य मोड अपघटन- डिट्रेंडेड उतार-चढ़ाव विश्लेषण (ईएमडी-डीएफए) लागू किया गया था। भारतीय मानसून वर्षा डेटासेट को एक सीमित संख्या में अनुभवजन्य मोड में विघटित किया गया था जिसे आंतरिक मोड फ़ंक्शन (आईएमएफ) कहा जाता है। प्राप्त आईएमएफ वर्षा की भविष्यवाणी में महत्वपूर्ण भूमिका निभाते हैं और विभिन्न मौसम संबंधी मापदंडों के साथ मानसून वर्षा को जोड़ने के लिए एक भौतिक आधार भी प्रदान कर रहे हैं, जो दिखाते हैं एक विशेष आईएमएफ के रूप में घटनाओं की एक समान अवधि। ईएमडी-डीएफए

दृष्टिकोण हमें यह पता लगाने में मदद करता है कि पहले कुछ अनुभवजन्य मोड नॉनलीनियर हैं और बाकी लीनियर हैं। आईएमएफ के विभिन्न तरीकों में, अंतिम मोड अवशिष्ट का प्रतिनिधित्व करता है, जो दिखाता है दीर्घकालिक जलवायु का औसत और हमेशा प्रत्येक संगत वर्ष के लिए स्थिर मान उत्पन्न करता है। वर्तमान अध्ययन के अनुसार, गैर-रेखीय मोड ई अंतर-वार्षिक परिवर्तनशीलता के 50% से अधिक को स्पष्ट कर सकते हैं । पहले कुछ आईएमएफ उच्च आवृत्ति मोड प्रदर्शित करते हैं। लीनियर आईएमएफ नॉन-लीनियर आईएमएफ की तुलना में धीमे होते हैं और कम रैंडमाइजेशन के साथ होते हैं। मॉडलिंग और भारतीय मानसून वर्षा की भविष्यवाणी के लिए लीनियर और नॉन-लीनियर दोनों रूपरेखाओं का विश्लेषण करने की आवश्यकता है। इस प्रकार इस अध्ययन में दोनों भागों का विश्लेषण किया गया और दोनों भागों के संयोजन के प्रभाव का भी प्रयोग किया गया है।

वर्तमान में, मशीन लर्निंग की तकनीक अन्य संदर्भों में लोकप्रियता प्राप्त कर रही है, परंतु जलवायु पूर्वानुमान के दृष्टिकोण से उनके कमी अनुप्रयोग चिंता का विषय है। मौसम विज्ञान में अंतर्निहित नॉन-लीनियर गतिशीलता का प्राथमिक लक्ष्य सटीक भविष्यवाणियां हैं जिन्हें सावधानीपूर्वक मॉडल चयन, ट्यूनिंग और सत्यापन की आवश्यकता होती है। वर्तमान अध्ययन इसी प्रयास का एक प्रयास है जिसमें भारतीय मानसून वर्षा की भविष्यवाणी के लिए मशीन लर्निंग तकनीकों के अनुप्रयोगों पर ध्यान केंद्रित किया गया है और शोध के परिणाम और सीखने के अनुभव के आधार पर इसे सुधारने के प्रयास किए गए हैं। इस अध्ययन में, हमने आर्टिफिशियल न्यूरल नेटवर्क (एएनएन), एक्सट्रीम लर्निंग मशीन (ईएलएम) का उपयोग किया है और पहली बार नियमित ऑनलाइन अनुक्रमिक रैंडम वेक्टर फंक्शन लिंक (आरओएस-आरवीएफएल), ईएमडी-डीएफए और डीप लॉन्ग शॉर्ट-टर्म मेमोरी (वर्षा पूर्वानुमान उद्देश्य के लिए LSTM) न्यूरल नेटवर्क (EMD-LSTM) तकनीकों और बाद में उनके सापेक्ष प्रदर्शन की जांच की गई है । पिछले वर्ष के आंकड़ों का उपयोग करके वर्षा की मौसमी भविष्यवाणी एक साल पहले की गई है। यह पाया गया है कि डीप ईएमडी-एलएसटीएम न्यूरल नेटवर्क मॉडल इस अध्ययन में प्रयुक्त अन्य तकनीकों की तुलना में अधिक सटीक भविष्यवाणी परिणाम उत्पन्न कर रहा है। आईएमएफ की गणना की गई केंद्रीय अवधियों के आधार पर, भारतीय मानसून का जलवायु पूर्वानुमानों को प्रभावित करने के साथ संभावित जुड़ाव का सुझाव दिया गया है। आईएमएफ में मानसून की वर्षा को विघटित करने की विधि और बाद में डीएफए और डीप लर्निंग तकनीकों का अनुप्रयोग भारतीय मानसून वर्षा भविष्यवाणी के लिए एक नया दृष्टिकोण है। प्रस्तावित विधि सांख्यिकीय रूप से वर्षा की भविष्यवाणी करने में प्रभावी है क्योंकि सत्यापन एक स्वतंत्र परीक्षण डेटासेट पर किया गया है। यह अध्ययन नया, अनूठा और भारतीय मानसूनी वर्षा की भविष्यवाणी के लिए एक नई दिशा देने की उम्मीद है।

Contents

List of Figures

List of Tables

List of Acronyms

ACF	Autocorrelation function
AMO	Atlantic Multi-decadal Oscillation
ANN	Artificial Neural Network
BP	Back-Propagation
BPNN	Back-Propagation Neural Network
CC	Correlation Coefficient
CEIND	Central India
CP	Central Period
CV	Climate Variability
DFA	Detrended Fluctuation Analysis
EBP	Error-Back Propagation
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
ENSO	El Niño- Southern Oscillation
FFNN	Feed-Forward Neural Network
FFT	Fast Fourier Transform
FSI (α)	Fractal Scaling Index
IAV	Inter-Annual Variability
IITM	Indian Institute of Tropical Meteorology
IMD	Indian Meteorological Department
ISMR	Indian Summer Monsoon Rainfall
IPO	Inter-decadal Pacific Oscillation
JJAS	June-July- August-September
KNEMR	Northeast Monsoon Rainfall the Kerala state
KSMR	Southwest Monsoon Rainfall for the Kerala state
LM	Levenberg-Marquardt
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MLP	Multilayer Perceptron
NEIND	Northeast India

NEMR	Northeast Monsoon Rainfall
NWIND	Northwest India
OND	October-November- December
PCA	Principal Component Analysis
PDO	Pacific Decadal Oscillation
PIND	Peninsular India
QBO	Quasi-Biennial oscillation
$\mathbf{R^2}$	Goodness of Fit or Coefficient of Determination
radbas	Radial Basis Activation Function
RNN	Recurrent Neural Network
RVFL	Random Vector Functional Link Neural Network
ROS-RVFL	Regularized version of Online Sequential Random Vector Functional Link Neural Network
SLFN	Single layer feed-forward neural network
SST	Sea Surface Temperature