# PROGRESS REPORT

| 1. Project Title:  **Forecasting of Rainfall in India Using Large-Scale Climate Indices through Advanced Deep Learning Approach** | DST No:  **SRG/2020/001871** |
| --- | --- |
| 2. PI(Name & Address):  **Dr. Vinit Jakhetiya**  **Assistant Professor, Department of CSE,**  **Indian Institute of Technology Jammu, J&K, India.** | Date of Birth:    19/07/1987 |
| 3. Co-PI(Name & Address): N/A | Date of Birth |
| 1. Broad area of Research   Deep Learning, Artificial Intelligence   * 1. Sub Area:   Predictions and Classifications   1. Approved Objectives of the Proposal: 2. Exploit the growth of latest Machine/Deep Learning techniques such as bi-directioanl LSTM, GRU, Deep Echo State Network (DeepESN) and fusing these techniques via ensemble learning. 3. Predict the rainfall in each sub-region in India, for analyzing the drought and flood in the particular part of the country. 4. Analyze forecasting of the rainfall in all subdivisions and classifying them into five homogenous regions of normal rainfall, excessive rainfall, flood, drought, less rainfall. | |
| Date of Start: 19/12/2020 | Total cost of Project:  Rs. **21,86, 840/-** |
| Date of completion:    18/12/2022 | Expenditure as on 31st march 2022:  **Capital – 8,00,000/- (Committed P.O released (24-04-2022)**  **General – Rs. 4,87,254/-** |

| 1. Methodology:   The process of data acquisition involves collecting samples of information in real-time and converting these samples into numeric values that can be manipulated by a computer. In this work, we have collected data from a public repository named IRI Data Library. Fine preprocessing is performed on collected data, and then it is arranged season-wise. Seasonal information is pity useful in characterizing the patterns of rainfall in any subdivision. An amalgam of the multiple neural networks is being designed for the rainfall forecasting problem. We hypothesize It can get the best features from the models and then assemble them into a single recursive network. Such techniques are being utilized for improving accuracy, better generalization, and providing optimized methods for the removal of prediction uncertainties. In order to get better results, we are using different combinations of loss functions, optimizers, and activations. The proposed algorithm is able to achieve a 0.91 value of pearson’s correlation coefficient (PLCC) between the predicted rainfall and observed rainfall. |
| --- |
| 1. Salient Research Achievements: 2. Data Acquisition:   The datasets were taken from [IRI/LDEO Data Library](https://iridl.ldeo.columbia.edu/). These datasets were in NetCDF(\*.nc) format. SLP and UWND data were provided at 2.5° × 2.5° grid locations. SST data were provided in 2° × 2° resolution. Data from each month from 1948-to 2020 are considered. Using simple python scripts, NetCDF files were converted to .csv files and the data was re-grided to 10° × 20° grid locations which about 324 and 192 grid boxes (rectangular) for UWND, SLP and SST respectively. It is interesting to observe that in this work, we have considered the data of the whole world instead of using the data of specific regions.   1. Model Designing:   for the rainfall forecasting problem, we have achieved a 0.91 value of correlations in predictions and ground truths, using an amalgam of multiple neural network architectures. Deep learning-based methods are ensembled together to achieve the objective of rainfall predictions.   1. Predictords identification:   By using a stacked autoencoder, we identified two categories of predictors: one based on variables specific to the climate and one based on variables that are combined together.  **7.1 Summary of Progress**  **Data Preprocessing**  Acquired data is arranged seasons-wise and finely pre-processed. Our analysis of the various parameters in the collected data focused only on the key features that could be used to predict rainfall. During data analysis, features are selected based on trends and variations. This can be enhanced by using some feature selection algorithms like recursive feature elimination, principal component analysis etc.  In order to identify new predictors for the Indian summer monsoon, certain climate variables at 200hPa pressure level have been examined, including sea level pressure (SLP), sea surface temperature (SST), and zonal U-wind (UWND).  From the dataset, a training and a testing set was created. By using stacked autoencoders, it is possible to uncover monsoon predictors and to create a weather prediction model. New predictors are evaluated using the test data set. From 1949 to 2015, training was conducted, and from 2010 to 2019, testing was conducted.  The monthly anomalies (denoted as Anomaly my for the month m and year y) were calculated, by the following equation, considering the years 1961-1990 as base years.  ***Anomaly my = X my - mean(Xm),***  where X my is the variable for the mth month and yth year and the mean(Xm) denotes the average of the mth month over base period.  **(a) Development of a Deep learning model**  Instead of taking a single‐model output, an amalgam of the multiple neural networks is being designed for the rainfall forecasting problem. We hypothesize It can get the best features from the models and then assemble them into a single recursive network. Such techniques are being utilized for improving the accuracy, better generalization and providing optimized methods for the removal of uncertainties. In order to get better results, we are using different combinations of loss functions, optimizers, and activations.  Basic backbone models are briefly described:   1. **Stacked Autoencoders**   Artificial neural networks constitute a class of architectures known as autoencoders. This architecture has three layers: Input, middle and output layer. With this approach, intricate feature are selected to reduce dimensionality. Autoencoders attempt to determine the input and output values by creating a representation based on the input values. Iterative training empowered the model to learn a nonlinear function. The representation discovered in the internal layer helped to generate outputs such that model aims to reduce the re-construction error.  A deep neural model is generated by heaping several single-layered autoencoders. This architecture continues until the desired depth is reached. The output from first encoder is feeded as input to the second. Fig 1 shows the structure and training method of autoencoder.    **Fig 1**. Set of three autoencoders is trained and then stacked to ensemble the complete architecture.  This is an unsupervised learning approach, since the predictand variable is unknown during the training process. One layer is trained at a time during pre-training. Autoencoders are trained to minimize errors in input reconstruction by separately training them in layers (see left portion of Fig. 1).  A gradient descent method is utilized to fine-tune the layers in stacked architecture, and the weights are adjusted in each layer systematically (see figure 1). With this learning phenomenon, deep autoencoders revealed novel features in their internal layers.  **(b) Identifying The Predictors**  In order to identify predictors, an unsupervised training and fine-tuning algorithm is implemented first, followed by thresholding of the acquired weights. After corelation study with Indian summer monsoon, necessary predictors are identified and, then use them to filter the remaining variables.  In the stacked autoencoder, we found two types of predictors, one based on individual climatic variables, the other based on multiple variables.  In a 2-D list, 5 different models are built based on 7 different input variables. Below is a diagram showing the architecture and the input variables. Fig 2 shows a picture of an autoencoder with the outermost layer being the outer autoencoder, the middle layer being the intermediate autoencoder, and the innermost layer being the inner autoencoder.   1. ***Monsoon predictor identification (Training autoencoder):***   Stack autoencoder architecture was developed by examining data from 1949 to 2015, for predictors identification. According to the input, there are total 330 instances (66 years \* 5 months(Jan-May)). The output of one of the autoencoders is then used as input for the next autoencoder during the unsupervised pre-training of the three autoencoders. Following the transfer of weights from the pre-trained autoencoders to the stacked model, the final fine-tuning of the stacked model is done. To train the layers, we reduced the error during the reconstruction process, following the principles of the autoencoder. Stacking the autoencoder enables discovery of sophisticated and composite nodes.   1. ***Fine Tuning of model:***   We then fine-tune the layers within the stacked model, obtaining three sets of predictors. A threshold adjustment allows for the inclusion of input nodes actively influencing the inner layer node, while excluding the rest. Based on our weight matrix, we can only consider the values, two times the standard deviation over the mean of the weights (learned). After applying the threshold method, three new climatic predictors are evaluated by adding up nodes from the input layer. The base of a new (or potential) monsoon predictors are what makes up the node in the internal layer.   1. ***Post-filtering of identified predictors:***   As a result of studying the correlation between the newly identified predictors corresponding to the nodes in the internal layer and the Indian summer monsoon, the predictors are ranked for each layer. Top ‘n’ ranked features are chosen to fit SVR. The Pearson correlation coefficient (PLCC) is used to find the month of highest correlation, based on the predictors identified. By analyzing results fo SVR, it is seen that May is highly correlated with the identified predictors. In evaluating the identified predictors, we chose the most promising lead month (May).   1. ***Post-treatment by weight thresholding***   As shown in table 2 below, PLCC and SRCC are higher in SLP and SST, where UWND has lower correlations than two. Combinations are also exhibiting good correlation between predicted and actual rainfall.   | **Predictor Combination** | **Outer Layer** | **Middle Layer** | **Inner Layer** | **Full Model** | | --- | --- | --- | --- | --- | | SLP | 324-97-324 | 97-29-97 | 29-9-29 | 324-97-29-9-29-97-324 | | SST | 209-52-209 | 52-21-52 | 21-6-21 | 209-52-21-6-21-52-209 | | UWND | 324-97-324 | 97-29-97 | 29-9-29 | 324-97-29-9-29-97-324 | | SLP-SST | 533-154-533 | 154-48-154 | 48-12-48 | 533-154-48-12-48-154-533 | | SST-UWND | 533-154-533 | 154-48-154 | 48-12-48 | 533-154-48-12-48-154-533 | | SLP-UWND | 648-194-648 | 194-58-194 | 58-17-58 | 648-194-58-17-58-194-648 | | ALL | 857-350-857 | 350-70-350 | 70-25-70 | 857-350-70-25-70-350-857 |   **Table 1:** In every row, three columns define three autoencoders with a mentioned number of nodes which are later stacked to form the autoencoder mentioned in the last column.   | **S. No.** | **Predictor Combination** | **PLCC** | **SRCC** | | --- | --- | --- | --- | | 1 | SLP | 0.9197 | 0.9030 | | 2 | UWND | -0.8803 | -0.8909 | | 3 | SLP-SST | -0.8875 | -0.9030 | | 4 | SST-UWND | -0.8045 | -0.8060 | | 5 | SLP-UWND | -0.8406 | -0.8545 | | 6 | SLP-SST-UWND | -0.8600 | -0.8909 |   **Table 2:** Correlations (between the predicted and original rainfall) of all combinations of predictors  **7.2 New Observations:**  As shown in diagrams, we can say that:   1. Predictors and precipitation (rainfall) showed strong correlations. 2. Predictions derived from deep layers (as shown in Fig1) are more accurate than predictions from shallow layers. 3. Stacked autoencoder is found advantageous over a single layer autoencoder, due to higher predictability of predictors from a deeper layer.     **7.3 Innovations:**   1. Utilization of global predictors instead of local (regional) predictor variables. 2. Every combination is initialized with random weights, and the learned weights are utilized. 3. Final feature vector is converted to sparse, before feeding it to regression model. 4. SVR is used along with Stacked Auto-Encoder, this combination brings novelty in terms of using highly correlated sparse feature vectors to train SVR. 5. Adding a dense layer after input layer of Auto-Encoder as shown in figure below.      1. Normalizaion scheme is applied on the predictions in order to fit the historical rainfall pattern. As seen in section 7.2.   **7.4 Application Potential:**   * + 1. **Long Term**   The designed models investigated through this project have lots of applications. The models which are developed to solve the proposed problems can be used as baseline forecasting tasks. Overall, the results to be obtained in this work will certainly increase the understanding of the inner relationships among several types of prediction/forecast modeling in the field of science and engineering. These identified predictors will be also useful in other related research problems such as climate change, drought forecasting, and flood forecasting.   * + 1. **Immediate**   The designed model is utilized in rainfall prediction in the Indian summer monsoon. A similar method can be replicated to predict the rainfall in 36 homogenous regions. |
| 7.5 Any other  N/A |
| Research work which remains to be done under the project (for on-going projects)   1. Explicit training of developed deep learning method to enable predictions for all 36 subdivisions. Althoug for most of the sub-divisions training has been completed and satisfactory results are obtained. 2. Fine-tuning (making adjustments in order to achieve better predictions) the developed model. |

| PhDs Produced no:  **1** | Technical Personnel trained:  **6** | Research Publications arising out of the present project:  **3** |
| --- | --- | --- |

| List of Publications from this Project(including title, author(s),journals & year(s)   1. Papers published only in cited Journals(SCI) 2. Mumtaz, Deebha, **Vinit Jakhetiya**, Karan Nathwani, Badri Narayan Subudhi, and Sharath Chandra Guntuku. "Non-Intrusive Perceptual Audio Quality Assessment for User-Generated Content Using Deep Learning." ***IEEE Transactions on Industrial Informatics*** (2021). 3. **Jakhetiya, Vinit**, Deebha Mumtaz, Badri Narayan Subudhi, and Sharath Chandra Guntuku. "Stretching artifacts identification for quality assessment of 3D-synthesized views." ***IEEE Transactions on Image Processing*** 31 (2022): 1737-1750. 4. **Jakhetiya, Vinit**, Shubham Chaudhary, Badri Narayan Subudhi, Weisi Lin, and Sharath Chandra Guntuku. "Perceptually Unimportant Information Reduction and Cosine Similarity-Based Quality Assessment of 3D-Synthesized Images." ***IEEE Transactions on Image Processing*** 31 (2022): 2027-2039. 5. Papers published in Conference Proceedings, Popular Journals etc.   NIL | | | | | |
| --- | --- | --- | --- | --- | --- |
| Patents filed/to be filed: NIL | | | | | |
| **Major Equipment(Model and Make)** | | | | | |
| SNo | Sanctioned List | Procured (Yes/No)  Model & make | Cost  (Rs in lakhs) | Working (Yes/No) | Utilization Rate(%) |
| 1 | Server | Purchase Order has been released (25-02-2022) | 8,00,000 | Committed | Comitted |

Upload Images of equipment’s.