# 3. High Performance Computing & Cyber Security

CSIR-4PI continues to serve computational scientists and researchers across CSIR by providing them 24x7 High Performance Computing (HPC) resources backed with high speed and redundant National Knowledge Network (NKN) connectivity to address grand challenge problems in frontier areas of science, technology and engineering. A grant is received from CSIR to establish CSIR centralised HPC, AI & ML Platform (CHAMP) to further cater to the computational requirements of scientists of various CSIR laboratories located across the country. A donation of about 1.2 PF compute power has been received from AMD, USA to carry out COVID related research across India. Cyber security and Cryptography continues to be the major research area of the group with notable scientific outcome. Two Grand-in-Aid R&D projects in the area of cyber security are being executed with external funding from Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology, under the Interdisciplinary Cyber Physical System (ICPS) mission of Government of India.

**Inside:**

- Cyber Security Research: Demonstration of Maliciously-induced-Bursts
- Multi-pronged approaches to short-term prediction of COVID19 positive cases in India
- Multidimensional Ensemble LSTM for Wind Speed Predictions
- Atmospheric Temperature Prediction Using Ensemble Deep Learning Techniques
- High Performance Computing

## 3.1 Cyber Security Research: Demonstration of Maliciously-induced-Bursts

Cyber security continues to be one of the major research areas of the group with notable scientific outcome. A cyber security R&D project sponsored by the Interdisciplinary Cyber Physical System programme of Department of Science and Technology (DST) has resulted in identification of an interesting traffic pattern called Maliciously-induced-Bursts (MiB) on a futuristic transport protocol called Multi Path Transmission Control Protocol (MPTCP). MiB is demonstrated by exploiting a newly exposed vulnerability called Data Sequence Signal Manipulation in MPTCP. The highlights of results from this research are shown in Figure 3.1 (a)-(d) using a MPTCP connection consisting of two sub flows in a network testbed environment. Figure 3.1 (a) and (b) corresponds to a normal scenario in which MPTCP Data Sequence Signal carrying acknowledgements (ACK) are used to clock out new data sequences (SEQ). On the other hand, Figure 3.1 (c) and (d) show the MiB, which are triggered in response to manipulated Data Sequence Signals. Notably, several manipulated segments do not instantly trigger any data sequence, rather they are accumulated over a period of time. Subsequently, at one point in time, one of them triggers a bunch of full-size back-to-back data sequences triggering a MiB. Such MiBs are harmful because they consume substantial amount of buffer space on bottleneck routers and switches, and can also result in sharp increase in queue length, excessive queuing delay and packet loss - the primary causes for performance degradation in packet switched networks.

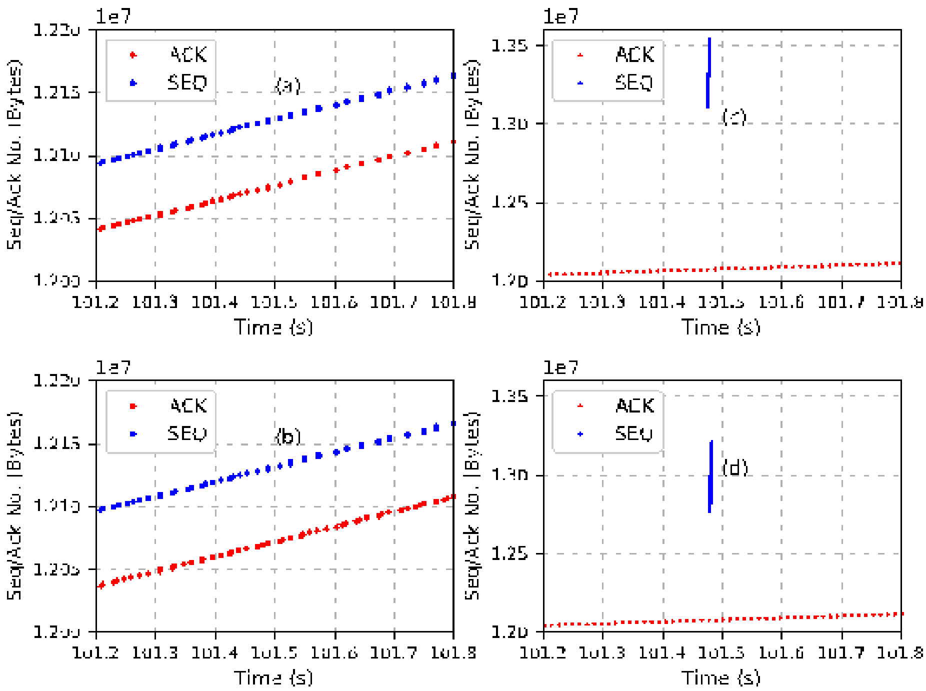


Figure 3.1: The transmission pattern of MPTCP sender at packet granularity under normal scenario and during MiB: Subplots (a) and (b) correspond to subflow1 and subflow2 under normal scenario, and subplots(c) and (d) correspond to subflow1 and subflow2 with MiB.

## 3.2 Multi-pronged approaches to short-term prediction of COVID-19 positive cases in India

The coronavirus disease 2019, known as COVID-19 pandemic is one of the worst human tragedies of recent times. One of the important aspects of preparedness for the pandemic was to establish a proper supply chain management and delivery system for medical equipments, medical consumables and medicines. One of the biggest challenges of supply chain management and delivery system is to distribute the resources in a fair and optimal manner. However, lack of sufficient prior information about the demands results in inefficient management. We aim to design models that can estimate the number of COVID-19 positive cases which can act as useful input for supply chain management and delivery system.

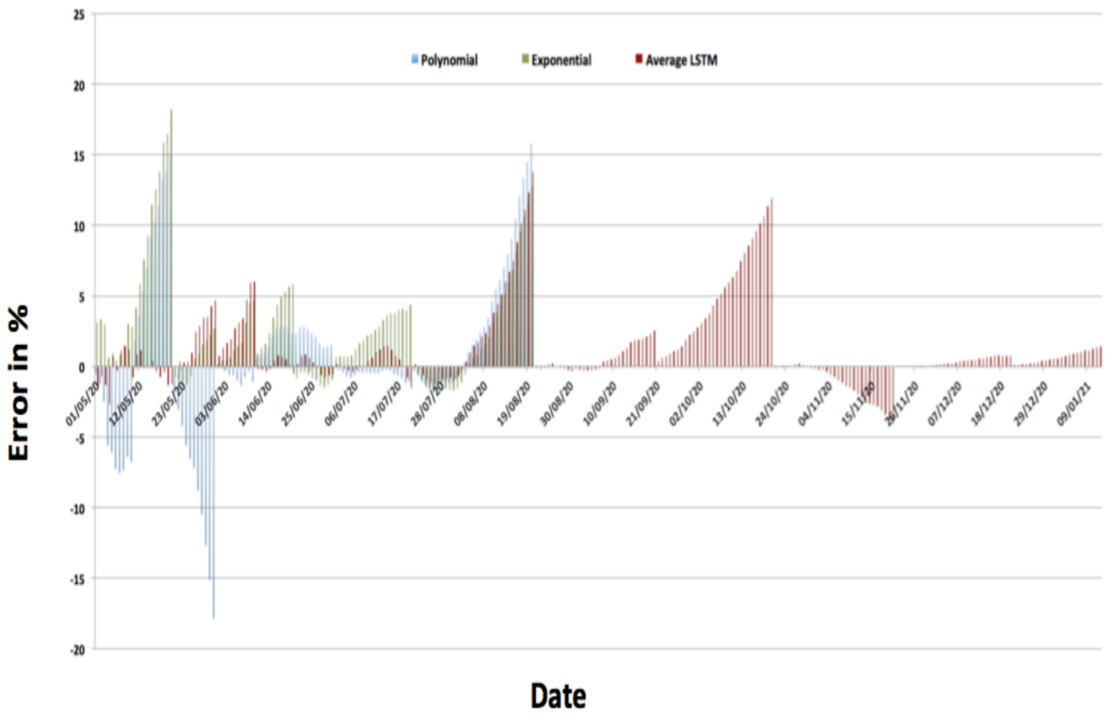


Figure 3.2: Relative error in percentage for prediction of cumulative COVID-19 cases in India using the regression (polynomial and exponential) and the deep learning models (variants of LSTM) at CSIR-4PI. The predictions were done for a block of 10 days starting from 1st May 2020 (i.e. 1st May, 11th May and so on for the next 10 days) and gradually the lead time was increased to one month.

We have used regression (polynomial as well as exponential) as well as deep learning framework based on Long Short Term Memory (LSTM) to provide the prediction on a weekly basis. A sample outcome is presented in Figure 3.2, which demostrates the skill of the prediction framework in terms of percentage of error. LSTM framework is useful for COVID-19 case prediction at national, state (Karnataka) and district (Bengaluru) level. It is possible to automate the data-gathering and prediction process and provide inputs for policy makers. These models cannot predict arrival of a peak – but it can certainly pick up the trend once it sets in.

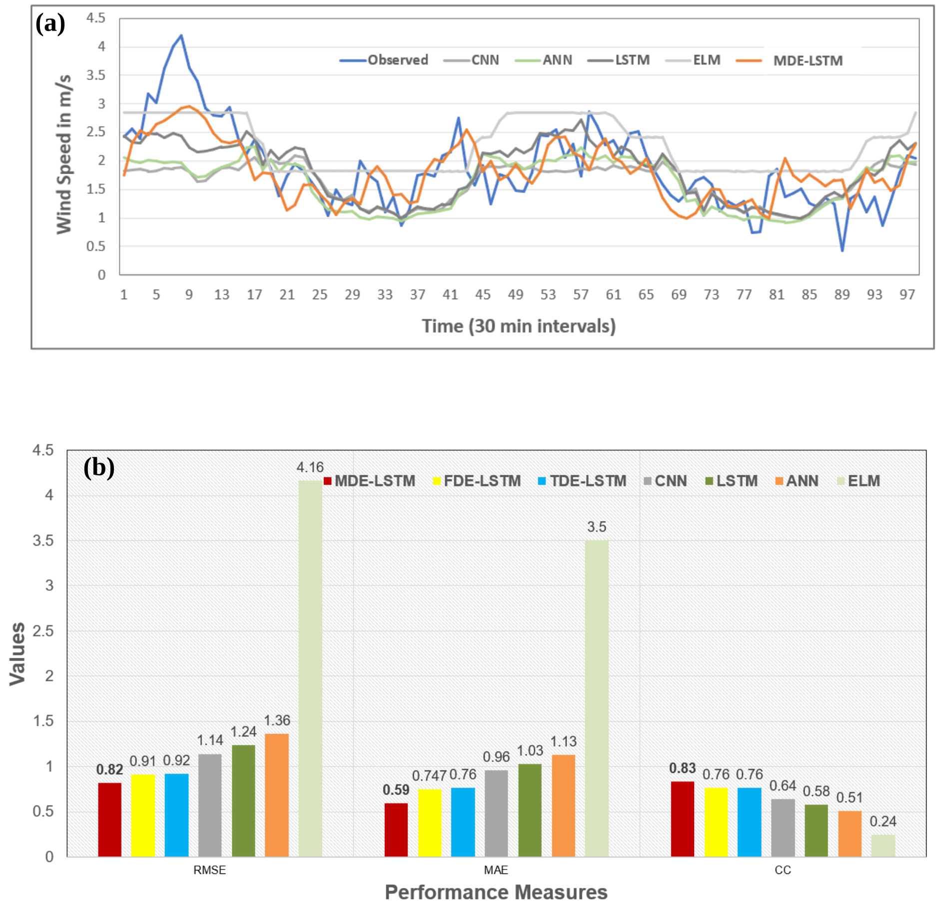


Figure 3.3: (a) Comparison of predicted wind speed from MDE-LSTM, CNN, LSTM, ANN, ELM vs observed wind speed. (b) Comparisons among different performance metrics obtained in various approaches.

## 3.3 Multidimensional Ensemble LSTM for Wind Speed Prediction

Wind energy is an environmental friendly green energy resource that can be well utilised to generate electricity for industrial and domestic purposes. Better prediction of wind speed in reasonable advance time is an essential requirement as it helps in determining the amount of electricity generated by a turbine. However, the accurate prediction of wind speed becomes challenging because of the stochastic behaviour of wind. Artificial intelligence techniques, especially deep learning algorithms, are recently been successful in addressing atmospheric prediction problems on time series data. A novel time series forecasting algorithm named Multidimensional Ensemble LSTM (MDE-LSTM) derived from Long Short-Term Memory (LSTM) technique has been proposed for predicting wind speed at 6 hours lead time. Here, the input dataset is first distributed over time dimension called (TDE-LSTM) and then over frequency dimension (FDE-LSTM) to prepare distinct varieties of dataset that are fed to individual LSTM, whose outputs are again given as inputs to ensemble LSTM to produce the final output. The proposed method is demonstrated with half an hour interval data for the duration 2010 to 2013 from a meteorological tower located in Bangalore. The potential efficiency of the proposed approach has been compared with existing AI based methods such as Convolutional Neural Network (CNN), LSTM, Artificial Neural Network (ANN), Extreme Learning Machine (ELM) in terms of different performance measures as shown in Figure 3.3(a) and 3.3(b). As plotted in Figure 3.3(a), it is evident that MDE-LSTM outperformed all other approaches. This has also been observed from the histogram drawn in Figure 3.3(b).

## 3.4 Atmospheric Temperature Prediction Using Ensemble Deep Learning Technique

In the area of climatology research, temperature is one of the pivotal atmospheric parameters that plays significant role in measuring climate changes. It is important to understand the climate change because, it impacts various activities such as agriculture, solar energy production, travel, climate conditions in extreme cold or hot places, etc. In order to deal with such activities, it is required to predict atmospheric temperature with higher accuracy at sufficient lead time. Generally, complex and dynamical climatology models are used to compute atmospheric temperature. However, these models are time and compute intensive. Therefore, we need alternate approaches to make prediction of different atmospheric variables. With advancement of artificial intelligence, deep learning techniques especially, LSTM is capable of providing better solutions with higher efficiency for time series prediction problems. We have proposed ensemble LSTM to predict atmospheric temperature in seven days lead time. We have considered the effects of other atmospheric parameters such as wind speed, pressure, humidity along with atmospheric temperature to predict temperature at future timestamps. We have used half an hour interval data for the duration 2010 to 2013 from a meteorological tower located in Bangalore to evaluate the efficiency of our proposed algorithm.

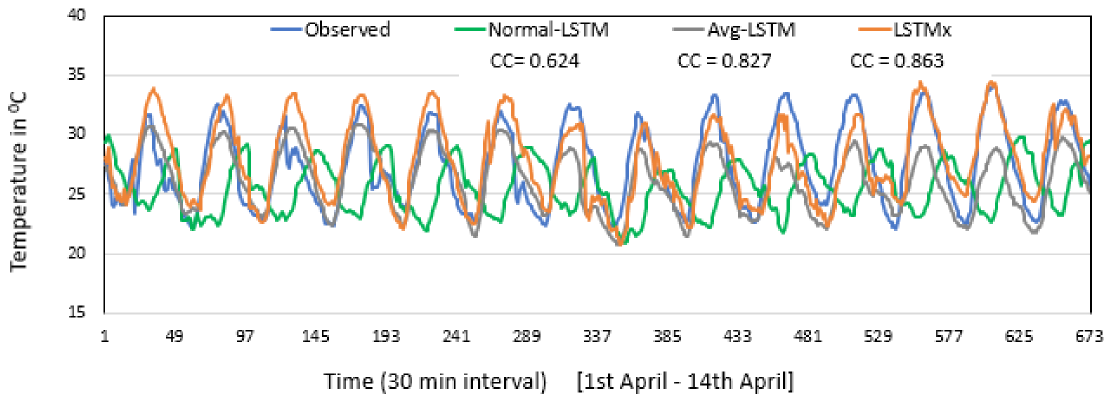


Figure 3.4: Comparison of predicted atmospheric temperature from Ensemble LSTM (LSTMx), Average LSTM, normal LSTM vs observed atmospheric temperature.

As shown in Figure 3.4, the result using proposed method (LSTMx) has performed other methods like normal LSTM and average LSTM. We have also compared different performance metrics such as RMSE, MAE, CC etc. and found the performance of our proposed model superior than other models. The comparison of experimental outcome of the proposed technique with results of existing approaches convincingly justifies the logic behind our technique.

## 3.5 High Performance Computing

CSIR-4PI continues to provide High Performance Computing (HPC) resources to the computational scientists across various CSIR laboratories. This facility is actively used for solving different compute intensive grand challenge problems in the areas of Aerospace, Computational Biology, Computational Chemistry, Material Science, Physical & Environmental Science etc.

Figure 3.5 represents the picture of the CSIR centralised 489 TF HPC Facility. This supercomputer, the Ananta, has a peak computing capability of 489 TeraFLOPS (TF). The system is tightly integrated using two generations of Intel processors i.e. Sandy Bridge, (362 TF of peak) and Skylake, (127 TF of peak).

The supercomputer Ananta is a cluster of 1136 computing nodes distributed over 18 racks, with 1088 nodes each having two eight-core Intel Xeon E5 2670 processors and another 48 nodes each having two eighteen-core Intel Xeon Gold 6140 processors. The total memory of the system is about 77 TB. The inter-node communication are powered by high-speed FDR / EDR InfiniBand interconnect. All the 1136 nodes access LUSTRE parallel file system of about 3 Petabytes that is capable of providing a minimum 20 Gbps simultaneous read and write performance. The PBSPro workload manager ensures efficient usage of the system.

Figure 3.5: CSIR centralised 489 TF HPC Facility.

The efficient support infrastructure plays a pivotal role in the smooth running of the HPC facility. The Ananta supercomputer is positioned in a Tier-3 equivalent state-of-the-art datacenter efficiently supported by a state-of-the-art energy farm. The most significant component of the datacenter is the water-based cooling mechanism with Rear Door Heat Exchangers (RDHx) that has enabled the datacenter as one of the high density and high power-efficient datacenter in the country having Power Usage Efficiency (PUE) of less than 1.5. The energy farm consists of two redundant compact substations having capacity of 1.25 MVA each, three 1010 KVA diesel generators, an underground diesel yard having capacity of more than 15000 litres, and three numbers of UPS (400 KVA each) with battery backup for ensuring 24x7 power supply to the datacenter.