Analysis and Forecasting of Carbon Emission in SAARC Countries using Attention-based LSTM

1st Anil Verma
Computer Science and Engineering
Indian Institute of Technology, Patna
Patna, India
anil_2321cs11@iitp.ac.in

2nd Harshit Dhankhar
Department of Mathematics
Indian Institute of Technology Patna
Patna, India
harshit_2101mc20@iitp.ac.in

3rd Prof Rajiv Misra
Computer Science and Engineering
Indian Institute of Technology, Patna
Patna, India
rajivm@iitp.ac.in

4th Prof. T.N. Singh
Director
Indian Institute of Technology, Patna
Patna, India
tnsiitb@gmail.com

5th Om Prakash Dhakal
Electronics and Computer Engineering
IOE, Purwanchal Campus
Dharan, Nepal
omprakash@ioepc.edu.np

2023 IEEE International Conference on Big Data (BigData) | 979-8-3503-2445-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/BigData59044.2023.10386385

***Abstract*—Climate change and global warming are urgent environmental issues demanding immediate action to safeguard future generations. The major contributor to the greenhouse effect, carbon dioxide (CO_2), primarily originates from industrial and transportation fossil fuel combustion. International agreements, like the Paris Agreement, call for a 30-35% reduction in CO_2 emissions compared to 2005 levels. This research aims to predict CO_2 emissions and raise awareness among SAARC nations and governments about the increasing trend. We introduce a novel predictive framework using Attention-based Long Short-Term Memory (A-LSTM) for CO_2 emissions analysis. The Attention mechanism assigns variable weights to input data, facilitating indirect connections between LSTM outputs and pertinent inputs. This enhances resource allocation in the A-LSTM model, overcoming computational constraints. We integrate input parameters encompassing CO_2 emissions from land-use changes, oil, natural gas, and coal combustion to forecast CO_2 emissions and correlate them with population and per capita GDP. Our comparative analysis conclusively demonstrates the superior performance of A-LSTM models over baseline LSTM models when applied to the CO_2 emission dataset sourced from Our World in Data (OWID) and World Bank Indicator database. Specifically, the LSTM model registers a MAPE of 24.968 and an RMSE of 0.34, whereas the Attention-based LSTM model showcases a marked improvement of 57% with a considerably lower MAPE of 10.5902 and an RMSE of 0.107.**

***Index Terms*—Climate Change, Green House Gas, Long Short Term Memory, Attention, CO_2 Emission**

I. Introduction

A. Background

Similar to other emerging nations, the member states of the South Asian Association for Regional Cooperation (SAARC) are confronted with various challenges linked to climate change. Mitigating these challenges necessitates coordinated global efforts. Accurate predictions of CO_2 emissions are essential to combat climate change, guide policy-making, improve air quality, fulfil international commitments, support economic development, raise public awareness, and advance environmental research and education. Ensuring precise CO_2 emission forecasts is imperative for these nations to attain a sustainable and resilient future. In 2015, a worldwide summit on climate change took place in France, with participation from representatives of 196 nations, including the United States [1]. This gathering led to an accord to incorporate nationally determined contributions (NDCs) into the framework of the Paris Agreement.

Developing countries, characterised by significant landmass, population, and expanding economies, possess substantial potential for carbon reduction [2]. However, they require substantial financial and technical support to leverage their physical assets for a global cause, as emphasized in the Paris Agreement. Numerous economic factors, such as foreign investment, population growth, land use changes, economic growth (EG), energy consumption (EC), labour force (LF), urban population (UP), inflation (INF), tourism, and transportation contribute to carbon emissions. Many scholars have conducted extensive analyses of the literature on the relationships between CO_2 emissions, environmental factors, and energy usage [3]. Due to the heavy reliance of SAARC nations on coal, cement, land use alterations, and oil, a significant increase in carbon emissions from these countries is anticipated in the future. This paper evaluates carbon emissions and their Granger causality correlations with population changes, GDP, energy consumption, and income for each SAARC nation (Afghanistan, Bangladesh, Bhutan, India, Nepal, Pakistan, and Sri Lanka), excluding Maldives due to data limitations. Whenever time-series data is available and there is a requirement for forecasting future trends, machine learning presents itself as a valuable and effective tool.

Traditional ML approaches have inherent drawbacks, including prolonged processing times and often inadequate forecasting capabilities [14], [15]. Therefore, this work creates time-series forecast models based on robust and efficient ML algorithms. The outcome is the attention-based long short-term memory (A-LSTM) network, which integrates the attention mechanism into the long short-term memory network. This A-LSTM network addresses the issue of poor predictive power by offering speed and precision. Like other developing nations, South Asian Association for Regional Cooperation (SAARC) nations face a variety of climate change-related difficulties, and addressing those challenges will need coordinated global action. To combat climate change, inform policy making, enhance air quality, fulfil international obligations, assist economic development, increase public awareness, and advance environmental research and education, CO_2 emissions ought to be predicted. For the nations to have a sustainable and resilient future, accurate CO_2 emission forecasts are crucial. Significant opportunity for carbon reduction exists in emerging economies that are characterised by huge landmasses, sizable people, and growing economic activity. As emphasized in the Paris Agreement, these developing countries need a lot of financial and technical support in order to utilise their physical resources for a greater good [2].

A number of economic factors contribute to Carbon emission, including foreign investment, population growth, changes in land use, economic growth (EG), energy consumption (EC), labour force (LF), urban population (UP), inflation (INF), tourism, and transportation, among others. Numerous scholars have conducted analyses and evaluations of the body of literature concerning the correlation between CO_2 emissions, environmental variables, and various energy consumption patterns [5]. Machine learning can be used to tackle any problem involving time-series data and a need to foresee the future. Traditional and conventional ML approaches have the primary flaw that they take a lot of time and frequently have poor forecasting ability. In order to forecast production using historical data, time-series forecast models based on effective and powerful ML algorithms are created in this work. The attention-based long short-term memory (A-LSTM) network is the result of integrating the attention mechanism into the long short-term memory network. Since the A-LSTM network is quick and precise, it can address the issue of poor predicting power.

The paper is structured as follows: Section 2 outlines the research gap and presents the contributions made by this work. In Section 3, the methodology is detailed, which includes an in-depth description of the deep learning models used and the performance metrics employed. Section 4 elaborates on the experimental setup, covering model configurations, performance metrics, and the results of the experiments. Section 5 is dedicated to a discussion of the findings and offers insights into potential research directions. It also outlines the conclusions drawn from the study and discusses policy implications.

Numerous variables might affect CO_2 emissions, thus forecasting is important. Because of this, new models to predict CO_2 emissions have recently been developed and some are in the works. The study [4] underscores the pivotal role of feature engineering and selection in improving the accuracy of ML and deep learning algorithms for CO_2 emission prediction. The superior performance of hybrid deep learning algorithms, as evidenced by lower RMSE, rRMSE, and MAPE values, offers promising prospects for effectively addressing the challenges associated with CO_2 emission forecasting, ultimately contributing to our efforts to combat climate change. In the study titled "Predicting the Environmental Change of Carbon Emission Patterns in South Asia," researchers employed a Bi-LSTM deep learning approach. Their analysis revealed that carbon emissions pose a substantial concern, with projections indicating that the problem is likely to exacerbate in the coming years [5]. This escalation is attributed primarily to the high levels of emissions from China and India, especially within the next decade.

The CO_2 emissions in Bangladesh are predicted using the logarithmic mean division index (LMDI) decomposition method, and the results show that while globalisation, foreign direct investment, and innovation have a negative impact on CO_2 emissions in terms of improving environmental quality, economic growth, trade, energy consumption, and urbanisation have a positive impact on CO_2 emissions and thereby promote environmental degradation both in the long run and the short term [6].

To assess the impact of GDP and different forms of energy consumption—comprising total, non-renewable, renewable, industrial, and residential energy—on CO_2 emissions at the state level, various quantitative methodologies have been employed. These approaches encompass both static models and dynamic models [7].

To determine the CO_2 emission trends in Iran, Canada, and Italy, The results demonstrate that the proposed method is more accurate in long-term CO_2 emissions forecasting when using a general regression neural network (GRNN) and grey wolf optimisation (GWO) [8]. By using a linear and nonlinear time series modelling approach, specifically ARIMA, naive, TBATS, ETS, NNAR, and MLP, an accurate estimate of future CO_2 emissions in Pakistan was shown [9].

Four deep learning algorithms—Convolution neural network (CNN), CNN long short-term memory (CNN-LSTM), long short-term memory (LSTM), and dense neural network (DNN)—are examined for formulating the multivariate time series CO_2 emissions forecasting. A statistically significant long-term cointegrating relationship between CO_2 emissions, electrical energy consumption, and GDP is presented [10].

II. RESEARCH GAPS AND CONTRIBUTIONS

Numerous empirical inquiries have delved into multifaceted aspects of carbon dioxide (CO_2) emissions, encompassing examinations of the causal linkages between CO_2 emissions and their determinants. Nevertheless, these endeavours have left several research lacunae that necessitate further investigation. The impetuses for pursuing this research endeavour are grounded in several considerations. Firstly, limitations inherent in the scale of datasets employed have constrained the comprehensiveness of prior studies. Secondly, certain pivotal determinants of CO_2 emissions have been omitted from prior investigations. Lastly, the South Asian Association for Regional Cooperation (SAARC) nations have not received adequate scrutiny within this context. Additionally, the conventional machine learning models employed in this domain have confronted formidable challenges related to overfitting due to the scarcity of data points available for training and testing.

The prevailing literature has centered on time series forecasting in the domain of CO_2 emissions, leveraging machine learning, regression methodologies, and the extensive adoption of deep learning techniques. This research, in particular, focuses on evaluating the efficacy of Attention-based Long Short-Term Memory (A-LSTM) in forecasting, with the objective of addressing gaps in prior research efforts. This research contributes to the body of knowledge in the following ways:

- Shedding light on the low level of environmental awareness, particularly regarding CO_2 emissions, in SAARC countries. The study investigates the intricate dynamics and causal relationships between CO_2 emissions and cement production, illuminating critical insights.
- While similar indicators have been explored in previous studies, minimal attention has been devoted to analyzing SAARC data using machine learning techniques. This research endeavours to enrich the literature by applying deep learning methodologies to forecast CO_2 emissions, particularly within the unique context of SAARC nations.
- Significantly, this research emphasizes the importance of forecasting CO_2 emissions using A-LSTM, demonstrating its practical advantages in terms of enhanced accuracy compared to traditional LSTM models.

III. Methodology

A. Data and Variables

The data used in this research are collected from the year 1858 to 2021 from datasets in our world in data and world bank indicators of the world bank, which is the most comprehensive and includes all the sectors along with gases [13]. The major influencing carbon monitor categories considered in this work are Power (Oil Power plants, Coal Power Plants, Gas Power Plants), Industry(Oil combustion and use for Manufacturing, Gas combustion and use for Manufacturing, Coal combustion use for Manufacturing, Cement and other) including GDP per capita and population. The cumulative CO_2 emission data was collected from Kaggle. In this work, multivariate time series data of SAARC Countries (including Afganistan, Bangladesh, Bhutan, India, Nepal, Pakistan and Srilanka, excluding Maldives) have been used, having an increasing CO_2 emission trend. Table 1 depicts the descriptive analysis of the dataset CO_2 emission data. The upward trend in CO_2 levels becomes evident upon examining the minimum, median, and maximum values presented in Table I. This study primarily focuses on monitoring carbon emissions across critical categories such as Power, Industry, Residential Consumption, Ground Transport, and Land Use Change. Our analysis centres on understanding the growing trend in CO_2 emissions in relation to two key factors: GDP Per Capita and Population, within the SAARC countries. This approach

Variable	**Year**	**CO_2 Emission**
Mean	5347.84137548842	1944.35294117647
Standard Error	79.5731157532357	3.67292915087126
Median	3258.583479	1945
Standard Deviation	6630.40330990176	45.4316245333571
Sample Variance	43962248.0519562	2064.03250773994
Kurtosis	4.77203074137582	-1.08436364978744
Skewness	2.03347130343245	-0.0838680587457871
Range	40568.739095459	163
Minimum	1.630634541	1858
Maximum	40570.36973	2021
Sum	37130062.6700161	297486
Count	6943	153

TABLE I
Descriptive Analysis of Data

not only provides insights into both short-term and long-term dynamics but also helps uncover the causal connections between industrial development, energy consumption, and carbon emissions.

To gain a deeper understanding of the evolving relationships between these variables and to make more informed predictions about their future trends, we express them in a growth form. This form allows us to track changes over time relative to the previous period, offering valuable insights into potential future movements. In contrast, a straightforward ratio form only provides data for the current period.

Our research employs time series-based models to forecast CO_2 emissions for the upcoming years in all SAARC nations. The dataset exhibits stationarity with a clear upward trend. In this study, we implement advanced deep learning models, including LSTM and attention-based LSTM, tailored to the unique characteristics of this dataset.

B. Data Preprocessing

The datasets encompass 26 factors such as GDP, population, cement-related CO_2 emissions, gas production, and land use changes, along with per-capita CO_2 production trends over time. We conducted data cleaning by removing values with missing data points exceeding 50 per cent, primarily those before 1950. For the remaining missing data, we utilised interpolation and next-value fill methods from the pandas library. To enhance the data for LSTM training, we upsampled it and converted it into monthly data using the 'M' keyword within the pandas resample() library. This transformation ensured the data's consistency and suitability for LSTM modelling. To predict next year's per capita CO_2 production, we utilised data from the preceding three years, converting it into training samples. Furthermore, we partitioned the data by country, introducing a separation between adjacent data elements. This approach resulted in a multivariate time series dataset, on which we meticulously applied LSTM and LSTM, as elaborated in the subsequent sections. During our analysis of carbon emissions across SAARC countries, we observed a consistent upward trend in carbon emissions across all countries within the SAARC region. However, it's important to note that these emission trends are not uniform across the region, which poses a significant challenge in developing a suitable model for joint

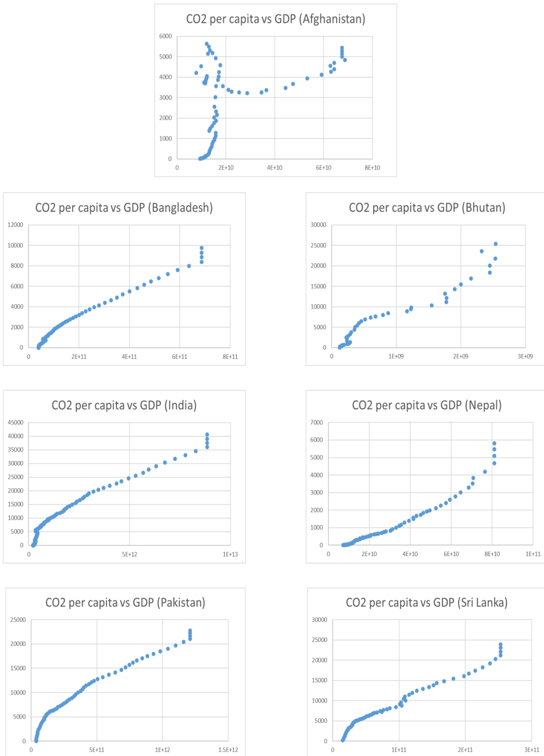


Fig. 1. Country wise GDP vs CO_2 emission per capita

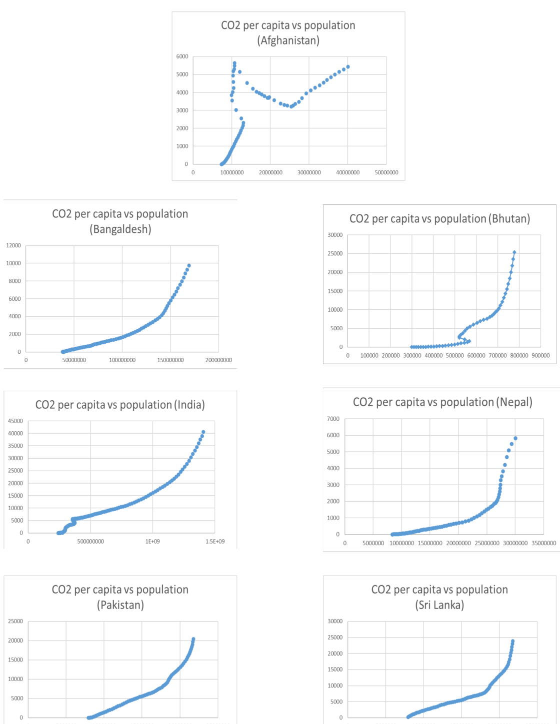


Fig. 2. Country wise Population vs CO_2 emission per capita

analysis and forecasting of carbon emissions for these diverse nations.

Figures 1 and 2 depict the correlation between carbon emissions and GDP per capita as well as population growth for the seven major SAARC countries.

Table 1 offers a descriptive analysis of the preprocessed CO_2 emission dataset. By examining the minimum, maximum, and median values, it becomes evident that carbon emissions are consistently increasing.

C. Long Short Term Memory

Recurrent neural networks (RNNs) are a class of artificial neural networks that are particularly good at processing time series data because the nodes between the hidden layers are connected. Although RNN may potentially capture long-distance dependence, when the time step is increased, it is simple to produce gradient explosion or disappearance when the backpropagation method updates the weights. Gradient explosion may hinder network convergence during training and may potentially result in network failure, and the gradient's removal will make it more challenging for networks to learn long-distance dependence. Gradient clipping can prevent gradient explosion, which is a pretty simple problem to manage. The key to understanding RNN and nearly all other deep learning techniques is understanding how to minimise gradient vanishing. Hochreiter and Schmidhuber (1997) first proposed the LSTM, and Alex Graves most recently made improvements to it. The gradient disappearance issue is resolved by LSTM by creating a complex network structure. Figure 3 illustrates the

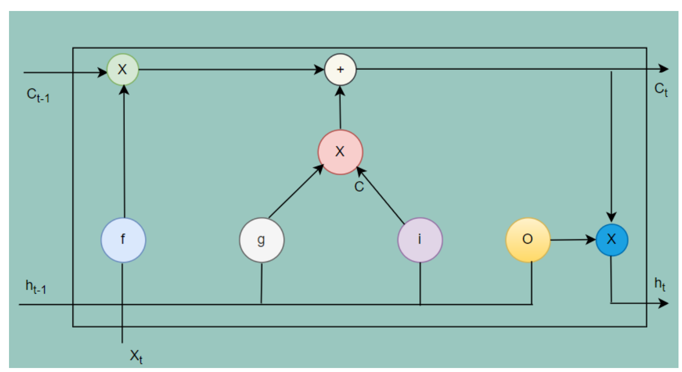


Fig. 3. LSTM architecture

fundamental architectural arrangement of an LSTM. Within the LSTM cell, it receives inputs: the current variable x_t, the preceding output $h_t - 1$, and the prior cell state $C_t - 1$. The components, namely the input gate (i_t), forget gate (f_t), output gate (o_t), and memory cell (C_t), are represented as small boxes [3]. These components are computed utilizing equations 1 through 5.

$$f_t = \sigma_g \left(W_f \cdot x_t + U_f \cdot h_t - 1 + b_f \right) \tag{1}$$

$$i_t = \sigma_g \left(W_i \cdot x_t + U_i \cdot h_t - 1 + b_i \right) \tag{2}$$

$$o_t = \sigma_g \left(W_o \cdot x_t + U_o \cdot h_t - 1 + b_o \right) \tag{3}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot x_t + U_C \cdot h_t - 1 + b_C\right) \quad (4)$$

In this context, the symbol $\cdot$ signifies the matrix multiplication operation, while b_f, b_i, b_o, and b_C correspond to four bias vectors. The weight matrices U_f, U_i, U_o, and U_C establish connections between the previous output and the three gates as well as the memory cell. Similarly, the weight matrices W_f, W_i, W_o, and W_C are associated with these gates. The gate activation function denoted as σ_g, is specifically a sigmoid function, and the function tanh() refers to the hyperbolic tangent function, as indicated in reference [3] and demonstrated in equations 5 and 6.

$$C_t = (f_t \oplus C_{t-1}) + i_t \oplus \tilde{C}_t \quad (5)$$

$$h_t = o_t \oplus \tanh\left(C_t\right) \quad (6)$$

where $\oplus$ denotes the element-wise matrix/vector, multiplication operator.

D. Proposed Framework

The task of predicting CO_2 emissions is a formidable challenge that engages climate change experts and meteorological departments worldwide. In recent years, a multitude of strategies has evolved to anticipate CO_2 output, and our research framework, depicted in Figure 4, encapsulates these endeavours. We commence with data collection from sources such as OWID and World Bank indicators, followed by a meticulous data preprocessing phase. This encompasses handling missing values, selecting relevant features, normalizing data, and implementing resampling techniques. Subsequently, we partition the dataset into an 80:20 split for training and testing purposes. The heart of our study lies in the development of predictive models, employing both LSTM and Attention-based LSTM architectures. These models undergo rigorous training and fine-tuning, culminating in a meticulous assessment of their performance through a range of performance evaluation metrics. Our overarching goal is to leverage the predictive capabilities of these models to forecast CO_2 emissions, an imperative facet of environmental research with far-reaching implications.

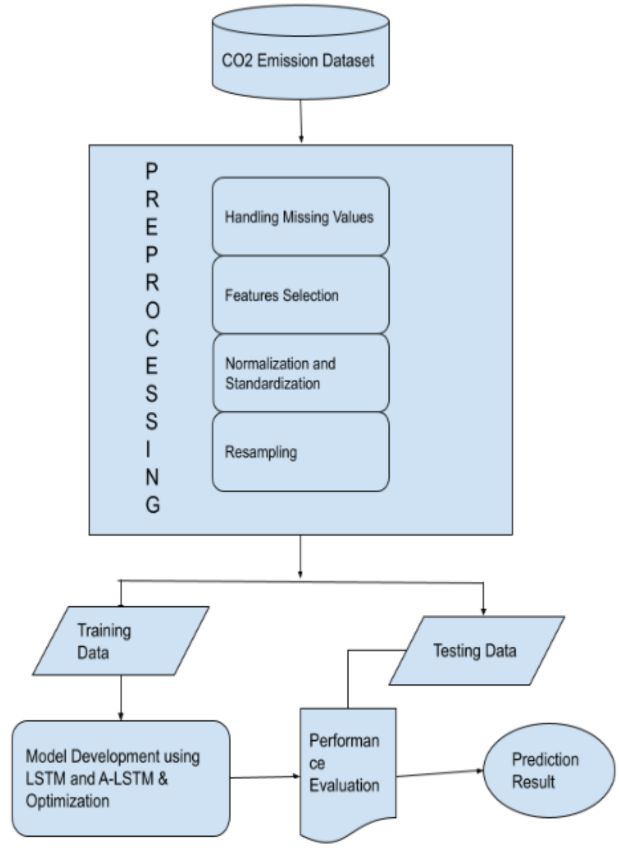


Fig. 4. Framework for CO_2 emission forecasting

E. Proposed Model

Human perception is unique in that it prioritises the relevant aspects of the environment in order to get the information they require rather than dealing with all external stimuli at once. Likewise, within financial markets, the significance of different trading information varies significantly. While certain information may be redundant, other data points could be of paramount importance [11]. In order to anticipate CO_2 emissions, we also need to prioritise important features over less important ones. ”Hence, we propose an Attention Mechanism-based Long Short-Term Memory model as a predictive tool capable of synthesizing and utilizing valuable information from multiple temporal contexts to inform decision-making.” The data suggests using an attention model to improve input sequences for CO_2 emission predictions. This mechanism maps queries and key-value pairs to outputs, all represented as vectors. [11]. The result is calculated as a weighted sum of the values, with each value’s weight determined by the query’s compatibility function with its corresponding key. The attention weight generation and attention-based new input features are summarised in Figures 5 and 6. In the initial step, we establish a mapping from x_t to f_t as follows: $\mathbf{h}_t = \mathbf{f}_1(\mathbf{f}_{t-1}, \mathbf{x}_t)$ Here, $\mathbf{f}_t$ represents a non-linear activation function, and $\mathbf{h}_t \in \mathbf{R}^s$ denotes the hidden state at time t, where ’s’ signifies the size of the hidden state. In this context, we employ LSTM as the function $\mathbf{f}_1$ to mitigate the long-term dependence issue often encountered in time series prediction. Subsequently, we construct an attention mechanism utilizing a deterministic attention model. For a given feature sequence $\mathbf{x}^k = (x_1^k, x_2^k, \ldots, x_m^k)^{\mathrm{T}} \in \mathbf{R}^m$, we define the input feature sequence using equation 7, while the attention weight is established in equation 8 by referencing the preceding hidden state $\mathbf{h}_{t-1}$ and cell state $\mathbf{C}_{t-1}$ within the LSTM unit.

$$\alpha_t^k = \mathbf{v}^{\mathrm{T}} \tanh\left(\mathbf{W}_1 \cdot [\mathbf{h}_{t-1}, \mathbf{C}_{t-1}] + \mathbf{W}_2 \mathbf{x}^k\right) \quad (7)$$

$$\beta_t^k = \operatorname{softmax}\left(\alpha_t^k\right) = \frac{\exp\left(\alpha_t^k\right)}{\sum_{i=1}^n \exp\left(\alpha_t^k\right)} \quad (8)$$

”The vector $\mathbf{v}$ and the matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are adjustable parameters within the model. The vector $\boldsymbol{\alpha}^k$ consists of ’m’ elements, where each ’i’-th element quantifies the significance

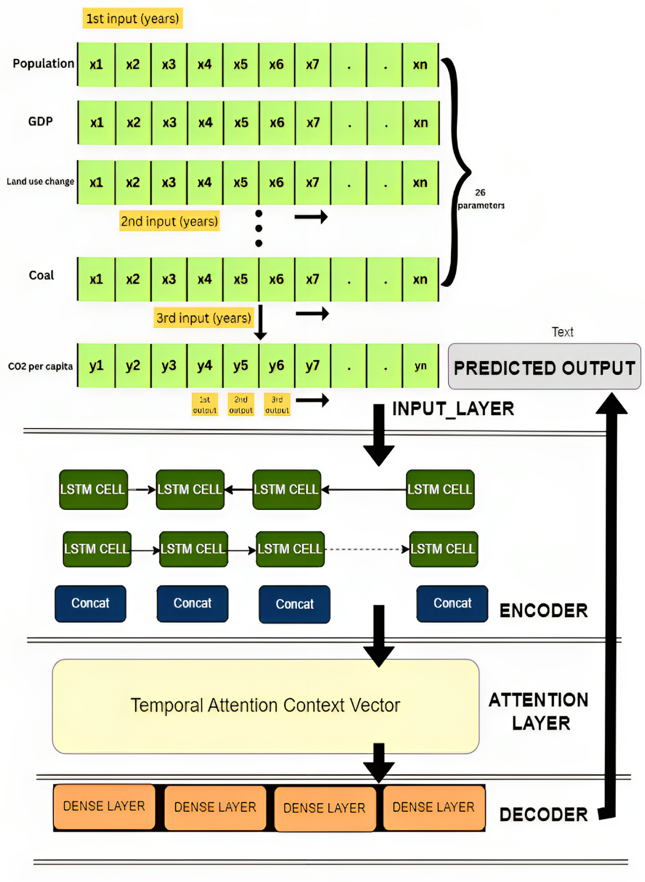


Fig. 5. Proposed Attention-based Long Short Term Memory

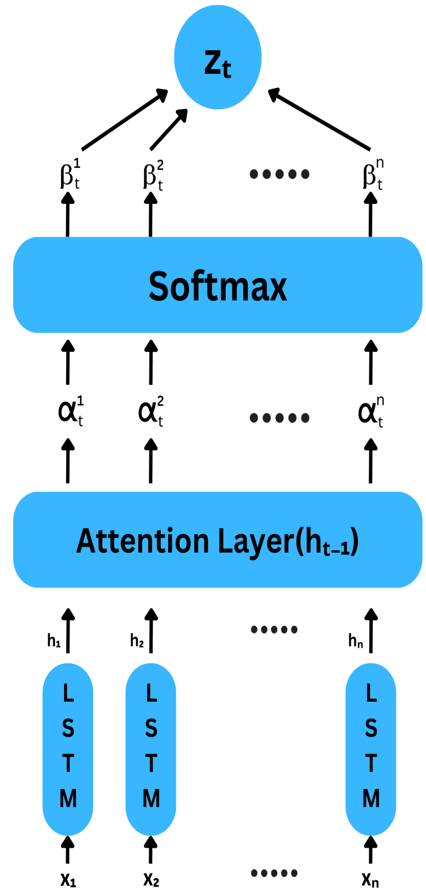


Fig. 6. Attention-based Long Short Term Memory

of the 'k'-th input feature sequence at time 't'. These values are normalised using softmax [12]. Additionally, β^k serves as an attention weight, encapsulating a score that determines the degree of attention directed towards the 'k'-th feature sequence. Equation 9 provides the expression for the weighted input feature sequence $\mathbf{z}_t$, at time 't,' which represents the output of the attention model.

$$\mathbf{z}_t = \left(\beta_t^1 x_t^1, \beta_t^2 x_t^2, \ldots, \beta_t^n x_t^n\right)^{\mathrm{T}} \tag{9}$$

Now $\mathbf{x}_t$ is replaced by the newly computed $\mathbf{z}_t$ to update the attention model and finally raw time series is converted into attention-based time series.

F. Performance metrics

In this section, the performance metrics are used to determine the effectiveness of the proposed model. There are many evaluation metrics; we have used four to evaluate the models' effectiveness. Training and test sets of the data are created where the Training set (80 per cent of the data set) was used to train all models and the remaining 20 per cent was used to test the model. The MSE, RMSE, MSLE, MAPE and R-2 score values for the training and test sets have been assessed. We have also showcased a comparative analysis between the performance evaluation metrics of the proposed A-LSTM model and those of the baseline LSTM model.

1) Mean Absolute Error (MAE): MAE is the average of the absolute difference between forecasted values and true values. The MAE shows us how much inaccuracy we should expect from the forecast on average. The model with the lowest MAE is superior when comparing many models. MAE can be calculated by using equation 10.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} \mid y_i - \hat{y}_i \mid}{n} \tag{10}$$

Where $\hat{y}$ is the predicted value and y_i is the actual value. The letter n represents the total number of values in the test set.

2) Mean Squared Error (MSE): Mean squared error is the average of the square of the difference between the predicted values and true values. It has the same units as the true and predicted values squared and is always positive. MSE can be calculated by using Equation 11.

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{11}$$

Where $\hat{y}$ is the predicted value, y_i is the actual value, and n is the total number of values in the data sets.

3) Mean Absolute Percentage Error(MAPE): The calculation of MAPE (Mean Absolute Percentage Error) plays a crucial role in enhancing the precision of forecasts for future projects. When a MAPE calculation indicates that the existing forecasting is inaccurate, it prompts the possibility of revising strategies or even adopting entirely new methods. Furthermore, MAPE calculations are valuable for pinpointing when forecasting is accurate, enabling more effective communication of forecasting results and the implementation of sound strategies. MAPE can be calculated by using equation 12.

$$\text{MPE}(y,\hat{y}) = \frac{100\%}{N}\sum_{i=0}^{N-1}\frac{y_i-\hat{y}_i}{y_i}. \tag{12}$$

4) Root Mean Square Error(RMSE): RMSE (Root Mean Square Error) is a widely utilised metric in fields such as climatology, forecasting, and regression analysis to validate experimental findings. It quantifies the standard deviation of prediction errors or residuals. These residuals represent the extent to which data points deviate from the regression line. RMSE, in essence, provides insight into how these residuals are distributed, indicating how closely the data points cluster around the line of best fit. RMSE can be calculated by using equation 13.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i-o_i)^2} \tag{13}$$

IV. Experiments and Results

A. Implementation Architecture

The experiments are carried out in Jupyter Notebook using Python 3.7.13. Open-source libraries like Pandas 2.1.1, Numpy 1.19.3, Keras and Tensorflow 2.0+ are used for experimentation. The experimental setup is based on a working environment having a Ryzen 7 5800-H processor with 16 GB RAM under a 64-bit Windows 11 Operating system. Time series forecasting of the CO_2 emissions dataset is modelled using two deep learning models, namely, LSTM and LSTM with Attention (referred to as A-LSTM). The optimum results are considered from each model's ten unique runs. The chosen architecture of both the models is briefed as follows:

The model architecture relies on LSTM units for sequence modelling. It consists of two fully connected layers, each with 64 neurons, followed by a dense layer. The ReLU activation function is used for each layer. Optimisation is done using ADAM with a learning rate of 1e-4 and a decay of 1e6. Training employs the MAE loss metric and a 20% validation split. To combat overfitting, a 0.1 dropout rate is applied after the second fully connected layer. Batch size 32 is found to be optimal for performance and robustness.

In the proposed A-LSTM model, we introduce a 64-unit attention layer between the LSTM layer and the dense layer. Hence, we have a 64-neuron LSTM layer followed by the attention layer and the dense layer. The learning rate, loss metric, optimiser, decay rate and batch size are the same as LSTM for a comparative study.

B. Performance Analysis

Table II displays the experimental outcomes achieved through the application of both the LSTM model and the proposed A-LSTM model. The results unmistakably indicate substantial enhancements in the proposed model compared to the baseline. Table III illustrates a comparison between the experimental results obtained and those acquired in [5]. The findings indicate that A-LSTM surpasses LSTM, GRU, and Bi-LSTM in both Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE).

	MAE	**MSE**	**MAPE**	**RMSE**
LSTM	0.227	0.0246	24.968	0.34
A-LSTM	**0.056**	**0.0115**	**10.5902**	**0.107**

TABLE II
EXPERIMENTAL RESULTS

Fine-tuning was performed on both LSTM and A-LSTM models, given that their performance is highly dependent on parameter adjustments. Particular attention has been given to calibrating the parameters of the LSTM forecasting model due to its more intricate structure compared to other deep learning models. We manually adjusted all hyperparameters through trial and error, optimizing the number of hidden layers and neurons for both deep learning models to achieve the best results. Across all models, a learning rate of 0.0001 consistently yields higher accuracy. To prevent over-fitting and enhance model efficiency, we've implemented an early stopping strategy.

Models	**Performance Parameters**	
	MSE	**MAPE**
LSTM	0.0246	24.968
GRU [5]	0.027	26.11
Bi-LSTM [5]	0.0242	24.72
A-LSTM [proposed]	**0.0115**	**10.5902**

TABLE III
COMPARISON OF EXPERIMENTAL RESULTS

We gave importance to dropout due to the over-fitting nature of the data. We tried different dropout settings in the LSTM architecture. We found out that a dropout of 0.1 after the second layer was the optimum setting with the loss metric as MAE as shown in Table 3. Hence, we used this in our final model architecture. In comparing the performance of the LSTM and A-LSTM models, we observed significant differences across multiple evaluation metrics. The LSTM model exhibited a Mean Absolute Error (MAE) of 0.227, a Mean Squared Error (MSE) of 0.0246, a Mean Absolute

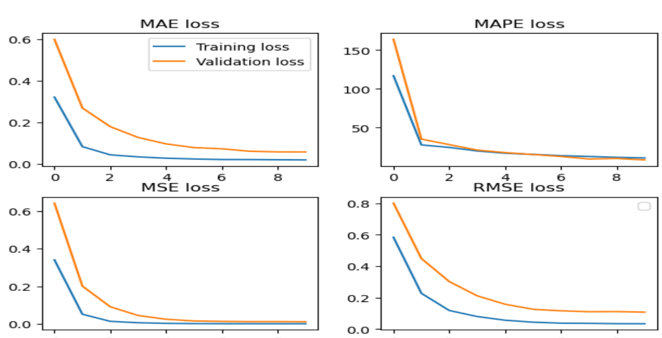
Dropout configuration	**Training MAE**	**Validation MAE**
Two dropouts (0.6 each)	0.148	0.244
Two dropouts (0.2 each)	0.1	0.23
Dropout of 0.2 after first layer	0.03	0.31
Dropout of 0.1 after first layer	0.06	0.27
Dropout of 0.1 after second layer	0.05	0.24
No Dropout	0.01	0.32

TABLE IV
LOSS METRIC (MAE) EVALUATION IN DIFFERENT DROPOUTS

Percentage Error (MAPE) of 43.6, and a Root Mean Squared Error (RMSE) of 0.34. In contrast, the A-LSTM model displayed notably superior results, with a substantially lower MAE of 0.056, a significantly reduced MSE of 0.0115, a remarkably improved MAPE of 10.5902, and a substantially lower RMSE of 0.107. These metrics collectively demonstrate that the A-LSTM model outperforms the LSTM model in terms of accuracy and precision, showcasing its superior predictive capabilities in forecasting CO_2 per capita across various countries over multiple years.

In Figure 7, 8, we presented the validation and training loss values for both LSTM and A-LSTM models. Furthermore, we evaluate the predictive proficiency of the models by analyzing how effectively they forecast CO_2 emissions per capita across the countries specified in the dataset over various years. The superior performance of the A-LSTM model is evident in the results illustrated in Figure 8,9. Furthermore, we provide a comprehensive comparison of the training and testing loss for both the LSTM and A-LSTM models in Figure 9,10 to evaluate their overall performance. Figure 11 displays the final predicted CO_2 emission outcomes for each of the seven SAARC nations.

Fig. 7. LSTM Loss



c

Fig. 8. A-LSTM LOSS

V. Conclusion and Future research directions

In conclusion, our study delved into the critical task of forecasting carbon dioxide (CO_2) emissions, an imperative endeavour to address the pressing issue of climate change and global

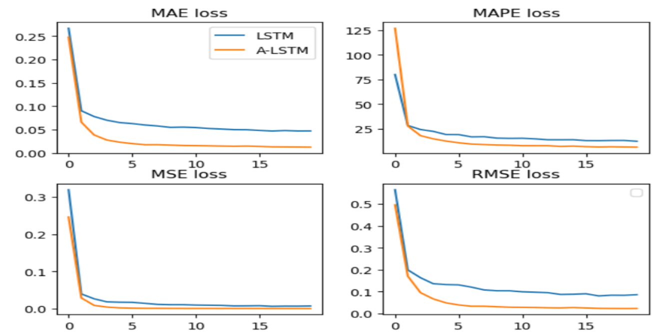


Fig. 9. Training loss comparison of LSTM and A-LSTM

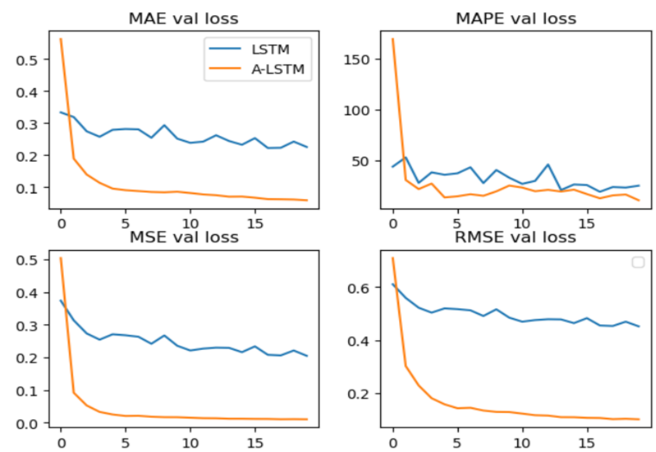


Fig. 10. Validation loss comparison of LSTM and A-LSTM

warming. With international commitments calling for substantial reductions in CO_2 emissions, our research introduced an innovative predictive framework leveraging an Attention-based Long Short-Term Memory (A-LSTM) architecture. This A-LSTM model, which incorporates an attention mechanism to allocate computational resources efficiently, outperformed the baseline LSTM model across various evaluation metrics. The LSTM model had limitations, with a MAPE of 24.968 and an RMSE of 0.34, while the A-LSTM model demonstrated superior performance with a lower MAPE of 10.5902 and RMSE of 0.107 with 57% improvement in MAPE loss. These results highlight a notable enhancement in predictive accuracy when compared with the existing models LSTM, GRU and Bi-LSTM [5], accomplished through the application of attention mechanisms within the context of the carbon emission dataset examined in this research. This outcome bears noteworthy implications and provides valuable insights, particularly for a diverse set of predictive modelling applications, with a strong focus on forecasting CO_2 emissions and understanding their impact on sustainable development.

While our study primarily centres on SAARC countries, its implications extend far beyond this region. The discoveries made herein can serve as a viable blueprint for reshaping energy consumption practices globally, ultimately leading to

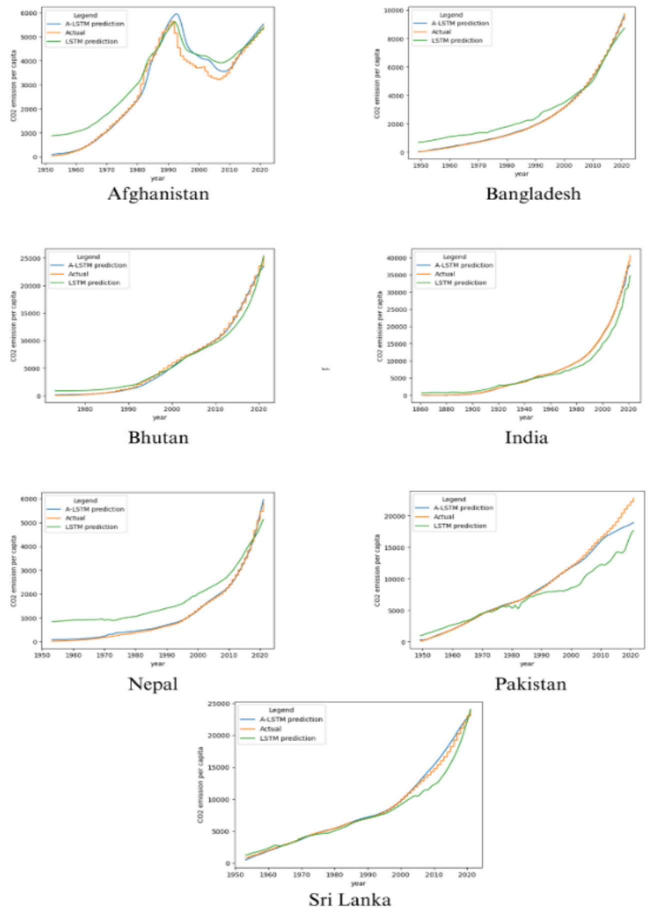


Fig. 11. Country wise prediction of CO_2 via LSTM and A-LSTM

a reduction in CO_2 emissions worldwide. In the future, this study has the potential for expansion, exploring the benefits and practical applications of quantum-enhanced LSTM and A-LSTM models for multivariate time series prediction.

References

[1] Morgan, Jamie. "Paris COP 21: Power that speaks the truth?." Globalizations 13.6 (2016): 943-951.
[2] Latief, Rashid, et al. "Carbon emissions in the SAARC countries with causal effects of FDI, economic growth and other economic factors: Evidence from dynamic simultaneous equation models." International Journal of Environmental Research and Public Health 18.9 (2021): 4605.
[3] Faruque, Md Omer, et al. "A comparative analysis to forecast carbon dioxide emissions." Energy Reports 8 (2022): 8046-8060.
[4] Nguyen, Van Giao, et al. "An extensive investigation on leveraging machine learning techniques for high-precision predictive modeling of CO_2 emission." Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 45.3 (2023): 9149-9177.
[5] Aamir, Muhammad, et al. "Predicting the Environmental Change of Carbon Emission Patterns in South Asia: A Deep Learning Approach Using BiLSTM." Atmosphere 13.12 (2022): 2011.
[6] Islam, Md Monirul, et al. "Impact of globalization, foreign direct investment, and energy consumption on CO_2 emissions in Bangladesh: Does institutional quality matter?." Environmental Science and Pollution Research 28.35 (2021): 48851-48871.
[7] Salari, Mahmoud, Roxana J. Javid, and Hamid Noghanibehambari. "The nexus between CO_2 emissions, energy consumption, and economic growth in the US." Economic Analysis and Policy 69 (2021): 182-194.
[8] Heydari, Azim, et al. "Renewable energies generation and carbon dioxide emission forecasting in microgrids and national grids using GRNN-GWO methodology." Energy Procedia 159 (2019): 154-159.
[9] Tawiah, Kassim, Muhammad Daniyal, and Moiz Qureshi. "Pakistan CO_2 Emission Modelling and Forecasting: A Linear and Nonlinear Time Series Approach." Journal of Environmental and Public Health 2023 (2023).
[10] Faruque, Md Omer, et al. "A comparative analysis to forecast carbon dioxide emissions." Energy Reports 8 (2022): 8046-8060.
[11] Zhang, Xuan, et al. "At-lstm: An attention-based lstm model for financial time series prediction." IOP Conference Series: Materials Science and Engineering. Vol. 569. No. 5. IOP Publishing, 2019.
[12] Kim, Yoon, et al. "Structured attention networks." *arXiv preprint arXiv:1702.00887* (2017).
[13] Ritchie, Hannah, Max Roser, and Pablo Rosado. "CO_2 and greenhouse gas emissions." Our world in data (2020).
[14] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[15] Rosas-Romero, Roberto, Alejandro Díaz-Torres, and Gibran Etcheverry. "Forecasting of stock return prices with sparse representation of financial time series over redundant dictionaries." *Expert Systems with Applications* 57 (2016): 37-48.