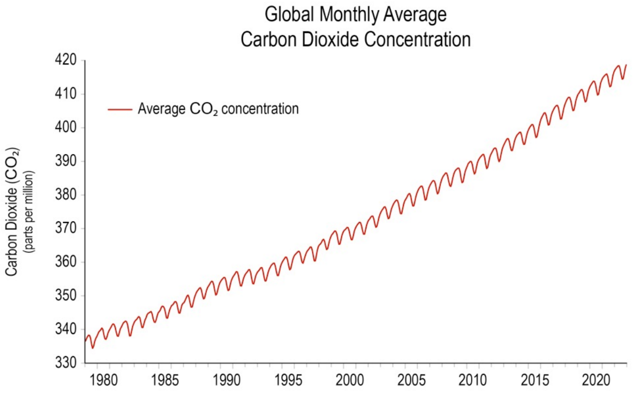
Improving Power Plant CO2 Emission Estimation with Deep Learning and Satellite/Simulated Data

Dibyabha Deb (Manipal Institute of Technology, Bengaluru),
Kamal Das (IBM Research, India)

NEURAL INFORMATION PROCESSING SYSTEMS

Background and Motivation

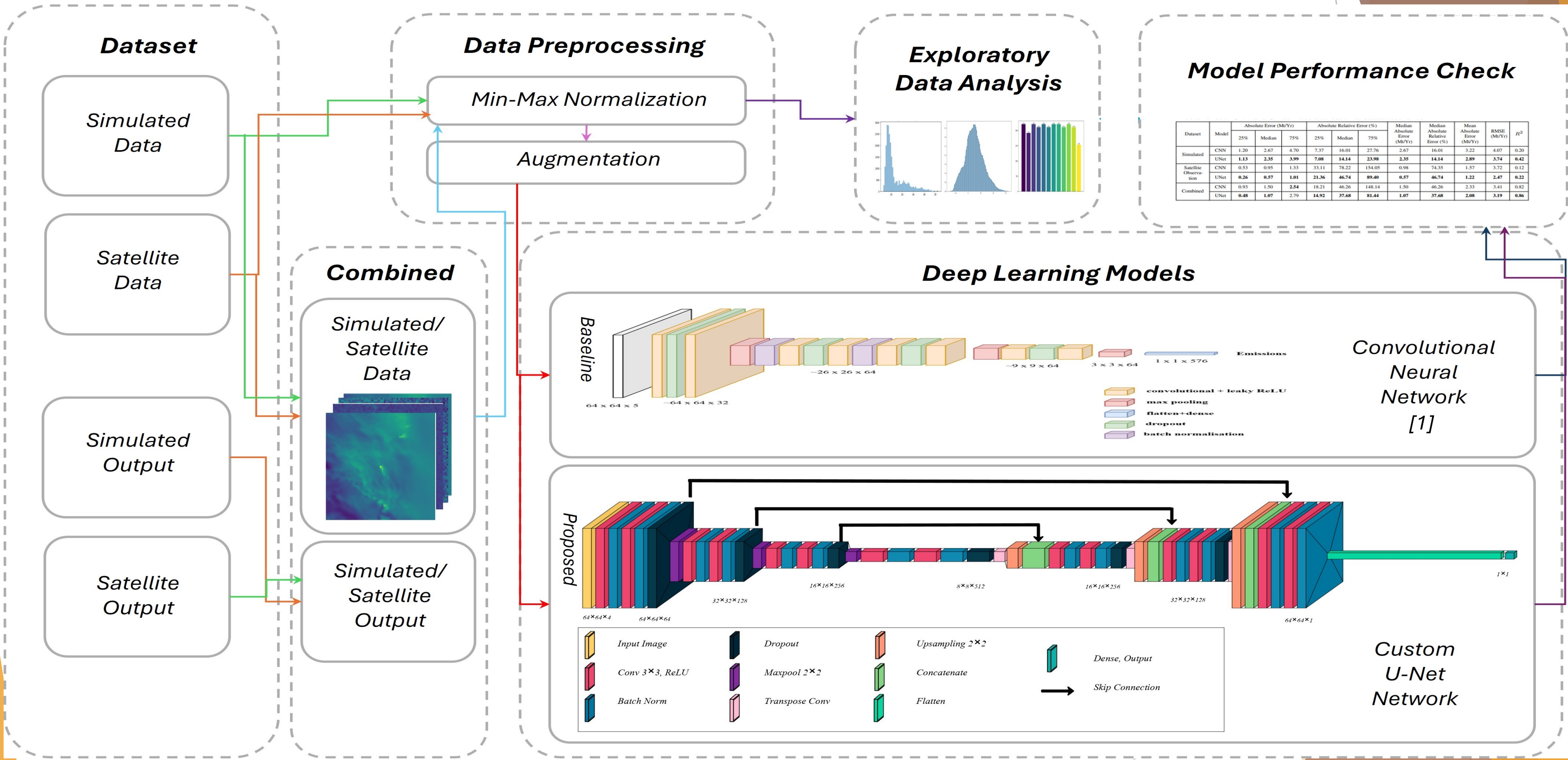
- Greenhouse gas emissions (mainly CO2) are a major driver of climate change, and therefore, they represent the cause of all its adverse effects, namely extreme weather events, wildfires, droughts, sea level rise, etc.
- To control the amount of greenhouse gas emissions, and therefore to mitigate the effect of climate change, emission trading systems have been implemented that aim to provide economic incentives for reducing the emission of pollutants.
- ESG reporting is no longer voluntary. Regulations are getting tighter across the globe.
- Investors are now extremely sensitive to ESG performance of the investee companies
- Many enterprises committed to to be net-zero by 2030
- Lack of emission data and use of average emission factor is one of the bottleneck of ESG reporting
- Average emission data is not up to date and has a typical lag of 2-3 years
- The need of the hour is to timely and accurately estimate emissions for emission intensive industries such as power plants.



NOAA Global Monitoring Laboratory

"In early 2025, countries must deliver new nationally determined contributions" – COP28

Overview of the proposed methodology and workflow



[1]: Joffrey Dumont Le Brazidec et al. "Deep learning applied to CO2 power plant emissions quantification using simulated satellite images" Geoscientific Model Development 17.5 (2024), pp. 1995–2014

Overview

- **Dataset**
  Simulated Data from SMARTCARB (We considered Lippendorf and Boxberg dataset)
  Satellite Data from Kingdom of Saudi Arabia (KSA) region [2]
- **Approach**
  Simulated Data and Satellite Data are normalized using min-max normalization
  Both these datasets are combined to form a single dataset
  In total, three datasets – Simulated, Satellite and Combined
- **Exploratory Data Analysis (EDA)**
  Our primary focus : Distribution of the data in all 3 Cases
- **Model Development and Performance**
  Both models evaluated on all datasets

[2] Ali Hamieh et al. "Quantification and analysis of CO2 footprint from industrial facilities in Saudi Arabia". In: Energy Conversion and Management: X 16 (2022), p. 100299.

Distribution of Datasets

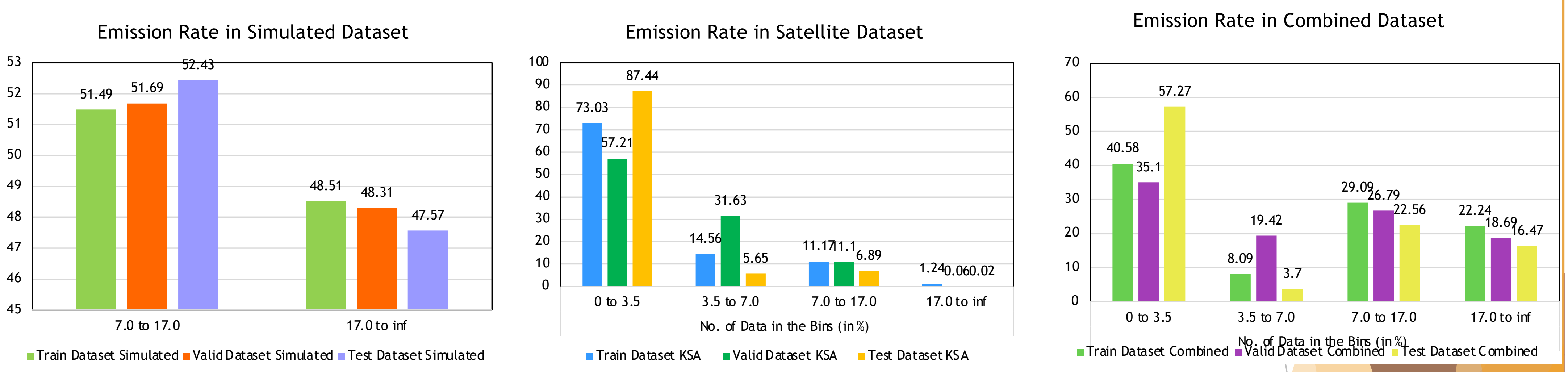
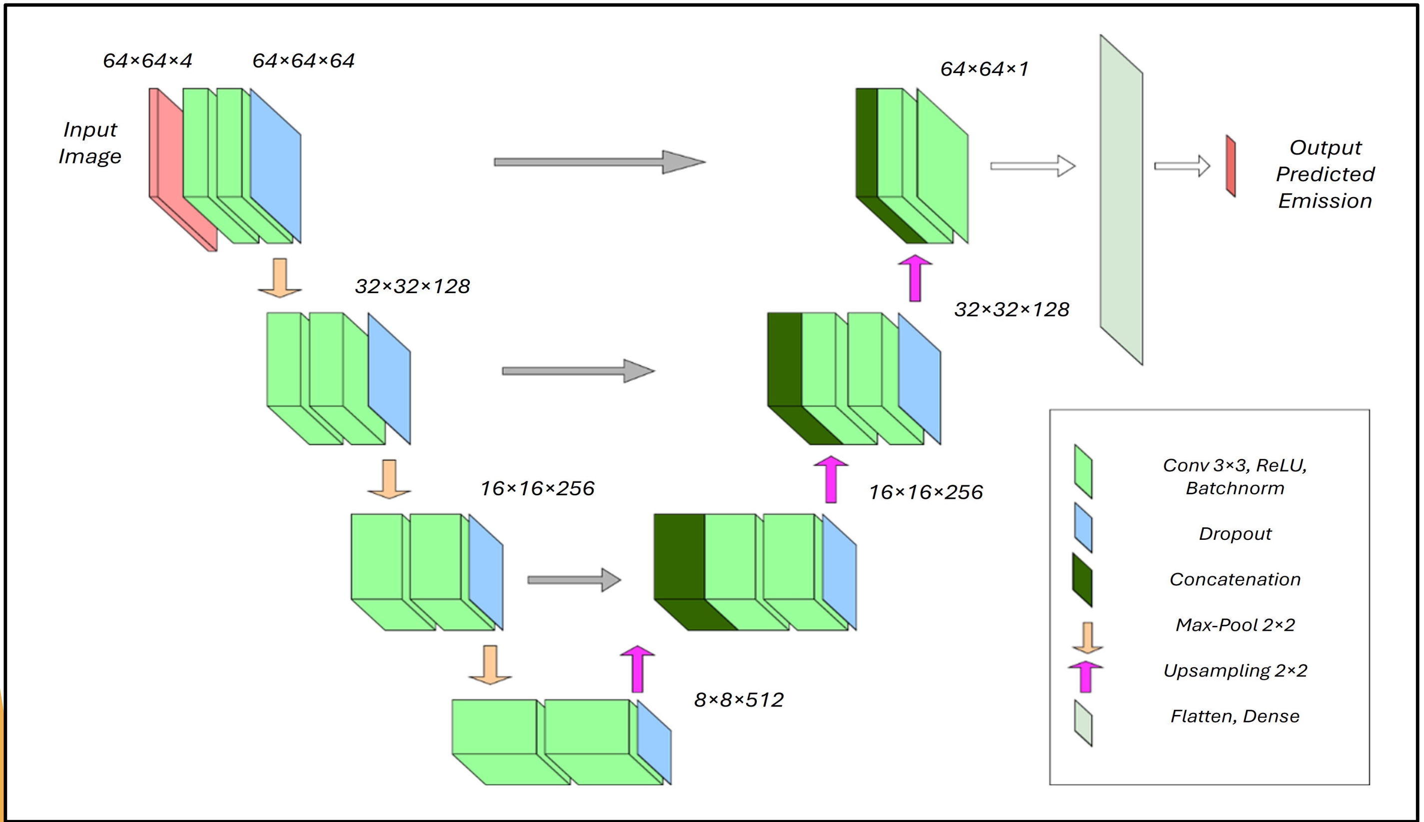


Chart 1: Emission Rate Distribution graph of Simulated, Satellite and Combined Datasets

- In Simulated dataset, emission data is not present from 0 to 7 Mt/yr
- Further, in Satellite dataset the distribution is concentrated towards 0 to 3.5 Mt/yr
- A better distribution of data is observed in the Combined dataset
- This solved two major challenges that we faced:
  - (1) No distribution of data in 0 – 7 Mt/Yr
  - (2) Concentrated distribution in few bins

Proposed Model for Emission Rate Estimation



Simplified Figure of the U-Net regression model

- Traditionally, U-Net is used for segmentation tasks
- Adoption of U-Net for regression task
- Unique approach: U-Net modelling technique for GHG emission rate estimation
- Considered four inputs: XCO2, NO2, u and v

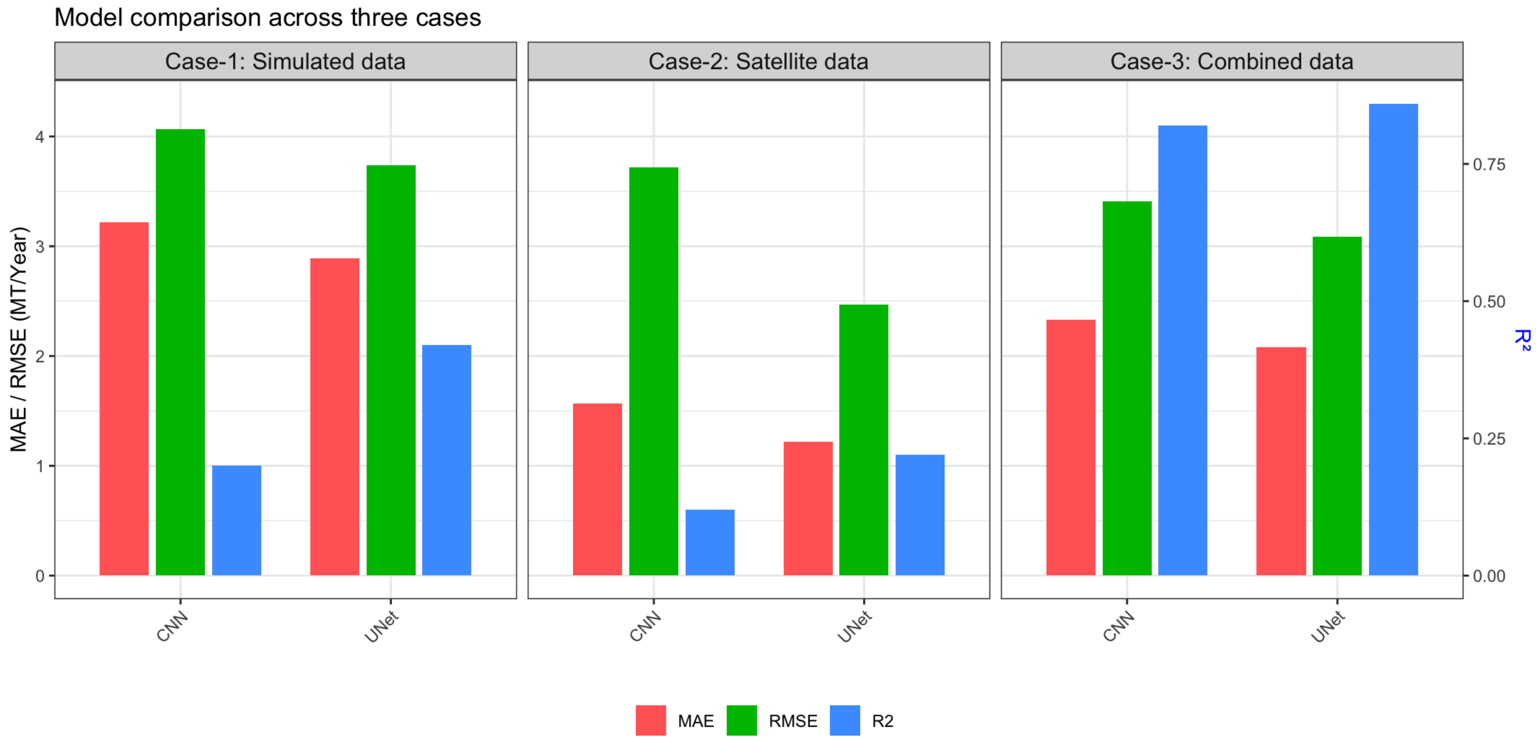
Model Performance Assessment

- Evaluated both models in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and R2 values
- Checked different percentile values in Absolute Error and Relative Absolute Error for both the models

Results from both the models evaluated on all datasets

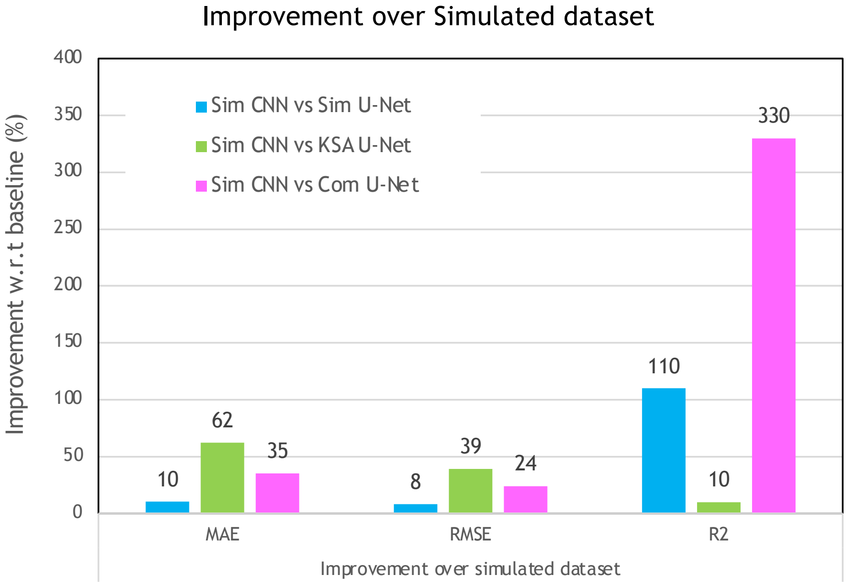
Dataset	Model	Absolute Error (Mt/Yr)			Absolute Relative Error (%)			Median Absolute Error (Mt/Yr)	Median Absolute Relative Error (%)	Mean Absolute Error (Mt/Yr)	RMSE (Mt/Yr)	R^2
		25%	Median	75%	25%	Median	75%					
Simulated Data	CNN	1.20	2.67	4.70	7.37	16.01	27.76	2.67	16.01	3.22	4.07	0.20
	U-Net	**1.13**	**2.35**	**3.99**	**7.08**	**14.14**	**23.98**	**2.35**	**14.14**	**2.89**	**3.74**	**0.42**
Satellite Data	CNN	0.53	0.95	1.33	33.11	78.22	154.05	0.98	74.35	1.57	3.72	0.12
	U-Net	**0.26**	**0.57**	**1.01**	**21.36**	**46.74**	**89.40**	**0.57**	**46.74**	**1.22**	**2.47**	**0.22**
Combined Data	CNN	0.93	1.50	**2.54**	18.21	46.26	148.14	1.50	46.26	2.33	3.41	0.82
	U-Net	**0.48**	**1.07**	2.79	**14.92**	**37.68**	**81.44**	**1.07**	**37.68**	**2.08**	**3.19**	**0.86**

Model Performance Assessment (continued…)



- In terms of MAE, U-Net model did much better in all the 3 datasets
- Similarly, in terms of RMSE the U-Net performed better than the baseline model
- A drastic increase in R2 value is also observed in combined dataset

Summary and Future work



- Results demonstrate that the U-Net model consistently outperforms the baseline CNN model across all datasets
- In Simulated dataset, U-Net achieved an improvement of 10%, 8% and 110% in MAE, RMSE and R2 values over baseline CNN model
- In Satellite dataset, U-Net exhibit superior performance in terms of MAE and RMSE, 1.22 and 2.47 respectively, although the R2 values are lower than those obtained with simulated data
- Finally, in the Combined dataset the U-Net model achieved a 330% improvement over the original benchmark value in R2 when compared to the best result obtained with individual datasets
- Combine more datasets and use different other encoders