Imperial College London

# Short-Term Rainfall Cell Forecasting using Machine Learning Techniques

**Yen Hann Yoo**

**Supervisor: Dr. Christian Onof**

*EWRE Section, Department of Civil and Environmental Engineering, Imperial College London*

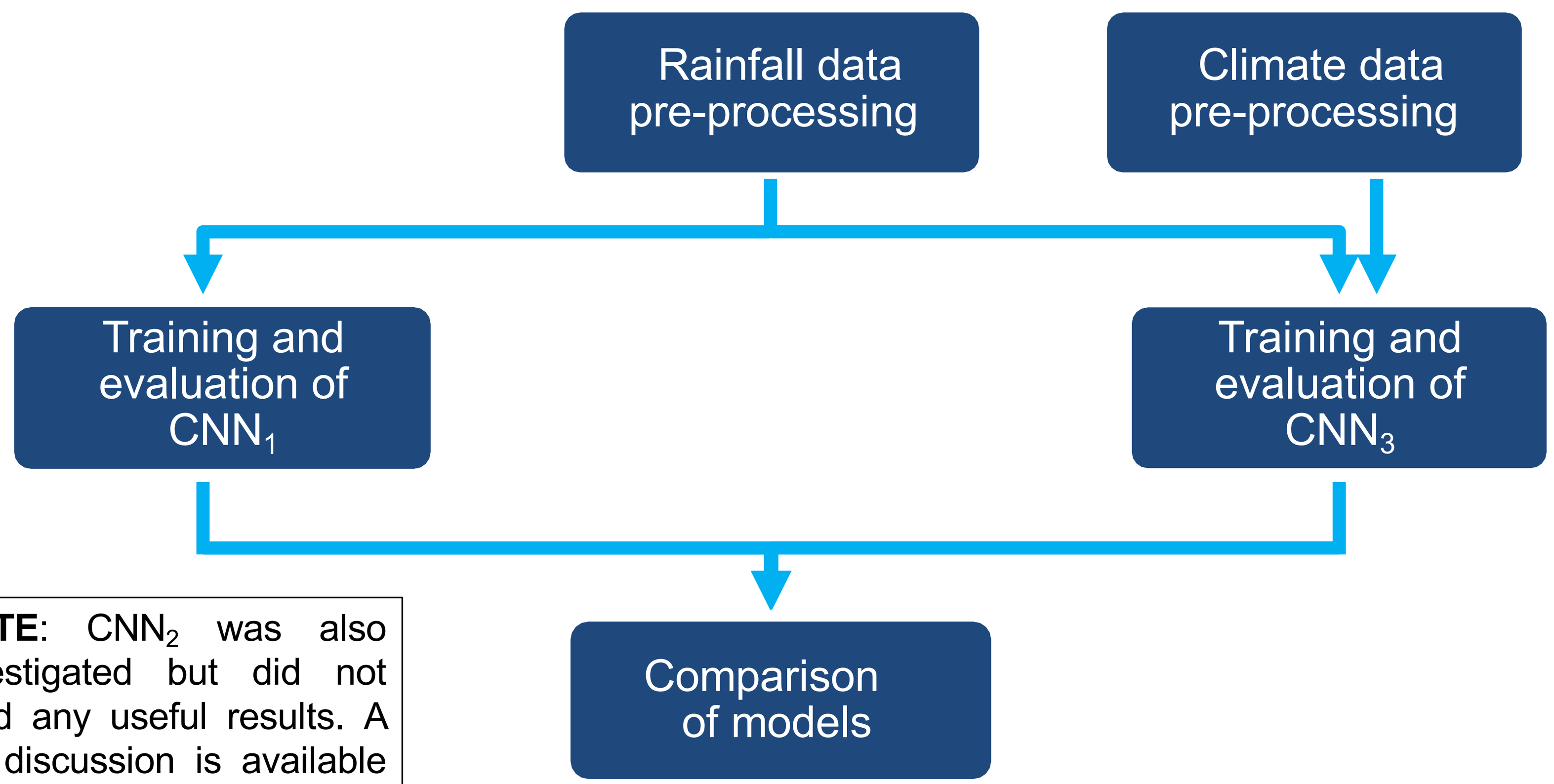
## 1. INTRODUCTION

Prolonged storm events of high rainfall intensities have the potential to induce destructive floods. One of the factors exacerbating the likelihood and destructiveness of these rainfall induced floods is the increasing areas of impervious paved ground due to urbanisation. Furthermore, the warming effects of climate change in recent years have resulted in increased rainfall intensities. According to Massam (2020), the Clausius-Clapeyron relationship dictates that atmospheric water content increases by 6 to 7% for every °C increase in surface temperature. This positive correlation is supported by recent evidence indicating that rain volumes from extremely wet days has increased by 17% when comparing periods between 1961 to 1990 and 2008 to 2017 (UK Met Office, 2021) with 19 of the warmest years recorded in history occurring since 2000 (NASA, 2021). It is for these reasons that the development of reliable rainfall forecasting models is imperative such that their outputs can be used for flood predictions.

Today, most rainfall forecasts are conducted using physics-based models (e.g., Numerical Weather Prediction – NWP) and data-driven models (statistical and Machine Learning, ML, models). However, the improvements in computational capacities and processing powers coupled with the growing popularity of AI (Artificial Intelligence), have fuelled the proliferation of ML models. Thus, this research aims to investigate the implementation of ML models, particularly Convolutional Neural Networks (CNNs), in rainfall forecasting and identification of high rainfall intensities for Birmingham as it is the 2nd most populous city in the United Kingdom and suffers from an outdated sewer network designed in the 18th century.

## 2. METHODOLOGY

An overview of the methodology is shown in Figure 1. Two main CNNs were developed. The first ($CNN_1$) was trained on a database of rainfall radar images only and the second ($CNN_3$) was trained on climatological features in addition to the same database of rainfall images to quantify any differences. Both models were quantitatively evaluated on the hit rates on high rainfall pixels and qualitatively evaluated on the predicted images.




**NOTE**: $CNN_2$ was also investigated but did not yield any useful results. A full discussion is available in the report

*Figure 1: Overview of the methodology*

## 3. RESULTS AND DISCUSSION

Several investigations were conducted for both models, which included but were not limited to: spatially aggregating pixels, varying the input setup, increased forecast lead times and combinations of loss functions and transformations.

| Combination | Loss Function | Transformation | Recall (R) | Precision (P) | F1-Score (F) |
|---|---|---|---|---|---|
| 1 | MAE | Raw | 18.95% | 60.87% | 28.90% |
| 2 | | Transformation 1 | 9.79% | 66.81% | 17.07% |
| 3 | | Transformation 2 | 12.92% | 61.54% | 21.36% |
| 4 | Logcosh | Raw | 13.60% | 68.79% | 22.70% |
| 5 | | Transformation 1 | 7.25% | 65.48% | 13.06% |
| 6 | | Transformation 2 | 10.45% | 53.25% | 17.47% |

| Setup | $t_{in}$ | Window Size | Samples | Recall (R) | Precision (P) | F1-Score (F) |
|---|---|---|---|---|---|---|
| 1 (No overlap) | 3 | 5-minutes | 22,580 | 0.64% | 40.51% | 1.26% |
| 2 (Overlap) | 12 | 5-minutes | 25,699 | 0.15% | 26.67% | 0.31% |

*Figure 2: Examples of quantitative results for $CNN_1$ (Top) and $CNN_3$ (Bottom)*

## 3. RESULTS AND DISCUSSION (*Cont'd*)

### $CNN_1$

The model exhibited a hit rate of approximately 20% across most investigations. This low hit rate was primarily due to the high proportion of zero intensity pixels that the model was trained on. The model performed the best when the inputs were sequential, and the temporal resolution was equal to the forecast lead time. The general spatial structure could be predicted but more localised high rainfall pixels could not.

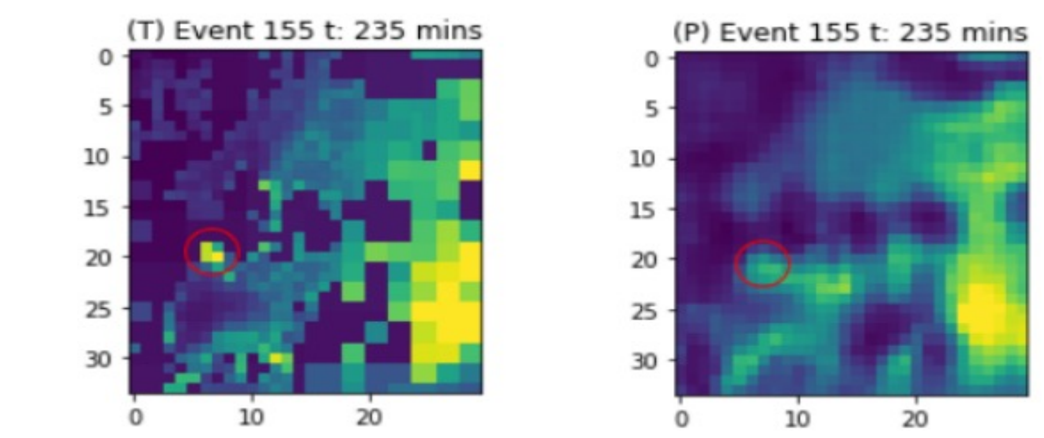



*Figure 3: Comparison of true and predicted images ($CNN_1$)*

### $CNN_3$

$CNN_3$ commonly exhibited a hit rate of $\leq$ 1%. This was attributed to the high proportion of zero intensity pixels, skewed climatological features and the input setup of the climate data. The model performed the best when the relative humidity, temperature and ozone mixing ratio were excluded as these were the most skewed climatological features. Qualitatively, the model predicted a more 'natural' (mostly oval) shape for the clusters of high rainfall pixels and in some cases, also overpredicted the spatial extent.

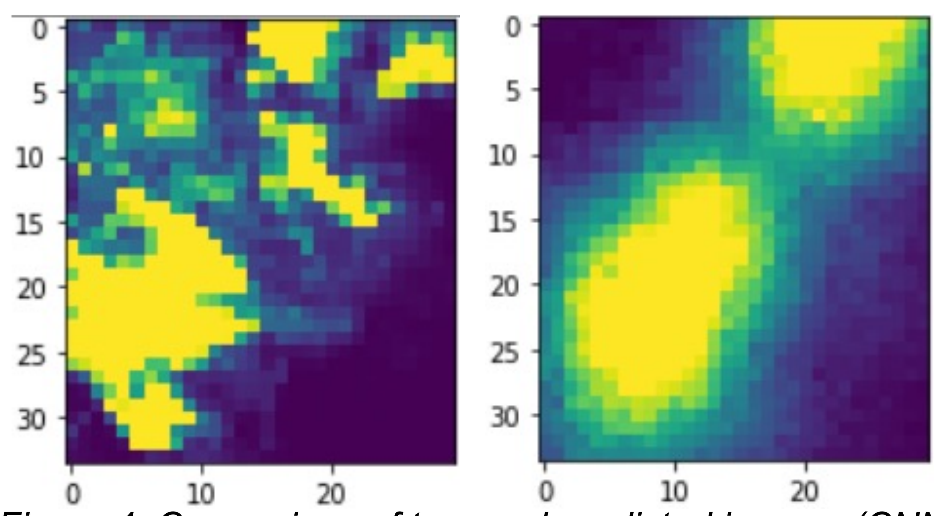


*Figure 4: Comparison of true and predicted images ($CNN_3$)*

## 4. COMPARISON OF MODELS

From both quantitative and qualitative standpoints, $CNN_1$ outperformed $CNN_3$. $CNN_1$ not only exhibited a higher hit rate but also depicted the spatial structure of the rainfall fields more accurately. Where both models suffered from the high proportion of zero intensity pixels, $CNN_3$ was also hindered by the skewness of certain climate data and the mismatching temporal resolutions between the climatological and rainfall data, which led to the duplication of climate data and doing so may have introduced unwanted errors. However, both models also struggled to identify localised high rainfall pixels and commonly depicted 'smoother' images. Ultimately, it is inconclusive whether the inclusion of climate data adversely affects predictions without a proper climatological setup.

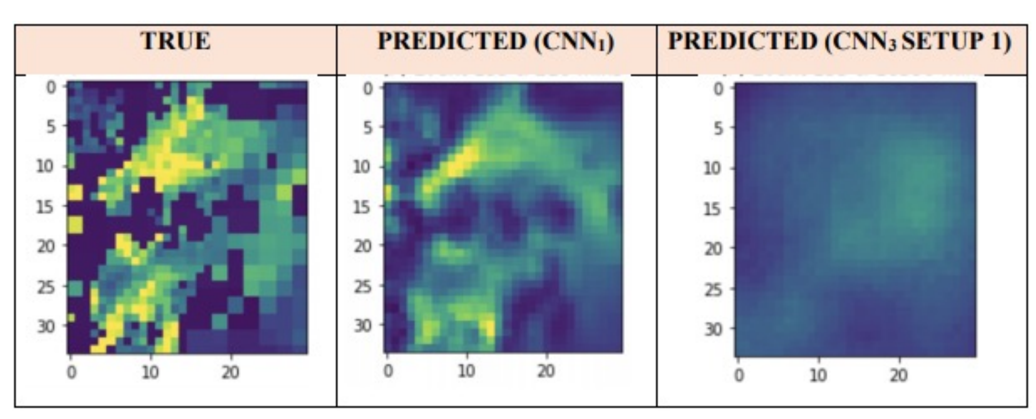



*Figure 5: Predicted images for $CNN_1$ and $CNN_3$*

## 5. CONCLUSIONS

In this study, two models were investigated – one was trained on rainfall images only, the other on rainfall images and climatological data. The results have shown the former outperformed the latter. However, the novelty of incorporating climatological data, duplication of climate data due to mismatching temporal resolutions, etc. means there is opportunity for further refinement. Solutions have been proposed in the report.

## ACKNOWLEDGEMENTS

The author would like to thank Dr. Christian Onof and Yuting Chen for their continued support, encouragement and insights. This research would not have been possible without them.

## REFERENCES

NASA (2021) *Global Climate Change Vital Signs of the Planet*. Available from: https://climate.nasa.gov [Accessed: 11th March 2021]

UK Met Office (2021) *UK Extreme Events – Heavy Rainfall and Floods*. Available from: https://www.metoffice.gov.uk/research/climate/understanding-climate/uk-extreme-events-_heavy-rainfall-and-floods [Accessed: 11th March 2021]