UNIVERSITY OF LEEDS

Deep Learning Predictions of Tropical Flood Images using Satellite Rainfall Data

Word Count: 17,360

Sulayman Birt
201342787

Supervised by Dr Robert G. Aykroyd

Submitted in accordance with the requirements for the
module MATH5872M: Dissertation in Data Science and Analytics
as part of the degree of

Master of Science in Data Science and Analytics

The University of Leeds, School of Mathematics

September 2023

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Academic integrity statement

You must sign this (typing in your details is acceptable) and include it with each piece of work you submit.

I am aware that the University defines plagiarism as presenting someone else's work, in whole or in part, as your own. Work means any intellectual output, and typically includes text, data, images, sound or performance.

I promise that in the attached submission I have not presented anyone else's work, in whole or in part, as my own and I have not colluded with others in the preparation of this work. Where I have taken advantage of the work of others, I have given full acknowledgement. I have not resubmitted my own work or part thereof without specific written permission to do so from the University staff concerned when any of this work has been or is being submitted for marks or credits even if in a different module or for a different qualification or completed prior to entry to the University. I have read and understood the University's published rules on plagiarism and also any more detailed rules specified at School or module level. I know that if I commit plagiarism I can be expelled from the University and that it is my responsibility to be aware of the University's regulations on plagiarism and their importance.

I re-confirm my consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to monitor breaches of regulations, to verify whether my work contains plagiarised material, and for quality assurance purposes.

I confirm that I have declared all mitigating circumstances that may be relevant to the assessment of this piece of work and that I wish to have taken into account. I am aware of the University's policy on mitigation and the School's procedures for the submission of statements and evidence of mitigation. I am aware of the penalties imposed for the late submission of coursework.

Name	Sulayman Birt
Student ID	201342787

Acknowledgements

Firstly, I would like to express gratitude to God for helping me to do this work in the best capacity I could. I would like to give a big thanks to my supervisor, Robert, for all the help in working through these ideas and for our enjoyable meetings and discussions about this subject. I want to also thank my mum, dad and Layla for all their love and support during this time. Finally, I want to give credit to close friends on my course who helped me a lot to get to grips with writing a good master's dissertation.

Abstract

In order to counteract and prepare for the rising global trend of extreme flood events, there is a pressing need to develop powerful predictive models that can effectively map potential flood hazards. This is so that effective mitigating measures can be undertaken and the amount of destruction and loss of life can be minimized. Existing flood prediction models are often localized and lack the global coverage that is needed to encompass affected areas. This study makes use of Tropical Rainfall Measuring Mission satellite rainfall data and Dartmouth Flood Observatory flood event data in order to try to predict tropical region flooding from rainfall patterns. The use of an image-to-image global prediction of flooding from rainfall appears to be an entirely novel technique. Several deep learning models are developed, including CNN, RNN, CNN-RNN and Convolutional RNN, to predict 72x20 flood images from rainfall images of the same resolution. The images are downscaled from the original satellite-grid resolution of 1440x400 in order to test the concept of a fully global, image-based technique. Furthermore, Convolutional RNN is used to predict future rainfall images based on the existing sequence of patterns. The CNN, RNN, and CNN-RNN show poor results in flood prediction, predicting no flood events across the entire test set of images. However, the Convolutional RNN seems to perform much better, with an average precision of 0.9252, an average recall of 0.8657, and an average f1-score of 0.8583. The Convolutional RNN for rainfall prediction also fares well with an average RMSE of 0.1243. an average MAE of 0.0461, and an average SSIM of 0.4756. Despite some limitations in this image-based sequence global prediction method, these models can act as a proof of concept from which more powerful models can be constructed. Future endeavours can aim to leverage additional climatic and topographical variables in a multi-modal analysis or utilize more powerful computational systems to encompass the true spatiotemporal extent of global rainfall and flooding, by considering all of the underlying anthropogenic and geophysical factors. This can be incorporated alongside existing localised efforts to manage and predict flooding hazards in areas already at risk, such that the damage and destruction caused by flooding events is best mitigated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

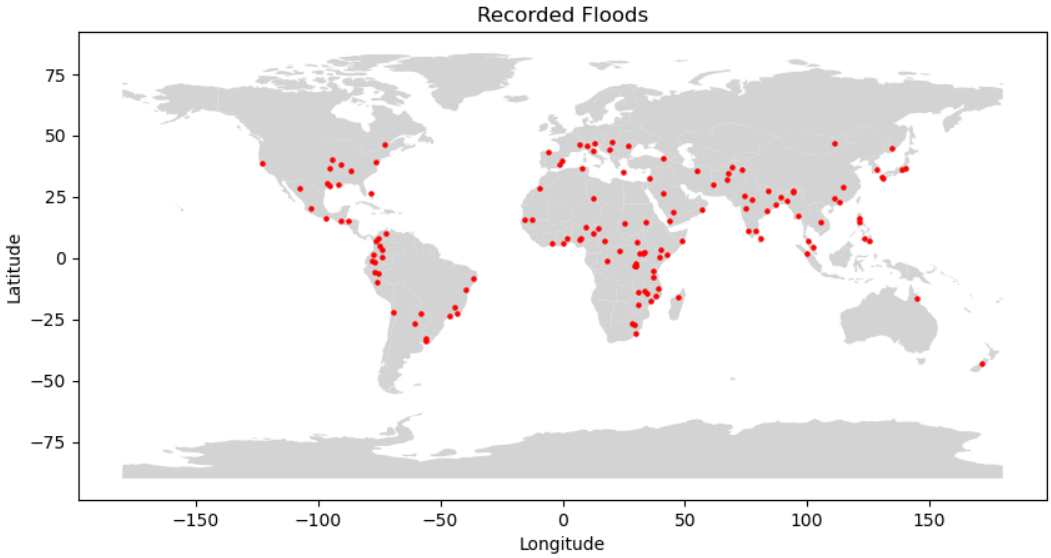

Figure 1.1: 50 Degrees North-South Flooding 1997-2019 (from Dartmouth Flood Observatory)

In the last several decades, the scientific community, and by extension the whole world, has been made aware of the increasing global temperature and extreme weather patterns resulting from climate change. Anthropogenic activities over the last few hundred years have directly altered atmospheric composition, leading to an intensified global water cycle, accompanying more intense rainfall, longer and more severe droughts, and the melting of glaciers and polar ice caps [1]. In particular, this has led to a rise in flooding events occurring worldwide, which can be seen over the twentieth century specifically. Rising sea levels from the thermal expansion of seawater and the melting of polar ice have facilitated tropical cyclone floods and hurricane storm surges [2]. This great spread of global flooding is demonstrated in Figure 1.1 and the trend of rising sea levels can be seen in Figure 1.2. Furthermore, local flash flooding has greatly risen, being attributed to the more frequent and intense heavy rainfall period caused by climate change, and this has been exacerbated by the disruption of natural floodplain processes from river channelization and dam construction [3].

It has been known for several decades that flooding is the natural disaster that affects the most people annually, specifically, flash floods kill the most people per event, although Asian river floods have a greater overall global impact [4] [5]. No matter if taking a generous or

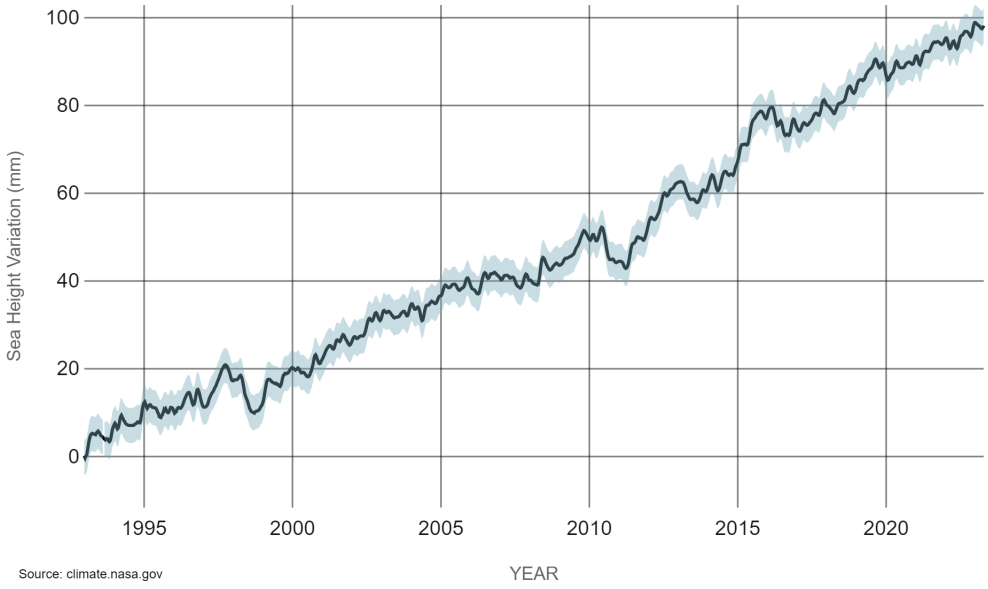

Figure 1.2: Sea Level Variation over the Last 30 Years (Source: NASA)

cautionary estimate of the global temperature rise, a minimum of 300 million people will be directly affected by floods each year, which has an immediate impact on a stricken country's social and economic development [6]. The fact that global exposure to floods can vary so much in a given year means that preparations must be made for the highest level flooding impacts even prior to significant global temperature rise [7].

Earth's tropical and subtropical regions are particularly at risk from large flooding disasters, due to their proximity to the equator and unique climate that feeds into high rainfall, saturated soil levels, and complex drainage systems [8]. They receive high amounts of rainfall that varies seasonally and intra-seasonally throughout the year, often caused by factors like the Intertropical Convergence Zone (ITCZ), irregular climate phenomena like El Niño and La Niña, and accompanying monsoon systems that bring high-moisture air masses, causing tropical rainfall storm surges and waterlogged soils.

This is exacerbated by the wealth of river systems that run through these bands, particularly in the western Amazon, and South and Southeast Asia, where extensive river systems can overflow due to the low-lying coastal or floodplain regions that have limited natural drainage processes [8]. The higher potential for extreme weather events and risk of coastal flooding in these regions means that countries in these regions, especially in Southeast Asian countries like Myanmar and Vietnam, face great risk from flooding events, made even worse by the limited infrastructure and flood mitigation measures present in both the low and high-income countries of these regions [9].

The link between rainfall and the water cycle, and by extension flooding, is a complex one, with many interrelated processes coming together through several feedback loops to determine the conditions in a single region of the globe. Climate change has led to global warmer air, which can hold more moisture, leading to heavier downpours that can overwhelm drainage systems and natural waterways. Precipitation intensity and duration, have a direct impact on the triggering or exacerbation of flooding events. Not only does a changing climate lead to greater flood risk, but tropical storms and droughts also play a part in increasing pressures on humanity

[10]. Several factors, like precipitation, evaporation cycles, soil moisture, tropospheric vapour, surface runoff, glacier sizes, and the changing seasons all have a direct impact on, whether a flood event, either through river or coastal means, can transpire.

It is difficult to track the movement and intensity of heavy rainfall areas due to the turbulent nature of the deep convective factors associated with thunderstorms and extreme weather events that trigger precipitation [4]. Heavy or intense rainfall over a short duration can lead to flash floods, where the rate of rainfall exceeds the rate of groundwater absorbance, leading to rapid surface water runoff and inundation. In this way, floods are not just a result of the weather patterns, but also the topography of the ground and the underlying geology, where water is in constant transfer and movement, meaning that the modelling of these complex interlinked processes is highly evasive. For example, even in regions like Australia where extreme rainfall events are increasing, changes in soil moisture levels have led to a decrease in flood events, again highlighting the confounding nature and variability of global flood predictions [11]. Continuous or prolonged rainfall over a longer period can saturate the soil past critical levels meaning that rivers overflow their banks and inundate adjacent areas.

Understanding the risk from floods, in the context of an individual country's needs and economic and social profile, is key to mitigating flood risk and damage, and saving as many lives as possible, through the development of tailored adaptation strategies and recovery plans [9]. Projections of risk must take into account socioeconomic development models, along with river and climate projections, to develop the best country-specific flood management plans, which often necessitates the use of Geographic Information System (GIS) modelling to integrate all techniques best [12]. There is also the problem of trying to scale up local risk models to the global paradigm, as global flood risk models (GFRMs) are often impractical in real-world usage, feeding into the need to develop more complex predictive techniques [13].

Unfortunately, there is a paradox in the fact that flood-prone regions are favourable for human settlement, due to the transportation accessibility and food production potential, meaning that humanity's regions of high asset value are often at the highest risk, with rapid urbanization and infrastructure development in flood-prone areas exacerbating flooding problems [14]. Increasing population densities and infrastructural buildup have led to the need for greater defensive measures, which again facilitates the increased global flood risk over the last several decades. The need for flood prediction and risk management in a spatiotemporal sense has therefore also increased, whereby economic risks are first set out before quantifying the levels of potential damage [15] [16]. Additionally, the clearing of forests and changes in land use can alter natural water flow patterns by increasing surface runoff and reducing groundwater storage, due to extensive impermeable surfaces like roads and buildings, which compounds the risk of flash floods.

Structural solutions like levees and dams that block or control water flow, combined with non-structural strategies like early warning systems and community preparedness measures alert, educate and equip the population in a timely fashion. These measures are vital to ensure the protection of people's livelihoods when faced with extreme events, especially where there has previously been a lack of investment in such schemes [15]. Zoning and mapping of flood-prone areas allows for informed land use decisions to reduce flood exposure, such as elevated and strengthened structures or community relocations. These strategies, along with developing improved and well-maintained drainage systems, and utilizing natural measures such as floodplain restoration and reforestation, can be used to siphon and guide rainwater away from critical areas.

Overall, to better prepare for the global flood and extreme event wave, there is a great need to develop powerful and sophisticated predictive models that take into account all of the

spatiotemporally changing socioeconomic and hydrological factors at play. Chapter 2 will break down the underlying satellite techniques underpinning the analysis, and the history of machine learning methods being made use of. Subsequently, Chapters 3 and 4 will delve deeper into the technical aspects of this project, explaining the methodology, data sources, utilized deep learning models, and the practical application of these models in predicting flooding events based on TRMM rainfall data and flood imagery from the Dartmouth Flood Observatory. The results of the study are highlighted and analysed in Chapter 5, leading onto a discussion of deep learning methods in flood prediction and the future of work in this field in Chapter 6, along with the conclusive statements following this analysis. This research endeavour aims to contribute to the global understanding of flood prediction and, ultimately, enhance our capacity to mitigate the growing threat of flooding in a changing climate. It greatly favours the fostering of collaborative efforts between the scientific community, governments and local communities to implement the most effective risk-reduction strategies.

Chapter 2

Background

2.1 Satellite Measurements

Rainfall patterns are too variable and intermittent to track entirely globally through conventional monitoring methods, such as with localised station rain gauge measurements, as these lack the spatiotemporal power required to effectively capture large-scale trends and patterns. This has led to the development of satellite missions and methodologies tailored towards capturing global rainfall distribution, whereby a combination of different data sources is efficiently brought together to generate the most accurate evaluations [17]. Other climatic parameters such as temperature are much better and more accurately recorded over the long term than rainfall, due to the long-term use of geological and paleoclimatic evidence, which has led to the development of several different satellite missions with unique measuring instruments and techniques [18].

The incorporation of satellite and remote sensing data has yielded great improvement in the field of flood modelling, filling the previous gaps in large-scale studies due to the potential for global data coverage [19]. Several factors, such as topography, flood coverage and water levels can be captured effectively using various sensing probe methods, and alongside precipitation monitoring can directly feed into hydraulic flood models.

The era of satellite-based rainfall measurement began with the launch of weather satellites in the early 1960s, which provided insightful imagery but had limited rainfall measurement capabilities. In the 1980s, the development of passive microwave sensors allowed satellites to measure rainfall by detecting microwave radiation emitted by raindrops. However, the key and most influential mission in this field has been the Tropical Rainfall Measuring Mission (TRMM), a collaboration between the U.S. and Japan set up to have a strong focus on constraining tropical rainfall, due to its great variability and associated latent heating effects [20]. Originally launched in 1997 for a planned mission length of three years, this satellite remained in constant orbit and tracking all the way until 2015, and stands as a milestone in the field, providing invaluable insights into global precipitation patterns.

Several climatic variables are captured, with rainfall being the main focus of attention, and tracked using various sensors of differing frequency bandwidths. These include a passive microwave radiometer to measure surface and atmosphere radiation, precipitation radar (PR) to vertically profile raindrop size and intensity, and a Visible and Infrared Radiometer System to monitor clouds. The sensor data combine to provide detailed information on rainfall patterns, intensity, and distribution in the tropical region, which is depicted in Figure 2.1 [21]. The precipitation radar is operated at 13.8 GHz, giving a cross-range spatial resolution of around

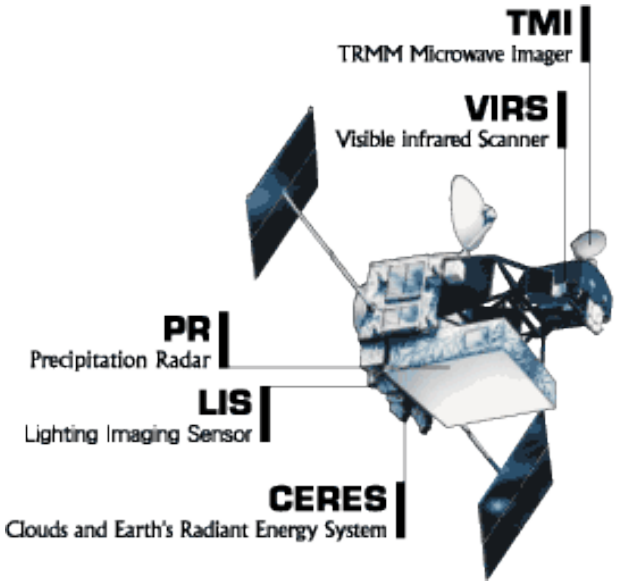

Figure 2.1: TRMM Satellite and Instruments (Source: NASA)

4.3km and therefore a rain range resolution of 250m, at the satellite's altitude of approximately 400 kilometres [22]. Over the mission span, the sensors have been recalibrated and rainfall algorithms redeveloped constantly in order to improve the rainfall quantification output [23].

The data outputs from this mission have been used in a variety of contexts, due to this being the first such rainfall data collection of its kind. For example, daily rainfall data has been utilised to simulate soil moisture contents to the same accuracy as conventional methods, feeding into global tracking of the water cycle [24]. The rainfall data is highly malleable and has been thoroughly statistically described and broken down into its probability density functions, allowing for further applications in numerical predictive models [25]. This mission has also been directly followed up on by the Global Precipitation Measurement (GPM) mission, which was launched in 2014 and carries advanced instruments like the Dual-frequency Precipitation Radar (DPR) and the GPM Microwave Imager (GMI) which provide continuous updates up to the present day. The new mission provides global coverage of rainfall and snowfall, covering the TRMM dataset for a further 4 years after its mission ended in 2015, leading to a full TRMM dataset coverage from 1998 to 2019.

The Dartmouth Flood Observatory (DFO) maintains a global database of major flooding events, being established by researchers at Dartmouth College in the U.S. with the goal of monitoring and analyzing global flood events using remote sensing and satellite imagery. They work to compare pre and post-flood imagery to identify changes in land cover and water extent, based on satellite imagery from sensors like MODIS (Moderate Resolution Imaging Spectroradiometer) and Landsat, as well as other data from radar and further remote sensing methods. The archive images originate mainly from passive 3 GHz microwave remote sensing data signals of water surface changes that suggest potential river inundation events [26].

2.2 Global Flood Modelling

In recent years, global flooding predictive modelling efforts have turned to increasingly powerful statistical and computational techniques that more often than not make use of satellite data prod-

ucts to improve the coverage of utilised data. These can encompass scales that are beyond those of traditional monitoring, in order to make direct requests to the appropriate response teams [27]. Early prediction efforts relied heavily on hydrological models that incorporate weather station rainfall data to estimate river discharge and potential flood risk based on measured and predicted precipitation. However, the emergence of satellite technology offers new tools for monitoring weather and climatic conditions, especially in the case of TRMM data, which has become greatly utilised in modelling efforts. Satellites can now describe land surface reflectivity variations in order to quantify flood water at a high resolution, which is directly applicable to hazard map generation [28].

High-resolution hazard models have been realized at a global scale and are especially useful in low-income countries and data-scarce areas. For example, simulated river discharges have been fed into a hydrological river network model and thereby can model local flood inundations, coming into fruition as continental hazard maps of local flood-prone regions [29]. This technique of utilizing globally connected river channel network simulations has also been applied to UK and Canadian datasets to generate high-resolution flood hazard maps for terrestrial areas [30]. Current operational flood forecasting systems use a combination of ground-based observations, remote sensing data, weather forecasts, and hydrogeological models to predict and monitor floods, acting as an ensemble approach.

Machine learning is the use of algorithms and statistical techniques to analyse and learn from patterns in inputted data, and such models have been shown to be very potent in accurately mapping flood hazards. For example, Sentinel-1 satellite imagery has been utilized in order to attempt to detect flooded pixels in images and finds that the unsupervised flood detection method shows a greater performance over traditional machine learning methods, therefore having more potential use in urban settings [31]. Techniques such as Artificial Neural Networks and Logistic Regression have also been applied to Nigerian flood records, further bolstering the use of Machine Learning techniques to accurately map flood risks and guide mitigation plans [32].

There is a noticeable global lack of hybrid machine learning and image processing usage in flood management scenarios, and future endeavours should consider these innovative methods for hazard risk prevention, especially where they can be integrated into dynamic GIS urban and meteorological models to better quantify flood risk [33] [34]. TRMM data has also been directly fed into global flood monitoring systems and improved their performative power, although limitations are revealed such as a lack of high-resolution river water runoff models and knowledge of dam positions [35]. Global systems should also aim to integrate local forecasting of extreme events, which can make accurate predictions using the same TRMM products at single river basins [36].

It appears that traditional hazard mapping and even newer machine learning methods, whilst showing potential in many cases, are not powerful enough, on the whole, to map floods and associated hazards at the variety of different spatiotemporal scales that are needed to create truly global prediction models. However, more innovative techniques are able to use deep learning methods to better capture the complex interrelated processes of weather and flooding, using a variety of different data sources to learn difficult relationships between the key variables. Ensemble machine learning methods such as Random Forest and Gradient Boosting are effective in capturing temporal dependencies, but they do not have the full spatiotemporal grasp of newer deep learning techniques.

2.3 What is Deep Learning?

Deep learning is an umbrella term, first coined in 2012, for a relatively new advancement in machine learning that has given rise to a variety of new statistical tools. These models can learn complex features from large, unstructured data sets in order to solve perceptual problems [37]. The various methods use processing layers in a neural network to capture patterns within a given dataset, where each successive layer builds on the previous knowledge about the data, without the need for any initial explicit information inputted regarding the dataset. Different models within a deep learning setting have long-lasting origins in older linear models, and their roots can be traced to artificial neural networks (ANNs), with there now being applications in image processing, speech recognition, robotics, cybersecurity, large language models, and many more [38] [39]. The concept of artificial ANNs was first coined in the 1950s and 1960s, when early single-layer perceptrons, were used for simple pattern recognition tasks. The backpropagation algorithm was developed in the 1980s and allows multiple-layer neural networks to learn complex patterns and representations over time. However, it was not until the 2010s that the so-called deep learning versions of these algorithms came into existence and rose in popularity.

The branch of deep learning that is most pertinent to satellite data handling is image handling, which is often done through the use of a Convolutional Neural Network (CNN), and it was a deep CNN called AlexNet that won the ImageNet Large Scale Visual Recognition Challenge and marked a turning point in the field. Recurrent Neural Networks (RNN) particularly Long Short-Term Memory (LSTM) networks, were subsequently introduced and achieved success in natural language processing (NLP) tasks. These are now also used for image-based tasks and made use of to capture temporal dependencies, and along with CNNs have a variety of uses on satellite data. These include weather forecasting, precipitation nowcasting, and spatiotemporal analyses, which can act as a form of next-frame prediction of future events based on past information [40] [41]. These techniques have generally demonstrated remarkable capabilities in image analysis and recognition, especially with the ability of convolutional layers to capture spatial hierarchies. The field of deep learning for images has seen much of its growth due to its use in medical MRI image analysis, where methods such as detection, registration, segmentation, generation, and classification have been well-documented and explored [42] [43]. Deep learning continues to evolve to the present day, with research focusing on areas like reinforcement learning, generative adversarial networks (GANs), and large-scale language models like GPT-3 and GPT-4. More details on the deep learning methods used in this study, in particular on CNNs and RNNs, will be given in chapter 4.

2.4 Deep Learning for Weather & Flooding

Traditional numerical weather prediction (NWP) models, based on physical equations, were the foundation of meteorology, utilising simpler statistical techniques and limited observational data, prior to the introduction of machine learning methods. In recent years, deep learning, particularly through the use of CNNs and RNNs, has emerged as a highly powerful technique to use in precipitation nowcasting and current time predictions. These methods are better able to handle the complex spatiotemporal dynamics at play and capture the volatile underlying features of radar and satellite imagery, allowing for the improved performance of global rainstorm warning systems and targeted risk management [44] [45]. Deep learning does not necessarily have to replace traditional methods and can work in tandem with traditional weather simulations through ensemble predictions, which make better forecasts at lower computational costs, now

being common in operational weather forecasting systems [46]. There is also the potential for improved data assimilation, or integrating observed data into existing numerical models, using deep learning to automatically extract the necessary and relevant features that facilitate high predictive capabilities. These methods are relevant to all aspects of weather management alongside precipitation, such as storm nowcasting and drought management modelling, the latter of which has seen utilisation on TRMM data in China to model a high soil moisture predictive accuracy [47] [48].

Deep learning methods have been found after a literature survey to be generally superior to traditional machine learning and shallow neural network methods in the context of weather prediction, and it is argued that they bypass the lack of depth of understanding of physical mechanisms and observed data which affects traditional numerical weather prediction models [49] [50]. CNNs have been adapted to handle gridded weather data, such as weather maps and satellite images and are useful for tasks like cloud classification, storm tracking, and weather pattern recognition. Often it is CNNs and RNNs that allow the automatic extraction of image features, which can be scaled up to have a three-dimensional information component to facilitate more accurate nowcasting predictions [51] [52].

Other aspects of the water cycle pertinent to weather and flooding include rainfall-runoff modelling, which is underpinned by complex precipitation and evaporation processes. These can result in highly nonlinear relationships, for which deep learning, in particular CNNs and RNNs, can be utilised over traditional numerical models to give an improved predictive power [53]. Hourly runoff can be predicted to a high accuracy, which feeds directly into flood warning systems and procedures, even in instances of extreme events taking place [54] [55].

Alongside the vast development in rainfall and weather predictive power, deep learning techniques have been employed directly in the management and mapping of extreme flood events. This is again in order to improve on the results of traditional hydrological and meteorological methods that struggle to capture complex interactions using physical equations and statistical approaches. CNNs and RNNs are made use of to predict and map these events for use in mitigation strategies, through their ability to capture the complex spatial patterns of flood risk. RNNs and variants like LSTM networks employed for time series forecasting, enabling the modelling of rainfall patterns and river discharge [56]. A recent survey has found that while predictive accuracies have greatly improved, the difficulty lies in developing deep learning models that generalize well to new, unseen flooding scenarios [57].

These methods have the ability to determine the specific predictors in a case study, like terrain ruggedness, slope or elevation, which most contribute to flood risk, or model additional factors that are specific to floods in a given region, such as temperature for monsoon floods in India [58] [59]. It seems that a novel approach is to combine rainfall-runoff simulations with satellite precipitation data in order to better account for the complex terrain and associated factors [60]. Deep learning can be used create new virtual scenarios with differing river basins, rainfall levels, flood durations, and time distributions, such that predictive models can learn to generalize better to new flood scenarios [61]. Hybrid models that combine physical models with deep learning techniques are key to leveraging both the specific domain knowledge embedded in physical models and the complex data-driven capabilities of deep learning.

2.5 Justifying This Study

After this review of rainfall and flood prediction methodologies over the last several decades, it appears that deep learning techniques can offer the most valuable means of capturing the

complex underlying interrelated factors and processes. However, it appears that studies focus on models in individual river basins or urban areas and do not aim to generalize over the entire globe. The relationship between rainfall and flooding is extremely difficult to characterize at a large scale, but perhaps the validated methods of CNN and RNN models can be used at a large scale to generate short-term predictions of floods from measured satellite rainfall data, using a novel image-to-image comparison of rainfall and flooding scenarios that can possibly act as an anchor point for new global image-based models to arise.

This study makes use of TRMM satellite rainfall data and pairs it with the DFO database of flood events from 1985 to the present day. Several deep learning models are utilized, including CNN, RNN, CNN-RNN and Convolutional RNN, to predict image-based flood maps from the TRMM satellite rainfall sequence of inputs, alongside the prior flood image sequence. Furthermore, Convolutional RNN will be used to predict next-day rainfall in order to test this image-based methodology and serve as a comparison to the flood prediction efforts. It is pertinent to focus on the tropics due to the regional high flood vulnerability and complex rainfall patterns, which are more important to capture than areas where the risk is less high. This can facilitate the development of specialized and localized flood prediction solutions tailored to these high-risk areas based on the initial large-scale predictive effort.

The study downscales the high-resolution data to one that is appropriate for large-scale flooding trends and the computational limits set, in the hope that future work can aim to localise and increase the breadth and depth of the predictions. It is a new approach that treats both rainfall and floods as linked spatial images in order to better characterise the complex relationship between them, laying out a framework by which additional features and spatial complexity can be implemented down the line, following the validation of a simpler model on a narrowed focus of data, with a plan to fully globalize once the method is proven. The aims of this study can be summarised as follows:

- Investigate the current state of deep learning methods for global rainfall and flood prediction
- Develop a new image-based deep learning strategy for large-scale tropical rainfall prediction
- Train CNN, RNN, CNN-RNN and Convolutional RNN models to compare flood prediction performance
- Train Convolutional RNN to study the potential for next-day rainfall prediction
- Discuss the potential use for these models and the future of deep-learning rainfall and flood predictions

Chapter 3

Data

3.1 TRMM & DFO Datasets

Within the TRMM data the primary measurement, Rainfall, is available alongside other climatic variables such as latent heat release, which are not utilized in this study [62]. The TRMM provides rainfall data at intervals of monthly, daily, and 3-hourly, at a spatial scale of 0.25x0.25 degrees [63]. A balance must be struck in the choice of data interval when considering that some flood events may be strongly influenced by short-term heavy rainfall, which could be better captured with daily data, whereas others could be shaped by longer-term precipitation trends, such as a wetter-than-average month. Monthly precipitation data provides a coarser level of granularity compared to daily data, meaning that the precipitation information is aggregated over a month, possibly resulting in the loss of some fine-grained temporal patterns. However, this monthly data has been found to be at a comparable range to gridded gauge data for 9 years of the 22-year period, which supports a high accuracy despite the aggregative nature present [64]. A monthly dataset would contain only 264 images over the 22-year period, which is nowhere near the appropriate amount for which to conduct a deep learning analysis. In contrast, it is still necessary to reduce the dimensionality of the input as compared to the 3-hourly data as this would result in a dataset too large. Some coarseness of granularity is accepted to give a dataset size that lies within the set computational limits. Therefore, daily data has been selected for this analysis as it seems to provide the best balance between dataset size and resolution. The critical temporal dependencies of rainfall can be captured here by the deep learning models, through their ability to track longer and shorter-term trends and weather patterns.

The TRMM data is known to over-detect heavy rainfall events, which means that events relating to potential flooding are likely to be present in the record [65]. There are also data available that quantify errors in the rainfall measurements due to rain detection issues, particularly in cases of complex terrain, but for the sake of simplicity, these errors are not considered [66]. Deep learning estimates of rainfall from the sensor rainfall products have also been made available, but for simplicity, only the estimates provided by the TRMM are utilized in this study [67]. Overall, this rainfall data can provide key information about the changing rainfall patterns across the globe which are necessary in developing a high-quality flood prediction model.

The DFO archive provides spatial and temporal information on river floods, flash floods, and other types of inundation events, including their location, duration, and magnitude. This data has been used prior in image form to validate existing flood hazard estimates, and even to investigate flood events at a pixel-level [68] [69]. It has the advantage of being focused on global flood events only, as opposed to other large-scale datasets like EM-DAT (Emergency Events

Database), which has a wider scope of extreme event coverage but loses the strong coverage of flooding events specifically. Many countries maintain flood records from their meteorological and hydrological services, although these are highly country-specific and lack the global coverage that is needed for this analysis. The process of bringing together several of these would be too time-consuming and pose technical challenges that are beyond the scope of this study. The DFO data here, when converted to image form, crucially acts as a ground truth against which flood predictions can be measured.

3.2 Data Selection & Processing

Spatial data has the advantage over numerical approaches as it can capture complex information in a more compact format. Each gridded pixel corresponds to a specific location and can encode key information at the set resolution level, which is often much more space-efficient than storing numerical values for each data point. Furthermore, using images for this analysis allows for the use of deep learning techniques such as CNNs and RNNs which are already known to have strong applications with such input modalities. CNNs especially are translation-invariant, meaning they can recognize patterns and features regardless of their position in the image. This is useful in mitigating the greatly unpredictable nature of spatial weather patterns.

This flood prediction approach is inherently image-based, relying on the visual representation of flood events from satellite imagery to capture spatial relationships. This allows for the provision of key contextual information crucial for accurate and reliable predictions, reducing the need for manual feature engineering. Utilizing spatial data allows for direct feature extraction techniques by the deep learning models, where the images can be upscaled or downscaled in resolution as needed to ensure balance in the temporal processing aspects of the training pipeline. The outputted prediction images can also be verified visually by the human eye to make inferences or observations from the results, and easily augmented to optimize training for the best predictive performance. Ideally, the trained models should see better generalization across different spatial contexts with variable locations and scales, through their ability to recognize similar patterns in new data.

In this study, the computational tasks can be parallelized efficiently as the use of virtual CPUs and GPUs is well-suited to speed up the processing stages. Trying to reshape data frames with an extremely high number of rows is highly taxing on system memory, even when utilizing virtual memory systems. The calculations for the data processing and analysis are performed using Python, via the use of Kaggle online notebooks. These are able to utilize Google's free online computing systems, here having 2 CPU cores and 13 GB of GPU memory. It is this level of computing power that limits the extent to which a very high-resolution spatiotemporal flood prediction model can be developed and utilized in this study.

The TRMM data is available on separate download links as individual HDF5 files for each recorded measurement, which are individually downloaded one by one using an automated script. These are converted from rainfall-latitude-longitude data to image data by concatenating each file along the time dimension. The maximum rainfall value across the entire set is 1036 mm/hr, which is an extreme outlier as the vast majority of readings are between 0 and 12.5 mm/hr, with this highlighted in Figure 3.1. The data have a resolution of 1440x400 squares of 0.25x0.25 degrees each, across a band between 50 degrees North and South. Across the entire dataset, there are 249,948 data points with missing values. This may seem like a lot, but in total there are 4,499,040,000 data points resulting from 8034 days of 1440x400 images, meaning that only around 0.00006% of the data is missing. The numerical data is converted into vectorized

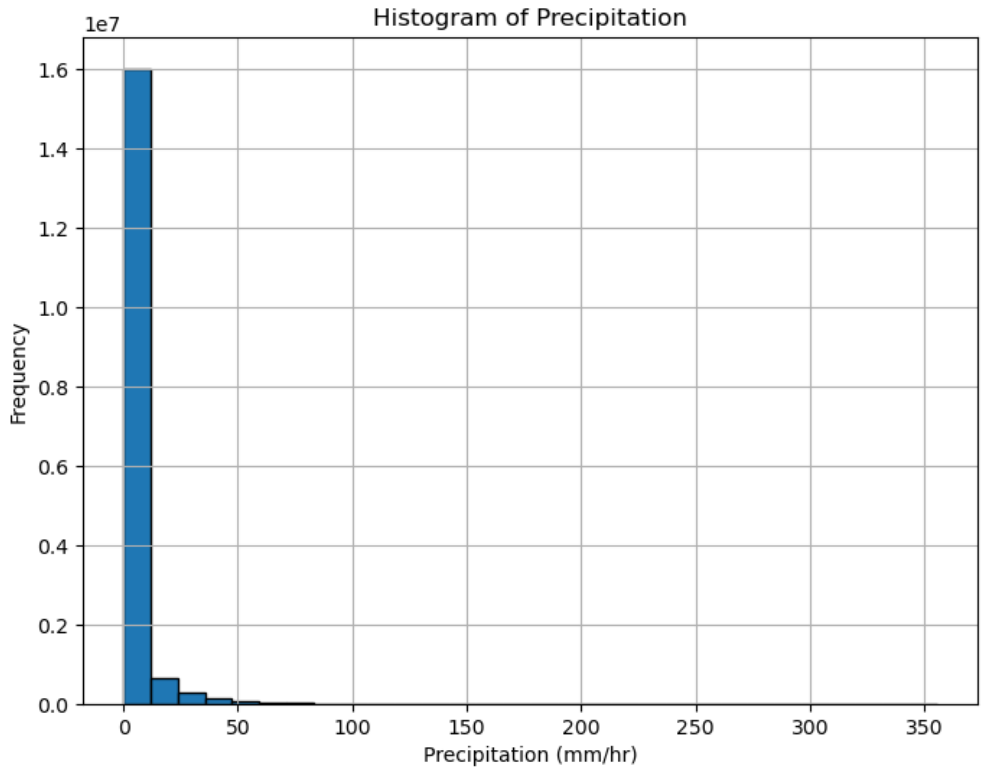

Figure 3.1: Tropical Region Precipitation in 2019

images, and in the cases of missing data, the most recent value in time for that pixel position is carried over, acting as a time series imputation. The rainfall data is normalized between 0 and 1, where the maximum rainfall of 1036 mm/hr is used to scale all the values, ensuring consistency across the time series of dataset images. White or lighter grey pixels represent zero or low rainfall whilst darker grey or black represent high rainfall at a given position.

The DFO data is available from 1985 to the present, so the downloaded spreadsheet data is filtered to have the same time coverage of 1997-2019 as the TRMM data. The file contains data on 5130 different floods, covering their position in latitude and longitude as well as several impact factors like the number of casualties, main underlying cause, and severity. However, most of these variables are not useful to this analysis and are excluded accordingly. The latitude and longitude positions of each flood are firstly mapped to the nearest values on the 1440x400 TRMM data grid. As the TRMM data is only available between a band from 50 degrees North to 50 degrees South, all other flood positions are eliminated from the data. One simplification that must be made here is the separation of flood events over multiple days. Each event is provided with a start and end date, which is used to generate identical events for each day across the flooding period. This is such that each flood image can be created with pixel positions representing the currently active floods according to the data. Another key assumption is that each flood represents a single pixel position given by the latitude and longitude data when in fact they have a greatly varying size and shape due to the hydrodynamics and geological makeup of each situation. Even though the DFO data provide an area of coverage measurement for each event, this study chooses to contain each event within a single pixel in order to simplify the deep learning prediction task and initial data processing steps. The hope is that future endeavors can

try to better represent the true spatial coverage of floods. Looking for repeats within the 139 remaining flood events shows that there is one repeat entry that must be removed, which may be a result of the rounding of latitude and longitude values. Figure 1.1 in Chapter 1 shows a great spread of floods across the dataset that is mostly concentrated in coastal or river flood-plain regions. The flood positions and dates are used to generate a set of 1440x400 images, where a white ”0” pixel represents no flood, and a black ”1” pixel represents a flood event. There are 8034 images that correspond to each of the 8034 days in the rainfall image set.

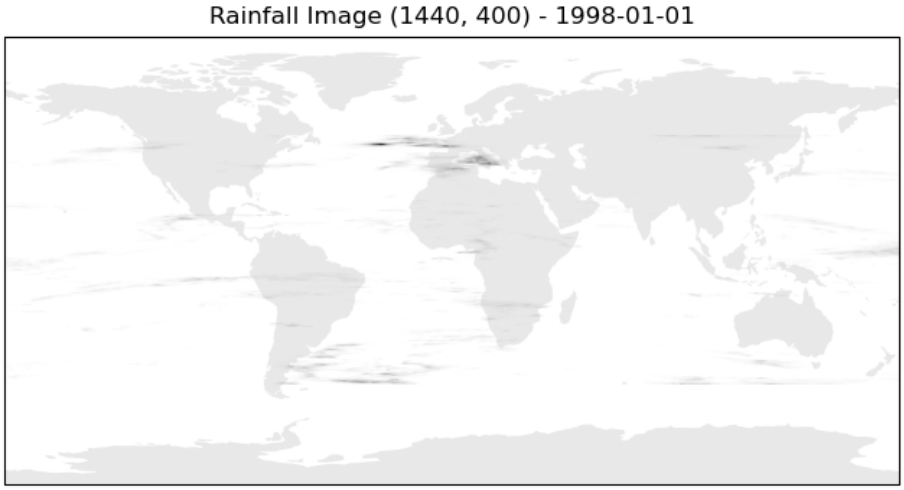

Figure 3.2: Tropical Rainfall before 20x Downscaling

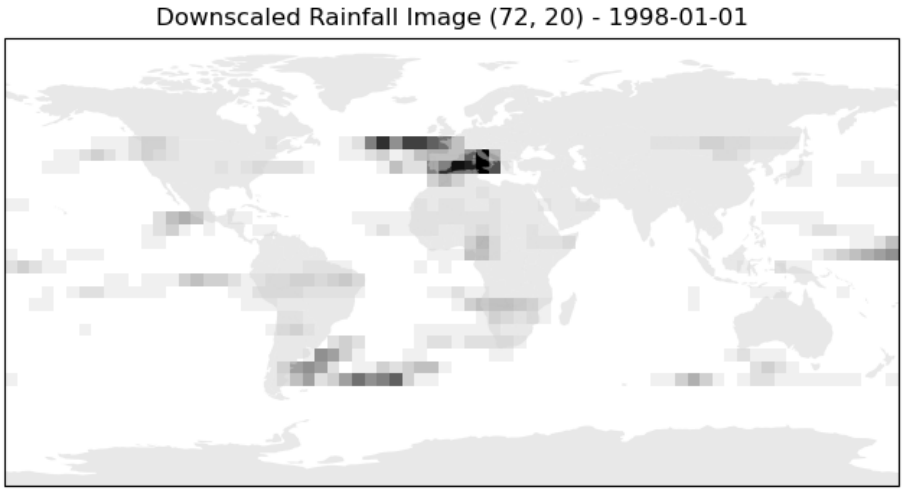

Figure 3.3: Tropical Rainfall after 20x Downscaling

There are only 40,247 flood event pixels out of a total of 4,499,040,000 data points, which means that this is a greatly imbalanced dataset of only around 0.000009% positive examples. Performing deep learning calculations on images with a resolution of 1440x400 is beyond the computational limits set by the hardware available to use. The rainfall and flood images must therefore be downscaled to reduce the computational burden and training time and allow for greater visual interpretability of results. For the rainfall images, Lanczos resampling is used

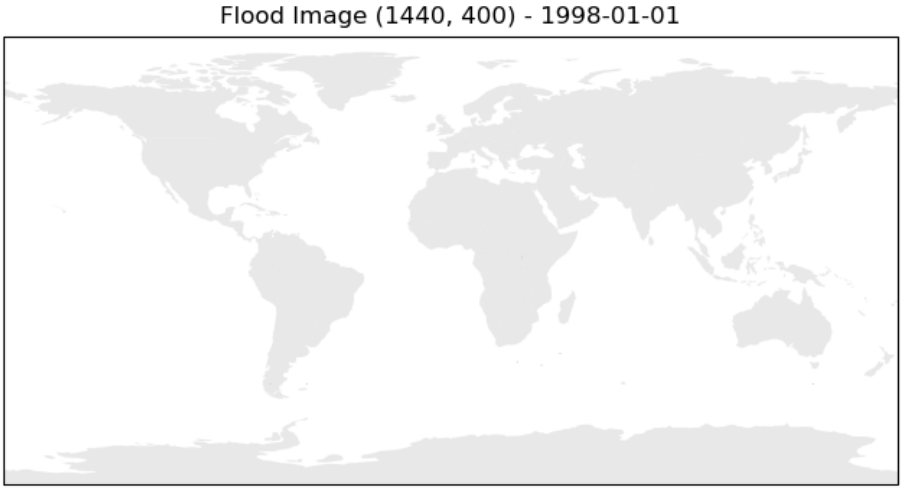

Figure 3.4: Tropical Flooding before 20x Downscaling

to downscale the resolution by increasing the sampling rate of the signal, acting as a form of multivariate interpolation, for which further mathematical details are not necessary to go into. The effects of this downsampling can be seen in figures 3.2 and 3.3. A different approach is required for the flood images to keep the black flood pixels when downscaling. A custom downscaling function is used to iterate through each new grouping of pixels and determine whether it should be black or white depending on if black pixels are present within. The effects of this downscaling can be seen in figures 3.4 and 3.5, where the original image's black pixels are not visible at the original resolution. A downscaled resolution of 72x20 is chosen here, which is a 20x reduction from the original size of 1440x400. This value is determined as a good balance between having a greater degree of certainty in a flood event location and considering the computational constraints and increased training time for higher-resolution images.

It is important to determine if the models require higher spatial resolution than those chosen here, in order to accurately classify flood pixels, or if the chosen resolution is appropriate. This level of downscaling seems to maintain coverage of the large-scale trends in tropical rainfall, with both dark-grey high-rainfall and light-grey medium-rainfall regions preserved at a large scale. For the flood images, it is clear that losing the high level of precision is worth the ability to greatly reduce the dataset imbalance. There are now 39,278 positive examples out of 11,568,960 data points, giving a ratio of around 0.003%. While this dataset is still very imbalanced on the whole, a deep learning model could still potentially be adjusted to be able to learn the positive examples. There are usually at least 1 or several floods within each image given to the model in sequence.

The size and extent of floods can vary greatly depending on the region, climate, and local topography, with some being localized and covering a small area. However, others are more extensive and affect larger regions, so it is important to select a pixel size that is relevant to the range of potential flood sizes. The new size of a flood area pixel prediction following downscaling is much larger, taking up close to 98000 km^2 rather than around 250 km^2 at the original resolution. The pixels can be seen visually to cover up a significant region of a country or even the majority of a smaller country. Without computational constraints, it would be ideal to better constrain the locations of potential floods, but here a country could nonetheless have some idea of there being flood risk in a certain region, and therefore able to make the necessary

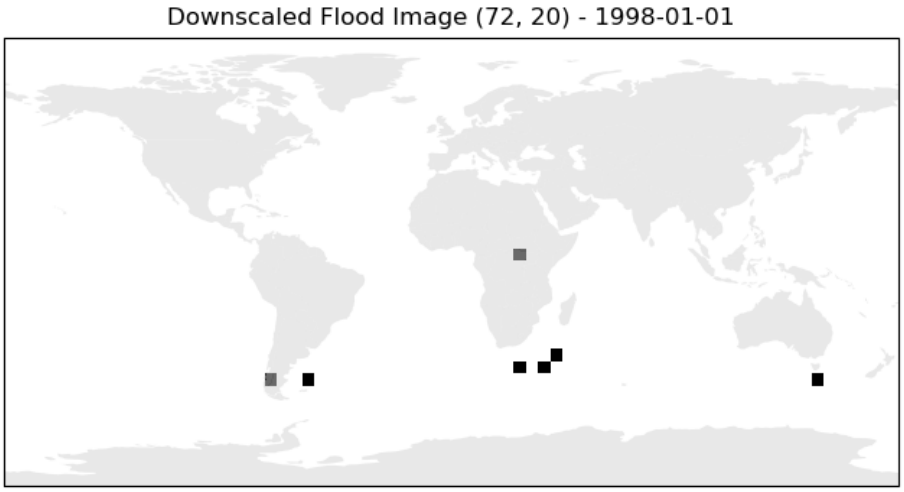

Figure 3.5: Tropical Flooding after 20x Downscaling

preparations and appropriate targeted actions. As an initial analysis of long-term flood patterns or trends, this broader spatial region is acceptable. These preprocessed images are directly supplied to the deep learning models, rather than downscaling during the loading process, as this is much more efficient.

One option for reducing the computational strain of this problem would be to limit the spatial region, either by selecting a narrower latitude band or perhaps integrating GIS data or flood hazard maps to spatially filter out oceanic regions or other areas where flood risk is less of a key factor. The DFO flooding data could also be used to create spatial regions surrounding points where floods have been recorded. However, this would be a very general representation lacking consideration of the complexities and nuances of spatiotemporal flood patterns in these areas. This would potentially complicate the use of image-based deep learning methods, although maybe favoring the use of more traditional machine learning methods. Contrastingly, filtering the data in time is another option for reducing the dataset size, although it would not address the high computational burden of processing each full-resolution image. Therefore, image downscaling remains the most logical option as it allows the full spatiotemporal scope of the TRMM data to be utilized, but at a high-level scale. Subsequent efforts can aim to use higher-resolution images or incorporate the fully global GPM dataset.

Both the input rainfall and target flood image groups are split into training, validation, and testing sets, whilst maintaining the temporal order of the data. It is essential to maintain the temporal structure when providing images to the model, such that the shift in weather patterns over the 22-year period is kept in chronological order and can be learned in sequence. The training set period is decided as the first 14 years, from 1998 to 2011. Additionally, the validation and test sets are created as four years each, from 2012 to 2015, and 2016 to 2019 respectively, giving a split ratio of 7:2:2. This appears an appropriate split for a dataset of 8034 images, as it leaves 5113 training images, 1461 validation images, and 1460 testing images. The validation images are subsequently used to iteratively improve the predictions based on the training set, whilst the test set remains unseen and is used to calculate the final model performance scores.

Chapter 4

Methodology

4.1 CNN Model

CNNs are well-suited for image-based tasks due to their ability to automatically learn hierarchical features from inputs and are known as one of the most representative types of networks in the field [70]. They have various applications, such as in visual and speech recognition, and natural language processing (NLP) [71]. Convolutional layers are used to apply filters, known as kernels, to inputted images to extract the key features and capture spatial hierarchies through backpropagation. This allows for the solving of complex image pattern recognition tasks in a variety of novel contexts [72]. There are also pooling layers, which downsample the generated feature maps in order to reduce spatial resolution while preserving the necessary information. Fully connected layers are implemented to make predictions or classifications at the end of the network, based on the previously extracted features, alongside non-linearity layers in some cases [73] [74]. This model, and all subsequent models, are trained on the provided data by iteratively feeding batches of input images and corresponding target images. The model's weights and biases are dynamically adjusted, using backpropagation and the optimization algorithm to minimize the loss function.

MobileNetV2 is used as the CNN feature extractor of choice, as its lightweight architecture and low number of parameters allow for the efficient processing of flood images whilst reducing the computational burden. Downscaled grayscale images of 72x20 should not be hard to learn in a deep image-analysis sense, despite the complex spatiotemporal rainfall dynamics underneath. In MobileNet architectures, convolutional layers take outputted feature maps directly from previous layers in dense blocks in order to efficiently generate a larger number of features for the learning process [75]. A depthwise separable convolution is used to adjust hyperparameters such that the tradeoff between model accuracy and latency can be optimized [76]. Models like VGG and ResNet are deeper and have more parameters, making them potentially more suitable for complex tasks and large datasets. However, the more lightweight and efficient architecture of MobileNet makes it a better choice for this relatively small dataset and potentially avoids the overfitting of a more complex model. The MobileNetV2 version of the MobileNet architecture has shown a wide range of high-performance applications when utilizing the pre-trained weights, such as with the recognition of grayscale palmprint images [77].

This pre-trained model has been trained on large image datasets and can therefore be fine-tuned for the flood prediction task in this study. Such a transfer learning approach can save computational resources and time while improving model performance. This dataset is relatively small and lacks diversity, so the pre-trained weights allow for the utilization of the knowl-

edge acquired from training on the set of various simple images. It uses depthwise separable convolutions, to significantly reduce the number of parameters and computations compared to traditional convolutional methods. The classifier part of the model is replaced with a simple identity layer and the first layer is modified to handle the single-channel grayscale input images rather than typical RGB images. In the forward pass, the model processes the input through the MobileNet layers and then applies Adaptive Average Pooling to obtain a 1x1 feature map, which is flattened into a feature vector to prepare it for classification. The convolution has a kernel size of 3, stride of 2, and a padding of 1.

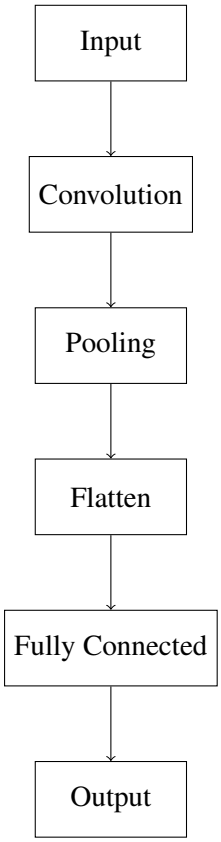

Figure 4.1: CNN Model Flowchart

The structure of the CNN model architecture can be seen in Figure 4.1. A fully connected layer with two output neurons (weight and bias) is utilized as the new classifier for predicting flood or non-flood pixels. It achieves this by transforming the feature vector to the desired resolution of 72x20. Batch normalization and dropout layers are added for regularization, with the former applied to the input image, and the latter applied to both the input image and the feature vector. A Multihead Attention layer with four heads is added to process the feature vector, allowing the model to capture complex relationships in the data. Implementing an attention mechanism allows the model to dynamically attend to different parts of the input. This means it is potentially able to focus selectively on the rainfall features that are most important to flood prediction. A batch size of 32 is used for data loaders with two workers, although shuffling of the data is disabled to ensure the maintenance of the temporal structure. These data loader and batch settings are kept for all the models utilized.

4.2 RNN Model

RNNs play a crucial role in time-series data analysis, making them a natural choice for rainfall prediction due to their highly temporal nature. An RNN model can capture the temporal dependencies within a set of data, enabling accurate predictions based on historical patterns. Unlike standard feedforward neural networks, where information flows unidirectionally, RNNs have connections that loop back on themselves, essentially providing a form of memory for processing sequences effectively. The RNN of choice here is the LSTM network, which addresses the vanishing gradient problem by incorporating specific memory cells and gating mechanisms. This is done in order to capture long-range dependencies in the sequential data [78]. LSTMs have a variety of applications, such as in the forecasting of Photovoltaic system output power, where there are significant temporal dependencies present [79]. Gated Recurrence Units are another RNN variant and simplify the LSTM architecture into an efficient solution while retaining key capabilities. However, this study opts to utilize the slightly greater performance of traditional LSTMs [80].

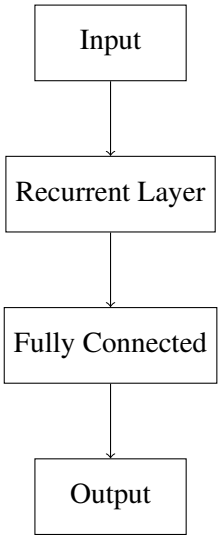

Figure 4.2: RNN Model Flowchart

Figure 4.2 depicts the architecture for this RNN model. The model contains two LSTM layers, one for processing rainfall data and another for processing flood data, with hidden sizes of 128. Multihead attention layers with four heads are added for both rainfall and flood data. A batch normalization layer is applied to the along-feature-dimension concatenated output of the two LSTM layers, to normalize the data along this dimension. A linear fully connected layer is used to generate binary predictions from the processed data at the appropriate resolution.

For this RNN model, along with the subsequent CNN-RNN and 2 convolutional RNN models, a sequenced input is given rather than an individual image input. This means that a grouping of five sequenced images is given rather than one alone. Along with the long-term temporal capabilities provided by each model, a 5-day window up to the day of prediction is given to the model. This means it can focus on and learn the key short-term rainfall trends leading up to the final day of rainfall or flooding, depending on the model in question. Furthermore, all but the last model with rainfall prediction combines either a rainfall image or feature map prediction with a prediction based on the previous 4 days of flooding. This is done in the hope that the model can learn the interlinked trends in rainfall and flooding that lead to the day

of flooding being predicted. In other words, the models can be thought of as predicting day 5 of flooding, based on days 1-5 of rainfall and days 1-4 of flooding. In the case of the final ConvRNN model, day 5 of rainfall is predicted based on a sequence of days 1-4 of rainfall.

The real-world implications of this method seem somewhat appropriate given the fact that global flooding information is likely to be readily available for previous days. This can be used in conjunction with global rainfall measurements for the last few days and the rainfall data thus far on the given day in question. The use of this 5-day sequence also seems reasonable given the fact that it takes time for the accumulated rainfall to result in significant changes in flood levels. A sliding window approach, such as this, where future values are predicted on a short-term scale sees a variety of different uses, like in stock and share price prediction [81]. It was also considered whether or not to make use of a date tensor input to the time-based models (i.e. all but the initial CNN model). However, it proved apparent that the temporal dependencies could be automatically learned through the sequenced nature of the input images.

4.3 CNN-RNN Model

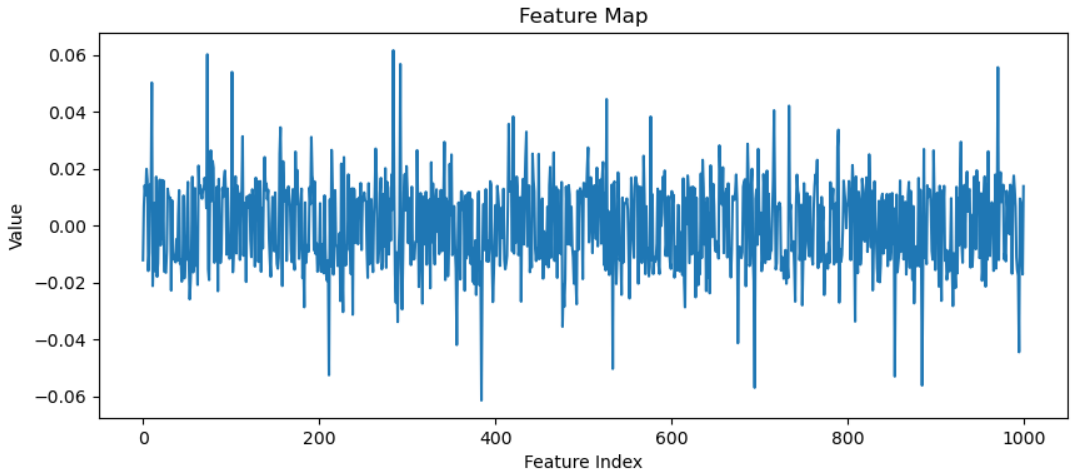

Figure 4.3: CNN Feature Map for 1st January 1998

The CNN-RNN hybrid model combines the strengths of both CNNs and RNNs to leverage both spatial and temporal information in flood prediction. Image data is processed through CNN layers, before feeding the features into RNN components to capture long and short-term temporal dependencies [82] [83]. The Convolutional feature extractor layers capture spatial patterns and hierarchical internal representations, with LSTM layers aiming to find sequenced dependencies and patterns within these feature representations. This method has shown higher predictive accuracy potential than CNN and RNN models alone in different scenarios, such as in daily stock price, housing energy consumption, and crop yield forecasting [84] [85] [86]. The feature vectors provided by the CNN have a length of 1000 values, which are not easy to find understandable meaning in, as can be seen in Figure 4.3. Using pre-trained CNN models within such hybrid architectures is not something novel, and for example, has been used prior to effectively classify blood cell images with higher accuracy than traditional CNN analyses [87].

The CNN-RNN architecture is highlighted in Figure 4.4. The model contains two LSTM layers, one for processing flood data and another for processing feature maps, with hidden sizes of 128. Similar to the RNN model, multi-head attention layers with 4 heads are used for both

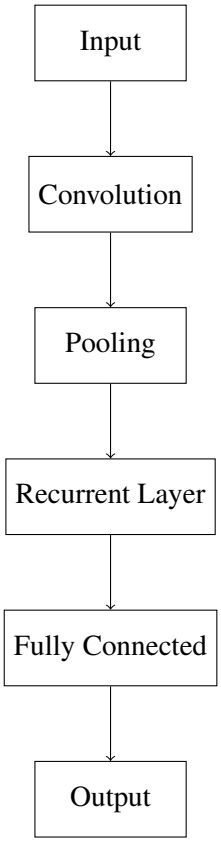

Figure 4.4: CNN-RNN Hybrid Model Flowchart

flood data and feature maps. There is a batch normalization layer applied to the concatenated output of the two LSTM layers and a linear layer to generate output predictions.

4.4 Convolutional RNN Models

Convolutional RNN models (ConvRNNs) integrate convolutional layers within the RNN architecture, allowing for the capturing of spatial and temporal dependencies simultaneously. This makes them particularly useful for tasks involving both image data and sequential information. In these situations, they can offer a potentially more powerful flood prediction capability, as there is a clear spatiotemporally varying component. ConvRNNs have widely been used for gesture recognition, where the model contributes to the temporal feature aggregation over the recurrent steps to learn the long-term spatiotemporal features [88]. ConvRNNs start with a CNN component that acts as a feature extractor, followed by RNN layers that are responsible for modelling temporal dependencies. However, they differ from the CNN-RNN models in that there is a single pipeline for the training process. In other words, there is not a distinct training process for the CNN that feeds into the same for an RNN, although there is the tradeoff of increased computational intensity. However, this technique has the potential for improved spatiotemporal representation of rainfall patterns and therefore increased flood predictive power. In the context of clinical diabetes detection, Convolutional LSTM (ConvLSTM) has been found to have higher predictive accuracy, than the CNN, LSTM, and CNN-LSTM counterparts [89]. In this study, the CNN and RNN models are not ones that can be expected to perform well, as they lack

full spatiotemporal coverage, so have been included in order to compare results with the more promising CNN-RNN and ConvRNN analyses.

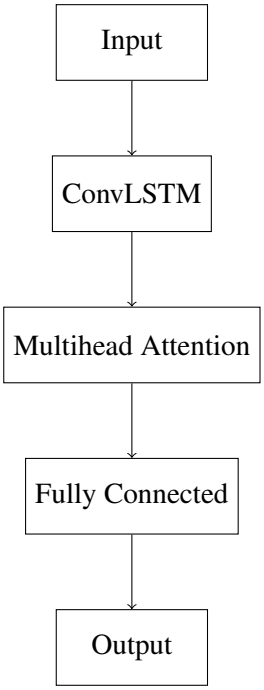

Figure 4.5: ConvLSTM Model Flowchart

The flood prediction ConvRNN model is shown in Figure 4.5, and contains two ConvLSTM layers, one for processing rainfall data and another for processing flood data. These have a hidden dimension of 256 multiplied by the number of layers and a kernel size of 3x3. There is also a convolutional layer to linearly map the ConvLSTM outputs to the desired output size, with a kernel size of one. Multihead attention layers with four heads are applied to both the rainfall and flood ConvLSTM outputs, and batch normalization is applied to the concatenated ConvLSTM output.

The rainfall prediction ConvRNN model processes the TRMM rainfall data to forecast future precipitation patterns. The incorporation of Convolutional RNN layers should ideally enhance the ability to predict rainfall, which is in turn a crucial input for flood forecasting. ConvRNNs have been employed in various similar applications, including predicting rainfall data in meteorology and hydrology. They have been shown to greatly outperform traditional precipitation nowcasting methods such as optical flow methods [90]. Such methods are optimal for treating both the input and predictor target as spatiotemporal sequences, allowing for better capturing of the correlations underpinning precipitation nowcasting [91]. ConvLSTM has also been used previously on TRMM data to predict short-term precipitation based on the known historical record to a degree of accuracy exceeding traditional numerical models [92]. The key difference here from the previous models described is in the fact that this is a continuous regression problem rather than a binary prediction, meaning that an appropriate choice of loss function must be made in training. The purpose of this analysis is to provide results alongside the flood prediction which can act as a potential validator of this downscaled-image predictive method using TRMM data. This is due to the use of ConvRNN in the context of rainfall prediction being well-documented, facilitating the development of more complex dynamic predictors

of rainfall alongside extreme flood events.

The model is similar to the previous Convolutional RNN model and contains a single ConvLSTM layer for processing the sequential rainfall data and a convolutional layer to linearly map the output to the 72x20 resolution. Once again, multi-head attention with four heads and batch normalization are applied to the ConvLSTM output. The use of multi-head attention, in this case, has the potential to improve the forecasting power significantly, as it has been shown prior to potentially highlight the key regions for which to base predictions on [93].

4.5 Overfitting & Loss Functions

Overfitting is a common challenge faced by deep learning analyses, where models perform very well on training data but fail to generalize to unseen data. This is combatted through the use of data augmentation, dropout layers, and regularization methods, ensuring the robustness of deep learning models. The level of overfitting is greatly dependent on the choice of activation function, network structure, learning rate, and the number of epochs trained for [94]. Continuous gradient updates can imbalance the highly sensitive loss function underneath [95]. Deep learning models are prone to fitting noise in the training data, so an overly complex model captures noise and fluctuations in the training data. This is especially a problem where the training dataset is small and the model does not have a diverse enough set of samples from which to effectively learn the underlying patterns. Additionally, training a deep learning model for too many epochs means that a model may continue to fine-tune its parameters until it fits the training data perfectly, losing generalization capability in the process.

The models in this study utilize regularization techniques in the form of adding L2 regularization terms to the model's optimizer, here using a value of 10^-5. This is done in order to penalize large parameter values and combat the fitting of noise by adding an L2 norm penalty term to the loss function, encouraging the model to reduce the squared values of the weight. Furthermore, dropout in training is implemented to randomly drop a portion of neurons in the network to prevent co-adaptation of neurons and promote more robust representations of the data, with the probability set to 0.5. The optimizer function used for all models is the Adam optimizer, which is a common choice that is particularly useful for models with a lot of noisy or sparse gradients, here utilizing an initial learning rate of 0.001. The rate of learning is dynamically reduced by a factor of 0.5 when the model stagnates or stops improving for three epochs, via the use of the ReduceLROnPlateau scheduler. This is able to decrease the learning rate appropriately as training progresses to fine-tune the model and help it converge to a better solution. Gradient clipping with a value of 1.0 is used to scale down the gradients if their norm exceeds a predefined threshold, ensuring that they stay within a reasonable range and do not become too large. Batch Normalization is also implemented to avoid the distribution of activation values changing during training, by subtracting the batch mean and dividing by the batch standard deviation. Scaling and shifting by learnable parameters allows for the optimization of each layer and thereby ensures regularization.

The use of a validation set is used to automatically stop training when the validation error starts to increase or shows no improvement over 10 epochs, which are markers of overfitting. Where possible the models here are kept simple with fewer layers and parameters to avoid overcomplexity and allow for a more fair comparison between them. Another important step to reduce overfitting is the normalization of each rainfall image to be between zero and one, such that the minimum and maximum rainfall values are mapped to these values, acting as a form of data augmentation. While this means that the relative rainfall quantities between sequential

images that were previously maintained are not preserved, the performance should be improved due to the image-based models being more able to pick out higher or lower rainfall areas from images that have higher overall contrast. The data is less affected by the extreme outlier rainfall values which give typical high rainfall values a poor contrast with lower values. Other commonly used forms of data augmentation for improving model generalization include rotations, translations, scaling, flipping, and adding random noise. However, these are utilized here as they would possibly have a detrimental effect on the model's understanding of the complex spatial rainfall trends, due to them not making physical sense and potentially introducing noise that does not align with the problem's characteristics. In this analysis, the models with the lowest validation loss are selected as the best versions to generate predictions following the use of automatic stopping. The parameter values for the model results given in Chapter 5 are the ones chosen following iterative retraining to determine the best solutions within the computational constraints set.

Selecting appropriate loss functions is key to training deep learning models effectively. They have a direct impact on model performance optimization and the ability of models to make accurate predictions whilst also minimizing errors. The first four flood analysis models predict binary flood images from a continuous rainfall input, which is a classification problem. Therefore, Binary Cross-Entropy Loss (BCE Loss) is used to measure the dissimilarity between the predicted values and the true binary labels. This encourages the model to predict high values for true positives and low values for true negatives. A sigmoid activation and then a rounding function are applied to the output layer to give binary predictions for the presence of a flood in a given pixel position. The weight of positive predictions is adjustable such that more or less weight can be given to flood pixel values in order to counteract any potential class imbalances in the data. Another option here would be to use Focal Loss which assigns a higher weight to hard-to-classify examples to combat any dataset class imbalances. In these models, the sigmoid activation and rounding are built into the loss, meaning that it does not need to be applied to outputs manually. However, the final predictions must be run through an explicit sigmoid activation and rounded to give a binary output.

The rainfall prediction task with ConvRNN is a continuous regression prediction of the next rainfall image in sequence. This therefore favours the use of Mean Squared Error Loss (MSE Loss) to measure the average squared difference between the predicted continuous rainfall values and true target image values, thereby penalizing larger errors more heavily. There is no activation function used to keep the values between 0 and 1, although the MSE Loss acts to prefer values that lie within this range. However, the final output values for the test set predictions are clipped to the range zero to one, with values below zero raised up to zero and values above one dropped down to one. This is done such that a visual and statistical comparison can be made with the true rainfall images.

Chapter 5

Results

5.1 Model Performances

Gaining insight into the predictive capabilities of the deep learning models sheds light on their potential performances across the various geographic regions and temporal scales of a global flood management effort. For the binary prediction problems, Confusion Matrices heatmaps can be plotted in order to visualize the proportions of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The retrieval of such scores allows for the calculation of metrics like precision, recall, and F1-score. Precision quantifies the proportion of true positive predictions against all positive predictions made, in order to try to emphasize the minimizing of false positives, calculated as $\frac{TP}{TP+FP}$. Recall similarly measures the proportion of true positive predictions amongst all actual positives, aiming to minimize false negatives, being calculated as $\frac{TP}{TP+FN}$. Finally, the F1-score combines precision and recall into a single metric in order to act as a balanced quantification of the model's performance. It is especially useful for imbalanced data and is calculated as $2 * \frac{precision*recall}{precision+recall}$. These metrics can be visualized using Precision-Recall (PR) Curves that plot precision and recall for different threshold values in order to investigate the various model's abilities to balance these two variables.

For the continuous rainfall prediction, Mean Absolute Error (MAE) is calculated as the sum of absolute errors divided by the sample size, to measure the average absolute difference between predicted and true flood images. This is alongside the calculation of Root Mean Square Error (RMSE), which is the square root of the mean squared error (the mean of the square of the errors), which penalizes larger errors due to the squared term. Furthermore, the Structural Similarity Index (SSIM) utilized here is generally used to quantify image quality degradation as a result of data compression or data transmission losses and acts as a different lens by which to view the sequenced rainfall prediction quality. It measures the structural similarity between the predicted image and the ground truth image by comparing luminance, and contrast, giving a score between -1 and 1, where 1 indicates a perfect match and -1 indicates a complete mismatch. The formula for SSIM is not necessary to go into further detail about but essentially calculates various matrix components for different windows of a given image. The RMSE can be utilized to plot a heatmap of average errors across the image sets in order to identify areas of high performance or concern across the predictions.

The results for the flood prediction modelling can be seen in Table 5.1. These models are trained and retrained such that the results shown here are the highest-scoring ones following the tweaking of the various models and hyperparameters after each iteration. However, it is important to note that these are not fully optimized solutions, due to automated hyperparame-

Model	Precision	Recall	F1-Score
CNN	1	0.0692	0.0692
RNN	1	0.0694	0.0694
CNN-RNN	1	0.0694	0.0694
ConvRNN	0.9252	0.8657	0.8583

Table 5.1: Flood Prediction Performance Metrics

ter searching techniques being beyond the technical limits of this study. The results from the first three models, CNN, RNN, and CNN-RNN are disappointing, as they predict entirely white images with zero flood pixels present. This suggests that they are not powerful enough to distinguish the rainfall patterns and their links to the flooding regions. Despite the fact that many anti-overfitting measures have been implemented, it seems these models have learned to predict entirely white images to minimize the BCE loss of predictions. As the data is highly imbalanced towards negative pixels, such monotone predictions would generally be very accurate, suggesting a problem in the methodology rather than the training processes undertaken if overfitting is strongly mitigated.

When making predictions on the entire test set, the CNN produces an average precision of 1.0000, which is misleading as it is a result of there being 0 true positives and 0 false positives as the model does not predict any positives. There are also poor average recall and f1-scores of 0.0692 for both, which result from the instances of false negatives due to the model always predicting no flood in the case of a flood. The results are very similar for both the RNN and CNN-RNN, with an average precision of 1.0000, and an average recall and f1-score of 0.0694 for both. The precision-recall curves and confusion matrix heatmaps for these models have been omitted as they show no useful information due to these poor performances.

The learning processes for the two ConvRNN models are improved, with the training and validation set losses converging over a longer period to give a smoother period of learning the data. This suggests that finer details of the predictions are improved on and fine-tuned over time. The results for the flood prediction ConvRNN are vastly improved over the previous three models, with some of the predictions having a 100% match to the ground truth flood images, although this is not the case across the entire set. The model seems to show reasonable success in predicting flood pixels and non-flood pixels with high accuracy as compared to before. Across the test set, the average precision is 0.9252, the average recall is 0.8657, and the average f1-score is 0.8583, which suggests that the model is very highly successful in predicting flood pixels correctly. These metrics indicate a high accuracy of positive predictions, and a high success in avoiding predicting non-flood pixels incorrectly, meaning a good ”completeness” of positive predictions.

The Precision-Recall curve for the Flood ConvRNN predictions in Figure 5.1 shows a flat line across the majority of the predictions. This suggests that the recall can be increased and more True Positives can be recognized without also increasing the number of misclassified negative results or the number of false positives. The number of correctly classified flood pixels is high and this appears to have no significant negative bearing on the number of flood scenarios that are missed out on, again supporting the fact that this model is performing well in this flood prediction analysis.

The confusion matrix plot for this model in Figure 5.2 shows that there is almost 100% correct classification of negative non-flood pixels, with an almost complete proportion of true negatives classified as compared to false negatives. This suggests that the model is very good at

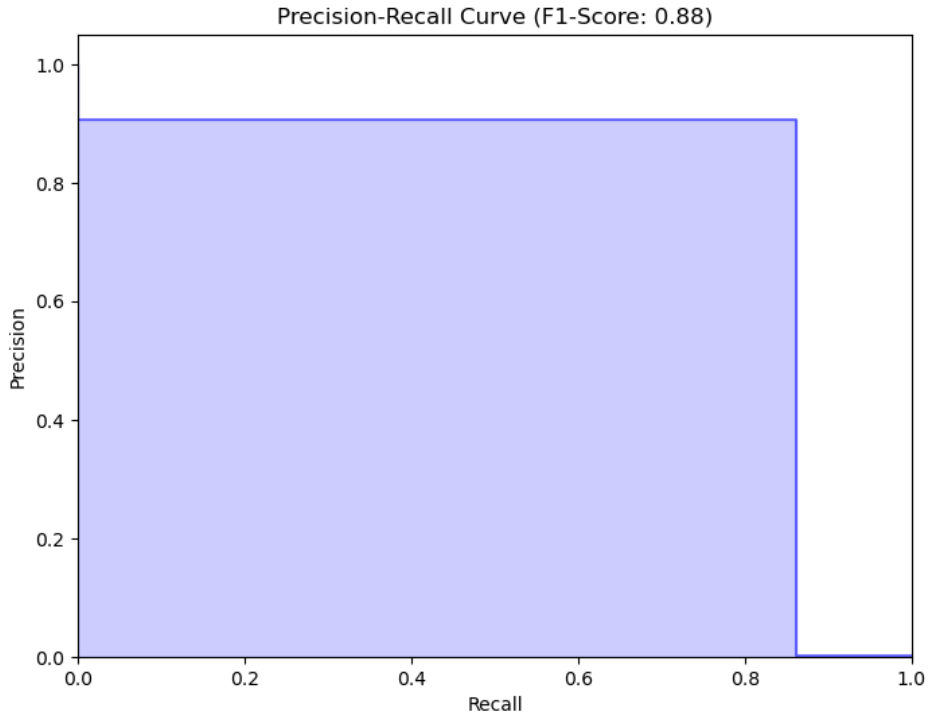

Figure 5.1: ConvRNN Precision-Recall Curve

determining where there is likely to be no flood, and thereby is unlikely to miss out on flooding events and giving an appropriate flood warning for that region. This would be a very good result in the real world as it would mean a model that is highly successful at warning against potential floods. However, it is important to remember that this dataset is highly imbalanced towards negative pixels, so these ratios are unsurprising when considering that the proportion of false negatives would inherently be very low. Therefore it is not entirely possible to claim that this model shows such great success based on these metrics alone. There also appears to be an 86% correct classification of positive flood pixels, leaving the remaining 14% as false positives. This means that in most cases, the model correctly predicts potential floods, which is another indicator of this model being able to correctly identify where potential flood risk exists. On the other hand, this proportion of false positives is still significant and shows that there are a number of cases where flood risk warnings could be given out in situations where there are no floods. Whilst it can be argued that false warnings are not as significant as situations where flood events are not predicted and mitigated, there would still be a significant real-world social and economic cost in cases of such false alarms. Overall these statistics imply that the model is a high-performing predictor in this specific context and image-based methodology but that there is room for improvement, especially given the fact that this data is inherently highly imbalanced.

Model	RMSE	MAE	SSIM
ConvRNN	0.1243	0.0461	0.4756

Table 5.2: Rainfall Prediction Performance Metrics

The results for the rainfall prediction modelling can be seen in Table 5.2, and are promis-

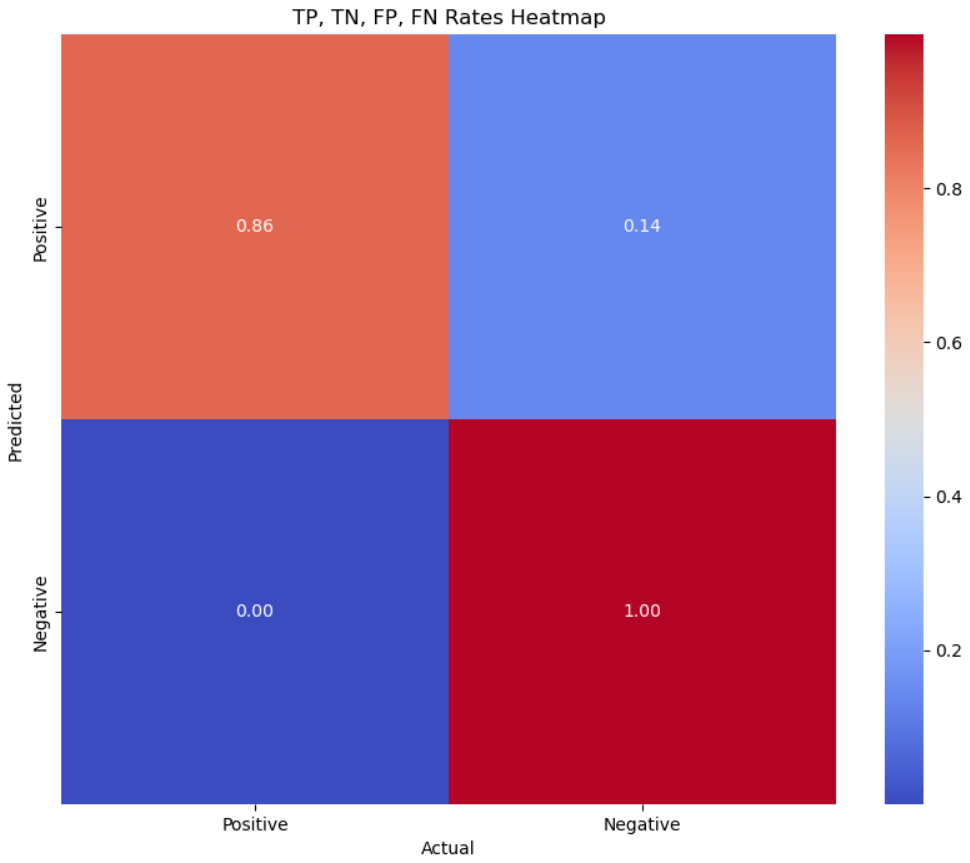

Figure 5.2: ConvRNN Confusion Matrix Heatmap Curve

ing, with the ConvRNN model appearing to accurately predict the rainfall images based on the sequenced inputs. Across the test set, there is an average RMSE of 0.1243 and an average MAE of 0.0461, which shows that there is little squared or absolute error between the predicted and true images. Additionally, there is an average SSIM of 0.4756, which suggests that the underlying quality of images generated is also reasonably good, with only a few minor differences in the produced rainfall patterns and intensities. The RMSE Heatmap depicted in Figure 5.3 shows that for a significant proportion of a given rainfall image, the errors are very low, being below 0.04 RMSE. There are only a few regions, such as a band in the north, and a larger cluster to the south of the images where errors exceed 0.1 RMSE on average. It is hard to comment on this spatial error distribution, but these areas may be harder to predict perhaps due to differences in the rainfall mechanisms in these areas that are not captured well by these image representations and deep learning modelling. However, the fact that these errors are all still low means that in practice this is not important.

5.2 Model Viability & Limitations

The results of each of the model's predictions depend on factors like architecture choices, data quality, and training procedures, where limitations could arise from data constraints, model complexity, or inherent challenges in flood prediction methodologies. Data biases, imbalances, or missing data may impact the model's performance and generalization. It must be considered

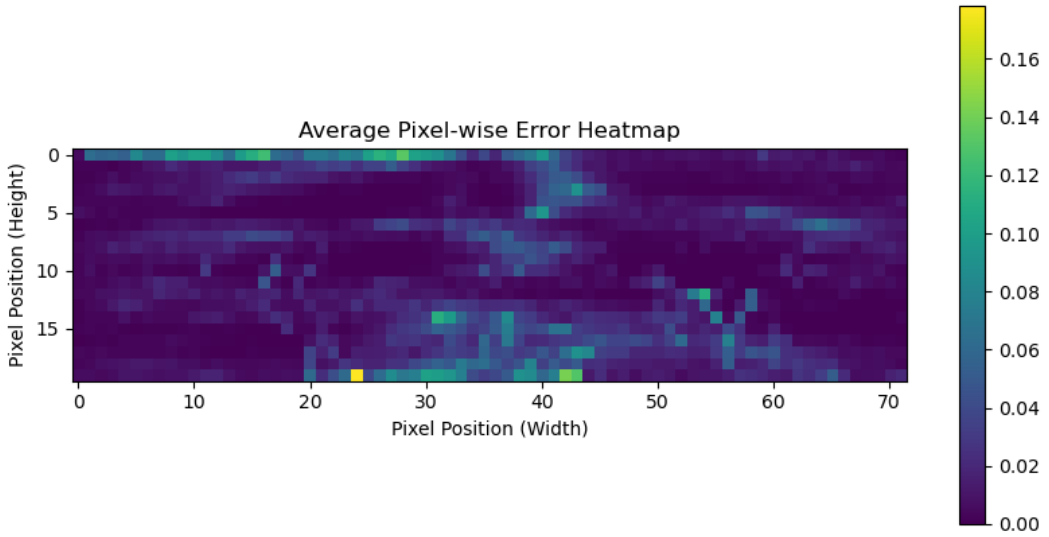

Figure 5.3: Rainfall ConvRNN RMSE Heatmap

whether there exists the necessary computational resources and infrastructure to train models that are effective at this specific flood prediction task, with there also being no potential issues present related to overfitting and generalization.

The training and validation losses have both decreased and converged over the training period for the first three models, but the predictions are entirely white images. If the binary loss between predicted and true images is being minimized by the model by generating such images, then these predictions are the ones which are favoured in training. During the model testing period, the positive weight of the loss function was adjusted in order to prioritize positive pixels. This was done at factors of 2, 3, 4, etc., as well as at a much higher value of 293.5405, which is the ratio of negative pixels to positive. Also, the focal loss method was implemented in order to take a different approach to mitigating data imbalance through the loss function. However, in all cases, the predictive scores were even worse than with no such measures, as the models generated seemingly random mixed black-and-white images with no consideration of potential flood locations. It can therefore be argued that these models are too simplistic to capture the complex spatiotemporal dynamics that underpin this problem, or that there are issues in the architecture which prevent proper learning of the data. CNN models are simply image-based, so have no concept of the time sequence, whilst RNN models have no convolutional mechanism by which to understand the relative pixel patterns of the rainfall images, so it is unsurprising that the performances are as such.

The CNN-RNN model could perhaps be expected to perform better but the method is potentially bottlenecked by either of the two component models. The CNN feature maps are directly utilized in the RNN segment of the model, and perhaps these latent spatial representations fed into LSTM layers are not able to capture the spatiotemporal trends. Furthermore, the use of MobileNet may be a bottleneck as it is a pre-trained model trained on a generic simple image set, so is not tailored specifically to rainfall spatiotemporal patterns. The features captured by the MobileNet model are general image features, so may not be relevant and useful for predicting these flood images from rainfall images. However, the fact that the accuracy scores for the RNN and CNN-RNN are identical might suggest that the RNN is the bottleneck for this problem, as it would seem unimportant whether rainfall images or latent representations are

provided to the model. The ConvRNN model seeming to perform well in rainfall prediction is not surprising, due to its common use in nowcasting as a result of the strong spatiotemporal capabilities present. However, the good performance of ConvRNN in the flood prediction scenario is a novel result, as previous studies do not seem to use the method effectively in such downscaled image-to-image predictions. There are of course caveats and limitations to this result as discussed previously, and an in-depth breakdown of the different predictive scenarios is required to critically analyze these results more effectively, which will be undertaken subsequently.

Visualizing some of the various model predictions and comparing them to the ground truth flood images allows for the identification of areas where the model struggles, such as misclassifications in image data or inaccuracies in regression predictions. It is important to examine specific instances where the model performs exceptionally well or poorly in different scenarios. This means an assessment can be made whether or not there are common patterns or challenges that determine whether a sequenced flood prediction is accurate or inaccurate. In order to further analyse the spread of predictions for both ConvRNN analyses, the best, middle and worst scoring predictions, are determined according to f1-score or RMSE respectively. These are plotted along with the sequences of rainfall and flood or rainfall inputs, in order to better understand each of the different scenarios.

Figure 5.4 shows the best prediction for the flood ConvRNN model. The rainfall trends do not seem to shift much between Days 1 and 3, but on Day 4 there is a reduction of rainfall in southern regions and an increase in the northern regions. This is followed up by an even bigger shift on Day 5, where there is a further rainfall increase in the north, as well as further emerging increases in the east and west. In terms of the flooding images, the floods remain in the south of the images across this period, and their positions remain similar, with identical flood patterns on Days 2 and 3. The Day 5 ground truth is identical to the Day 4 pattern of floods, and the model manages to predict this image perfectly, despite the changing patterns in rainfall over the previous couple of Days. It is interesting that there are no emerging floods in the north over this sequence, despite the increase in rainfall, with the model also not attempting to predict any floods in this area.

In Figure 5.5, the middle scoring flood ConvRNN prediction is depicted, and again, this is a perfectly scoring example as this model has performed very well in a significant proportion of the scenarios. Across the 5-day sequence, the rainfall patterns appear generally the same, with a few shifts in position in the lower rainfall bands across the image and a slight shift in the position of the high rainfall portions of the southern region. Flood images 1 and 2 show a cluster of floods in this same region of high rainfall, as well as two smaller clusters to the west and northeast. However, on Days 3 and 4 these smaller flood clusters are no longer present, leaving only the main southern cluster. The Day 5 ground truth image is identical to the 2 prior Days, and the model correctly predicts this cluster of floods. In this scenario, it appears that the high rainfall in the southern region over a number of Days has led to continuous flooding over the sequence period.

Finally, in Figure 5.6, the worst scoring flood ConvRNN prediction can be seen. In this sequence, there is a large spread of rainfall across roughly the entire eastern side of the image, with a similar amount of rainfall from Days 1 to 3, and a slight increase in intensity during Days 4 and 5. The flood patterns are harder to interpret, with one southwestern and two southeastern floods on Days 1 and 2 and then no floods appearing on Day 3, followed up by one of the eastern floods reappearing on Day 4. The Day 5 ground truth sees an additional flood region appear to the south of the image alongside the existing southeastern flood. However, the prediction is of no floods across the entire image, giving an f1-score of 0.0 for this scenario. It is difficult to comment on why this scenario has resulted in a poor prediction, but perhaps the large spread

of rainfall along with the unusual flood sequence has made this a hard situation to analyse in a spatiotemporal sense. Overall, it seems that the ConvRNN model struggles in some scenarios, perhaps where the rainfall patterns are more mixed and the flood sequence is harder to predict, but on the whole the model appears to generalize well to different scenarios.

The best-scoring ConvRNN rainfall sequence shown in Figure 5.7 depicts a region of high rainfall in the south and southeast. There are also some more spread-out patterns to the west and northeast of the images, and these patterns remain generally similar across the first three Days. On Day 4, there appears to be a reduction in the intensity of all the bands of rainfall, made even more clear in the ground truth image for Day 5, where most of the high rainfall has disappeared aside from a small area in the southeast. Surprisingly, the model manages to accurately predict this rainfall image despite the quick changes in the global patterns, and an RMSE score of 0.0694 is derived.

Figure 5.8 contains the middle-scoring rainfall ConvRNN prediction, for which the rainfall patterns seem to shift very dramatically. Day 1 shows bands of high rainfall in the west of the image, which becomes more intense on Day 2, especially in the south of the image. By Day 3 the rainfall seems to become more present in the centre of the image and less so in the west, with the same central bands appearing but at a lower intensity. This is followed by a shift slightly in an eastwards direction as well as an increased intensity on Day 4. The ground truth for Day 5 shows a great reduction in the rainfall in these central bands, and a new emerging band of rainfall appearing in the western region, a big shift as compared to the previous Days. It may be expected that the model could have an issue predicting this image given the previous sequence, but the patterns are matched very accurately in the prediction, which is reflected in the RMSE score of 0.1236.

In the worst scoring scenario shown by Figure 5.9, the first image contains a few horizontal rainfall bands, mainly to the west, along with some that are smaller in the east. Days 2 and 3 show similar intensity of rainfall in the west, but these western bands disappear increasingly over time, although Day 4 does not show much significant change. However, the 5th-day ground truth image shows a significant increase in intensity in these western bands of rainfall. This image is nonetheless accurately captured by the modelled prediction, even given the fact that this is the worst-scoring scenario with an RMSE score of 0.2005. This score is not too far in magnitude from the best RMSE score of 0.0694. Overall, this particular analysis appears to be successful, with all of the predictions in a range of low RMSE scores that indicate this model generalizes well. The model appears able to predict the next day of rainfall very accurately based on ConvLSTM and a sequence of prior rainfall images.

It appears that the ConvRNN for rainfall prediction has generalized well to all possible scenarios of recent rainfall trends and next-day prediction, although this needs to be implemented at a much greater spatial scale for real-world use. The ConvRNN analysis, despite the existing limitations that would affect the usefulness of deploying the model in its current state, seems to have performed well in most scenarios. Upon closer reflection on the spread of the predictions, it does seem that the model performs better in instances where it has predicted the same flood image as the cay prior. Whilst this may not matter if the performance is high, it must still be mentioned, as it is contingent on the splitting of each flood event into several in sequence for the sake of the spatiotemporal image analysis. It could be argued when looking at the best and middle-performing scenarios that the model is only performing well at predicting the continuation of existing floods. It must therefore be established whether the model is able to correctly predict novel floods based on the rainfall input. The fact that the model also takes the previous flood images as input means it can be questioned whether the model is taking into account more of the flood sequence in its prediction as compared to the rainfall sequence trends.

However, it is important that the model is able to see some of the flood patterns in its input, in order to make use of the history of flood events. This is utilized alongside the rainfall history in order to establish a predictive link between the two and better capture the underlying dynamics at play. It could be argued that it does not matter which input contributes more to a potential prediction if the performance is high. Deep learning models have an inherent black-box nature, meaning a high-performing model can hypothetically be praised no matter what input modalities are explored in training. Furthermore, the fact that the model does not always learn to predict the same flood image as the day prior suggests that it has a higher level of understanding of the link between rainfall and flooding. This indicates the model is in fact building a strong link between the two such that predictions are dynamic, perhaps only faltering where the flood sequence is unstable such as in the worst-performing example. A full analysis of all of the different predictive performances is beyond the scope of this project. However, the fact that the overall performance statistics are high across the test set favours this model being a high performer, especially given the true positive prediction rate being so high.

The viability of these methods in the real world is dependent on several factors, such as computational resources, data requirements, and model training time. The ease of deployment and maintenance in global flood prediction scenarios is directly altered by the adaptability of the deep learning models to different geographic regions and scales. The feasibility of deploying deep-learning flood prediction models in such production environments is dependent on the scalability, latency, and time-based requirements. There is also the need for periodic updates and retraining to maintain model accuracy over the long term. Gathering feedback from users and risk mitigation strategists to determine whether the solution meets their needs and expectations for accurately predicting flooding events is key here. It is important to assess whether the benefits gained from using these predictive models outweigh the costs in terms of data, resources, and infrastructure.

It does not seem that CNN and RNN models are spatiotemporally powerful enough alone for real-world global flood mitigating predictions such as this. However, CNN-RNN has potential perhaps if a custom CNN model was utilized along with a more refined RNN component. Building and training a custom CNN is computationally intensive but could be tailored to the specifics of this flood prediction task. This would especially be supported if there were no computational constraints limiting the use of auto-parameterization to optimize the solution. Convolutional RNN for rainfall prediction is well documented, but the global image method used here would have to be scaled to higher resolutions in order to determine its potential usefulness. Rainfall prediction at this resolution of 72x20 is not viable for real-world nowcasting, as much greater precision and number of data points are needed to forecast different weather zones as required. Convolutional RNN flood prediction, even with the promising results shown here, would be at a scale such that potential floods could only be identified to lie within large portions of an individual country or countries. Whilst the relatively high model predictive power can act as a first step in developing a potential real-world model, the resolution of 72x20 is not high enough to meet the necessary requirements for deployment. A certain level of accuracy is needed for the optimal distribution of resources and localized management strategies in cases of flood occurrences. Potentially, a more computationally intensive version of this analysis must be undertaken at a resolution appropriate to give more accurate predictions, although this does not necessarily have to be at the original data resolution of 1440x400.

One great assumption in this analysis comes from the fact that real-world floods are highly complex in size and shape. This means that reducing them to a single square pixel is not appropriate if trying to create a model with true spatial interpretability. Ideally, the predictions would be dynamic and able to generate flood pixel regions that represent the true shape and extent of

an event. However, this would potentially require the use of a much more rigorous dataset than the DFO archive, which is itself an estimated set of data from satellite observations. As both the TRMM and DFO data are remote sensing estimates, their limited data coverage cannot be used alone to create a highly dynamic flood prediction model, and should ideally be paired with other resources and predictive techniques to give the best outcome. In order to conduct research on rainfall and flooding over a period of many years, such that predictive models can cover the entire spatiotemporal range of weather and extreme event trends, the full scope of rainfall and flooding coverage must be incorporated into modelling. The underlying problem here comes when trying to further globalize the scale of flood modelling. Available data becomes increasingly scarce as one moves further away from the highly focused localized hazard mapping and monitoring efforts. Perhaps in the future, efforts could be undertaken to unify and verify many local flood datasets with more large-scale efforts such as the DFO archive in order to bolster the data coverage and spatiotemporal potential. This would mean that predictive models could have a wealth of data from which to learn the complex spatiotemporal nature of extreme flooding events.

Another limitation that arises in this methodology is in the temporal assumption that there would be rainfall monitoring data available up until the moment of flood prediction, in order to allow for instantaneous or very quick predictions. In the flooding analysis, this study provides Days 1-5 of rainfall, alongside Days 1-4 of flooding, to predict flooding on Day 5. It must therefore be considered if such rainfall and flooding data would be readily available to enable the next day of prediction. Furthermore, if floods are already present in a given location, it may not then be useful to predict the same flood location again in the outputted image, as local authorities would already be privy to such events. This originates in the fact that the DFO data has been simplified for this sequence-based analysis by treating each date period for a flood as a sequence of several floods, one for each day within the period. Whilst this has allowed for a direct image-to-image analysis between the rainfall and flood image sequences, it means that the potential real-world use of these models is put into question, until these spatiotemporal implications can be addressed. Deep learning methods require very consistent and rigorous inputs in order to work, but the real geophysical and hydrological nature of flooding is complex. Flooding events perhaps cannot be represented within a single input, whether an image or any other representation. Maybe this study can then only serve as the initial foundation for further multi-modal techniques to follow, which integrate more of the underlying nature of flood scenarios to bolster the application potential in real-world settings.

The fact that the flood target images are binary has the effect of reducing the highly complex nature of flood risk to a simple yes/no formula. Ideally, the potential flood risk would be able to be gauged by a continuous dynamic flood risk score that quantifies the risk level of probability of a flood in order for mitigating measures can be made. Rounding to binary values discards the model's confidence in its predictions, making it difficult to assess prediction certainty. This is a limitation of the data and use of deep learning methodologies, as a consistent style and makeup of the target variable is required. The BCE loss function would not be able to iteratively improve on the predictions when compared to continuous images. Furthermore, interpreting varying prediction values can be more complex, especially for non-experts. The DFO flood archive data can only quantify the history of estimated flood events that have taken place rather than giving information on regions that have been at risk of flooding. Perhaps a period of focused data gathering and preparation would be favoured in order to produce a set of data that quantifies locational risk factors based on the underlying geological, topographical and societal infrastructural features. This would allow the training of deep learning models which could automatically detect at-risk regions based on both non-shifting factors and turbulent

global weather patterns. This would allow for an improvement in flood predictive power in a very localised sense despite the goal of global scale modelling.

To conclude the analysis of these results, despite these different considerations and underlying assumptions, it appears fair to say that this ConvRNN rainfall and flood prediction methodology shows reasonable success, even where the CNN, RNN and CNN-RNN models have performed poorly. Although this is within the limited confines placed on such a down-scaled image-to-image predictive scenario, it can act as a promising foundation from which further models can incorporate higher resolution multi-modal elements on top of the existing image-based representation alongside other novel techniques.

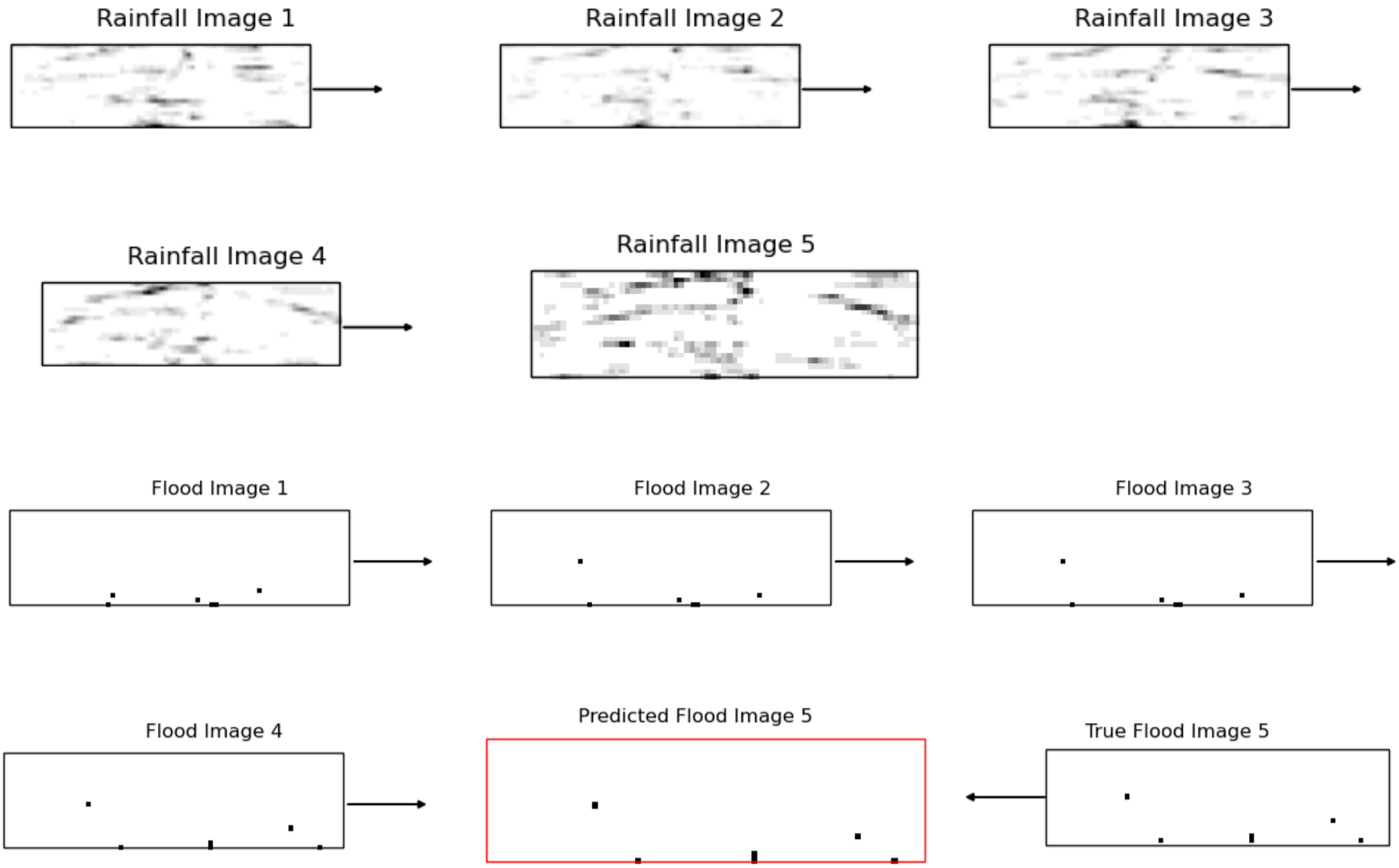

Figure 5.4: ConvRNN Best Prediction Sequence

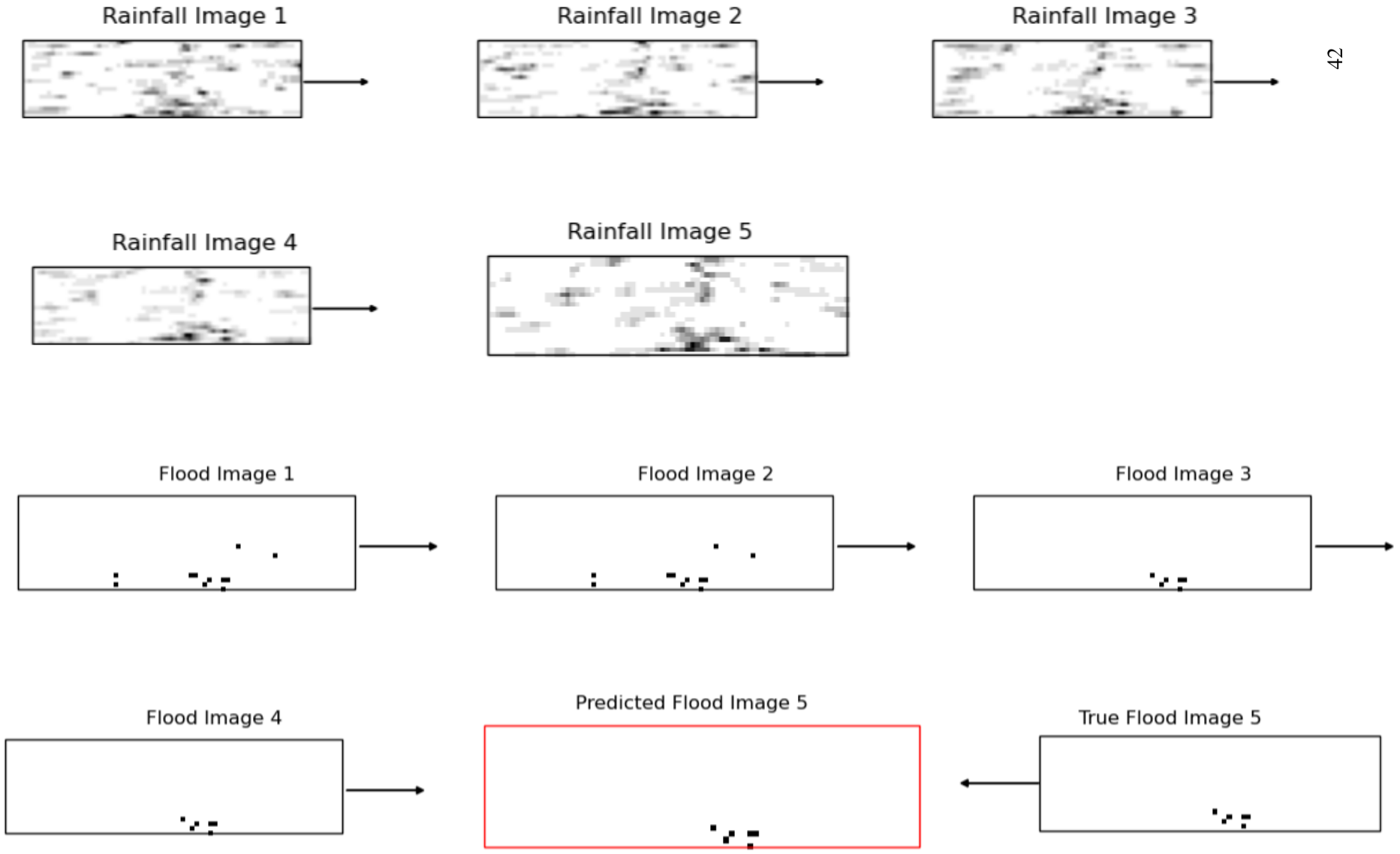

Figure 5.5: ConvRNN Middle Prediction Sequence

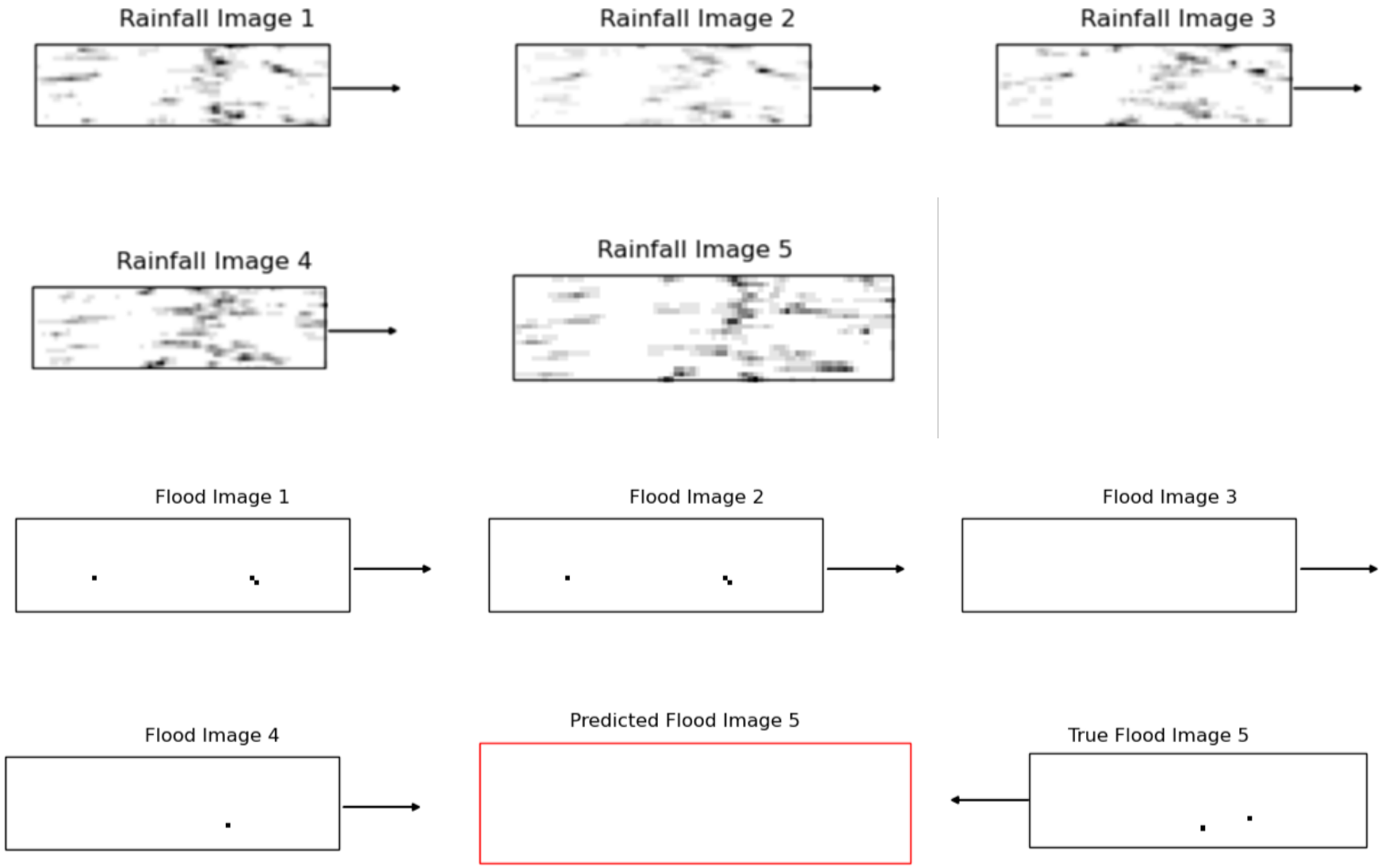

Figure 5.6: ConvRNN Worst Prediction Sequence

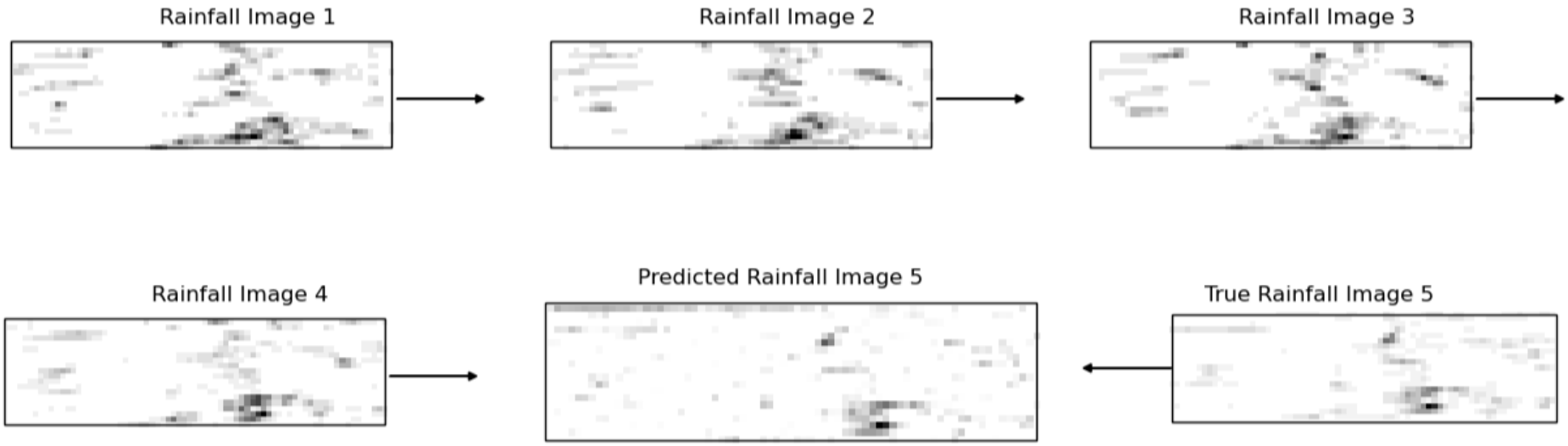

Figure 5.7: Rainfall ConvRNN Best Prediction Sequence

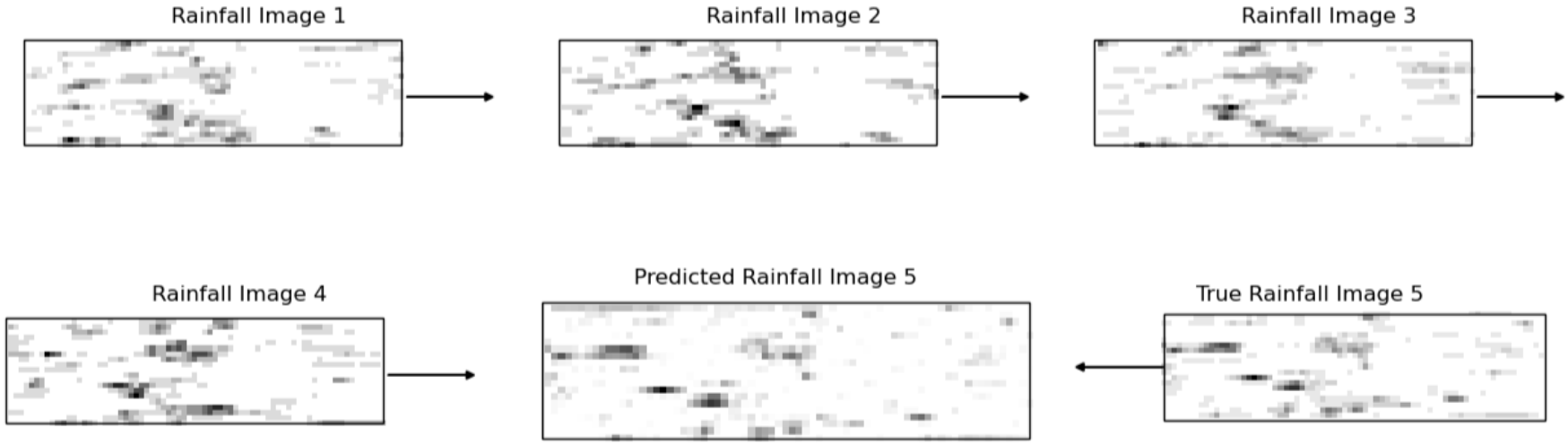

Figure 5.8: Rainfall ConvRNN Middle Prediction Sequence

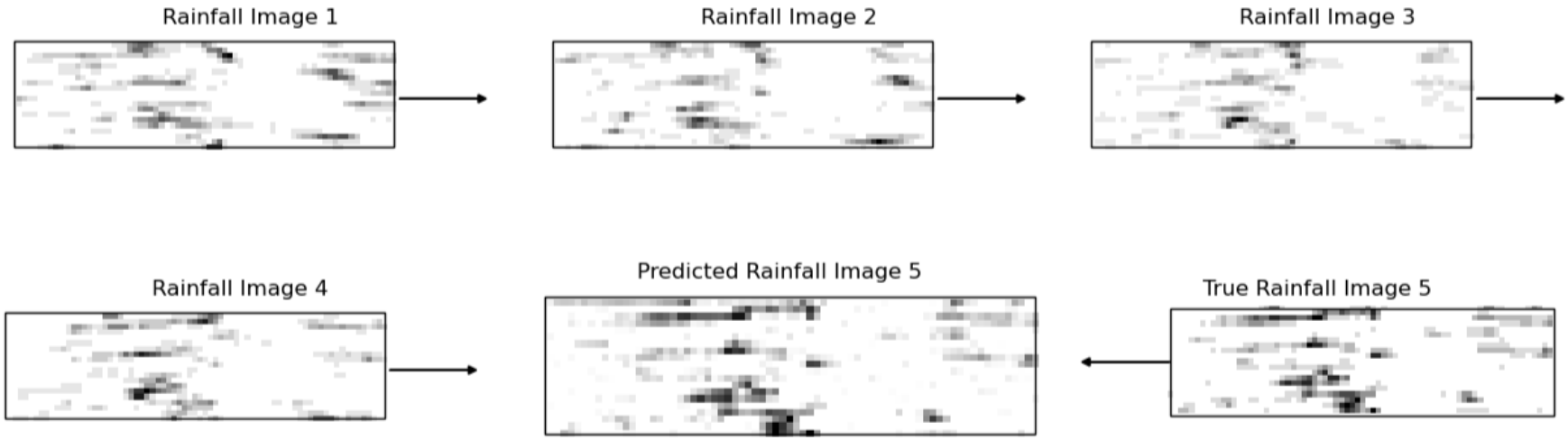

Figure 5.9: Rainfall ConvRNN Worst Prediction Sequence

Chapter 6

Discussion

6.1 Improving the Method

For this image-based study specifically, it may be useful to incorporate more meteorological data, such as temperature, hydrological data, like river flow and water levels, or topographic data, including slope and land cover. This would allow for more variables from which the deep learning models can learn the determining variables for flooding events. The great imbalance present in this dataset could perhaps be mitigated through the use of techniques such as resampling to oversample the minority class and balance the class distribution or data augmentation techniques to generate synthetic examples, although this would be difficult given the use of an existing flood archive dataset. These methods could consider this an anomaly detection problem and use specialized algorithms to identify instances deviating from the majority no-flood class, although this would greatly complicate the technical aspects of the analysis over the image-based analysis undertaken here. Instead of treating each entire image as a single data point, the images could potentially be divided into smaller or overlapping patches of different areas of flood regions. These could be trained to give more examples of flood regions and address the data imbalance, although doing this would complicate the analysis of the spatiotemporal sequence. Autoencoders are unsupervised learning models that can learn a compressed representation of the input data. An autoencoder version of the flood prediction models could be implemented to generate fake flood images from rainfall images in order to try to fool the model and thereby learn the representations. Convolutional and encoder-decoder models have been utilized to improve coverage of data-sparse zones of the TRMM data, which would allow for a more robust input sequence to base the flood prediction analysis on [96]. However, in the case of this study it would seem to overcomplicate the sequenced rainfall problem and would potentially create noisy images, as the autoencoder's objective is focused on self-generation rather than predicting floods.

There are several strategies that can bolster the method's robustness and adaptability for real-world use, including considerations for incorporating real-time data. The models could be made more dynamic by integrating additional environmental factors and leveraging advancements in artificial intelligence and machine learning. There is also the potential for combining rainfall and flood prediction models into an integrated system to provide more comprehensive and accurate flood forecasts. However, this would require extensive data coverage of the underlying spatiotemporal dynamics underpinning the link between the two. It is important to ensure the collection and incorporation of more diverse and representative data sources, with strong data quality control measures to filter out problematic and noisy data points. Utilizing real-time

satellite and weather data feeds would ensure high-quality and time-relevant data is being used for predictions and to calibrate the models effectively to a given situation, although it would pose a challenge to incorporate such methods into this image-based study.

Scaling up the method spatially would involve a greater computational burden due to the processing of higher-resolution images. The use of global GPM rainfall data rather than tropical TRMM data could increase the coverage to all potential flood risk zones, although this is limited by the much shorter set of data to train on. Increasing the temporal coverage to the 3-hourly data product would be a lot more intensive to process but allow for a much larger dataset to utilize. This would potentially reduce the risk of overfitting and allow for better learning of the long and short-term changes in rainfall and the resulting impact on flood images. However, with the current limitations of the DFO data and the method of converting it to being image-based, 3-hourly data would have to be paired with a much richer flood dataset. This would need to be able to quantify flood risk and temporal extent more precisely than at the daily scale, which would again contribute to high computational costs. The method could also be made "smarter" through the identification of areas already prone to flood risk. GIS data or existing flood hazard knowledge could be used to map out zones of economic and geophysical risk, allowing for the focusing of computational resources on the real-time data feed for these locations. This would require an extensive period of data management and filtering to create inputs suitable for deep learning analysis.

The rainfall data could also be dynamically downscaled and upscaled to generate high-resolution data in the required locations. Deep learning models have been shown to effectively generate higher-resolution upscaled rainfall data than methods such as linear interpolation, allowing for greatly improved distribution of predictive resources to the most at-risk locations [97]. Another way to dynamically focus on certain spatial regions for flood risk modelling would be through the web scraping of current social media images and metadata to extract visual and semantic features congruent with active flood events [98]. This would allow for the focus of attention on prediction in certain regions, as well as potentially giving an additional warning sign in a dynamic ensemble predictive system, also perhaps incorporating AI-generated virtual scenarios to increase model generalization. However, these techniques are again complex, and perhaps would see better use in a more complex representation than in this image-based methodology.

6.2 Future Flood Predictions

There is also the possibility of incorporating this method or others into an ensemble of forecasting systems to enhance the overall combined accuracy and reliability of flood predictions in various scenarios, by leveraging their collective predictive power. Ensemble models are often used to combine physics-based models with traditional forecasting methods, and deep learning of NWP ensembles has been shown to have high predictive power for heavy rainfall events [99]. This could be alongside the incorporation of domain-specific knowledge, physical models, or hydrological models like rainfall-runoff simulations to complement the predictions. Historical flooding data has often been incorporated into GIS models in order to increase the spatial and visual representation power of machine learning flood predictions [100]. CNN-LSTM has been used previously to predict spatiotemporal variations in groundwater storage accurately from remote sensing data, which could feed into a more complex hydrogeological flood prediction model [101]. LSTM networks have also been used in predictions of both snowmelt and rainfall-induced flooding events, showing that flood predictions using the range of hydrological

processes are necessary for a global flood prediction effort [102]. Incremental and transfer learning approaches allow models to adapt to changing conditions and incoming data, and to adapt pre-trained models to these specific flood prediction tasks. The use of attention mechanisms in models to focus on relevant features or periods, allows them to be more adaptive and dynamic. This works well in tandem with dynamic thresholding mechanisms that adjust regional flood risk warnings based on current conditions, river levels or recent weather patterns. Graph Neural Networks could also utilize graph-structured data representing hydrological networks and interactions between water bodies in order to better capture the underlying feedback processes.

It seems there is a need to develop complex four-dimensional models of the entire globe such that at-risk locations could be identified and then the data stream coverage of rainfall and other climatic variables for such locations could be increased. This could be integrated with existing hydrological models to predict flood risk at these locations. This appears a very difficult feat to accomplish both in terms of computational and engineering requirements, and it is hard to comment on how deep learning would be incorporated into such a model. Contrastingly, it is often the simplest solution which can have the most impact when deployed at scale in the real world. Future models may incorporate 3D atmospheric models, satellite data with higher spatio-temporal resolutions, and multi-sensor data to try to better capture the complex nature of global hydrodynamics and weather. Perhaps additional climatic variables, GIS datasets and flooding representations could be introduced alongside the initial image representations utilized in this study in order to determine the characteristics best defining a potential flood prediction. This image-to-image could act as the middle segment of a complex multi-layer model. Layers of rainfall, weather and hydrogeology knowledge could be built into the start of the pipeline to determine the precipitative or water input to the network, and feed into the simple image representations. Then the flood pixel output could be dynamically upscaled to more complex representations that feed directly into real-time risk mapping and decision-making schemes. Utilizing a simpler image-based representation alongside more powerful multi-dimensional or multimodal models could be key to developing truly global flood prediction models.

When ignoring the high computational costs, there is a great potential for the use of distributed computing, cloud resources, and specialized hardware like Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), to overcome the limitations of the traditional deep learning hierarchies. It is pertinent to aim to optimize model deployment for cloud-based or edge computing environments, such that predictions can be accessible in real-time, or use HPC systems for parallel processing, which can significantly reduce model training time. On the other hand, increasingly complex models with a large number of parameters and data points can require extensive computational resources and time for training and deployment. It is important to consider the trade-off between computational complexity and real-world model performance and how hardware advancements can influence the scalability of flood prediction systems. This is particularly true when models are implemented across large geographic regions or multiple river basins. Especially in real-time applications, such as flood prediction and early warning systems, models may not meet the stringent latency requirements where low-latency inference is crucial. HPC systems have already proven to be useful in combining an ensemble of low-resolution predictions into a higher-resolution representation of weather forecasting. This potentially gives a way for image-based analyses to act as an intermediary for linking complex risk management and hydrological models [103]. 3D convolutional and ConvLSTM neural networks have been used previously to learn high-level spatiotemporal features from 2D feature maps, which could allow for more complex weather representations to be embedded in the less computationally taxing intermediate stages [104]. 3D GRU has also been used in high-resolution rainfall forecasting from radar echo maps, and perhaps this could see an implementation in a

flood prediction context [105]. This could make use of a similar encoder-decoder methodology to capture the 3D spatiotemporal features. Overall, there are many different ways in which a global and powerful flood prediction model could be constructed, but it remains to be seen if it these methods are suitable for helpful real-world usage and deployment. It could be argued that localised solutions with greater data coverage are more potent in risk mitigation, although a global model would allow for countries and regions with poor strategies in place to have flood hazard risk prevention also take place.

6.3 Concluding Remarks

Investigating the literature has shown that Deep learning models offer promising results in flood prediction tasks and enhanced accuracy compared to traditional methods, especially with the global access to large and diverse datasets which are crucial for training such models and capturing the intricate spatial and temporal patterns. However, deep learning may not always be the best or only solution, and the choice of technique should be dependent on the specific context and consider factors such as data availability, computational resources, and scenario-specific challenges. This can ensure that data is accessible to all parties and that all ethical considerations are met [106]. It must be considered whether a globalised solution with large covererage to predict short term instances of flooding, or localised solutions in already flood prone areas is more effective long-term in managing flood risk. A combination of deep learning with ensemble and interdisciplinary approaches is likely to be the most effective way forward, offering both accuracy and interpretability in flood prediction systems. It would also take yet further advancements in the field of deep learning before numerical weather models are made entirely obsolete [107].

Deep-learning flood predictions should be able to provide early warnings to communities and authorities, in order to trigger timely responses, such as evacuations and flood defences. This can enhance decision-making for disaster response agencies, local governments, and emergency services. The results of this study should therefore ideally feed directly into applications in disaster preparedness, response, and recovery efforts. It appears that whilst the ConvRNN models for rainfall and flood prediction seem to show high performances, these are underpinned by limitations present in the downscaled and image-based sequence method that limit their use in real-world scenarios. However, these models can act as a building block from which additional layers of complexity can be incorporated, such as by acting as an intermediate stage within a fully globalised and high-powered flood hazard model that considers all of the underlying geophysical and societal factors. Future efforts can aim to utilize additional climatic and topographical variables in a multi-modal dynamic analysis or make use of more powerful computational systems to try to encapsulate the true spatiotemporal extent of global rainfall and flooding. This can be implemented alongside the already active usage of deep learning prediction at a localised scale in areas where risk management is in place, to ensure that the destruction and negative effects of floods is best mitigated.

Bibliography

[1] P. C. D. Milly, R. T. Wetherald, K. A. Dunne, and T. L. Delworth, "Increasing risk of great floods in a changing climate," vol. 415, 2002.

[2] J. D. Woodruff, J. L. Irish, and S. J. Camargo, "Coastal flooding by tropical cyclones and sea-level rise," *Nature*, vol. 504, pp. 44–52, Dec. 2013.

[3] Z. W. Kundzewicz, S. Kanae, S. I. Seneviratne, J. Handmer, N. Nicholls, P. Peduzzi, R. Mechler, L. M. Bouwer, N. Arnell, K. Mach, R. Muir-Wood, G. R. Brakenridge, W. Kron, G. Benito, Y. Honda, K. Takahashi, and B. Sherstyukov, "Flood risk and climate change: global and regional perspectives," *Hydrological Sciences Journal*, vol. 59, pp. 1–28, Jan. 2014.

[4] R. S. Schumacher, "Heavy Rainfall and Flash Flooding," in *Oxford Research Encyclopedia of Natural Hazard Science*, Oxford University Press, May 2017.

[5] S. N. Jonkman, "Global Perspectives on Loss of Human Life Caused by Floods," *Natural Hazards*, vol. 34, pp. 151–175, Feb. 2005.

[6] Y. Hirabayashi and S. Kanae, "First estimate of the future global population at risk of flooding," *Hydrological Research Letters*, vol. 3, pp. 6–9, 2009.

[7] Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae, "Global flood risk under climate change," *Nature Climate Change*, vol. 3, pp. 816–821, Sept. 2013.

[8] R. Eccles, H. Zhang, and D. Hamilton, "A review of the effects of climate change on riverine flooding in subtropical and tropical regions," *Journal of Water and Climate Change*, vol. 10, pp. 687–707, Dec. 2019.

[9] H. C. Winsemius, J. Aerts, L. Van Beek, M. Bierkens, A. Bouwman, B. Jongman, J. Kwadijk, W. Ligtvoet, P. Lucas, D. van Vuuren, and P. Ward, "Global drivers of future river flood risk," *Nature Climate Change*, vol. 6, pp. 381–385, Apr. 2016.

[10] T. G. Huntington, "Evidence for intensification of the global water cycle: Review and synthesis," *Journal of Hydrology*, vol. 319, pp. 83–95, Mar. 2006.

[11] C. Wasko and R. Nathan, "Influence of changes in rainfall and soil moisture on trends in flooding," *Journal of Hydrology*, vol. 575, pp. 432–441, Aug. 2019.

[12] T. E. Ologunorisa and M. J. Abawua, "Flood Risk Assessment - A Review.pdf," 2005.

[13] P. J. Ward, B. Jongman, P. Salamon, A. Simpson, P. Bates, T. De Groeve, S. Muis, E. C. De Perez, R. Rudari, M. A. Trigg, and H. C. Winsemius, "Usefulness and limitations of global flood risk models," *Nature Climate Change*, vol. 5, pp. 712–715, Aug. 2015.

[14] K.-J. Douben, "Characteristics of river floods and flooding: a global overview, 1985–2003," *Irrigation and Drainage*, vol. 55, pp. S9–S21, July 2006.

[15] B. Jongman, P. J. Ward, and J. C. Aerts, "Global exposure to river and coastal flooding: Long term trends and changes," *Global Environmental Change*, vol. 22, pp. 823–835, Oct. 2012.

[16] H. Apel, A. H. Thieken, B. Merz, and G. Blöschl, "Flood risk assessment and associated uncertainty," *Natural Hazards and Earth System Sciences*, vol. 4, pp. 295–308, Apr. 2004.

[17] E. C. Barrett and M. J. Beaumont, "Satellite rainfall monitoring: An overview," *Remote Sensing Reviews*, vol. 11, pp. 23–48, Oct. 1994.

[18] C. Kidd, "Satellite rainfall climatology: a review," *International Journal of Climatology*, vol. 21, pp. 1041–1066, July 2001.

[19] K. Yan, G. Di Baldassarre, D. P. Solomatine, and G. J.-P. Schumann, "A review of low-cost space-borne data for flood modelling: topography, flood extent and water level: A REVIEW OF LOW-COST SPACE-BORNE DATA FOR FLOOD MODELLING," *Hydrological Processes*, vol. 29, pp. 3368–3387, July 2015.

[20] J. Simpson, C. Kummerow, W. K. Tao, and R. F. Adler, "On the Tropical Rainfall Measuring Mission (TRMM)," *Meteorology and Atmospheric Physics*, vol. 60, no. 1-3, pp. 19–36, 1996.

[21] C. Kummerow, W. Barnes, T. Kozu, J. Shiue, and J. Simpson, "The Tropical Rainfall Measuring Mission (TRMM) Sensor Package," *Journal of Atmospheric and Oceanic Technology*, vol. 15, pp. 809–817, June 1998.

[22] T. Kozu, T. Kawanishi, H. Kuroiwa, M. Kojima, K. Oikawa, H. Kumagai, K. Okamoto, M. Okumura, H. Nakatsuka, and K. Nishikawa, "Development of precipitation radar onboard the Tropical Rainfall Measuring Mission (TRMM) satellite," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 102–116, Jan. 2001.

[23] C. Kummerow, J. Simpson, O. Thiele, W. Barnes, A. T. C. Chang, E. Stocker, R. F. Adler, A. Hou, R. Kakar, F. Wentz, P. Ashcroft, T. Kozu, Y. Hong, K. Okamoto, T. Iguchi, H. Kuroiwa, E. Im, Z. Haddad, G. Huffman, B. Ferrier, W. S. Olson, E. Zipser, E. A. Smith, T. T. Wilheit, G. North, T. Krishnamurti, and K. Nakamura, "The Status of the Tropical Rainfall Measuring Mission (TRMM) after Two Years in Orbit," *Journal of Applied Meteorology*, vol. 39, pp. 1965–1982, Dec. 2000.

[24] M. Gupta, P. K. Srivastava, T. Islam, and A. M. B. Ishak, "Evaluation of TRMM rainfall for soil moisture prediction in a subtropical climate," *Environmental Earth Sciences*, vol. 71, pp. 4421–4431, May 2014.

[25] H.-K. Cho, K. P. Bowman, and G. R. North, "A Comparison of Gamma and Lognormal Distributions for Characterizing Satellite Rain Rates from the Tropical Rainfall Measuring Mission," *Journal of Applied Meteorology*, vol. 43, pp. 1586–1597, Nov. 2004.

[26] Z. Kugler, "The Global Flood Detection System," 2007.

[27] G. Schumann, G. Brakenridge, A. Kettner, R. Kashif, and E. Niebuhr, "Assisting Flood Disaster Response with Earth Observation Data and Products: A Critical Assessment," *Remote Sensing*, vol. 10, p. 1230, Aug. 2018.

[28] S. R. Proud, R. Fensholt, L. V. Rasmussen, and I. Sandholt, "Rapid response flood detection using the MSG geostationary satellite," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, pp. 536–544, Aug. 2011.

[29] F. Dottori, P. Salamon, A. Bianchi, L. Alfieri, F. A. Hirpa, and L. Feyen, "Development and evaluation of a framework for global flood hazard mapping," *Advances in Water Resources*, vol. 94, pp. 87–102, Aug. 2016.

[30] C. C. Sampson, A. M. Smith, P. D. Bates, J. C. Neal, L. Alfieri, and J. E. Freer, "A high-resolution global flood hazard model: A High-Resolution Global Flood Hazard Model," *Water Resources Research*, vol. 51, pp. 7358–7381, Sept. 2015.

[31] A. H. Tanim, C. B. McRae, H. Tavakol-Davani, and E. Goharian, "Flood Detection in Urban Areas Using Satellite Imagery and Machine Learning," *Water*, vol. 14, p. 1140, Apr. 2022.

[32] E. H. Ighile, H. Shirakawa, and H. Tanikawa, "Application of GIS and Machine Learning to Predict Flood Areas in Nigeria," *Sustainability*, vol. 14, p. 5039, Apr. 2022.

[33] H. S. Munawar, A. W. Hammad, and S. T. Waller, "A review on flood management technologies related to image processing and machine learning," *Automation in Construction*, vol. 132, p. 103916, Dec. 2021.

[34] M. Motta, M. De Castro Neto, and P. Sarmento, "A mixed approach for urban flood prediction using Machine Learning and GIS," *International Journal of Disaster Risk Reduction*, vol. 56, p. 102154, Apr. 2021.

[35] H. Wu, R. F. Adler, Y. Hong, Y. Tian, and F. Policelli, "Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model," *Journal of Hydrometeorology*, vol. 13, pp. 1268–1284, Aug. 2012.

[36] P. K. Yeditha, V. Kasi, M. Rathinasamy, and A. Agarwal, "Forecasting of extreme flood events using different satellite precipitation products and wavelet-based machine learning methods," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, p. 063115, June 2020.

[37] N. Rusk, "Deep learning," *Nature Methods*, vol. 13, pp. 35–35, Jan. 2016.

[38] H. Wang and B. Raj, "On the Origin of Deep Learning," Mar. 2017. arXiv:1702.07800 [cs, stat].

[39] Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, and M. S. Nasrin, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," 2018.

[40] W. R. Moskolaï, W. Abdou, A. Dipanda, and Kolyang, "Application of Deep Learning Architectures for Satellite Image Time Series Prediction: A Review," *Remote Sensing*, vol. 13, p. 4822, Nov. 2021.

[41] Y. Zhou, H. Dong, and A. El Saddik, "Deep Learning in Next-Frame Prediction: A Benchmark Review," *IEEE Access*, vol. 8, pp. 69273–69283, 2020.

[42] Jin Liu, Yi Pan, Min Li, Ziyue Chen, Lu Tang, Chengqian Lu, and Jianxin Wang, "Applications of deep learning to MRI images: A survey," *Big Data Mining and Analytics*, vol. 1, pp. 1–18, Mar. 2018.

[43] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical Image Synthesis with Context-Aware Generative Adversarial Networks," in *Medical Image Computing and Computer Assisted Intervention MICCAI 2017* (M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, eds.), vol. 10435, pp. 417–425, Cham: Springer International Publishing, 2017. Series Title: Lecture Notes in Computer Science.

[44] L. Chen, Y. Cao, L. Ma, and J. Zhang, "A Deep Learning-Based Methodology for Precipitation Nowcasting With Radar," *Earth and Space Science*, vol. 7, Feb. 2020.

[45] Z. Gao, X. Shi, H. Wang, D. Yeung, W. Woo, and W. Wong, "Deep Learning and the Weather Forecasting Problem: Precipitation Nowcasting," in *Deep Learning for the Earth Sciences* (G. Camps-Valls, D. Tuia, X. X. Zhu, and M. Reichstein, eds.), pp. 218–239, Wiley, 1 ed., Sept. 2021.

[46] P. Grönquist, C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, and T. Hoefler, "Deep learning for post-processing ensemble weather forecasts," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, p. 20200092, Apr. 2021.

[47] J. Cuomo and V. Chandrasekar, "Developing Deep Learning Models for Storm Nowcasting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[48] R. Shen, A. Huang, B. Li, and J. Guo, "Construction of a drought monitoring model using deep learning based on multi-source remote sensing data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 79, pp. 48–57, July 2019.

[49] J. Hussain and C. Zoremsanga, "A Survey of Rainfall Prediction Using Deep Learning," in *2021 3rd International Conference on Electrical, Control and Instrumentation Engineering (ICECIE)*, (Kuala Lumpur, Malaysia), pp. 1–10, IEEE, Nov. 2021.

[50] X. Ren, X. Li, K. Ren, J. Song, Z. Xu, K. Deng, and X. Wang, "Deep Learning-Based Weather Prediction: A Survey," *Big Data Research*, vol. 23, p. 100178, Feb. 2021.

[51] W. Fang, Q. Xue, L. Shen, and V. S. Sheng, "Survey on the Application of Deep Learning in Extreme Weather Prediction," *Atmosphere*, vol. 12, p. 661, May 2021.

[52] J. Liu, L. Xu, and N. Chen, "A spatiotemporal deep learning model ST-LSTM-SA for hourly rainfall forecasting using radar echo images," *Journal of Hydrology*, vol. 609, p. 127748, June 2022.

[53] S. P. Van, H. M. Le, D. V. Thanh, T. D. Dang, H. H. Loc, and D. T. Anh, "Deep learning convolutional neural network in rainfall–runoff modelling," *Journal of Hydroinformatics*, vol. 22, pp. 541–561, May 2020.

[54] H. Han, C. Choi, J. Jung, and H. S. Kim, "Deep Learning with Long Short Term Memory Based Sequence-to-Sequence Model for Rainfall-Runoff Simulation," *Water*, vol. 13, p. 437, Feb. 2021.

[55] J. M. Frame, F. Kratzert, D. Klotz, M. Gauch, G. Shalev, O. Gilon, L. M. Qualls, H. V. Gupta, and G. S. Nearing, "Deep learning rainfall–runoff predictions of extreme events," *Hydrology and Earth System Sciences*, vol. 26, pp. 3377–3392, July 2022.

[56] M. Panahi, A. Jaafari, A. Shirzadi, H. Shahabi, O. Rahmati, E. Omidvar, S. Lee, and D. T. Bui, "Deep learning neural networks for spatially explicit prediction of flash flood probability," *Geoscience Frontiers*, vol. 12, p. 101076, May 2021.

[57] R. Bentivoglio, E. Isufi, S. N. Jonkman, and R. Taormina, "Deep learning methods for flood mapping: a review of existing applications and future research directions," *Hydrology and Earth System Sciences*, vol. 26, pp. 4345–4378, Aug. 2022.

[58] X. Lei, W. Chen, M. Panahi, F. Falah, O. Rahmati, E. Uuemaa, Z. Kalantari, C. S. S. Ferreira, F. Rezaie, J. P. Tiefenbacher, S. Lee, and H. Bian, "Urban flood modeling using deep-learning approaches in Seoul, South Korea," *Journal of Hydrology*, vol. 601, p. 126684, Oct. 2021.

[59] S. Sankaranarayanan, M. Prabhakar, S. Satish, P. Jain, A. Ramprasad, and A. Krishnan, "Flood prediction based on weather parameters using deep learning," *Journal of Water and Climate Change*, vol. 11, pp. 1766–1783, Dec. 2020.

[60] X. Tang, Z. Yin, G. Qin, L. Guo, and H. Li, "Integration of Satellite Precipitation Data and Deep Learning for Improving Flash Flood Simulation in a Poor-Gauged Mountainous Catchment," *Remote Sensing*, vol. 13, p. 5083, Dec. 2021.

[61] H. Moon, S. Yoon, and Y. Moon, "Urban flood forecasting using a hybrid modeling approach based on a deep learning technique," *Journal of Hydroinformatics*, vol. 25, pp. 593–610, Mar. 2023.

[62] J. Simpson, R. F. Adler, and G. R. North, "A Proposed Tropical Rainfall Measuring Mission (TRMM) Satellite," vol. 69, no. 2, 1988.

[63] G. J. Huffman, D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker, "The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales," *Journal of Hydrometeorology*, vol. 8, pp. 38–55, Feb. 2007.

[64] F. Su, Y. Hong, and D. P. Lettenmaier, "Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and Its Utility in Hydrologic Prediction in the La Plata Basin," *Journal of Hydrometeorology*, vol. 9, pp. 622–640, Aug. 2008.

[65] S. Prakash, A. K. Mitra, D. Pai, and A. AghaKouchak, "From TRMM to GPM: How well can heavy rainfall be detected from space?," *Advances in Water Resources*, vol. 88, pp. 1–7, Feb. 2016.

[66] V. Maggioni, P. C. Meyers, and M. D. Robinson, "A Review of Merged High-Resolution Satellite Precipitation Product Accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era," *Journal of Hydrometeorology*, vol. 17, pp. 1101–1117, Apr. 2016.

[67] H. Chen, V. Chandrasekar, H. Tan, and R. Cifelli, "Rainfall Estimation From Ground Radar and TRMM Precipitation Radar Using Hybrid Deep Neural Networks," *Geophysical Research Letters*, vol. 46, pp. 10669–10678, Sept. 2019.

[68] H. C. Winsemius, L. P. H. Van Beek, B. Jongman, P. J. Ward, and A. Bouwman, "A framework for global river flood risk assessments," *Hydrology and Earth System Sciences*, vol. 17, pp. 1871–1892, May 2013.

[69] L. Lin, L. Di, E. G. Yu, J. Tang, R. Shrestha, M. S. Rahman, L. Kang, Z. Sun, C. Zhang, L. Hu, G. Yang, and Z. Yang, "Extract flood duration from Dartmouth Flood Observatory flood product," in *2017 6th International Conference on Agro-Geoinformatics*, (Fairfax, VA, USA), pp. 1–4, IEEE, Aug. 2017.

[70] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, pp. 6999–7019, Dec. 2022.

[71] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, May 2018.

[72] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," Dec. 2015. arXiv:1511.08458 [cs].

[73] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, pp. 611–629, Aug. 2018.

[74] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, (Antalya), pp. 1–6, IEEE, Aug. 2017.

[75] W. Wang, Y. Li, T. Zou, X. Wang, J. You, and Y. Luo, "A Novel Image Classification Approach via Dense-MobileNet Models," *Mobile Information Systems*, vol. 2020, pp. 1–8, Jan. 2020.

[76] H.-Y. Chen and C.-Y. Su, "An Enhanced Hybrid MobileNet," 2017.

[77] A. Michele, V. Colin, and D. D. Santika, "MobileNet Convolutional Neural Networks and Support Vector Machines for Palmprint Recognition," *Procedia Computer Science*, vol. 157, pp. 110–117, 2019.

[78] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, Mar. 2020.

[79] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural Computing and Applications*, vol. 31, pp. 2727–2740, July 2019.

[80] A. Shewalkar, D. Nyavanandi, and S. A. Ludwig, "Performance Evaluation of Deep Neural Networks Applied to Speech Recognition: RNN, LSTM and GRU," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 9, pp. 235–245, Oct. 2019.

[81] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, (Udupi), pp. 1643–1647, IEEE, Sept. 2017.

[82] I. E. Livieris, E. Pintelas, and P. Pintelas, "A CNN–LSTM model for gold price time-series forecasting," *Neural Computing and Applications*, vol. 32, pp. 17351–17360, Dec. 2020.

[83] W. Zha, Y. Liu, Y. Wan, R. Luo, D. Li, S. Yang, and Y. Xu, "Forecasting monthly gas field production based on the CNN-LSTM model," *Energy*, vol. 260, p. 124889, Dec. 2022.

[84] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-Based Model to Forecast Stock Prices," *Complexity*, vol. 2020, pp. 1–10, Nov. 2020.

[85] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sept. 2019.

[86] S. Khaki, L. Wang, and S. V. Archontoulis, "A CNN-RNN Framework for Crop Yield Prediction," *Frontiers in Plant Science*, vol. 10, p. 1750, Jan. 2020.

[87] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining Convolutional Neural Network With Recursive Neural Network for Blood Cell Image Classification," *IEEE Access*, vol. 6, pp. 36188–36197, 2018.

[88] L. Zhang, G. Zhu, L. Mei, P. Shen, S. A. A. Shah, and M. Bennamoun, "Attention in Convolutional LSTM for Gesture Recognition," 2018.

[89] M. Rahman, D. Islam, R. J. Mukti, and I. Saha, "A deep learning approach based on convolutional LSTM for detecting diabetes," *Computational Biology and Chemistry*, vol. 88, p. 107329, Oct. 2020.

[90] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, "Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model," 2017.

[91] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," 2015.

[92] B. Kumar, N. Abhishek, R. Chattopadhyay, S. George, B. B. Singh, A. Samanta, B. Patnaik, S. S. Gill, R. S. Nanjundiah, and M. Singh, "Deep learning based short-range forecasting of Indian summer monsoon rainfall using earth observation and ground station datasets," *Geocarto International*, vol. 37, pp. 17994–18021, Dec. 2022.

[93] Q. Yan, F. Ji, K. Miao, Q. Wu, Y. Xia, and T. Li, "Convolutional Residual-Attention: A Deep Learning Approach for Precipitation Nowcasting," *Advances in Meteorology*, vol. 2020, pp. 1–12, Feb. 2020.

[94] H. Li, J. Li, X. Guan, B. Liang, Y. Lai, and X. Luo, "Research on Overfitting of Deep Learning," in *2019 15th International Conference on Computational Intelligence and Security (CIS)*, (Macao, Macao), pp. 78–81, IEEE, Dec. 2019.

[95] S. Salman and X. Liu, "Overfitting Mechanism and Avoidance in Deep Neural Networks," Jan. 2019. arXiv:1901.06566 [cs, stat].

[96] X.-H. Le, D. H. Nguyen, and G. Lee, "Performance Comparison of Bias-Corrected Satellite Precipitation Products by Various Deep Learning Schemes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.

[97] B. Kumar, R. Chattopadhyay, M. Singh, N. Chaudhari, K. Kodari, and A. Barve, "Deep learning–based downscaling of summer monsoon rainfall data over Indian region," *Theoretical and Applied Climatology*, vol. 143, pp. 1145–1156, Feb. 2021.

[98] L. Lopez-Fuentes, J. van de Weijer, M. Bolaños, and H. Skinnemoen, "Multi-modal Deep Learning Approach for Flood Detection," 2017.

[99] P. Hess and N. Boers, "Deep Learning for Improving Numerical Weather Prediction of Heavy Rainfall," *Journal of Advances in Modeling Earth Systems*, vol. 14, p. e2021MS002765, Mar. 2022.

[100] S. Khatri, P. Kokane, V. Kumar, and S. Pawar, "Prediction of waterlogged zones under heavy rainfall conditions using machine learning and GIS tools: a case study of Mumbai," *GeoJournal*, Aug. 2022.

[101] J. Y. Seo and S.-I. Lee, "Predicting Changes in Spatiotemporal Groundwater Storage Through the Integration of Multi-Satellite Data and Deep Learning Models," *IEEE Access*, vol. 9, pp. 157571–157583, 2021.

[102] S. Jiang, Y. Zheng, C. Wang, and V. Babovic, "Uncovering Flooding Mechanisms Across the Contiguous United States Through Interpretive Deep Learning on Representative Catchments," *Water Resources Research*, vol. 58, p. e2021WR030185, Jan. 2022.

[103] E. Rocha Rodrigues, I. Oliveira, R. Cunha, and M. Netto, "DeepDownscale: A Deep Learning Strategy for High-Resolution Weather Forecast," in *2018 IEEE 14th International Conference on e-Science (e-Science)*, (Amsterdam), pp. 415–422, IEEE, Oct. 2018.

[104] L. Zhang, G. Zhu, P. Shen, J. Song, S. A. Shah, and M. Bennamoun, "Learning Spatiotemporal Features Using 3DCNN and Convolutional LSTM for Gesture Recognition," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, (Venice), pp. 3120–3128, IEEE, Oct. 2017.

[105] D. Sun, J. Wu, H. Huang, R. Wang, F. Liang, and H. Xinhua, "Prediction of Short-Time Rainfall Based on Deep Learning," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–8, Mar. 2021.

[106] V. Kumar, H. M. Azamathulla, K. V. Sharma, D. J. Mehta, and K. T. Maharaj, "The State of the Art in Deep Learning Applications, Challenges, and Future Prospects: A Comprehensive Review of Flood Forecasting and Management," *Sustainability*, vol. 15, p. 10543, July 2023.

[107] M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari, and S. Stadtler, "Can deep learning beat numerical weather prediction?," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, p. 20200097, Apr. 2021.