Multi-Input ConvLSTM for Flood Extent Prediction

Leo Muckley and James Garforth

University of Edinburgh
{l.j.muckley, james.garforth}@ed.ac.uk

Abstract. Flooding is among the most destructive natural disasters in the world. The destruction that floods cause has led to an urgency in developing accurate prediction models. One aspect of flood prediction which has yet to benefit from machine learning techniques is in the prediction of flood extent. However, due to the many factors that can cause flooding, developing predictive models that can generalise to other potential flooding locations has proven to be a difficult task. This paper shows that a Multi-Input ConvLSTM can exploit several flood conditioning factors to effectively model flood extent while generalising well to other flood locations under certain conditions. Furthermore, this study compares the sub-components of the system to demonstrate their efficacy when applied to various flood types.

Keywords: Flood Prediction · Remote Sensing · Deep Learning

1 Introduction

Floods are among Earth's most common and most destructive natural disasters. According to the Organisation for Economic Cooperation and Development, floods cause more than $40 billion in damage worldwide annually [10]. There has been a lot of effort put into detecting floods ahead of time using machine learning techniques with varying success, however one common shortcoming is the inability of models to generalise well to other flood events [9]. This can be attributed to the various types of flooding that exist and the various factors which determine the location and extent of these events. For example, countries can be prone to fluvial flooding (i.e. river flooding), pluvial flooding (i.e. flash floods) and coastal flooding due to storm surges. As a consequence of this, developing a general model that can incorporate all the underpinning features of every variant of flooding to accurately detect a specific flood event ahead of time is difficult.

In this study, we propose an approach to predicting flood extent by using a novel deep learning technique, namely a Multi-Input ConvLSTM. This approach is developed to model both the spatial and temporal elements of flooding, where a limited quantity of training data is available while utilising the spatial autocorrelation in the dataset. Furthermore, the results show the ability of the Multi-Input ConvLSTM technique to generalise to a variety of flood events across several countries in Africa, under certain conditions.

2 Related Work

Flood extent prediction is the task of predicting the level of inundation for a specific location based on a set of flood conditioning factors. In flood extent prediction, the goal is to determine *where* flooding is going to happen and to what extent. This contrasts with flood forecasting where the main goal is to determine *when* a flood is going to happen, to aid flood warning systems. The task of flood forecasting is a widely studied area and various type of statistical and machine learning models have been developed.

Flood Forecasting techniques mostly utilise precipitation and hydrological data to forecast water levels in river basins to aid flood warning systems. Generally, these predictive models are specifically tuned for one area or for a specific type of flood (i.e. pluvial, fluvial flooding etc.) resulting in the inability of the model to generalise well to other flood events. Nevertheless, machine learning techniques have shown some success in flood forecasting tasks [9].

Early applications of feed-forward neural networks have been applied to modelling river water level to effectively forecast riverine flooding [1] in Tagliamento, Italy. In this scenario, the neural network used information from various rain gauges along the river as the features for the model. The results of this model were quite accurate with a mean square error less than 4% when used within a 1-hour time window. However, when the time window is increased, the prediction accuracy decreases as the feed-forward neural network is not able to model the time lag between the water level and the rainfall, which is an important aspect of flood forecasting.

More recent developments in flood forecasting utilise recurrent neural networks to model the temporal element of flooding. An example of this showed the efficacy of LSTM networks when utilised with discharge and rainfall information in the Da River basin in Vietnam [7]. The LSTM network modelled one-day, two-day and three-day flow rate forecasting with Nash–Sutcliffe efficiency (NSE), a metric used to assess the efficacy of hydrological models, reaching 99%, 95%, and 87% respectively. These results demonstrate how well-suited LSTM networks are for flood forecasting due to their ability to effectively deal with time lags and exploit the temporal dependencies in the data. However, the LSTM network does not consider the spatial relationships between the different hydrological stations.

Flood Extent Prediction is one area of environmental research that has yet to benefit from machine learning techniques. However, a similar problem to flood extent prediction is the modelling of flood susceptibility. This modelling is closely related as factors that determine flood susceptibility are likely to have predictive power in modelling flood extent. A GIS-based spatial prediction approach which modelled flood prone areas in the Xing Guo region of China was proposed using an ensemble of techniques [12]. The goal was to accurately classify the flood prone areas using a test set of ground truth levels of flood susceptibility. The ensemble of Logistic Regression and Weight of Evidence models achieved the highest performance of accuracy 90.36% when modelled on the following conditioning factors: (1) altitude, (2) slope, (3) aspect, (4) geology, (5) distance from river, (6) distance from road, (7) distance from fault, (8) soil type, (9) LULC, (10) rainfall, (11) Normalized Difference Vegetation Index (NDVI), (12) Stream Power Index (SPI), (13) Topographic Wetness Index (TWI), (14) Sediment Transport Index (STI) and (15) curvature. The result of this report supports the prior expectation that less vegetated areas at low altitude with high levels of rainfall are more susceptible to flooding. Flood susceptibility is not equivalent to flood extent, however, for an area with a high susceptibility of flooding it would be expected that there would be a positive correlation between the two values. As a result of this, the features used to model susceptibility can be expected to be useful for modelling flood extent.

One example of machine learning for predicting flood extent in an urban context uses both an LSTM and a Logistic Regression model [5]. This methodology first utilises the LSTM to predict the overflow of the storm water management system. The output of the total overflow is then used as a feature for the Logistic Regression model which determines the presence of flooding in each grid of the urban map. The results of this study show the potential for improving urban flood extent prediction using machine learning techniques. However, the methodology developed was specifically for the purpose of urban mapping and is not suitable for generalising to various types of flood events.

Deep Learning architectures that incorporate both recurrent and convolutional structures have been developed for the purpose of spatio-temporal modelling. The ConvLSTM has shown success in precipitation forecasting by utilising sequence of satellite imagery [15,6]. Deep learning models that incorporate a ConvLSTM architecture have the ability to predict the future state of a cell based on previous states of that cell and the states of the neighbours of that cell. As a result of this, the ConvLSTM is particularly useful for spatio-temporal modelling when adequate training data is available so that the spatial dependencies in the data can be modelled over time.

The effectiveness of the ConvLSTM has been shown in other problems beyond precipitation forecasting, such as traffic accident prediction, where heterogeneous urban data is utilised for training [16]. The main challenges highlighted was the rarity of traffic accidents and the spatial heterogeneity of the data. Spatial graph features were developed with an adjacency matrix to overcome the challenge of spatial heterogeneity. This allowed the ConvLSTM to handle the spatial heterogeneity and temporality at the same time while incorporating all the most important features for predicting traffic accidents. Therefore, this study showed that the ConvLSTM is suitable to problems where a trade-off between spatial heterogeneity and rarity of events exists, if the unique data properties are handled properly.

In computer vision, one method for handling unique data properties is to adopt a multi-input approach which is used for deep learning models that receives input of different data types [11,13]. These models adopt a multi-input CNN architecture to make use of a variety of different images or viewing planes as more than one data input is necessary to fully represent the target feature. A similar multi-input approach is utilised in the derivation of soil clay content from a synergy of multi-temporal optical and radar imagery data [14]. The results of this research demonstrated the applicability of multi-input architectures when utilising a variety of satellite imagery for the training of deep learning models.

3 Method

The prediction of flood extent and location is a task of trying to predict the level of inundation y, where $0 \leq y \leq 1$, at time t based on M features for the previous k points in time. In this problem, the level of inundation is the fraction of a region (i.e. over a 1km sq distance) that is covered in flood water at time t and each feature $m \in M$, is a sequence of k timesteps (e.g. sequence of daily total precipitation). It is these M sequences that are then utilised for the prediction of flood extent.

The flood conditioning factors utilised for modelling are the following: (a) constant features, consisting of *distance to water*, *elevation*, *slope*, *clay content* and *soil organic carbon*; (b) temporal features, consisting of *max precipitation*, *total precipitation*, *std precipitation*, *normalised difference vegetation index* and *soil moisture*. When utilising satellite imagery for machine learning tasks, it is often the case that the features extracted have a mix of data types. In addition, these images are generally not available for every time step due to satellite repetition rates. For this problem, each input to the network is either temporal, and therefore a sequence of the previous t-1 timesteps, or constant, therefore the input is a single timestep. To summarise, an effective flood extent prediction model needs to be able to deal with the following two cases: (1) data that is temporal by nature but the measurements may not be available for every timestep (e.g. soil moisture); (2) data measurements that have a very long repetition rate (e.g. 365 days) and can be thus considered a constant feature (e.g. elevation). So a model that is robust to different data types is desirable.

Multi-Input architecture is a type of network where the network receives input of different data types. Inspired by [13] which adopts a multi-input architecture for the purpose of flower grading using computer vision. In this context, each sample contains three images which is fed into a CNN model. Here the multi-input type architecture is favoured as one image cannot fully represent the flower for grading as one image equates to only a partial region of the flower. The motivation for using multi-input architecture on flood extent prediction is similar: the problem cannot be sufficiently modelled based on one type of data input.

In flood extent prediction, input features are generally a mixture of both constant features and temporal features, extracted from satellite imagery. The architecture of the Multi-Input network allows the model to effectively exploit the specific characteristics of each data type and thus, eradicating the need to model non-temporal or constant features in a temporal model. Introducing constant features, that do not vary over time, into a recurrent layer will not correlate with the output and over many timesteps the weights will converge to zero [8]. Therefore, adding constant features to a LSTM would add redundant information or noise at the potentially at the cost of greater model performance. Considering this, the design consists of a recurrent neural network for the temporal input and a feed-forward network for constant input before being concatenated at an intermediate step. Separating the two models based on input and then combining at a subsequent layer has the benefit of learning from each set of input most effectively and then learning interactions between features at the subsequent layer.

ConvLSTM is a type of architecture that was devised for the purpose of modelling spatio-temporal relationships in the short-term prediction of precipitation [15]. This method was shown to surpass state-of-the-art predictions by utilising satellite imagery. In order to achieve this, lots of data was available because rainfall is not a rare weather event, unlike flood prediction where a training set may consist of only a single image. As a consequence of not having sufficient image data for training, it is not feasible to train a conventional ConvLSTM on full image data as this type of architecture generally uses a two-dimensional filter to extract spatial features from the data.

A ConvLSTM layer is similar to an LSTM layer as both are recurrent except that in the ConvLSTM architecture the internal operations now utilise convolutions. This would generally mean that the internal matrix operations of the LSTM are now replaced with convolutional operations, with the input dimension, d, a three-dimensional tensor instead of a one-dimensional tensor that is seen in the LSTM. Generally, for convolutions the filter size is $n \times n$, which effectively extracts spatial dependencies with a sufficient amount of training. However, for the purpose of flood extent prediction, a filter size of $1 \times n$ is proposed. The benefits of this is that the modelling can still exploit local spatial dependencies across multiple features while still being modelled in a sequence. Therefore, the ConvLSTM method utilises a one-dimensional filter to learn high-level local spatial features in the data.

Multi-Input ConvLSTM is now proposed to effectively model both the temporal and spatial dependencies in the data for the purpose of flood extent prediction [1]. This novel approach solves some of the issues involving modelling extreme weather events using satellite imagery, such as: the ability to model sequential problems with a mixture of data types without redundant information being included in the modelling; the ability to exploit local spatial dependencies in the absence of large high-dimensional training sets. Figure 1 outlines the flow of operations for the Multi-Input ConvLSTM. This architecture is also divided into two components: (1) for modelling the temporal features, the ConvLSTM model

[1] `https://github.com/leomuckley/malawi-flood-prediction`

will be utilised; (2) for modelling the constant features a feed-forward network (MLP) will be used. These two components will be combined at a subsequent layer for before being propagated to the output layer.

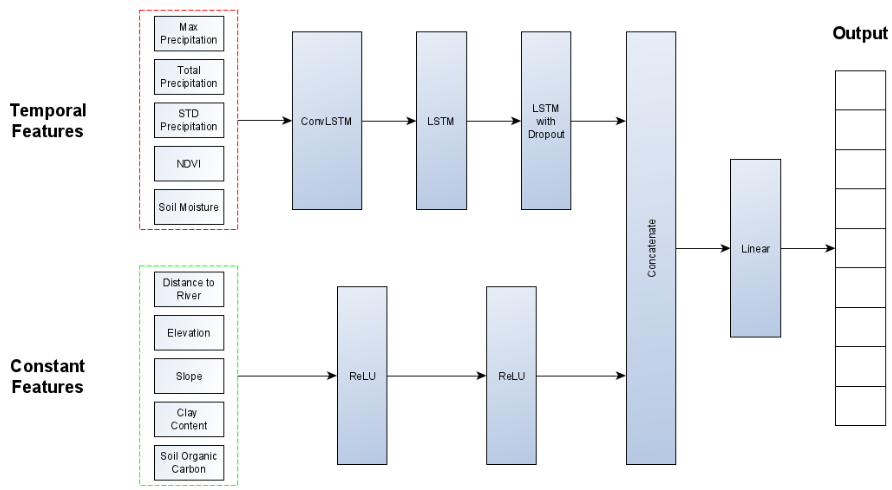

Fig. 1: Multi-Input ConvLSTM.

For the temporal modelling, 3 hidden layers were chosen: a ConvLSTM layer, an LSTM layer and an LSTM layer with dropout. For the ConvLSTM layer, the number of filters and the size of the kernel are to be considered. For the filter size, 64 was chosen and considering the ConvLSTM layer will be used for sequential data the kernel size must be one-dimensional. For this a kernel size 1×5 was chosen as the model could capture a greater amount of the spatial dependencies in the data. For the LSTM layers 2 hidden layers with 64 hidden units was chosen for the first layer and 32 hidden units for the second was chosen. To reduce chances of overfitting, the second layer applied a dropout rate of 0.10 was used. For the constant features, the feed-forward network with two hidden layers, containing 32 and 16 hidden units with a ReLU activation functions, respectively.

For the modelling, choice of loss function used for training the models was the Huber loss is robust to outliers and in the context of flood extent prediction as it would better handle the infrequent cases of high-valued flood extent.

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x)| \leq \delta \\ \delta |y - f(x)| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases} \tag{1}$$

The Adam optimiser [4] was chosen for the learning process as it has an adaptive learning rate which reduces the need to optimise the learning rate hyperparameter. In addition, the models in this study all used mini-batching which updates the error gradient after seeing a only a subset of the data.

4 Experiments

The data collection process used for creating the training and test sets involves extracting the relevant features from raster images, downloaded from Google Earth Engine [2]. For a given flood event, the flood extent target feature can be calculated through a change detection algorithm, where the resulting raster layer highlights the pixels flagged as flooding [3]. The fraction of flooding can then be extracted from this layer to be utilised for experimentation.

To determine the accuracy of the models developed, each model was cross-validated (K-Fold=3) with the results of the experiments reporting the Root Mean Squared Error (RMSE) and the standard deviation to allow for a measure of confidence. The data sets used for training and testing were carefully selected. The reason for careful selection was to test if the models could generalise to various flood events. To test this the following data sets were used:

1. Malawi used for training and testing
2. Mozambique used for training and Malawi used for testing
3. Mozambique used for training and Kenya used for testing

First, the models were trained and tested on homogeneous data to better understand the ability of the models to perform on data of a similar nature. This is a similar to the setup for a Zindi competition[2], as both the training and test set were extracted from identical coordinates in Malawi. Thus, the majority of the underlying features in both training and test sets being similar (e.g. elevation; distance to water).

Second, the models were trained and tested on heterogeneous data, which was comprised of two different datasets. These datasets were selected due to their different geographic features (e.g. land cover type; elevation etc.) but also due to the varying types of flood events. For instance, if a model was trained using pluvial flood data and the target flood event was due to fluvial flooding, this would make the prediction problem a much more difficult task. Therefore, this would test the ability of the models to generalise well to other flood events.

For validation, each dataset was further split into a training set and a validation set. This consisted of the following split: 80% for training the model and 20% for validation. The purpose of this split is to test the performance of a model in parallel allowing for a predefined stopping condition for the training process to be set. Furthermore, each model that was trained, a specific validation strategy was used. Each model was trained using a max epoch value of 1000 and early-stopping applied.

The benchmark to be utilised in this study is based a Light Gradient Boosting Model (LGBM) with model averaging to allow for appropriate comparison in

[2] `https://zindi.africa/competitions/2030-vision-flood-prediction-in-malawi`

the experiments [3]. This is based on the winning solution for a Zindi competition where the goal was to predict the flood extent in Malawi.

4.1 Homogeneous Data

Table 1 exhibits the results for the models utilised in this study. These models have been trained and tested on homogeneous data, where the training and testing data are sourced from the same flood event and type. In this case, the models were trained and tested on Malawi flood data. The results report the RMSE for each model and standard deviation.

Model	**Dataset**
	Malawi
LSTM	0.0525 ± 0.0004
LSTM-Autoencoder	0.0542 ± 0.01
Multi-Input LSTM	0.0532 ± 0.001
ConvLSTM	**0.0519** ± **0.0002**
Multi-Input ConvLSTM	0.0572 ± 0.001
LGBM	0.1021 ± 0.0001

Table 1: Results of training on Malawi flood data and testing on Malawi flood data using RMSE.

In the scenario of homogeneous data, the ConvLSTM surpasses the other models. It achieves the lowest RMSE of 0.0519 with the second lowest RMSE coming from the LSTM model achieving 0.0525. The LSTM-Autoencoder and Multi-Input LSTM report slightly higher RMSE than these models, with 0.0542 and 0.0532 respectively. The Multi-Input ConvLSTM is the worst performing deep learning technique with an RMSE of 0.0572. However, the deep learning techniques developed for this study all surpass the LGBM model as that model reports an RMSE of 0.1021. From Table 1 we see for all the deep learning techniques developed for this study had a RMSE range of between 0.0519 and 0.0572.

The results for the homogeneous dataset demonstrate the ability of each variant of LSTM network to predict flood extent as each model outperforms the LGBM model. This evidence suggests that the LSTM network is well-suited to the problem of flood extent prediction due to the ability of the recurrent layers to model the temporal features. This contrasts with the LGBM model where all the features utilised in the model were considered constant.

4.2 Heterogeneous Data

Table 2 presents the results for testing the ability of the models developed to generalise to other flood events. These models have been trained and tested

[3] `https://github.com/belkhir-aziz/Flood-Prediction-in-Malawi-winning-solution-`

on heterogeneous data, where the training and testing data are sourced from the different flood events with different flood types. In this case, the models were trained on Mozambique data and tested on both Malawi and Kenya flood data. The results in the table report the RMSE for each model with standard deviation.

Model	**Dataset**	
	Malawi	**Kenya**
LSTM	0.1471 ± 0.03	0.2270 ± 0.01
LSTM-Autoencoder	0.1117 ± 0.02	0.2333 ± 0.01
Multi-Input LSTM	**0.0803 ± 0.01**	**0.2120 ± 0.01**
ConvLSTM	0.1350 ± 0.02	0.2180 ± 0.01
Multi-Input ConvLSTM	0.0890 ± 0.01	0.2169 ± 0.02
LGBM	0.2815 ± 0.002	0.2554 ± 0.01

Table 2: Results of training on Mozambique flood data and testing on Malawi and Kenya flood data.

For the Malawi data, the Multi-Input LSTM achieves the lowest RMSE. This model has a RMSE of 0.0803 and the second lowest RMSE coming from the Multi-Input LSTM model achieving 0.0890. The LSTM-Autoencoder reports a slightly higher RMSE of 0.1117. The LSTM and ConvLSTM report the worst RMSE results with 0.1471 and 0.1350, respectively. However, the deep learning techniques developed for this study all surpass the LGBM model as that model reports an RMSE of 0.2815.

For the Kenya data, the Multi-Input LSTM again achieves the lowest RMSE. This model has a RMSE of 0.2120 and the second lowest RMSE coming from the Multi-Input ConvLSTM model with a RMSE of 0.2169. The ConvLSTM reports a slightly higher RMSE of 0.2180. The LSTM and LSTM-Autoencoder report the worst RMSE results with 0.2270 and 0.2333, respectively. However, the deep learning techniques developed for this study all surpass the LGBM model as that model reports an RMSE of 0.2554.

One reason why the models that utilise convolutions do not perform any better here is due to Malawi having less spatial autocorrelation. The Moran's I statistic was computed for each dataset used in the study and the results of these statistical tests showed that Malawi dataset has a lower amount of spatial autocorrelation (0.68), leading to less spatial dependencies in the data to exploit.

The ability of each model to generalise to different types of floods is assessed. The general trend here is that both the Multi-Input LSTM and Multi-Input ConvLSTM outperform the other models for both the Malawi and Kenya datasets. When comparing the performance across datasets, we see the stronger performance on the Malawi dataset in comparison to the Kenya dataset.

5 Conclusion

In this paper, we develop a novel solution for the flood extent prediction problem by proposing a Multi-Input ConvLSTM network which can exploit the spatio-temporal aspect of flooding. In the proposed solution, we split and separate the input features based on their temporality. This solves some issues related to the modelling of extreme weather events using satellite imagery, such as handling conditioning factors consisting of mixed data types, the varying repetition rates of the satellite imagery and the ability to exploit local spatial dependencies in the absence of large training sets.

The results of the experiments show that converting the original flood extent prediction problem into a temporal problem outperforms the current state-of-the-art model in generalising to various types of flood events. Furthermore, we compare the sub-components of the Multi-Input ConvLSTM model to demonstrate their efficacy separately. These comparisons show that the ConvLSTM is particularly effective when the data has high spatial autocorrelation and that the Multi-Input architecture is more effective when the data has higher levels of spatial heterogeneity.

The work presented in this study showed the ability of deep learning techniques when applied to flood extent prediction. Not only has this work provided a useful resource to help mitigate the damage inflicted by flooding, but also provides a foundation for future research in the area of deep learning for flood extent prediction.

References

1. Campolo, M., Andreussi, P., Soldati, A.: River flood forecasting with a neural network model. Water Resources Research **35**(4), 1191–1197 (1999). https://doi.org/10.1029/1998WR900086, `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1998WR900086`
2. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R.: Google earth engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment (2017). https://doi.org/10.1016/j.rse.2017.06.031, `https://doi.org/10.1016/j.rse.2017.06.031`
3. Huang, M., Jin, S.: Rapid flood mapping and evaluation with a supervised classifier and change detection in shouguang using sentinel-1 sar and sentinel-2 optical data. Remote Sensing **12**(13), 2073 (2020)
4. Jais, I.K.M., Ismail, A.R., Nisa, S.Q.: Adam optimization algorithm for wide and deep neural network. Knowl. Eng. Data Sci. **2**(1), 41–46 (2019)
5. Kim, H.I., Han, K.Y., Lee, J.Y.: Prediction of urban flood extent by lstm model and logistic regression. Journal of The Korean Society of Civil Engineers **40**(3), 273–283 (2020)
6. Kim, S., Hong, S., Joh, M., Song, S.k.: Deeprain: Convlstm network for precipitation prediction using multichannel radar data. arXiv preprint arXiv:1711.02316 (2017)
7. Le, X.H., Ho, H.V., Lee, G., Jung, S.: Application of long short-term memory (lstm) neural network for flood forecasting. Water **11**(7), 1387 (2019)

8. Ma, K., Leung, H.: A novel lstm approach for asynchronous multivariate time series prediction. In: 2019 International Joint Conference on Neural Networks (IJCNN). pp. 1–7 (2019). https://doi.org/10.1109/IJCNN.2019.8851792
9. Mosavi, A., Ozturk, P., Chau, K.w.: Flood prediction using machine learning models: Literature review. Water **10**(11), 1536 (2018)
10. OECD: Financial Management of Flood Risk (2016). https://doi.org/https://doi.org/https://doi.org/10.1787/9789264257689-en, `https://www.oecd-ilibrary.org/content/publication/9789264257689-en`
11. Oktay, O., Bai, W., Lee, M., Guerrero, R., Kamnitsas, K., Caballero, J., de Marvao, A., Cook, S., O'Regan, D., Rueckert, D.: Multi-input cardiac image super-resolution using convolutional neural networks. In: International conference on medical image computing and computer-assisted intervention. pp. 246–254. Springer (2016)
12. Shafapour Tehrany, M., Shabani, F., Neamah Jebur, M., Hong, H., Chen, W., Xie, X.: Gis-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques. Geomatics, Natural Hazards and Risk **8**(2), 1538–1561 (2017)
13. Sun, Y., Zhu, L., Wang, G., Zhao, F.: Multi-input convolutional neural network for flower grading. Journal of Electrical and Computer Engineering **2017** (2017)
14. Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G.: Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. Remote Sensing **12**(9), 1389 (2020)
15. Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems. pp. 802–810 (2015)
16. Yuan, Z., Zhou, X., Yang, T.: Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 984–992 (2018)