

Research article

An alternative for predicting real-time water levels of urban drainage systems

Pin-Chun Huang ^a, Kwan Tun Lee ^{a,*}

^a Department of Harbor and River Engineering, National Taiwan Ocean University, Taiwan

ARTICLE INFO

Handling Editor: Jason Michael Evans

Keywords:

Urban floods

Stormwater management

Artificial intelligence model

Sewer system

Street flow simulation

ABSTRACT

Storm Water Management Model (SWMM) developed by the United States Environmental Protection Agency (EPA) has been widely applied throughout the world for analysis associated with stormwater runoff, combined sewers, and other drainage facilities. To appropriately manage the runoff in urban areas, an integrated system including the simulations of sewer flow, street flow, and regional channel flow, called the 1D/1D SWMM model, was advocated to be performed. Nevertheless, the execution efficiency of this integrated system still needs to be promoted to meet the demand for real-time forecasting of urban floods. The objective of this study is to seek an alternative for predicting water levels both in the sewer system and on the streets within an urban district during rainstorms by utilizing a dynamic neuron network model. To strengthen the physical structure of the artificial intelligence (AI) model and simultaneously make up for the lack of measured data, simulation results of the 1D/1D SWMM model are provided as labels for the training of the proposed model. The novelty of this research is to propose a new methodology to effectively train the AI model for predicting the spatial distributions of depths based on the hydrologic conditions, geomorphologic properties, as well as the network relation of the drainage system. A two-stage training procedure is proposed in this study to consider more possible inundation conditions in both sewer and street (open channel) drainage networks. The research findings show that the proposed methodology is capable of reaching satisfactory accuracy and assisting the numerical-based SWMM model for real-time warning of drainage systems in the urban district.

1. Introduction

Urban flooding is usually induced by prolonged periods of intense rainfall, river overtopping, or storm surge since land surfaces are unable to effectively absorb excess water (Hammond et al., 2015). In recent decades, environmental changes have already increased flooding across many countries around the world, particularly in coastal and low-lying areas (Miller and Hutchins, 2017). It has been indicated that atmospheric warming leads to more evaporation and more water available for rain, which further contributes to changing weather patterns and flood risks (Kim et al., 2018; Darabi et al., 2019). Previous studies also point out that the duration, intensity, and frequency of extreme rainfall events have an obvious upward trend and this phenomenon causes more urban and flash floods (Bertilsson et al., 2019). Thus, flooding simulation and real-time early warning in urban areas have become more urgent and necessary to mitigate the threat of flood disasters to life and property. To pinpoint the most suitable dataset for validating numerical simulation models, available studies of urban flooding based on laboratory

experiments were implemented and reviewed. The typical ranges of scale factors used were also highlighted in this research and they depend mainly on the extent of the studied area (Mignot et al., 2019).

For the prevention of flooding in urban areas, the design of sewer systems connected with the street and channel networks is an important issue. Therefore, an integrated planning and management tool to achieve effective management of urban drainage systems is continuously developed (Schmitt et al., 2004; Liu et al., 2015). The EPA Storm Water Management Model (SWMM) is a dynamic rainfall-runoff model used for simulations of both water quantity and quality from primarily urban areas, and it has been widely used for planning and analysis related to multiplex drainage systems in urban districts (Rossman, 2009; Gironás et al., 2010). As low-impact development (LID) techniques have been advocated, the function of SWMM has been expanded for associated analysis on maintaining the pre-development hydrology of a site, thereby reducing negative effects on receiving waters (Rosa et al., 2015; Campisano et al., 2017; Baek et al., 2020). To accurately assess flood inundation extents in urban areas, SWMM was further linked with a

* Corresponding author.

E-mail address: ktle@ntou.edu.tw (K.T. Lee).

variety of surface runoff models to simulate the interaction between the sewer network and the overland flow (Leandro and Martins, 2016; Chen et al., 2017). The process of developing an automated sub-catchment generator, which divides the watershed into multiple sub-catchments and connects the grid cells to the underlying stormwater network, was also established to accelerate the setup of a study case for SWMM modeling (Warsta et al., 2017). In recent years, a geographic information system (GIS)-based SWMM model was developed and promoted to enhance the feasibility of the SWMM as a simulator applied in a river basin that experiences large-scale damage due to frequent flooding (Rai et al., 2017). A new tool for multi-objective automatic calibration was subsequently integrated with the SWMM for determining the optimal set of parameters (Behrouz et al., 2020).

Although a variety of programs have been developed to extend the functions of SWMM and reinforce the model's applicability, the SWMM still faces challenges in dealing with real-time flood simulations due to shortcomings in efficiency and stability, especially when the model is applied to an urban district possessing complicated drainage systems. To provide a quick and stable forecast of water depths in a network of sewer systems, this study intends to utilize machine-learning techniques to perform stormwater simulations and simultaneously reinforce the physical significance of the model by paying attention to the data processing and the training approach. To achieve this purpose, in this study, the simulated results of the SWMM model which is based on the theoretical governing equations are adopted to facilitate the training of the artificial intelligence model. Moreover, the model inputs, namely the influential factors, and the prediction procedure proposed in this study mainly depend on the hydrologic conditions, the geomorphological properties, and the network structure of drainage systems. Since the spatial variation of hydraulic characteristics is well considered in the proposed methodology, the machine learning-based model established by a novel operation method can effectively predict water depths of multiple locations in a network of urban drainage systems as long as it has been trained completely.

2. Description of study area and drainage conditions

An urban township, named Lugang, in northwestern Changhua County of Taiwan was adopted as the study area in this study. As shown in Fig. 1, this township is a seaport located on the west coast of Taiwan, facing the Taiwan Strait. Construction of the stormwater sewer systems in this area planned by the Taiwan government was completed in 1981. The rainfall intensity of the 1-year return period was adopted as the protection standard for the design of sewer pipes. With the development

of industry, commerce, and economy, the urban population has gradually increased, hence, the planning of overall land use in this urban area has been adjusted significantly. Moreover, in recent years, due to the rapid change of climate, rainfall pattern tends to be more concentrated and intensified. Rainfall events have caused a substantial increase in storm runoff and increased the load of existing drainage systems.

As shown in Fig. 1, there are currently three drainage channels (marked by the bold red line), named New Lukang (N.L.), Old Lukang (O.L.), and Fuxing (F.X.), distributed in this urban area to receive and drain water bodies from the stormwater sewer systems. The total drainage area of this urban township is about 452.58 ha. The number of each manhole is also marked in this figure to assist subsequent analysis and explanation. To meet the need for the overall investigation of drainage capability, the peripheral catchment area, regional drainage channels, stormwater sewer systems, and street network having drainage functions in this urban area are collectively considered for the simulations and subsequent training of the AI-based model.

3. Model establishment

All the related tools and approaches applied in this study are introduced accordingly. In the following subsections, the theoretical governing equations embedded in the SWMM are first addressed to offer reliable data as the target values for the training of the AI model. Inputs and outputs of the AI-based model are subsequently discussed to clarify the physical mechanism considered in the proposed method. Then, the operation process of the artificial neuron network algorithm selected in this study is also explained. In the last subsection, the complete procedure with new skills to train the AI model is illustrated in detail. It should be noted that the main contribution of this study, highlighting the novelties of the proposed methodology, is disserted in sub-sections 3.2 and 3.4.

3.1. Numerical-based SWMM

SWMM models the conveyance of a drainage system as a network of conduit links that are connected by nodes. In general, the movement of water through a conveyance network of channels and pipes is governed by the conservation of mass and momentum equations for gradually varied, unsteady free surface flow. The dynamic wave analysis has been indicated to solve the complete form of these equations, hence, it can produce the most theoretically accurate results by accounting for channel storage, backwater effects, flow reversal, and pressurized flow. Since SWMM couples together the solution for both water levels at nodes

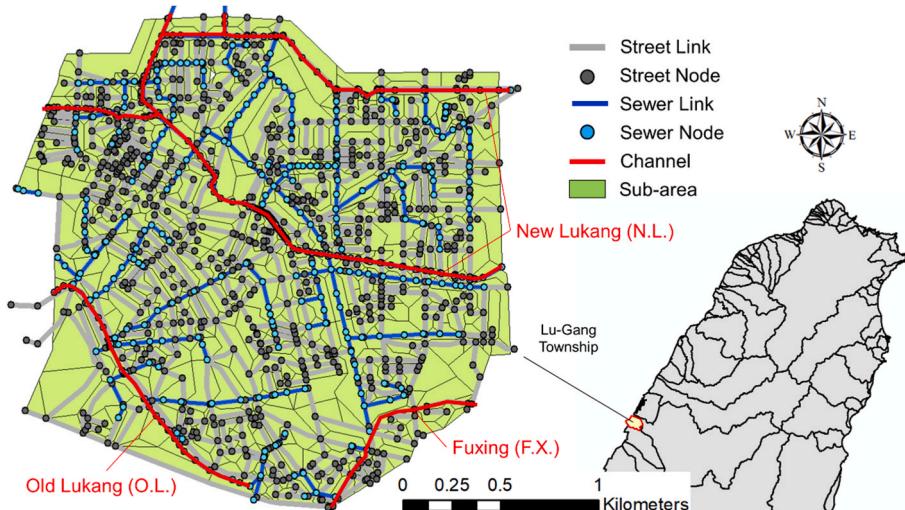


Fig. 1. Geographical location and drainage network condition of the study urban area, Lugang township.

and flows in conduits, it can be applied to any general network layout, even those containing multiple downstream diversions and loops (Rossman, 2017). The conservation of mass and momentum for unsteady free-surface flow through a channel or pipe, known as the St. Venant equations, can be expressed as:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \quad (1)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial(Q^2/A)}{\partial x} + gA \frac{\partial H}{\partial x} + gAS_f = 0 \quad (2)$$

where x denotes distance; t denotes time; A denotes flow cross-sectional area; Q denotes flow rate; H denotes hydraulic head of water in the conduit; S_f denotes friction slope (head loss per unit length); g denotes acceleration of gravity. Since a coupled set of partial differential equations shown in Eqs. (1) and (2) can not be solved analytically, the current version of SWMM adopts an implicit backward Euler method to solve for flow Q in the conduits and head H at the nodes of the conveyance network (Ascher and Petzold, 1998). However, it has been indicated that the numerical stability of SWMM's dynamic wave results can be affected by the choice of the simulation time step. This generality comes at the cost of using small time steps to maintain numerical stability (Rossman, 2017).

3.2. Inputs and outputs of the AI-based model

In this study, the influential factors, adopted as inputs of the machine learning model, are roughly classified into three categories, including hydrological conditions, geomorphological characteristics, and drainage network connection relationships. As shown in Fig. 2, to predict the water levels in both (underground) sewer and (aboveground) street drainage systems smoothly, two-stage training and prediction for the AI-based model are advocated in this study. As shown in the right side of Fig. 2, two kinds of output data are devised in this study, including the water level of a node in the sewer system (E_{SEW}) and the inundation depth of a node on the street (D_{STR}) for a designated location of a manhole (node). To adequately manage different situations causing urban flooding, these two types of output are suggested to be generated sequentially from the proposed prediction model to accommodate possible conditions, such as excess water overflowing to streets from the manhole when sewers are already full, immediate flooding caused by extreme rainfall events that cannot be effectively drained into sewers, or street flooding due to low-lying areas when sewers are not full.

For the first-stage prediction, as shown in the upper part of Fig. 2, input data related to the hydrological conditions contain the rainfall intensities (I_R) within the past T_R hours and the downstream boundary condition at the present state, namely the water depth at the

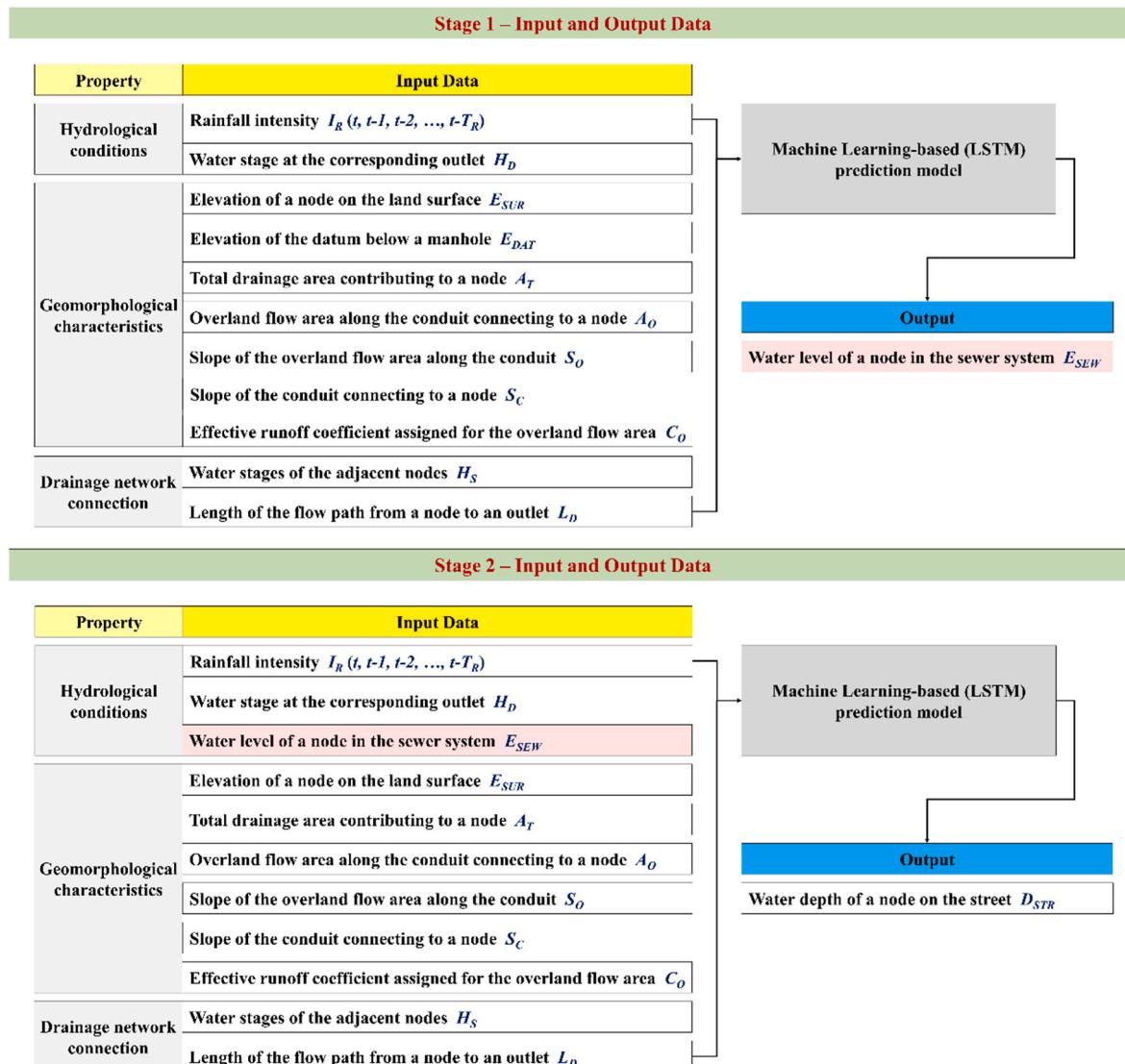


Fig. 2. Input and output data for the two-stage prediction by using the proposed AI models.

corresponding outlet (H_D) for the location of a manhole. It should be noted that the parameter T_R can be thought of as the length of antecedent rainfall that may affect the flow depth at the present state, hence, it is closely associated with the time of flow concentration. As shown in the middle part of upper Fig. 2, input data related to geomorphological characteristics contain the elevation of a manhole on the land surface (E_{SUR}), the sewer datum elevation below a manhole (E_{DAT}), the total drainage area contributing to a manhole (A_T), the overland flow area along the conduit connecting to a manhole (A_O), the average slope of the overland flow area along the conduit (S_O), the slope of the conduit connecting to a node (S_C), and the effective runoff coefficient assigned for the overland flow area along the conduit (C_O), which is determined by the type of land use. As shown by the diagram in the left graph of Fig. 3, input data and target values that are associated with the elevation measured in the sewer system are illustrated. As shown in the right graph of Fig. 3, the overland flow area (lateral flow area) along each conduit connecting to a manhole (A_O) can be partitioned by utilizing the digital elevation model according to the locations of all manholes in the drainage network and the high-resolution elevation dataset. As shown in the lower part of Fig. 2, input data related to the drainage network connection contain the water stages (in the sewer system) of the adjacent manholes directly connected by surrounding conduits (H_S), and the total length of the flow path from a manhole (node) to an outlet following the sewer drainage network (L_D). As shown by the schematic diagram in the right graph of Fig. 3, the water stages separately at the No. 16, No. 18, and No. 22 nodes are adopted as the inputs of H_S for the No. 17 manhole. As shown by the bold green line in the right graph of Fig. 3, the parameter L_D for the No. 17 manhole is determined by the principle of the steepest gradient of surrounding conduits to extract a single flow path connecting to the outlet.

For the second-stage prediction that is performed to forecast the inundation depth of a node on the street (D_{STR}), as shown in the lower graph of Fig. 2, input data related to the hydrological conditions additionally contain the predicted water level of a node in the sewer system (E_{SEW}), which is generated in the first stage. Moreover, the sewer datum elevation below a manhole (E_{DAT}) can be omitted in the input data of the second stage for predicting the (aboveground) inundation depth on the street (D_{STR}).

To justify the selected variables, in this study, Pearson correlation analysis was conducted to screen out these important environmental

factors for predicting water levels. The Pearson correlation can measure the strength of the relationship between two variables. All the input variables considered in the proposed model satisfy a condition – the absolute value of the correlation coefficient between the influential factor and the target output is greater than 0.3. The period of antecedent rainfall assigned in the model was also determined by analyzing the correlation coefficient between the antecedent rainfall intensity and the present water level. The results of the correlation analysis for each input variable including antecedent rainfall intensity can be seen in Table 1.

3.3. Mathematical operations of the AI-based model

To establish the relationship between the input and output data mentioned above, a type of recurrent neural network, named long short-term memory (LSTM), is adopted in this study. LSTM is featured by its feedback connections, which allow a recurrent neural network to process not only single points of data but also entire sequences of data. This characteristic makes LSTM algorithm suitable for predicting information that can be deemed as time series (Hochreiter and Schmidhuber, 1997). It has been indicated that the network architecture of LSTM is capable of providing short-term memory for a recurrent neural network that can last thousands of timesteps. As shown in Fig. 4, a typical LSTM unit is composed of a cell, an input gate, an output gate, and a forget gate. The compact forms of the equations for the forward pass of an LSTM cell can be expressed as

$$i_t = \sigma(W_i x_t + R_i c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_f x_t + R_f c_{t-1} + b_f) \quad (4)$$

$$o_t = \sigma(W_o x_t + R_o c_{t-1} + b_o) \quad (5)$$

in which, i_t is the input gate's activation vector; f_t is the forget gate's activation vector; o_t is the output gate's activation vector; σ is the sigmoid function; x_t is the input vector of the LSTM unit; W and R are respectively the weight matrices of the input and recurrent connections; b denotes the bias vector parameters which need to be learned during training; c_t denotes the cell state vector, which can be calculated by

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \quad (6)$$

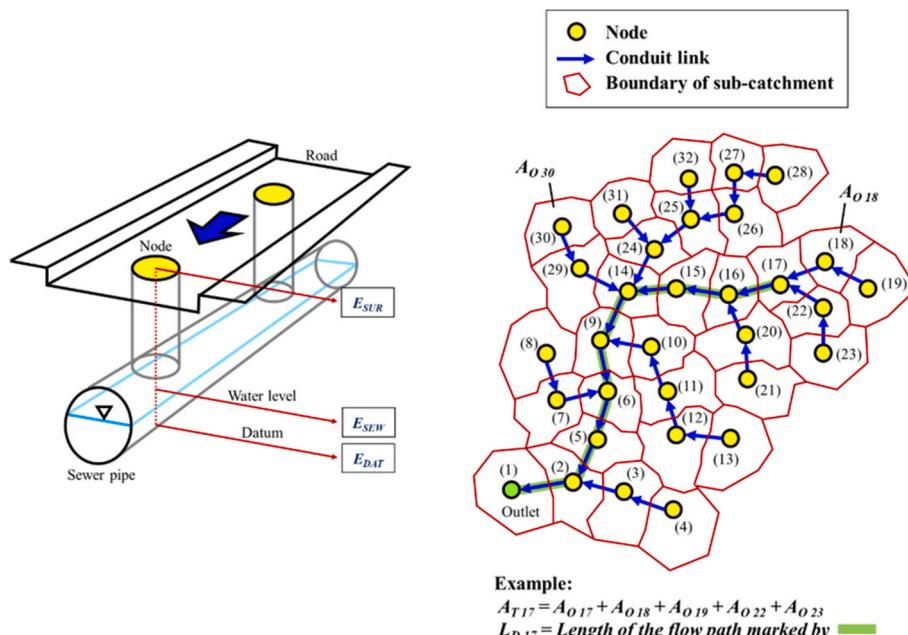
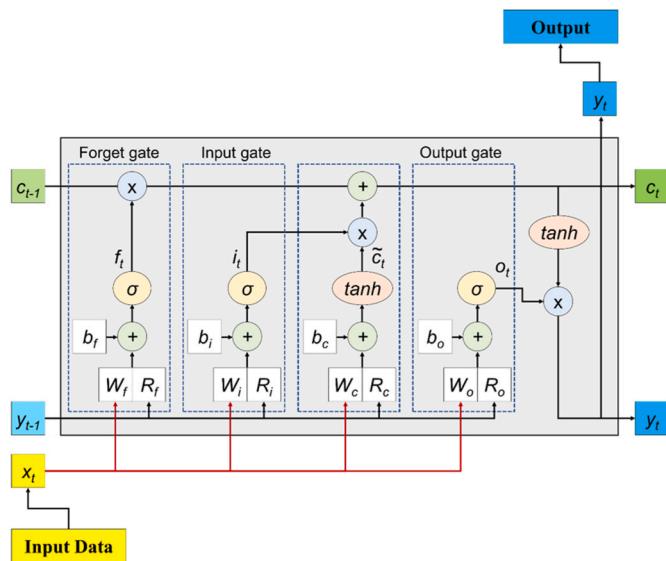


Fig. 3. Schematic diagram of a variety of related input and output data needed for describing the urban drainage conditions.

Table 1

Results of the correlation analysis for each input variable considered in the proposed model.

Property of Factors	Description of the Input Variable	Notation	Absolute Value of the Correlation Coefficient
Hydrological conditions	Rainfall intensity	$I_R(t)$	0.53
	$I_R(t-30\text{min})$		0.64
	$I_R(t-60\text{min})$		0.83
	$I_R(t-90\text{min})$		0.77
	$I_R(t-120\text{min})$		0.65
	$I_R(t-150\text{min})$		0.58
	$I_R(t-180\text{min})$		0.44
	$I_R(t-210\text{min})$		0.32
	H_D		0.52
	Water stage at the corresponding outlet		
Geomorphological characteristics	Elevation of a node on the land surface	E_{SUR}	0.74
	Elevation of the datum below a manhole	E_{DAT}	0.65
	Total drainage area contributing to a node	A_T	0.61
	Overland flow area along the conduit connecting to a node	A_O	0.55
	The slope of the overland flow area along the conduit	S_O	0.39
	The slope of the conduit connecting to a node	S_C	0.47
	Effective runoff coefficient assigned for the overland flow area	C_O	0.49
	Water stages of the adjacent nodes	H_S	0.81
	Length of the flow path from a node to an outlet	L_D	0.42

**Fig. 4.** Neural network structure of the long short-term memory (LSTM) model.

in which, \circ denotes the element-wise product; \tilde{c}_t denotes the cell input activation vector, which can be obtained by

$$\tilde{c}_t = \sigma(W_c x_t + R_c c_{t-1} + b_c) \quad (7)$$

The hidden state vector, which is also known as the output vector of the LSTM unit, can subsequently be derived by

$$y_t = o_t \tanh(c_t) \quad (8)$$

The predicted water level of a node in the sewer system (E_{SEW}) and the predicted inundation depth of a node on the street (D_{STR}) can respectively be updated as follows:

$$E_{SEW} \text{ (or } D_{STR}) = y_t \quad (9)$$

3.4. Training procedure and model setup

In this study, hyetographs of 24-h duration design rainfall separately with 5-year and 10-year return periods are calculated and assigned in the SWMM to generate maps that show the spatial distributions of potential water levels of all nodes in the sewer system of the study urban area. The difference between the elevation of the land surface (E_{SUR}) and the water level of the sewer system (E_{SEW}) at each node can be obtained by

$$\Delta E_i = E_{SUR_i} - E_{SEW_i} \quad (10)$$

The value of ΔE is positive when the land surface is higher than the water level of the sewer system under the criterion of a return period. In this study, the elevation difference (ΔE) of each node respectively according to 5-year and 10-year return periods is calculated and plotted in the upper graph of Fig. 5. It can be found that the distribution of data points is uneven and has multiple clusters with significant gaps, therefore, a clustering analysis is subsequently conducted to partition all data points into multiple groups, which are displayed by different colors. It should be noted that the data points distributed on the lower left side with smaller elevation differences (ΔE), such as the points belonging to Class 1 and Class 2, denote a higher risk of road flooding. Such a finding motivates an idea – if the nodes with a closer extent of flooding risk can be screened out and classified in advance and trained together, the resolution or accuracy to describe the water level by considering the effective input variables can be higher than that training all nodes together without data classification. A density-based clustering method (Ester et al., 1996), known as the density-based spatial clustering of applications with noise (DBSCAN), is chosen in this study to classify all nodes (the locations of manholes) into 6 groups. The spatial distribution of each data group in the sewer drainage network can be seen in the lower graph of Fig. 5. According to the classification result, nodes classified into the same group number have a relatively close scale of flooding risk during rainfall.

To reinforce the predicted accuracy of water depths in a drainage system, the training and prediction procedures of the proposed AI-based model for each class region are suggested to be performed separately. As shown by the flowchart in Fig. 6, all input data has to be classified beforehand according to the result of area division by investigating the spatial distribution of potential water levels in the sewer system as shown in Fig. 5. This step is subsequently followed by the respective training of the LSTM model once the sequence inputs and corresponding target outputs (simulation results yielded by SWMM) of each class are assigned. In this study, the Holdout Cross-validation (CV) method is adopted to ensure that the LSTM model has been trained completely and this method is executed by randomly selecting 25% of the whole data as the testing data and using the remaining data for model training. There is a total of 120 rainfall events that are available for cross-validation, in which 30 events are extracted for the model testing. The rainfall events collected for model training occurred between 1960 and 2020, and they are evenly distributed in four equal ranges of rainfall intensity between 10 mm/h and 110 mm/h. According to the frequency analysis of the

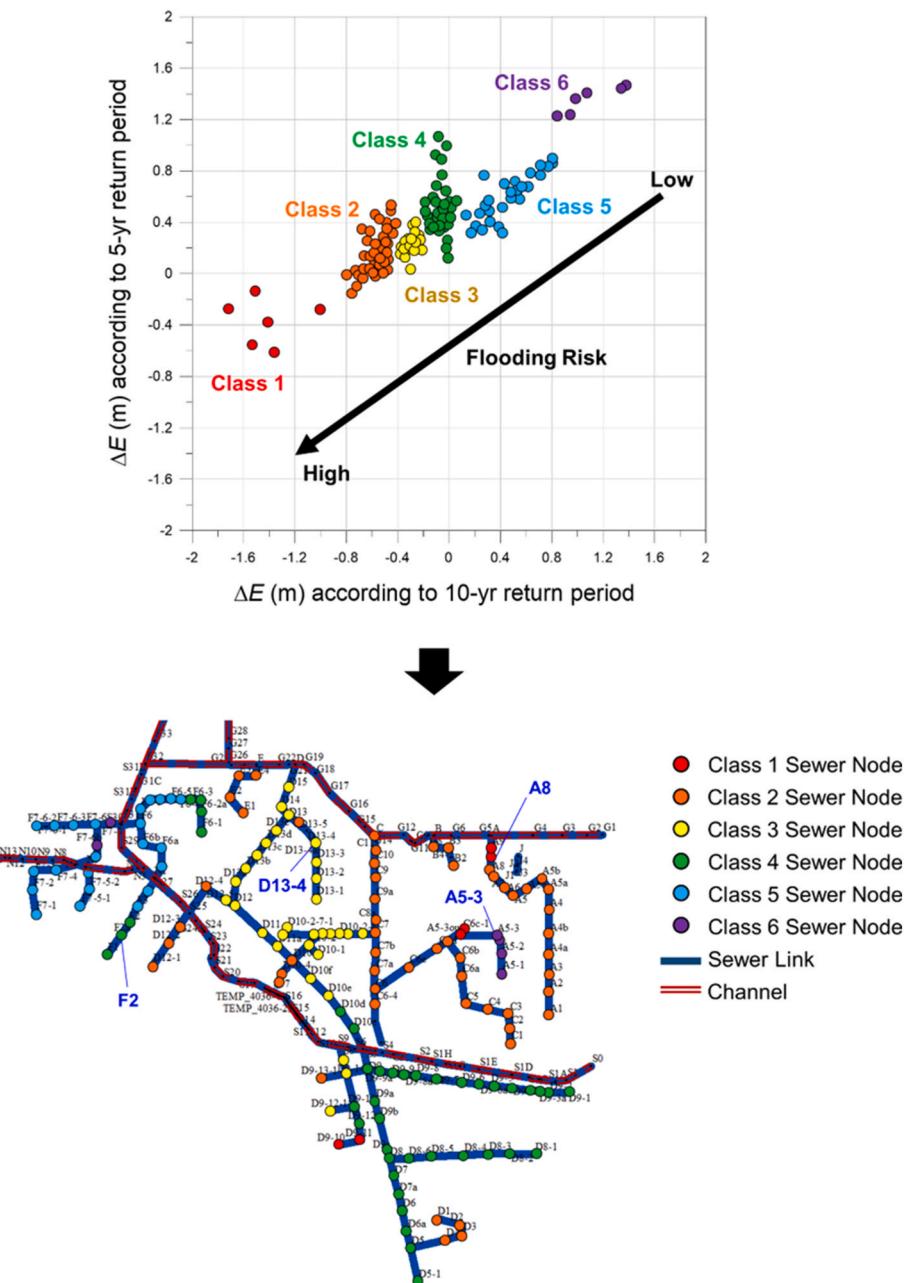


Fig. 5. Classification of input data according to the scale of flooding risk at each node by applying the clustering analysis.

historical rainfall condition observed in the study watershed, the rainfall intensities corresponding to 2-year and 10-year return periods with a duration of 60 min are 56 mm/h and 95 mm/h, respectively.

4. Model application and discussion

In the following content, the numerical results generated by SWMM are verified first to ensure that the simulated values are reliable to be adopted as labels for the training of the LSTM prediction model. The overall performance of the proposed methodology by integrating the LSTM algorithm and a new training procedure is subsequently evaluated in detail.

4.1. Calibrating the numerical-based SWMM and verifying the simulation results

In this study, the records of water stages at several manholes (nodes)

located in a sewer drainage network of the study urban area were adopted to assess and verify the simulation results of SWMM. As shown in the lower graph of Fig. 5, observed water stages of four locations (nodes), which are marked in blue font and installed with water level gauges to offer time series data every 10 min, were collected. A total of 12 rainfall events were utilized to calibrate parameters that are required in SWMM. It should be noted that the 12 rainfall events used in calibrating the SWMM are not from the pool of the 120 rainfall events collected for the training of the proposed LSTM model, and they are evenly distributed in four divided equal ranges of rainfall intensity between 10 mm/h and 110 mm/h. The matrix of mean relative error (denoted as MRE) was selected to determine the optimal set of parameters. The roughness coefficients assigned for the open channel paved with concrete material are between 0.014 and 0.018; the roughness coefficients assigned for the sewer are between 0.013 and 0.016 according to the texture and material of the pipe. As for the sub-areas, the roughness coefficients assigned for the impervious surface are between

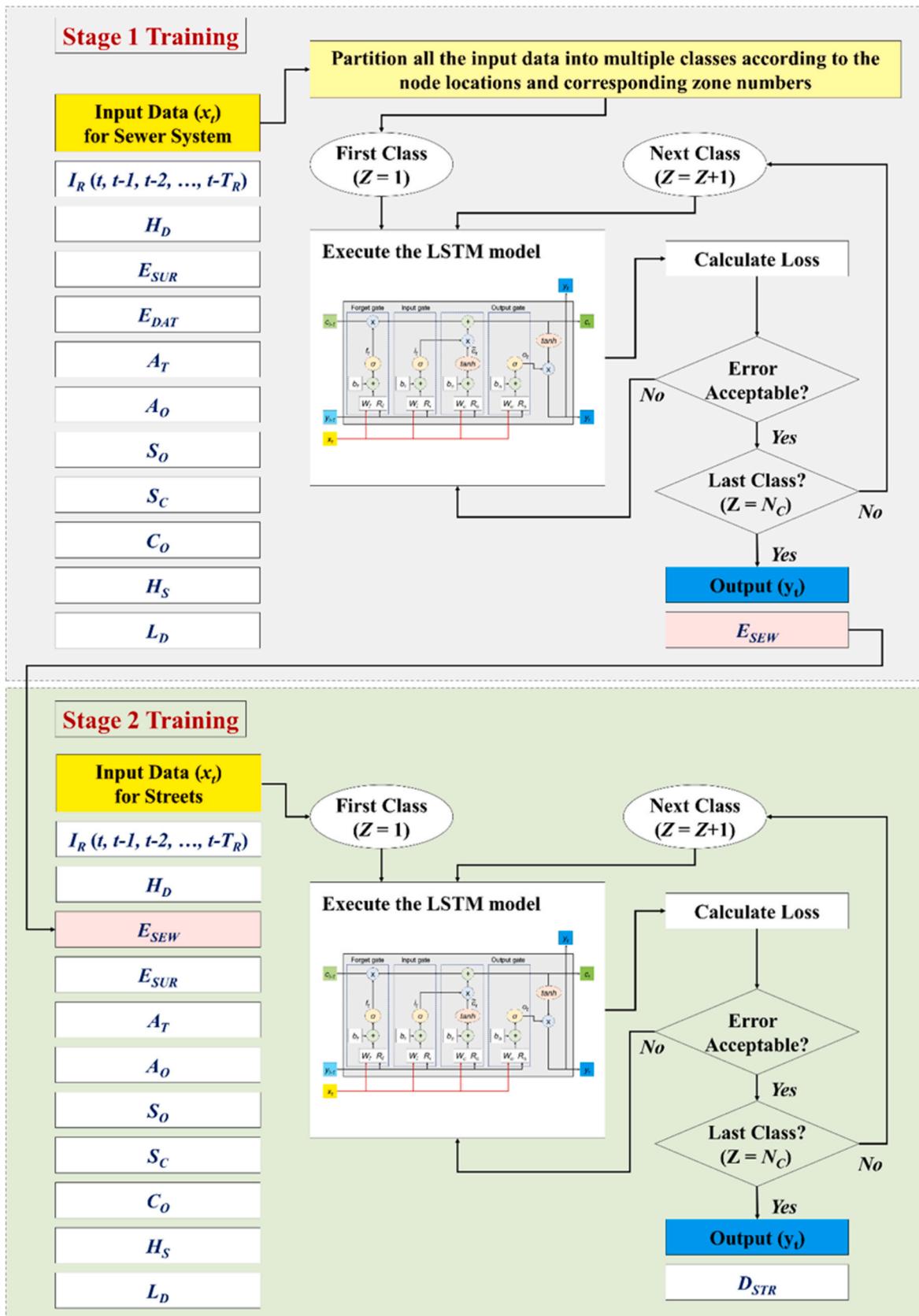


Fig. 6. Flowchart that illustrates the two-stage training procedure of the proposed LSTM prediction model.

0.02 and 0.03; the roughness coefficients assigned for the permeable surface are between 0.1 and 0.2.

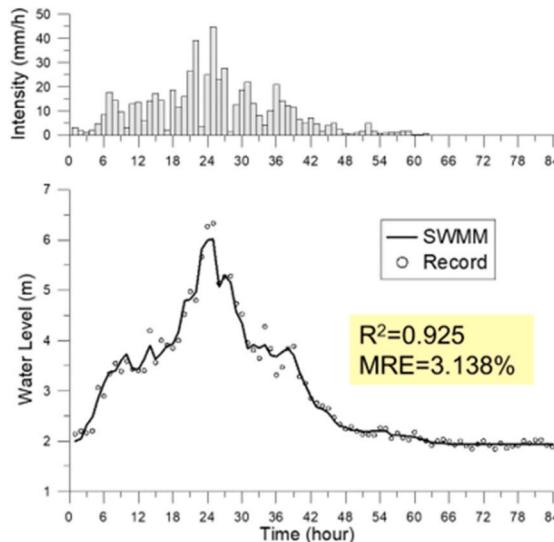
The simulated hydrographs of water levels and available records of water stages caused by different rainfall conditions at the four locations of nodes in the New Lukang (N.L.) district are displayed in Fig. 7. The mean relative error (MRE) and coefficient of determination (R^2) between the observed data and the simulated hydrograph generated by SWMM are also provided in the figure. It can be found that SWMM is capable of yielding satisfactory simulated results fitting with the observed water levels in terms of hydrograph shape (similarity), flood peak (magnitude), and quantified accuracy. Moreover, according to the quantified evaluation of the selected 12 rainfall events, the value of MRE (mean relative error) between the simulated and the observed results can be controlled within 7.742%. Consequently, the above results can preliminarily confirm that the physical mechanism of SWMM that is based on theoretical governing equations can provide reliable simulation results as the target values (labels) to assist the training and

establishment of the proposed AI-based model to predict the real-time water stages in the urban drainage network.

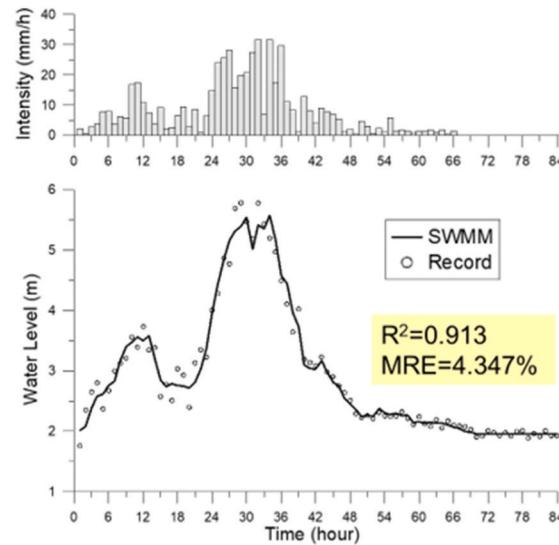
4.2. Evaluating the performance by executing the AI-based prediction model

As mentioned in the previous section, in this study, a two-stage training procedure is proposed to generate water levels of nodes respectively for the sewer and street drainage networks. Classification of all input data by applying the clustering method with the calculated difference of elevation (between the land surface and water level of a 5-year (10-year) return period) is also promoted before the separative training of the AI model to further enhance the predicted accuracy. Fig. 8 shows simulated water stages produced by SWMM and the predicted water levels yielded by LSTM models for the sewer drainage network of the New Lukang (N.L.) district under four of the testing rainfall events. It should be noted that two types of LSTM models that

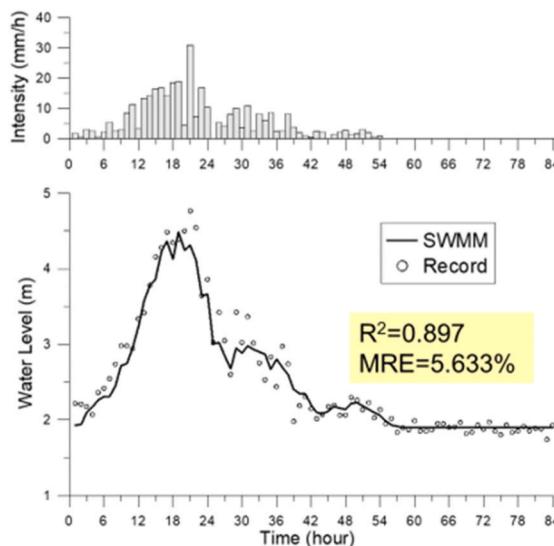
Class 1 – Node No. A8



Class 2 – Node No. A5-3



Class 3 – Node No. D13-4



Class 4 – Node No. F2

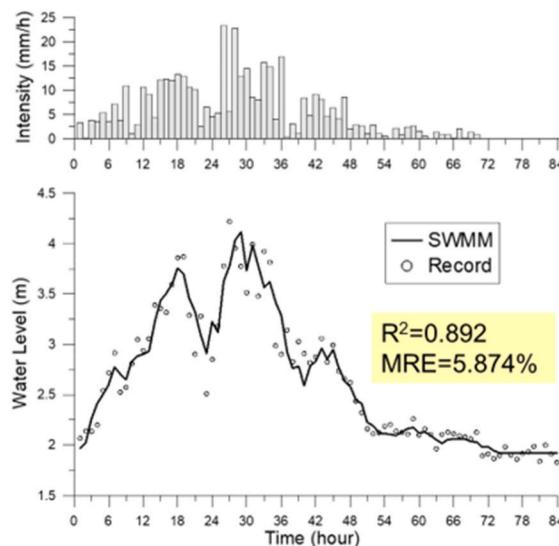
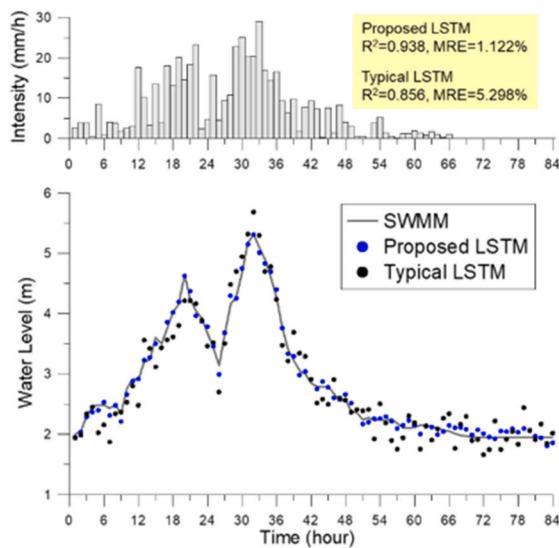
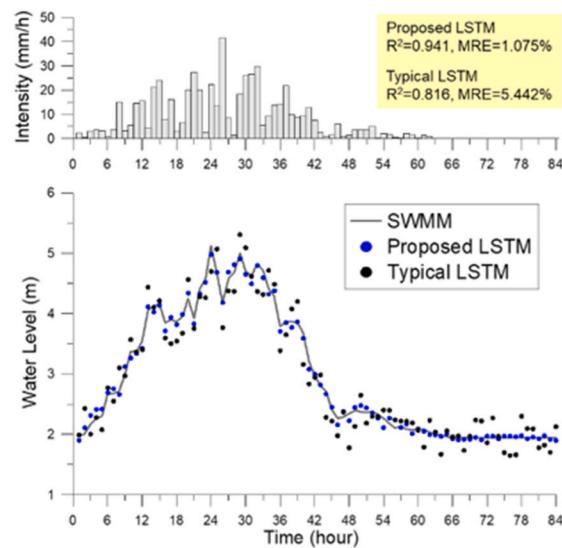


Fig. 7. Simulated and observed hydrographs of water levels caused by rainfall events at four locations of nodes.

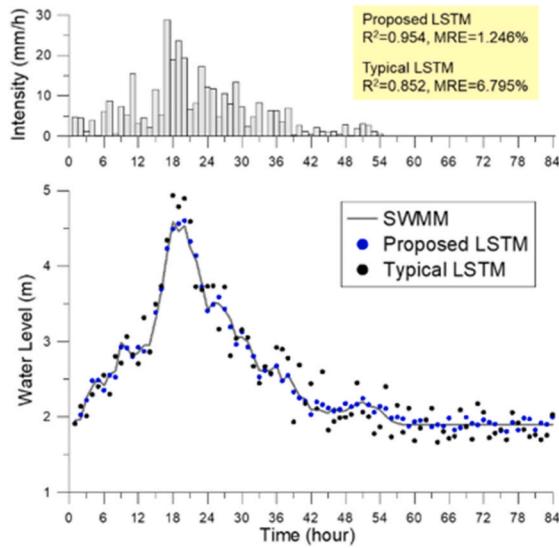
Class 1 – Node No. A8



Class 2 – Node No. A5-3



Class 3 – Node No. D13-4



Class 4 – Node No. F2

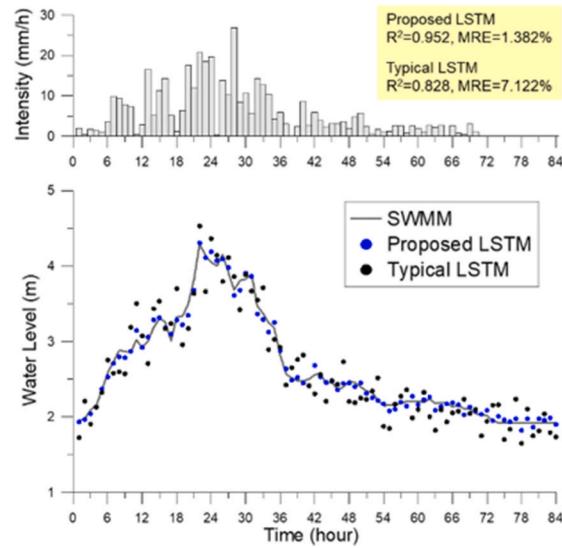


Fig. 8. Simulated water stages produced by SWMM and the predicted water levels yielded by LSTM models for the sewer drainage network of the New Lukang (N.L.) district.

respectively (a) apply the data classification for separative training and (b) assign all available data for collective training are performed to compare their difference in predicted accuracy. As shown in Fig. 8, “Proposed LSTM” denotes the method performed by data classification and separative training, and “Typical LSTM” denotes the method performed by collective training and prediction. The mean relative error (MRE) and coefficient of determination (R^2) between the two hydrographs respectively generated by the SWMM and LSTM model are also provided in figure. It can be seen that when the numerical solutions generated by the SWMM are provided as a reference criterion (target values), the proposed methodology that combines the LSTM model with the treatment of data classification for respective training can significantly offer a more accurate prediction than the other LSTM model that neglects the process of separative training.

The graphs that show the comparisons between the simulated water stages (produced by SWMM) and the predicted water stages (yielded by

LSTM models) are also provided in Fig. 9 to facilitate the explanation of the performance. The data points generated by the proposed methodology (marked in blue color) show a more concentrated distribution around the 45-degree diagonal than the data points generated by the typical LSTM model (marked in black color) without following the suggested training procedure. This indicates that the proposed method can provide predictions that are more similar to the simulated results of SWMM. These research findings show that adopting a variety of influential factors as input data is still not enough to effectively preserve the predicted accuracy if the proposed new training procedure (as illustrated in the flowchart of Fig. 6) is omitted. This situation also demonstrates the importance and necessity of conducting the suggested training procedure in which data classification and separative training of LSTM models are jointly implemented to achieve a better performance of water level predictions in the sewer network.

In this study, three indicators, mean relative error (MRE), coefficient

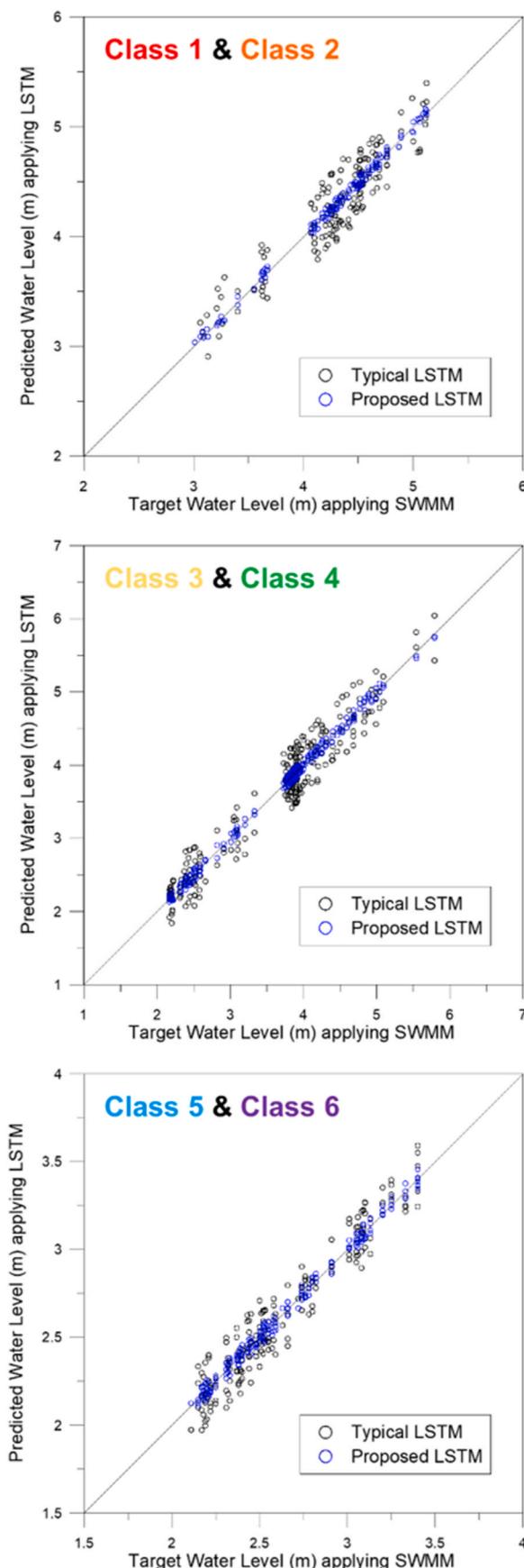


Fig. 9. Comparisons of the target water stages (produced by SWMM) and the predicted water stages (yielded by LSTM models).

of determination (R^2), and Pearson correlation coefficient (PC), are adopted to detect and quantify the accuracy in predicting water levels by applying the proposed LSTM model combined with a new training procedure. These three indicators can be calculated by

$$MRE = \frac{1}{N} \sum_{t=1}^N \frac{|y_t^p - y_t^m|}{y_t^m} \quad (11)$$

$$R^2 = 1 - \frac{\sum_{t=1}^N (y_t^p - y_t^m)^2}{\sum_{t=1}^N (y_t^p - \bar{y}^p)^2} \quad (12)$$

$$PC = \frac{\sum_{t=1}^N (y_t^p - \bar{y}^p)(y_t^m - \bar{y}^m)}{\sqrt{\sum_{t=1}^N (y_t^p - \bar{y}^p)^2 \sum_{t=1}^N (y_t^m - \bar{y}^m)^2}} \quad (13)$$

in which, N denotes the total number of predicted values (sample points); y_t^m denotes the target water stage generated by SWMM; y_t^p denotes the predicted water stage generated by the LSTM model; \bar{y}^m denotes the mean of all target water stages; \bar{y}^p denotes the mean of all predicted water stages. MRE is adopted to determine the average magnitude of the absolute error of the predictions to the target values. R^2 can measure the strength of the relationship between the target and predicted values on a convenient scale from 0 to 1. PC is a measure of linear correlation between two sets of data and it possesses a value between -1 and 1 . The predicted and simulated water stages at sewer nodes (manholes) for a total of 30 testing rainstorm events separately generated by the SWMM and LSTM models are collected for analyzing the predicted accuracy. Table 2 shows the quantified results of the three above indicators for each class of nodes based on the division shown in Fig. 5 by respectively applying the proposed LSTM model integrated with the suggested training procedure and the typical LSTM model without the process of separative training. According to the estimated results of all testing events, it can be found that when the proposed methodology is implemented, the MRE can be controlled within 1.542%; the R^2 can be consistently larger than 0.932; the PC can be greater than 0.955 overall. These results indicate the adequateness of using the proposed method to imitate the simulated water stages yielded by SWMM. Additionally, the table also shows that merely applying the typical LSTM model without training it through data classification in advance significantly compromises the predicted accuracy in the three aspects of indicators.

To discuss the capability of the proposed methodology in predicting the spatial distribution of water levels at sewer nodes and the inundation depths on the street, the flooding maps separately for the sewer and street drainage networks under the rainfall condition that is one of the

Table 2

Evaluation of predicted accuracy for the typical and proposed LSTM models by using three indicators.

Data Class	MRE		R^2		PC	
	Typical LSTM	Proposed LSTM	Typical LSTM	Proposed LSTM	Typical LSTM	Proposed LSTM
1	5.371%	1.115%	0.853	0.941	0.924	0.974
2	5.546%	1.068%	0.813	0.944	0.919	0.963
3	6.872%	1.239%	0.849	0.958	0.938	0.976
4	7.116%	1.375%	0.825	0.956	0.932	0.983
5	7.325%	1.542%	0.764	0.932	0.925	0.961
6	8.442%	1.489%	0.781	0.947	0.927	0.955

Note.

“MRE” denotes the mean relative error.

“ R^2 ” denotes the coefficient of determination.

“PC” denotes the Pearson correlation coefficient.

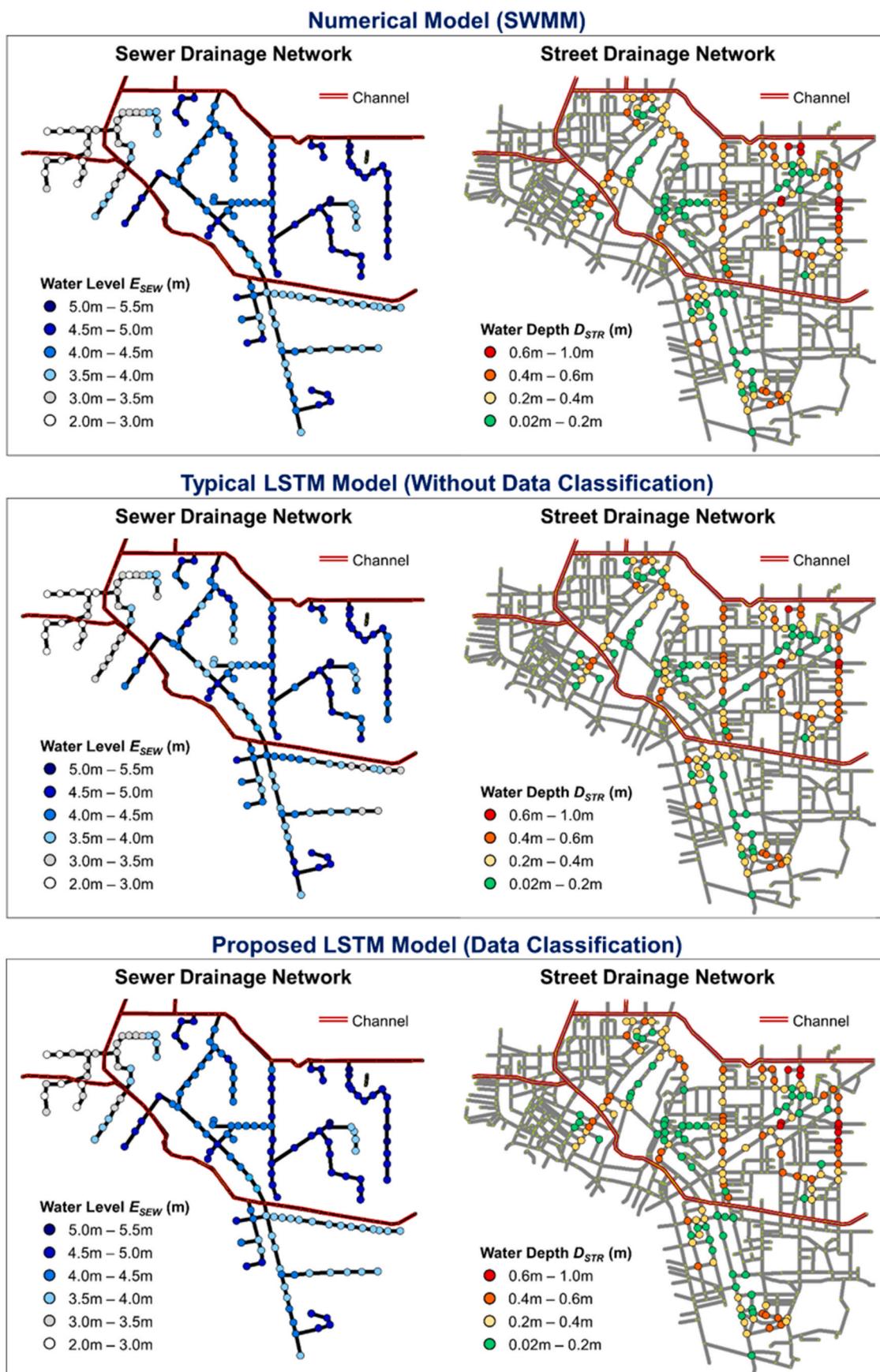


Fig. 10. Flooding maps separately for the sewer and street drainage networks under a testing rainfall event.

testing events are also provided in Fig. 10, in which the numerical simulation results yielded by SWMM are also displayed as benchmarks. The left graph shows the results of the first stage prediction, generating the sewer water levels in space that are classified into 6 intervals from 2.0 m to 5.5 m. It can be found that compared with the typical LSTM model, the proposed LSTM model established by a new training procedure can offer results that more closely match the numerical solutions produced by SWMM. The similarity between the two maps that are respectively generated by the LSTM model and SWMM can be improved from 71.8% to 98.5% as long as the new training approach is additionally applied to build the LSTM model. The right graph of Fig. 10 shows the results of the second stage prediction, generating the street inundation depths in space that are classified into 4 intervals from 0.02 m to 1.00 m. The proposed methodology again manifests a better ability to generate inundation maps than the typical LSTM model. This means the proposed method can also provide more accurate predictions of water depths on streets when the numerical solutions (produced by SWMM) are adopted as target values. The similarity between the two maps that are respectively generated by the LSTM model and SWMM can be enhanced from 65.4% to 96.6% if the proposed training approach is additionally implemented to establish the LSTM model.

To manifest the improvement of computational efficiency by applying the proposed methodology to replace the SWMM, the comparison of computational cost between the SWMM and the proposed LSTM model is also provided. As shown in Table 3, the running times of five rainfall events from the testing set by separately conducting the SWMM and the proposed LSTM model are listed. It can be founded that if the proposed LSTM model is adopted as an alternative to performing urban flooding prediction, about 9.1 times computational efficiency can be enhanced according to the overall evaluation. It should be noted that in addition to the improvement of computational efficiency, another important merit of applying the proposed LSTM model to replace the SWMM for real-time forecasting of urban flooding is that the problem of numerical instability can be completely avoided. The results of a practical application in this study show that the predicted accuracy is still not acceptable by merely integrating the typical LSTM with SWMM. Only by conducting the separative training based on the data classification and following the two-stage prediction procedure proposed in this study can the predicted accuracy for urban flooding forecasting be improved overall. Therefore, the importance and necessity of executing the proposed methodology to facilitate the training of the typical LSTM model can be demonstrated.

5. Conclusions

To enhance the model's efficiency and stability for achieving real-time flood forecasting of the urban drainage system, this study proposes an alternative to avoid the direct execution of numerical-based SWMM during a rainstorm. The purpose of this study is to establish another type of urban flood prediction model by combining the artificial neuron network algorithm with a new training procedure that is based on the hydrological conditions, geomorphological characteristics, and drainage network connection. The numerical results generated by SWMM are still required but they are merely provided as labels for the training of the AI-based model. To comprehensively consider more situations causing urban flooding, a combined drainage system including the sewer, streets, and channel networks, is advocated to be performed in this research. Therefore, the novelty of this study is not only to reinforce the physical significance of the LSTM model by extensively using possible characteristic factors but also to train the model with two primary stages that accommodate predictions in both sewer and street drainage systems. Moreover, to further enhance the accuracy of predicted water stages, a clustering method, named DBSCAN, is suggested to be applied to classify all input data into multiple classes in advance according to the difference between the land surface elevation and the water level of each node (under specific return periods) in the sewer

Table 3

Evaluation of computational efficiency by respectively applying the SWMM and the proposed LSTM models.

Event Number	Duration of the Flood Event (hours)	Running Time of the Model for Prediction (seconds)	
		SWMM	Proposed LSTM
(1)	24	496	59
(2)	49	994	115
(3)	66	1341	148
(4)	75	1519	165
(5)	84	1708	178

network before conducting the training process. The research findings show that the proposed LSTM model can provide satisfactory predictions, in which the RMSE is within 1.542%, R^2 is greater than 0.932, and the PC is greater than 0.955 in comparison with the numerical solutions yielded by SWMM, as long as the suggested routing procedure with data classification and separative training is implemented. On the other hand, the similarity of inundation depth for aboveground street nodes between the two flooding maps that are respectively generated by the LSTM model and SWMM can be enhanced from 65.4% to 96.6% if the suggested training approach can be performed to establish the LSTM model. Such a performance confirms that the proposed methodology is comparable to the SWMM in terms of simulated accuracy and simultaneously offers a more efficient and stable way to predict water levels of the urban drainage system without the risk of numerical oscillations during the process of runoff routing.

CRediT authorship contribution statement

Pin-Chun Huang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Review & Editing, Visualization, Data curation, Supervision. **Kwan Tun Lee:** Writing–Review, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Kwan Tun Lee reports financial support was provided by National Science and Technology Council, Taiwan. Kwan Tun Lee reports a relationship with National Science and Technology Council that includes: funding grants.

Data availability

The authors do not have permission to share data.

References

- Ascher, U.M., Petzold, L.R., 1998. Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations. SIAM, Philadelphia.
- Baek, S., Ligaray, M., Pachepsky, Y., Chun, J.A., Yoon, K.S., Park, Y., Cho, K.H., 2020. Assessment of a green roof practice using the coupled SWMM and HYDRUS models. *J. Environ. Manag.* 261, 109920.
- Behrouz, M.S., Zhu, Z., Matott, L.S., Rabideau, A.J., 2020. A new tool for automatic calibration of the Storm Water Management Model (SWMM). *J. Hydrol.* 581, 124436.
- Bertilsson, L., Wiklund, K., de Moura Tebaldi, I., Rezende, O.M., Veról, A.P., Miguez, M.G., 2019. Urban flood resilience—A multi-criteria index to integrate flood resilience into urban planning. *J. Hydrol.* 573, 970–982.
- Campisano, A., Catania, F.V., Modica, C., 2017. Evaluating the SWMM LID Editor rain barrel option for the estimation of retention potential of rainwater harvesting systems. *Urban Water J.* 14 (8), 876–881.
- Chen, W., Huang, G., Zhang, H., 2017. Urban stormwater inundation simulation based on SWMM and diffusive overland-flow model. *Water Sci. Technol.* 76 (12), 3392–3403.
- Darabi, H., Choubin, B., Rahmati, O., Haghghi, A.T., Pradhan, B., Kloke, B., 2019. Urban flood risk mapping using the GARP and QUEST models: a comparative study of machine learning techniques. *J. Hydrol.* 569, 142–154.

- Ester, M., Kriegel, H.P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *kdd 96* (34), 226–231.
- Gironás, J., Roesner, L.A., Rossman, L.A., Davis, J., 2010. A new applications manual for the storm water management model(SWMM). *Environ. Model. Software* 25 (6), 813–814.
- Hammond, M.J., Chen, A.S., Djordjević, S., Butler, D., Mark, O., 2015. Urban flood impact assessment: a state-of-the-art review. *Urban Water J.* 12 (1), 14–29.
- Kim, S.E., Lee, S., Kim, D., Song, C.G., 2018. Stormwater inundation analysis in small and medium cities for the climate change using EPA-SWMM and HDM-2D. *J. Coast Res.* 85 (10085), 991–995.
- Leandro, J., Martins, R., 2016. A methodology for linking 2D overland flow models with the sewer network model SWMM 5.1 based on dynamic link libraries. *Water Sci. Technol.* 73 (12), 3017–3026.
- Liu, L., Liu, Y., Wang, X., Yu, D., Liu, K., Huang, H., Hu, G., 2015. Developing an effective 2-D urban flood inundation model for city emergency management based on cellular automata. *Nat. Hazards Earth Syst. Sci.* 15 (3), 381–391.
- Mignot, E., Li, X., Dewals, B., 2019. Experimental modelling of urban flooding: a review. *J. Hydrol.* 568, 334–342.
- Miller, J.D., Hutchins, M., 2017. The impacts of urbanisation and climate change on urban flooding and urban water quality: a review of the evidence concerning the United Kingdom. *J. Hydrol.: Reg. Stud.* 12, 345–362.
- Rai, P.K., Chahar, B.R., Dhanya, C.T., 2017. GIS-based SWMM model for simulating the catchment response to flood events. *Nord. Hydrol.* 48 (2), 384–394.
- Rosa, D.J., Clausen, J.C., Dietz, M.E., 2015. Calibration and verification of SWMM for low impact development. *JAWRA J. Am. Water Resources Associat.* 51 (3), 746–757.
- Rossman, L.A., 2017. *Storm Water Management Model Reference Manual Volume II – Hydraulics*. U.S. Environmental Protection Agency.
- Rossman, L.A., 2009. *Storm Water Management Model User's Manual Version 5.0*. National Risk Management Research Laboratory, United. EPA/600/R-05/040.
- Schmitt, T.G., Thomas, M., Ettrich, N., 2004. Analysis and modeling of flooding in urban drainage systems. *J. Hydrol.* 299 (3–4), 300–311.
- Warsta, L., Niemi, T.J., Taka, M., Krebs, G., Haahti, K., Koivusalo, H., Kokkonen, T., 2017. Development and application of an automated subcatchment generator for SWMM using open data. *Urban Water J.* 14 (9), 954–963.