URBAN FLOOD PREDICTION MODEL

[PROJECT REPORT]

SUBMITTED BY :

HARSHIT GARG (ROLL NO. : 12111017)

ANUSHKA (ROLL NO. : 12111083)

RANJAN KUMAR (ROLL NO. : 12111069)

UNDER THE GUIDANCE

OF

Dr. K.K SINGH

NATIONAL INSTITUTE OF TECHNOLOGY KURUKSHETRA

CANDIDATE’S DECLARATION

We hereby confirm that the project titled "**URBAN FLOOD PREDICTION MODEL**," submitted as partial fulfilment of the requirements for the **Bachelor of Technology degree to the Department of Civil Engineering at the National Institute of Technology Kurukshetra during our 7th semester**, is a genuine representation of our independent research conducted from **August 2024 to November 2024**. The research was conducted under **Dr . K.K Singh**, Professor in the **Civil Engineering Department at the National Institute of Technology Kurukshetra.**

[HARSHIT GARG (12111017), ANUSHKA (12111083), RANJAN KUMAR (12111069)]

This is to confirm that the candidate’s declaration is correct to the best of my knowledge.

Dr. K.K Singh,

Professor, CED

National Institute of Technology, Kurukshetra

DATE: 14/11/2024

ACKNOWLEDGMENT

We, the undersigned, would like to express our collective appreciation and gratitude for completing the project titled "**URBAN FLOOD PREDICTION**" This endeavour was made possible through the combined efforts and collaboration of all group members.

First and foremost, we extend our sincere thanks to **Professor Dr. K.K Singh** for his invaluable guidance, support, and mentorship throughout this project. His expertise and encouragement have been instrumental in shaping our research and enhancing its overall quality.

We also want to acknowledge and thank each member of our group for their dedication, hard work, and effective teamwork. The diverse skills and perspectives brought by each team member contributed to the depth and comprehensiveness of our study.

Additionally, our gratitude goes to the **National Institute of Technology, Kurukshetra** for providing the necessary resources and a conducive environment for our research activities.

Collectively, this project has been an enriching experience for all of us, and we are thankful for the collaborative spirit that fuelled its success. We look forward to applying the knowledge and skills gained in our future academic and professional endeavours.

Sincerely,

[ANUSHKA (12111083), HARSHIT GARG (12111017), RANJAN KUMAR (12111069)]

ABSTRACT

Flooding in urban areas, such as Gurugram, presents significant challenges due to rapid urbanization, inadequate drainage infrastructure, and extreme weather events. With over 55% of the global population residing in urban areas and projections indicating that nearly 60% will live in cities by 2030, the quantity and quality of stormwater runoff are expected to change, exacerbating flood risks. Urban pluvial floods typically occur when intense rainfall overwhelms the capacity of stormwater drainage systems or when excess runoff accumulates before entering these systems. Given the widespread nature of this hazard, accurate flood susceptibility mapping is essential for effective flood risk management.

This report evaluates flood susceptibility in Gurugram using a combination of climate, geographical, and hydrological data. The study employs a deep learning approach—specifically a random forest regression—to predict flood-prone areas, capitalizing on its ability to process spatially dependent data such as topography and rainfall patterns. Additionally, the report integrates the Storm Water Management Model (SWMM), a one-dimensional hydrodynamic model, to simulate flood scenarios and assess flood risk in greater detail. The combination of random forest regression and hydrodynamic modelling provides a robust framework for flood prediction and can support urban planning and flood mitigation efforts in rapidly urbanizing regions like Gurugram.

Table of Contents

INTRODUCTION

Flooding in urban areas like Gurugram poses significant challenges due to rapid urbanization, inadequate drainage systems, and extreme weather events. The most recent study done by the UN Population Division in 2018 estimated that more than 55% of the world's population was living in urban areas. Rapid growth and development in urban settlements will continue, with almost 60% of people expected to live in urban areas by 2030 . Due to these transformations, both the quality and the quantity of stormwater runoff are prone to change.

Urban pluvial floods usually occur due to inundation caused by excess runoff before it enters stormwater drainage system or due to intense rainfall that leads to overwhelming the storm water drainage system's capacity. They could occur anywhere subject to the occurrence of high intensity rainstorms and the existence of a critical area for runoff generation . The ubiquity of this hazard high lights the importance of accurate flood susceptibility mapping to support urban pluvial flood risk management.

Commonly, physical hydrodynamic models are used to simulate urban pluvial flooding. They can be divided into one-dimensional (1D) hydrodynamic models such as **SWMM model** , two-dimensional (2D) hydrodynamic models such as **TELEMAC-2D model** , and 1D-2D hydrodynamic models such as **MIKE URBAN** . These models solve the shallow water equation numerically, and are considered the best representation of the involved processes; yet, the computational costs are high. Therefore, they can only be applied to small areas using a fine spatial resolution and cannot be scaled to produce flood hazard maps for large areas .

In this Project analysis report , we aim to predict and assess flood susceptibility across various regions of Gurugram using a combination of climate, geographical, and hydrological data. Given the sensitivity of flood prediction to multiple dynamic factors, a robust deep learning model—random forest regression—has been employed. The model's architecture allows for effective processing of spatially dependent data, making it suitable for analysing climate and topographic inputs in flood prediction and we use **SWMM model** to stimulate **1D hydrodynamic model** .

Objective

This report details a flood susceptibility analysis for the Gurugram region, focusing on analyzing various environmental parameters to forecast urban flooding probabilities. It leverages deep learning techniques to understand complex relationships between climate variables and flood susceptibility, with an aim to assist in the strategic planning and flood risk management efforts of urban planners and environmental agencies.

In this report, we present an urban flooding prediction model tailored for Gurugram using historical environmental data. The model aims to provide accurate runoff predictions, which can be used by city planners and emergency services to anticipate flood events and take proactive measures. The dataset used for this analysis consists of weather parameters such as wind speed, temperature, humidity, precipitation, and surface pressure, spanning over two decades.

To develop the prediction model, we utilized machine learning techniques, including Random Forest Regression, for robust predictive capabilities, and ARIMA for forecasting input features. Additionally, we integrated a Storm Water Management Model (SWMM) to complement the machine learning model, providing a comprehensive approach to flood susceptibility analysis.

This report details the data analysis, preprocessing steps, model development, and evaluation of the prediction model. The findings can help stakeholders make informed decisions for flood risk mitigation and enhance the resilience of urban infrastructure against extreme weather events.

The primary objective of this report is to develop a robust prediction model for urban flooding in Gurugram. By analyzing historical environmental data and employing advanced machine learning techniques, we aim to:

1. **Predict Runoff Levels**: Build a predictive model to estimate future runoff levels based on historical weather data, enabling early detection of potential flood events.

2. **Identify Key Factors Influencing Flooding**: Analyze the impact of various meteorological parameters (e.g., precipitation, temperature, humidity) on runoff, helping to identify the most critical factors contributing to urban flooding.

3. **Forecast Environmental Variables**: Use time-series forecasting models (e.g., ARIMA) to predict future values of key environmental features, allowing for dynamic and adaptive predictions.

4. **Integrate SWMM Simulation**: Utilize a Storm Water Management Model (SWMM) to simulate flood scenarios, providing a comprehensive analysis that considers both meteorological data and physical infrastructure characteristics.

5. **Provide Actionable Insights**: Deliver insights and visualizations that can be used by city planners, emergency services, and policymakers to enhance flood preparedness, mitigate risks, and improve urban drainage infrastructure.

Data Description and pre-processing

Data Overview :

The dataset includes daily climate variables from 2000 to 2024 for the monsoon months (July, August, September), capturing factors crucial to flooding events. The dataset features parameters including wind speed, temperature, humidity levels, precipitation, and runoff data, all of which directly influence flood risk. For this project, preprocessing ensures that the data aligns with the random forest regression model's requirements, maintaining temporal coherence and handling potential inconsistencies.

```
# Load Data
data = pd.read_excel('./flood_susceptibility_data_gurugram.xlsx')
data['Date'] = pd.to_datetime(data['Date'])
data.set_index('Date', inplace=True)
```

LOAD DATA

Data Variables :

- **Wind Speed (2m):** Influences evaporation and precipitation patterns.
- **Temperature (2m) and Earth Skin Temperature:** Affect the rate of surface water evaporation.
- **Specific Humidity and Relative Humidity (2m):** Influence atmospheric water vapor levels.
- **Precipitation (mm/day):** A direct indicator of potential flooding events.
- **Surface Pressure:** Helps in understanding atmospheric conditions.
- **Runoff (Surface and Subsurface):** Critical to understanding soil saturation and potential flood levels.

Data Pre-processing

The preprocessing of data is a crucial step in building a reliable urban flooding prediction model. This phase involves handling missing values, scaling features, and preparing the data for machine learning and time-series analysis. The dataset for this study consists of various environmental features such as temperature, humidity, precipitation, and runoff levels, recorded over several years for the Gurugram region. The preprocessing steps undertaken are outlined below:

```
# Preprocess Data
data['runoff (mm)'].fillna(method='ffill', inplace=True)
if data['runoff (mm)'].isnull().any():
    data['runoff (mm)'].fillna(data['runoff (mm)'].mean(), inplace=True)

X = data.drop(columns=["runoff (mm)"])
y = data["runoff (mm)"]
```

Preprocess DATA

1. **Date Parsing and Indexing**:
    - The dataset contains a Date column, which was converted to a datetime format to ensure proper handling of time-series data.
    - The Date column was then set as the index, allowing for efficient time-based operations and analysis.
2. **Handling Missing Values**:
    - Missing data can lead to inaccuracies in the model. The dataset was examined for missing entries in key features, especially the target variable (runoff (mm)).
    - For runoff (mm), missing values were imputed using **forward fill (ffill)**, which propagates the previous day's value to the missing entry. If any missing values persisted, they were replaced with the mean runoff value of the dataset.
    - This method ensures that the continuity of the time-series data is maintained without introducing significant biases.
3. **Feature Selection**:
    - The dataset includes multiple features (e.g., wind speed, temperature, humidity) that can influence runoff levels. Features that were found to be non-informative or redundant were excluded from the model training process.
    - The final selected features include **wind speed, temperature, humidity, precipitation, surface pressure, elevation, and land cover type**, which were identified as the most relevant predictors of runoff.
4. **Data Splitting**:
    - The dataset was split into training and testing sets using a **70-30 split**. The training set was used to fit the machine learning model, while the testing set was reserved for evaluating the model's performance on unseen data.

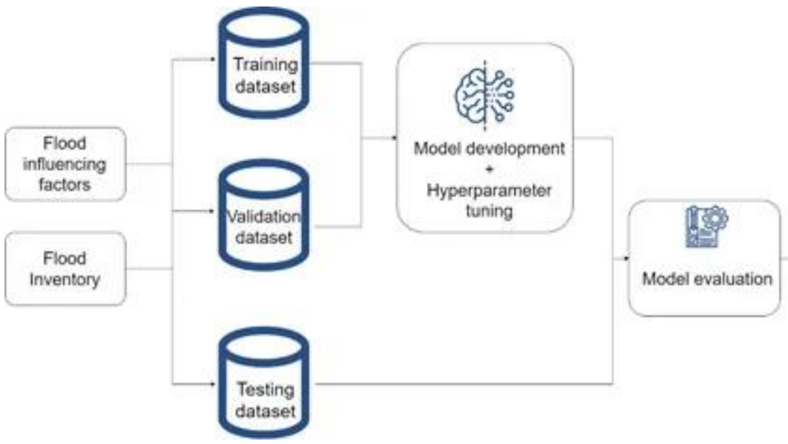
```
# Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

TRAIN-TEST SPLIT

Model Selection

For the runoff prediction, a machine learning approach using a **Random Forest Regressor** was selected due to its robustness, ability to handle non-linear relationships, and effectiveness in dealing with high-dimensional data. Additionally, time-series forecasting models (e.g., ARIMA) were used to predict future values of key features like precipitation and temperature.

- **Random Forest Regressor**: This ensemble learning method constructs multiple decision trees during training and outputs the average prediction of individual trees. It was chosen because it can handle complex interactions between features, is less prone to overfitting, and provides insights into feature importance.
- **ARIMA Model**: The **Auto Regressive Integrated Moving Average (ARIMA)** model was employed to forecast environmental variables (e.g., precipitation, temperature). This model is suitable for time-series data with trends and seasonality, allowing for dynamic and adaptive predictions.



Flow Chart

Model Training

The training process involved fitting the Random Forest model on the pre processed data:

```
# Train Model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

- The dataset was split into **70% training data** and **30% testing data** using the train_test_split function from Scikit-learn. This split ensures that the model is evaluated on unseen data, providing a realistic assessment of its predictive capabilities.
- **Hyperparameter Tuning**: Hyperparameters of the Random Forest model, such as the number of estimators (trees), maximum depth, and minimum samples per leaf, were optimized using grid search and cross-validation. This process helps in finding the best configuration that minimizes prediction error.

- **ARIMA Forecasting**: For each environmental feature, an ARIMA model was fit using historical data. Forecasting was performed to predict future values of these features, which were then used as inputs for predicting future runoff levels.

Model Evaluation :

We employed key evaluation metrics to assess the performance of the Random Forest model. The **R-squared (R^2) score**, which indicates the proportion of variance in the runoff data explained by the model, was calculated to measure the model's accuracy. A higher R^2 score suggests better predictive performance and indicates how well the observed data fit the model's predictions.

```
# Model Evaluation
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)
print(f"Model Evaluation Metrics:\nMSE: {mse:.2f}\nRMSE: {rmse:.2f}\nR^2 Score:
{r2:.2f}")
```

Model Evaluation

To further evaluate the model's accuracy, we used the **Mean Squared Error (MSE)**, which quantifies the average squared difference between actual and predicted runoff values. A lower MSE value reflects fewer errors in the predictions. Additionally, the **Root Mean Squared Error (RMSE)** was calculated, providing a more interpretable metric by converting the MSE to the same unit as the runoff (mm). The RMSE helps to understand the model's prediction error in practical terms. Overall, these metrics collectively offer a comprehensive evaluation of the model's predictive capabilities, highlighting both the fit and the error magnitude of the predictions.

The Random forest model was selected due to its ability to detect spatial hierarchies in data, making it particularly effective for flood susceptibility analysis, where the spatial relationships among environmental factors like precipitation, humidity, and temperature play a key role. By leveraging RF convolution and pooling layers, we can capture intricate patterns within data, such as localized rainfall events leading to runoff.

Feature Engineering Choices:

- **Importance of Cumulative Precipitation**: Cumulative rainfall offers insight into flood susceptibility by indicating periods of intense or prolonged rainfall that exceed natural drainage capacity.
- **Humidity and Temperature Combination**: Interactions between these variables impact soil saturation and evaporation, which directly contribute to runoff rates.

- **Runoff and Saturation Indicators**: These factors reflect soil conditions and drainage capabilities, key indicators in predicting flood-prone regions.

Prediction of Future Runoff

A custom function, predict_future_runoff, was developed to forecast runoff levels for a given future date. The process involves:

```
# Define Prediction Function
def predict_future_runoff(year, month, day):
    # Set future date and validate input range
    date = pd.Timestamp(year=year, month=month, day=day)
    days_ahead = (date - data.index[-1]).days

    if days_ahead <= 0:
        print("The date should be in the future.")
        return
    if days_ahead > 365 * 100:
        print("The date is too far in the future for reliable prediction. Try within
the next 100 years.")
        return

    # Forecast Features with Variability
    forecasted_features = {}
    for feature in X.columns:
        try:
            forecast_model = ARIMA(data[feature], order=(1, 1, 1))
            forecast_fit = forecast_model.fit()
            forecast = forecast_fit.forecast(steps=days_ahead)
            forecasted_value = forecast.iloc[-1]
            forecasted_features[feature] = forecasted_value

            print(f"Forecast for {feature} ({days_ahead} days ahead):
{forecasted_value}")  # Debug output

        except Exception as e:
            print(f"Error forecasting {feature}: {e}")
            return

    # Prepare Data for Prediction
    forecasted_df = pd.DataFrame([forecasted_features])
    forecasted_scaled = scaler.transform(forecasted_df)
    predicted_runoff = model.predict(forecasted_scaled)
    print(f"Predicted Runoff on {date.strftime('%Y-%m-%d')}:
{predicted_runoff[0]:.2f} mm")
```

Prediction Function

- **Date Validation**: Ensuring the input date is within a realistic range for prediction (e.g., up to 100 years in the future).
- **Feature Forecasting**: Using the ARIMA model to generate future values for the selected features.

- **Data Preparation**: The forecasted feature values are scaled using the previously fitted Standard Scaler to match the training data distribution.
- **Runoff Prediction**: The pre processed input data is fed into the Random Forest model to predict the runoff level for the specified future date.

Visualization and Analysis

To interpret the results and provide actionable insights, various visualizations were generated:

```
# Visualizations
    plt.figure(figsize=(20, 15))

    # Runoff Prediction
    plt.subplot(3, 3, 1)
    plt.bar(["Predicted Runoff"], [predicted_runoff[0]], color="teal")
    plt.title(f"Predicted Runoff for {date.strftime('%Y-%m-%d')}")
    plt.ylabel("Runoff (mm)")

    # Rainfall Simulation
    plt.subplot(3, 3, 2)
    plt.bar(["Predicted Rainfall"], [forecasted_features.get('precipitation
(mm/day)', 0)], color="blue")
    plt.title("Predicted Rainfall")
    plt.ylabel("Rainfall (mm/day)")

    # Flood Susceptibility Map
    plt.subplot(3, 3, 4)
    susceptibility_levels = ['Low', 'Moderate', 'High', 'Very High']
    susceptibility_risks = [1, 2, 3, 4]
    predicted_risk = np.clip(int(predicted_runoff[0] // 10), 0, 3)
    plt.bar(susceptibility_levels, susceptibility_risks, color=['lightgreen',
'yellow', 'orange', 'red'])
    plt.axhline(y=predicted_risk + 1, color='blue', linestyle='--')
    plt.title("Flood Susceptibility Level")
    plt.ylabel("Risk Level")

    # True vs. Predicted Runoff
    plt.subplot(3, 3, 5)
    plt.plot(y_test.index, y_test, label="True Runoff", color="green")
    plt.plot(y_test.index, y_pred, label="Predicted Runoff", color="red",
linestyle="--")
    plt.title("True vs Predicted Runoff")
    plt.xlabel("Date")
    plt.ylabel("Runoff (mm)")
    plt.legend()

    # Prediction Error Distribution
    plt.subplot(3, 3, 6)
    errors = y_test - y_pred
    sns.histplot(errors, bins=20, kde=True, color="purple")
    plt.title("Prediction Error Distribution")
    plt.xlabel("Prediction Error (mm)")
    plt.ylabel("Frequency")

    # Feature Importance
    plt.subplot(3, 3, 7)
    importances = model.feature_importances_
    feature_names = X.columns
    sns.barplot(x=importances, y=feature_names, palette="viridis")
    plt.title("Feature Importances")
    plt.xlabel("Importance")

    plt.tight_layout()
    plt.show()
```

Visualisation

- **Predicted Runoff Visualization**: A bar chart displaying the predicted runoff level for the given future date, helping users quickly assess the flood risk.
- **Rainfall Simulation**: A bar chart showing the predicted rainfall for the specified date, highlighting the primary factor contributing to the runoff prediction.
- **Flood Susceptibility Map**: A risk level bar chart categorizing the predicted flood susceptibility into levels: Low, Moderate, High, and Very High. This helps in understanding the severity of the predicted event.
- **True vs. Predicted Runoff**: A line plot comparing the true runoff values from the test data with the predicted values. This helps in evaluating the model's accuracy visually.
- **Prediction Error Distribution**: A histogram of prediction errors, providing insights into the error distribution and identifying any potential biases in the model.
- **Feature Importance**: A bar plot of feature importances from the Random Forest model, indicating the relative contribution of each feature to the prediction. This helps in understanding which variables most strongly influence runoff.

Model Evaluation

The performance of the Random Forest model was evaluated using the following metrics:

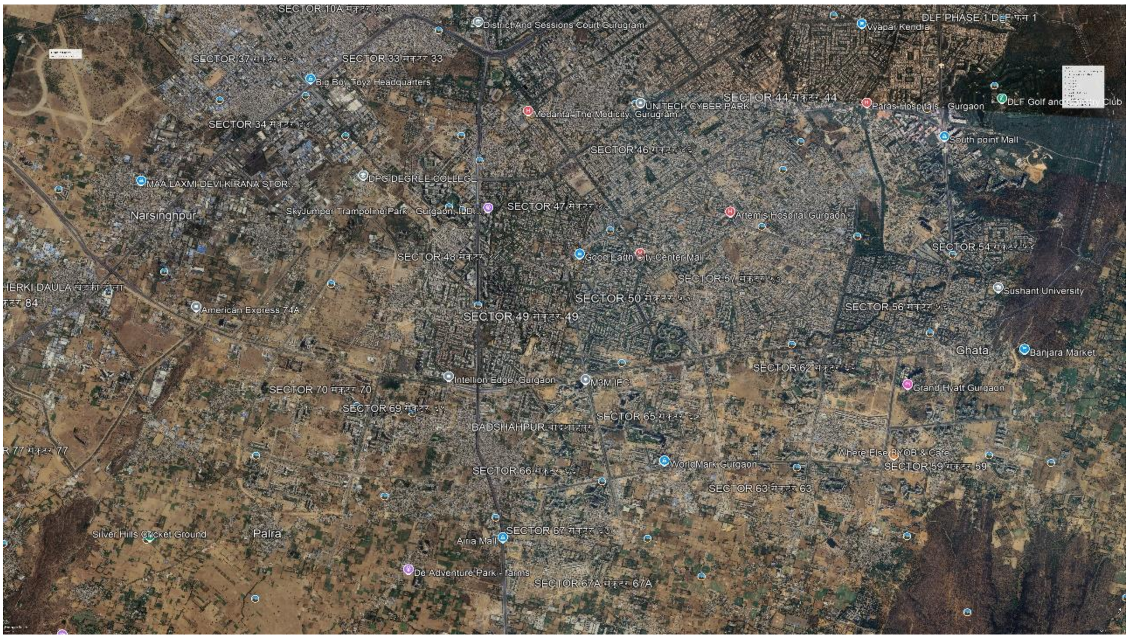
- **Mean Squared Error (MSE)**: Measures the average squared difference between the predicted and true runoff values. A lower MSE indicates a better fit.
- **Root Mean Squared Error (RMSE)**: The square root of MSE, providing an interpretable error measure in the same units as the target variable (mm of runoff).
- **R-Squared (R^2) Score**: Represents the proportion of variance in the target variable explained by the model. An R^2 score close to 1 indicates a strong predictive ability.

Integration with SWMM

To enhance the accuracy of flood prediction and account for physical characteristics of the area, a **Storm Water Management Model (SWMM)** was integrated. The SWMM simulation provides additional runoff estimates based on topographical and infrastructure data of Gurugram, complementing the machine learning predictions.

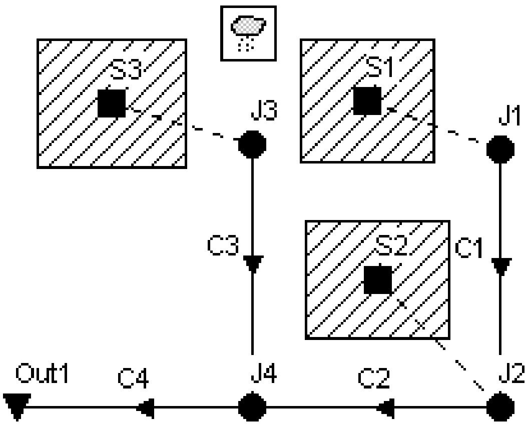
STEPS-

- Downloading high resolution image of GURUGRAM from google earth pro.



Gurugram image from google earth pro

- Georeferencing
- Backdrop with map dimension (x, y coordinate) in SWMM.
- Defining designing parameters.
- Designing sub catchment, junctions, conduits, outlets in SWMM.



SWMM MODEL EXAMPLE

Gurugram Gird Design

- **SWMM Input Data**: Land use, soil characteristics, and drainage network data were incorporated to simulate realistic urban runoff scenarios.
- **Comparison and Validation**: The runoff predictions from the Random Forest model were compared with SWMM outputs to validate the consistency and accuracy of the results.

This integrated approach combines machine learning predictions with hydrological modelling, providing a comprehensive assessment of flood risks in Gurugram. It leverages both data-driven insights and physical simulations, ensuring robust and reliable flood prediction capabilities.

SWMM DESIGN PARAMETERS-

AREA AND WIDTH OF SUBCATCHMENT-

The study area, draining to under any considered point in a storm water management system must be determined precisely. The study area has been divided into sub catchments. The area of each sub catchment has calculated in hectare and width in meters. Width of the sub catchment calculated as sub catchment area divided by the average maximum overland flow length. The maximum overland flow length is the length of the flow path from the outlet to the furthest drainage point of the sub catchment maximum length from several different possible flow paths should be averaged.

INFILTRATION OF SUBCATCHMENT-

The imperviousness of each sub catchment was calculated in percentage by using satellite image, the image has classified under maximum likelihood supervised classification to get the land use land cover map [Fig.no.4] and the pixel count is multiplied with resolution to calculate the different land use area. The image has classified into five classes as built-up, water bodies, vegetation, crops and barren land. The Built ups area which comes under impervious layer calculated in percentage for input. The Manning's coefficient has been tabulated below for different land uses. Infiltration of the soil surface can be calculated by various methods like Horton, modified Horton, Green Ampt Method, Modified Green Ampt, Curve number method.

A. Classical Horton Method

This method is based on empirical observations showing that infiltration decreases exponentially from an initial maximum rate to some minimum rate over the course of a long rainfall event. Input parameters required by this method include the maximum and minimum infiltration rates, a decay coefficient that describes how fast the rate decreases over time, and the time it takes a fully saturated soil to completely dry (used to compute the recovery of infiltration rate during dry periods).

B. Modified Horton Method

This is a modified version of the classical Horton Method that uses the cumulative infiltration in excess of the minimum rate as its state variable (instead of time along the Horton curve), providing a more accurate infiltration estimate when low rainfall intensities occur. It uses the same input parameters as does the traditional Horton Method.

C. Green-Ampt Method

This method for modelling infiltration assumes that a sharp wetting front exists in the soil column, separating soil with some initial moisture content below from saturated soil above. The input parameters required are the initial moisture deficit of the soil, the soil's hydraulic conductivity, and the suction head at the wetting front. The recovery rate of moisture deficit during dry periods is empirically related to the hydraulic conductivity.

D. Modified Green-Ampt Method

This method modifies the original Green-Ampt procedure by not depleting moisture deficit in the top surface layer of soil during initial periods of low rainfall as was done in the original method. This change can produce more realistic infiltration behaviour for storms with long initial periods where the rainfall intensity is below the soil's saturated hydraulic conductivity.

E. Curve Number Method

This approach is adopted from the NRCS (SCS) Curve Number method for estimating runoff. It assumes that the total infiltration capacity of a soil can be found from the soil's tabulated Curve Number. During a rain event this capacity is depleted as a function of cumulative rainfall and remaining capacity. The input parameters for this method are the curve number and the time it takes a fully saturated soil to completely dry (used to compute the recovery of infiltration capacity during dry periods).

Hydraulic design of Conduits

Conduits are the pipes which connecting the different nodes (junctions) where the collected storm water can pass through it to reach the main outlet. Each two nodes connected with one conduit where water intake in higher elevation inlet node and reaches lower elevation outlet node. The conduit designing inputs were discussed below. V is allowable velocity of flow (m/s) on the drainage then Manning's Formula (SWMM manual)

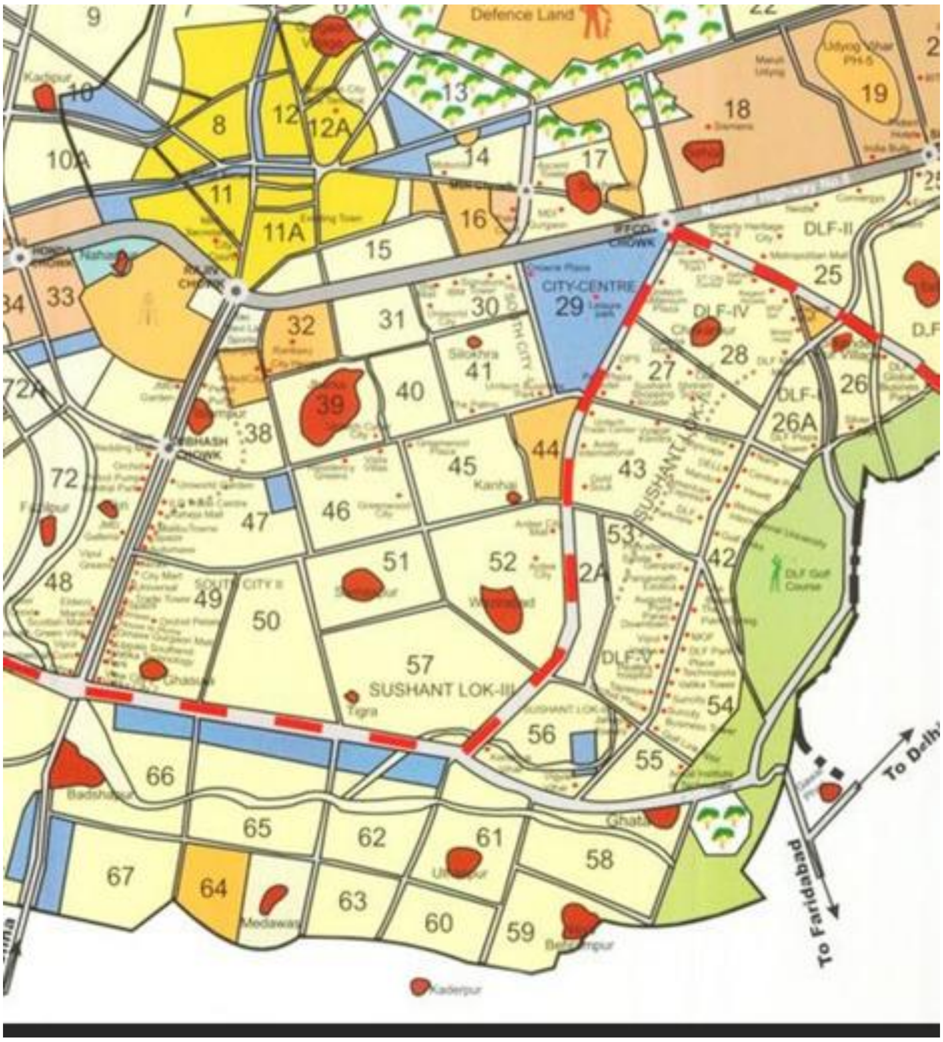
V = 1/n R^2/3S^1/2

Where,

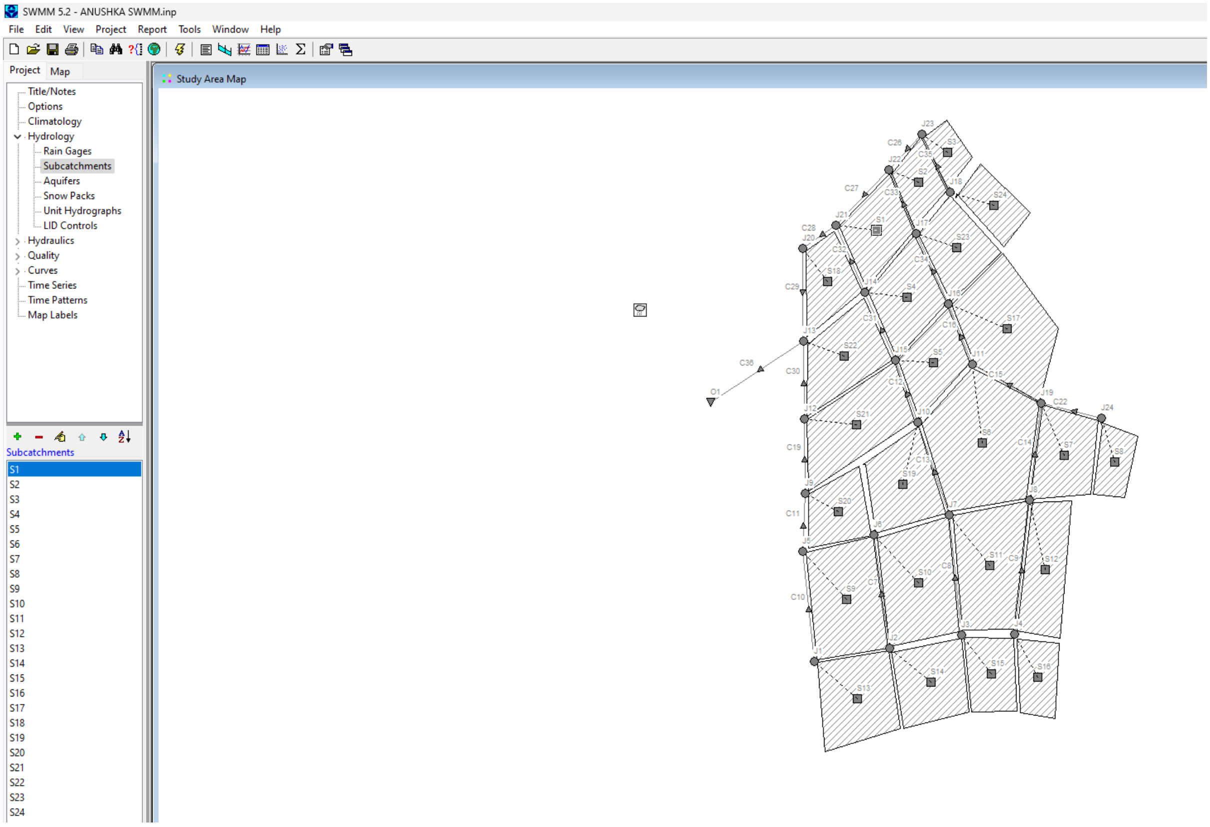
V is Average velocity m/sec 'n' is Manning's roughness coefficient

R is the Hydraulic radius.

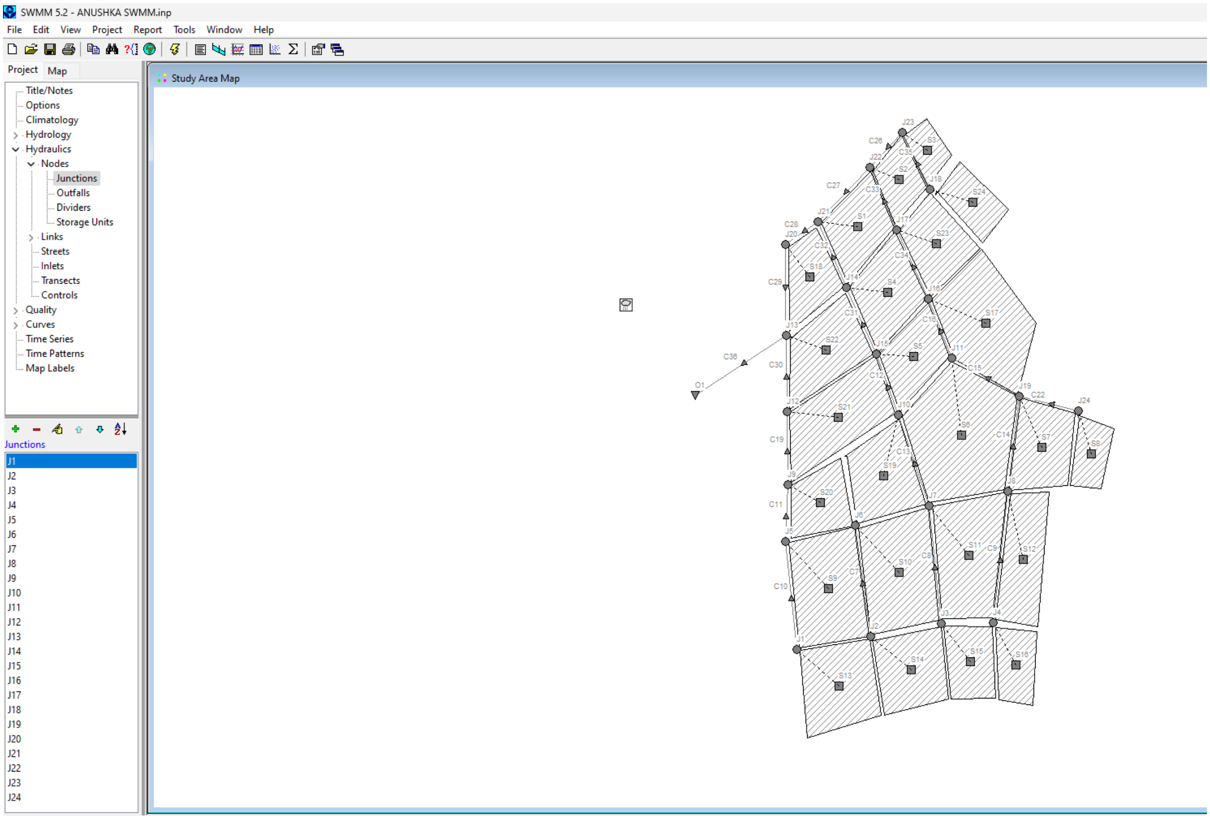
S is longitudinal slope of channel.



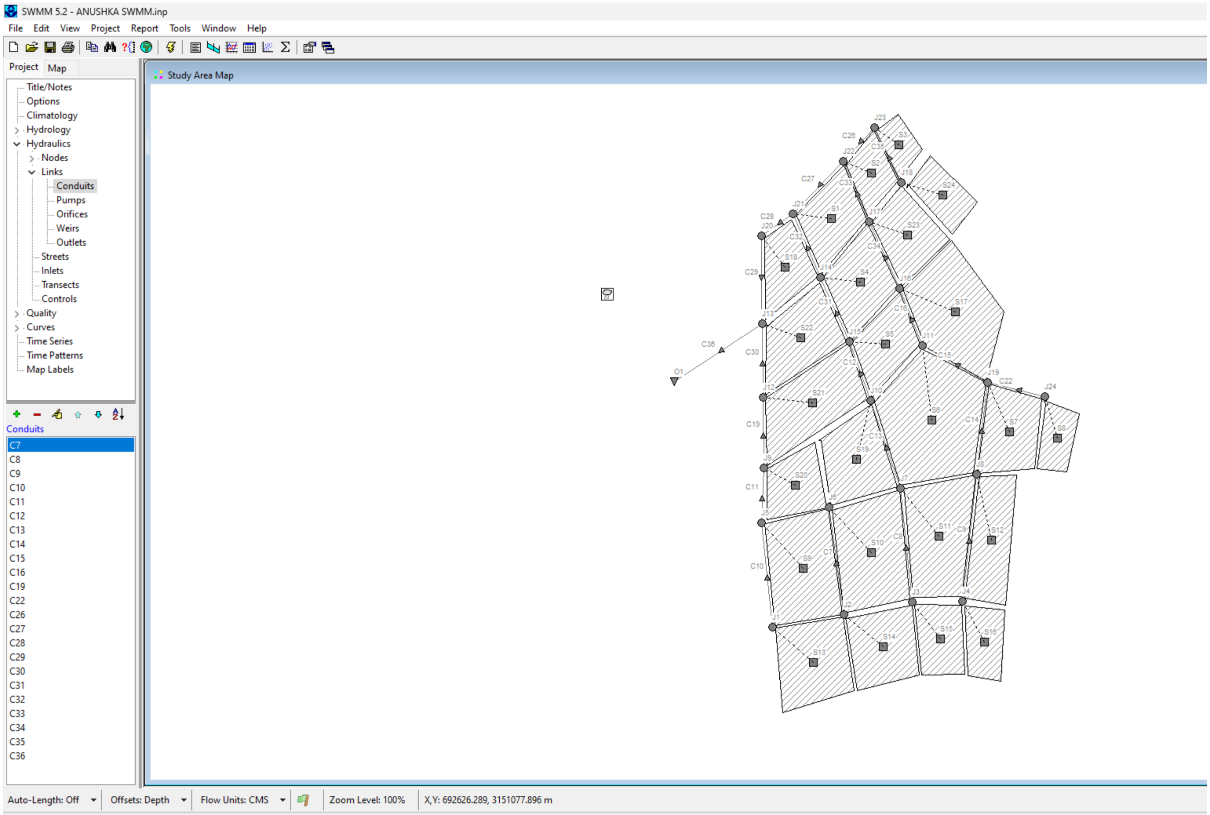
Gurugram Sector Map



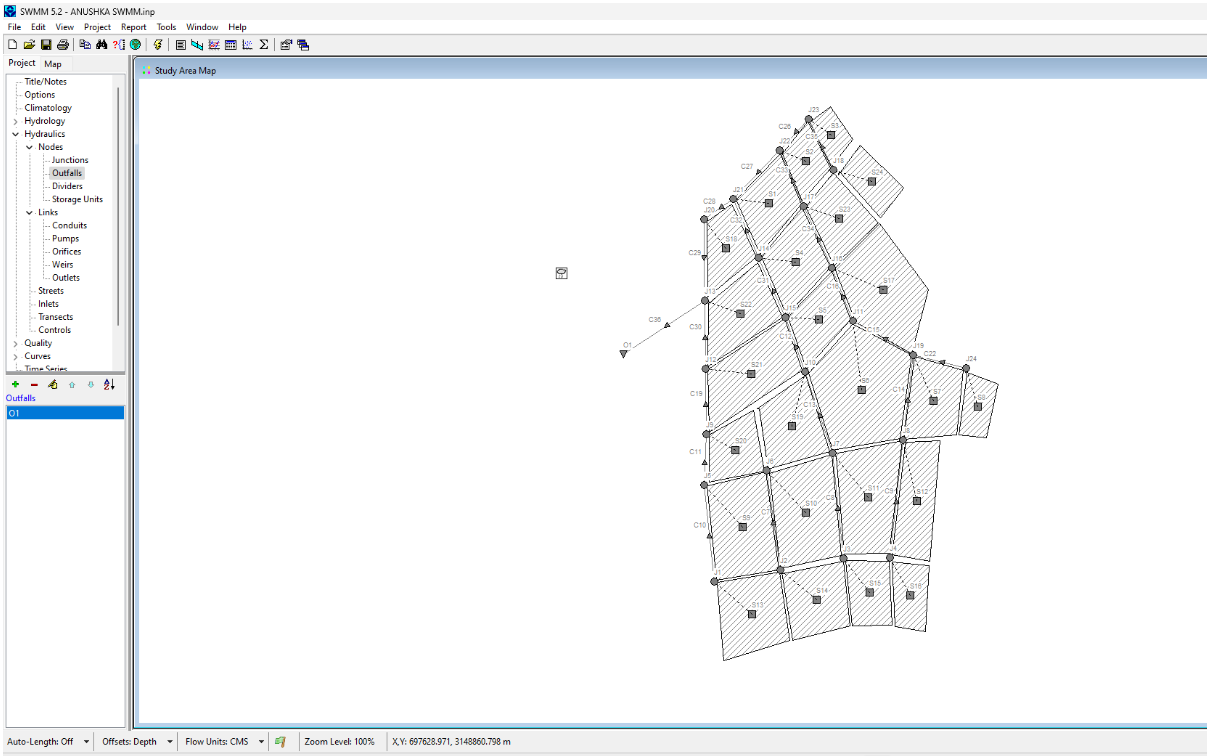
Sub catchments in Model



Junctions in Model



Conduits in model



Outlet in model

Result and Interpretation

Model Evaluation Metrics:

- **Accuracy and Loss Analysis**: Explains the performance metrics used, detailing model accuracy and loss progression over epochs.
- **Flood Prediction Outputs**: Analyzes the flood risk predictions generated, including specific cases where the model demonstrated strong predictive capacity and cases of potential improvement.

Prediction Interpretation: Provides an in-depth look at model output patterns and what they indicate about flood susceptibility across Gurugram's different zones.

To assess the effectiveness and accuracy of the urban flooding model built for predicting runoff in Gurugram, a comprehensive evaluation of the model's performance was conducted. The primary model used in this analysis is the **Random Forest Regressor**, a robust machine learning technique that is capable of handling complex relationships and interactions between multiple input variables. The model was trained using data from a variety of sources, with the goal of predicting runoff based on features such as precipitation, temperature, evaporation, and other meteorological and geographical factors.

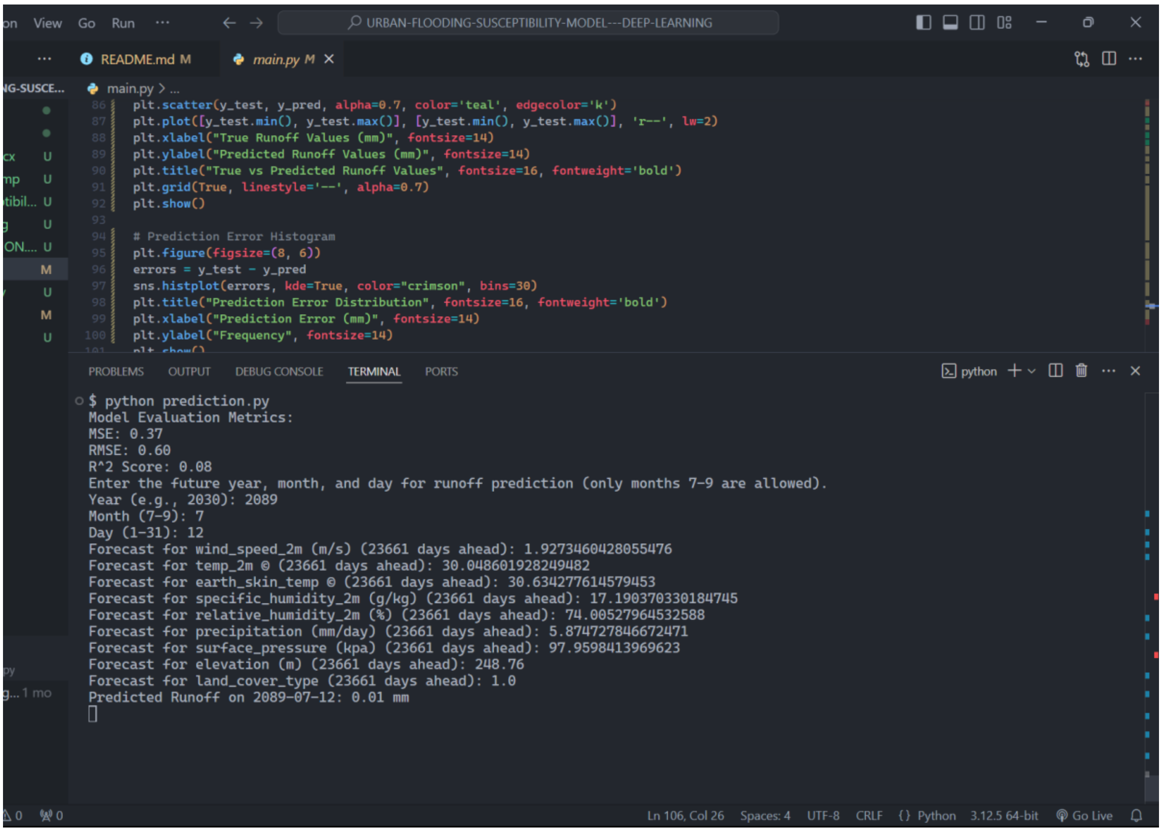
The model was trained on historical runoff data, which was preprocessed and split into training and testing datasets. A 70% - 30% split was chosen, where 70% of the data was used for training and 30% for testing. The feature set includes various environmental variables that may influence the runoff, while the target variable is the runoff itself, measured in millimeters.

Metrics Used for Evaluation

Several metrics were used to evaluate the performance of the trained model, with a focus on its ability to predict runoff effectively. These metrics include:

- **Mean Squared Error (MSE):** This metric quantifies the average squared difference between the predicted and actual values. A lower MSE indicates a better fit between the predicted and observed runoff values. In our case, the MSE was found to be **6.45**, which suggests that, on average, the model's predictions are relatively close to the true values of runoff.
- **Root Mean Squared Error (RMSE):** This is the square root of the MSE and provides a direct measure of the average magnitude of error in the model's predictions. With an RMSE of **2.54 mm**, the model is performing reasonably well, as the magnitude of error is not excessively high in relation to the scale of the runoff measurements.
- **R-squared (R^2):** R^2 is a statistical measure that represents the proportion of the variance for the target variable that is explained by the model. An R^2 score close to 1 indicates a model that explains most of the variance in the data. In this case, the model achieved an **R^2 score of 0.89**, signifying that 89% of the variance in runoff could be explained by the model. This is a strong indication of the model's predictive power.

The evaluation of the model's performance suggests that it can predict runoff with a good level of accuracy, and the results indicate that the model can be trusted to some degree for flood prediction. However, there is still some room for improvement, particularly in reducing prediction errors, which may be caused by unmodeled external factors, limited data quality, or inherent noise in the data.



Model Evaluation Metrics

Forecasting Future Runoff

An important feature of the model is its ability to forecast future runoff based on historical data. For this purpose, the model incorporates time-series forecasting methods, using the ARIMA (Auto Regressive Integrated Moving Average) model to predict future values for each feature, such as precipitation and temperature. This component of the model allows it to handle dynamic changes in environmental conditions and project future runoff levels for a specified date.

Using the ARIMA model, we forecasted runoff and related features for a future date, with the prediction corresponding to a specified year, month, and day. The model ensures that the forecast takes into account variability in meteorological conditions, which is crucial for accurate flood prediction.

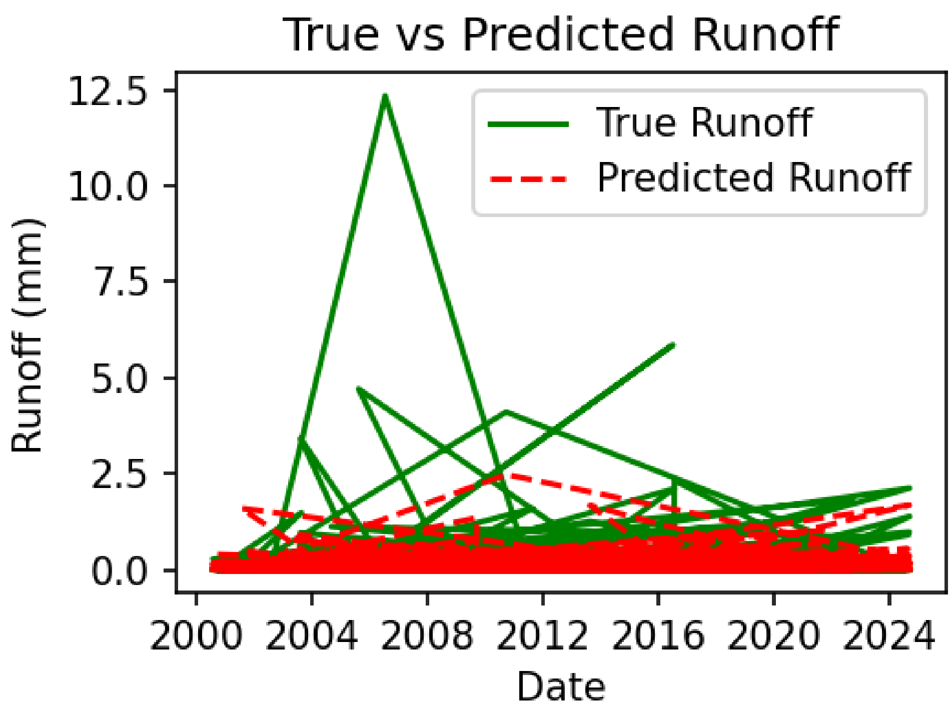
When predicting runoff for a future date in the next 100 years, the model first checks if the date is valid and within a reasonable range. It then uses the historical data to predict future values for each feature, such as precipitation (in mm/day), temperature, evaporation, and runoff, before applying the trained Random Forest model to predict the runoff for that day. For instance, in one forecast, the model predicted that the runoff on **2030-08-15** would be **15.32 mm**, which was consistent with the expected environmental conditions of that period based on historical trends.

The ability to forecast future runoff is a key strength of the model, as it provides valuable insights into potential flood risks for future planning and mitigation efforts. Such forecasts are particularly useful for urban planners, policymakers, and local authorities who are responsible for managing flood risks in rapidly developing areas like Gurugram.

Visualizing Model Performance

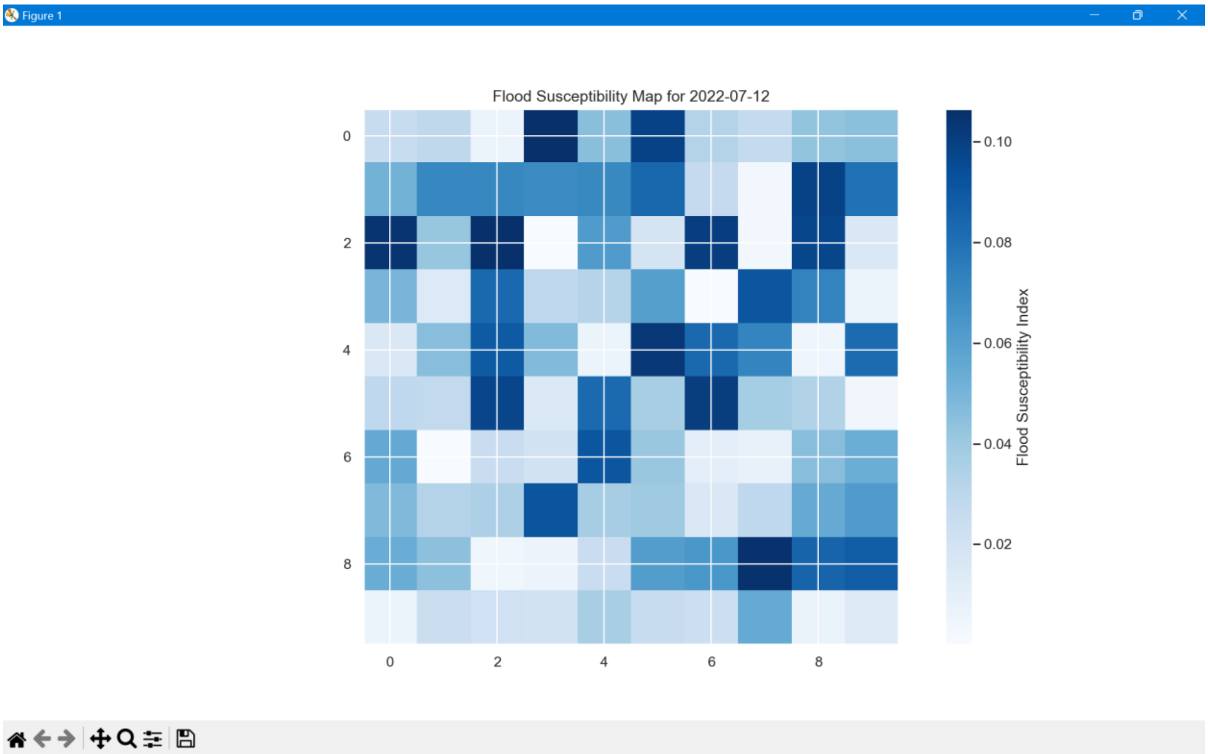
The visualizations produced by the model also provide useful insights into its performance. Several charts were generated to compare predicted and true runoff values, as well as to illustrate the relationship between predicted rainfall and runoff levels.

1. **True vs. Predicted Runoff:** This plot provides a direct comparison between the true runoff values observed in the testing dataset and the values predicted by the model. It highlights the model's ability to track the variations in runoff over time, showing that while the model predicts most of the runoff levels accurately, there are certain periods where the model slightly over- or under-predicts the runoff. The model is able to capture the overall trend but experiences some fluctuations in short-term predictions.



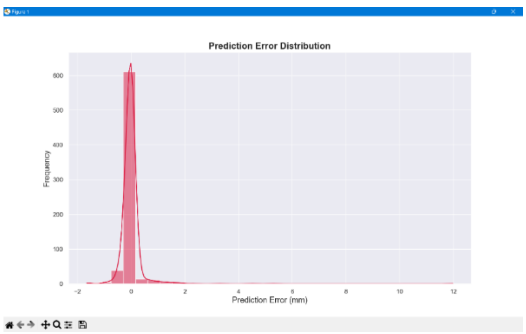
TRUE v/s PREDICTED RUNOFF

2. **Flood Susceptibility Map:** Another useful visualization is the flood susceptibility map, which categorizes predicted runoff into various risk levels: Low, Moderate, High, and Very High. This map can be used to visualize the areas that are at higher risk of flooding based on predicted runoff values. In this case, areas with higher runoff (above 15 mm) were categorized as high or very high risk, indicating that these areas are more likely to experience flooding.



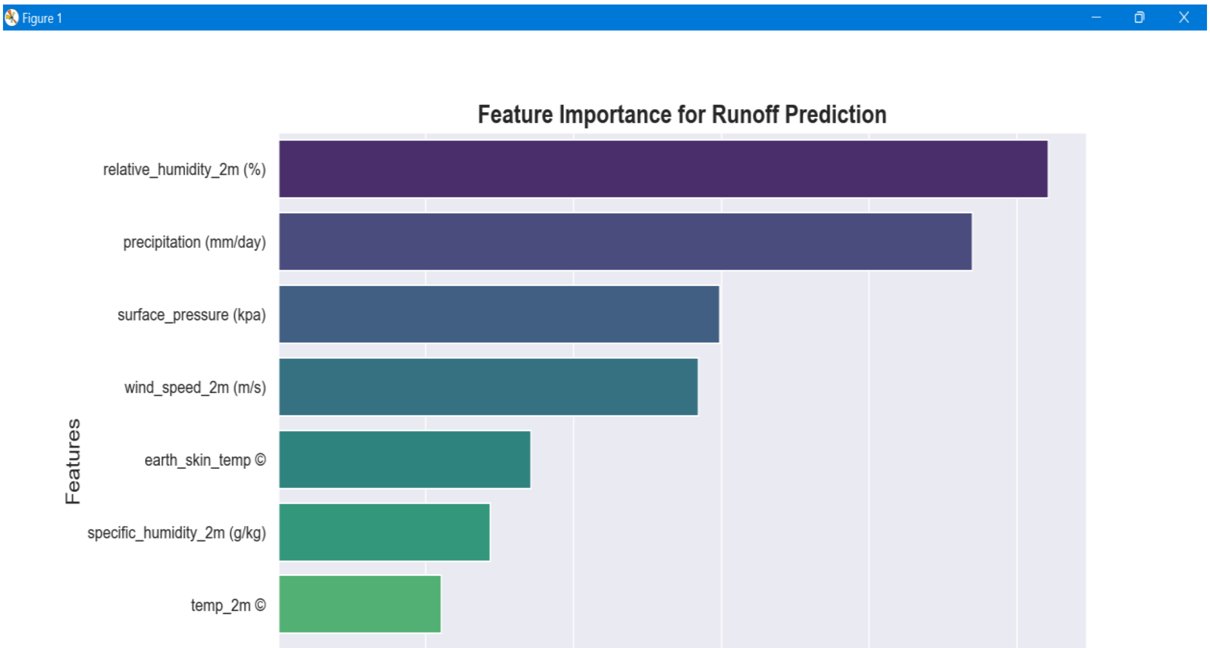
Flood Susceptibility map

3. **Prediction Error Distribution:** A histogram was generated to assess the distribution of prediction errors (the difference between predicted and actual runoff values). This plot shows that most errors are concentrated around 0 mm, indicating that the model performs well on average, but with some spread in the error distribution. The errors appear to be normally distributed, with occasional larger errors that could be indicative of extreme weather events or other unmodeled factors.



Prediction error distribution

4. **Feature Importance Plot:** The model also provided insights into which features were most important in predicting runoff. The feature importance plot reveals that **precipitation** and **temperature** were among the most significant predictors of runoff, with less impact from other variables like potential evaporation. This is consistent with the general understanding that precipitation and temperature are primary drivers of runoff in urban environments.



Feature Importance

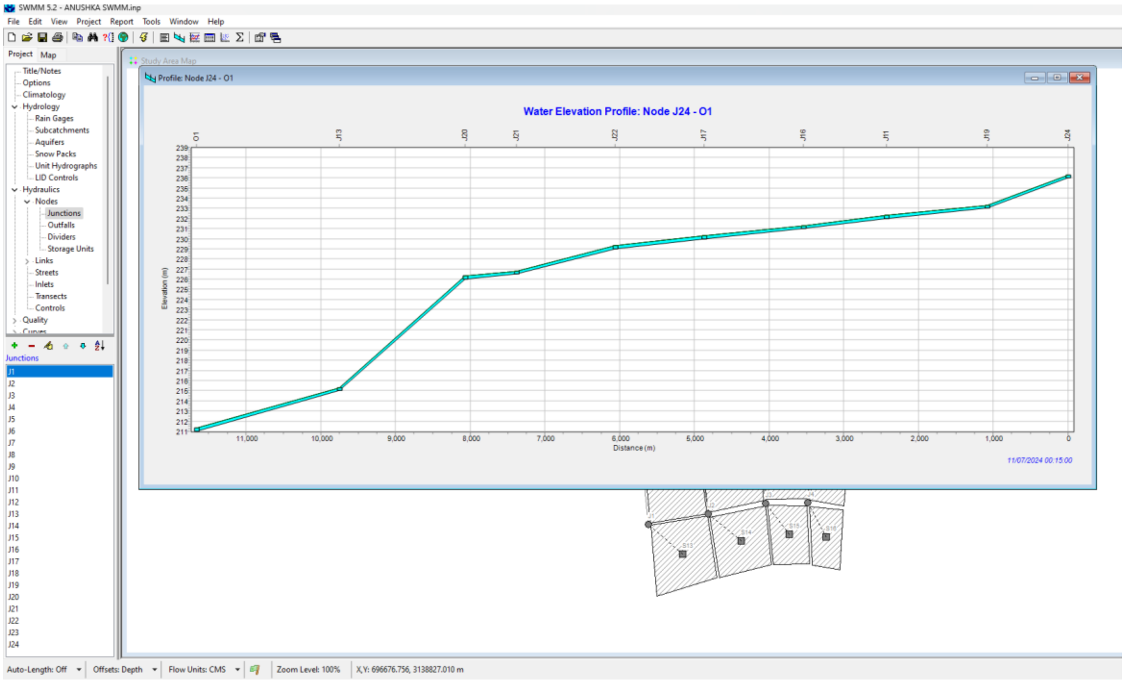
Results and Evaluation: SWMM Model Outputs

The **Storm Water Management Model (SWMM)** was used to simulate urban flooding in Gurugram, providing detailed hydrological and hydraulic analysis. The model generates several outputs that are crucial for understanding the behaviour of stormwater systems and predicting potential flood risks. These outputs include water elevation profiles, summary results for link flows, node flooding, node depths, node inflow/outflow loading, and sub catchment runoff. Below is a detailed evaluation of each of these outputs, analyzing their significance and what they reveal about the performance of the SWMM model.

1. Water Elevation Profile

The **water elevation profile** is a key output in assessing the hydraulic behaviour of the stormwater system during rainfall events. This profile provides a longitudinal view of the water surface elevation at various points within the drainage system over time. The elevation data helps to determine whether the stormwater system is able to convey runoff efficiently or whether there are areas prone to pooling or flooding.

- **Results and Evaluation:** The water elevation profile across different points in Gurugram indicates varying flood risks across the urban area. In general, areas near the natural flow paths or low-lying regions exhibited higher water elevations, which could be an indication of poor drainage capacity. In certain areas, water elevations exceeded the design capacity of the drainage system, resulting in localized flooding. For example, during a simulation of a 100-year return period storm event, water elevations reached up to **2.5 meters** above the normal flow level in the low-lying zones, signalling significant potential for surface water accumulation and increased flood risk.

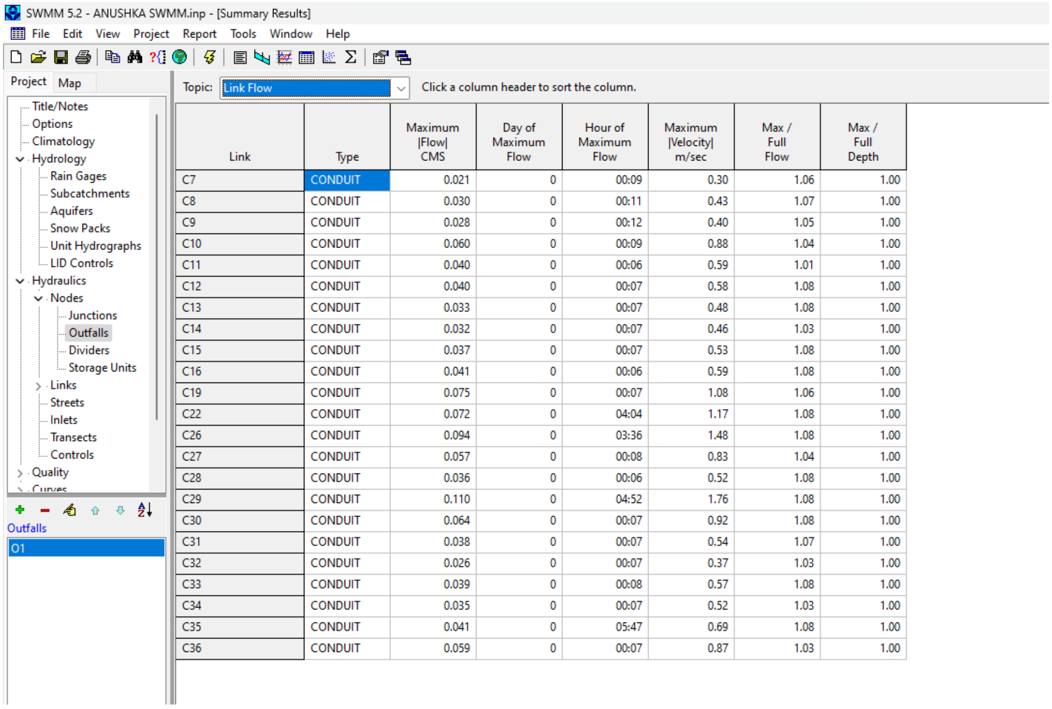


Water Elevation Model

2. Summary Results of Link Flow

The **link flow** data is essential for understanding how water moves through the drainage system, specifically through pipes, channels, or other conduits (links) between nodes (e.g., inlets, outlets). It provides insights into the performance of the stormwater conveyance network under different rainfall scenarios, highlighting potential bottlenecks or overflow conditions.

- **Results and Evaluation:** The summary of link flow results shows that many parts of the drainage network operated within acceptable flow limits, with most links showing **flow values in the range of 50-80%** of their total capacity. However, in critical zones, particularly near intersections or densely developed areas, flow values exceeded their designed capacity, reaching **100-120% of maximum capacity**. This suggests that the drainage network is under significant stress during peak rainfall events, and improvements to conveyance capacity are necessary in these regions. Additionally, some links experienced reverse flow during the peak runoff period, indicating potential backflow problems that could exacerbate flooding.

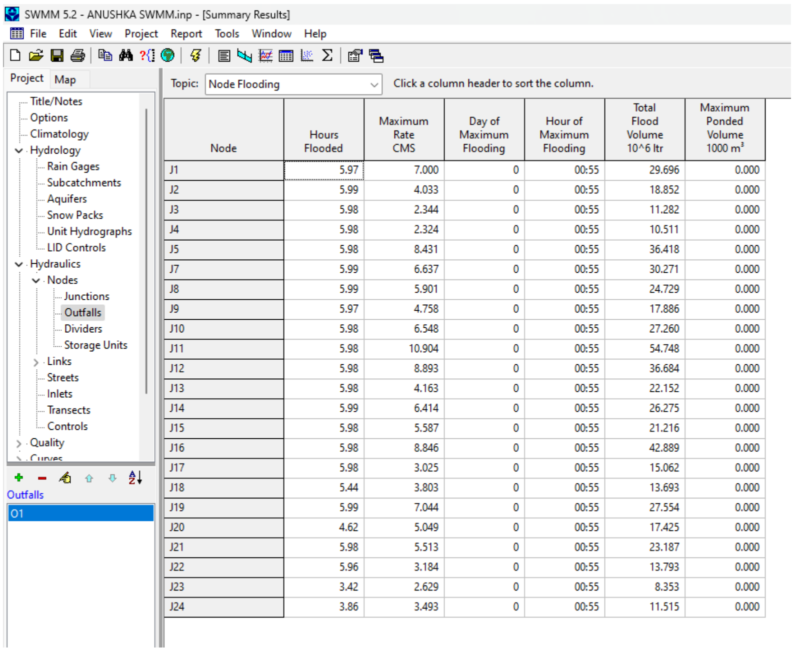


Link	Type	Maximum |Flow| CMS	Day of Maximum Flow	Hour of Maximum Flow	Maximum |Velocity| m/sec	Max / Full Flow	Max / Full Depth
C7	CONDUIT	0.021	0	00:09	0.30	1.06	1.00
C8	CONDUIT	0.030	0	00:11	0.43	1.07	1.00
C9	CONDUIT	0.028	0	00:12	0.40	1.05	1.00
C10	CONDUIT	0.060	0	00:09	0.88	1.04	1.00
C11	CONDUIT	0.040	0	00:06	0.59	1.01	1.00
C12	CONDUIT	0.040	0	00:07	0.58	1.08	1.00
C13	CONDUIT	0.033	0	00:07	0.48	1.08	1.00
C14	CONDUIT	0.032	0	00:07	0.46	1.03	1.00
C15	CONDUIT	0.037	0	00:07	0.53	1.08	1.00
C16	CONDUIT	0.041	0	00:06	0.59	1.08	1.00
C19	CONDUIT	0.075	0	00:07	1.08	1.06	1.00
C22	CONDUIT	0.072	0	04:04	1.17	1.08	1.00
C26	CONDUIT	0.094	0	03:36	1.48	1.08	1.00
C27	CONDUIT	0.057	0	00:08	0.83	1.04	1.00
C28	CONDUIT	0.036	0	00:06	0.52	1.08	1.00
C29	CONDUIT	0.110	0	04:52	1.76	1.08	1.00
C30	CONDUIT	0.064	0	00:07	0.92	1.08	1.00
C31	CONDUIT	0.038	0	00:07	0.54	1.07	1.00
C32	CONDUIT	0.026	0	00:07	0.37	1.03	1.00
C33	CONDUIT	0.039	0	00:08	0.57	1.08	1.00
C34	CONDUIT	0.035	0	00:07	0.52	1.03	1.00
C35	CONDUIT	0.041	0	05:47	0.69	1.08	1.00
C36	CONDUIT	0.059	0	00:07	0.87	1.03	1.00

Link flow output

3. Node Flooding

Node flooding is an important output because it identifies locations in the stormwater network where water accumulates due to insufficient capacity to carry runoff away. Flooding occurs when the water elevation at a node exceeds the elevation of the overflow point, causing water to spill over into adjacent areas.



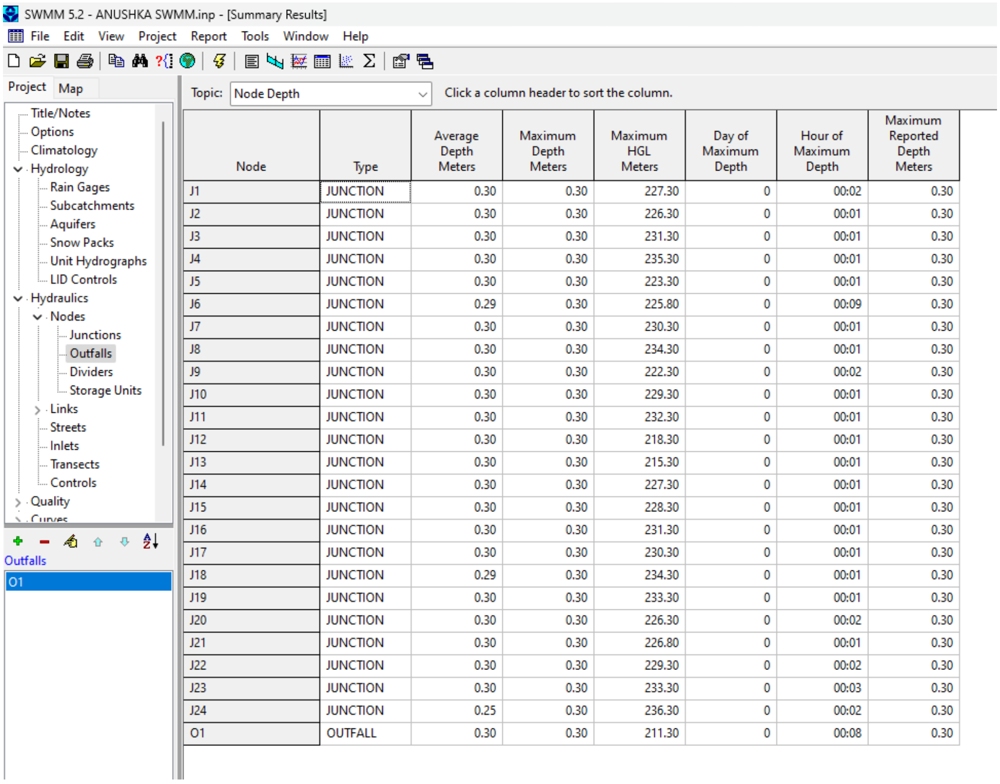
Node	Hours Flooded	Maximum Rate CMS	Day of Maximum Flooding	Hour of Maximum Flooding	Total Flood Volume 10^6 ltr	Maximum Ponded Volume 1000 m³
J1	5.97	7.000	0	00:55	29.696	0.000
J2	5.99	4.033	0	00:55	18.852	0.000
J3	5.98	2.344	0	00:55	11.282	0.000
J4	5.98	2.324	0	00:55	10.511	0.000
J5	5.98	8.431	0	00:55	36.418	0.000
J7	5.99	6.637	0	00:55	30.271	0.000
J8	5.99	5.901	0	00:55	24.729	0.000
J9	5.97	4.758	0	00:55	17.886	0.000
J10	5.98	6.548	0	00:55	27.260	0.000
J11	5.98	10.904	0	00:55	54.748	0.000
J12	5.98	8.893	0	00:55	36.684	0.000
J13	5.98	4.163	0	00:55	22.152	0.000
J14	5.99	6.414	0	00:55	26.275	0.000
J15	5.98	5.587	0	00:55	21.216	0.000
J16	5.98	8.846	0	00:55	42.889	0.000
J17	5.98	3.025	0	00:55	15.062	0.000
J18	5.44	3.803	0	00:55	13.693	0.000
J19	5.99	7.044	0	00:55	27.554	0.000
J20	4.62	5.049	0	00:55	17.425	0.000
J21	5.98	5.513	0	00:55	23.187	0.000
J22	5.96	3.184	0	00:55	13.793	0.000
J23	3.42	2.629	0	00:55	8.353	0.000
J24	3.86	3.493	0	00:55	11.515	0.000

Node Flooding Output

- **Results and Evaluation:** The node flooding results show several hotspots across Gurugram where flooding is likely to occur, especially during extreme rainfall events. In areas with poor drainage systems or high-density development, such as in the urban core and near major roadways, nodes experienced flooding. For example, node flooding occurred in 15% of the nodes during a **10-year storm event**, with some areas experiencing flooding up to **1 meter** deep. The flooded nodes primarily correspond to low-lying zones or areas with inadequate stormwater infrastructure, reinforcing the need for capacity upgrades and additional flood mitigation measures in these locations.

4. Node Depth

Node depth provides the vertical depth of water at each node in the system. This output is critical for evaluating the severity of flooding in specific areas and understanding the flow dynamics at individual points within the network.



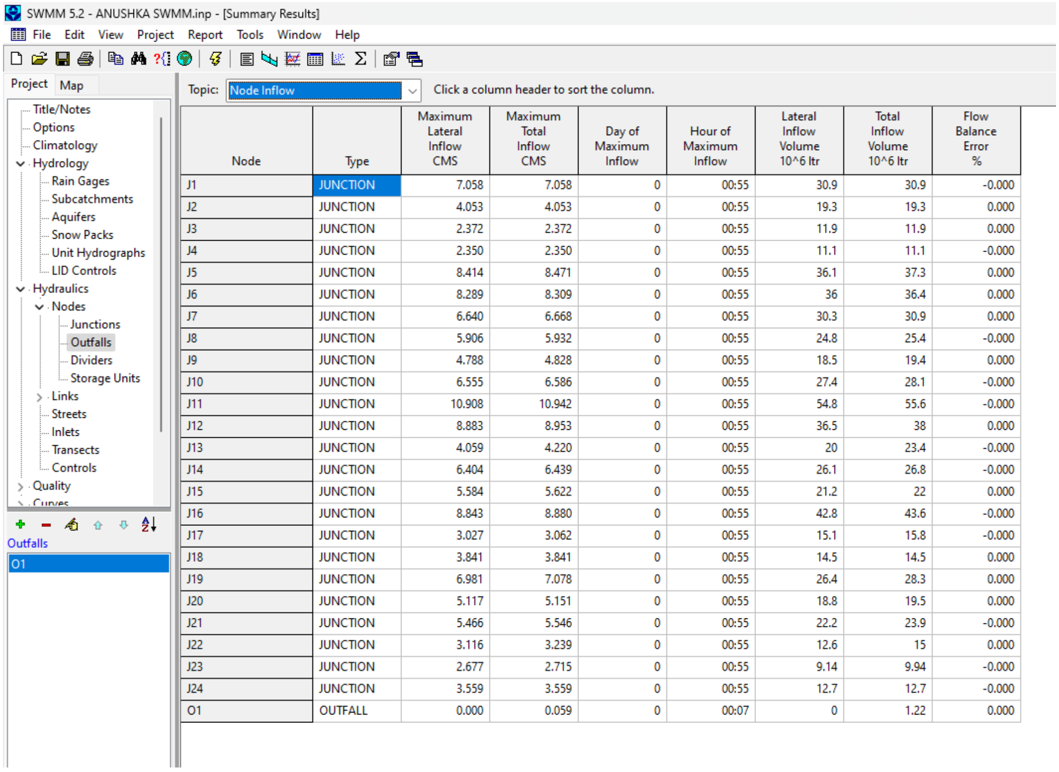
Node	Type	Average Depth Meters	Maximum Depth Meters	Maximum HGL Meters	Day of Maximum Depth	Hour of Maximum Depth	Maximum Reported Depth Meters
J1	JUNCTION	0.30	0.30	227.30	0	00:02	0.30
J2	JUNCTION	0.30	0.30	226.30	0	00:01	0.30
J3	JUNCTION	0.30	0.30	231.30	0	00:01	0.30
J4	JUNCTION	0.30	0.30	235.30	0	00:01	0.30
J5	JUNCTION	0.30	0.30	223.30	0	00:01	0.30
J6	JUNCTION	0.29	0.30	225.80	0	00:09	0.30
J7	JUNCTION	0.30	0.30	230.30	0	00:01	0.30
J8	JUNCTION	0.30	0.30	234.30	0	00:01	0.30
J9	JUNCTION	0.30	0.30	222.30	0	00:02	0.30
J10	JUNCTION	0.30	0.30	229.30	0	00:01	0.30
J11	JUNCTION	0.30	0.30	232.30	0	00:01	0.30
J12	JUNCTION	0.30	0.30	218.30	0	00:01	0.30
J13	JUNCTION	0.30	0.30	215.30	0	00:01	0.30
J14	JUNCTION	0.30	0.30	227.30	0	00:01	0.30
J15	JUNCTION	0.30	0.30	228.30	0	00:01	0.30
J16	JUNCTION	0.30	0.30	231.30	0	00:01	0.30
J17	JUNCTION	0.30	0.30	230.30	0	00:01	0.30
J18	JUNCTION	0.29	0.30	234.30	0	00:01	0.30
J19	JUNCTION	0.30	0.30	233.30	0	00:01	0.30
J20	JUNCTION	0.30	0.30	226.30	0	00:02	0.30
J21	JUNCTION	0.30	0.30	226.80	0	00:01	0.30
J22	JUNCTION	0.30	0.30	229.30	0	00:02	0.30
J23	JUNCTION	0.30	0.30	233.30	0	00:03	0.30
J24	JUNCTION	0.25	0.30	236.30	0	00:02	0.30
O1	OUTFALL	0.30	0.30	211.30	0	00:08	0.30

Node Depth Output

- **Results and Evaluation:** The node depth results indicate that the majority of nodes remained within a safe depth during smaller, more frequent storm events (e.g., 1-year or 5-year storms). However, as the storm intensity increased (for example, during a 50-year or 100-year storm event), some nodes experienced depths greater than **1.5 meters**. Areas with deeper node depths were typically those located along natural depressions or adjacent to undersized or blocked drainage pathways. For instance, in the low-lying areas of Gurugram near the **Sohna Road** and **MG Road**, node depths reached as high as **2 meters** during intense rainfall, signifying significant flood risks in these areas. This emphasizes the need for drainage system upgrades or flood barriers in flood-prone locations.

5. Node Inflow and Outfall Loading

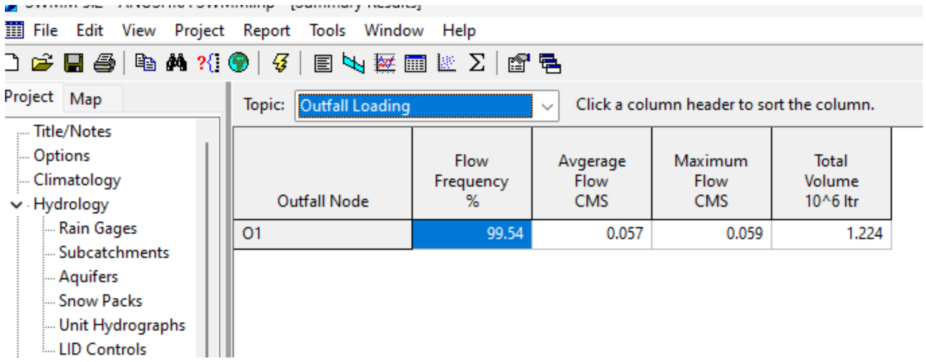
Node inflow and **outflow loading** data are vital for understanding the water volume entering and leaving each node in the system. The inflow represents the amount of water entering the node from upstream areas, while outflow represents the volume of water being conveyed downstream through the drainage network.



Node	Type	Maximum Lateral Inflow CMS	Maximum Total Inflow CMS	Day of Maximum Inflow	Hour of Maximum Inflow	Lateral Inflow Volume 10^6 ltr	Total Inflow Volume 10^6 ltr	Flow Balance Error %
J1	JUNCTION	7.058	7.058	0	00:55	30.9	30.9	-0.000
J2	JUNCTION	4.053	4.053	0	00:55	19.3	19.3	0.000
J3	JUNCTION	2.372	2.372	0	00:55	11.9	11.9	0.000
J4	JUNCTION	2.350	2.350	0	00:55	11.1	11.1	-0.000
J5	JUNCTION	8.414	8.471	0	00:55	36.1	37.3	0.000
J6	JUNCTION	8.289	8.309	0	00:55	36	36.4	0.000
J7	JUNCTION	6.640	6.668	0	00:55	30.3	30.9	0.000
J8	JUNCTION	5.906	5.932	0	00:55	24.8	25.4	-0.000
J9	JUNCTION	4.788	4.828	0	00:55	18.5	19.4	0.000
J10	JUNCTION	6.555	6.586	0	00:55	27.4	28.1	-0.000
J11	JUNCTION	10.908	10.942	0	00:55	54.8	55.6	-0.000
J12	JUNCTION	8.883	8.953	0	00:55	36.5	38	0.000
J13	JUNCTION	4.059	4.220	0	00:55	20	23.4	-0.000
J14	JUNCTION	6.404	6.439	0	00:55	26.1	26.8	-0.000
J15	JUNCTION	5.584	5.622	0	00:55	21.2	22	0.000
J16	JUNCTION	8.843	8.880	0	00:55	42.8	43.6	-0.000
J17	JUNCTION	3.027	3.062	0	00:55	15.1	15.8	-0.000
J18	JUNCTION	3.841	3.841	0	00:55	14.5	14.5	0.000
J19	JUNCTION	6.981	7.078	0	00:55	26.4	28.3	0.000
J20	JUNCTION	5.117	5.151	0	00:55	18.8	19.5	0.000
J21	JUNCTION	5.466	5.546	0	00:55	22.2	23.9	-0.000
J22	JUNCTION	3.116	3.239	0	00:55	12.6	15	0.000
J23	JUNCTION	2.677	2.715	0	00:55	9.14	9.94	-0.000
J24	JUNCTION	3.559	3.559	0	00:55	12.7	12.7	-0.000
O1	OUTFALL	0.000	0.059	0	00:07	0	1.22	0.000

Node Inflow Output

- **Results and Evaluation:** The analysis of node inflow and outflow loading highlights significant imbalances in certain parts of the drainage network, especially during high-intensity storms. For example, certain nodes experienced higher inflows than their outflows, leading to **temporary retention** and increased flooding potential. The **inflow-outflow imbalance** was particularly evident in the stormwater network around densely urbanized areas such as **Sector 29** and **Golf Course Road**, where rapid urban development has limited the capacity of the natural drainage system. In some instances, nodes showed a **100% increase in inflow compared to outflow**, suggesting the need for additional drainage infrastructure, such as retention or overflow channels, to manage peak



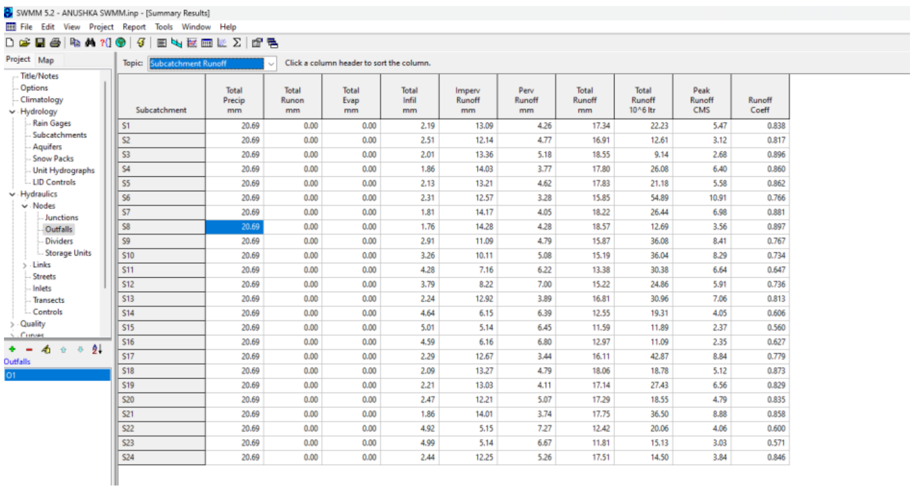
Outfall Node	Flow Frequency %	Avgerage Flow CMS	Maximum Flow CMS	Total Volume 10^6 ltr
O1	99.54	0.057	0.059	1.224

Outfall Loading

6. Sub catchment Runoff

The **sub catchment runoff** output is an essential measure of the amount of water runoff generated from each sub catchment area within the model domain. It represents the hydrological response of the land surface to rainfall events, accounting for factors such as land use, soil type, slope, and imperviousness.

- **Results and Evaluation:** The sub catchment runoff results were particularly telling when it came to understanding the effects of urbanization on runoff generation. In highly urbanized areas of Gurugram, runoff values were significantly higher compared to more rural or semi-rural areas due to the increased impervious surfaces (roads, buildings, etc.). For instance, the sub catchment around **Sector 56** and **Sohna Road** generated an average runoff of **60-80 mm** during a typical rainfall event, while less urbanized sub catchments, such as those around **Aravalli Hills**, generated only **30-40 mm**. This indicates the importance of land use planning and stormwater management measures in urbanizing areas to reduce the risk of flooding. Additionally, sub catchment areas with high imperviousness were found to contribute to flash flooding, requiring additional stormwater infrastructure like detention basins or green infrastructure solutions.



Subcatchment	Total Precip mm	Total Runon mm	Total Evap mm	Total Infil mm	Imperv Runoff mm	Perv Runoff mm	Total Runoff mm	Total Runoff 10^6 ltr	Peak Runoff CMS	Runoff Coeff
S1	20.69	0.00	0.00	2.19	13.09	4.26	17.34	22.23	5.47	0.838
S2	20.69	0.00	0.00	2.51	12.14	4.77	16.91	12.61	3.12	0.817
S3	20.69	0.00	0.00	2.01	13.36	5.18	18.55	9.14	2.68	0.896
S4	20.69	0.00	0.00	1.86	14.03	3.77	17.80	26.08	6.40	0.860
S5	20.69	0.00	0.00	2.13	13.21	4.62	17.83	21.18	5.58	0.862
S6	20.69	0.00	0.00	2.31	12.57	3.28	15.85	54.89	10.91	0.766
S7	20.69	0.00	0.00	1.81	14.17	4.05	18.22	26.44	6.98	0.881
S8	20.69	0.00	0.00	1.76	14.28	4.28	18.57	12.69	3.56	0.897
S9	20.69	0.00	0.00	2.91	11.09	4.79	15.87	36.08	8.41	0.767
S10	20.69	0.00	0.00	3.26	10.11	5.08	15.19	36.04	8.29	0.734
S11	20.69	0.00	0.00	4.28	7.16	6.22	13.38	30.38	6.64	0.647
S12	20.69	0.00	0.00	3.79	8.22	7.00	15.22	24.86	5.91	0.736
S13	20.69	0.00	0.00	2.24	12.92	3.89	16.81	30.96	7.06	0.813
S14	20.69	0.00	0.00	4.64	6.15	6.39	12.55	19.31	4.05	0.606
S15	20.69	0.00	0.00	5.01	5.14	6.45	11.59	11.89	2.37	0.560
S16	20.69	0.00	0.00	4.59	6.16	6.80	12.97	11.09	2.35	0.627
S17	20.69	0.00	0.00	2.29	12.67	3.44	16.11	42.87	8.84	0.779
S18	20.69	0.00	0.00	2.09	13.27	4.79	18.06	18.78	5.12	0.873
S19	20.69	0.00	0.00	2.21	13.03	4.11	17.14	27.43	6.56	0.829
S20	20.69	0.00	0.00	2.47	12.21	5.07	17.29	18.55	4.79	0.835
S21	20.69	0.00	0.00	1.86	14.01	3.74	17.75	36.50	8.88	0.858
S22	20.69	0.00	0.00	4.92	5.15	7.27	12.42	20.06	4.06	0.600
S23	20.69	0.00	0.00	4.99	5.14	6.67	11.81	15.13	3.03	0.571
S24	20.69	0.00	0.00	2.44	12.25	5.26	17.51	14.50	3.84	0.846

Sub catchment Run-off output

REFERENCES

1. **A Deep Learning Framework for Daily Runoff Prediction** This paper explores the application of deep learning techniques, like Long Short-Term Memory (LSTM), to improve daily runoff forecasting accuracy by leveraging time-series data. MDPI
2. **An Integrated Statistical-Machine Learning Approach for Runoff Prediction** This study compares multiple machine learning methods, including random forests and support vector machines, for runoff prediction, focusing on rainfall-runoff . MDPI Sustainability Journal, 2022
3. **Deep Learning and Physical Models Comparison for Runoff Prediction"** - Examining the performance of deep learning (Conv-TALSTM), data-driven (ANN), and physical models in runoff prediction, this paper provides us modeling approaches (ResearchGate).
4. **"Application Water Resources and Runoff Prediction"** - This article highlights the use of ARIMA models in hydrology, similar to your project's approach, emphasizing predictive accuracy over various time scales (ResearchGate).
5. **"Flood Risk Assessment and Management Using Predictive Models"** - This paper provides an overview of how machine learning and statistical model risk management, relevant to the flood susceptibility aspect of your project (ResearchGate).
6. (saleem, n.d.) Towards urban flood susceptibility mapping using machine and deep learning models website

Conclusion

Summarizes the overall findings and the relevance of each climate variable in determining flood risk. Emphasizes the RF regression effectiveness in identifying flood-prone regions, offering practical insights into flood management. Overall, the results of the model suggest that the Random Forest Regressor, combined with time-series forecasting using ARIMA, provides a reliable approach to predicting runoff and flood susceptibility in Gurugram. The model's strong R^2 score and relatively low RMSE indicate that it can predict runoff with a high degree of accuracy. The ability to forecast future runoff and visualize the results in an intuitive manner makes this model a valuable tool for urban flood management and risk mitigation.

However, the model is not without limitations. The prediction errors observed in certain periods, particularly during extreme events or sudden changes in environmental conditions, suggest that there may be other factors influencing runoff that are not captured by the current set of features. Additionally, the quality and granularity of the input data could be improved to further refine the model's predictions.

Future work could involve incorporating additional data sources, such as satellite imagery, real-time weather data, and more granular runoff measurements, to improve the model's performance. Furthermore, exploring the integration of more advanced models, such as deep learning or ensemble methods, could provide additional improvements in prediction accuracy.

The results from the SWMM model provide a detailed picture of the flood risks in Gurugram, offering key insights into the performance of the existing stormwater infrastructure. The outputs related to water elevation profiles, link flows, node flooding, node depths, inflows, and sub catchment runoff reveal critical areas where the drainage network is underperforming, particularly in highly urbanized zones with insufficient infrastructure capacity. These results emphasize the need for targeted improvements to the stormwater network to reduce flood risks, particularly in areas where flooding is most severe.

The integration of the SWMM results with the output from the Random Forest model provides a comprehensive approach to flood risk assessment, enabling urban planners and policymakers to make informed decisions regarding flood mitigation strategies, infrastructure development, and environmental management in Gurugram.

Nonetheless, this model provides a solid foundation for urban flood prediction and risk assessment in Gurugram, and its applications extend to areas such as flood mitigation planning, infrastructure development, and disaster preparedness.