1. Data Collection: Obtain a dataset of URLs labeled with their corresponding types (e.g., phishing, benign).

2. Data Preprocessing:

• Tokenize the URLs: Convert the URLs into sequences of tokens (words or characters).

• Padding: Ensure that all sequences are of the same length by padding shorter sequences with zeros.

• Encode Labels: Encode the categorical labels (e.g., phishing, benign) into numerical format.

3. Split the Data: Split the dataset into training and validation sets to evaluate the model's performance.

4. Build a Deep Learning Model:

• Define the model architecture: Choose the type of deep learning model (e.g., LSTM, CNN) and the number of layers.

• Add an Embedding layer: Map the tokens to dense vectors of fixed size.

• Add LSTM layers: LSTM (Long Short-Term Memory) layers are effective for sequence processing tasks.

• Add Dense layers: Use one or more Dense layers for classification.

• Compile the model: Specify the optimizer, loss function, and metrics to use during training.

5. Train the Model:

• Feed the training data into the model.

• Specify the number of epochs (iterations over the entire dataset) and batch size.

• Monitor the training process and adjust hyperparameters if needed.

6. Evaluate the Model:

• Use the validation set to evaluate the model's performance.

• Calculate metrics such as accuracy, precision, recall, and F1-score.

7. Make Predictions:

• Use the trained model to predict the types of URLs in new data.

8. Deploy the Model:

• Once satisfied with the model's performance, deploy it to detect malicious URLs in real-world applications.