Robustness Testing: Data Drafting to Model Defending
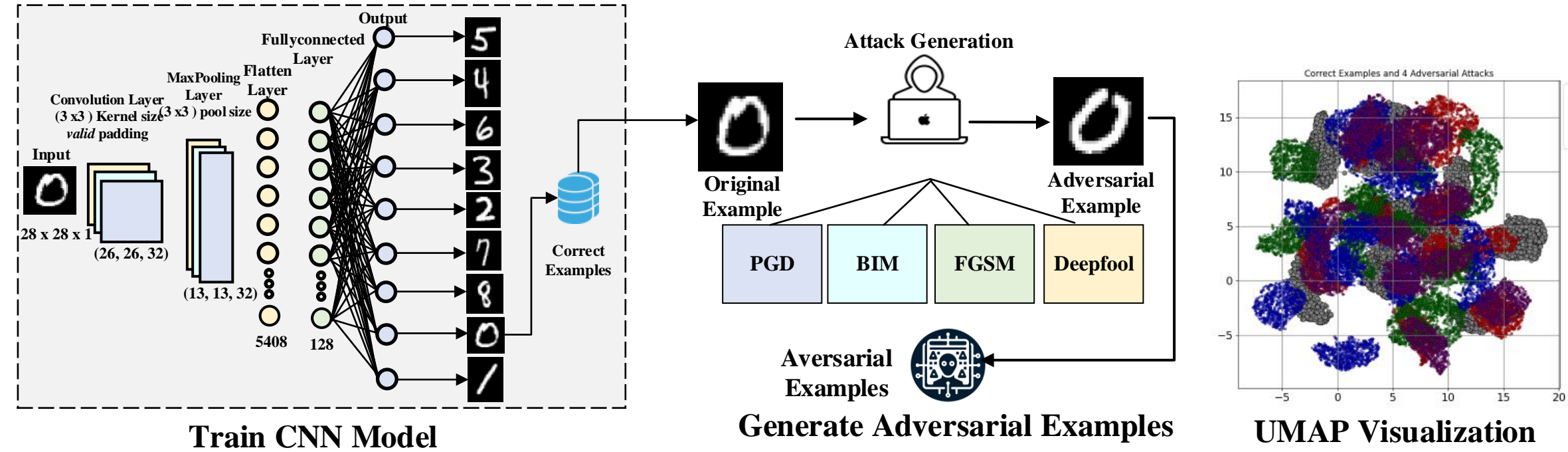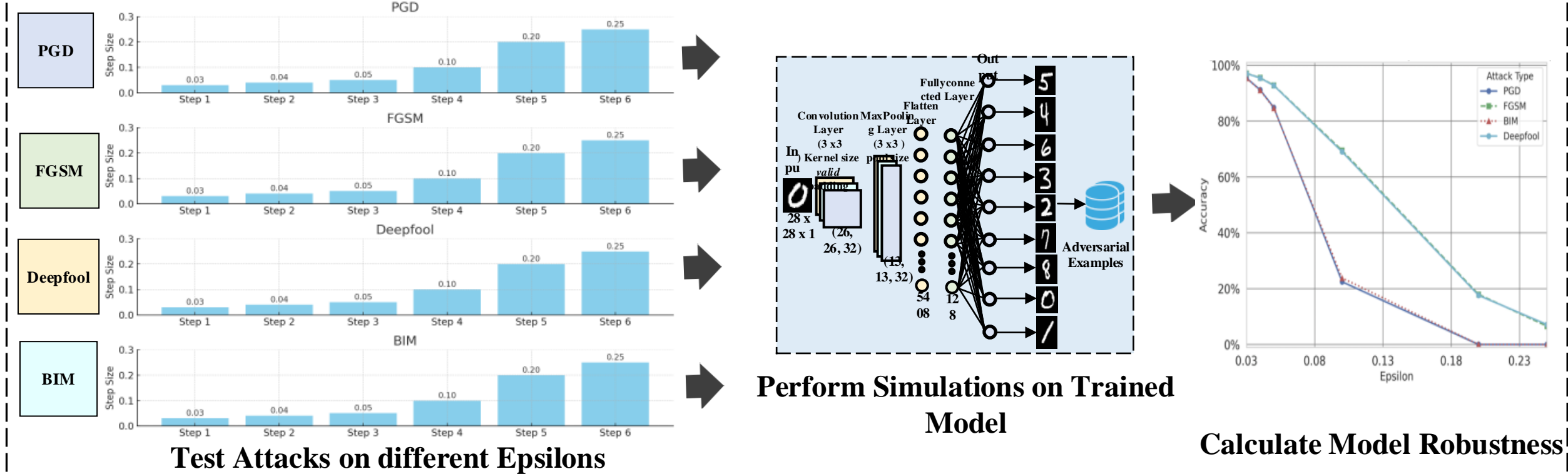
Stage 1: Data Preparation (Correct & Adversarial Examples)
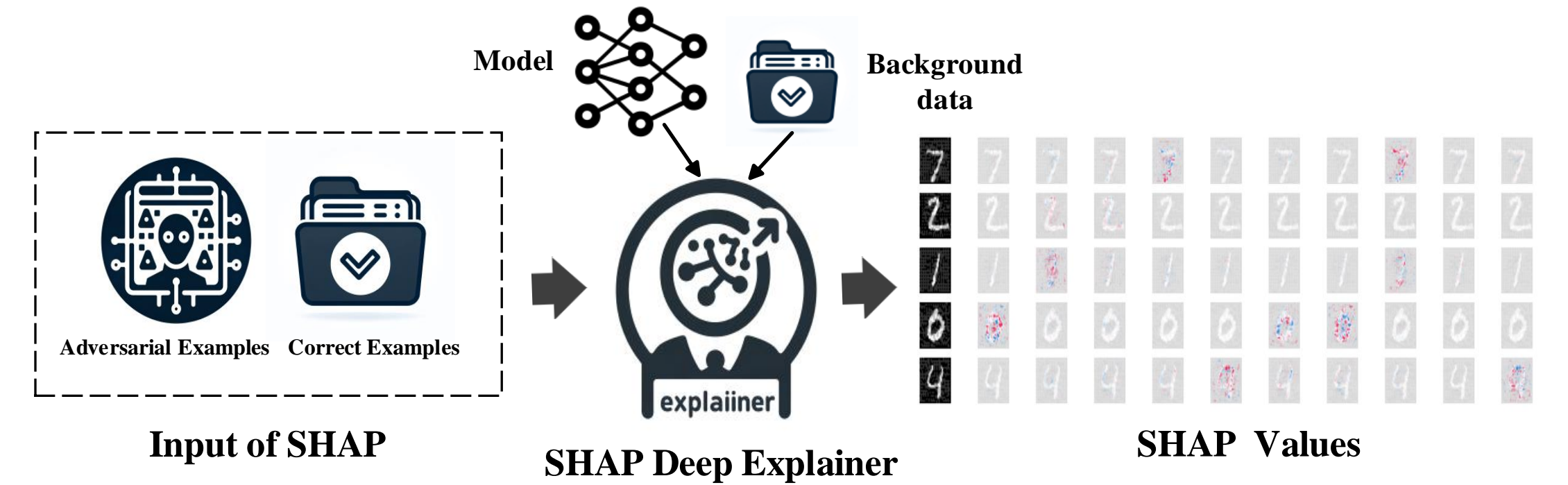
Train CNN Model

Generate Adversarial Examples

UMAP Visualization

Stage 2: Model Robustness Against Atatcks

Test Attacks on different Epsilons

Perform Simulations on Trained Model

Calculate Model Robustness

Comprehensive SHAP Pixel Analysis and Validation

Stage 3: Calculate SHAP Signatures

Input of SHAP

SHAP Deep Explainer

SHAP Values

Stage 4: Pixel-Level Analysis

Influence Insights: Pixel Power at 70%, 80%, 90%

Pixel Power: Tweaking Images Between Normal and Adversarial

Pixel Impact Verification via T-Test