The purpose of our research is to develop a meta-framework for anomaly detection in time series data. This framework is capable of performing anomaly detection across various domains.

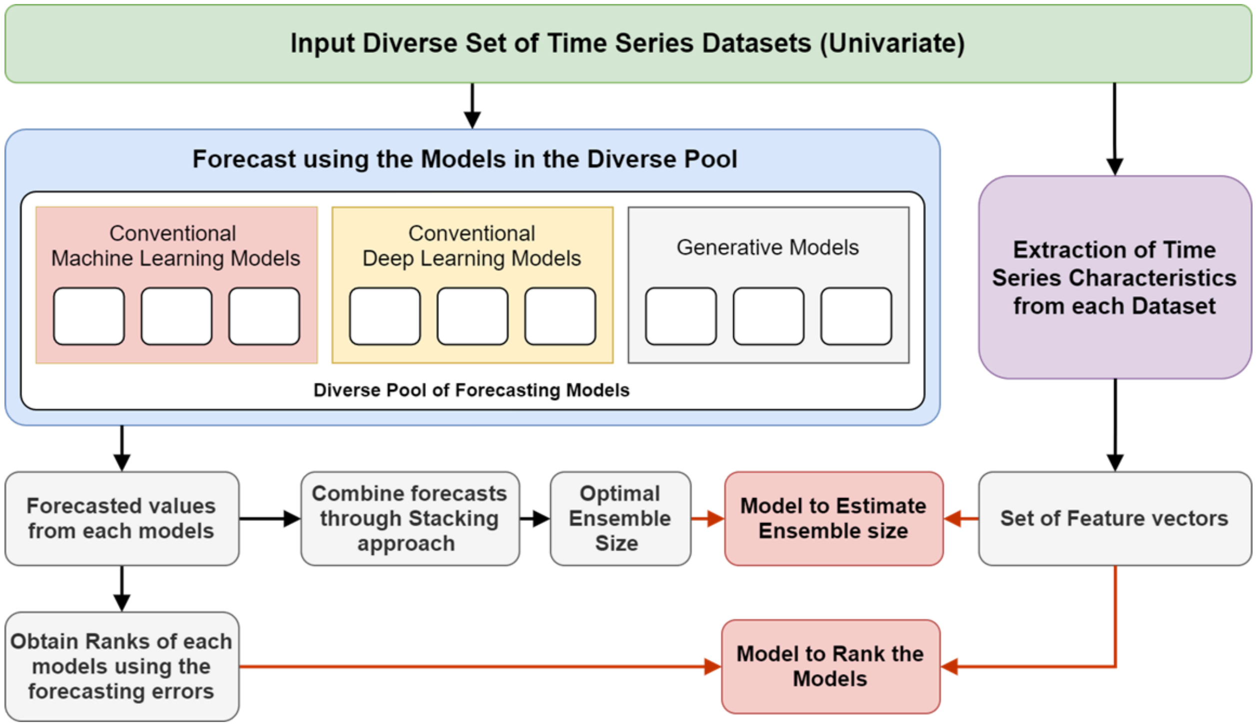

Figure 1

Figure 1 illustrates the training process of our system. We collected approximately 50 different time series datasets from a variety of domains, such as temperature sensors, real-time traffic data, online advertisement click rates, computer usage patterns, etc. From each time series, we extracted relevant features and characteristics.

Next, we selected the best forecasting models from three categories: conventional machine learning, conventional deep learning, and generative models. Each time series was then trained using all the top-performing models from each of these three categories. This resulted in a set of forecasted values from each model, which we compared against the ground truth values to evaluate accuracy.

Using this information, we were able to rank the models based on their forecasting accuracy across all 50 time series. We also applied a stacking approach to combine the forecasts, enabling us to determine the optimal ensemble size that maximizes prediction accuracy.

As a result, we obtained both the rankings and the optimal ensemble size for specific time series characteristics. Based on this data, we developed two predictive models: one to estimate the ideal ensemble size and another to rank forecasting models.

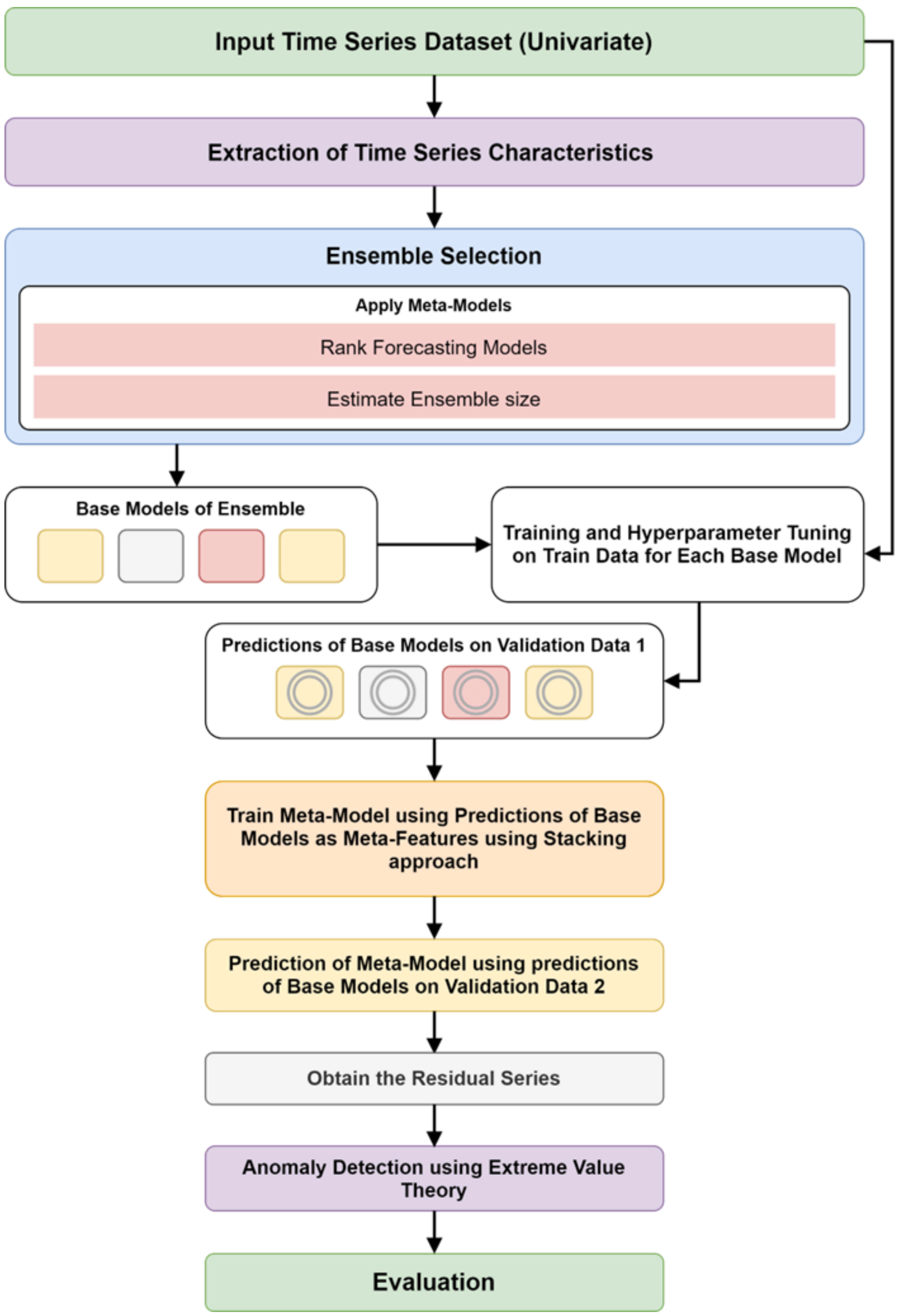

Figure 2

Figure 2 illustrates how our meta-framework operates on new time series data. When a new time series is input, the system first extracts its characteristics. Based on these extracted features, the framework uses the trained models to predict both the model rankings and the optimal ensemble size. Using these predictions, the system selects the best models for the ensemble and generates the forecasting output. Finally, it applies extreme value theory (EVT) to the forecasted values for anomaly detection.