Time-series,
Spring, 2026

Time-series Classification (TSC)

Faculty of DS & AI
Spring semester, 2026

Trong-Nghia Nguyen

Business AI Lab

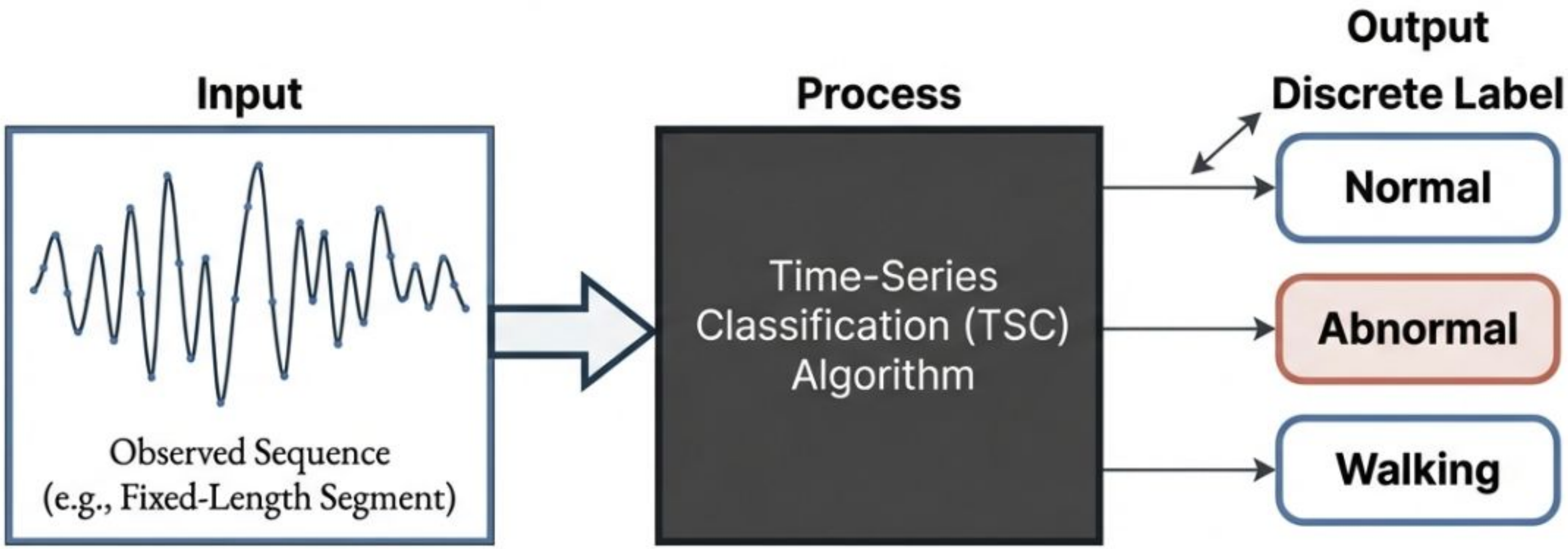
Content

- **Time-series Classification (TSC)**
- Application of TSC
- Standard Pipeline
- Advanced Techniques

Time-series Classification (TSC)

Assigning a Discrete Label to Temporal Data

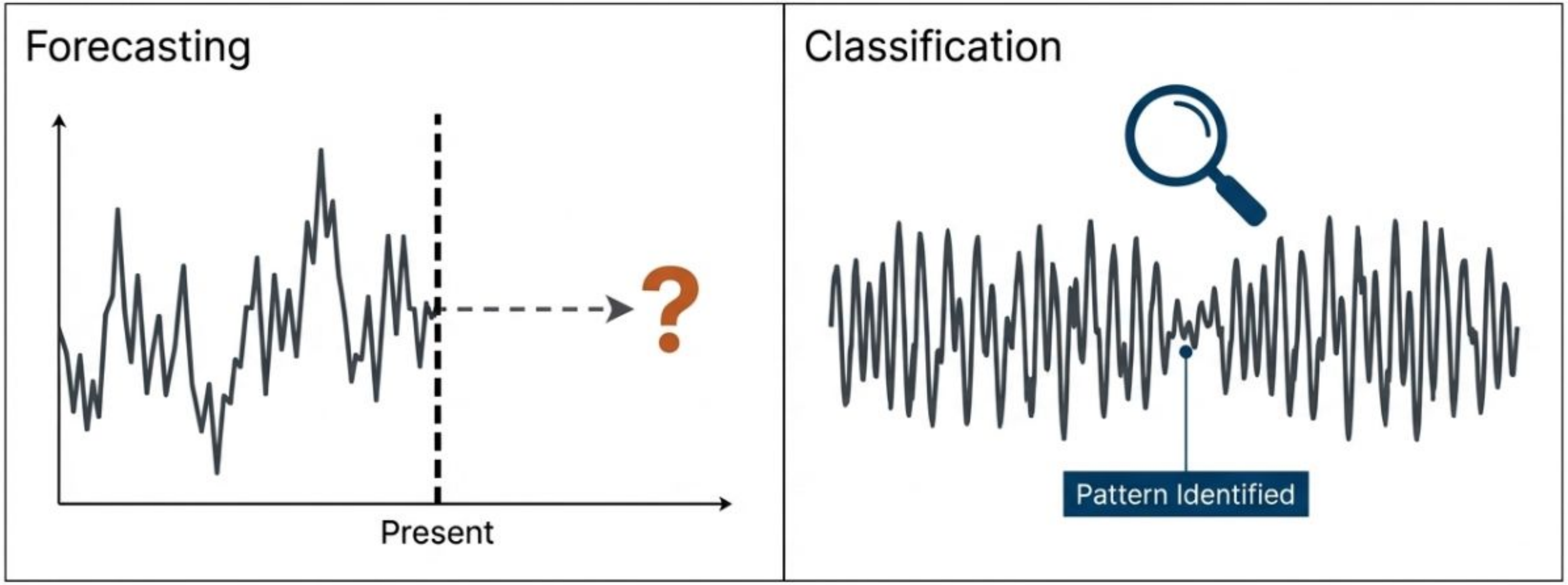
- Time-Series Classification (TSC) focuses on assigning a categorical label to an entire sequence or a specific segment of data.
- The system evaluates observed sequences (often fixed-length segments) to determine their class.



Time-series Classification (TSC)

Discovering "Meaning" Rather Than the "Next Value"

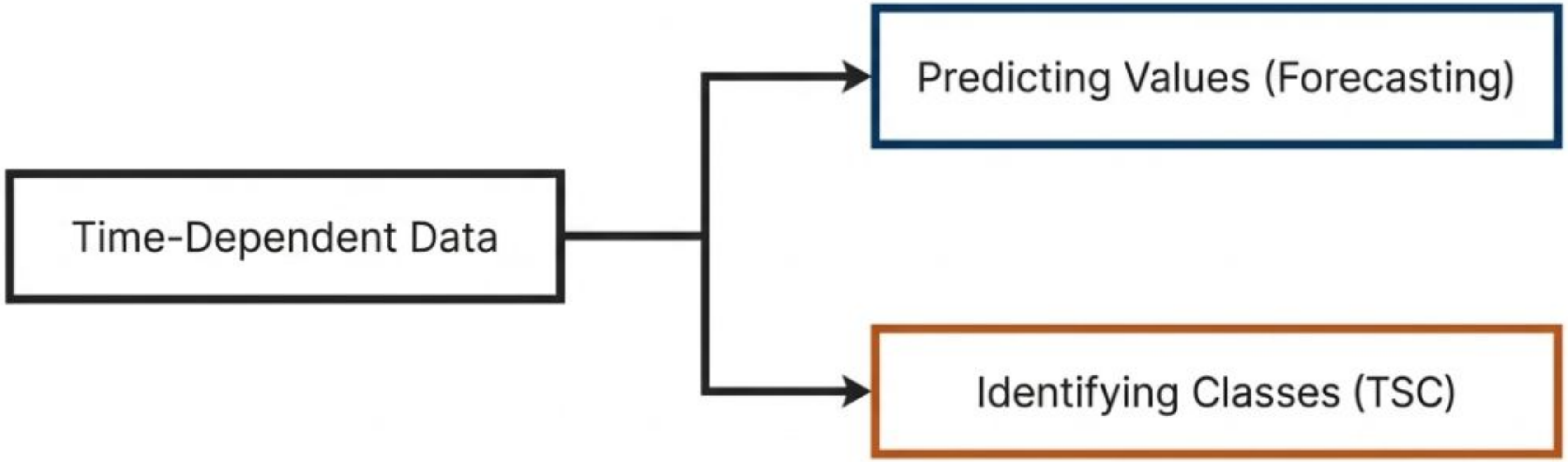
- Forecasting seeks the future.
- Classification seeks the nature and inherent pattern of the data.



Time-series Classification (TSC)

The Shared Foundation of Time-Dependent Data

A common misconception arises because both forecasting and classification operate on time-dependent data. However, they diverge significantly in their mathematical structure and evaluation objectives.



Time-series Classification (TSC)

Diverging Objectives and Outputs

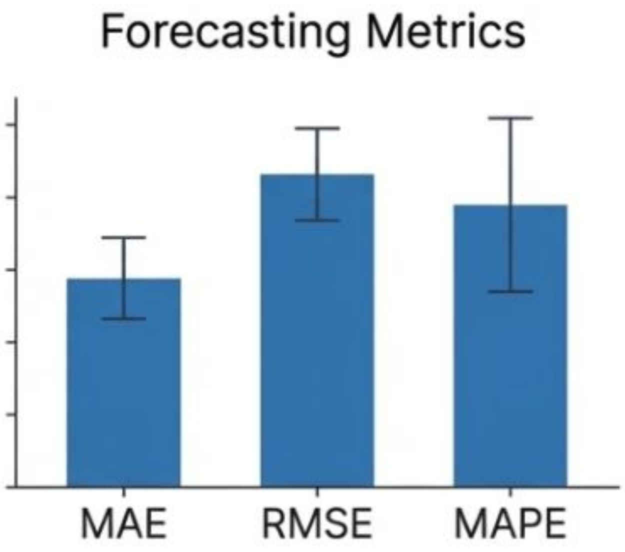
- Forecasting: Finds the conditional expected value at a future point. Yields a continuous value.
- Classification: Determines the probability that a sequence belongs to a specific class. Yields a categorical label.

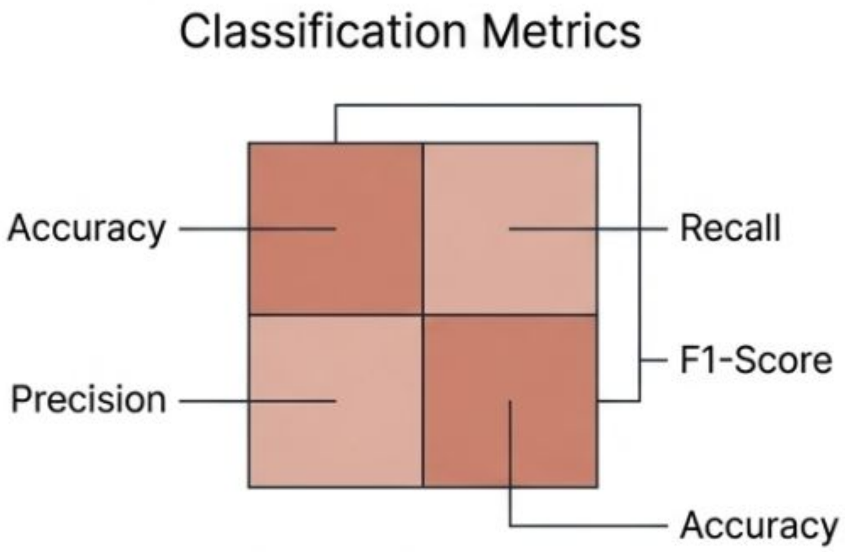
Forecasting	Classification
$\hat{X}_{n+h\mid n} = E[X_{n+h} \mid X_1, \ldots, X_n]$	$P(C = \text{“Class”} \mid X)$
Continuous Value	Categorical Label

Time-series Classification (TSC)

Distinct Evaluation Metrics for Distinct Goals

- **Forecasting:** Measures the distance between actual and predicted values using error metrics.
- **Classification:** Evaluates the ability to correctly identify classes using accuracy metrics.





Time-series Classification (TSC)

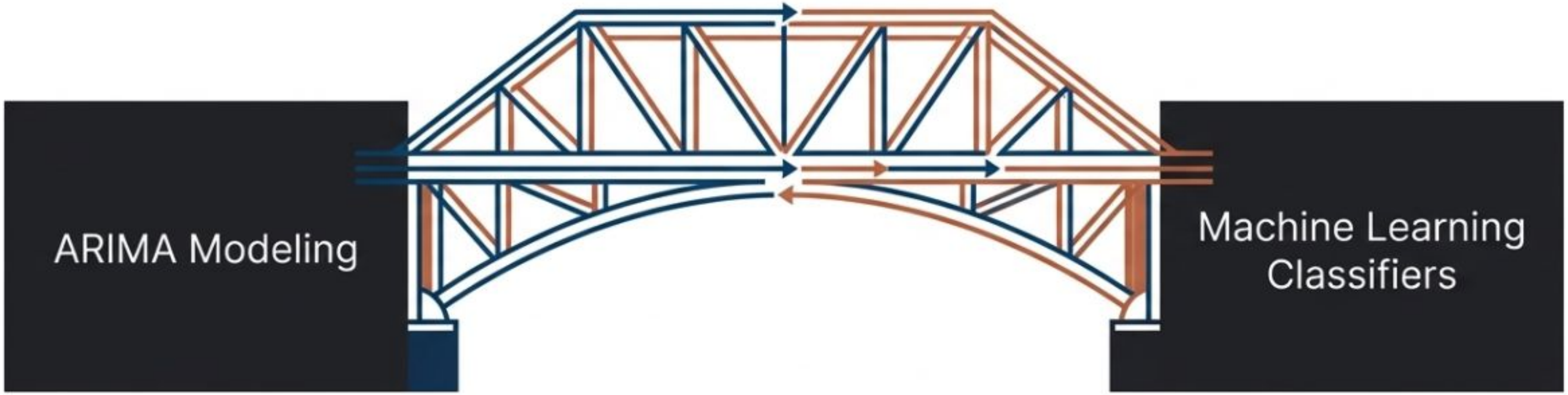
Synthesizing the Domains

Criteria	Forecasting/Regression	Classification (TSC)
Output	Continuous value (X_{n+h})	Categorical label (Class)
Objective	Predict the next step	Pattern Recognition
Key Metrics	MAE, RMSE, MAPE	Accuracy, F1-Score, Confusion Matrix

Time-series Classification (TSC)

Bridging the Gap: Model-Based Feature Extraction

Time-Series Classification can powerfully leverage existing forecasting models. By utilizing learned ARIMA parameters, we extract structured features to feed into classification algorithms.

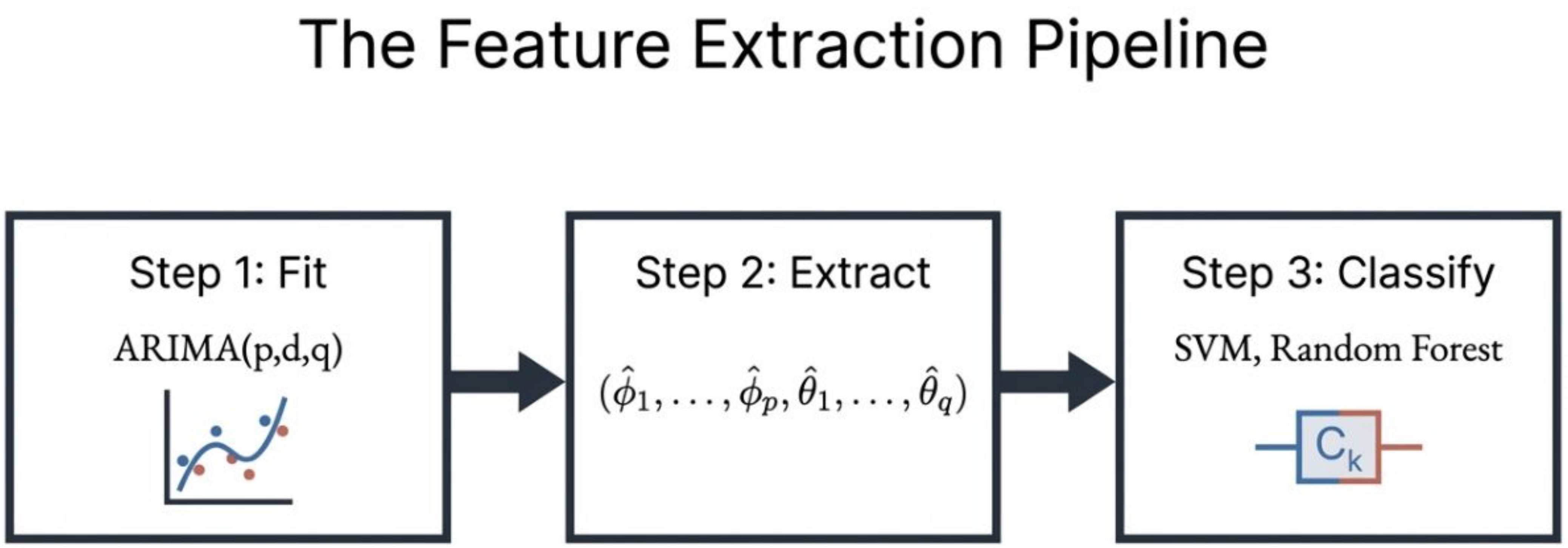


Time-series Classification (TSC)

ARIMA Coefficients as System “Fingerprints”

Auto-Regressive (\\ϕi) and Moving Average (*theta*) parameters represent the structural ‘memory’ and response to ‘shocks’ of the data, acting as unique identifiers for the series.

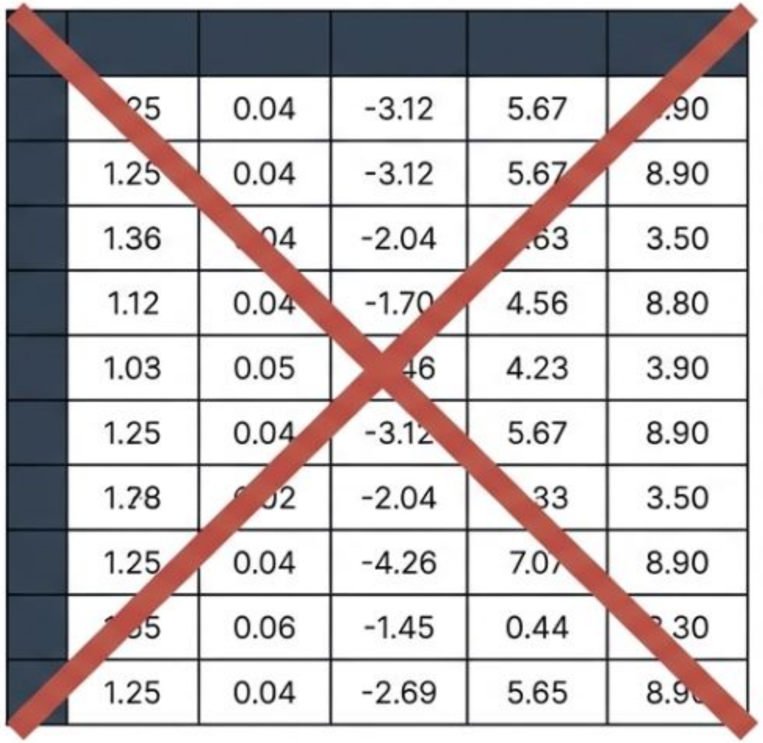
Time-series Classification (TSC)



Time-series Classification (TSC)

The Incompatibility of Standard Tabular Logic

Classifying time-series data presents unique structural hurdles. Temporal data breaks the core assumptions relied upon by standard machine learning algorithms built for static, tabular data.

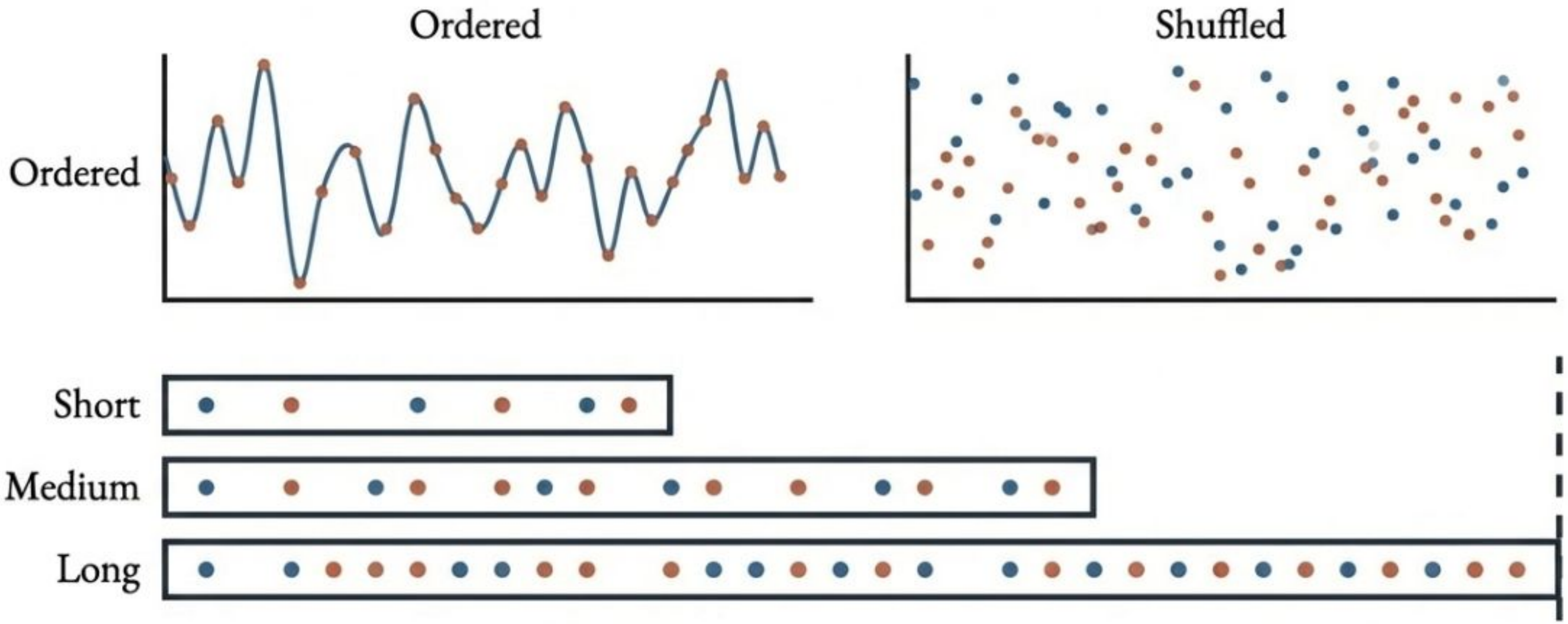


[illegible]	0.04	-3.12	5.67	[illegible]
1.25	0.04	-3.12	5.67	8.90
1.36	[illegible]	-2.04	[illegible]	3.50
1.12	0.04	-1.70	4.56	8.80
1.03	0.05	[illegible]	4.23	3.90
1.25	0.04	-3.12	5.67	8.90
1.78	[illegible]	-2.04	[illegible]	3.50
1.25	0.04	-4.26	[illegible]	8.90
[illegible]	0.06	-1.45	0.44	[illegible]
1.25	0.04	-2.69	5.65	[illegible]

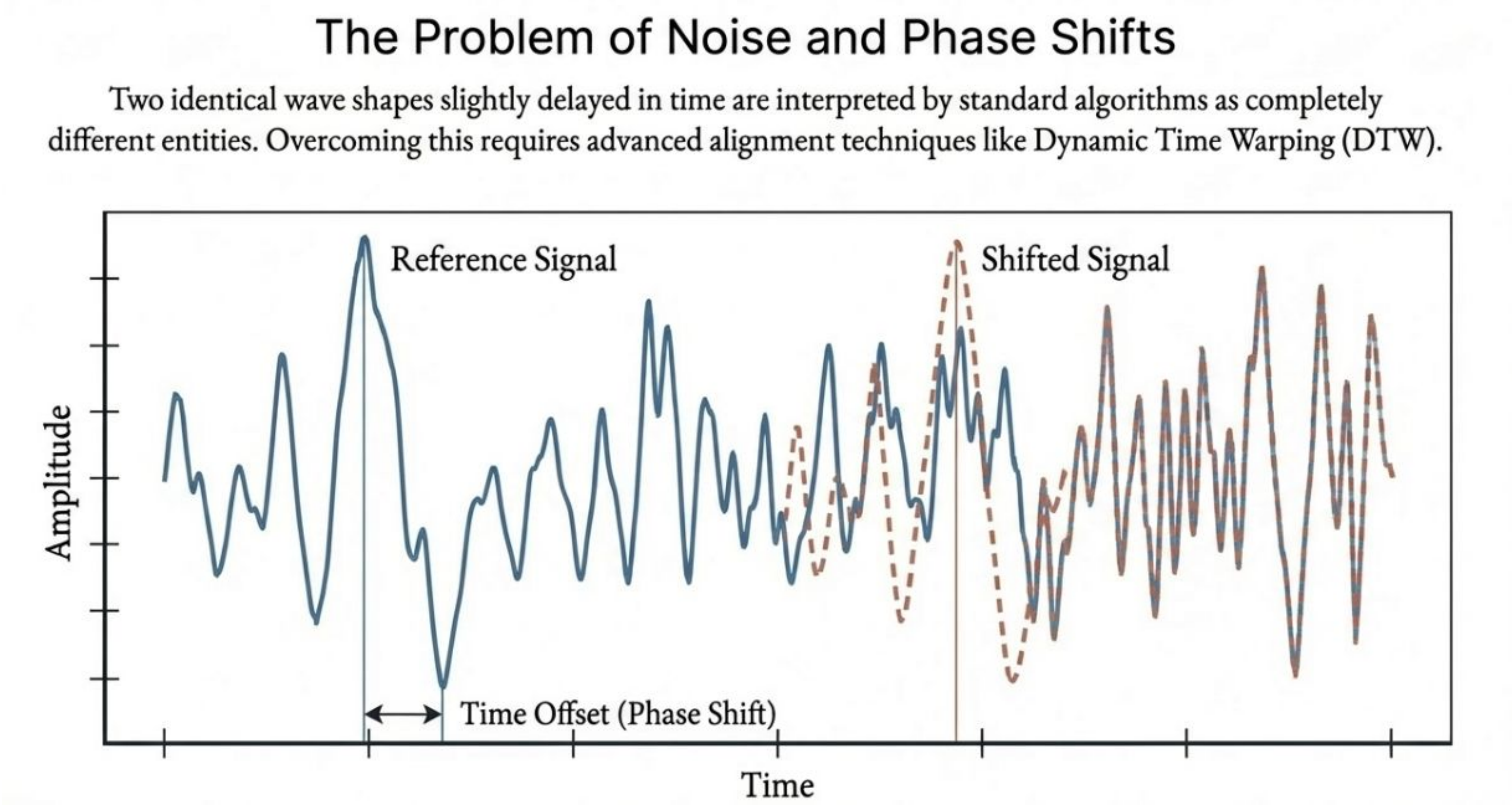
Time-series Classification (TSC)

Challenges of Sequence and Structure

- Temporal Dependence: Order holds the meaning; shuffling observations destroys the data structure.
- Variable Length: Real-world sequences differ in duration, breaking algorithms that require fixed-length input vectors.



Time-series Classification (TSC)



Time-series Classification (TSC)

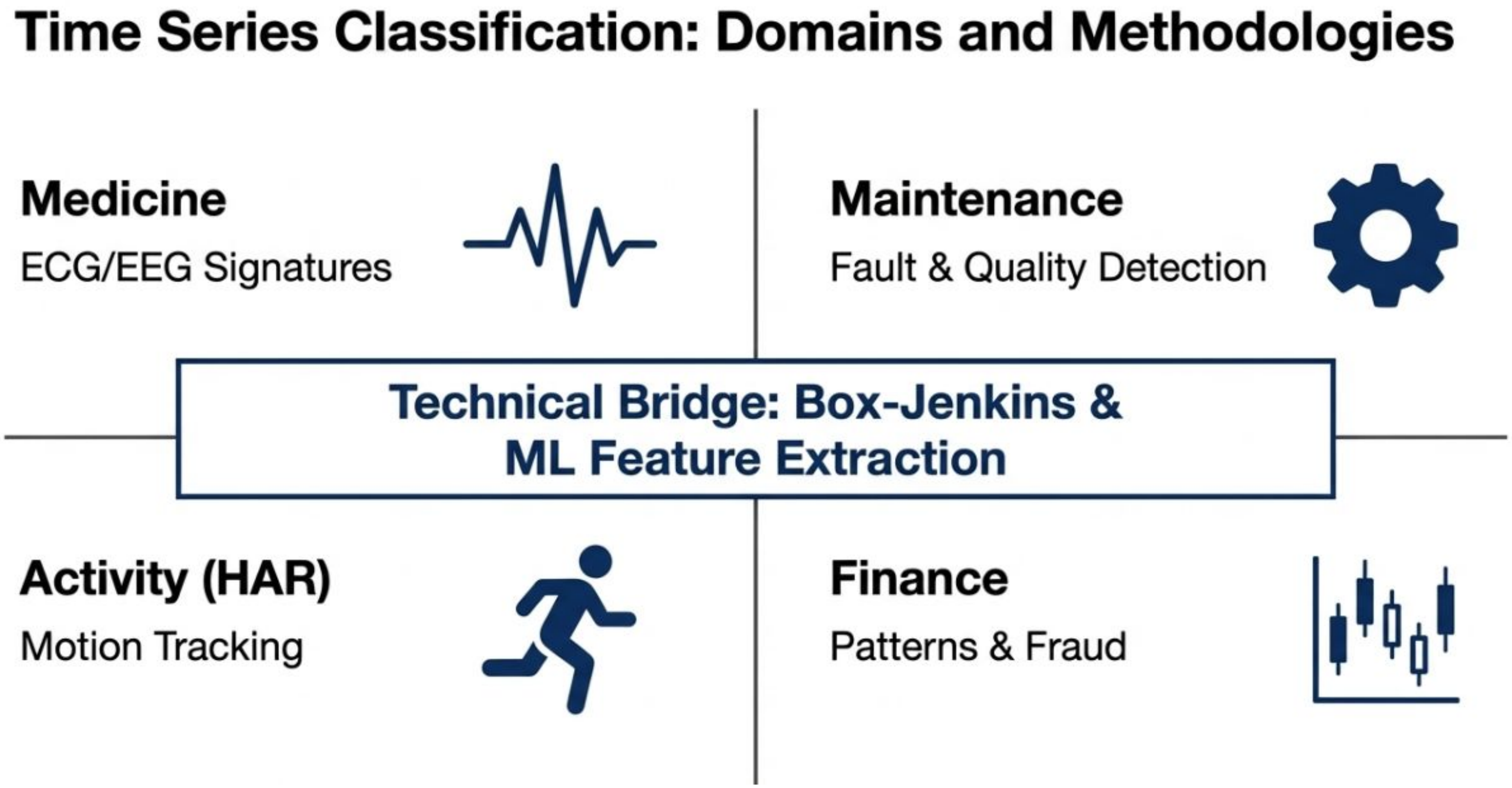
Core Concepts in Time-Series Classification

- The Paradigm: TSC assigns discrete, categorical labels to temporal sequences, prioritizing pattern recognition over predicting future values.
- The Distinction: Unlike forecasting (evaluated by error metrics like RMSE), TSC acts structurally like standard classification, relying on Accuracy and F1-Scores.
- The Complexity: Temporal dependence, variable lengths, and phase shifts require specialized logic—either structural extraction (via ARIMA) or dynamic alignment (DTW).

Content

- Time-series Classification (TSC)
- **Application of TSC**
- Standard Pipeline
- Advanced Techniques

Application of TSC

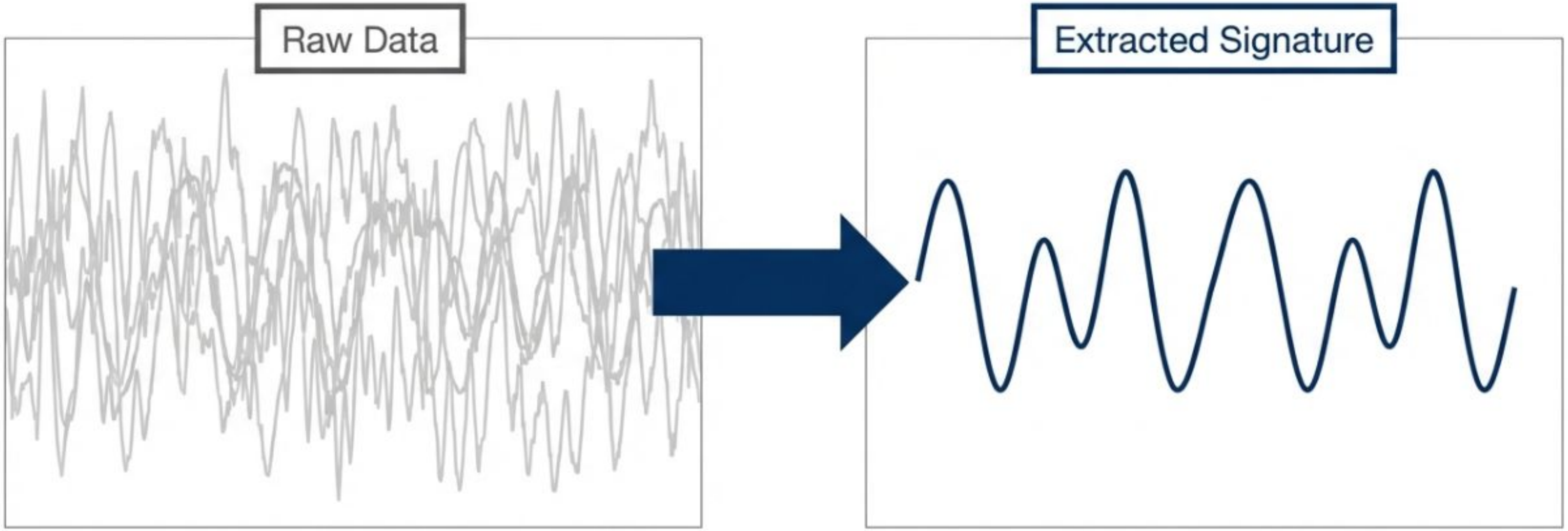


Application of TSC

Time series data contains unique behavioral signatures

The core task of Time Series Classification (TSC) is to identify embedded signatures to draw categorical conclusions.

Unlike forecasting, the goal is state recognition.



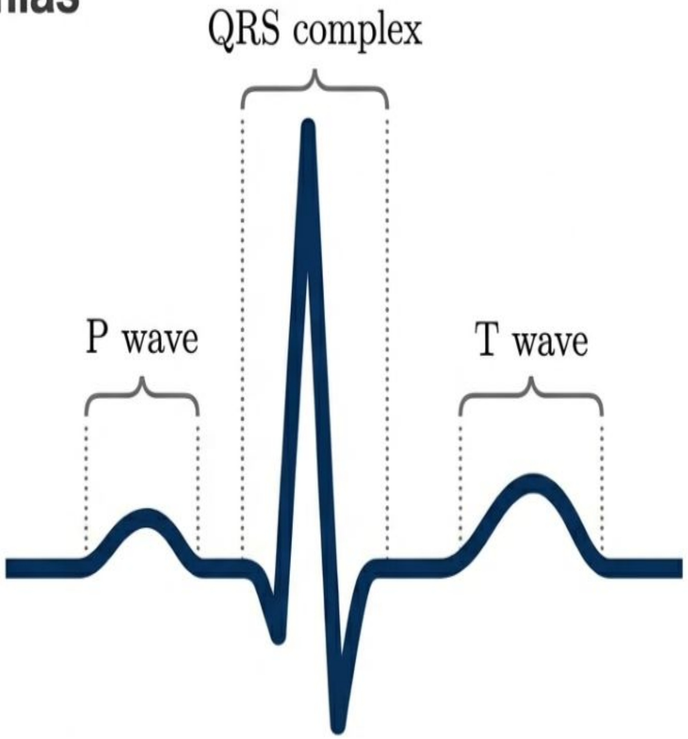
Application of TSC

Cardiology relies on waveform shapes to detect arrhythmias

Data: Electrical pulse sequences measured via body electrodes (ECG).

Problem: Classifying Normal Sinus Rhythm versus Arrhythmias (e.g., atrial fibrillation, extrasystole).

Role of Time: The explicit shape and distance between P, QRS, and T waves determines the classification label.

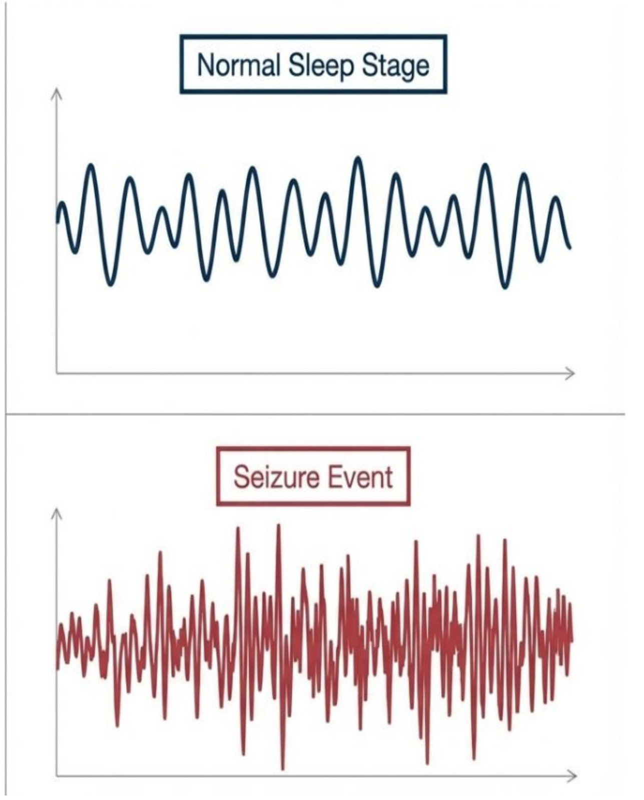


Neurology analyzes brainwave frequencies for state classification

Data: Brain waves recording neuronal electrical activity (EEG).

Problem: Seizure detection and sleep stage classification.

Integration: Combined with Survival Analysis to classify patient risk levels based on vital sign progression over time.



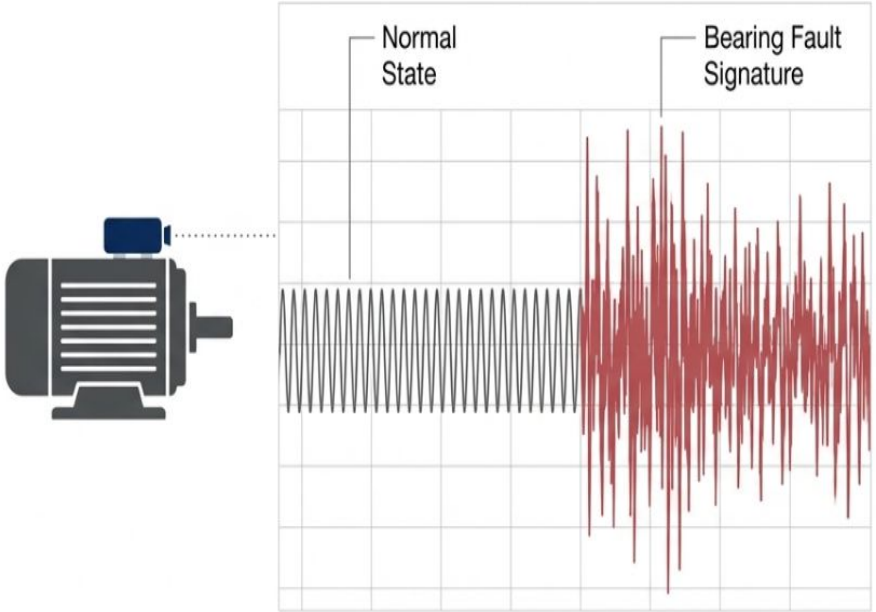
Application of TSC

Machinery transitions from reactive repair to proactive monitoring

Data: Sequential readings from accelerometer and vibration sensors mounted on motors or bearings.

Problem: Classifying operational states into specific categories: Normal, Misaligned, Mechanical Looseness, or Bearing Fault.

Role of Time: Recognizing degrading vibration frequencies before catastrophic failure occurs.

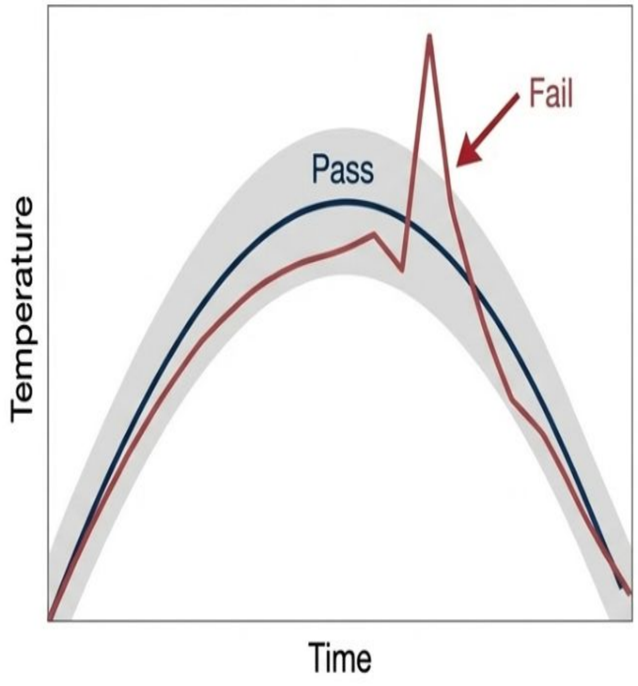


Production quality is classified dynamically during manufacturing cycles

Data: Pressure and temperature parameters tracked across a plastic injection or metal casting cycle.

Problem: Instantly assigning a Pass or Fail label to an individual product the moment the cycle finishes.

Role of Time: Ensuring thermodynamic curves stay within acceptable limits for the duration of the mold.



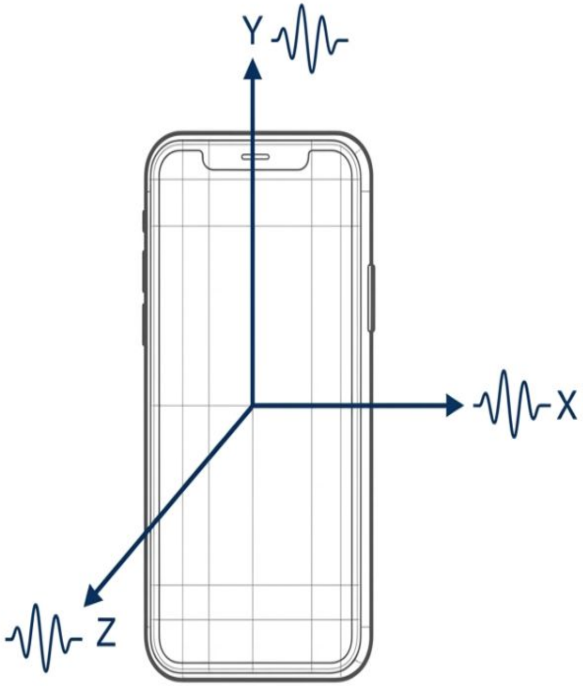
Application of TSC

Mobile sensors generate ubiquitous data for activity recognition

Domain: Human Activity Recognition (HAR) in smartphones and wearable devices.

Data: Continuous sequential data from internal Accelerometers and Gyroscopes.

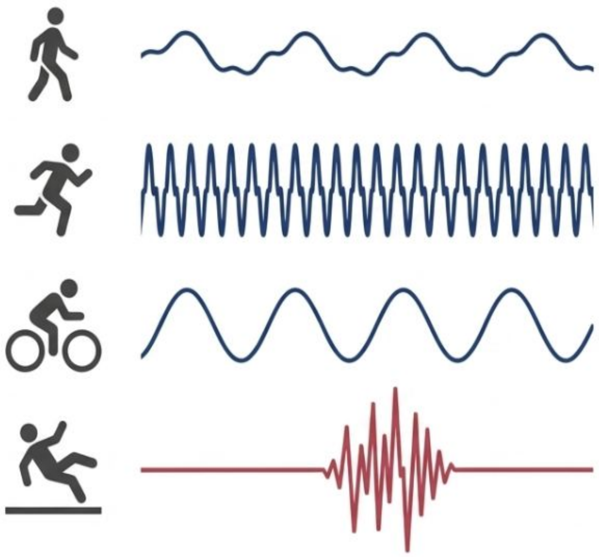
Role of Time: Tracking human motion mechanics in three-dimensional space over continuous intervals.



Activity classification enables continuous health and fitness tracking

Problem:
Classifying the user's current physical action (Walking, Running, Cycling, or Falling).

Practical Value:
Elderly care monitoring (immediate fall detection) and sports tracking (precise calorie expenditure).



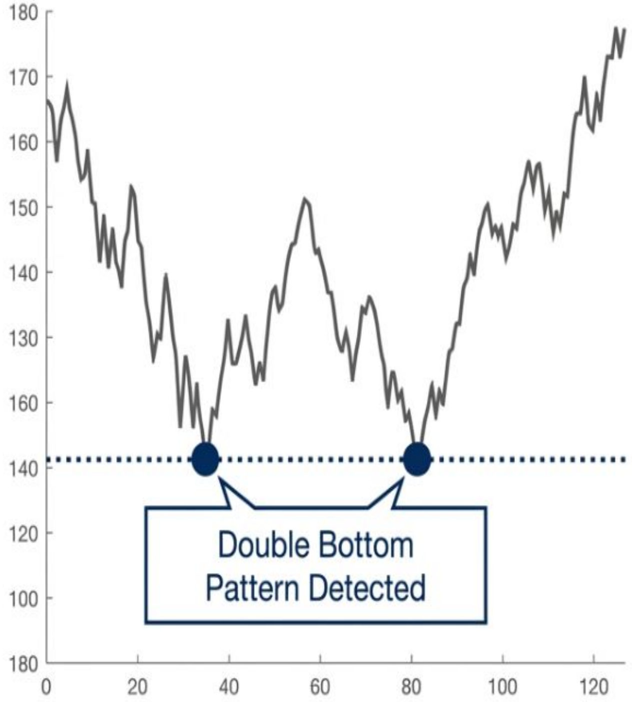
Application of TSC

Financial markets utilize classification to identify strategic chart patterns

Data: Sequential closing prices within a defined time frame.

Problem: Classifying current graph segments into known technical patterns (Double Bottom, Head and Shoulders, Rising Wedge).

Role of Time: Defining the geometry of market sceme enisting the geometry of market psychology over historical periods.

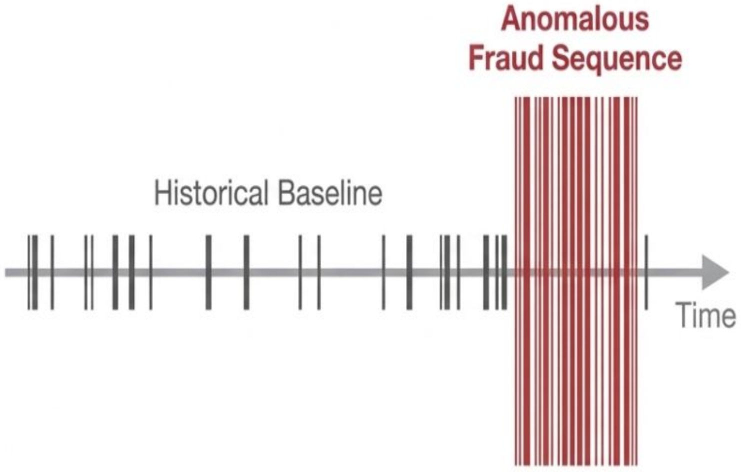


High-frequency transaction sequences reveal anomalous behaviors

Data: Short-term sequences of transactions tied to a specific account.

Problem: Classifying an action sequence as Valid or Fraud/Attack.

Role of Time: Detecting extreme deviations in transaction velocity and volume compared to historical baseline profiles.



Application of TSC

The Box-Jenkins process transforms raw sequences into structured features

Instead of feeding raw sequences directly into classifiers, traditional forecasting tools extract foundational features.

Step 1: Stationarity: Utilizing the ADF test to determine the order of differencing (d).
Step 2: Parameter Estimation: Applying Maximum Likelihood Estimation (MLE) to find ϕ and \theta coefficients.

Raw Data → ADF Test / Differencing d → MLE Estimation → ϕ & θ Extraction

Application of TSC

ARIMA coefficients serve as critical inputs for Machine Learning classifiers

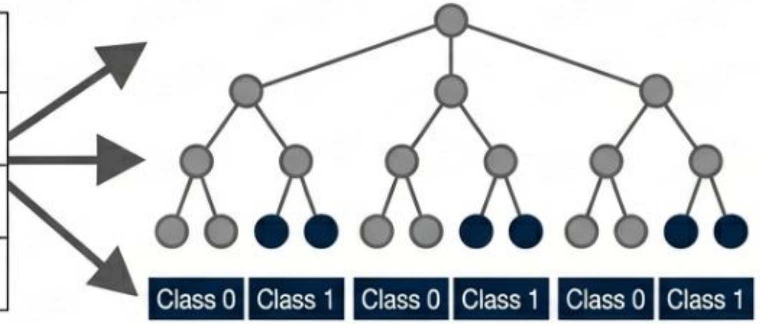
The Features:

- ϕ (Autoregressive): Reflects the memory of the series.
- θ (Moving Average): Reflects the impact of shocks.

The Models: These coefficients become the exact input vectors for ML classifiers (SVM, Random Forest) to assign the final domain labels.

ϕ & θ

ϕ_1	ϕ_2	θ_1	Label
0.45	0.21	-0.33	0
-0.12	0.67	0.15	1
0.08	-0.04	0.52	0

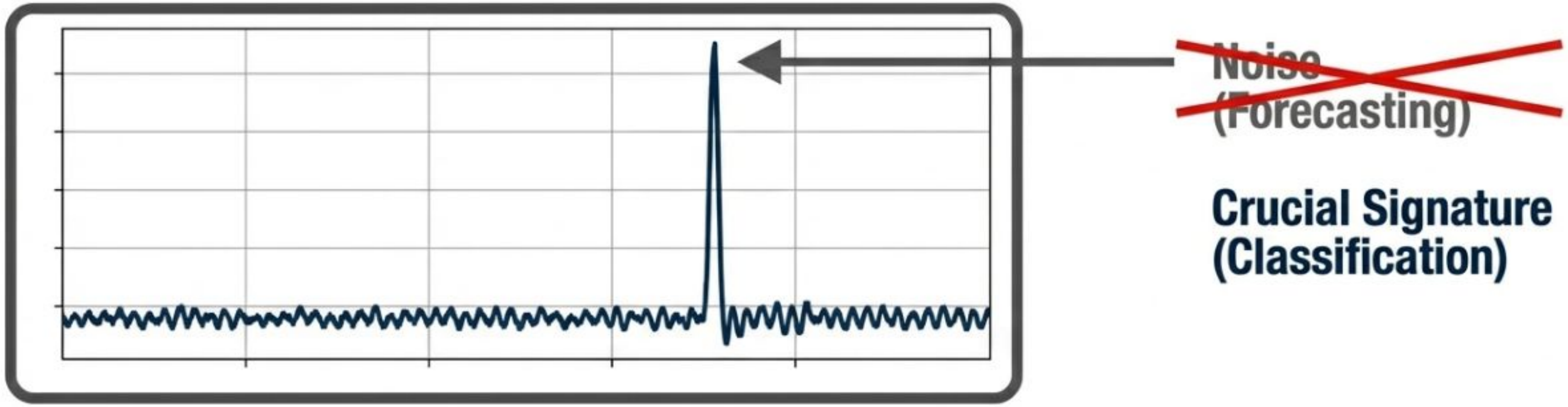


Application of TSC

In classification, there are no outliers if they form a pattern

A massive spike in vibration sensor data might be discarded as random noise by an ARIMA forecasting model.

In **Time Series Classification, that exact same shock is the most critical signature for identifying a state.**



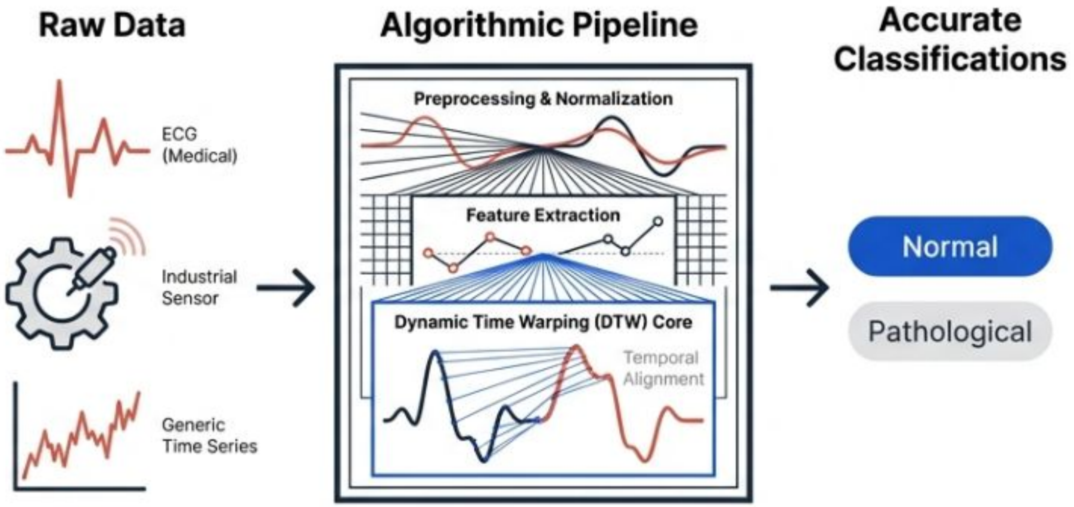
Content

- Time-series Classification (TSC)
- Application of TSC
- **Standard Pipeline**
- Advanced Techniques

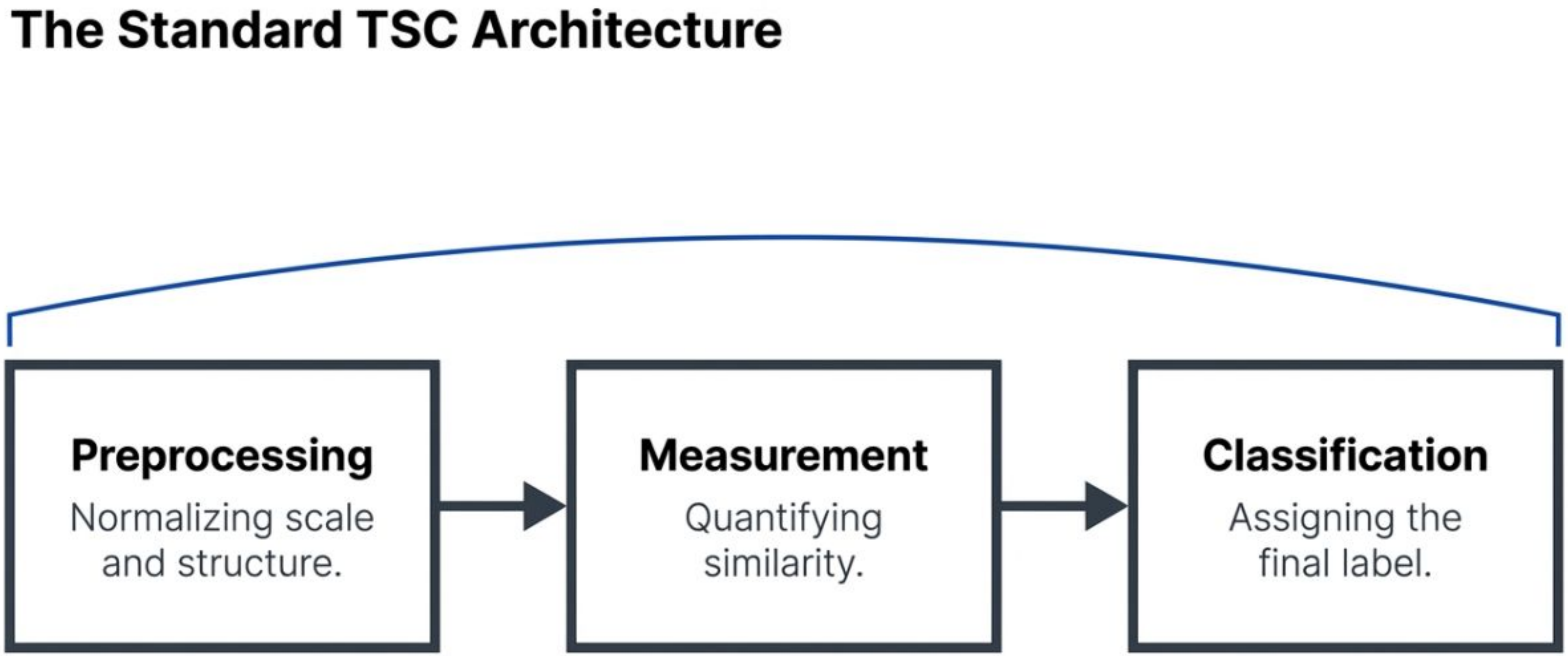
Standard Pipeline

Transforming Sensor Data into Actionable Labels

- Raw data from sensors and medical records is inherently noisy and subject to scale variations.
- Achieving accurate classification requires a strict, sequential pipeline.
- At the core of this pipeline lies the golden algorithm: Dynamic Time Warping (DTW), enabling machines to comprehend waveform similarity despite temporal distortions.



Standard Pipeline

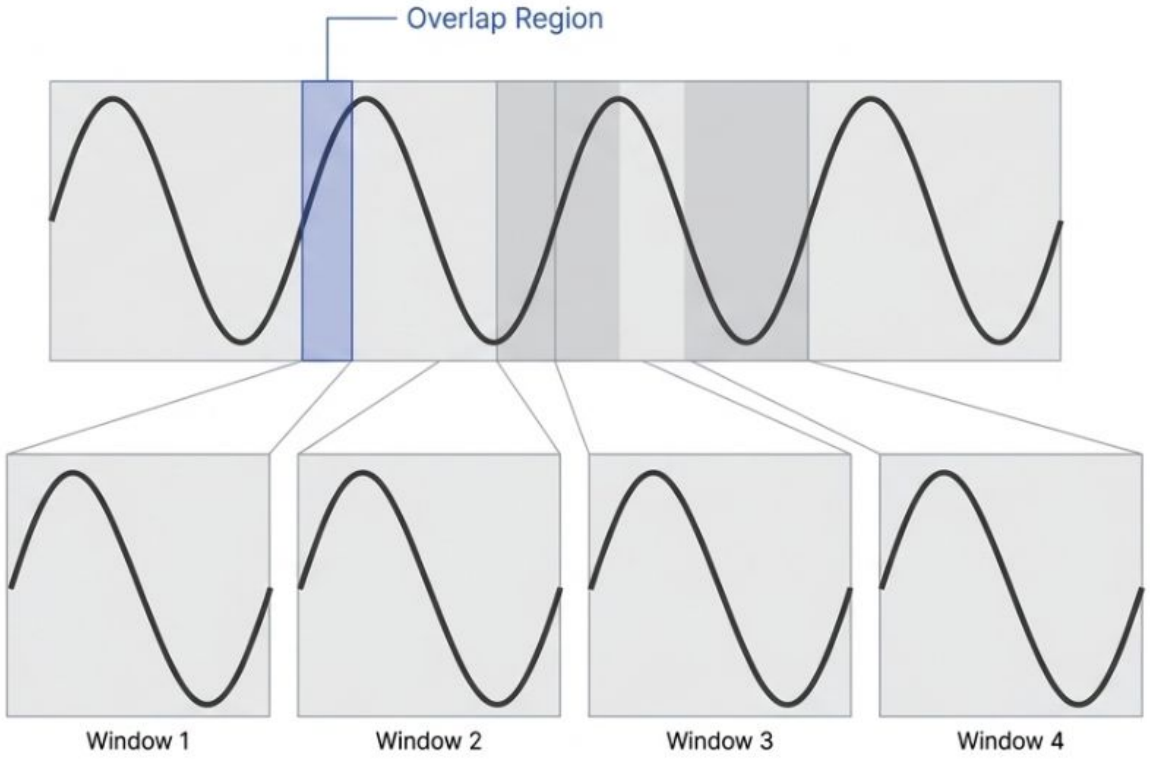


Standard Pipeline

Structuring Data via Windowing

Action:
Slicing continuous history into fixed-length segments (windows).

Purpose:
Unlike ARIMA forecasting which utilizes complete historical sequences, TSC requires every input sample to maintain an identical vector structure. Sliding windows with overlap prevent data loss at segment boundaries.



Standard Pipeline

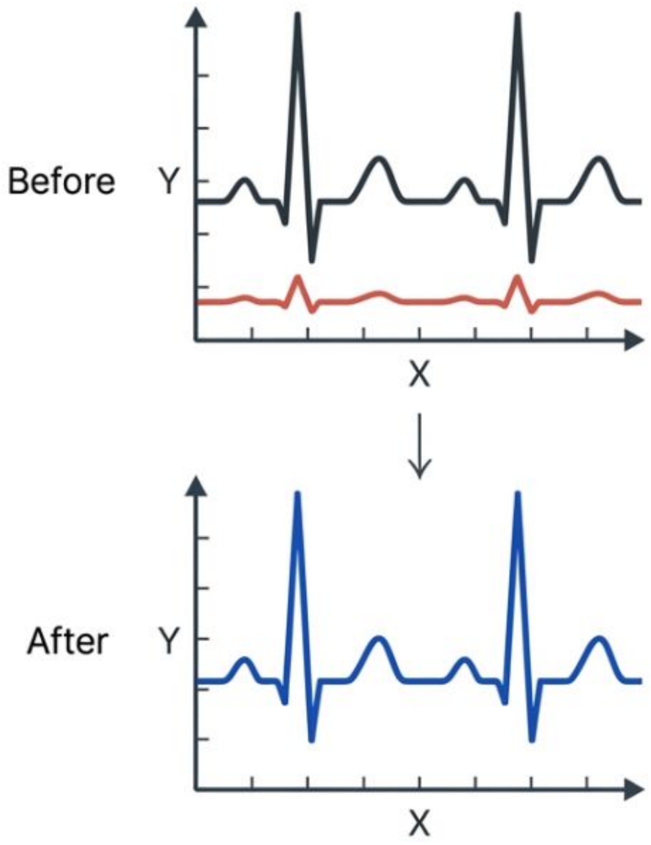
Isolating Shape with Z-Normalization

$$z = \frac{x - \mu}{\sigma}$$

\mu = mean, \sigma = standard deviation

Action: Standardizing the scale of every window. A mandatory TSC step.

Purpose: Models focus on shape, not amplitude. Identical events may differ in height due to sensor calibration. Normalization forces comparative scale.

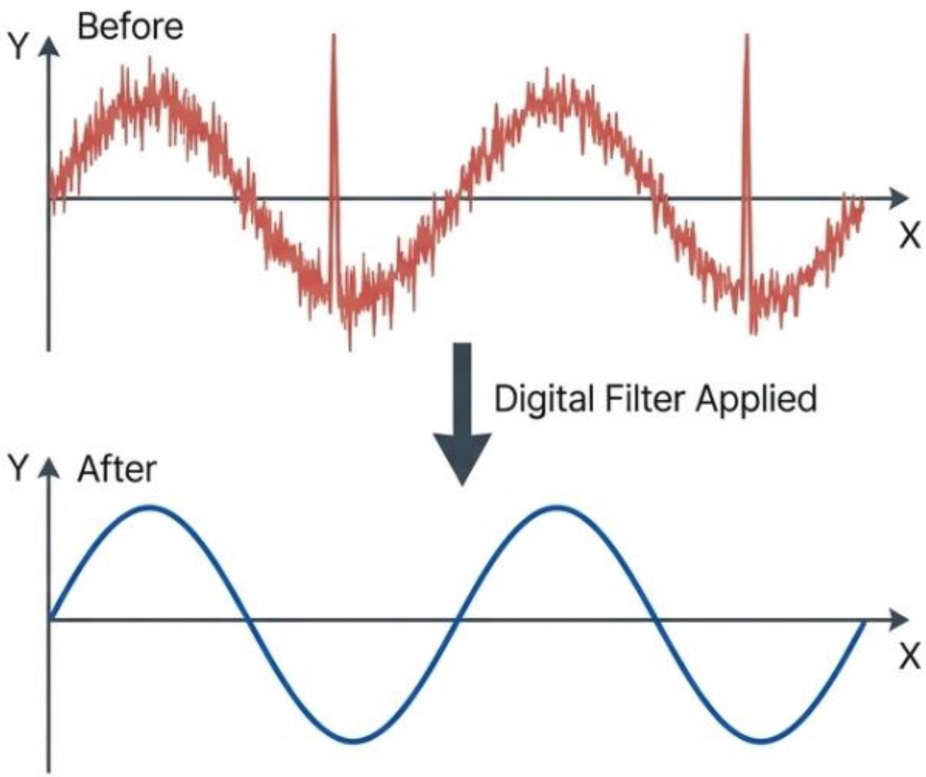


Standard Pipeline

Eliminating Artifacts through Filtering

Action: Applying digital filters like Butterworth or Median filters.

Purpose: To systematically remove unwanted high-frequency fluctuations and extreme outliers caused by temporary sensor errors. Ensures models learn actual structural patterns rather than random equipment noise.



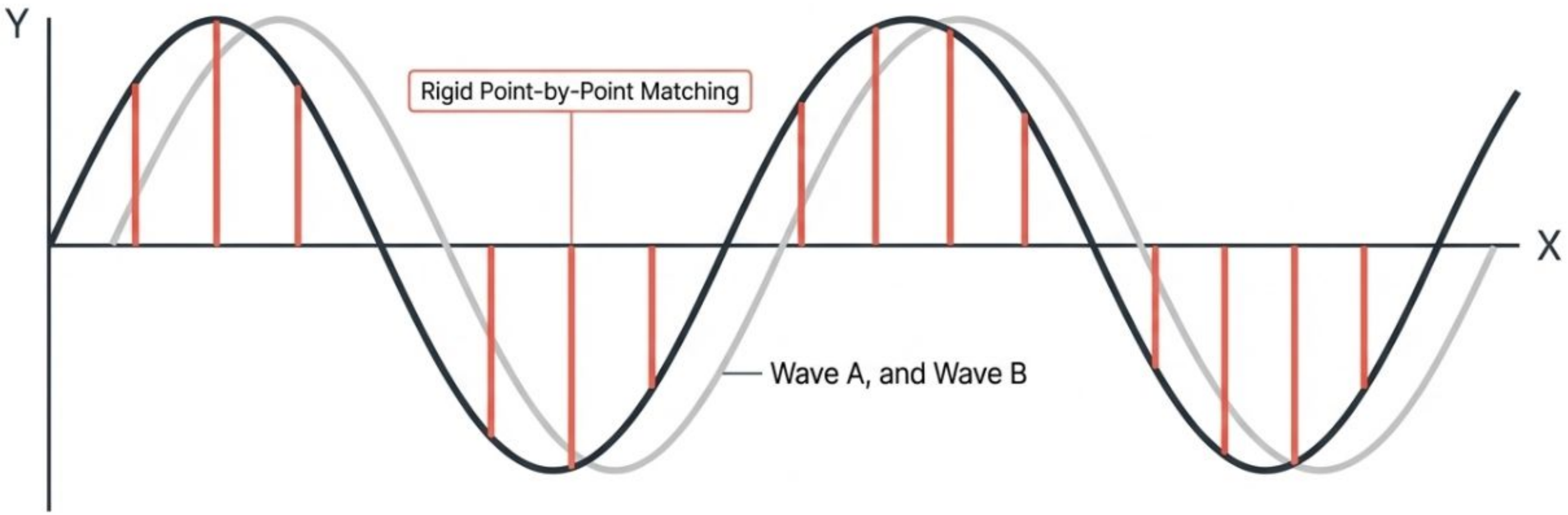
Standard Pipeline

The Phase-Shift Vulnerability of Euclidean Distance

Limitation: Euclidean distance forces rigid, point-by-point temporal comparison.

The Problem: If two sequences share an identical shape but experience a slight time delay (phase shift), Euclidean measurement reight registers them as vastly different.

Impact: Disastrous in applications—interpreting two identical cardiac events as completely distinct simply because one occurred a fraction of a second slower.



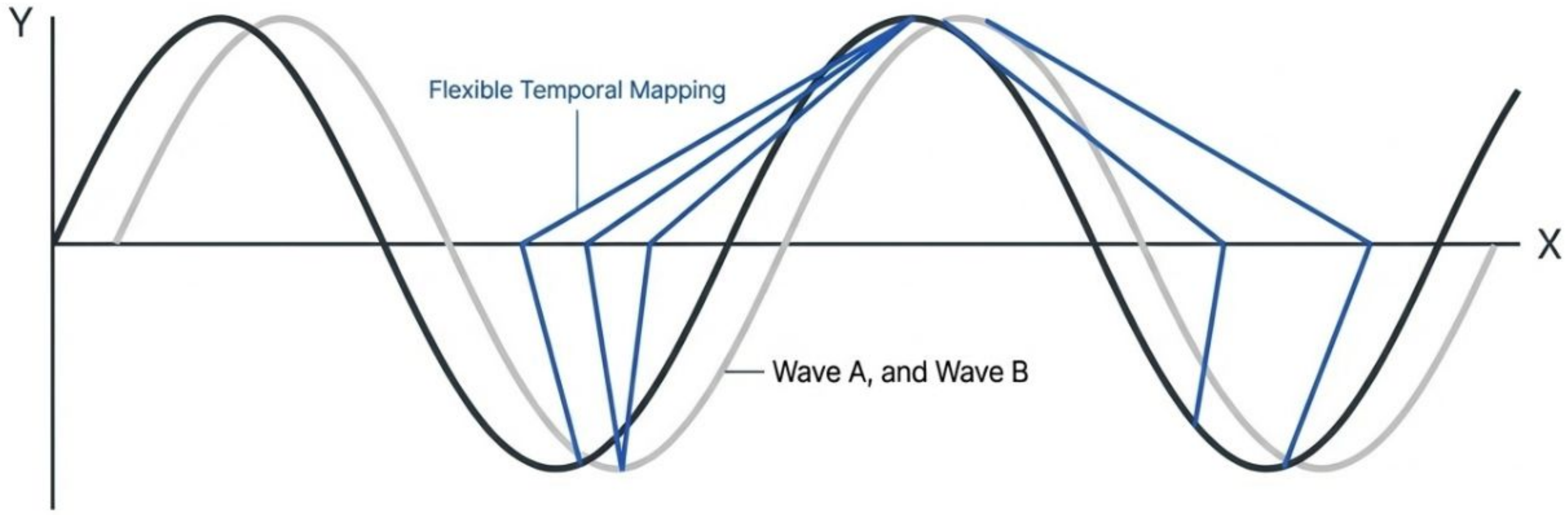
Standard Pipeline

Dynamic Time Warping (DTW)

Principle: Non-linear temporal stretching.

The Solution: DTW abandons rigid temporal alignment, allowing the time axis to stretch or compress.

Impact: Recognizes identical underlying patterns regardless of phase shifts or varying speeds. A single point on Sequence A can map to multiple points on Sequence B, and vice versa.

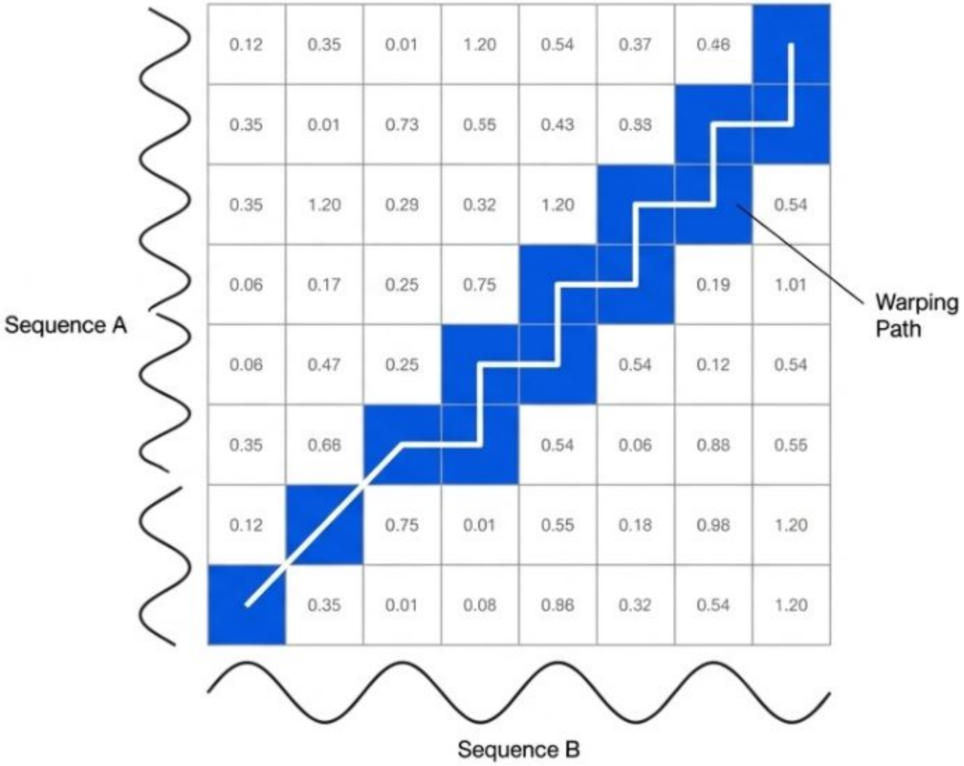


Standard Pipeline

The Mechanics of the Warping Path

1. **Cost Matrix**: The algorithm calculates the distance between every possible combination of points across two sequences in a comprehensive grid.

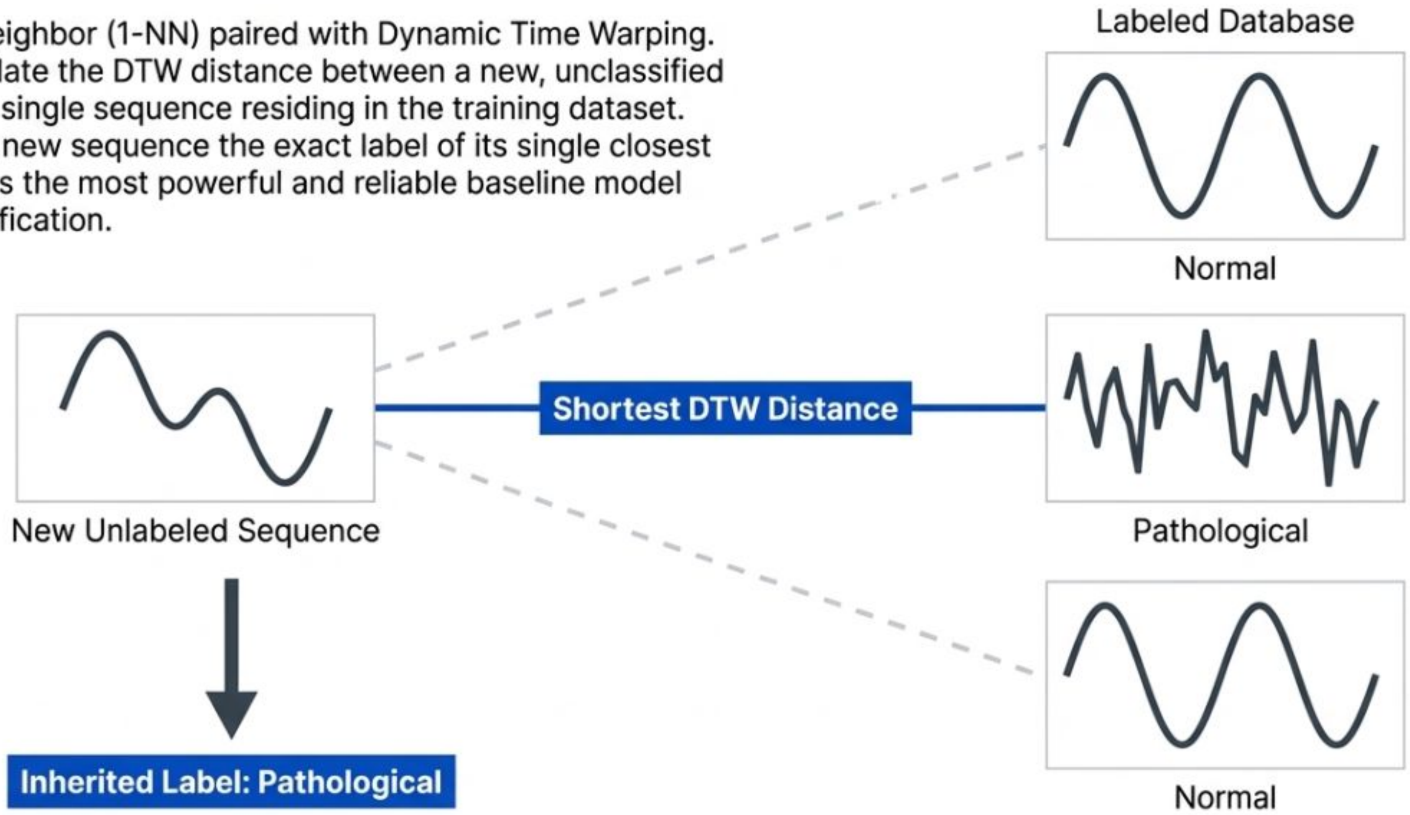
2. **Optimal Path:**
   It evaluates all alignments to discover the single Warping Path yielding the lowest total cumulative cost.
3. **Result:** The lowest cumulative cost serves as the final, highly accurate similarity score between the distinct waveforms.



Standard Pipeline

Distance-Based Classification: 1-NN DTW

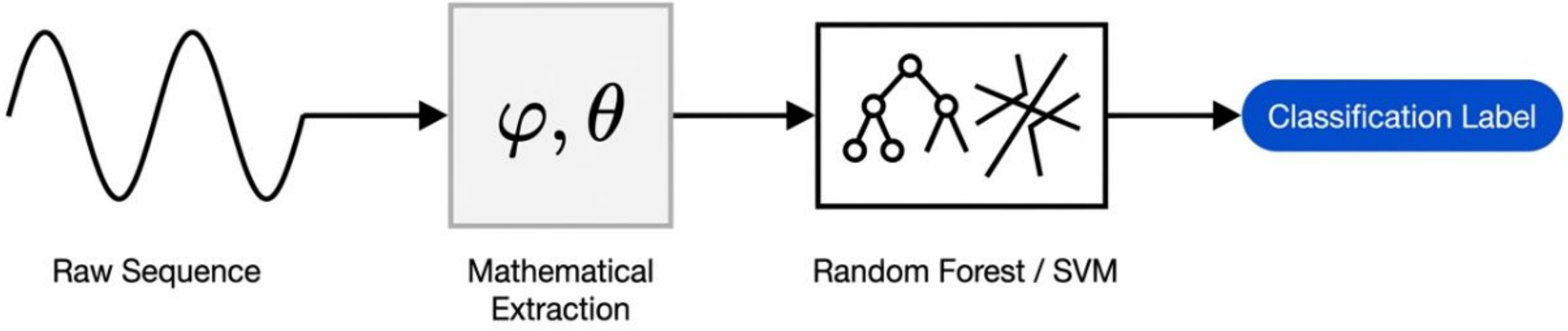
Action: 1-Nearest Neighbor (1-NN) paired with Dynamic Time Warping.
Methodology: Calculate the DTW distance between a new, unclassified sequence and every single sequence residing in the training dataset.
Purpose: Assign the new sequence the exact label of its single closest match. This serves as the most powerful and reliable baseline model in Time Series Classification.



Standard Pipeline

Feature-Based Classification: ARIMA Encodings

Action: Utilizing established forecasting models for feature extraction.
Methodology: Estimate ARIMA coefficients via Maximum Likelihood. Treat these coefficients as encoded summaries of the sequence's autocorrelation rules.
Purpose: Instead of feeding raw waveform data, input these summarized statistical features into robust machine learning algorithms.



Standard Pipeline

Pipeline Synthesis

Phase	Action	Goal
● Preprocessing	Windowing + Z-normalization	Eliminate noise and unify amplitude scales.
● Measurement	Apply Dynamic Time Warping (DTW)	Recognize shapes despite temporal phase shifts.
● Classification	1-NN DTW or ML on ARIMA coefficients	Assign final categorical label (e.g., Normal vs Pathological).

Content

- Time-series Classification (TSC)
- Application of TSC
- Standard Pipeline
- **Advanced Techniques**

Advanced Techniques

Real-World Applications Demand Advanced Methodologies

Basic TSC

Evaluates single, isolated wave segments. Insufficient for modern multidimensional environments.

Advanced TSC

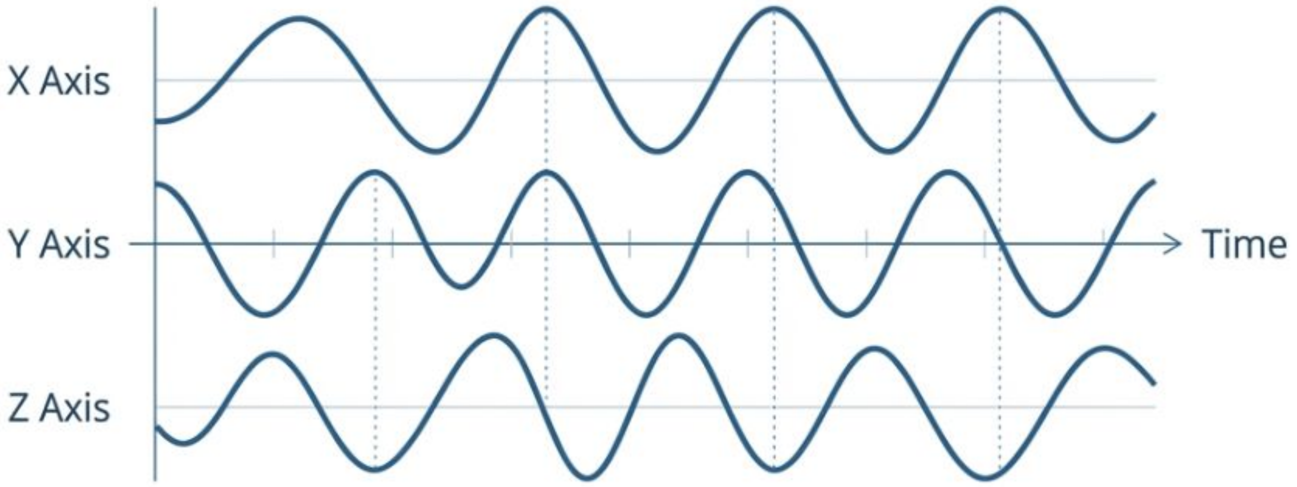
Requires handling complex architectures across four critical domains:

- Multisensor data integration
- Explainable diagnostic models
- Detection of rare anomalies
- Real-time analytical reactions

Advanced Techniques

Multivariate Data Introduces Dimensional Complexity

Modern devices capture multiple channels simultaneously (e.g., 3-axis accelerometers, patient vitals monitoring).



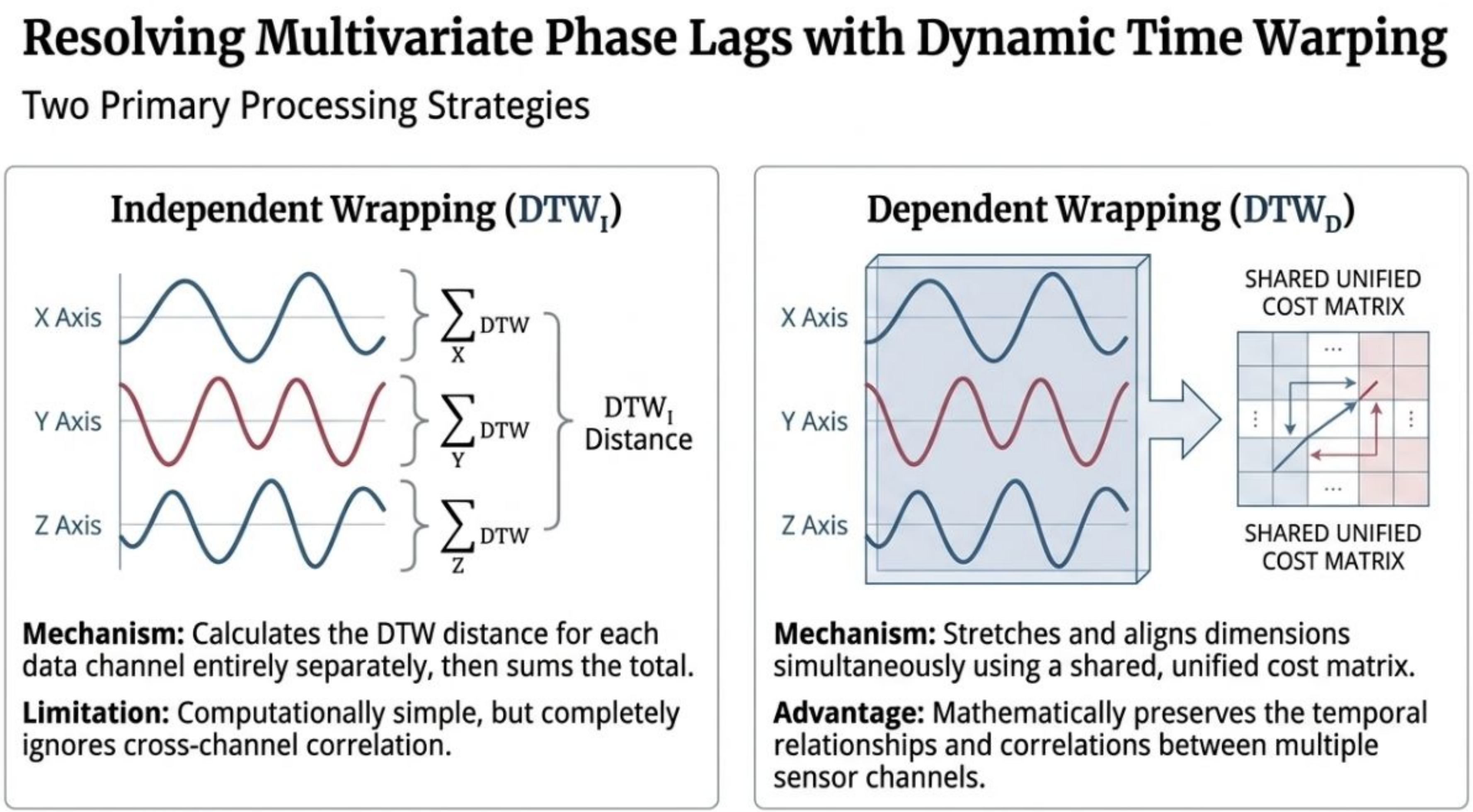
Challenge 1: Dimensional Correlation

Variables do not operate independently. A shift in one dimension actively triggers shifts in others.

Challenge 2: Phase Lag

Correlated signals from different sensors may not align perfectly in time, despite originating from the exact same event.

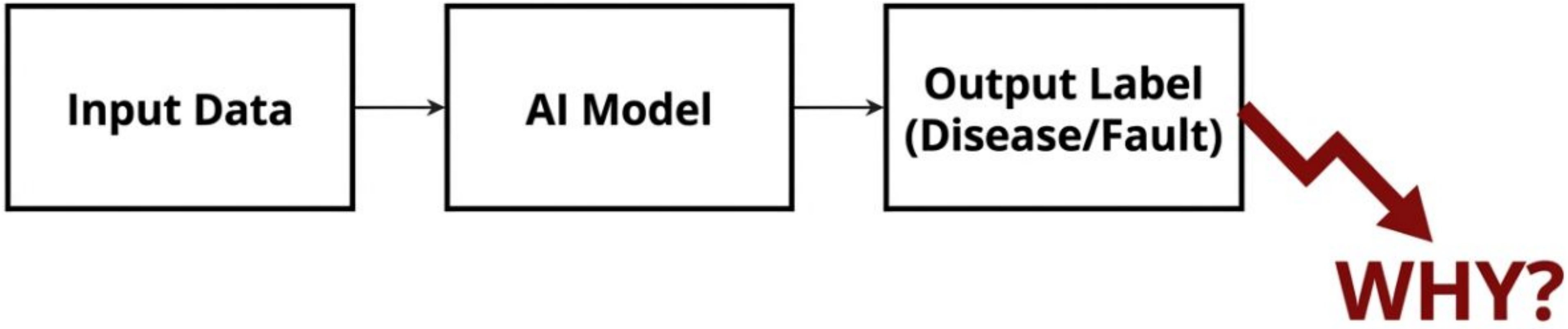
Advanced Techniques



Advanced Techniques

The Necessity of Interpretability Beyond the Black Box

In critical deployments (AI disease diagnosis, industrial machinery monitoring), an output label of ‘Abnormal’ or ‘Pathological’ is fundamentally insufficient.



The Requirement: Doctors and industrial engineers require concrete, interpretable evidence. The model must explain why a classification was made to be trusted and utilized in the real world.

Advanced Techniques

Shapelets Isolate Deterministic Temporal Evidence

Definition:

Shapelets are the most mathematically characteristic subsequences that successfully distinguish between data classes.

Interpretive Value:

When a model classifies a sequence as a 'Fault', Shapelets allow it to pinpoint the exact temporal segment containing the specific 'fault signature'.

Application:

Enables industrial engineers to identify the precise millisecond machinery began to fail.

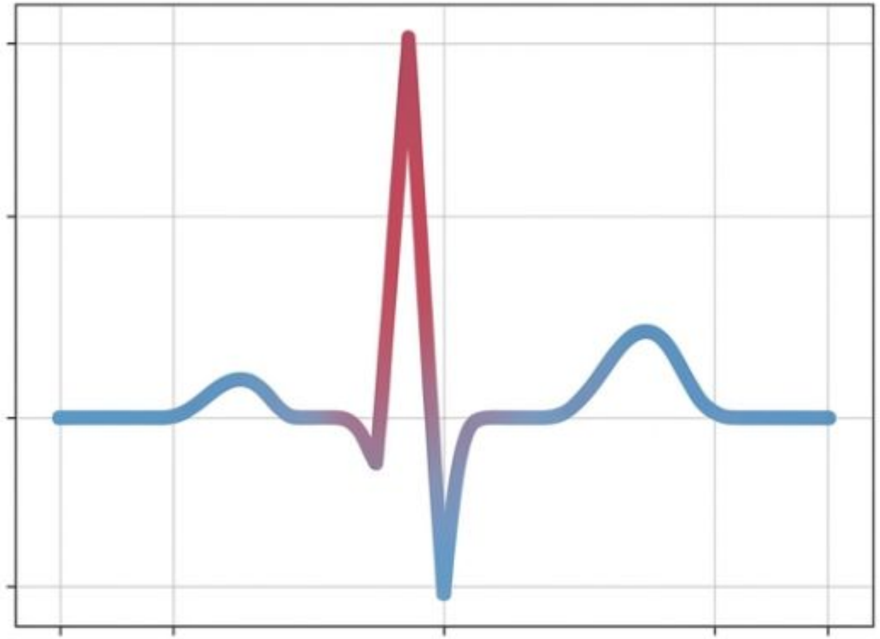
Advanced Techniques

Class Activation Mapping (CAM) Visualizes Deep Learning Decisions

Applied specifically to Deep Learning architectures like 1D CNNs.

Mechanism:
Generates and overlays probability heatmaps directly onto the wave segments that exerted the highest influence on the final class probability calculation.

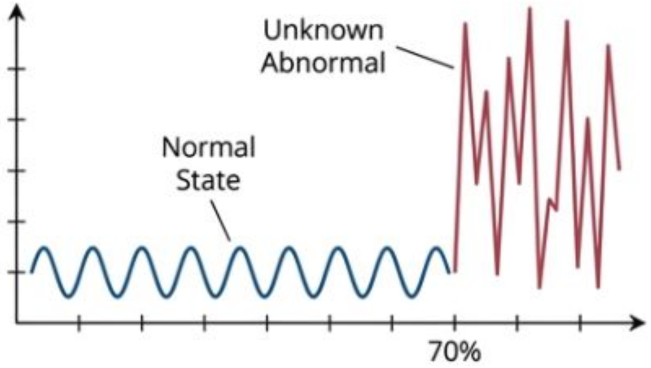
Clinical Application:
In an electrocardiogram (ECG), CAM visually highlights the QRS complex or ST segment, providing direct visual proof for a 'Myocardial Infarction' diagnosis.



Advanced Techniques

Addressing Rare Faults via Anomaly Detection

The Data Deficit: Standard classifiers require massive datasets of both normal and faulty data. Real-world catastrophic failures are rare, meaning “Fault” datasets are often non-existent.



The Solution: Instead of training a model to recognize specific faults, we rigorously model the definition of “Normal” behavior.

The ARIMA Synergy: AutoRegressive Integrated Moving Average (ARIMA) models provide the ideal mathematical foundation for defining this baseline expectation.

Advanced Techniques

Residual-Based Classification Mechanics

Step 1: Baseline Modeling

Fit a standard ARIMA model strictly on 'Normal' historical data.

Step 2: Residual Analysis

Feed new data into the model and analyze the resulting residuals (the mathematical difference between expected vs. actual data).

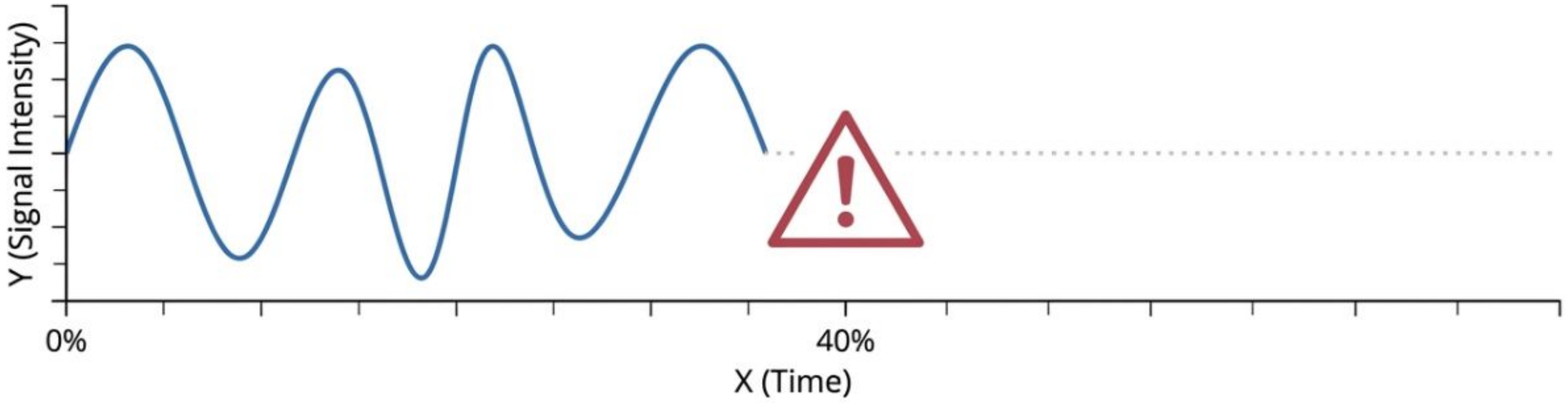
Detection Criteria

A new sample is definitively classified as 'Abnormal' if either condition is met:

- Residuals strictly exceed the statistical confidence band of $\pm 1.96/\sqrt{n}$.
- The Ljung-Box test yields $p < 0.05$ (proving the residuals reject the null hypothesis and are no longer mere white noise).

Advanced Techniques

Early Classification in Real-Time Environments



The Constraint:

Waiting for a sequence to finish collecting data is often too late to prevent a catastrophe.

The Objective:

Achieve high-accuracy classification using only 30% to 50% of the total temporal sequence length.

Clinical Application:

Issuing a definitive stroke warning the moment abnormal brain signals are detected, enabling life-saving intervention rather than waiting for the biological event to conclude.

Advanced Techniques

Synthesis Matrix: Advanced TSC Methodologies

Technique	Core Application	Primary Benefit
Multivariate DTW	Multi-axis sensors, IoT	Identifies and preserves dimensional correlation across data channels.
Shapelets	Industrial fault diagnosis	Isolates direct visual evidence of faulty temporal subsequences.
ARIMA Residuals	Intrusion & Anomaly Detection	Classifies purely on mathematical deviation from baseline models.
Early TSC	Medical emergencies, Cybersecurity	Rapidly intercepts events via preemptive real-time classification.

Thank you