# Chapter 2

# Background and Related Works

## 2.1 Background

Nowadays, time series data is generated in every field, such as finance, healthcare, agriculture, network security, etc. In the field of data mining and machine learning, various problems are studied based on time series data, including forecasting, indexing, clustering, and classification. In this thesis work, we focus on the problem of early classification on time series that is the special case of the conventional TSC approach [7]. This chapter discusses the fundamentals of time series, including an overview of representation techniques, distance/similarity measures, classification approaches, and early classification, followed by a detailed literature review of early classification methods.

### 2.1.1 Time series representation

Typically, time series data is high dimensional because each data point in the time series is considered as one dimension or as a feature in a feature vector. High dimensionality introduced the complexity in learning the model in the area of data mining [26]. Therefore, time series data may not be appropriate in its raw form to perform tasks such as indexing, process, query, store, etc. In this context, various representation meth-

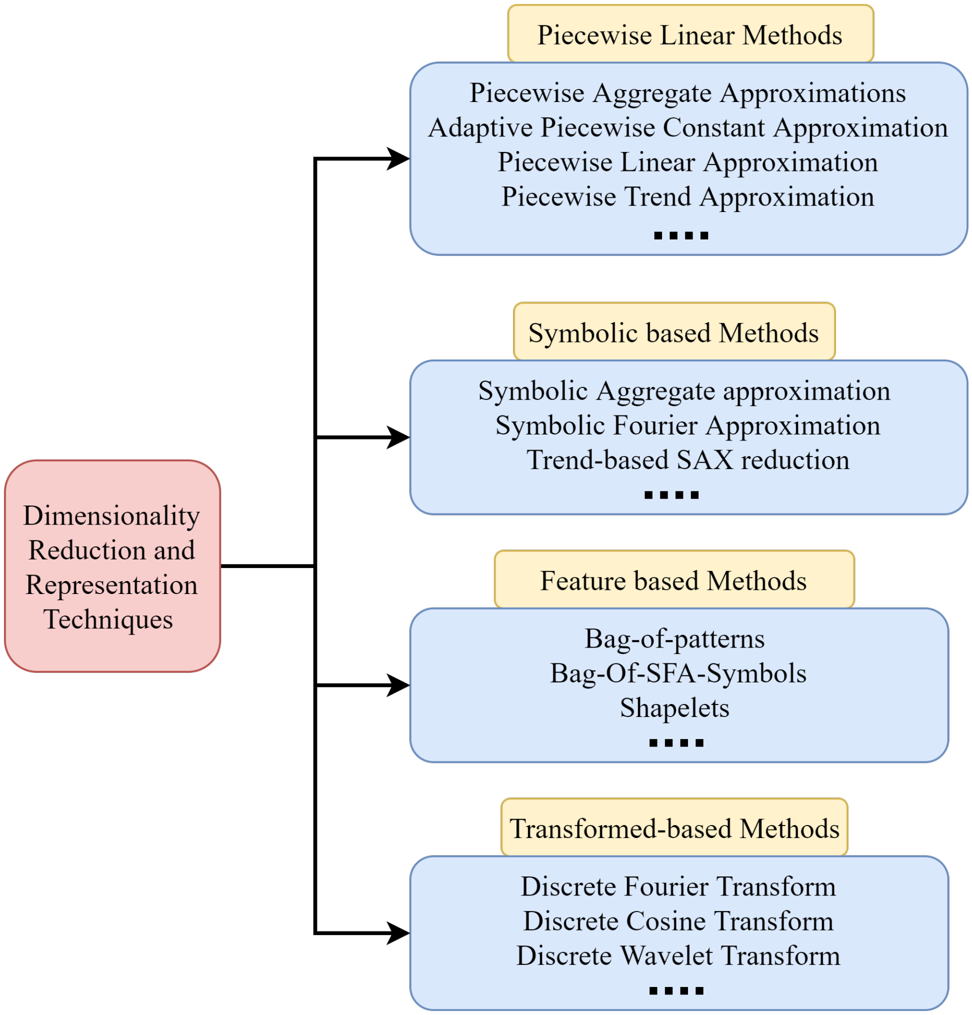


**Figure 2.1**: Taxonomy of time series representation and dimensionality reduction methods

ods have been introduced in the literature, which includes piecewise linear methods, symbolic-based methods, feature-based methods, and transformation-based methods. Piecewise linear methods usually divide the time series into different segments and then approximate the segments to represent the series. It includes a Piecewise Aggregate Approximation (PAA) [27], Piecewise Linear Approximation (PLA) [28], and Piecewise Trend Approximation (PTA) [29]. Symbolic-based methods represent the time series into a symbolic format to provide a higher level of an approximation than piecewise linear methods, for example, Symbolic Aggregate approXimation (SAX) [30] and Symbolic Fourier Approximation (SFA) [31]. Feature-based methods first learn the unique features set from the training data and then represent the time series in terms of these features. It includes a Bag Of Patterns (BOP) [32], Bag-Of-SFA-Symbols (BOSS) [33], and shapelets [34]. Finally, transformation-based methods transform the time series from one domain to another domain, such as Discrete Fourier Transformation (DFT) [35], Discrete Cosine Transformation (DCT) [36], etc. The taxonomy of representation methods [37] is presented in Figure 2.1.

### 2.1.2 Time series distance/similarity measures

Similarity measures are used to quantify the degree of similarity or dissimilarity between the time series. When these similarity measures are applied to the same problem, they capture the different aspects of the similarity [38]. Euclidean distance is the most simple and effective similarity measure which requires no parameter from the user. However, it has some limitations, such as not adaptive to noise, phase changes, and unequal length time series. Various elastic similarity measures also suggested to overcome these problems such as Dynamic Time Warping (DTW) [39, 40], Longest Common SubSequence (LCSS) [41], Edit distance with Real Penalty (ERP) [42], etc. DTW is a widely used similarity measure, which is more robust against distortion but computationally expensive. The LCSS is robust against the noise and outlier, but threshold setting with care is required for similarity measure. In this line, variants of edit distance are also suggested, such as ERP, Edit Distance on Real sequence (EDR) [43], Extended Edit Distance (EED) [44] to handle the different aspect of applications.

**Definition 2.1 Time series** *is defined as the ordered sequence of values, typically recorded at equal-space time intervals. It is denoted as* $X = \{x_1, x_2, \ldots, x_T\}$*, where* $T$ *is the length of the time series and* $x_t \in \mathbb{R}^V$ *for* $1 \leq t \leq T$*. When* $V \geq 2$*, time series is referred as multivariate otherwise univariate. In general, the term time series is referred to* univariate *until specified.*

### 2.1.3 Time series classification

TSC has received significant attention in the data mining community due to its applicability in various domains and also the availability of a large number of labeled datasets such as UCR [24], UCI [45], and MTS datasets [46]. TSC is a supervised learning task, in which a classifier $h$ is build based on given training set $\mathcal{D} = \{(X^i, y^i), 1 \leq i \leq M\}$ where $X^i$ is the $i^{th}$ time series with corresponding class label $y^i \in \mathcal{Y}$ for some discrete set of class labels $\mathcal{Y}$ and $M$ is the number of samples in the training set. The classifier learns the mapping function between time series and class label, which is formally defined as $h : X_T \rightarrow y$, where $X_T \in \mathbb{R}^T$ denotes the complete time series. The main objective of classifier $h$ is to classify the time series $X'_T$ (new time series with an unknown class label, probably with same domain) as accurately as possible. The TSC approach is depicted in Figure 2.2:

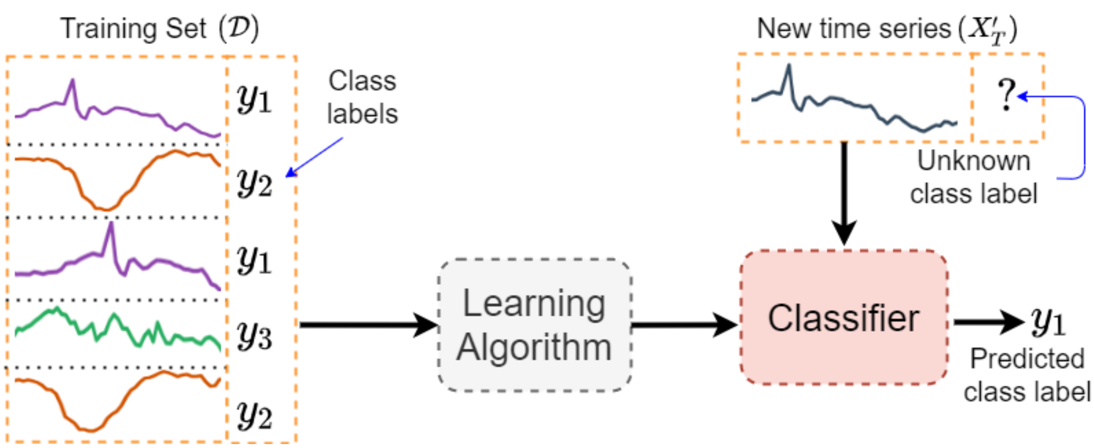


**Figure 2.2**: Traditional TSC approach.

TSC methods broadly divided into three categories [47]. The first category is the *distance-based classification*. The distance function measures the similarity between the new time series and all the time series in the training set. The class label is assigned to the time series $X'$ based on the closest time series in the training set. Hence, it is also called instance-based classification because the classification result depends on K nearest neighbour instances (time series) in the training set. The one nearest neighbour

(1-NN) classifier with Euclidean distance is used as the baseline method for TSC [7]. It is formally defined as:

$$dist(X, X') = \sqrt{\sum_{t=1}^{T}(X_t - X'_t)^2} \tag{2.1}$$

$$h(X') = \{y^i | \underset{i \in [1 \ldots M]}{\arg\min}\; dist(X^i, X')\} \tag{2.2}$$

1-NN with Euclidean distance has provided competitive performance [48]. However, the Euclidean distance similarity measure is suffered from noise and phase dispersion in the time series. In literature, various similarity/distance measures have been proposed including, Manhattan distance, DTW, ERP, and LCSS.

The second category is the *features-based method* in which each time series is transformed into a feature vector and then it fed into any conventional classifier such as neural network and decision tree for traning. In this line, some spectral feature representation methods are also included, such as DFT, discrete wavelet transform, and singular values decomposition. Distance-based feature transformation is also found in the literature that can be global or local [49]. In the global distance feature, time series is transformed into feature vector by computing the distance from all the time series in the training set. Hence, attributes in the feature vector depend on the number of time series in the training set. Kate in [50], utilized various distance measures as a feature within Support Vector Machine (SVM) for TSC. In the local distance feature, instead of computing distances between the entire time series, distance is computed using local patterns that are maximally class representative. These local patterns are the sub-sequences of time series, so-called shapelets [34]. Basically, Lines et al. [51] introduced the concept of shapelet transformation, in which firstly shapelets are discovered by employing the sliding window on each time series in the training set, and then maximally informative top $K$ shapelets are selected for feature transformation.

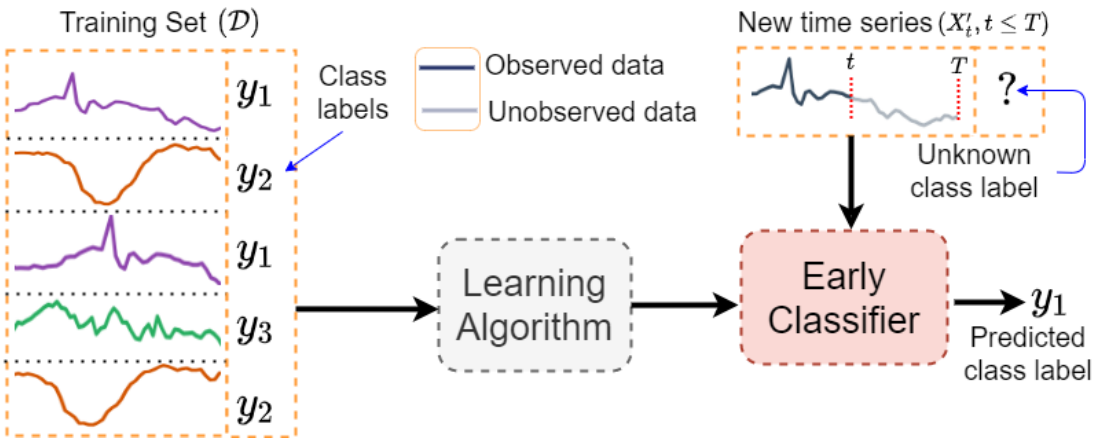


**Figure 2.3**: Early classification on time series.

The third category is the *model based classification* methods, in which it is assumed that the underlying model generates all the time series in a class. In the training step, the model learns the probability distributions of time series in a class defined by a set of parameters. The class label is assigned to the new time series that best fits the model in the testing step. Some of the examples of this approach are Naive Bayes model [52], auto-regressive model [53] and Hidden Markov model [54].

**Definition 2.2 Incomplete time series** *is considered a incomplete or partially observed, if it is having only initial $t$ data points of full-length time series. It is formally defined as $X_t = \{x_1, x_2, \ldots, x_t\}$, where $t \leq T$ is the length of incomplete time series.*

### 2.1.4 Early classification on time series

Although the concept of early prediction is sometimes vague in the literature and even used equivalently by some authors for forecasting, they are two different modeling problems with different objectives and characteristics. Early classification on time series is a special case of the traditional TSC problem where the objective is to maximize the quality of prediction with the added property of minimizing the prediction time [9]. Formally, early classifier $h$ is able to provide the class prediction for incoming time series $X'$ with $t$ data points only, where $t \leq T$. It is formally defined as $h : X_t' \rightarrow \hat{y}$, where $X_t'$ is the incomplete time series and $\hat{y}$ is the predicted class label. Figure 2.3 demonstrates the early classification of time series.

The traditional classification model required complete time series to perform the classification task. In contrast, the early classification model process incomplete time series to predict the class label at an early stage. It means the input of the classifier always has some missing values. Thus, two challenges arise in the designing of early classification model [55].

- First, to define the *classifier strategy* so that early classifier became adaptable to missing values.
- Second, the implementation of *decision policy* so that it could provide reliable class prediction at an early stage.

**Classifier strategy for handling missing value** :

Basically, three types of strategies have been suggested for an early classifier to handle incomplete time series [55].

1) *Adapting to missing values:* The methods in this category do not use all the information contained in the complete training time series. These methods deal directly with an incoming time series and predict its class label without carrying out any operation to complete it. These methods are implemented as distance-based models or as a series of classifiers.
2) *Imputation of missing values:* The methods in this category utilize the full-length time series information to make a class prediction on incomplete time series. Basically, these methods perform implicit or explicit imputation to make the time series complete.
    - Implicit imputation methods leverage the information included in the complete series data to make a class prediction. For example, the cluster-based approach utilizes the closest cluster of incoming time series for imputing the missing values.

- Explicit imputation methods first complete the incomplete time series with explicit imputation, for example, single imputation, machine learning imputation [56], imputation with forecasting [57], etc. The former method fills the missing values with zero, the conditional mean value, or the last observation carry forward approach. The latter methods explicitly learn the model from training data for imputing the missing values of incomplete time series.

3) *Representation of missing values:* The methods in this category implicitly use all the information contained in the complete time series. Basically, these methods change the representation of incomplete time series in another time-invariant domain in order to make it complete. Moreover, the prediction of the class label is performed on the transformed data. Let incomplete time series $X_t \in \mathbb{R}^t$ is in time domain. Then it is transformed in another domain by some function $\Phi(X_t) \in \mathbb{R}^K$.

**Decision strategy**:

Decision strategy is the heart of the early classification model, which takes the following conditions [55].

- A decision criterion decides when to stop measuring additional information and predict the final output.
- Optimization between time and quality in prediction trade-off helps in finding the optimal balance between these two objectives, i.e., accuracy and earliness.
- Adaptability refers to the handling of incomplete time series $X_t$. An early classification approach is said to be adaptive if it is able to make decisions at any time point $t$ while collecting the data points over time.

## 2.2 Related works

TSC has a valuable impact on solving various applications in data mining and machine learning. In the last few decades, numerous traditional TSC methods have been studied [7], whereas, in recent times, early classification on time series data has received great research interest [58, 59, 60, 61, 11, 12, 10, 62]. In the literature, early classification methods have been designed for UTS as well as MTS. However, early classification on UTS has attracted researchers more than MTS. We group the early classification approaches into three categories, namely instance-based [63, 64, 65, 9, 61], shapelet-based [66, 67, 68, 59, 69, 70, 13, 62, 71], and model-based [72, 60, 12, 73, 74, 75, 76, 11]. *Instance-based* early classification methods basically learn the Minimum Predictive Length (MPL) of time series in the training set and further use it to make reliable class predictions on incoming time series. The methods that fall in this category look for matching the incomplete time series in the training data set and classify the time series when the MPL condition is satisfied. *Shapelet-based* early classification methods mainly focus on extracting the key shapelets from the training set and use them to classify the new time series. The shapelets are the subsequences of the time series, having high discriminative power. Thus they act as a class representative. Shapelet-based early classification approaches discover the local shapelet with the help of some utility measures to provide support in early decision making. *Model-based* early classification methods utilize discriminative or generative classifiers to provide the classification results and predict the class label when the defined reliability threshold or decision function is satisfied. If any early classification method does not fall in any of these three categories, then it is considered in other category.

### 2.2.1 Instance-based methods

Initially, the methods for early classification on time series were developed by considering a fixed set of time points to train the set of classifiers and predict the class label of the test sample based on the incomplete time series [63]. Bregón *et al.* [63] is the first instance-based method that uses a case-based reasoning approach for early fault classification in the laboratory plant. The method utilizes the K-NN classifiers with different distance measures such as Euclidean, Manhattan, and DTW for analysis. The above-mentioned methods classify the time series at the prefixes of time series, and as a result, no adaptive decision policy has been designed for early classification. In [64], the authors presented the two methods, namely sequential classification rule (SCR) and generalized sequential decision tree (GSDT), to tackle the problem of early classification on symbolic sequences. These methods extract a large number of sequential rules from the different lengths of prefix space and select top $K$ rules with the help of support and prediction accuracy.

Xing *et al.* [9], formally defined the early classification of the time series problem and presented a 1-NN based early classification approach that analyses the nearest neighbour stability relationship in the training set. The authors present the MPL concept that was learned for each time series in the training set. This approach classifies the new time series based on the MPL of matching time series in the training set. It also tries to achieve decent early classification accuracy compared to the conventional 1-NN approach on full-length time series. A similar method for MTS, namely Multivariate Time Series Early Classification based on Piecewise aggregate approximation (MTSECP), has been presented in [61] by including two additional pre-processing steps. First, MTS is converted into UTS by computing center sequence. Then dimensionality reduction is performed with the help of the piecewise aggregate approximation technique. MTSECP does not utilize variable information in MTS effectively. These instance-based methods basically define the MPL for each time series in the training set and do not consider the earliness in their learning process. Typically these methods only consider the accuracy as decision criteria while defining the MPL for reliable class prediction.

### 2.2.2 Shapelet-based methods

Shapelet-based methods for early classification are highly adopted, notably in the domain of medical and health informatics, due to its interpretability. Basically, shapelets are the sub-sequences of time series that have the discriminating power to differentiate the time series among multiple classes. Moreover, shapelets represent distinct class patterns, and hence, they are called interpretable features. The first baseline of this type of approach is presented in [66] for early classification on UTS and named as Early Distinctive Shapelet Classification (EDSC). Firstly, EDSC adopted two methods, namely kernel density estimation, and Chebyshev's inequality, for learning the shapelet threshold. Then they mined the best local shapelets using defined utility measures, which is highly effective for early classification.

Ghalwash *et al.* [67] proposed an extension of the EDSC with an additional uncertainty estimation property since EDSC does not have any assessment of the uncertainty while deciding on the class prediction. Furthermore, Ghalwash *et al.* [68] extended the concept of shapelets for early classification on MTS and proposed a method named Multivariate Shapelet Detection (MSD). The MSD extracts the local key shapelets from N-dimensional MTS in the training set and classifies the new incoming MTS based on the best matching key shapelet. The limitation of this method lies in the employment of a sliding window in mining the shapelets. As an effect, all the sub-sequences in a multivariate shapelet have the same start and endpoint. However, in many realistic scenarios, the variable's informative patterns can lie in a different part of MTS and need not be synchronous. He *et al.* [59] have tackled this issue and introduced a Mining Core Feature method for Early Classification (MCFEC). The MCFEC method first extracts the shapelets for each variable independently and then selects the core shapelets by proposing a utility measure termed as generalized extended F-measure. Finally, two classification strategies were proposed rule-based and query by a committee to classify the incoming MTS.

In [69], the authors proposed an approach called Interpretable Patterns for Early Diagnosis (IPED). This method formulated the problem as an optimization-based binary classification to extract the key multivariate shapelets from the training dataset having different start and endpoints compared to MSD. The IPED has been evaluated for viral infection in humans based on gene expression data. Lin *et al.* [70] developed a Reliable EArly ClassificaTion (REACT) method for heterogeneous MTS data, including categorical and numerical attributes. REACT generates the shapelets after discretizing the categorical time series and uses the concept of equivalence classes mining to avoid redundant shapelets. This method builds an early serial classifier that ensures accuracy stability compared to the full-length time series classifier.

Data imbalance is also a common problem in many real-world applications. In this regard, an ensemble-based early classification framework is presented in [13], called Early Prediction on Imbalanced Multivariate Time Series (EPIMTS), that can effectively handle inter and intra class imbalance for early classification. EPIMTS considers the correlation among the multiple variables while learning the key shapelets. Later, He *et al.* [62] extended this work by considering confidence estimation for reliable early class prediction on MTS. Recently, Zhao *et al.* [71] developed an early classification approach for patient monitoring in intensive care unit. They extracted multivariate shapelets called *MEShapelet* and predicted asynchronous MTS with interpretability. The above-given literature reveals the fact that the shapelet-based methods are highly interpretable for class prediction. However, two critical issues exist with this approach. Firstly, the shapelet's threshold is tough to define if the class-wise patterns are not well distinguishable. Secondly, the process of extracting informative shapelets is highly time-consuming and complex.

### 2.2.3 Model-based methods

Model-based early classification methods learn the mathematical model from the data and design decision criteria or rules to perform the classification task. Most of these methods define the decision criteria using conditional class probabilities. These probabilities are either computed by the generative model or the discriminative model. The authors in [72] presented the early classification approach using a generative model, which guarantees the reliability in assigning the class label to incomplete time series. Basically, they develop the decision rule for early classification based on Quadratic Discriminant Analysis (QDA) classifier by assuming that training data follow Gaussian distribution. Parish et al. in [77] further extended the approach and proposed a more tractable decision rule by using different classifiers, including linear SVM, Linear Discriminant Analysis (LDA).

A simple and effective model-based early classification approach has been presented in [60], based on discriminating the classes over time. This model develops the set of probabilistic classifiers at different timestamps and computes the reliability threshold for each class label. Moreover, the model also defines the discriminative safeguard point for each class. It classifies the incoming time series only if the reliability threshold (the difference between the two highest class probabilities) and specified safeguard points for the predicted class level are satisfied. Reference [12] developed a relatively similar framework, in which the confidence threshold was defined by fusing the classifier's true prediction probabilities at successive time steps. This framework is adaptable for both probabilistic as well as discriminative classifiers. A distance transformation-based framework for early classification on time series was put forth by Yao *et al.* [73]. Firstly, they represent the time series into distance space using informative local patterns and then train the probabilistic classifiers. Finally, they proposed a confidence area as a criterion for early decision-making on incoming time series. Li *et al.* [74] applied an early classification approach for human activity recognition based on partially observed activity information. They modeled the 3D action recognition problem as a stochastic process called a dynamic marked point process. Hsu *et al.* [75] proposed deep learning-based early classification on MTS through attention mechanism, which helped in identifying best-performing segments in variables of MTS.

Meanwhile, Dachraoui *et al.* in [58] consider the trade-off between two conflicting objectives, i.e., accuracy and earliness. Basically, they proposed a meta-algorithm for early classification on time series that is non-myopic in nature. In this approach, the authors included the misclassification cost and delaying decision cost in its optimization function for balancing the accuracy and the earliness. Thus, during testing, incoming time series was classified if only if the estimated cost at the current timestamp is less than the estimated cost at all future timestamps. However, this method used a clustering approach for evaluating future costs, which caused a lack of clarity in the overall process. Later, Tavenard *et al.* [76] introduced two new strategies (*NoCluster* and *2Step*) by eliminating the clustering step for further improvement. Mori *et al.* [11], introduced a framework for early classification on time series by defining the stopping rules as decision criteria and learned the rules by optimizing the accuracy as well as earliness simultaneously.

### 2.2.4 Other methods

This section presented the methods that do not fall in the above three categories. The first time, Rodríguez and Alonso [78], come up with the concept of early classification as a natural finding. They presented boosting interval-based literals method to classify the variable-length time series. This method first partitioned the time series into different time intervals and defined its states as increasing, decreasing, staying, etc. Later each interval is replaced by predicates that signal the presence or absence of the states. Finally, the predicate's output is used for classification at defined intervals. Thus, they did not attempt to optimize the trade-off between accuracy and earliness. In [79], The authors presented a reinforcement learning-based early classification framework. They introduced an early classifier agent and reinforcement learning agent as deep Q-network to perform the early classification task. The proposed framework defines the decision function as a learning agent that uses the reward function to balance accuracy and earliness.

A brief summary of early classification methods has been provided in Table 2.1. It has been presented based on three key factors:

- They way it handles the incoming time series.
- The type of decision strategy it considers.
- The trade-off between accuracy and earliness is considered or not while learning the model.

**Table 2.1**: Comparative analysis of early classification approaches

| ID | Year | Classifier used | Handling missing values | Trade off | Decision strategy | Method type | Series type |
|---|---|---|---|---|---|---|---|
| [78] | 2004 | Adaboost ensemble classifier | Not adaptive | No | No explicit function was defined for decision-making. | Other | MTS |
| [63] | 2006 | 1-NN with Euclidean, Manhattan, and DTW distances | Adaptive to missing values, distance-based method | No | The decision function is defined as a user-defined threshold. | Instance-based | MTS |
| [64] | 2008 | Decision tree and Rule-based classifier | Representation, dictionary based method | No | The decision function is triggered when the user expected accuracy achieved. | Instance-based | UTS |
| [65] | 2009 | 1-NN | Adaptive to missing values, distance-based approach | No | Minimum prediction length is defined as decision criteria. | Instance-based | UTS |
| [66] | 2011 | Closest shapelet using Euclidean | Missing values representation, dictionary-based approach | No | Early decision is performed when the defined shapelet is matched with the incoming time series. | shapelet-based | UTS |
| [9] | 2011 | 1-NN | Adaptive to missing values, distance-based approach | No | Minimum prediction length is defined as decision criteria. | Instance-based | UTS |

*Continued on next page*

Table 2.1 – *Comparative analysis of early classification approaches*

| **ID** | **Year** | **Classifier used** | **Handling missing values** | **Trade off** | **Decision strategy** | **Method type** | **Series type** |
|---|---|---|---|---|---|---|---|
| [68] | 2012 | closest multivariate shapelet using Euclidean | Representation to missing values, dictionary-based approach | No | Variate-wise shapelet threshold has been defined, and a classification task is performed when variate-wise thresholds are satisfied. | shapelet-based | MTS |
| [80, 72] | 2012 | Linear SVM and Local QDA | Missing values imputation conditioned on incoming time series as well as the distribution of training samples | No | The decision function has been defined as the probability threshold. | Model-based | UTS |
| [72] | 2012 | QDA | Missing values imputation conditioned on incoming time series as well as the distribution of training samples | No | The decision function has been defined as the probability threshold. | Model-based | UTS |
| [17] | 2013 | SVM | Adapting to missing values, A series of classifiers developed | No | The decision criterion has been defined as confidence score, considering accuracy only. | Model based | MTS |
| [69] | 2013 | Closest multivariate shapelet using Euclidean | Representation to missing values, dictionary-based approach | No | Formulated the problem as convex optimization and extracts the key shapelets to classify the time series. | shapelet-based | MTS |
| [81] | 2013 | Closest multivariate shapelet using Euclidean | Representation to missing values, dictionary-based approach | No | Core shapelets are extracted using a defined utility measure that takes earliness to account. | shapelet-based | MTS |

*Continued on next page*

Table 2.1 – *Comparative analysis of early classification approaches*

| ID | Year | Classifier used | Handling missing values | Trade off | Decision strategy | Method type | Series type |
|---|---|---|---|---|---|---|---|
| [77] | 2013 | Linear and quadratic discriminants | Missing value imputation | No | Early decision criteria are defined as reliability thresholds. | Model-Based | UTS |
| [82] | 2014 | Rule-based and Query by Committee classifiers | Representation to missing values, dictionary-based approach | No | Core shapelets are extracted using a defined utility measure. | shapelet-based | MTS |
| [67] | 2014 | Closest shapelet using ED | Representation to missing values, dictionary-based approach | No | Decision function has been defined based on shapelets matching with confidence threshold, computed using uncertainty in predictions. | shapelet-based | UTS |
| [83] | 2015 | Linear SVM | adapting to missing values, a series of classifier developed with an ensemble classification approach | No | Ensemble classifier developed with minimization of empirical risk and response time simultaneously. | Model-based | UTS |
| [58] | 2015 | Naïve Bayes and Multi-layer Perceptrons | Imputing the missing values | yes | The decision policy has been defined based on cost estimation at current as well as all future time steps. This approach is non-myopic in nature. | Model-based | UTS |
| [70] | 2015 | Decision tree | Representation to missing values, dictionary-based approach | No | The decision criteria are defined as heterogeneous shapelets with a defined threshold. Earliness is not taken into consideration while learning the shapelets. | shapelet-based | MTS |

*Continued on next page*

Table 2.1 – *Comparative analysis of early classification approaches*

| ID | Year | Classifier used | Handling missing values | Trade off | Decision strategy | Method type | Series type |
|---|---|---|---|---|---|---|---|
| [16] | 2015 | Hybrid model using HMM and SVM | Adaptive to missing values | No | Early decision is performed when the difference between two class probabilities is higher than the certain threshold value. | Model-based | MTS |
| [76] | 2016 | SVM | Imputing the missing values | yes | The decision function has been defined based on cost optimization between current and future predictions. | Model-based | UTS |
| [60] | 2016 | GP classifier | Adaptive to missing values by developing a series of classifiers. | No | The decision criteria have been defined based on the difference between the first two highest class probabilities, conditioned on user-defined parameters. | Model-based | UTS |
| [84] | 2016 | CNN | Implicit imputation of missing values | No | No explicit decision function is defined for early classification. Hence no trade-off optimization in decision learning. | Model-based | UTS |
| [61] | 2017 | 1-NN | Adaptive to missing values with distance-based classifier | No | Minimum required length has been defined for making decisions. | Instance-based | MTS |
| [11] | 2018 | GP and SVM | Adaptive to missing values; a series of classifiers | yes | Stopping rules are defined to classify incomplete time series. | Model-based | UTS |
| [79] | 2018 | Reinforcement learning agent | adapting to missing values | yes | A reinforcement learning agent has been defined to make early decisions. | other | UTS |
| [74] | 2018 | Stochastic process | Not adaptive to missing values | No | No specific decision policy is defined for early classification. | other | MTS |

*Continued on next page*

Table 2.1 – *Comparative analysis of early classification approaches*

| ID | Year | Classifier used | Handling missing values | Trade off | Decision strategy | Method type | Series type |
|---|---|---|---|---|---|---|---|
| [85] | 2018 | Combination of CNN and LSTM | No adaptive to missing value | No | No decision policy has been defined for making a reliable class prediction. | Model-based | MTS |
| [71] | 2019 | Decision tree and random forest | Representation based approach | No | Shapelet Matching | shapelet-based | MTS |
| [62] | 2019 | Multivariate shapelet with rule-based classifier | Representation based approach | No | Decision criteria have been defined as shapelet matching with Cumulative confidence. | shapelet-based | MTS |
| [12] | 2019 | Dictionary classifier with WEASEL | Representation to missing values, dictionary-based approach | yes | Confidence threshold is defined to classify the incoming time series. | Model-based | UTS |
| [86] | 2019 | Combination of CNN and LSTM | Explicit imputation of missing values | yes | The decision criteria are defined as probability threshold. Trade-off optimization has been taken into consideration while learning the model. | Model-based | UTS |
| [10] | 2019 | GP and SVM | adapting to missing values | yes | Rules have been designed for making an early decision. | Model-based | UTS |
| [73] | 2019 | Closest shapelet using Euclidean | Representation to missing values, dictionary-based approach | No | The decision criteria have been defined as confidence areas. The framework considers the earliness in selecting the shapelets only. | shapelet-based | UTS |

### 2.2.5 Research Gap

Certain limitations have been observed in the existing early classification approaches, which are as follows:

- The methods proposed in [78, 63, 17, 60] learn the classification model and design various mechanisms to determine whether the prediction is accepted or rejected at different time points to classify the time series. Moreover, these methods define some fixed MPL or confidence threshold to perform early classification. Hence these methods have not considered the trade-off optimization between earliness and accuracy. Even this is a desirable property for early classification problems.
- The method in the references [64, 87, 9, 66, 60, 11] specifically designed for UTS and are not suitable for MTS data directly. Developing early classification methods for MTS is a challenging task compared to UTS because of multiple variables, e.g., each variable in MTS represents a UTS. Often, these variables may be of different lengths and have hidden interconnected relationships. For example, in the health monitoring system, various patient parameters are collected simultaneously, such as temperature, blood pressure, oxygen label, and heartbeats, to measure the health of the patient.
- A very few notable works have been found in the literature that tried to provide early classification on MTS data. These works have been accomplished using shapelet-based methods [82, 67, 69, 59, 13, 70, 88]. These methods utilize the local shapelets as interpretable features that are generated from the MTS training set. Thus, irrespective of its interpretability, these methods demand intensive computation to extract informative shapelets [34]. Moreover, the existing approaches for early classification on MTS have not adequately addressed conflicting objectives: accuracy and earliness.