Preface

Getting real-time information becomes convenient and effortless with the availability of low-cost sensors. These sensors generate multiple sequences of measurements taken at regular intervals. Each sequence is a time series of measurements (data points). A Multivariate Time Series (MTS) consists of more than one such time series that are collected from the sensors. If a time series is a part of MTS then it is referred as component. Early classification of an MTS is mainly an extension of traditional classification with an ability to classify the MTS with limited number of data points. One can achieve better accuracy by using more data points but the response time will be increased. In other words, earliness will be reduced. It indicates that there exists a tradeoff between accuracy and earliness. Now, the main objective of an early classification approach is to optimize this tradeoff while building the classifier.

In the field of data mining and machine learning, early classification has been an important area of research for many years. It is a primarily studied topic for minimizing class prediction delay in various applications including human activity recognition, gene expression based health diagnostic, industrial process monitoring, and so on. The existing literature solved the early classification problem with following assumptions about MTS: all the components are of equal length, each component is reliable (*i.e.,* no faulty component), and all the class labels are known (seen). In order to have equal length components, the sensors are required to record the data at equal sampling rate. However, it is an unrealistic assumption in real-world applications where the sensors are purposefully set to record the fine-grained information at different sampling rate. An example of such applications is intelligent transportation system where the sensors (*e.g.,* temperature, accelerometer, vibration, *etc.*) are desired to have different sampling rate to observe different inside and outside conditions of vehicle. Next, the faulty components may be present in an MTS if one or more sensors are faulty (unreliable). Further, in some applications such as appliance monitoring, early classification of the MTS is desirable for fault identification. As an appliance may encounter several types of faults during its lifetime, it is impractical to have training instances for all of them (*i.e.,* some faults are unseen). Thus, the early classifier should be able to classify an MTS (corresponding to a fault) even if it belongs to an unseen class.

In this thesis, we study the early classification problem for sensors generated MTS with the relaxation of aforementioned assumptions. The first contribution of this thesis is to classify an incomplete MTS which is generated from the sensors of different sampling rate. This work proposes a probabilistic approach for early classification of MTS where a class forwarding method is developed to handle the varying length of components of the MTS. Next contribution of this thesis is to classify an MTS even when some of its components are faulty. We propose a fault-tolerant early classification approach that first identifies the faulty components of a testing incomplete MTS with the help of training data and then employs a probabilistic classifier to predict its class label. The approach also incorporates a concept of partial order set to prune out the irrelevant components. In the last contribution, we address the early classification problem in the presence of unseen class for which no training instance is available. As a solution, this work proposes a probabilistic early classification approach that can identify an unseen class by utilizing the semantic information or attributes of the seen classes.