Abnormalities of cardiac rhythms are correlated with significant morbidity. For example, atrial fibrillation, affecting at least 2.3 million people in the United States, is associated with an increased risk of both stroke and mortality; supra-ventricular tachycardia, detected in approximately 90,000 cases annually in the United States, results in hospitalization in about 25% of all emergency department visits for supra-ventricular tachycardia; ventricular arrhythmias cause 75% to 80% of the cases of sudden cardiac death; bradyarrhythmias and other forms of conduction disease may cause syncope, fatigue from chronotropic incompetence, or sudden death from asystole or ventricular tachycardia. Due to the time-sensitive nature of cardiac events, it is of upmost importance to ensure that medical intervention is provided in a timely manner, which could benefit greatly from a cardiac arrhythmia monitoring system that can detect and preferably also predict the abnormal cardiac events.

In recent years, with the development of cardiac and other types of medical monitoring devices, vast amounts of physiological signal data have been collected and become available for analysis. Physiological signals such as electrocardiogram (ECG) have many clinical applications in cardiac arrhythmia, including diagnosis confirmation, drug effect monitoring and rate control. However, the extraction of the relevant information from physiological signals—despite its great value—is hindered by the complexity and variability within signal morphology, which leads to vague definitions and ambiguous guidelines, causing difficulties even for a well-trained medical expert.
Such variability-related issues are ubiquitous and manifested in different ways: via the ECG signals themselves, the measurements derived from such signals, and the diagnostic interpretations based on such measurements.

In order to address the variability-related issues, most traditional methods for physiological signal analysis depend heavily on pre-processing to identify specific morphology types (such as R peaks in ECG) and extract the related features. Despite many successes, one of the drawbacks of these methods is that they often require signal data of high quality and tend to be less effective in the presence of noise which could significantly distort the signal morphology. Although not an issue in almost noiseless situations such as bedside ECG monitoring, such pre-processing–based methods have become insufficient with the advent of portable arrhythmia monitoring devices in recent years capable of collecting physiological signals in real time, albeit with more noise. Therefore, in order to enable automated clinical decision making using such ECG sources, it is desirable to introduce new methods that require minimal pre-processing prior to analysis and are robust to noise.

This thesis aims to develop a cardiac arrhythmia monitoring and prediction system applicable to portable arrhythmia monitoring devices. The analysis is based on a novel algorithm which does not rely on the detailed morphological information contained within each heartbeat, thus minimizing the impact of noise. Instead, the method works by analyzing the similarity and variability within strings of consecutive heartbeats, relying only on the broad morphology type of each heartbeat and utilizing the computer’s ability to store and process a large number of heartbeats beyond humanly possible. The novel algorithm is based on deterministic probabilistic finite-state automata which have found great success in the field of natural language processing by studying the relation among different words in a sentence rather than the detailed structure of the individual words.