1 BACKGROUND AND RELATED WORK

This section provides background information on time series and ECGs, as well as methods to analyze them.

1.1 Time Series and Time Series Analysis

TODO explain what univariate and multivariate time series are This subsection will provide background information on time series and time series analysis methods. A time series is a set of values recorded at specific times. A common form of time series are discrete-time time series (often simply called discrete time series). Discrete time series are time series whose values are recorded at discrete points in time, the most common example of this are time series with values recorded at fixed intervals. Continuous-time time series are time series that are recorded continuously over a certain interval [brockwell2016]. Time series are used in many disciplines to record information on time-dependent processes, e.g. stock prices in economics, the sun's activity in physics, or the heart's activity in medicine. Time series can be recorded digitally, physically, or, if they were recorded physically, can later be digitized. The recorded data can then be used to gain insight into the processes that were studied. To gain insight using a time series, the relevant information needs to be extracted from it–a process that is often called data mining. Data mining of time series is a vast discipline that, among others, includes [lin2003,aghabozorgi2015]:

- visualization (graphical representation),
- forecasting (predicting future behavior),
- indexing (finding the most similar time series to a given one),
- clustering (dividing time series into groups of similar ones),
- anomaly detection (detecting parts that are not "normal" or do not fit certain parameters),
- classification (assigning a label based on its features, e.g. "sick" and "not sick"), and
- summarization (reducing the complexity–often length–while preserving important features).

Challenges for time series analysis include the often very large data sets that are difficult for humans to analyze and take up considerable digital storage space. Analyzing very large data sets requires a large amount of computational power because most data mining algorithms become less efficient with larger data sets [lin2003]. To mitigate this issue, time series dimension reduction (also known as dimensionality reduction or time series representation) is used. Dimension reduction transforms a "raw" (unmodified) time series into a representation that is simpler but nonetheless resembles the raw time series. This can be achieved by either using a method that reduces the number of values in a time series, or by extracting only the relevant features from the time series. According to [aghabozorgi2015], there are four types of dimension reduction methods:

dr-methods:01
1. non-data adaptive,
2. data adaptive,
dr-methods:03
3. model-based, and
4. data dictated.

Methods 1–3 have their dimension reduction factors set by user-defined parameters. This means that the user can determine how much the dimension of the data should be reduced [aghabozorgi2015].

mentioned above have varying compression ratios, based on user-given input.

TODO fix this next segment, add more citations, maybe short descriptions

TODO turn these into subsections

TODO also cite shieh2008 here, has table and explanations too

TODO maybe just mention the categories, explain what they mean, then do a "in this research the focus is on SAX, a …"

TODO cite all this shit and rephrase

TODO add more information, maybe cut the sax thing short by referring to the methods section

1.1.1 Non-data adaptive representation

Non-data adaptive methods operate on time series segments with a fixed size to reduce the dimension and they are useful for comparing multiple time series with each other. These methods include wavelets, the Discrete Wavelet Transform, and the Piecewise Aggregate Approximation (PAA) [aghabozorgi2015]. - DFT - DWT - DCT - PAA - PIP - IPLA

1.1.2 Data adaptive representation

Data adaptive methods use non-fixed size segments and aim to fit the raw data most closely. Examples of data adaptive methods are the Piecewise Polynomial Approximation, Piecewise Linear Approximation, and Piecewise Constant Approximation [aghabozorgi2015]. - SVD - PLA - APCA - SAX

1.1.3 Model-based representation

Model based methods use stochastic methods such as Markov Models and Hidden Markov Models, and the Auto-Regressive Moving Average (ARMA) [aghabozorgi2015].

1.1.4 Data dictated representation

Data dictated methods derive their compression ratios from the data automatically, the most common form of this method is Clipped [aghabozorgi2015]. - Clipped Data

TODO create some type of transition here

1.1.5 The SAX representation

:section_sax

TODO add some paragraphs to this mess

A particular dimension reduction method is SAX. Introduced by lin2003 Lin, Keogh, Lonardi, and Chiu, SAX is a symbolic time series representation method for univariate time series. The authors felt that the symbolic methods available in lin2003 2003 did not provide the desired dimension reduction, did not correspond to the raw data accurately enough, and could not be applied to a subset of the total data. SAX uses the averaging of a user-defined number of segments and the labeling of segments with letters to reduce the dimension of the time series data. The number of letters, called the alphabet size, can also be chosen by the user and influences the dimension reduction. The distance between two time series in the SAX representation is guaranteed to resemble the distance between the two raw time series, this is called the distance measure. Since its creation, SAX has found widespread use in data mining and many researchers have attempted to modify and improve it.

The SAX distance measure has been improved to include the standard deviation [zan2016] and a measure of the trend of each averaged segment [sun2014,yu2019]. Extended SAX modifies SAX to include the minimum and maximum values of each segment for improved representation of the raw data [lkhagva2006] while 1d-SAX incorporates a linear regression over each segment into SAX [malinowski2013]. A combination of SAX and a polynomial approximation was used to speed up the SAX method [fuad2010]. To improve the indexing performance of SAX, iSAX introduced convertible alphabet sizes, allowing SAX representations with different alphabet sizes to compared with each other and indexed into a tree structure [shieh2008]. iSAX 2.0 improves the iSAX index by reducing its computational complexity, enabling it to index a time series that has one billion elements, something that SAX or iSAX cannot do [camerra2010a]. To perform time series anomaly detection using SAX, Heuristically Ordered Time series using SAX (HOT SAX) was introduced. HOT SAX sorts segments of a SAX-represented time series by their distance to other segments, effectively identifying the most abnormal segments of a time series [keogh2005].

SAX and its variants have also been used for the analysis of multivariate time series. SAX-ARM combines the SAX representation with association rule mining (identifying rules and implications found in the data, i.e. parameter a influences parameter b) to analyze multivariate time series and discover the rules underlying the data [park2020]. MSAX expanded the use of SAX to multivariate time series by utilizing multivariate normalization with the help of a covariance matrix and a modified distance measure [anacleto2020]. SAX has also been used to visualize multivariate medical test results and enable their analysis [ordonez, 2008]. Resource-aware SAX is a SAX variant developed to analyze ECG using a mobile device like a mobile phone. The method takes advantage of the computational efficiency of SAX to perform the ECG analysis on the device and even preserve its battery life. Another application of the SAX method to ECGs is [zhang2019], which uses SAX with an added binary measure of the trend of each segment to detect ECG anomalies, achieving a recall value of 98%.

1.2 ECGs and ECG Analysis

The following subsection covers the ECG and methods used in its analysis. Luigi Galvani noted the electrical activity in muscles 1786, but the history of the ECG only started in 1842, when Carlo Matteucci showed the electrical activity of a frog's heartbeat. In the 1870s, it was discovered that each heartbeat is characterized by electrical changes. Willem Einthoven was the first to publish an ECG waveform with the now standard annotations P, Q, R, S, and T for the different features. Then, in **TODO** refer to the figure of the heart beat here

In 1901-1902, Einthoven created the first ECG recording of a human heartbeat using using 3 leads connected to the limbs of the patient. He would receive the 1924 Nobel Prize in medicine for his invention of the electrocardiograph. As a result of further development, 12-lead ECG that we know today was created [alghatrif2012,fye1994].

TODO cite becker2006 on how ECGs work

TODO write a section on how the 12 leads work

TODO how does it work

TODO different methods of recording ECGs

TODO annotated ecg for parts, which diseases are apparent

TODO look at which ECG methods were covered in first lit review and take some of those

As mentioned above, SAX has already successfully been applied to ECGs, but there has not been much use of the method in that respect? **TODO** is that true?

MSAX has not yet been applied to ECG analysis.

1.2.1 ECG databases

TODO cite this shit

TODO turn this into at least one full paragraph

ECG data is patient data and thus not freely accessible in most cases. Online databases, most of them on Physionet, are an exception to this rule. Physionet provides databases on many types of medical data. They are all freely available and licensed for research and educational use.

TODO to put in:
- previous research in this area
- current research in the area
- show why more work is still important
- enough background to really clear up the problem
- stuff relevant to question, and some background

TODO Sections:
- overview of methods for ECG analysis
- main elements
- current foci
- use the confusion matrix at all?
- support assertions made in introduction
- arrive at natural conclusion that SAX/MSAX/HOTSAX should be investigated
- describe all 4 statistical measures in some detail

TODO Structure:
- follow the same overall structure as the introduction, order of topics
- background on history of ECG
- how an ECG works in more technical terms
- how do cardiologists detect heart diseases?
- move the annotated graph down here and leave the general graph in intro?
- which time series methods are being applied to ECGs? strengths an weaknesses
- what are current hot topics in this field?
- more research on sax, msax, hotsax and what people have done with it
- ecg applications; why is it relevant
- RESEARCH MORE ABOUT ”ecg discord discovery algoorithm”
- use some of their arguments to support my choice
- try to make it flow so that my choices seem to come from the literature analysis
- maybe add a section that talks about how to evaluate these types of algorithms

2 STATE OF COMPUTERIZED ECG ANALYSIS

Recent advances in computer technology have enabled the use of computers in every aspect of ECG acquisition, processing, analysis, and storage. In light of these developments, the American Heart Association published recommendations for the interpretation and standardization of the ECG. They recommend that the low-frequency cutoff for low-frequency filtering of an ECG should be 0.05 Hz or 0.67 Hz for filters that do not exhibit phase distortion. For high-frequency filtering they recommend a cutoff of at least 150 Hz. For the storage of digital ECG samples (at 500 samples per second), it is recommended use use compression with an error of less than 10 microvolt [kligfield2007].

xie2020 provide an overview of the current approaches to computerized ECG analysis. The standard approach to using computerized methods in ECG analysis is comprised of four steps (1) denoising of the raw ECG signal(s), (2) feature engineering, (3) dimensionality reduction, and (4) classification. To denoise an ECG, digital filters are often used. Their drawbacks are that they only filter out very specific frequencies. Because noisy ECGs contain different types of contaminations, digital filters can be inaccurate. Using wavelet transforms for denoising has the advantage that noise can be more precisely targeted and the clean signal reconstructed afterwards. Choosing appropriate wavelet parameters can be challenging and methods to optimize this process have been proposed. Empirical mode decomposition is the third option generally employed to denoise an ECG. It does not require the user to set parameters but it can lead to a mixing of oscillations of different time scales.

After the signal has been appropriately denoised, feature engineering is performed. Feature engineering is the process of extracting features that are relevant for diagnosis from the many points the ECG signal contains. The main features targeted for extraction are the PQRST features mentioned in the introduction. The fast Fourier Transform provides a way of analysing the frequency domain of the ECG signal, enabling the detection of the QRS complex and other features. The missing time information in the fast Fourier Transform can lead to difficulties in detecting time-dependent features. The short-time Fourier Transform adds time information to the fast Fourier Transforms data. This can increase the accuracy of the feature extraction. This transform has the drawback that there is a tradeoff between the time and frequency resolutions. Wavelet transforms can also be used for feature extraction. They have the advantage that they are suitable for all frequency ranges. Choosing the right wavelet base for the desired application can be a challenge. The discrete wavelet transform is the most widely used wavelet transform, thanks to its computational efficiency. Statistical methods are also used to extract features from ECGs; those methods are generally less affected by noise in the signal.

After the features of the ECG have been extracted, it is often necessary to reduce the number of features. The reason for this is that a large number of features, despite their high accuracy, require a high amount of computation to classify. This lengthy computation can negate the advantages gained by high accuracy. This process sacrifices a certain amount of information and sometimes precision, but significantly speeds up the classification. Feature selection is a process that attempts to select a subset of the original data that adequately describes the whole data. Feature selection can be performed by a filter that filters out unnecessary attributes based on some metric. This methods is relatively simple, but the filtering process removes data and thus negatively impacts the precision of further steps. Feature extraction on the other hand uses dimensionality reduction methods to keep as much of the original information as possible. Principal component analysis preserves as much of the variance in the original data as it can. Other algorithms focus on separating classes of data, pattern recognition, or retaining the structure of the original data.

The final stage of the ECG processing is the classification stage. In this stage judgements are made based on the prepared input data and the result should be a disease diagnosis. In the early stages of computerized ECG analysis classification was performed by algorithms based on human actions when reading an ECG. Those algorithms were basic and not particularly accurate. Currently, the classification at the end of the preparation process is performed by a machine learning algorithm. Such models include the k-nearest-neighbors model which classifies points into groups but which is very expensive to calculate for high-dimensional data. Support vector machines are used for pattern recognition and are able to work with small samples. Artificial neural networks are robust and can work with complex problems, they are generally more accurate than support vector machines. The newest approach is to forego the stages discussed here and use a single neural network to perform all the required tasks ”end-to-end”. These networks are fed raw data and the denoising, feature extraction, selection, and classification is performed internally by the model [xie2020].

The end-to-end approach to ECG analysis is a relatively new development and is being actively researched. The more traditional method using denoising, features engineering, and classification as separate steps is also still relevant. The combination of denoising and feature extraction with a machine learning classifier can lead to very good results. **prasad2018** use the fast Fourier Transform to extract features from an ECG and then employ a multi-objective genetic algorithm to detect abnormal ECG signals with high accuracy. **vaneghi2012** compare 6 common feature extraction techniques with respect to their detection of ventricular late potentials. The compared methods are the autoregressive method, wavelet transform, eigenvector, fast Fourier Transform, linear prediction, and independent component analysis. **valupadasu2012** use the fast Fourier Transform to analyze the energy level in different frequencies in the ECG of patients with IHD. They find that the energy is distributed differently, allowing the distinction of ECGs with IHD from those without IHD. kaur2016 Kaur, Rajni, and Marwaha [2] analyzed ECG signals with both the wavelet transform and principal component analysis. They found that the wavelet transform outperformed principal component analysis for the detection of heart beats in an ECG. Their model achieved an error rate of 0.221% of incorrectly classified heart beats