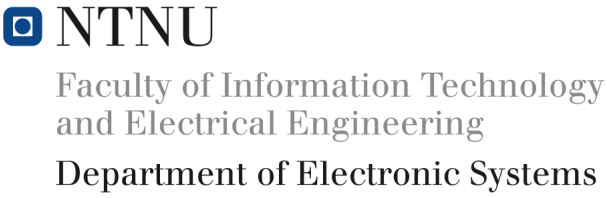
Auto-strain master thesis

Written by
Yohann Jacob Sandvik

Master thesis EMNEKODE

Supervised by
Lasse Løvstakken



Department of electronic systems
Faculty of Information Technology and Electrical Engineering
Norwegian University of Science and Technology
May 28, 2020

Abstract

This is the abstract.

Acknowledgements

These are my acknowledgements.

Contents

List of Abbreviations

ARI Adjusted Rand Index. 18

BMI Body Mass Index. 4, 10, 11

EF Ejection Fracture. 4, 10, 12, 13, 15, 18

GLS Global Longitudinal Strain. 4, 12, 13, 15–18

ML Machine Learning. 17

RLS Regional Longitudinal Strain. 17, 18

RNN Recurrent Neural Network. 18

TSC Time-series clustering. 18

List of Figures

List of Tables

Chapter 1

Introduction

This is the introduction.

1.1 Motivation

This will be the section on the motivation for the assignment.

1.2 Objective

This will be the section where i outline the objective of the assignment.

1.3 Structure of Thesis

Here the outline for the rest of the assignment will be given.

Chapter 2

Myocardial Imaging and Echocardiography

This will be a kind of theory section about echocardiography, and strain imaging.

2.1 Basic Cardiology

2.2 Introduction to Echocardiography

2.3 Myocardial Strain

Chapter 3

Machine Learning Theory

This section will act as a theory section for the machine learning models used.

Chapter 4

Review of The Literature

This chapter will contain the review of the literature.

Chapter 5

Data Exploration

In this chapter the variability, distribution and type of data used in the assignment will be explored. The exploration is divided into three sections corresponding to the three main groups of variables: The *patient meta-data*, the *input variables* and the *target variables*. The *meta-data* is the data about the patients which is not used in the classification models, but can be used to give a description of the patient demographich which makes up the dataset. The *input variables* are the variables that are inputed into the machine learning models in order to train them, and later used to make predictions about the patients' *target variables*. The target variables are then variables that the models will be trained to predict. Target variables are used both in training to correct erroneus predictions that models make during training, and to evaluate the accuracy of the model after training.

5.1 Patient Meta-data

The patient meta-data that will be considered in this section are age, gender, Body Mass Index (BMI) and blood pressure.
Figure 5.1 shows the patient distributions with regard to age, gender and BMI. As evident from the figure the patients that make up the dataset is made up of 138 males and 57 females. The majority of the patients are in the age group 60-80 years with a number of patients in the range 80-90 years (AGE SECTION SUBJECT TO CHANGE). The BMI distribution of patients is centered around 26 kg/m^2. Figure 5.2 shows the joint distribution of systolic and diastolic blood pressure among the patients.

5.2 Input variables

As mentioned earlier in section REF the different machine learning models that will be applied will apply two types of input data, time-series data in the form of longitudinal strain curves, and point-values in the form of peak systolic global longitudinal strain and patient EF.

5.2.1 Peak values

As mentioned in section REFERENCE EF values below 40-50% is regarded as unhealthy with regard to probability of heart failure. Keeping that in mind, one should note that the distribution of EF values among the patients shown in figure 5.3 is centered at approximately 40% with tails going as low as 8% and as high as 70%. Figure 5.4 shows the distribution of peak systolic

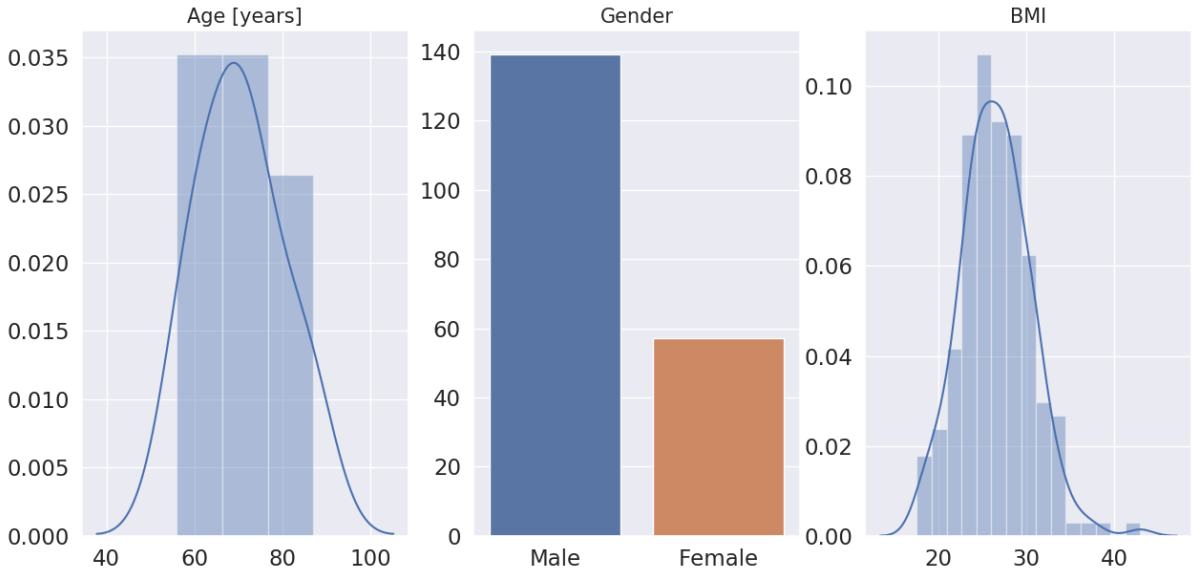


Figure 5.1: Distribution of age, gender and BMI.

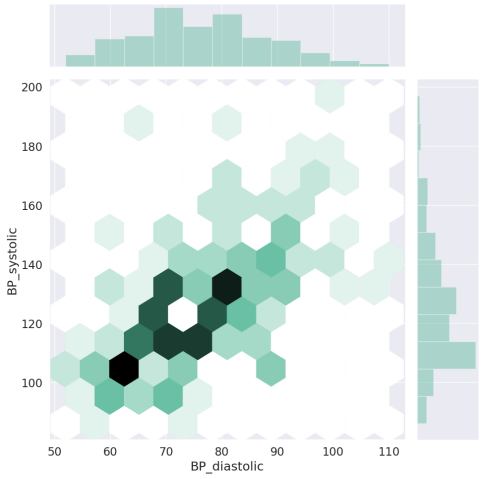


Figure 5.2: A joint distribtion plot of systolic and diastolic blood pressure of the patients.

GLS values, four the three different views. As evident from the figure, the values are centered around -12.5 with tails going as low as -29, and as high as -2.5.

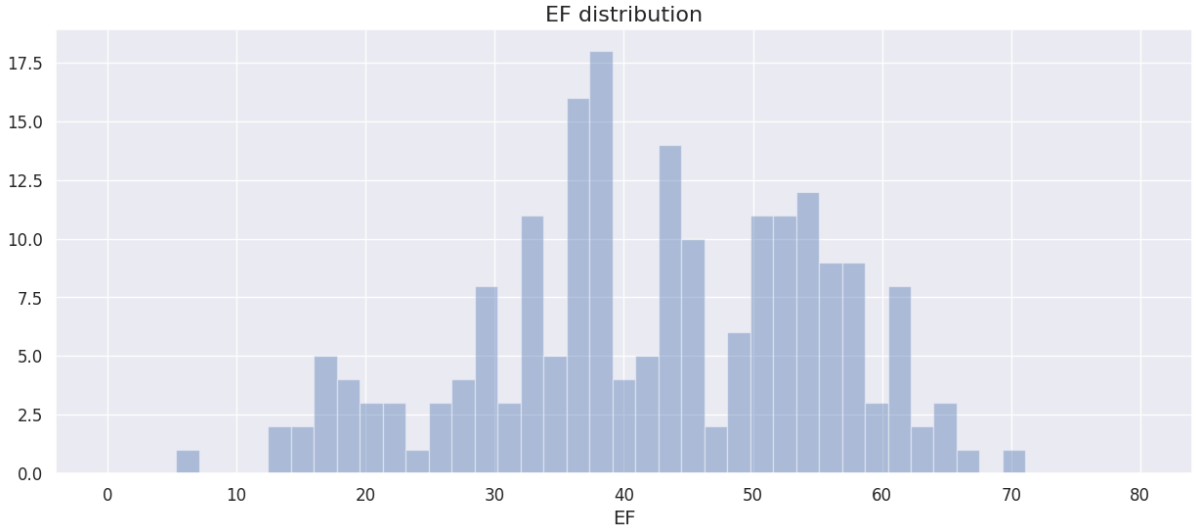


Figure 5.3: Distribution of patient EF values.

5.2.2 Strain curves

Figure 5.5 shows what a typical set of strain curves look like for a patient. Only the six regional strain curves, and the one global strain curve from the 4CH view have been included as they are fairly similar across the different views. Since the data from the different patients have been taken at different times, and possibly with different ultrasound machines factors such as number of samples per strain curve, and the frame rate of the particular ultrasound machine during an examination. Each strain curve has a standardized length of one heart cycle, due to this different curves have different number of samples. Figure 5.6 shows the distribution of frame rates, and number of samples among the total number of strain curves.

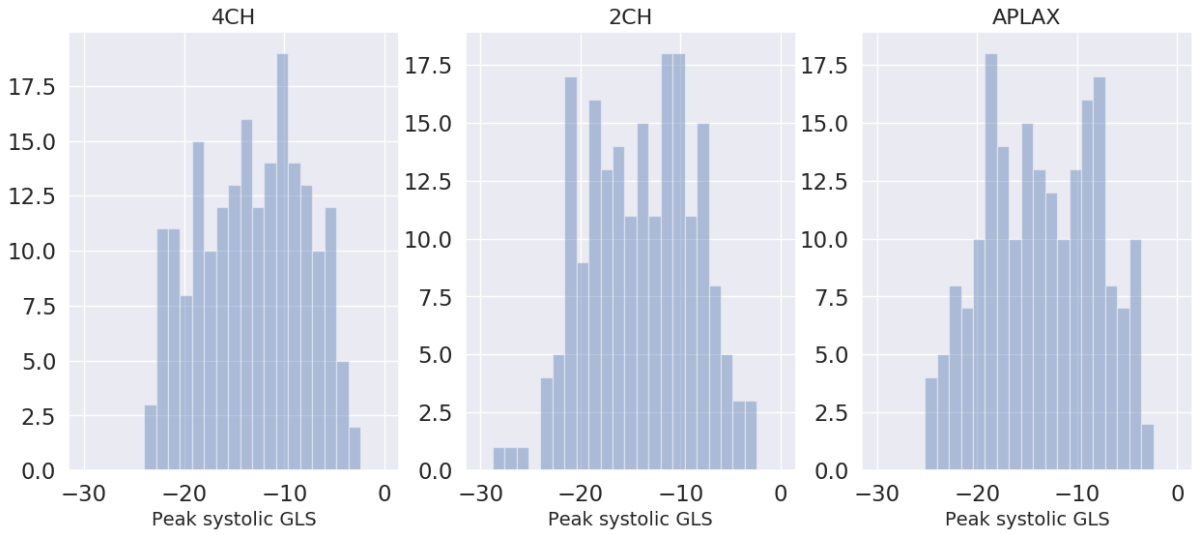


Figure 5.4: Distribution of peak systolic global longitudinal strain.

5.3 Target variables

Figure 5.7 shows the distribution of heart failure among patients (left), and the distribution of different indications (right). Since the dataset has approximately as many patients with a heart failure diagnosis as without, it can be considered balanced in that regard. With regard to the different patient diagnosises, their rate of occurance can is not uniform in this dataset. The control group of healthy individuals consists of 31 patients. The groups of patients with STEMI, and NSTEMI indications consist of 60 and 39 patients respectively. Finally, the group of patients with heart failure, but with a non-stemic indication (labelled OTHER in left barplot in figure 5.7) consists of 69 patients.

To illustrate the diagnostic power of EF, and peak strain values figure 5.8 shows the distribution of EF for patients with and without heart failure (left), and the distribution of EF for patients in the control group and the other patients (right). Figure 5.9 shows the distribution of peak systolic GLS values for patients with and without heart failure, and figure 5.10 shows the distribution of peak systolic GLS values for patients in the control group and the rest of the patients. From the samples used to produce the left plot in figure 5.8 and figure 5.9 it seems as though the heart failure patients are more separable with the EF values than with the GLS values. With regard to separability of patients with diagnoses and patients in the control group it seems as though the right plot in figure 5.8, and figure 5.10 follows the same distribution as the heart failure patients. However, it is hard to make an evaluation on this since the sample size of the control group is much smaller than the group of diagnosed patients.

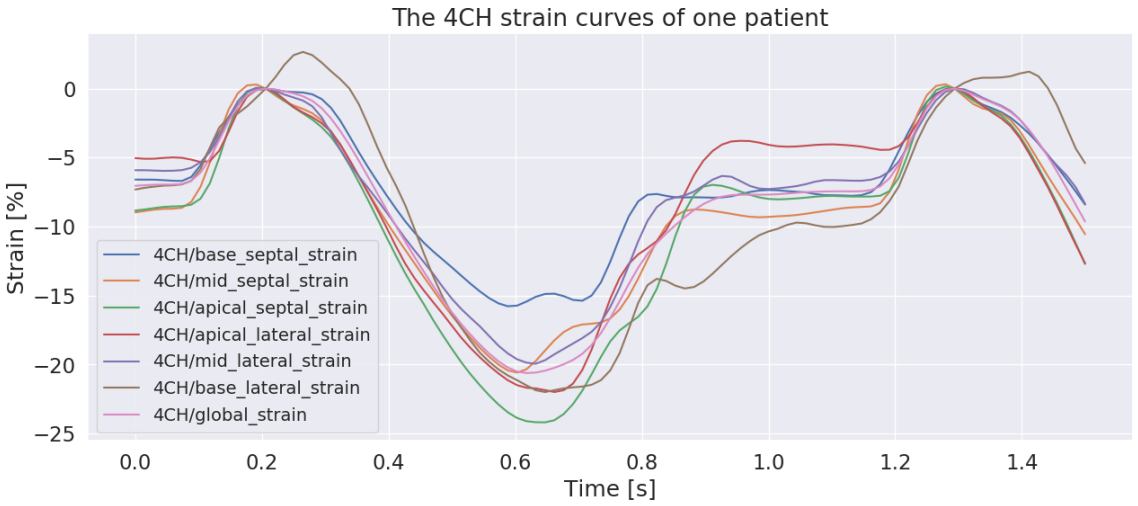


Figure 5.5: Plot of the global and regional longitudinal strain curves of one patient in the 4CH view.

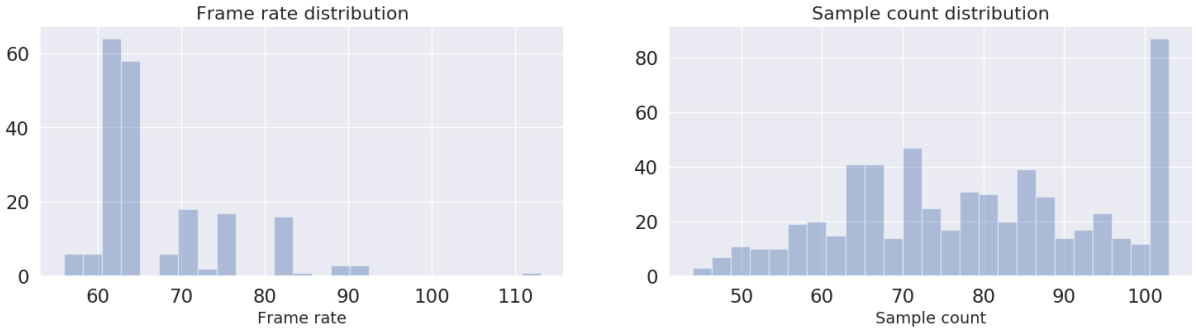


Figure 5.6: Distribution of the frame rate used in the ultrasound imaging used to obtain the strain curves (left), and sample count of the different strain curves (right).

Figure 5.11 shows the distribution of the different segment indications, for all the left ventricle segments of all the patients in the dataset. Since the occurance of indications other that "normal" and "hypokinetic" are very rare, the occurance axis has been used as logarithmic. The imbalance of segment-indication labels illustrated in figure 5.11 means that it will challenging for any statistical model to perform well in the classes with low occurance. To counteract this one can change the taxonomy of the labels such that the classification problem becomes binary with the labels *Normal* and *Not normal*. The dataset is then fairly evenly distributed with 1695 *Normal* labels and 1818 *Not normal* labels.

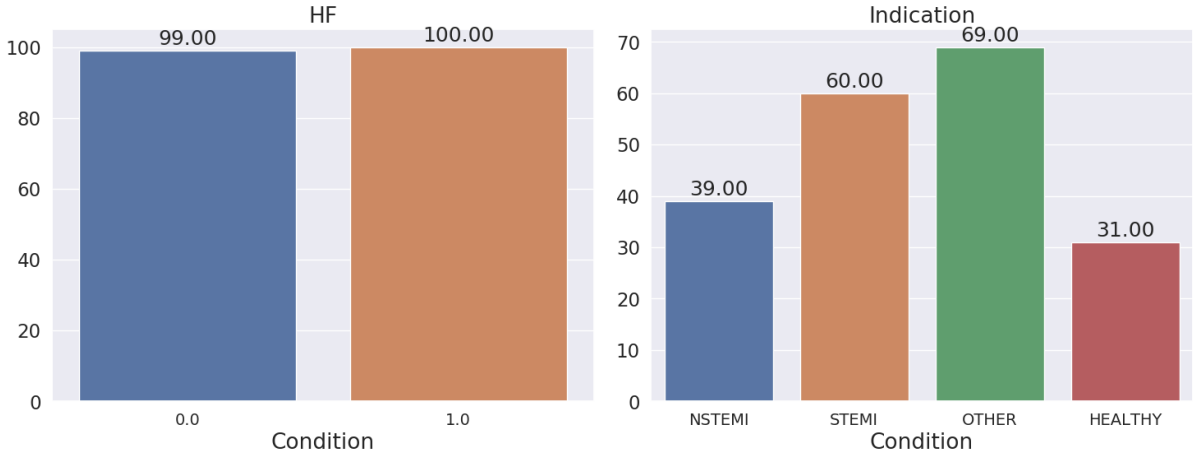


Figure 5.7: The distribution of heart failure and different indications within patients.

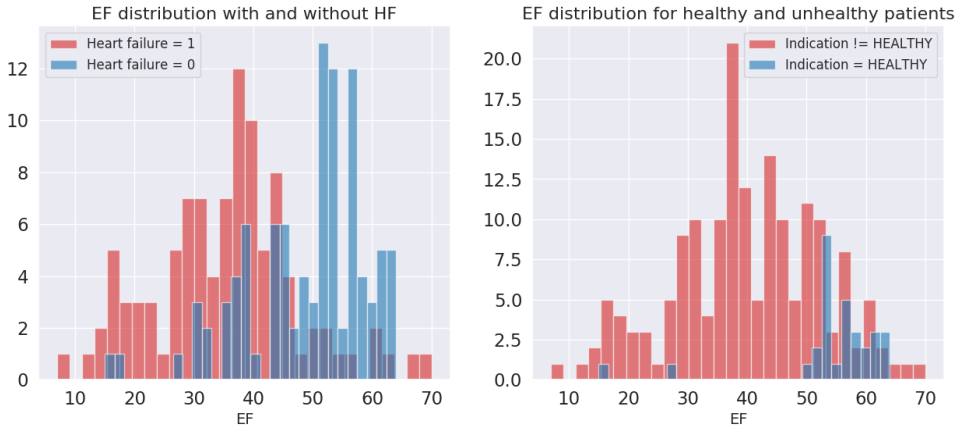


Figure 5.8: Distribution of EF for patients with and without heart failure (left), and distribution of EF for patients in the control group, and patients with a diagnosis.

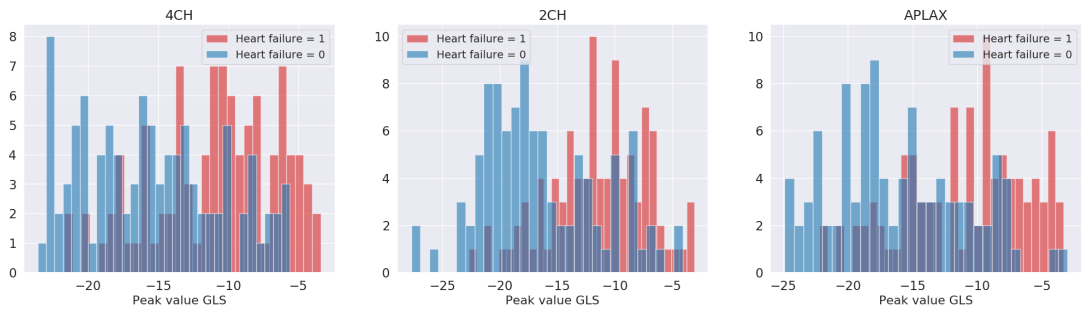


Figure 5.9: Distribution of GLS for patients with and without heart failure.

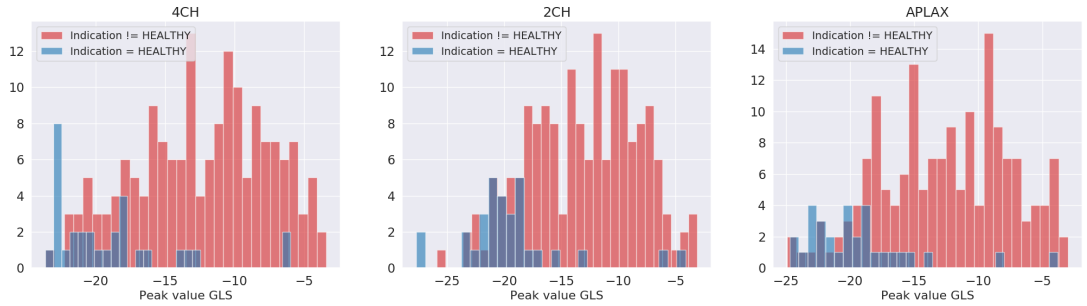


Figure 5.10: Distribution of GLS for patients in the healthy control group, and the other patients.

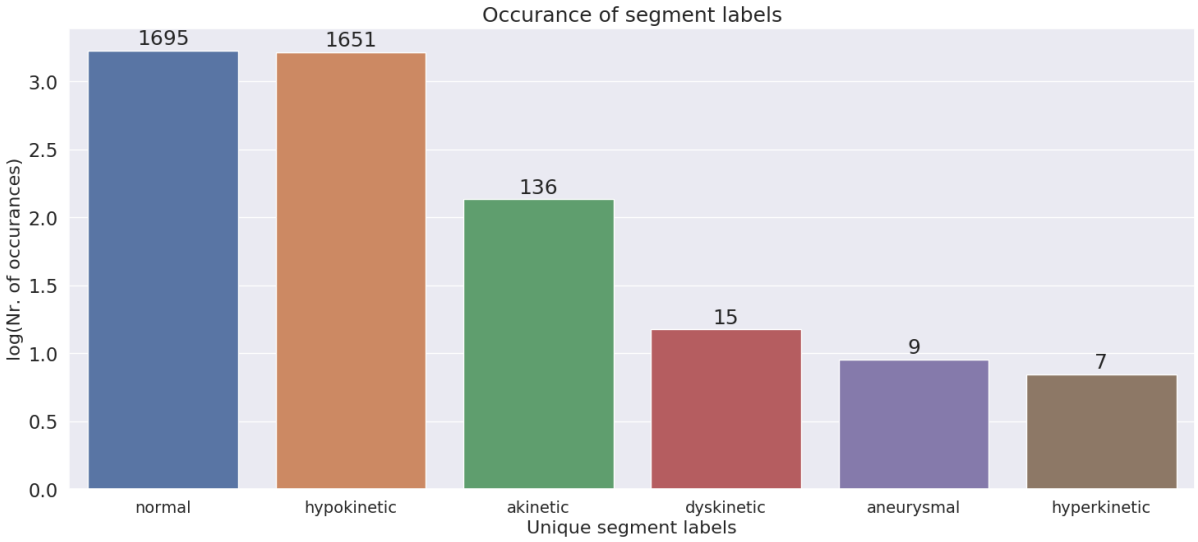


Figure 5.11: Distribution segment indication labels.

Chapter 6

Method

6.1 Models

This is the section where we detail the specific models used.

6.1.1 Time-series clustering

6.1.2 Peak-value clustering

6.1.3 Recurrent Neural Network

6.1.4 Supervised Peak-value Classifiers

6.2 Description of The Datasets

Since the different ML models detailed in chapter REFERENCE require different types of input data the, datasets have been divided into two main categories: The peak-value datasets and the time-series datasets.

6.2.1 Time-series Datasets

Nr	Input variables	Shape
1	Single RLS curves	$(3600, 1)$
2	RLS curves	$(200, 18)$
3	GLS curves	$(200, 3)$
4	Strain curves	$(200, 21)$

Table 6.1: Time-series datasets. The "Shape" parameter is indicates: (Number of objects in the dataset, Number of curves in each individual object). The curve length is not included in the shape parameter because it differs for different curves.

Table 6.1 shows the different time-series datasets that will be used. All the datasets except *Single RLS curves* will be used to predict whether or not the patient is diagnosed, and whether the patient has heart failure. Recall that the different diagnosises are described in section REFERENCE, and there occurance rate are illustrated in figure 5.7. *Single RLS curves* will be used to

predict the segment indications shown in figure **??** and described in section REFERENCE. The point of classifying individual segments of a patients left ventricle is that if a single segment is found to be *not normal*, this would also mean that the patient can be considered as *not healthy*. As mentioned in the description of table 6.1 the "Shape" parameter shows how many objects each dataset has, and how many curves are associated to each object. Since each ultrasound examination takes ultrasound inspections from three views (four chamber, two chamber, and APLAX chamber), each patient has three views to estimate a GLS curve from. Since each GLS curve, also can be divided into six RLS curves, there is a total of 21 strain curves per patient. Since each patient has 18 RLS curves, there are $18 \times 200 = 3600$ curves that make up dataset number 1. Both the RNN, and the TSC model are applied on the datasets listed in table 6.1,

6.2.2 Peak-value Datasets

Nr	Input variables	Shape
1	Single peak systolic RLS values	$(3600, 1)$
2	Peak systolic RLS values	$(200, 18)$
3	Peak systolic GLS values	$(200, 3)$
4	Peak systolic strain values	$(200, 21)$
5	Peak systolic RLS, and EF values	$(200, 19)$
6	Peak systolic GLS, and EF values	$(200, 4)$
7	Peak systolic strain, and EF values	$(200, 22)$

Table 6.2: Peak-value datasets. The "Shape" parameter is indicates: (Number of objects in the dataset, Number of dimensions of each individual object).

Table 6.2 shows the different peak-value datasets. All the datasets with exception of *Single peak systolic RLS values* will be used to predict the diagnosis of patients, and whether the patient has heart failure. *Single peak systolic RLS values* is also the only peak-value dataset that is not suited for clustering, since a minimum of two dimensions is required to cluster a point-value dataset. The reason that there are more peak-value datasets than there are time-series datasets, is that the peak-value version of three datasets in table 6.1 have been combined with EF to determine whether a combination of peak systolic strain, and EF can have a higher predictive power than strain alone.

6.3 Case Studies

The results will be presented in the form of three case studies. Each case study will focus on a single target variable, and aims to find which model group performs best at predicting the target variable in question. Recall that the three target variables that will be considered in this thesis are: Heart failure, patient diagnosis, and the indication of individual left ventricle segments. As mentioned earlier in the chapter, four model groups will be tested. The case studies will first deal with each model group individually, where variants of the models with different hypermarameters will be tested on the different datasets. Then, the best model within each model group will be used to compare the four model groups. Since the unsupervised models will be evaluated with regard to ARI in addition to the other evaluation metrics, the performance of these models will be evaluated on the different datasets individually. However, since the supervised models will not be evaluated with regard to ARI the performance of the

models on the different datasets are aggregated, and the best performing combination of dataset and model variation is taken further to compare the four models.

Chapter 7

Results

7.1 Case Study: Heart Failure

7.1.1 Peak-value Clustering

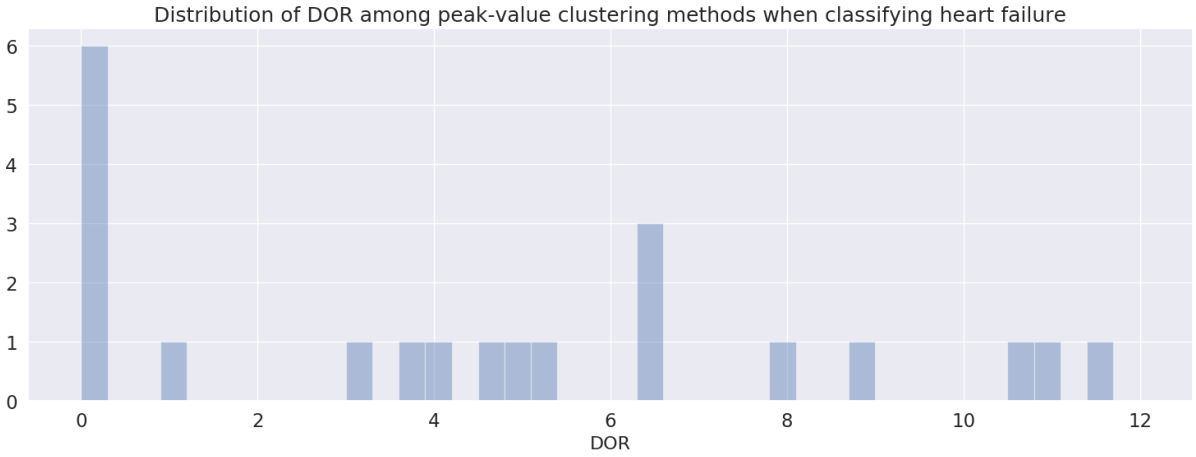


Figure 7.1: DOR distribution of PVC methods when classifying heart failure.

7.1.2 Time-series Clustering

7.1.3 Deep Neural Network

7.1.4 Peak-value Classifiers

7.1.5 Comparisons

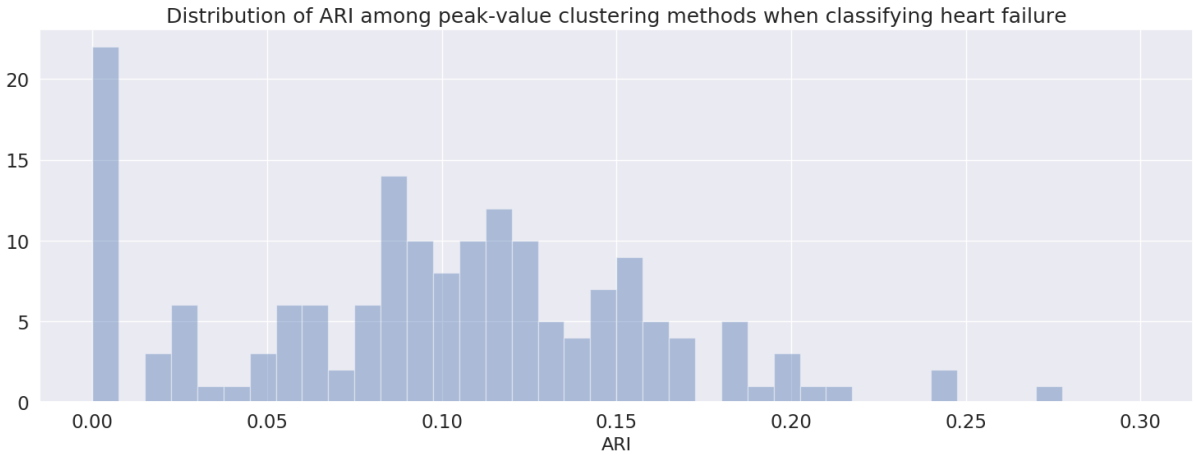


Figure 7.2: ARI distribution of PVC methods when classifying heart failure.

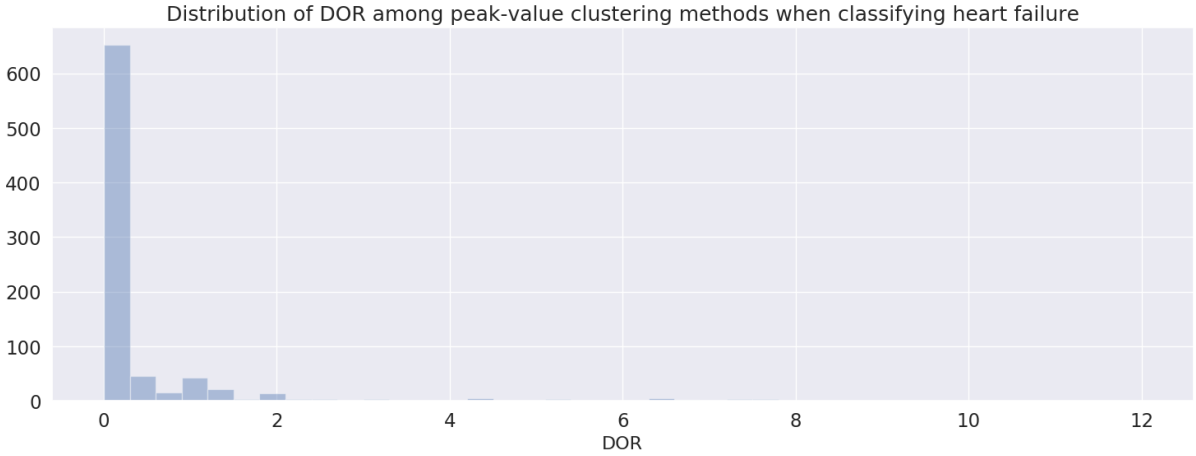


Figure 7.3: DOR distribution of TSC methods when classifying heart failure.

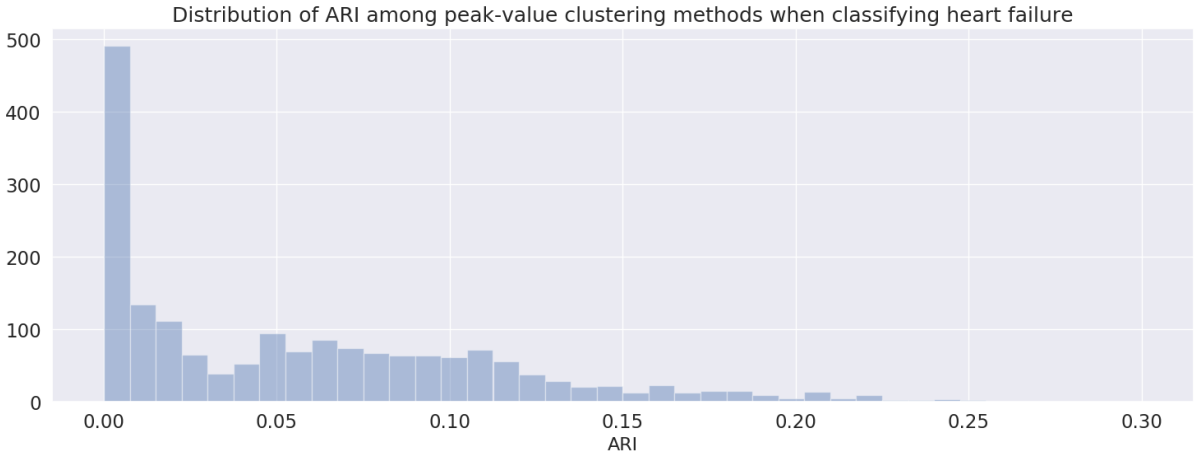


Figure 7.4: ARI distribution of TSC methods when classifying heart failure.

7.2 Case Study: Patient Diagnosis

7.2.1 Peak-value Clustering

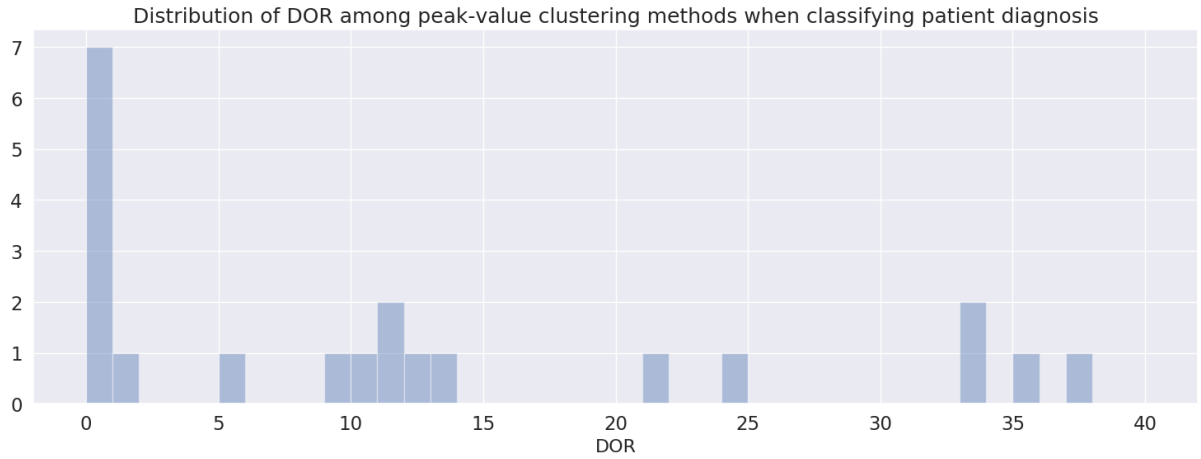


Figure 7.5: DOR distribution of PVC methods when classifying patient diagnoses.

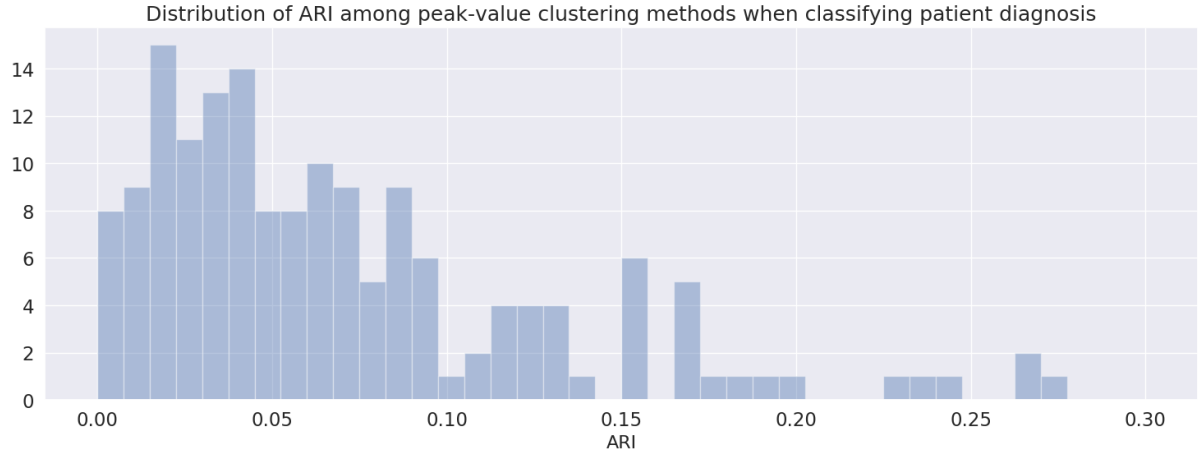


Figure 7.6: ARI distribution of PVC methods when classifying patient diagnoses.

7.2.2 Time-series Clustering

7.2.3 Deep Neural Network

7.2.4 Peak-value Classifiers

7.2.5 Comparisons

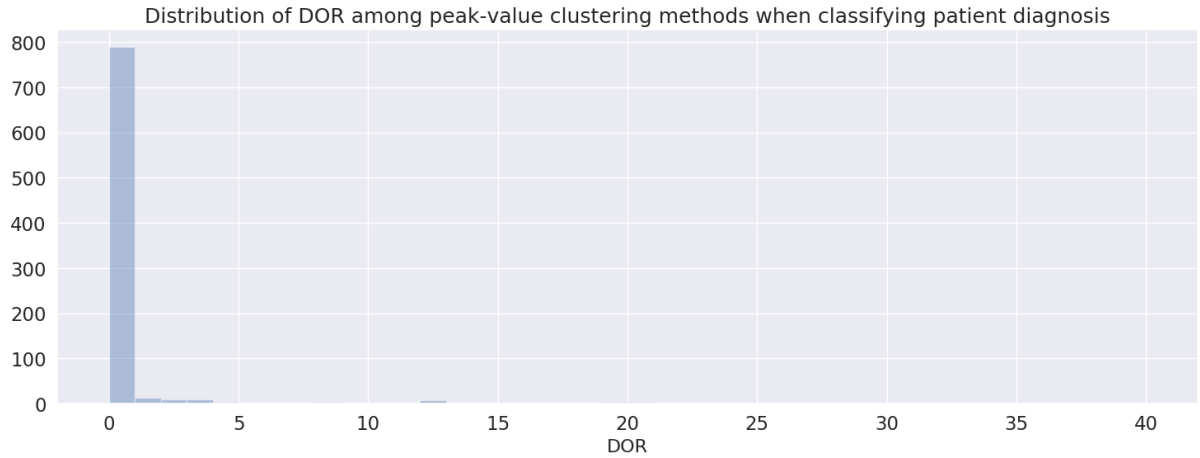


Figure 7.7: DOR distribution of TSC methods when classifying patient diagnoses.

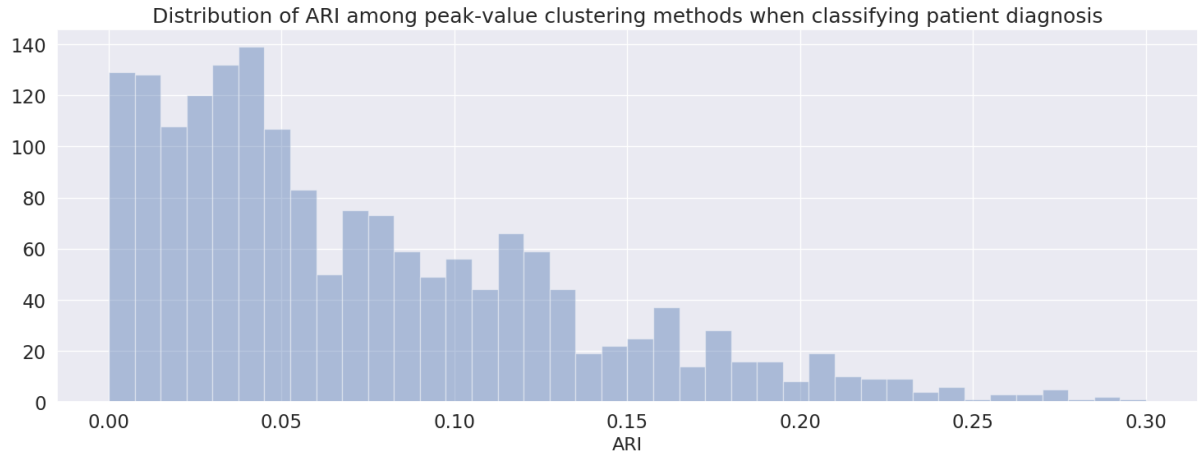


Figure 7.8: ARI distribution of TSC methods when classifying patient diagnoses.

7.3 Case Study: Segment Indication

7.3.1 Time-series Clustering

7.3.2 Deep Neural Network

7.3.3 Peak-value Classifiers

7.3.4 Comparisons

7.4 Chapter Summary

Chapter 8

Discussion

This is the discussion.

Chapter 9

Conclusion

This is the conclusion.

9.1 Future Work

This is the future work section.