

Auto-strain master thesis

Written by

Yohann Jacob Sandvik

Master thesis EMNEKODE

Supervised by

Lasse Løvstakken



Faculty of Information Technology
and Electrical Engineering

Department of Electronic Systems

Department of electronic systems
Faculty of Information Technology and Electrical Engineering
Norwegian University of Science and Technology
May 10, 2020

Abstract

This is the abstract.

Acknowledgements

These are my acknowledgements.

Contents

1	Introduction	4
1.1	Motivation	4
1.2	Objective	4
1.3	Structure of Thesis	4
2	Myocardial Imaging and Echocardiography	5
2.1	Basic Cardiology	5
2.2	Introduction to Echocardiography	5
2.3	Myocardial Strain	5
3	Machine Learning Theory	6
4	Review of The Literature	7
5	Data Exploration	8
5.1	Patient Meta-data	8
5.2	Input variables	9
5.2.1	Peak values	9
5.2.2	Strain curves	10
5.3	Target variables	11
6	Method	14
7	Results	15
8	Discussion	16
9	Conclusion	17
9.1	Future Work	17

List of Figures

5.1	Distribution of age, gender and BMI.	8
5.2	A joint distribution plot of systolic and diastolic blood pressure of the patients. . .	9
5.3	Distribution of patient EF values.	9
5.4	Distribution of peak systolic global longitudinal strain.	10
5.5	Plot of the global and regional longitudinal strain curves of one patient in the 4CH view.	10
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). . . .	11
5.7	The distribution of heart failure and different indications within patients. . . .	11
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. . .	12
5.9	Distribution of GLS for patients with and without heart failure.	12
5.10	Distribution of GLS for patients in the healthy control group, and the other patients.	12
5.11	Distribution segment indication labels.	13

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.

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 erroneous 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.

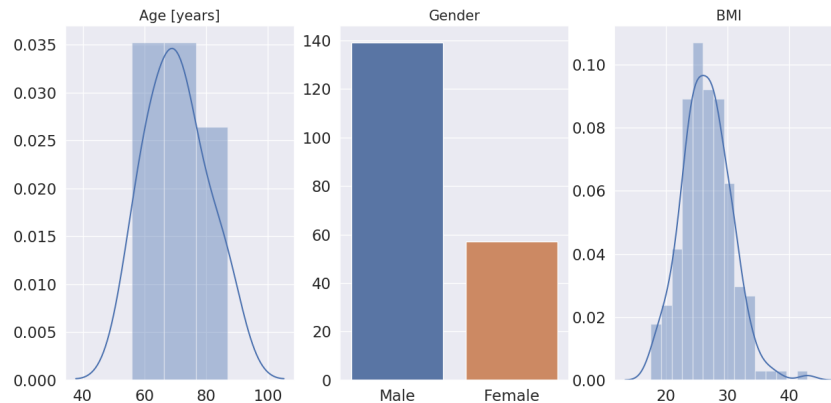


Figure 5.1: Distribution of age, gender and BMI.

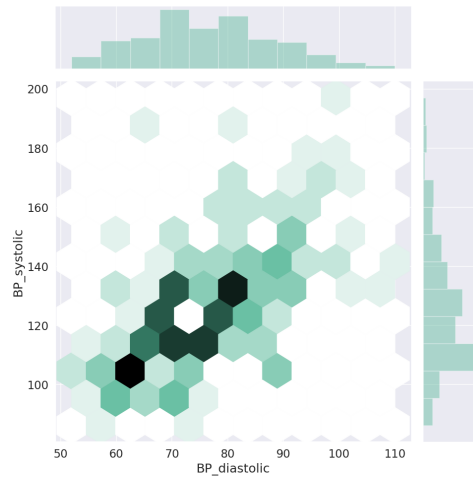


Figure 5.2: A joint distribution plot of systolic and diastolic blood pressure of the patients.

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 GLS values, for 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 .

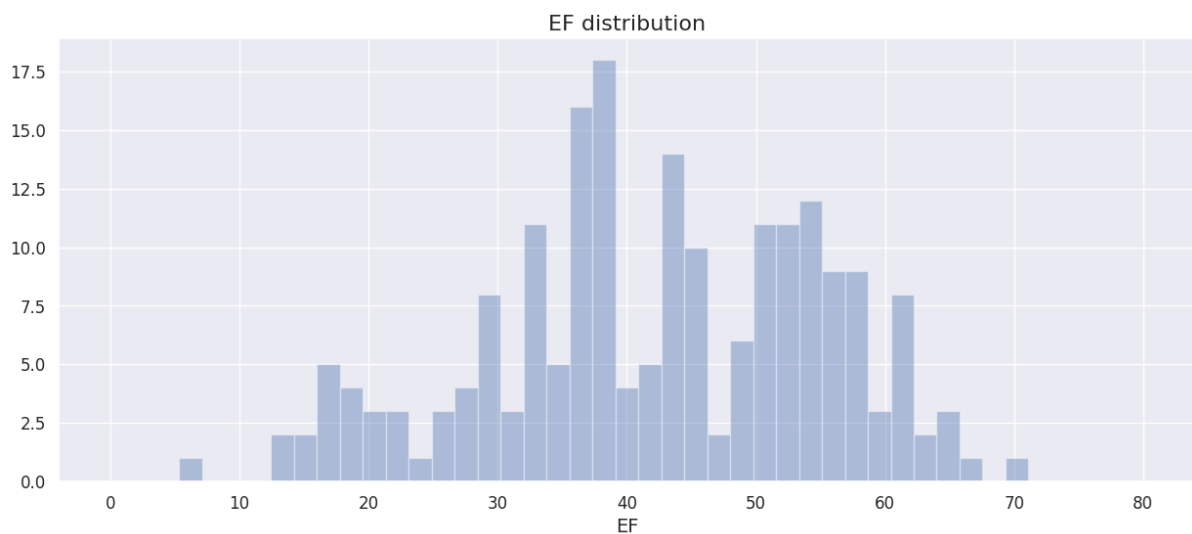


Figure 5.3: Distribution of patient EF values.

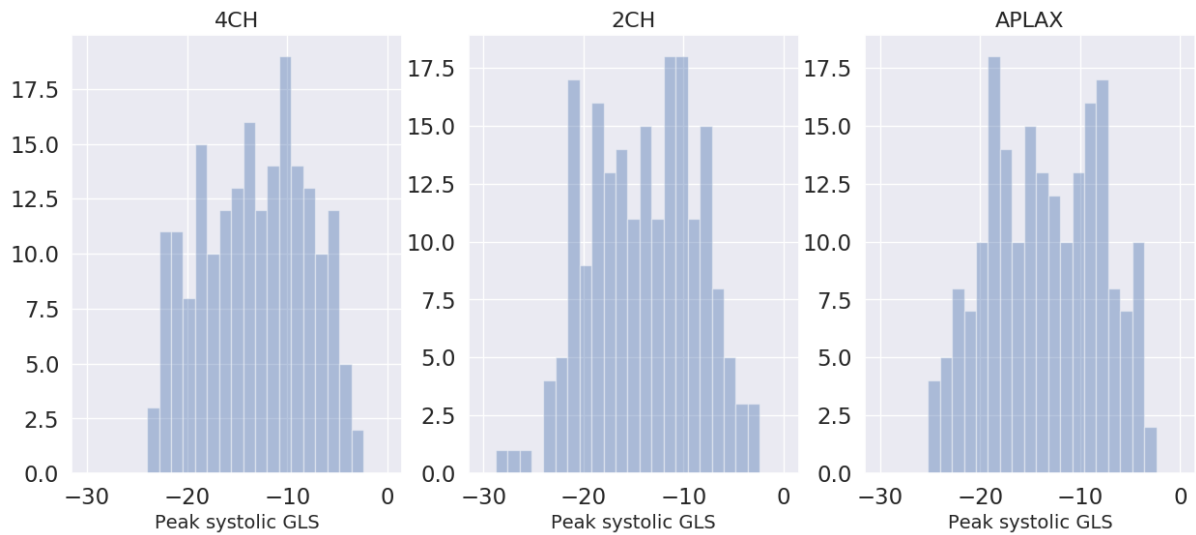


Figure 5.4: Distribution of peak systolic global longitudinal strain.

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.

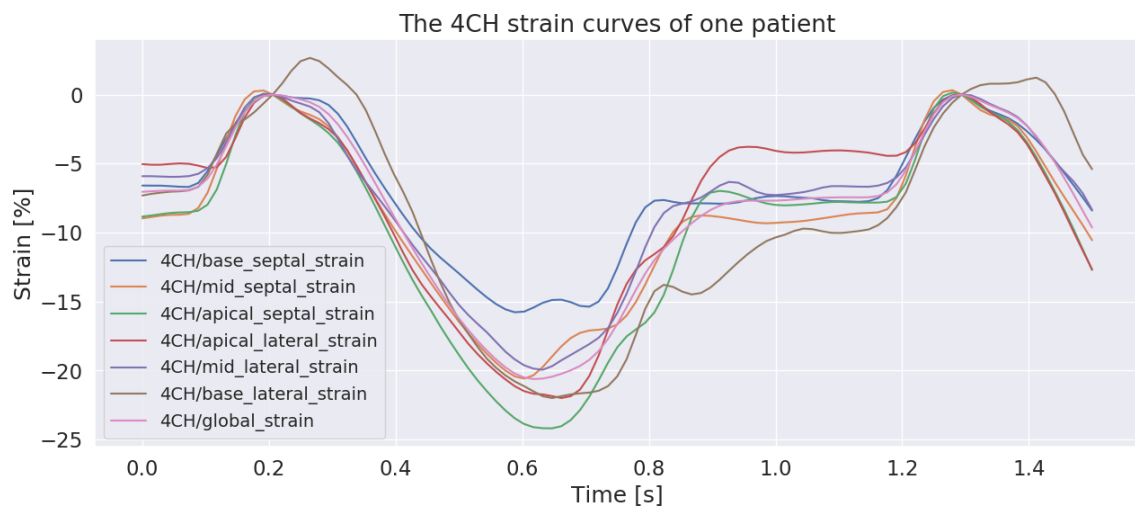


Figure 5.5: Plot of the global and regional longitudinal strain curves of one patient in the 4CH view.

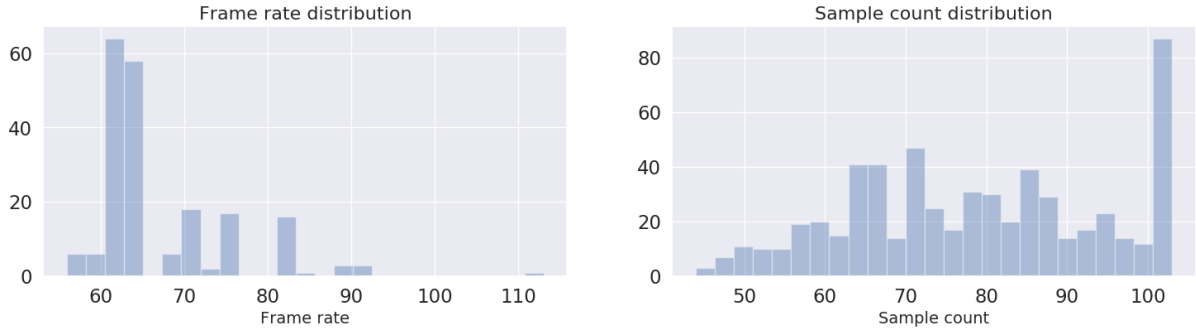


Figure 5.6: Distribution of the frame rate used in the ultrasound imaging used to obtain the strain curves (left), and sample count of the different strain curves (right).

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 indications, their occurrence is not uniform. 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.

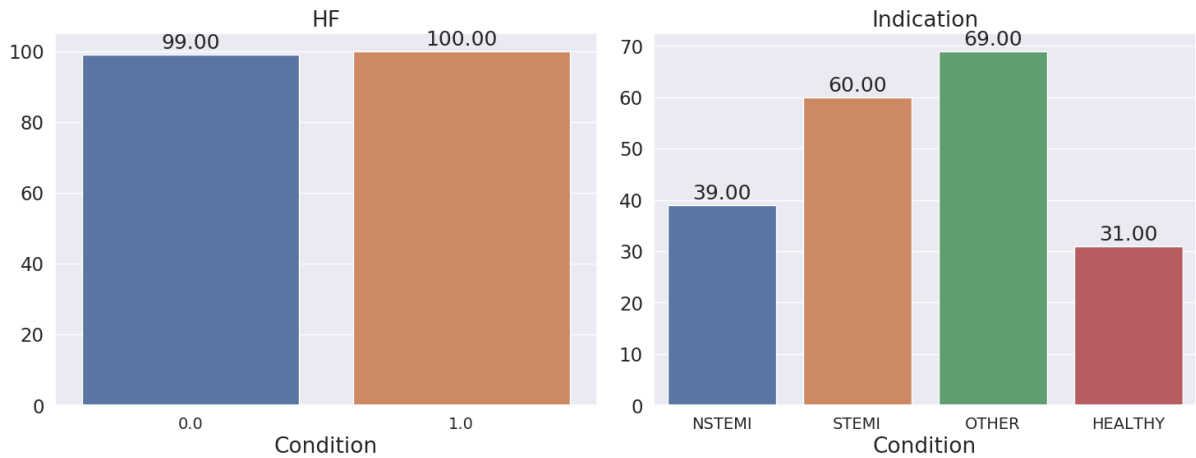


Figure 5.7: The distribution of heart failure and different indications within patients.

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 so much smaller than the group of diagnosed patients.

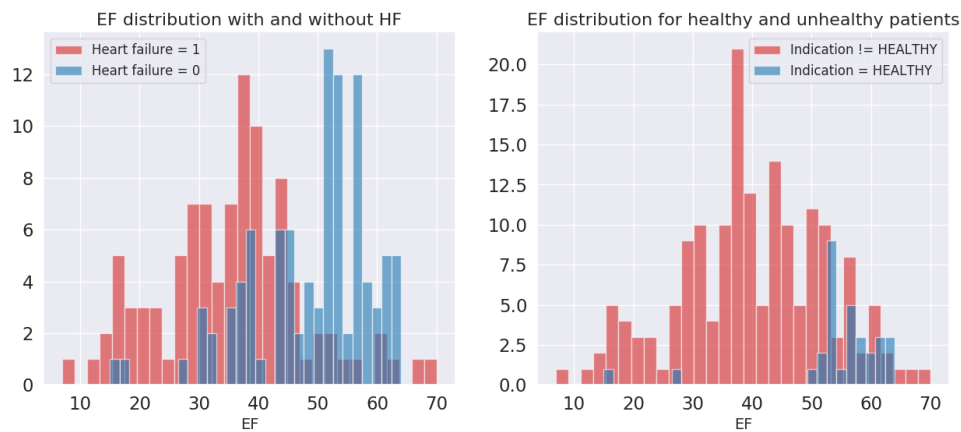


Figure 5.8: Distribution of EF for patients with and without heart failure (left), and distribution of EF for patients in the control group, and patients with a diagnosis.

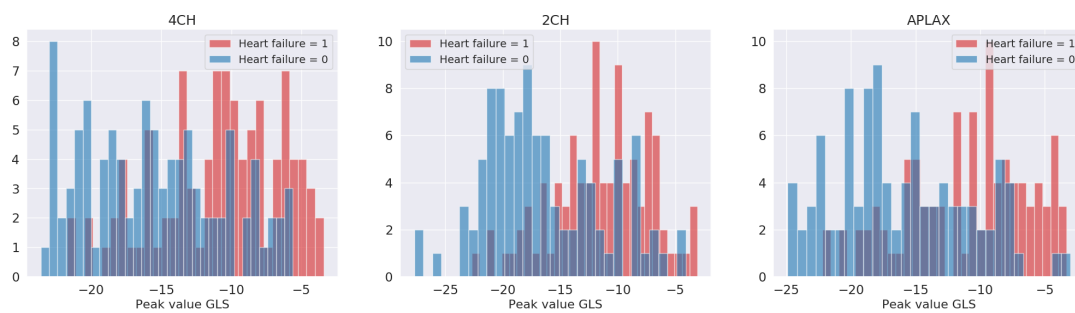


Figure 5.9: Distribution of GLS for patients with and without heart failure.

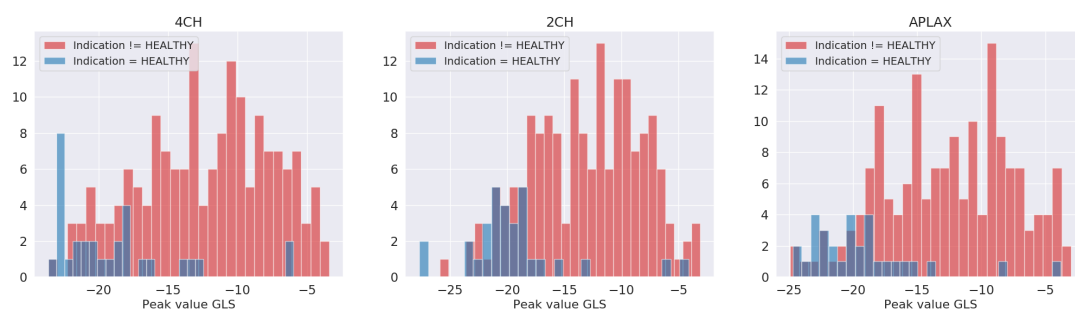


Figure 5.10: Distribution of GLS for patients in the healthy control group, and the other patients.

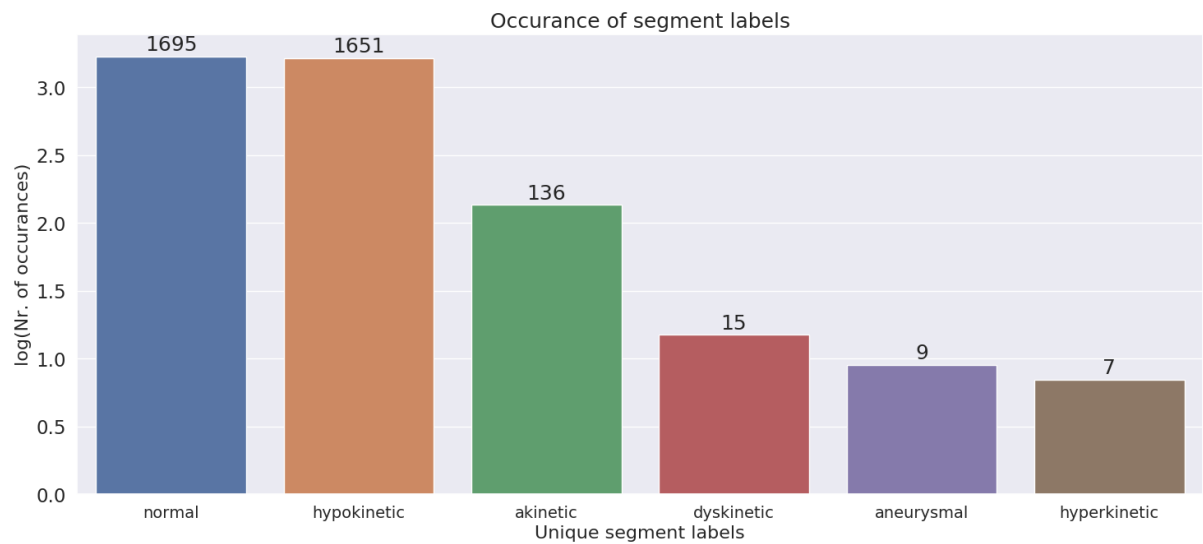


Figure 5.11: Distribution segment indication labels.

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 occurrence of indications other than "normal" and "hypokinetic" are very rare, the occurrence axis has been used as logarithmic.

Chapter 6

Method

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

Chapter 7

Results

Chapter 8

Discussion

This is the discussion.

Chapter 9

Conclusion

This is the conclusion.

9.1 Future Work

This is the future work section.