AALBORG UNIVERSITY
STUDENT REPORT

Detection and classification of heart opening snaps

MSc Thesis, 4th semester Masters
Biomedical Engineering & Informatics - Spring 2019

Project group: 19gr10412
Simon Bruun and Oliver Thomsen Damsgaard

Aalborg University
School of Medicine and Health

AALBORG UNIVERSITY
STUDENT REPORT

4th Semester (ST10), Master Thesis

School of Medicine and Health

Biomedical Engineering & Informatics

Fredrik Bajers Vej 7A

9220 Aalborg

Project period:
Spring 2019
01/02/2019 - 06/06/2019

Project group:
19gr10412

Collaborators:

Simon Bruun

Oliver Thomsen Damsgaard

Supervisor:
Samuel Schmidt

Pages: 46
Appendixes: 0
Completed: 06/06/2019

***Abstract* -** Cardiovascular diseases (CVD) are the primary cause of death around the world, but the methods for detection still rely heavily on subjective observations during auscultation followed by, in some cases, invasive examinations. New methods based on neural network and classifiers for automatic detection of heart disorders through phonocardiography (PCG) are being tested to overcome the subjective classifications within current methods. The PCG can be used to represent the hearts state, as the mechanic nature of CVD's result in unique abnormal heart sounds. Opening snaps (OS) followed by murmurs are caused by mitral stenosis, where the changed mechanical properties of the leaflets cause a snapping sound followed by a murmur due to blood turbulence. This study examines the cause of OS without an accompanying murmur, to find if this relates to calcification in the heart. This study implements parallel Fully Convolutional Networks (FCN) coupled with Long Short-Term Memory (LSTM) neural networks followed by a support vector machine (SVM) classifier to determine if an OS is present. Three networks will operate on either a filtered signal, Mel Frequency Cepstral Coefficients (MFCC) or Discrete Wavelet Transforms (DWT), as the last mentioned features have been proved useful for sound classification. In contrary to other studies, analysis of the heart cycle will only be performed on the relevant area for the specific abnormal heart sound rather than the entire cycle. Our results show that this approach is useful with a best average accuracy of 92% and an area under curve of 0.9288. No significant results were found for the cause of an OS without accompanied murmurs for the factors examined in this study.

Detection and classification of heart opening snaps

Simon Bruun, Oliver Damsgaard
Supervisor: Samuel Schmidt
MSc Thesis, Aalborg University, 2019

Abstract - Cardiovascular diseases (CVD) are the primary cause of death around the world, but the methods for detection still rely heavily on subjective observations during auscultation followed by, in some cases, invasive examinations. New methods based on neural network and classifiers for automatic detection of heart disorders through phonocardiography (PCG) are being tested to overcome the subjective classifications within current methods. The PCG can be used to represent the hearts state, as the mechanic nature of CVD's result in unique abnormal heart sounds. Opening snaps (OS) followed by murmurs are caused by mitral stenosis, where the changed mechanical properties of the leaflets cause a snapping sound followed by a murmur due to blood turbulence. This study examines the cause of OS without an accompanying murmur, to find if this relates to calcification in the heart. This study implements parallel Fully Convolutional Networks (FCN) coupled with Long Short-Term Memory (LSTM) neural networks followed by a support vector machine (SVM) classifier to determine if an OS is present. Three networks will operate on either a filtered signal, Mel Frequency Cepstral Coefficients (MFCC) or Discrete Wavelet Transforms (DWT), as the last mentioned features have been proved useful for sound classification. In contrary to other studies, analysis of the heart cycle will only be performed on the relevant area for the specific abnormal heart sound rather than the entire cycle. Our results show that this approach is useful with a best average accuracy of 92% and an area under curve of 0.9288. No significant results were found for the cause of an OS without accompanied murmurs for the factors examined in this study.

1 Introduction

The opening and closing of heart valves and blood rushing through the heart produce sounds. These sounds are audible to the ear without aid, but stethoscopes are used by physicians when evaluating patients, to aid in listening to specific sounds. During a normal heart cycle two heart sounds are present. The first heart sound (S_1) is associated with the closing of the left and right AV valves (mitral and tricuspid valves) during the beginning of systole. Individually, the sounds of the closing mitral and tricuspid valves are denoted as M_1 and T_1 respectively. The second heart sound (S_2) is associated with the closing of the aortic and pulmonary valves as the ventricles begins to fill during diastole. The sounds are denoted A_2 and P_2. [1]

Besides the four heart sounds other abnormal sounds can be present, such as murmurs, rumbles, clicks and snaps. These abnormal heart sounds can be present with heart disease and can be classified into three general groups, relating to different complications within the mechanical function of or blood flow in the heart. Heart murmurs will usually stem from a flow related complication, where the blood is being pushed through a valve that has not opened or closed completely, or a narrowed blood vessel close to the heart. This causes turbulence in the blood flow, resulting in a longer lasting sound during auscultation. [1, 2]

Cardiovascular diseases that can produce these abnormal heart sounds are the number one cause of death on a global scale. [3, 4] Thus, many people are referred to screenings for suspected CVD based on the presence of CVD symptoms. Of all CVDs coronary artery disease (CAD) causing ischaemia of the heart, is the leading cause of death [3, 4, 5].

An interesting and predominant sound is the OS, which, when accompanied by diastolic murmurs, have been found to appear in patients with mitral stenosis. Here the sounds intensity correlates with the valve mobility, until the point where the valve becomes immobile. The OS occurs at the point of maxi-

mal mitral valve opening, where the decline and level of pressure in the left ventricle affects the time difference between the sound and A_2. An OS follows A_2 by an interval of 30 to 150 ms and can be measured as a high frequency signal. Stenosis of the tricuspid valve can also be able to produce an OS. [1]

Currently used diagnostic methods in the clinical assessment of CAD are expensive and invasive, and has shown a low diagnostic yield. [6] This presents a risk factor to the patients being unnecessarily tested. Resent studies utilizing a diagnostic method based on acoustic systems to detect CAD have reported diagnostic accuracies of 74% [7] and 82% [8], measured in area under the curve (AUC), when comparing to coronary CT angiography (CCTA) or invasive coronary angiography (ICA) as golden standard. This reveals an interesting topic of utilizing acoustic based systems to detect CVD.

Redlarski et al. [9] developed a system utilizing a Linear Predictive Coding algorithm for phonocardiography (PCG) segmentation and a Support-Vector Machine (SVM) for classification of 12 different abnormal heart sounds. This system achieved a 93% best average accuracy, which is the mean of specificity and sensitivity. [9] A study by Low et al. [10] aimed to design a convolutional neural network (CNN) and provide it with as raw a time-sequence signal of PCG as possible to classify periodic heart sounds. The use of a CNN on raw data overcomes the need to perform feature extraction and achieved a 75% accuracy. [10] These studies show that heart sounds can be classified with good accuracy.

In recent years the use of Artificial Neural Networks (ANN) have gained increased interest within the medical field. To analyse PCGs studies by Castro et. al [11] and Lai et. al [12] have achieved interesting results in detecting heart murmurs. Castro et. al achieved a sensitivity of 69.67% and a specificity 46.91%, while Lai et. al achieved 87% sensitivity, and 100% specificity. This shows ANNs to be a viable method for analysing PCGs.

Clifford et al. [13] states that proper detection and classification of valve pathologies like mitral and aortic stenosis still presents a challenge. The presence of mitral stenosis is associated with the occurrence of OS. [14, 15, 16] Development of an acoustic based system analysing recordings of heart sounds to detect and classify OSs could be an important method in clinical evaluations of patients with suspected mitral pathologies.

Thus, the projects aim is to answer the question: How can detection of heart opening snaps be automated, and what relation does this abnormal heart sound have to the diagnosis of the patient?

2 Methods and Materials

This study will design and implement a combination of NNs to use for detection and classification of OS in a group of subjects. The group will be consisting of subjects with and without OS. Subjects will be evaluated to either have OS or not, by a manual approach which follows criteria made based on available literature on the physiology and auscultation of OS.

Neural Networks

Convolutional Neural Networks (CNN) have been widely used for feature extraction. [17, 18, 19] A CNN functions by passing the input through a number of filters. With increasing number of filter layers the CNN can detect higher levels of abstract features where the output from one filter is passed into the next. The filter sizes determine how many inputs are used in calculating a single feature. This functions as a "scanning" process where features are calculated for each filter by the following: [19]

$$A(x) = \sigma(W * x + b) \tag{1}$$

, where $A(x)$ is the node output, σ is the activation function, W is a vector of all weights, x is input and b is bias. If the input to filters and between filters are all connected, the network is considered fully connected. CNNs have been widely used for analysing images, however can also be used to analyse sequential data. When analysing images the filters sizes are defined in 2D to create a matrix to sweep over the image to cal-

culate features. This enables the CNN to learn to recognize features of the image like orientations of edges and changes in colours. When analysing sequential data the filter size would be in 1D, because the data is only progressing in time. Still the CNN will be able to learn to recognize features of the input. [19]

A long short-term memory (LSTM) neural network is a type of recurrent neural network (RNN), expanding on the basic RNN architecture to overcome shortcomings of a vanishing gradient during backpropagation. A RNN function by passing information in a hidden node h, back to itself, thus keeping information from prior inputs to use in later calculations. This enables RNNs to handle varying sizes of input data, as well as working well with sequential data. The core strength of a RNN is the ability for it to store information, giving it the ability to find meaning in information progressing over time. [20, 21, 22]
However, the basic RNN suffers from short-term memory, where it will gradually forget information from earlier states as it processes more data. When updating node weights and biases of the network during backpropagation, the RNN will only be able to perform the update for the latest processed data, as it has forgotten what came earlier. The LSTM have been invented to overcome this issue by having a separate cell state C which function as a memory. This memory can be updated and used or ignored in later calculations. [20] The memory is a combination of four gates in the LSTM layer expressed as:

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} sigm \\ sigm \\ sigm \\ softsign \end{pmatrix} * W * \begin{pmatrix} h_t \\ h_{h-t} \end{pmatrix} + b \quad (2)$$

, where the vectors i, f and o are controlling the input, forget and output states. g is a vector used to modify the memory content in the cell state. $sigm$ and $softsign$ are activation functions for the gates. W is the vector of all node weights, learnt through backpropagation. h_t and h_{t-1} are the current and old cell output. b is bias. The update function for the cell state C_t is calculated as: [21]

$$C_t = f * C_{t-1} + i * g \quad (3)$$

The final output h_t of the LSTM layer is as follows: [21]

$$h_t = o * tanh(C_t) \quad (4)$$

Heart Sound Data

A total of 600 subjects were included in this study. Data was granted by Acarix A/S as a dataset containing information gathered from acoustic recordings, coronoary artery calcium score (CACS), coronary computed tomographic angiography (CCTA), invasive coronary angiography (ICA) and patient interviews and reviews of patient medical recordings. All patients in the dataset were referred for suspected CAD. [8, 23]
The CADScor®System designed by Acarix A/S, consists of a microphone that is fastened to the subject using adhesive patches, and the recordings were done at the fourth intercostal space left of the sternum. Subjects were in a supine position during recordings, where they were asked to hold their breath for eight seconds four times within the three minutes of recording. [24]
Recordings were done with a 8000Hz sampling rate, and were subjected to segmentation into systolic and diastolic parts. This enables alignment of S_2 within the subjects, leading to an easier segmentation of specific heart sounds in relation to S_2 later on. The number of heart cycles varied between 5 and 30 for the subjects, with an average of 16.8. [24]

For the labelling process both auscultation and PCGs were examined. Based on the information on OS and splitting of S_2 the following criteria was made for the labelling process of OS:

- The distance between onset of A_2 and a following sound must be greater than 30ms [1]
- A_2 and a following sound must be both visibly and audibly divided [1]
- A sound following A_2 must be no more than 150 ms from the onset of A_2 [1]

The criteria ensures that the sounds located are OS, and not splitting of S_2 or early S_3 as these sound occur close the the interval in which an OS can occur. [1]

Both authors individually went through the entire dataset labelling subjects to have an OS or no OS. Afterwards all subjects labelled with OS was re-evaluated and a final labelling was decided. Examples of the PCGs are shown in figure 1.

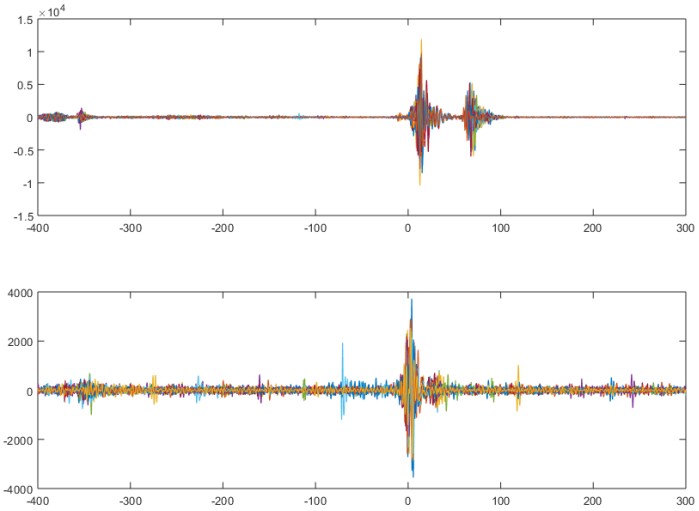

Fig. 1: Examples of PCGs analysed during the labelling. The figures show several superimposed PCGs from two different subjects. The topmost PCG is an example of an OS, where S_2 and the OS is easily differentiable. It is visible that the individual PCGs vary little in timing for the occurrence of the OS. The lower PCG is an example of S_2 splitting. Here the individual PCGs vary more over time as the timing of the split is affected by respiration.

System Design

In order to classify these OS automatically, a system was designed combining ideas from previous studies into a single, complex combination of different NN's, classifiers and threshold functions. The system consists of four steps; preprocessing, NN's, classifier and threshold function. The connection between these are shown in figure 2.

To the authors knowledge, a setup like the one used in this paper has not earlier been implemented for heart sound classification. The setup consists of branches of multiple parallel FCN-LSTM NN's with various inputs feeding probability outputs to a classifier followed by a threshold function determining the condition based on the percentage of OS detected for the subject. Additionally, the input is only based on the specific part of the heart cycle where OS can occur, rather than examining the entire cycle. This means that instead of letting the FCN-LSTM network decide what part of the cycle separates OS from normal recordings, it was only fed data in a window from the onset of S_2 and the following 1500 samples (187 ms), which is a change compared to other studies implementing NN.

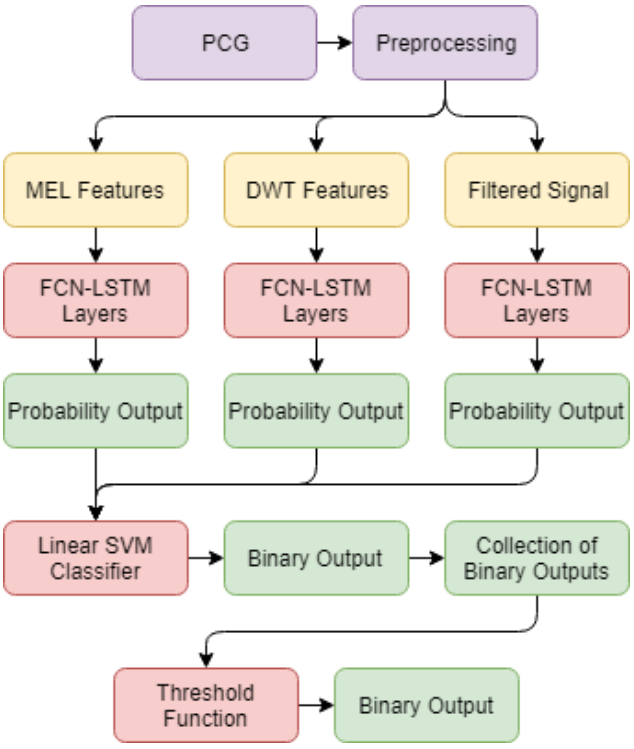

Fig. 2: Overview showing the final setup and connection through the system of preprocessing, NN's, classifier and threshold function.

Input data was filtered with a fourth order Butterworth

bandpass filter between 250 and 1200Hz to remove irrelevant information and noise, as the OS is heard up to 400Hz with murmurs varying between 30 and 400Hz. [25] This also means that the recordings comply with the Nyquist theorem, ensuring a correct representation of the chosen frequency spectrum. Further examination through fast Fourier transform (FFT) showed that there were no significant frequency activity above 600Hz.

The system was trained with approximately 1800 heart cycles from each group (OS and NOS), meaning a total of 110 subjects distributed close to evenly between the two groups. NN's were trained using approximately 1100 cycles while the classifier was trained using 700 samples. Test data consisted of an even number of new OS and NOS subjects, which where not used in the training process, with approximately 600 samples distributed evenly between the two groups, which represented 36 subjects in total. The exact numbers varied slightly depending on the random selection of subjects, as the number of heart cycles were not equal for every subject.

Neural Network Setup

In a study by Karim et al. [17], a LSTM and FCN were run in parallel achieving state-of-the-art results when analysing time-series data. The final network design of this study combined the FCN and LSTM networks to run in sequence to form a FCN-LSTM model. Three such architectures were then run in parallel to handle the Discrete Wavelet Transform (DWT) and Mel's Frequency Cepstral Coefficients (MFCC) features and the filtered signal as input data. DWT and first five MFCC were extracted from the signal using MATLAB's (R2019a, MathWorks Inc.) built in functions, and were chosen due to being useful for classification of sound. [26, 27]

The FCN block consisted of three filter layers, respectively with 80, 100 and 80 filters, with individually decided filter sizes for each NN branch. The filter layers were followed by a batch normalization layer (epsilon of 0.00001) and a ReLU activation layer. The LSTM block consisted of three LSTM layers with 256, 512 and 256 hidden nodes, respectively. The state activation function was softsign and gate activation function was set to sigmoid. To combat overfitting each LSTM layer was followed by a dropout layer with dropout probability set at 20%. Lastly, the output from the final LSTM layer gets passed to a fully connected layer and a regression layer to produce a continuous probability output. This was then passed to the Support Vector Machine (SVM) classifier. All inputs were scaled between 0 and 1 in order to improve performance. Every network was trained with a mini batch size of 100 samples.

Mel's Frequency Cepstral Coefficients Network

MFCC is a way of linearising the frequency range under 1000 Hz, to mimic the human ears ability to detect minor changes in pitch for lower frequencies. It works by splitting the signal into segments which are then subjected to a Fast Fourier Transform (FFT), and subjected to Mel-frequency scaling through the Mel filter bank. After this log of the power at Mel frequencies are taken, followed by a discrete cosine transform that transforms the signal into MFCC's. The first coefficients are good for expressing the general structure of the signal, while higher coefficients describe less important parts of the signal, such as noise or other small changes. The window length was 30 samples (approximately 4ms) with a 50% overlap to ensure a more detailed representation of the signal. [26, 27]

After initial testing, it was found that the most representative MFCC's were the first and third MFCC, whereas the other coefficients decreased performance of the network, leading to exclusion of these. The optimal performance for the MFCC network was found with an initial learn rate of 0.001 and a drop factor of 0.8 after every fourth epoch. The filter size was found to be the most optimal at 10 samples. Running 120 epochs resulted in an RMSE of 0.21 and a loss of 0.022.

Discrete Wavelet Transform Network

The DWT is a rather new alternative to the FFT when converting signals to the time-frequency domain, hereby creating a representation of the frequencies in time. Outputs consists of two different sequences describing the high and low frequencies contained in the signal that represent details and overall shape respectively. The DWT is obtained with a finite number of wavelet transforms over the signal obtained by moving a scalable window and calculating the spectrum for each step. This is done multiple times with different window sizes to create the time-frequency representation. [26, 27]

Optimizing the DWT network resulted in an initial learning rate of 0.0005 with a learn rate drop factor of 0.8 over 4 epochs. The optimal filter size was found to be 10 samples, which gave an RMSE of 0.19 and a loss of 0.018 after 120 epochs.

Signal Based Network

As previously described the signal was subjected to a bandpass filter between 250 and 1200Hz, after which it was scaled between 0 and 1 before being fed to the FCN-LSTM network. Optimizing the signal based NN led to an initial learning rate of 0.0005 with a drop factor of 0.05 every 10 epochs. The best filter size was found to be 10 samples and after 100 epochs the RMSE settled around 0.20 with a loss of 0.018.

Classifier

SVM classifiers are based on the principle of creating a hyperplane between the classes to separate them with highest possible precision. The hyperplane is created based on the support vectors, that are the data points closest to the hyperplane. These affect the position and shape of the plane, as the optimizer attempts to minimize the error while increasing the margin between samples and the hyperplane. [26]

Using the Classification Learner in MATLAB a Linear SVM classifier was chosen due to its accuracy and trained with the outputs from the three NN branches and set to have binary output classes. Training samples were 800 heart cycles distributed equally between subjects with and without OS, while 10 fold cross validation was implemented in order to decrease overfitting. The classifier verification accuracy with the chosen setup was 81.5%.

Threshold Function

At the end a threshold function was implemented, in order to determine if the subject was supposed to be classified with an OS. This function determines the outcome based on the percentage of snaps in the heart cycles for each subject, with a threshold of 50% for a subject to be classified with a OS. This threshold has been chosen in order to achieve the highest possible sensitivity while keeping a reasonable specificity.

Relation between OS and subject health

The dataset used for this project contains medical information on each subject, mainly in relation to cardiovascular pathologies. [23] This information is used to compare the groups of OS subjects with no-OS (NOS) subjects. The parameters chosen for comparing are based on clinical characteristics affecting the heart, like the Duke risk score (sex, age, diabetes, tobacco use, history of myocardial infarction, and symptoms of angina pectoris) and Morise risk score (sex, age, diabetes, tobacco use, symptoms of angina pectoris, hypercholesterolemia, hypertension, family history of CAD, obesity, and estrogen status). [8, 28, 29]

A case-control study was also conducted where subject's cardiac computerized tomography angiographies (CCTA) and echocardiographies were further examined for details on heart pathologies. The case-control study included a total of 83 subject divided in two groups of OS (n = 50) and NOS (n = 33) subjects. The groups were compared on clinical characteristics of the mitral and tricuspid valves and for the presence of atrial septum aneurysm.

Data of clinical characteristics were separated into categorical or non-categorical groups. Non-categorical data was tested for distribution using one-

sample Kolmogorov-Smirnov test. Data of Gaussian distribution was compared using two sample t-test. Non-Gaussian distributed data was compared using Mann-Whitney test (Wilcoxon rank sum). Categorical data was compared using Chi-squared test. All tests were performed with a 5% level of significance and tests for general difference in group means (two-sided). All statistic analyses was be performed using MATLAB.

3 Results

The manual labelling process of subjects in the dataset resulted in 77 subjects (12.83%) out of the total 600 subjects, were evaluated to have OS. Zero subjects were excluded. The two groups of OS subjects (n = 77) and no-OS (NOS) subjects (n = 523) were used for training and testing for the NNs and later comparisons were made between the two groups. The subjects were all over 40 years of age and nearly evenly distributed in sex. The details for the two groups are presented in table 3.

System Accuracy

Precision of the individual networks are as described in table 1 through AUC.

Network Features	AUC
MFCC	0.7388
DWT	0.8161
Signal	0.7433

Tab. 1: Accuracy of individual NN branches measured in AUC.

The classifier accuracy is described in figure 3 with a precision of 81.1% for single cycles, with approximately the same specificity and sensitivity.
The system accuracy can be seen in table 2, where the optimal threshold was found to be 45% providing 94.5% sensitivity, 89.5% specificity and 92% best average accuracy (BAC). To calculate AUC of the threshold function, the mean of predictions for each subject was set as the classifier output while the label for that subject was the supposed class. Thereby the overall systems AUC was found to be 0.9288 for the classification of a test group of 36 subjects mixed equally between heart cycles containing OS or no OS.

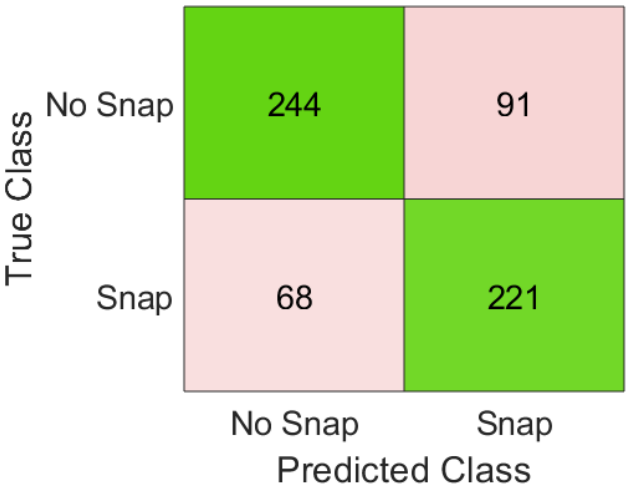

Fig. 3: Confusion matrix for the classifier accuracy. These results are for the classification of whether heart beats contain an OS or not, before the threshold function.

The accuracy for correctly and wrongly classified subjects for the overall system can be seen in figure 4.

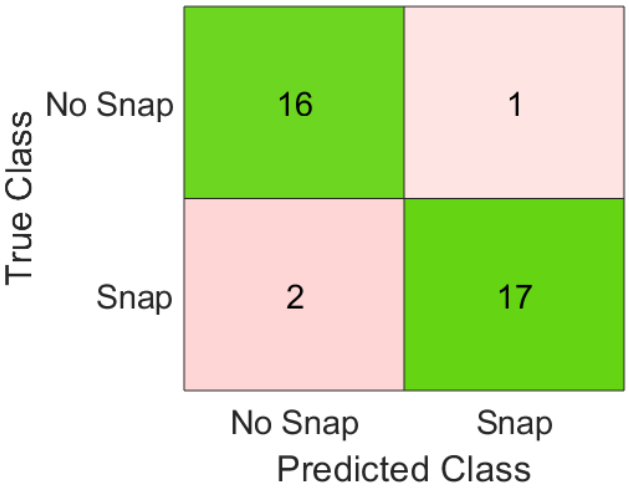

Fig. 4: Confusion matrix for the overall system. These results are the classification on whether subjects have OS or not.

Threshold	Sensitivity	Specificity	BAC
35%	94.5%	78.9%	86.7%
45%	94.5%	89.5%	92%
55%	70.6%	89.5%	80.1%

Tab. 2: System accuracy with various thresholds.

The Receiver Operator Curve (ROC) for the system performance are shown in figure 5.

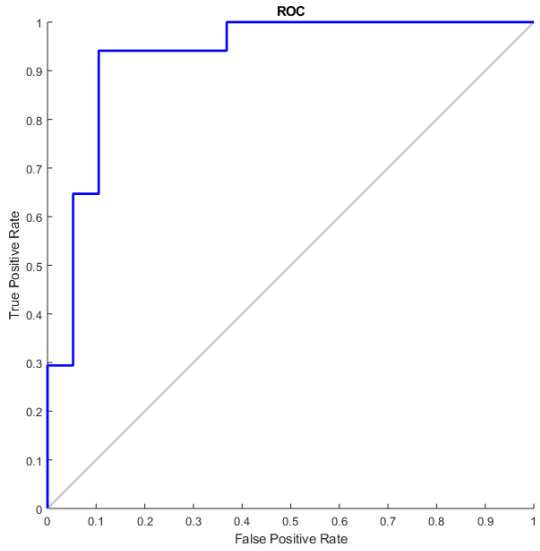

Fig. 5: ROC for the overall system performance.

Comparison of OS and NOS subjects

The Kolmogorov-Smirnov test proved all non-categorical characteristics to be of a non-Gaussian distribution, thus all non-categorical data was tested with a Mann-Whitney test (Wilcoxon rank sum). All categorical data were tested with Chi-squared test. The chosen characteristics are shown along with statistical results for the comparison in table 3. Non-Gaussian distributed values are shown with a standard deviation ($\pm$), while categorical values are reported with frequencies (percentages). Several of the chosen characteristics had missing entries in the dataset, because not every subject have undergone the same procedures and clinical tests. Characteristics which have missing entries are annotated and the number of missing entries are noted in the bottom of table 3. A significant difference were found for age between the groups ($p < 0.05$). No other significant differences were found.

Comparison of CCTA and Echocardiography

A selection of OS and NOS subjects, who had had CCTA and echocardiography performed, were examined for conditions and pathologies of the heart. The Kolmogorov-Smirnov test showed that all non-categorical values were from a non-Gaussian distribution. This data was tested with a Mann-Whitney test (Wilcoxon rank sum). Categorical data was tested with Chi-squared test. The results are shown in table 4. Non-Gaussian distributed values are shown with a standard deviation ($\pm$), while categorical values are reported with frequencies (percentages). Missing entries are annotated and noted at the bottom of table 4. A significant difference between the groups were found for subjects which had been diagnosed with mild mitral insufficiency ($p < 0.05$). No other significant differences were found.

	OS (n = 77)	NOS (n = 523)
Age (Years)	54.42 $\pm$ 9.54*	57.38 $\pm$ 8.75*
Sex		
- Female	39 (51%)	289 (55%)
- Male	38 (49%)	234 (45%)
Weight (kg)	80.75 $\pm$ 16.75	79.14 $\pm$ 14.21[1]
Height (cm)	173.75 $\pm$ 7.67	172.06 $\pm$ 8.98[2]
Pulse (BPM)	64.57 $\pm$ 12.05	65.45 $\pm$ 10.81[3]
Blood pressure (mmHg)		[4]
- Systolic	135.21 $\pm$ 19.72	139.06 $\pm$ 17.97
- Diastolic	82.83 $\pm$ 12.82	84.12 $\pm$ 10.90
Smoker		
- Active	15 (19%)	95 (18%)
- Former	25 (32%)	182 (35%)
- Never	37 (48%)	246 (47%)
Diabetes		
- Has diabetes	1 (1%)	27 (5%)
- No diabetes	76 (99%)	496 (95%)
CADScore	19.77 $\pm$ 9.18[5]	21.02 $\pm$ 9.85[5]
Agatston score	136.13 $\pm$ 335.61	118.70 $\pm$ 302.13
P-cholesterol (mmol/L)	5.37 $\pm$ 0.95[6]	5.38 $\pm$ 1.02[6]

Tab. 3: Data is missing for several categories: [1] Weight NOS: 3, [2] Height NOS: 1, [3] Pulse NOS: 2, [4] Systolic Blood Pressure NOS: 2, [5] CADScore OS: 1 NOS: 18, [6] P-cholesterol OS: 7 NOS: 34.
Significant differences between the two groups are indicated with * for $p < 0.05$ and ** for $p < 0.01$.

4 Discussion

The results show that it is possible to estimate OS precisely using a combination of NN's and classifiers.

	OS (n=50)	NOS (n=33)
Mitral Plague		
- No	49 (98%)	31 (94%)
- Yes	1 (2%)	2 (6%)
Mitral Valve Thickening		
- No	49 (98%)	32 (97%)
- Yes	1 (2%)	1 (3%)
MR		
- No	47 (94%)	3 (6%)
- Yes	32 (97%)	1 (3%)
Mitral insufficiency	[1]	[1]
- None	29 (66%)	13 (43%)
- Mild	12 (27%) *	17 (57%) *
- Moderate	3 (7%)	0
- Severe	0	0
Mitral stenosis	[2]	[2]
- None	45 (98%)	30 (100%)
- Mild	1 (2%)	0
- Moderate	0	0
- Severe	0	0
Mitral Restrictive	[3]	[3]
- No	44 (96%)	31 (100%)
- Yes	2 (4%)	0
Mitral Flow (m/s)	0.71 ± 0.17[4]	0.73 ± 0.21[4]
Mitral Dec (ms)	216.05 ± 52.18[5]	213.33 ± 60.95[5]
Mitral E (m/s)	0.10 ± 0.03[6]	0.11 ± 0.03[6]
Tricuspid insufficiency	[7]	[7]
- None	22 (61%)	7 (41%)
- Mild	13 (36%)	10 (59%)
- Moderate	1 (3%)	0
- Severe	0	0
Tricuspid stenosis	[8]	[8]
- None	35 (100%)	17 (100%)
- Mild	0	0
- Moderate	0	0
- Severe	0	0
Atrial Septum Aneurysm	[9]	[9]
- No	19 (95%)	13 (93%)
- Yes	1 (5%)	1 (7%)

Tab. 4: Data is missing for several categories: [1] Mitral insufficiency OS: 6 NOS: 3, [2] Mitral stenosis OS: 4 NOS: 3, [3] Mitral Restrictive OS: 4 NOS: 2, [4] Mitral Flow OS: 13 NOS: 12, [5] Mitral Dec OS: 13 NOS: 12, [6] Mitral E OS: 14 NOS: 11, [7] Tricuspid insufficiency OS: 14 NOS: 16, [8] Tricuspid Stenosis OS: 15 NOS: 16, [9] Atrial Septum Aneurysm OS: 30 NOS: 19.
Significant differences between the two groups are indicated with * for $p < 0.05$ and ** for $p < 0.01$.

This is done with a high accuracy when examining only the relevant areas for the specific heart sound rather than the entire signal. One of the leading causes for this high accuracy could potentially be the combination of three different NN's, as this expands the variables which can describe the OS, while also examining the different variables in specific ways, optimized for each feature.

The comparisons between OS subjects and NOS subjects and for the CCTA and echocardiographies did show very little relation between the occurrence of OS and subject health. Between OS and NOS subjects a relation was found only for the age of the subjects, where OS subjects are significantly ($p < 0.05$) younger than NOS subjects. It is known that some heart sounds are more present in younger subjects, as is also the case for S_3. [1]

Between CT-scans and echocardiographies for both groups a significant difference was found for the number of subjects diagnosed with a mild case of mitral insufficiency, where the OS group have significantly ($p < 0.05$) fewer cases than the NOS group, meaning the OS group is more healthy than the NOS group. This find is more controversial as pathologies of the mitral valve, specifically mitral stenosis, have been studied and found to be connected with the presence of OS. [14, 16] However, to indicate mitral stenosis diastolic murmurs must also be present following the OS. [30, 31] In this study only the OS was object for investigation which might explain why no relation was found between the OS group and CVDs. It raise the question if mitral stenosis is more related to diastolic murmurs than to the presence of OS.

The number of heart cycles for the subjects varied between 5 and 30 with an average of 16.8 which is a factor to consider if this is to be implemented for easier examination of subjects in the future. A higher number of heart cycles will most likely lead to a more confident prediction, which could be examined in further studies using this method to find the optimal number of heart cycles for accurate classification. Another valid point to bring for further studies could be the se-

lection of a specific area of the heart cycle, where the abnormal sound occurs, rather than examining the entire heart cycle.

Limitations

In this study PCGs where evaluated manually and later used for training the NNs, and a limitation is the level of human performance. Evaluation of the individual heart cycles done by qualified physicians could potentially improve the model robustness, as the current approach, despite being systematic and based on specific parameters, could potentially result in a few misclassified subjects as the authors were not educated within the field of auscultation. The parameters and classification was based on literature describing the OS and recordings found on websites meant for medical education within auscultation. Another limitation is the amount of data made available and used for training and testing, as NN's improve with more data.

The data used in this study were a subset from a larger dataset obtained for patients referred for suspected CAD to have coronary angiography performed. This could be the cause for why only few subjects have been found with OS or diastolic murmurs since these events are not related to CAD. This can also be an important factor for why so few cases of mitral stenosis were fund.

5 Conclusion

It can be concluded that finding OS through the use of NN's combined with a classifier is rather effective, with a high AUC of 0.9288, making it an effective tool for detecting OS. It can also be concluded that a high accuracy can be found when examining the specific area in relation to S_2 where an OS can occur.
A relation between the presence of OS in subjects and cardiovascular diseases cannot be drawn from this study.

Acknowledgement

We would like to thank Samuel Schmidt for excellent supervision and guidance. We also greatly appreciate the help of Simon Winther and Thomas Lyngaa for analysing a large number of CT-scans and echocardiographies. Lastly, we would like to thank Acarix A/S for granting us the data for this project.

References

[1] Valentin Fuster et al. *Hurst's The Heart*. 12th ed. McGraw-Hill Companies, Inc., 2008, p. 2473.

[2] Linda E. Pinsky and Joyce E. Wipf. *Learning and Teaching at the Bedside*. 1998.

[3] WHO. "Cardiovascular Diseases (CVDs)". In: *WHO* (2017), pp. 1–9.

[4] E Jougla et al. *Atlas on mortality in the European Union*. 2002.

[5] EM Flachs et al. *Sygdomsbyrden i Danmark*. Tech. rep. København: Statens Institut for Folkesundhed, Syddansk Universitet, 2015, p. 384.

[6] Eva Prescott et al. "Low diagnostic yield of non-invasive testing in patients with suspected coronary artery disease: results from a large unselected hospital-based sample". In: *European Heart Journal - Quality of Care and Clinical Outcomes* 4.4 (2017), pp. 301–308.

[7] Amgad N. Makaryus et al. "Utility of an advanced digital electronic stethoscope in the diagnosis of coronary artery disease compared with coronary computed tomographic angiography". In: *American Journal of Cardiology* 111.6 (2013), pp. 786–792.

[8] Simon Winther et al. "Diagnosing coronary artery disease by sound analysis from coronary stenosis induced turbulent blood flow : diagnostic performance in patients with stable angina pectoris". In: *The International Journal of Cardiovascular Imaging* 32.2 (2016), pp. 235–245.

[9] Grzegorz Redlarski, Dawid Gradolewski, and Aleksander Palkowski. "A system for heart sounds classification". In: *PLoS ONE* 9.11 (2014).

[10] Jia Xin Low and Keng Wah Choo. "Automatic Classification of Periodic Heart Sounds Using Convolutional Neural Network". In: *Journal of Electrical and Computer Engineering* 12.1 (2018), pp. 18–21.

[11] Ana Castro and Tiago T V Vinhoza. “Automatic Heart Sound Segmentation and Murmur Detection in Pediatric Phonocardiograms”. In: *IEEE Engineering in Medicine and Biology Society* (2014), pp. 2294–2297.

[12] Lillian S W Lai et al. “Computerized Automatic Diagnosis of Innocent and Pathologic Murmurs in Pediatrics: A Pilot Study”. In: *Congenital Heart Disease* 11.5 (2016).

[13] Gari D. Clifford et al. “Recent advances in heart sound analysis”. In: *Physiological Measurement* 38.8 (2017), E10–E25.

[14] E. Sejdić and E. Veledar. “An algorithm for classification of opening snaps and third heart sounds based on wavelet decomposition”. In: *IFMBE Proceedings* 37 (2011), pp. 404–407.

[15] Donna Boorse and Jill Stunkard. “Uncovering the Secret of Snaps, Rubs, and Clicks”. In: *Understanding Heart Sounds, Nursing94*. 1994.

[16] Tatiana Mularek-Kubzdela et al. “First heart sound and opening snap in patients with mitral valve disease. Phonocardiographic and pathomorphologic study”. In: *International Journal of Cardiology* 125.3 (2008), pp. 433–435.

[17] Fazle Karim et al. “LSTM Fully Convolutional Networks for Time Series Classification”. In: *IEEE Access* 6 (2017), pp. 1662–1669.

[18] Zhiguang Wang, Weizhong Yan, and Tim Oates. “Time series classification from scratch with deep neural networks: A strong baseline”. In: *Proceedings of the International Joint Conference on Neural Networks* 2017-May (2017), pp. 1578–1585.

[19] Christopher Olah. *Conv Nets: A Modular Perspective*. 2014.

[20] Christopher Olah. *Understanding LSTM networks*. 2015.

[21] Andrej Karpathy, Justin Johnson, and Li Fei-Fei. “Visualizing and Understanding Recurrent Networks”. In: *ICLR* (2015), pp. 1–12.

[22] Umberto Michelucci. *Applied Deep Learning*. 2018.

[23] Louise Nissen et al. “Danish study of Non-Invasive testing in Coronary Artery Disease (Dan-NICAD): study protocol for a randomised controlled trial”. In: *Trials* (2016), pp. 1–11.

[24] Simon Winther et al. “Diagnostic performance of an acoustic-based system for coronary artery disease risk stratification”. In: *Heart* 104.11 (2018), pp. 928–935.

[25] Aldo A. Luisada. *The Sounds of the Diseased Heart*. 1st Edit. Warren H. Green, 1973, pp. 93–100.

[26] Gui-young Son and Soonil Kwon. “Classification of Heart Sound Signal Using Multiple Features”. In: *Applied Sciences* (2018).

[27] Wenjie Fu, Xinghai Yang, and Yutai Wang. “Heart sound diagnosis based on DTW and MFCC”. In: *Proceedings - 2010 3rd International Congress on Image and Signal Processing, CISP 2010* 6.2 (2010), pp. 2920–2923.

[28] David B. Pryor et al. “Estimating the likelihood of significant coronary artery disease”. In: *The American Journal of Medicine* 75.5 (1983), pp. 771–780.

[29] Anthony P. Morise, W. John Haddad, and David Beckner. “Development and validation of a clinical score to estimate the probability of coronary artery disease in men and women presenting with suspected coronary disease”. In: *American Journal of Medicine* 102.4 (1997), pp. 350–356.

[30] Alexander M.D. Margolies and Charles C. M.D. Wolferth. “(CLAQUEMENT D ’ OUVERTURE DE LA”. In: *The American Heart Journal* (1930).

[31] Andrew C Oehler, Peter D Sullivan, and André Martin Mansoor. *Mitral Stenosis*. 2017.

- Worksheets -

Contents

1 Problem Analysis **1**
1.1 The Heart 1
1.2 Diagnostic Methods 5
1.3 Automatic Detection of Heart Sounds 6
1.4 Problem definition 7

2 Methods **9**
2.1 Data presentation 9
2.2 Neural Network Models 9
2.3 Neural Network Design 13
2.4 Manual Labelling and Human-Level Performance 17
2.5 Statistical Analysis 19

3 Implementation **20**
3.1 Neural Network Setup 20

4 Results **24**

1 | Problem Analysis

1.1 The Heart

The human heart is responsible for pumping blood around the body's circulatory system, supplying the body with oxygenated blood while moving the deoxygenated blood back to the lungs. At the same time it moves nutrients, waste products and toxins to the appropriate organs for further processing. This makes the heart one of the most vital organs of the human body, as the rest of the system depends on its functionality.

1.1.1 Anatomy and Physiology of the Heart

The heart has four chambers; two atriums at the superior part and two ventricles at the inferior part of the heart. Specific chambers are referred to as the left or right atrium or ventricle, also left or right side of the heart. Valves are located between the chambers in each side and at the base of the pulmonary artery and base of aorta. The valves are responsible for controlling the blood flow to only go in one direction. The tricuspid valve is located between the right atrium and ventricle in the right side of the heart. The bicuspid, or mitral valve, is located between the atrium and ventricle in the left side. The valves leading out of the heart at the base of the pulmonary artery and aorta, are called the pulmonary valve and aortic valve respectively figure 1.1. The pumping action of the heart is caused and controlled by electric impulses like the skeletal muscles of the body. The heart muscle differs from skeletal muscles in a number of ways, most noticeably in its metabolism as it is never rests. The heart is controlled autorhythmically by pacemaker cells. The heart is supplied with blood by the coronary circulation consisting of the the coronary arteries and cardiac veins. Damage to or narrowing of the arteries or veins of the coronary circulation is a common cause of heart disease and death. [1]

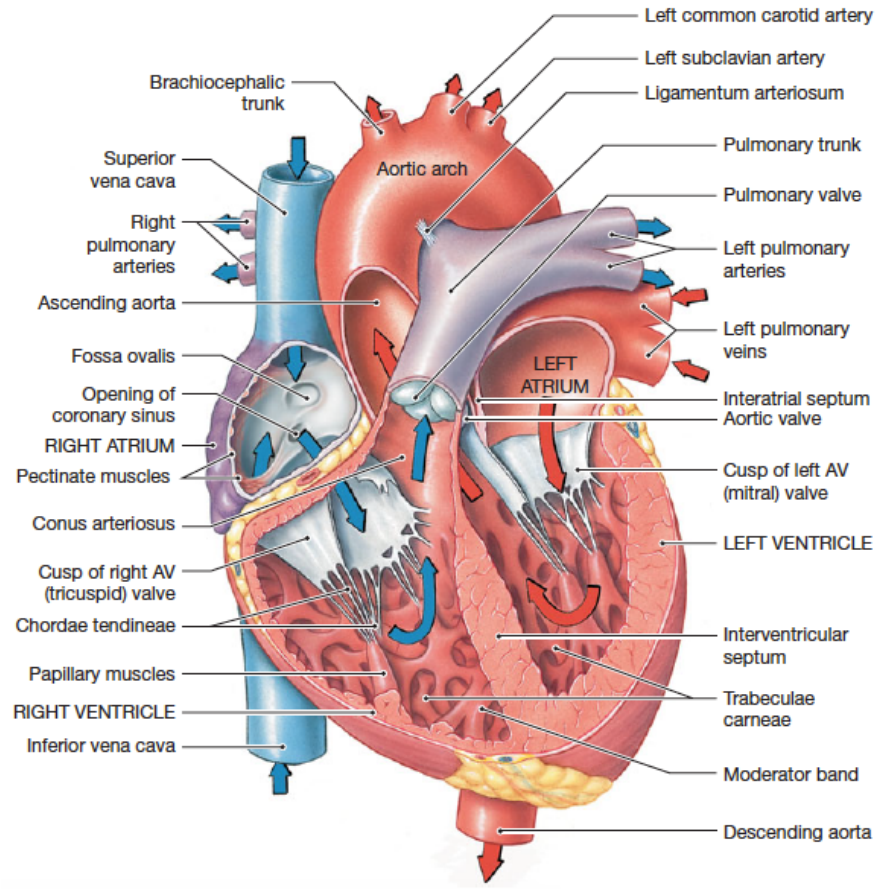

Figure 1.1: Frontal section of the heart showing chambers, valves, veins, arteries and the blood flow through the heart marked by arrows. [1]

A heart beat, the contraction of the heart muscle, is caused by electric impulses originating from the sinoatrial (SA) node at the top wall of the right atrium close to the superior vena cava opening. The SA node contains pacemaker cells which establish the heart rate. The electric impulse travels downwards towards the atrioventricular (AV) node at the top of the right ventricle between the right atrium and ventricle. Activation of the AV node causes the atriums to contract pumping blood from the atriums to the ventricles. The electric impulse travels down the AV bundle in the interventricular septum extending towards the apex and is divided between the bundles leading to the left and right ventricles. Reaching the apex the impulse spreads along the Purkinje fibers going towards the base of the heart, causing the ventricles to contract from apex and up, pushing blood into the aorta and pulmonary trunk. The spread of the electric potential causing through the heart can be recorded using electrocardiography (ECG). [1]

1.1.2 Cardiovascular Diseases

The most common cardiovascular disease in Denmark is atherosclerosis, which is a build up of plaque, a mix of fat and calcium, in the blood vessels. This blockage occurs in the arteries,

meaning that it affects the flow of oxygenated blood to specific organs, such as the brain, which can cause a stroke. If the blockages occur in the arteries supplying the heart with blood it is called coronary artery disease and can lead to ischaemia of the heart. The build up of plaque can result in heart attacks, which leaves the patient with a high risk of either dying or suffering from severe complications following the event. [2, 3, 4]

The second largest group of cardiovascular patients suffer from atrial fibrilation, which is an irregular activation of the atria, resulting in blood flowing back and forth between the chambers of the heart. This leaves the patient with an increased risk of other cardiovascular diseases, such as heart attacks or failure. Strokes are also common. [2, 3, 4]

Patients suffering from heart valve diseases is the third biggest group of CVD patients in Denmark. [2, 5] This category covers both valvular stenosis and insufficiency. The heart valves can stiffen over time due to calcification, leading to a smaller opening of the heart valves. This puts an additional stress on the heart, as it has to create a larger pressure over a prolonged time to move the same amount of blood as a healthy heart. [3]

In cases of valvular insufficiency, the valves will not create a functional seal, leaving an opening where blood can flow in the opposite direction. This will cause the blood to go back and forth, while the heart is put under more stress to support the necessary supply to the body. The combination of these factors will increase the risk of blood cloths and heart failure. [3]

1.1.3 Normal Heart Sounds

The opening and closing of heart valves and blood rushing through the heart produce sounds. These sounds are audible to the ear without aid, but stethoscopes are used by physicians when evaluating patients, to aid in listening to specific sounds. During a normal heart cycle two heart sounds are present. The first heart sound (S_1) is associated with the closing of the left and right AV valves (mitral and tricuspid valves) during the beginning of systole. Individually, the sounds of the closing mitral and tricuspid valves are denoted as M_1 and T_1 respectively.

The second heart sound (S_2) is associated with the closing of the aortic and pulmonary valves and the ventricles begins to fill during diastole. The sounds are denoted A_2 and P_2. A phonocardiogram (PCG) of a normal heart is shown at the bottom in figure 1.2. The first two heart sounds are the easy to hear, while the third and fourth heart sounds are much more faint as they are caused by blood flow and atrial contraction, rather than valve action. The third sound (S_3) is produced by blood flowing into the atriums and follows A_2 by an interval of 120 to 200 ms and is a low frequency event. The fourth (S_4) is also of low frequency and is caused by contraction of the atriums with an onset approximately 70 ms after the P wave in the ECG. In healthy subjects both the S_3 and S_4 sounds are rarely audible. [3]

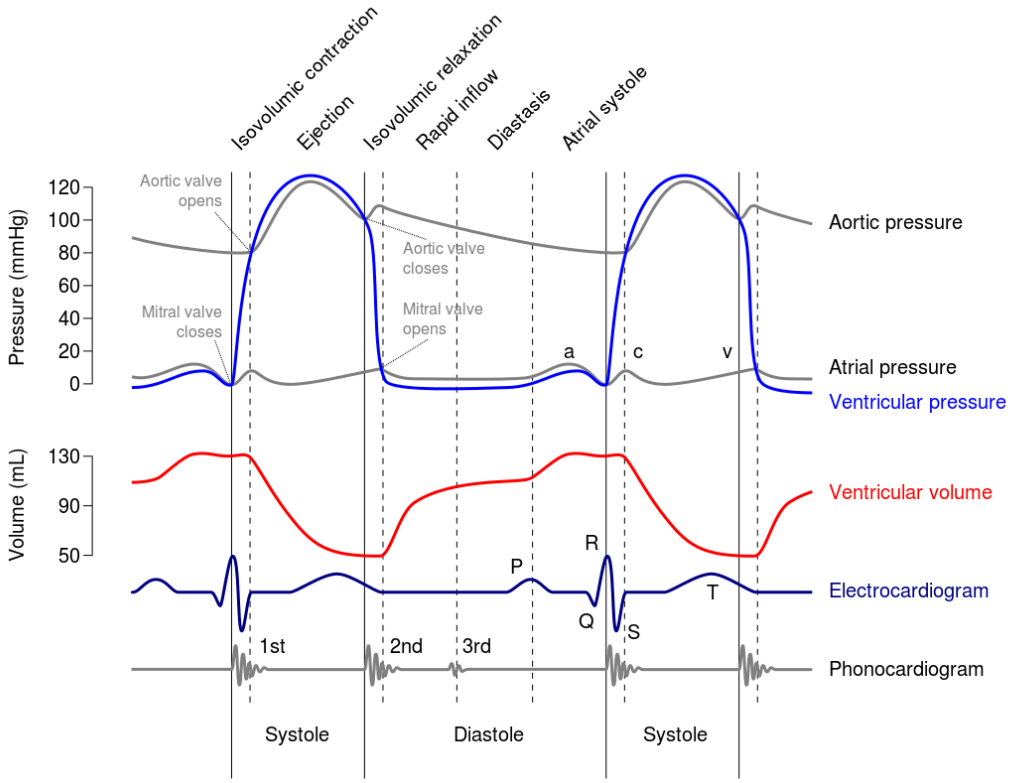

Figure 1.2: Wiggers diagram showing the pressure in the heart chambers, along with the ECG and PCG of the heart cycle. ©Wikimedia Commons User: DanielChangMD / CC-BY-SA-2.5

1.1.4 Abnormal Heart Sounds

Besides the four normal heart sounds other abnormal sounds can be present, such as murmurs, rumbles, clicks and snaps. These abnormal heart sounds can be present with heart disease and can be classified into three general groups, relating to different complications within the mechanical function of or bloodflow in the heart. Heart murmurs will usually stem from a flow related complication, where the blood is being pushed through a valve that has not opened or closed completely, or a narrowed blood vessel close to the heart. This will cause turbulence in the blood flow, resulting in an often longer lasting sound during auscultation. [3, 6]

Rubs are caused by pericardial rub, where the two layers of the pericardium rub against each other, or pleural rub, stemming from friction within the pleural cavity. These complications often stem from inflammations in the membranes that decreases the regular level of lubrication between the layers of the membranes. Clicks and snaps of the heart relates to the mechanical function of the heart, where clicks are associated with the closing of valves while snaps happen in relation to openings. [3, 6]

Opening snaps (OS) have been found to appear in patients with mitral stenosis, where the sounds

intensity correlates with the valve mobility, until the point where the valve becomes immobile. The OS occurs at the point of maximal mitral valve opening, where the decline and level of pressure in the left ventricle affects the time difference between the sound and A_2. An OS follows the onset of A_2 by an interval of 30 to 150 ms and can be measured as a very high frequency signal. Stenosis of the tricuspid valves can also be able to produce an OS. [3]

The characteristics of S_1 can in some cases also indicate possible complications. Examples of this can be found in cases of severe mitral regurgitation, causing an absent or attenuation of S_1. A delay and increased intensity of M_1 along with the loud OS is an indication of mitral stenosis, where severe cases with calcific fixation of the mitral valve will soften the sound of M_1 while the OS will become absent. Acute aortic regurgitation changes the intensity of or completely removes the M_1 sound during auscultation. [3]

Change in S_2 can also be an indicator of disease. The abnormal splitting of S_2 can be classified into different categories, where a short but audible split of S_2 is called narrow splitting, which relates to pulmonary hypertension. A wider interval between A_2 and P_2 is described as wide splitting, often caused by a delay in the right ventricle activation, but it can also relate to pulmonary hypertension or pulmonic stenosis. The last example of abnormal S_2 split is reversed splitting. This can be a result of left bundle branch block, where the heart activates from right to left during septal depolarization. [3]

Another useful characteristic during auscultation are the systolic ejection sounds, that indicates if there is an obstruction of the ventricular outflow or if the patient suffers from pulmonary hypertension. The valvular sounds are caused by deformed aortic or pulmonic valves, where the sudden deceleration of the blood results in vibrations of the entire system. Aortic root ejection sounds will often be a consequence of systemic arterial hypertension, while pulmonary root ejection sounds are mainly a result of a widened pulmonary artery. [3]

Listening for the four heart sounds and the abnormal sounds can be used as a mean to evaluate patients heart condition and possibly diagnose patients based on the sounds produced by the heart. [1]

1.2 Diagnostic Methods

Several methods have been developed to assess heart function and conditions. Methods used most frequently in Denmark are here described briefly. [4]

1.2.1 Invasive Diagnostic Methods

In cases of CAD with suspected blockages and flow related complications, a coronary angiography might be recommended. This method can provide a view of potentially weakened blood vessels around the heart, while also showing blockages caused by deposits of calcium or fat. This image will be created with the use of x-ray and a specialized catheter made for depositing contrast agent at specific locations in the blood vessels. During the coronary angiography, a

catheter will be inserted into an artery at the groin, threading it to the heart and injecting the contrast agent in coronary arteries while recording the area with x-ray. This means the patient risks complications such as puncture of the arteries, myocardial infarction or allergic reactions to the contrast dye. [3]

Coronary angiography have been further developed to avoid the need for catheters by use of computed tomography (CT) scans. CT angiography is less invasive as no catheter is used, but still relies on the use of a contrast agent to visualize blood vessels on the scan and still expose patients to radiation. [3]

1.2.2 Non-invasive Diagnostic Methods

The CT scan have also been used for detection of calcium deposits in the heart, especially in the coronary arteries, to determine coronary artery calcification (CAC). Contrary to CT angiography, coronary CT (CCT) for coronary calcium scores (CCS) use no contrast agent and is non-invasive. [7] The outcome of CCT is a score which determine severity of calcification based on the Agatston score ranging from 0 to 400, where the higher the score the more calcification. [8] Function elucidation of the heart is also possible by use of SPECT (single-photon emission computed tomography) or cardiac Magnetic Resonance imaging (MRI) to locate possible dysfunctions. MRI provides high resolution images and enables physicians to locate specific sites in the heart affected by disease. [4, 9]

Electrocardiography (ECG) of the heart beat is also used as an diagnostic method for detection of heart disease. ECG is easy to measure and has low cost, however has poor accuracy in detection of heart diseases. [10, 3]

Echocardiography is a non-invasive way to create images of the heart using ultrasound. As the method provides a view of the hearts structures it is used to check heart structures like valves and champers and to investigate blood flow. Echocardiography is also used in stress echocardiography, where the heart is put under stress with either physical activity or by means of pharmacological stress. This method is primarily used to determine blood supply to the heart and the heart strength in relation to heart valve stenosis. [3]

Of all the current diagnostic methods used in cardiology, one of the oldest and primary diagnostic methods for cardiovascular diseases is auscultation of the heart. In Denmark most patients with abnormal heart sounds are found randomly as a part of routine clinical check-ups. Auscultation can provide basic information of heart function and can be the basis for further examinations using above described methods. [11]

1.3 Automatic Detection of Heart Sounds

In an attempt of overcoming the problem with patients being sent in for further examination after auscultation there has been made attempts of automatic detection of heart sounds. The Computing in Cardiology Challenge of 2016 proved that many different combinations of fea-

tures and classification methods can yield sensitivities around 80% and higher, with a specificity above 80% for most of the entries in the challenge as well. The main aim for the participants were to classify the recordings in to three groups; normal, abnormal and unsure. [9, 12]

Automatic detection technologies have also found their way into the medicinal industry, where companies such as Acarix A/S have created a system capable of ruling out CAD with automatic detection of the abnormal sounds usually related to this disease. The CADScor®System aims to classify CAD by analysing phonocardiography recordings. A study by Winther et al. [13] achieved a 72% diagnostic accuracy with this system, measured by Area Under the Curve (AUC). The system functions by analysing PCG recordings made with a digital stethoscope, and calculating four measures based on both frequency and amplitude of the heart sounds, which is then combined to find a final CAD-score. [13]

Murmur-detection have been a focus for studies trying to create automatic detection methods for abnormal heart sounds as well. This has been done both to detect murmurs in patients, and to distinguish innocent and pathologic murmurs in children. Here the methods vary from neural networks (NN) to classification and signal analysis algorithms to determine whether there is a murmur, and if the cause is non-pathologic or caused by a cardiovascular disease. [14, 15]

Automatic detection have also been implemented to find multiple heart sounds like opening snaps and ejection clicks simultaneously with the use of support vector machines (SVM) as in a study by Redlarski et al. [16], while other studies focus on differentiating in classifying if a heart sound is S_3 or an OS. [17] This is relevant, as an OS accompanied by murmurs and presence of mitral stenosis is widely believed to be connected. [18, 3, 17].

1.4 Problem definition

In Denmark most patients with heart diseases are identified at routine clinical check-ups. In case the general practician physician suspects a patient has a heart disease, basic examination will be conducted to make an initial assessment. This includes checking blood pressure, pulse and respiration. The heart and lungs are examined with auscultation. If the practician physician has further reason to suspect heart disease, the patient will be included to evaluate and decide on the forward process of treatment. Dependent upon the suspected heart disease the patient will be referred to the according specific cardiology department. [5]

To improve on the evaluation a practician physician makes during routine clinical check-ups, it could be favourable to implement a method or system in the test battery. Here a system could be implemented to assist during the auscultation process when examining the heart and lungs. A system capable of automatically detecting and classifying abnormal heart sounds, could possibly improve on the assessment where the physician is suspecting heart disease. This would ensure early detection of heart disease patients and sort out persons who would unnecessarily be sent to further examinations at the hospital.

Studies by Castro et al. [14] and Lai et al. [15] have achieved interesting results in detecting

heart murmurs using neural networks analysing PCG. Castro et. al achieved a sensitivity of 69.67% and a specificity 46.91%, while Lai et al. achieved 87% sensitivity, and 100% specificity. Similarly utilizing an acoustic system and signal processing of PCGs, Winther et al. [13] has achieved 72% accuracy (AUC) in detecting CAD.

As mentioned in the previous section, the presence of OS and mitral stenosis is believed to be connected. [18, 3, 17]. The use of NN to analyse PCG could be relevant in the analysis of OS without murmurs and, in case these OS are clinically relevant, also provide a method of finding the cause automatically.

Based on the problem analysis, this leads to the following problem definition:

How can detection of heart opening snaps be automated, and what relation does this abnormal heart sound have to the diagnosis of the patient?

2 | Methods

This study will design and implement a combination of NNs to use for detection and classification of OS in a group of subjects. The group will be consisting of subjects with and without OS. Subjects will be evaluated to either have OS or not, by a manual approach which follows criteria made based on available literature on the physiology and auscultation of OS.

2.1 Data presentation

A total of 600 subjects were included in this study. Data was granted by Acarix A/S as a dataset containing information gathered from acoustic recordings, coronoary artery calcium score (CACS), coronary computed tomographic angiography (CCTA), invasive coronary angiography (ICA) and patient interviews and reviews of patient medical recordings, as stated in [13]. Out of the 600 subjects, 77 (12.83%) were evaluated to have OS. Zero subjects were excluded.
The CADScor®System designed by Acarix A/S consists of a microphone that is fastened to the subject using adhesive patches, and the recordings were done at the fourth intercostal space left of the sternum. Subjects were in a supine position during recordings, where they were asked to hold their breath for eight seconds four times within the three minutes of recording. [19]
Recordings were done with a 8000Hz sampling rate, and were subjected to segmentation into systolic and diastolic parts. This enables alignment of S_2 within the subjects, leading to an easier segmentation of specific heart sounds in relation to S_2 later on. The number of heart cycles varied between 5 and 30 for the subjects, with an average of 16.8. [19]

2.2 Neural Network Models

Neural networks (NN) have been used to analyse big data sets. [20, 21, 22] This study will use an approach of designing a deep neural network for detection and classification of heart OS. The following section describes methods of three different types of NN which will all be used in this project.

2.2.1 Fully Connected Network

Fully Connected Neural Networks are one of simpler types of NN. Fully Connected Networks expands on the classic feedforward network, which was the first and is the simplest model of NN to be invented. [23] The simplest types of feedforward NN only have one layer, and are called single-layer perceptron networks. They work by taking a input and passing it through a single layer with a function to produce a output. The function can be generalized as follows:

$$\hat{y} = f(w_1 * x_1 + w_2 * x_2 + ... + w_{n_x} * x_{n_x} + b) \tag{2.1}$$

, where $\hat{y}$ is the estimated output of the node, f is an activation function. w is weight, x is the node input and b is bias. Often used activation functions are the squashing functions sigmoid σ or hyperbolic tangent *tanh*. More layers can be introduced arranged after each other, with interconnected nodes. This is a multi-layer perceptron (MLP) network, as shown left in figure 2.3. If every node in one layer is connected to every node in the next layer the network is a fully connected network, hence the name. A Fully Connected Network can be seen right in figure 2.3:

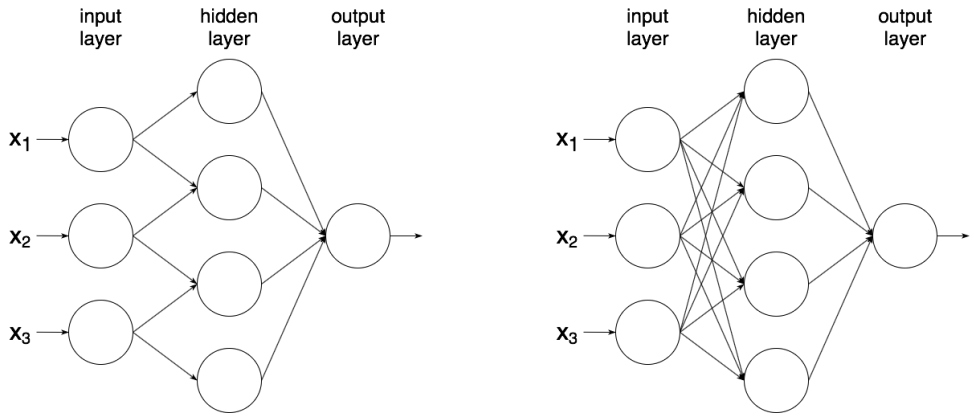

Figure 2.3: Left: A simple Multi-Layer Perceptron neural network. Right: A simple Fully Connected neural network.

2.2.2 Convolutional Network

A 2017 study by Karim et al. [21], propose a NN as a combination of a LSTM and Fully Convolutional Network (FCN) called an LSTM-FCN. Karim et al. use a FCN part in their network to extract features from time series data and concatenate the FCN output with the output of a LSTM part. This type of setup produced state-of-the-art results. [21] Convolutional Neural Networks (CNN) have been widely used for feature extraction, because its architecture gives it an inherent ability to extract features. [21, 24, 25] A CNN functions much like a feedforward NN, where inputs are fed forward through the network. In a CNN the input is fed through filters of one convolutional layer, instead of the whole network. A CNN calculates features of inputs by "scanning" inputs with a filter. The size of the filter determines how many inputs are used for calculating a single feature. With addition of several filter layers, the calculated features of the inputs can be passed on to a new filter layer of feature calculations. With increasing number of filter layers the CNN can detect higher levels of abstract features. [25] The function for a CNN layer can be expressed as in equation (2.2):

$$A(x) = \sigma(W * x + b) \tag{2.2}$$

, where $A(x)$ is the node output, W is a vector of all weights, x is input and b is bias. As it can be seen in equation (2.2) the function is very similar to the basic function of a feedforward node

as shown in equation (2.1). Consider a one dimension fully CNN layer with a filter size of 2, and with 2 filters, see figure 2.4. Consider an input x_n. The CNN layer A in figure 2.4 have filter size 2 and thus takes in two inputs at a time. Each filter in filter layer A will produce an feature according to equation (2.2). These features from A are fed to filter layer B taking two inputs to calculate a new feature based on the two previous features from A. Lastly, the outputs from B is passed to the layer output F. The layer shown figure 2.4 is considered fully connected because every filter node of B is connected to the layer output F. Constructing a NN with an architecture like this would make a Fully Convolutional Neural Network (FCN). [25, 24]

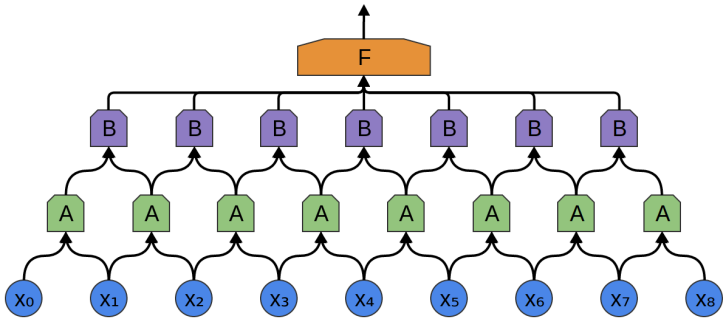

Figure 2.4: A one dimension fully Convolutional Neural Network layer with eight inputs (x), a filter size of 2 and 2 filters (A and B). The output of the CNN layer is F. [25]

2.2.3 Recurrent Neural Networks and Long Short-Term Memory

Many different models of neural networks have been invented to work on different types of data. The use of neural networks in computer vision and image analysis have had much focus in the last decade, especially in relation to the development of self driving cars. Analysis of time series has however not attracted as much attention as image analysis, but have been gaining increased interest and with this NN models which work specifically with sequential data have been developed. [21, 26]

On such model is the Long Short-Term Memory (LSTM) network. A LSTM is a type of recurrent neural network (RNN), however expanding on the basic RNN architecture to overcome shortcomings of a vanishing gradient during backpropagation. Contrary to a feedforward NN, which have fixed in- and output sizes, RNNs can handle input data of sequences of varying sizes. A RNN function by passing the information in a node, back to itself, using the processed output from the prior input to calculate the output for the next input of the sequence. The basic architecture of a RNN can be seen in figure 2.5. The recurring structure of the RNN enables information to persist in the network, giving it the ability to find meaning in data progressing over time. [27, 26, 28]

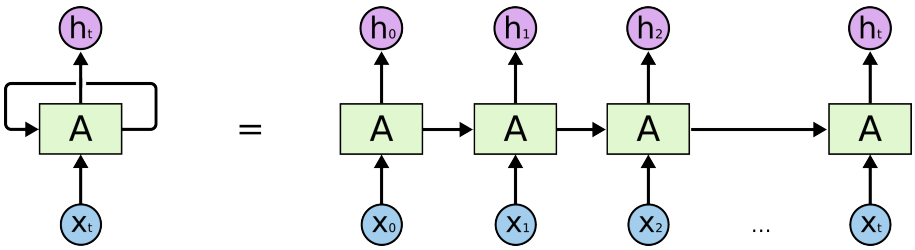

Figure 2.5: Left: the rolled structure of an RNN. Right: unrolled structure of an RNN. *X* is the sequence input, *A* is a node and *h* is the output. [27]

Because the output from each node is passed onto the next step, the network lose information from earlier steps over time. This is the issue of short-term memory resulting in a vanishing gradient which proves a problem when backpropagating through the network. Backpropagation is the process enabling networks to learn. This is achieved by calculating a loss function, after data has passed through the network. A loss function calculates an error to estimate how well the network have performed of predicting the desired output. Then the error is used to calculate the gradients for each node in the network which are used to adjust the weights of the nodes to minimize the loss function. This process repeats until a minimum of the loss function have been found. Finding the global minimum for the loss function will produce the best predictions from the network. [27]

Because basic RNNs have the problem of a vanishing gradient they are not well suited for analysing long sequences. Luckily the LSTM have overcome this issue. The structure of a node in a basic RNN will have only one simple function like the squashing hyperbolic tangent function (*tanh*). The LSTM node is more complex having four layers with gates interacting within the node. (See figure 2.8) The principle behind the LSTM node is that it has a cell state *C* which function as a memory which can be updated and used or ignored in later calculations. The cell state is manipulated by the four gate units in the node. [27]

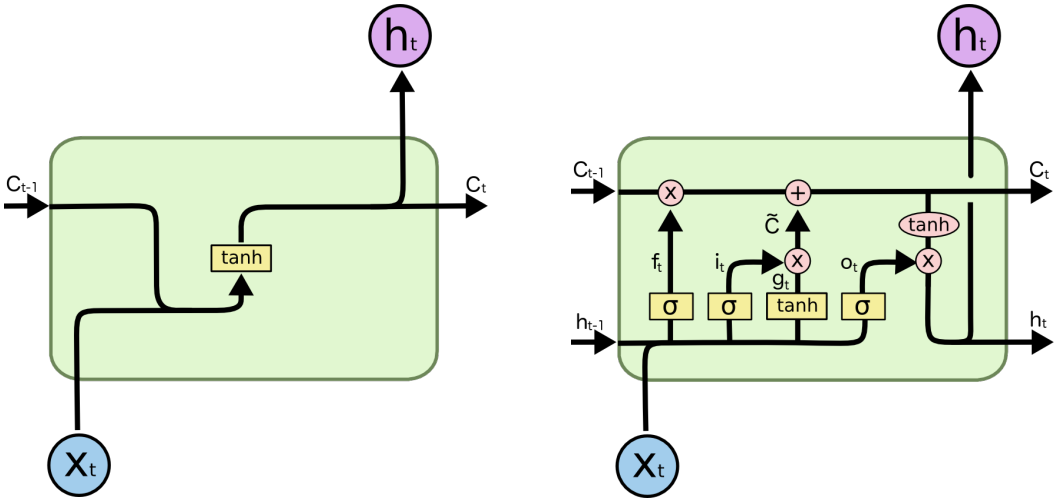

Figure 2.8: Left: The inner layer of a RNN node. Right: The inner layers of a LSTM node. Modified from [27]

The LSTM nodes first gate unit is the forget gate. Based on the earlier output (h_{t-1}) and the new input (x_t), it calculates a value between 0 and 1 for every entry in the cell state, to decide if this value should be forgotten or kept.
The next gate is the input gate which have two steps. First a hyperbolic tangent function ($tanh$) calculates new values ($\widetilde{C}$) to update the cell state. Second a sigmoid function (σ) decides which of the new values to keep or throw away. These two steps are combined in the update gate to update the old cell state (C_{t-1}) to the new cell state (C_t) (see equation (2.4)).
Lastly, the output (h_t) from the node is calculated. This is a combination of the the current cell state (C_t) through a tanh function and the earlier node output (h_{t-1}) through a sigmoid function. (see equation (2.5)) [27, 26] The gate functions for the LSTM layer can be expressed as follows: [26]

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} sigm \\ sigm \\ sigm \\ tanh \end{pmatrix} * W * \begin{pmatrix} h_t \\ h_{h-t} \end{pmatrix} + b \tag{2.3}$$

, where the vectors i, f and o are controlling the input, forget and output states. g is a vector used to modify the memory content in the cell state. W is the vector of all node weights, learnt through backpropagation. h_t and h_{t-1} are the current and old cell output. b is bias. The function for the update of the cell state (C_t) relies on the gates i, f and g and is as follows: [26]

$$C_t = f * C_{t-1} + i * g \tag{2.4}$$

The final output h_t of the LSTM layer is as follows: [26]

$$h_t = o * tanh(C_t) \tag{2.5}$$

Some variations of the LSTM NN have been developed since 1997, among others the GRU (Gated Recurrent Unit) NN invented by Cho et al. in 2014. The GRU simplifies the LSTM node by only having two gate units compared to four in the LSTM. [29] The GRU NN have been gaining an increase of interest as a simpler alternative to the LSTM, however a 2017 study by Greff et al. [30] analysed eight variants of LSTM networks, including the GRU, where they found that none of the tested variants performed significantly better than the classic LSTM. [26, 30]

2.3 Neural Network Design

When designing a NN there are several different parameters that must be defined such as the number of layers, nodes in the layers and possibly how the nodes are connected, as in a Fully

Connected Network. For each model of NN there also exist hyperparameters, a set of parameters used when training and optimizing the network to produce the best results.

Many of these hyperparameters are connected to the learning process of the NN. The network learning to correctly recognize data is dependent on backpropagation. As briefly described in section 2.2, backpropagation is initialized by calculation a loss function for the performance of the network. Different loss functions are used for different network models dependent on whether the network is doing classification or regression. For sequence-to-classification models a cross entropy loss is calculated as: equation (2.6)

$$loss = -\sum_{i=1}^{N}\sum_{j=1}^{K} *t_{ij} * ln * y_{ij} \tag{2.6}$$

, where N is the number of samples, K is the number of classes, t_{ij} is an indicator for the ith sample to the jth class. y_{ij} is the ith output for the jth class.

For sequence-to-regression models variations of a mean-square-error (MSE) loss function is used, dependent on whether the network is doing image-to-image, image and sequence-to-one or sequence-to-sequence regression. This project will not be working with images, thus the only loss functions to possibly be used are for sequence-to-one or sequence-to-sequence regression. Respectively, the functions are as follows:

$$loss = 1/2\sum_{i=1}^{R}(t_i - y_i)^2 \tag{2.7}$$

, where R is the number of responses, t_i is the target output, y_i is the predicted output to response i. [31]

$$loss = \frac{1}{2S}\sum_{i=1}^{S}\sum_{j=1}^{R}(t_{ij} - y_{ij})^2 \tag{2.8}$$

, where S is the sequence length, t_{ij} is an indicator for the ith sample to the jth class. y_{ij} is the ith output for the jth class. [31]

When the loss function have been calculated the networks optimizer algorithm will update weights and bias in the network to gain a better result of the loss function for the next iteration. Optimizing the network and making it learn happens through the process of updating node weights and biases to minimize the loss function. The result of the loss function can be thought of as a topographic map with mountains and valleys, as maximum and minimum. On this map

there might exist several local minima, however it is the global minimum which is desired. To find the global minimum an optimizer algorithm is used. [28]

All optimizer algorithms are based on the idea of gradient descent. Gradient descent is the process of finding the point in the weight space (w) where the loss function has the lowest value. This is achieved by simple steps of first choosing a random point on the loss function and then calculating the direction of the steepest gradient for that point, and taking a step in the other direction of that gradient, which will be towards a minima. The step size is defined by the learning rate. Choosing a good learning rate determines how fast and well the model will converge towards a global minimum for the loss function. Choosing a learning rate too low will make a model which might be precise in finding a minimum as it will not miss one by stepping over it. However, it could converge to a local minimum and it will be very slow in learning meaning it could end up not reaching a minimum during the number of iterations. Setting the learning rate too high will make the model unable to find a minimum as it could be stepping over it at every iteration. An illustration of learning rates too low and high are shown in figure 2.11:

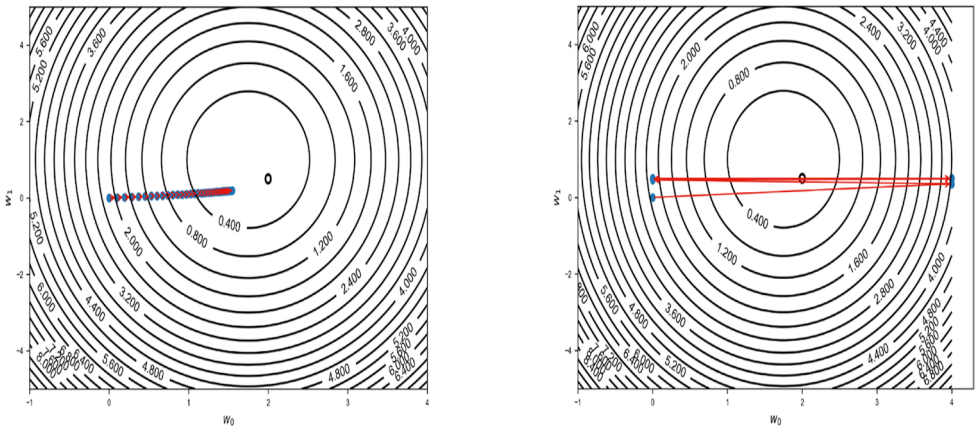

Figure 2.11: Left: Example of a learning rate too low. The model converges too slowly and does not reach a minimum. Right: A learning rate too high. The model never reach the minimum because it continuously steps past it.[28]

Several different optimizer algorithms exist, however, the Adam (Adaptive Moment estimation) optimizer have been most widely used, as it is generally considered faster and better than other optimizers. [28] The Adam optimizer combines the ideas of the RMSProb and Momentum optimizers. From RMSProb it uses exponentially weighted averages from earlier squared derivatives, and from Momentum it uses exponential weighted averages from earlier derivatives. Because Adam is adaptive it enable a network to start out with a higher learning rate, which can be adapted to the problem during the iterations. This enables the model to converge faster but still maintain precision. [28]

A commonly used method in deep learning is the use of mini-batch size. This hyperparameter splits the input dataset into smaller sets of batches which makes the model only perform weight

updates after each iteration, which is when a batch have been passed through the network. Without the use of mini batches weights and biases would only be updated after each epoch, which is when the entire dataset have passed through the network. Not using mini batches normally results in a more stable gradient and convergence, however the entire dataset must then be in the computers memory, making the process slow and computationally expensive on large datasets. With the use of mini batches the convergence is more robust and takes less time and computations. [28]

The training period for the network is decided by the hyperparameter of number of epochs. An epoch is when the entire dataset have passed through the network once. With the use of mini-batches the concept of iterations is introduced. Without mini-batches, the weights and biases update will only happen after an epoch. With mini-batches the input dataset is divided into smaller datasets and when each of the smaller datasets have passed the network, one iteration will have passed and weights and biases will be updated. Thus, if the input dataset has 100.000 samples and is divided into batches of 250 samples, there will be $100.000/250 = 400$ iterations per epoch. Setting the number of epochs decide for how long the network should train. It is desirable that the network should find the highest accuracy and lowest value for the loss function within this time. However, in some scenarios the network can begin to overfit to data. This can happen of several reasons, one being if training for too long. [28]

2.3.1 Overcoming Overfitting

To combat a network overfitting to data there are several methods often used in deep learning. Implementation of the hyperparameter regularization helps a network to better generalize to new data. Regularization adds an additional term to the loss function effectively making weight updates shift towards zero when backpropagating. This makes weights which have little effect on the network output approach zero making them ignorable. With many weights and node effectively being zero, they can be ignored, making the network less complex which hinders the networks ability to overfit to data. [28] The most often used regularization is l_2 regularization. This adds the term in equation (2.9) to the loss function.

$$\frac{\lambda}{2m}||W||_2^2 \tag{2.9}$$

, where, m is the number of observations, λ is a constant regularization parameter and W is the vector of all weights. [28]. Another method used to fight overfitting is the use of a dropout layer, which randomly selects connections between nodes and disables them. This prevents the network nodes to co-adapt too much and constantly use the same node connections which force the network to generalize and not overfit. [32, 28] Stopping the training of a NN can also be used as a means to combat overfitting by stopping the network training before overfitting becomes too much. Early stopping can be implemented through a validation process, where, during training, the network will validate the network output an a specified set of data for accuracy and loss. A hyperparameter can be set to stop the training if the validation loss becomes larger than a loss

found for a previous iteration. This can avoid overfitting and cut training time considerably if the network is quick to converge. [28, 33]

2.4 Manual Labelling and Human-Level Performance

In supervised training a NN must have data which is already labelled. This labelling process is most often performed by humans, who go through the data and manually analyse and evaluate how each observation should be classified. However, it is not reasonable to expect a 0% error rate for humans. Some observations might be too poor in quality to be able to classify and label. When a network is trained on data where some labelling error exist, the network will have trouble in exceeding the level of the human performance. [28]

The data used in this project contained PCGs for 600 subjects which had to be manually labelled. As the authors of this project had zero experience with classification of heart sounds and PCG, this process is a weak link. To strengthen the process as much as possible and avoid subjectivity for the manual classification the following was considered.
According to [3] an OS occurs 30 to 150 ms after S_2. As described in section 1.1.3 the second heart sound, S_2, is a combination of the sounds produced by the closing of the aortic and pulmonary valves (A_2 and P_2). With a normal heart cycle during expiration P_2 will occur after A_2 within less than 30ms. A_2 and P_2 will be heard as a single sound. During inspiration the splitting widens and both sounds become audible. However, this interval is no longer than 40 to 50ms in young subjects. After the age of 40 the interval shortens so that S_2 will be a single sound in both expiration and inspiration. [3] However, the splitting of S_2 can be abnormal. Abnormal splitting exists in three different conditions: wide physiologic splitting, reverse splitting and narrow physiologic splitting.
Wide physiologic splitting is an increase in the interval between the occurrence of A_2 and P_2. It is caused by a delayed electrical activation of the right ventricle caused by right bundle branch block (RBBB). It is often found in patients with pulmonic dilation, mild pulmonic stenosis and atrial septal defect, a hole in the wall between the atriums.
Reverse splitting is where P_2 will occur before A_2. The delay of A_2 is most commonly caused by complete left bundle branch block (LBBB). LBBB is a condition where the left side of the heart contracts after the right side because of a electric conduction defect. Narrow physiologic splitting are similar to normal physiologic splitting in every way except that P_2 has increased intensity and higher frequency. A_2 and P_2 still occur in the normal 30 ms interval. The condition is most often found in subjects with severe pulmonary hypertension. [3]
Wide and narrow physiologic splitting present the most problems in identifying OS, as each condition can make P_2 more audible, either by extending the interval to A_2 or by increasing its intensity and frequency, making P_2 more alike to an OS.

For the labelling process both auscultation and PCGs were examined. Based on the information on OS and splitting of S_2 the following criteria was made for the labelling process of OS:

- The distance between A_2 and a following sound must be greater than 30ms [3]

- A_2 and a following sound must be both visibly and audibly divided
- A sound following A_2 must be no more than 150 ms from A_2 [3]

Both authors individually went through the entire dataset labelling subjects to have an OS or no OS. Afterwards all subjects labelled with OS was reevaluated and a final labelling was decided.

It proved difficult to differentiate between S_2 splitting and early OS when only listening to the recordings. Noise also proved a hindrance in evaluating many subjects. Figure 2.12 shows two PCGs from the dataset. Out of the total 600 subjects 77 was labelled with having OS.

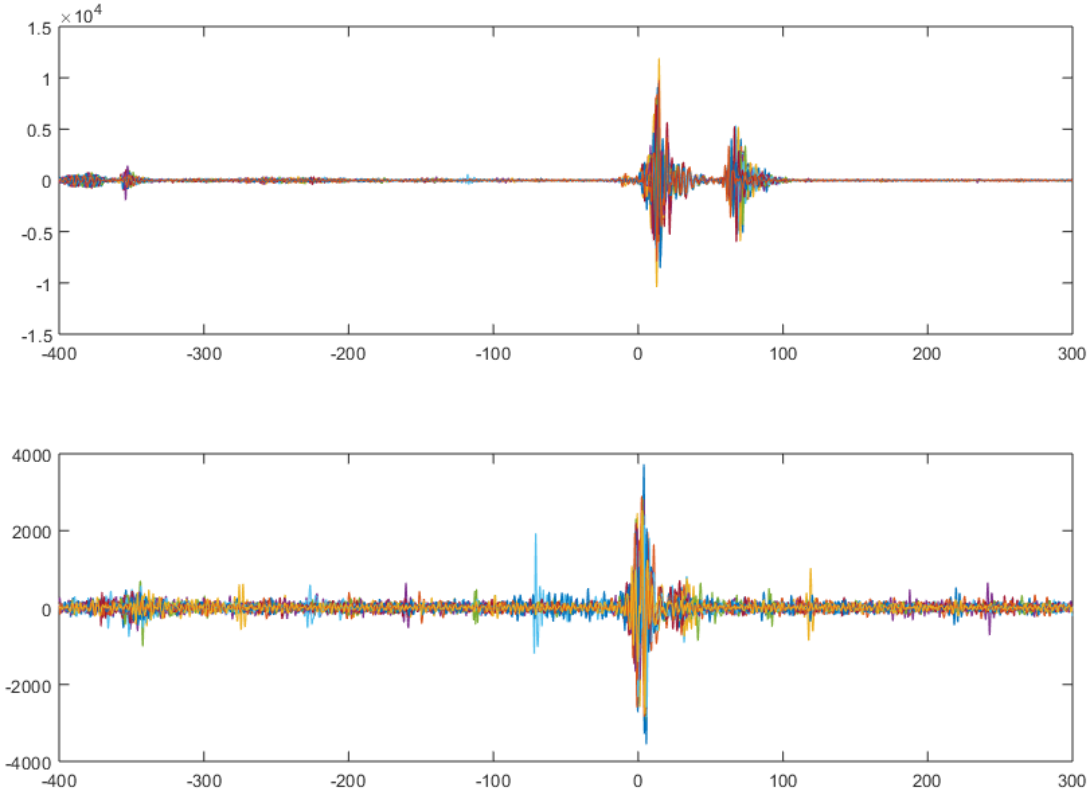

Figure 2.12: Examples of PCGs analysed during the labelling. The figures show several superimposed PCGs from one subject. The topmost PCG is an example of an OS, where S_2 and the OS is easily differentiable. It is visible that the individual PCGs vary little in timing for the occurrence of the OS. The lower PCG is an example of S_2 splitting. Here the individual PCGs vary more over time as the timing of the split are affected by respiration.

The correct way to go about labelling of OS in the dataset would be to recruit a group of physicians, preferably heart specialists, to achieve a higher human-level performance. This could improve the networks training when using this data.

2.5 Statistical Analysis

The dataset used for this project contains medical information on each patient, mainly in relation to heart conditions and pathologies. This information is used to compare the groups of OS subjects with no-OS (NOS) subjects. The parameters chosen for comparing are based on clinical characteristics affecting the heart, like the Duke risk score (sex, age, diabetes, tobacco use, history of myocardial infarction, and symptoms of angina pectoris) and Morise risk score (sex, age, diabetes, tobacco use, symptoms of angina pectoris, hypercholesterolemia, hypertension, family history of CAD, obesity, and estrogen status). [13, 34, 35]

A case-control study was also conducted where subject's cardiac computerized tomography angiographies (CCTA) and echocardiographies were further examined for details on heart pathologies. The case-control study included a total of 83 subject divided in two groups of OS (n = 50) and NOS (n = 33) subjects. The groups were compared on clinical characteristics of the mitral and tricuspid valves and for the presence of atrial septum aneurysm.

Data of clinical characteristics are separated into categorical or non-categorical groups. Non-categorical data is tested for distribution using one-sample Kolmogorov-Smirnov test. Data of Gaussian distribution will be compared using two sample t-test. Non-Gaussian distributed data will be compared using Mann-Whitney test (Wilcoxon rank sum). Categorical data will be compared using Chi-squared test. All tests are performed with 5%, 1% and 0.1% levels of significance and tests for general difference in group means (two-sided). All statistic analyses will be performed using MATLAB (R2019a, MathWorks Inc.)

3 Implementation

3.1 Neural Network Setup

The setup chosen for this project is a, to the authors knowledge, different setup compared to the currently used methods for heart sound classification, as it consists of multiple parallel NN's, classifier steps and a threshold function. The input is based on a windowed feature extraction for the LSTM networks and a filtered signal as an input for the FCN-LSTM combination. At the same time this study only examines the signal within a relevant time span for the opening snaps in relation to S_2, rather than looking at the entire signal for a heart beat.

The final system consists of four steps; preprocessing, NN's, classifier and threshold function. The connection between these are shown in figure 3.1.

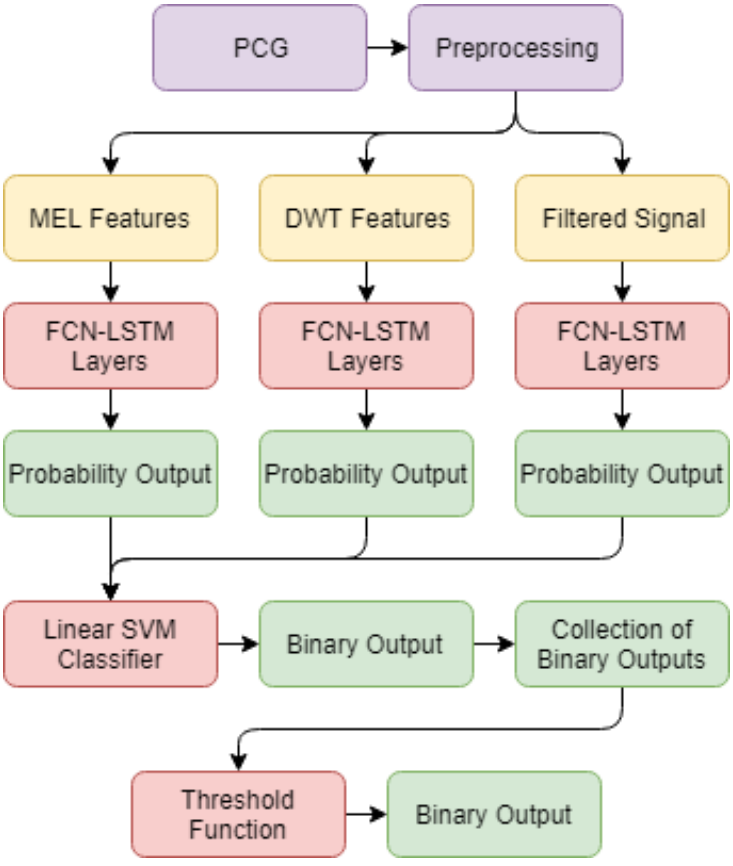

Figure 3.1: Overview showing the final setup and connection through the system of preprocessing, NN's, classifier and threshold function.

This shows that the preprocessed data is fed to neural networks responsible for examining different domains of the data, thereby giving a probability output. This is then fed to the classifier

as three individual values, which then classifies if the specific heart cycle contains a snap. Results from the classifiers examination of single heart cycles is then collected for each patients and subjected to the threshold function in order to determine if the patient should be classified with a snap.

3.1.1 Initial Data Processing

The data was subjected to filtering using a Butterworth bandpass filter between 250 and 1200 Hz to remove unwanted noise and irrelevant information, as the opening snap is a high pitched sound following S_2.

In order to extract features, a multitude of different extractions were implemented. This includes both time, frequency and time-frequency domain features, where the simple time- and frequency based features were extracted from 100 sample windows with a 50% overlap for the first 1500 samples following S_2, as this showed the most precise individual results for each NN and highest overall precision compared to including samples representing S_2 as well. The time-frequency domain features, Mel Feature Cepstral Coefficients (MFCC) and Discrete Wavelet Transform (DWT), were extracted from the 1500 samples using build in functions from MATLAB.

All features were normalized between 0 and 1 in relation to the measured max and min values within the data. This is done in order to improve performance of the NN's, as varying values due to small variations in microphone placement and attenuation of the heart sounds through the body can affect the amplitude of measurements, and indirectly affecting the calculated features.

3.1.2 Network Implementation

LSTM networks were chosen for the feature based NN's, where each network has its own responsibility for predicting the probability of a snap based on features from a certain domain. The design was implemented with two individual NN's for the time and frequency domain. As the MFCC and DWT inputs are essentially transformed signal sequences, these were fed into a LSTM FCN as an experiment, which showed significantly improvement in the RMSE compared to LSTM networks handling these inputs, leading to this being implemented in the final system.

A higher overall accuracy of the system was seen with the implementation of frequency and time dimension based features in two individual NN's, rather than collecting them into one. Therefore the overall system was built with two networks for these two dimensions.

Input	MFCC	Signal	DWT	Frequency Domain	Time Domain
Epochs	120	100	120	30	50
Mini Batch Size	100	100	100	100	100
Init. Learn Rate	0.001	0.0005	0.0005	0.0002	0.0001
Learn Rate Drop Factor	0.8	0.05	0.8	0.5	0.5
Learn Rate Drop Period	4	10	4	3	8
Filter Size	10	10	10	-	-
RMSE	0.21	0.20	0.19	0.5	0.41
Loss	0.022	0.018	0.018	0.1	0.084
Area Under Curve	0.74	0.81	0.74	0.51	0.66

When implementing the signal based networks, one of the main focus points were the filters chosen in the FCN part of the NN, as these have a high effect on the accuracy. A combination of different filter sizes and numbers were tested, with the one yielding highest performance being a combination of three layers consisting of 80, 100 and 80 filters with a size of 10 samples. A higher or lower number of filters and changes in filter size resulted in a decreased performance. The networks were trained with 1988 samples distributed between OS and NOS, and the layers consisted of 256, 512 and 256 neurons.

3.1.3 Feature Selection

When examining the time domain features, it was found that a higher precision was found for the network using only MAV, STD, RMS, VAR and WFL, rather than including ZC, SKW and SSC as well. This improved RMSE for the network from 50% to 41%, while decreasing loss as well. No changes were found when excluding specific features for the frequency domain.

The MFCC based network showed significant improvement from 50% to 21% RMSE when excluding everything but the first and third MFCC, while loss also decreased with this design.

3.1.4 Classifier

Using the Classification Learner in MATLAB it was found that SVM classifiers outperformed other methods, where the most effective was found to be a Linear SVM classifier. This was trained using outputs from the five NN's with classes being either 0 or 1 for the heart cycle be normal or contain a snap. The classifier was trained with 994 samples distributed equally between cycles with and without snaps using 10 fold cross validation in order to decrease overfitting.

The results showed that a classifier based on the DWT, MFCC and signal based networks excluding NN's based on simple features proved the best results with a classifier verification accuracy of 81.9%. .

3.1.5 Threshold Function

At the end a threshold function was implemented, in order to determine if the patient was supposed to be classified with an opening snap. This function determines the outcome based on the percentage of snaps in the heart cycles for each patient, with a threshold of 50% for a person to be classified with a snap. This threshold has been chosen in order to achieve the highest possible sensitivity while keeping a reasonable specificity.

4 | Results

The manual labelling process of subjects in the dataset resulted in 77 subjects (12.83%) out of the total 600 subjects, were evaluated to have OS. Zero subjects were excluded. The two groups of OS subjects (n = 77) and no-OS (NOS) subjects (n = 523) were used for training and testing for the NNs and later comparisons were made between the two groups. The subjects were all over 40 years of age and nearly evenly distributed in sex. The details for the two groups are presented in table 4.3.

4.0.1 System Accuracy

Precision of the individual networks are as described in table 4.1 through area under curve (AUC).

Network Features	AUC
MFCC	0.7388
DWT	0.8161
Signal	0.7433

Table 4.1: Accuracy of individual NN branches measured in AUC.

The classifier accuracy is described in figure 4.1 with a precision of 81.1% for single cycles, with approximately the same specificity and sensitivity.

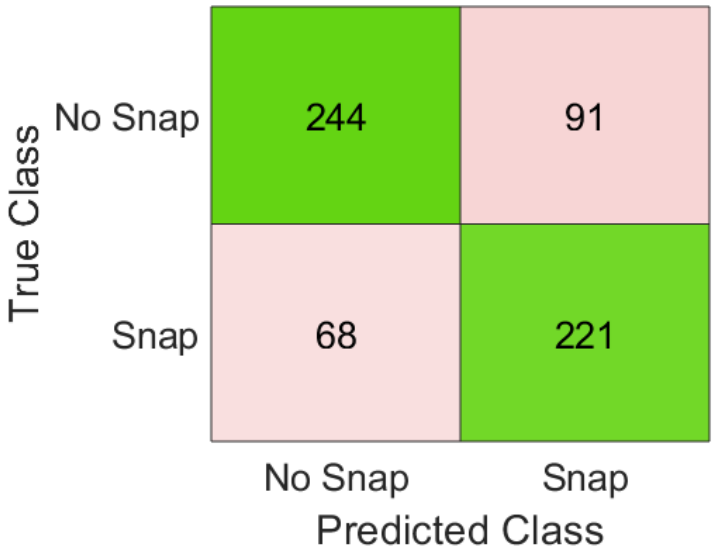

Figure 4.1: Confusion matrix for the classifier accuracy. These results are for the classification of whether heart beats contain an OS or not, before the threshold function.

The receiver operating characteristics (ROC) for the individual networks are shown in figure 4.7.

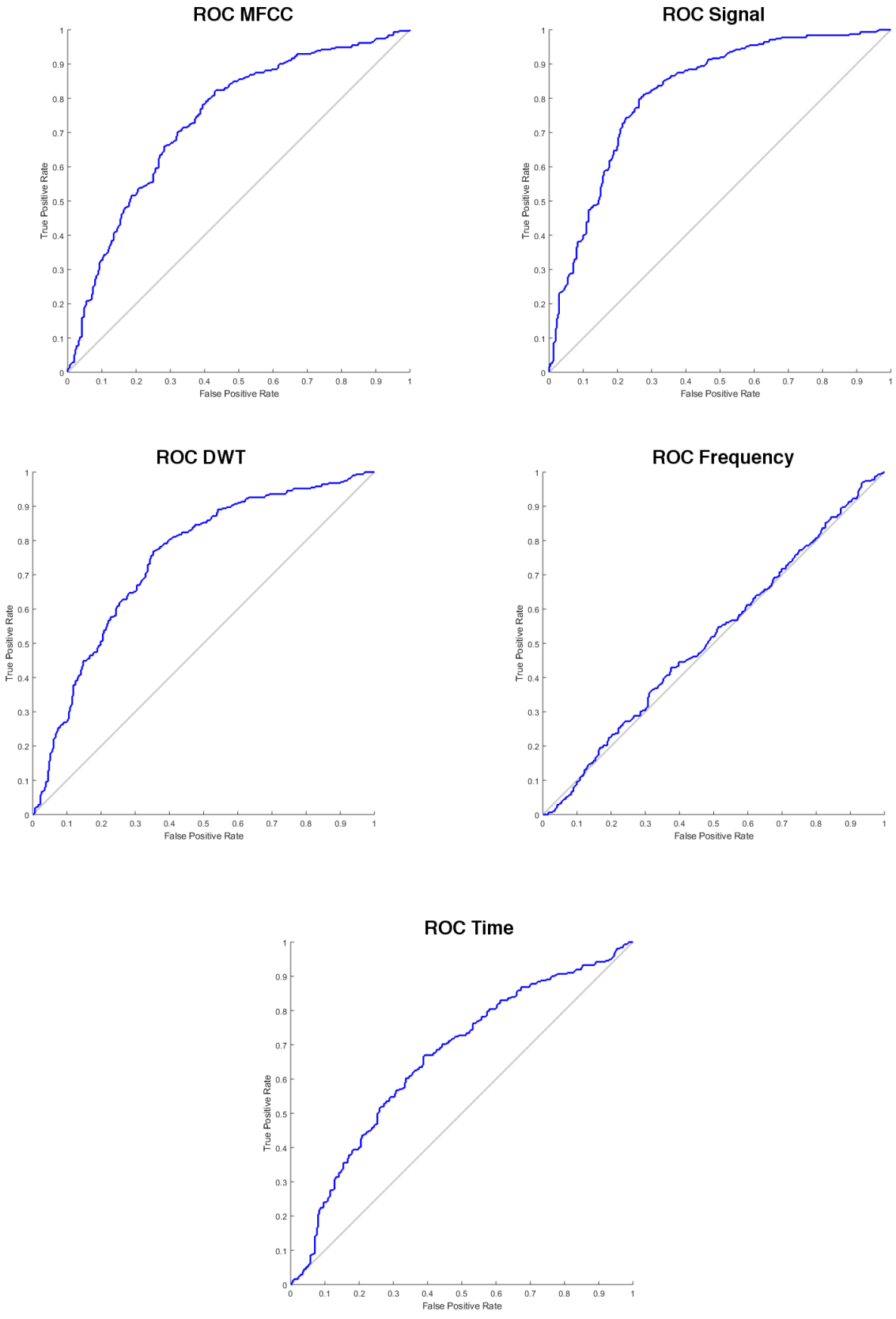

Figure 4.7: ROC curves for the networks: MFCC, signal, DWT, frequency and time based networks are shown in that order.

The system accuracy can be seen in table 4.2, where the optimal threshold was found to be 45% providing 94.5% sensitivity, 89.5% specificity and 92% best average accuracy (BAC). To calculate AUC of the threshold function, the mean of predictions for each subject was set as the classifier output while the label for that subject was the supposed class. Thereby the overall systems AUC was found to be 0.9288 for the classification of a test group of 36 subjects mixed equally between heart cycles containing OS or no OS..

Threshold	Sensitivity	Specificity	BAC
35%	94.5%	78.9%	86.7%
45%	94.5%	89.5%	92%
55%	70.6%	89.5%	80.1%

Table 4.2: System accuracy with various thresholds.

The accuracy for correctly and wrongly classified subjects for the overall system can be seen in figure 4.8.

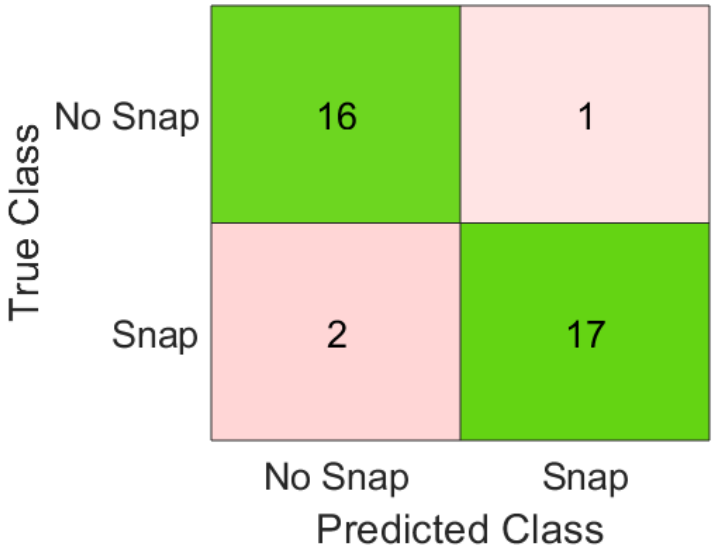

Figure 4.8: Confusion matrix for the overall system. These results are the classification on whether subjects have OS or not.

The ROC curve for the threshold function is shown in figure 4.9.

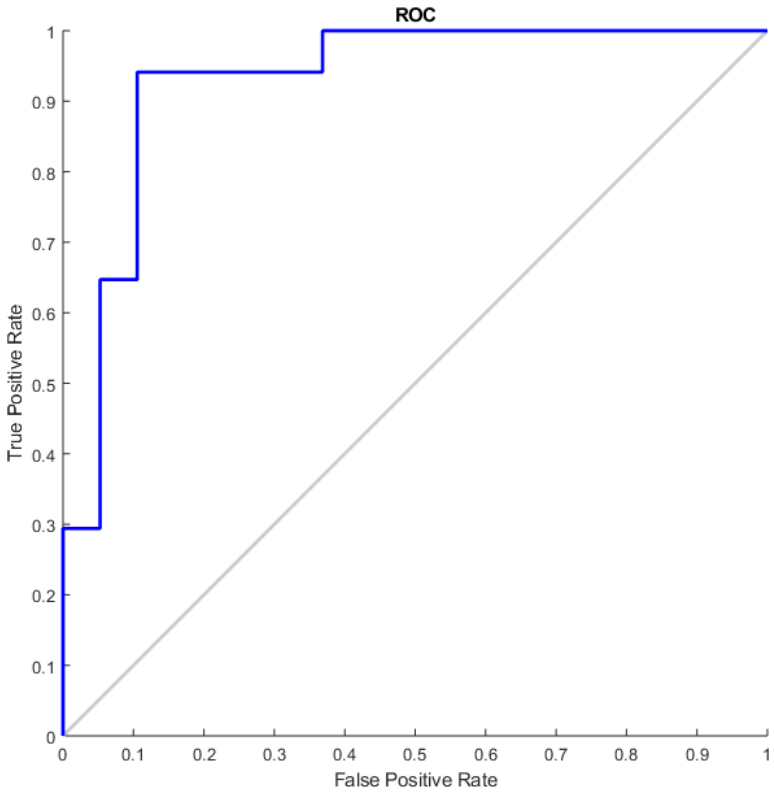

Figure 4.9: ROC curve for the threshold function.

4.0.2 Statistical Results for comparison of OS and NOS subjects

As mentioned in section 2.5 the two groups of subjects, OS subjects and no-OS subjects (NOS) are compared on clinical characteristics affecting the heart, like the Duke risk score and Morise risk score. [13, 34, 35] The Kolmogorov-Smirnov test proved all non-categorical characteristics to be of a non-Gaussian distribution, thus all non-categorical data was tested with a Mann-Whitney test (Wilcoxon rank sum). All categorical data were tested with Chi-squared test. The chosen characteristics are shown along with statistical results for the comparison in table 4.3. Non-Gaussian distributed values are shown with a standard deviation ($\pm$), while categorical values are reported with frequencies (percentages). Several of the chosen characteristics had missing entries in the dataset, because not every subject have undergone the same procedures and clinical tests. Characteristics which have missing entries are annotated and the number of missing entries are noted in the bottom of table 4.3. A significant difference were found for age between the groups ($p < 0.05$). No other significant differences were found.

	OS (n = 77)	NOS (n = 523)
Age (Years)	54.42 ± 9.54*	57.38 ± 8.75*
Sex		
- Female	39 (51%)	289 (55%)
- Male	38 (49%)	234 (45%)
Weight (kg)	80.75 ± 16.75	79.14 ± 14.21[1]
Height (cm)	173.75 ± 7.67	172.06 ± 8.98[2]
Pulse (BPM)	64.57 ± 12.05	65.45 ± 10.81[3]
Blood pressure (mmHg)		[4]
- Systolic	135.21 ± 19.72	139.06 ± 17.97
- Diastolic	82.83 ± 12.82	84.12 ± 10.90
Smoker		
- Active	15 (19%)	95 (18%)
- Former	25 (32%)	182 (35%)
- Never	37 (48%)	246 (47%)
Diabetes		
- Has diabetes	1 (1%)	27 (5%)
- No diabetes	76 (99%)	496 (95%)
CADScore	19.77 ± 9.18[5]	21.02 ± 9.85[5]
Agatston score	136.13 ± 335.61	118.70 ± 302.13
P-cholesterol (mmol/L)	5.37 ± 0.95[6]	5.38 ± 1.02[6]

Table 4.3: Data is missing for several categories: [1] Weight NOS: 3, [2] Height NOS: 1, [3] Pulse NOS: 2, [4] Systolic Blood Pressure NOS: 2, [5] CADScore OS: 1 NOS: 18, [6] P-cholesterol OS: 7 NOS: 34.
Significant differences between the two groups are indicated with * for $p < 0.05$ and ** for $p < 0.01$.

4.0.3 Comparison of CCTA and Echocardiography

A selection of OS and NOS subjects, who had had CCTA and echocardiography performed, were examined for conditions and pathologies of the heart. The Kolmogorov-Smirnov test showed that all non-categorical values were from a non-Gaussian distribution. This data was tested with a Mann-Whitney test (Wilcoxon rank sum). Categorical data was tested with Chi-squared test. The results are shown in table 4.4. Non-Gaussian distributed values are shown with a standard deviation (±), while categorical values are reported with frequencies (percentages). Missing entries are annotated and noted at the bottom of table 4.4. A significant difference between the groups were found for subjects which had been diagnosed with mild mitral insufficiency ($p < 0.05$). No other significant differences were found.

	OS (n=50)	NOS (n=33)
Mitral Plague		
- No	49 (98%)	31 (94%)
- Yes	1 (2%)	2 (6%)
Mitral Valve Thickening		
- No	49 (98%)	32 (97%)
- Yes	1 (2%)	1 (3%)
MR		
- No	47 (94%)	3 (6%)
- Yes	32 (97%)	1 (3%)
Mitral insufficiency	[1]	[1]
- None	29 (66%)	13 (43%)
- Mild	12 (27%) *	17 (57%) *
- Moderate	3 (7%)	0
- Severe	0	0
Mitral stenosis	[2]	[2]
- None	45 (98%)	30 (100%)
- Mild	1 (2%)	0
- Moderate	0	0
- Severe	0	0
Mitral Restrictive	[3]	[3]
- No	44 (96%)	31 (100%)
- Yes	2 (4%)	0
Mitral Flow (m/s)	0.71 ± 0.17[4]	0.73 ± 0.21[4]
Mitral Dec (ms)	216.05 ± 52.18[5]	213.33 ± 60.95[5]
Mitral E (m/s)	0.10 ± 0.03[6]	0.11 ± 0.03[6]
Tricuspid insufficiency	[7]	[7]
- None	22 (61%)	7 (41%)
- Mild	13 (36%)	10 (59%)
- Moderate	1 (3%)	0
- Severe	0	0
Tricuspid stenosis	[8]	[8]
- None	35 (100%)	17 (100%)
- Mild	0	0
- Moderate	0	0
- Severe	0	0
Atrial Septum Aneurysm	[9]	[9]
- No	19 (95%)	13 (93%)
- Yes	1 (5%)	1 (7%)

Table 4.4: Data is missing for several categories: [1] Mitral insufficiency OS: 6 NOS: 3, [2] Mitral stenosis OS: 4 NOS: 3, [3] Mitral Restrictive OS: 4 NOS: 2, [4] Mitral Flow OS: 13 NOS: 12, [5] Mitral Dec OS: 13 NOS: 12, [6] Mitral E OS: 14 NOS: 11, [7] Tricuspid insufficiency OS: 14 NOS: 16, [8] Tricuspid Stenosis OS: 15 NOS: 16, [9] Atrial Septum Aneurysm OS: 30 NOS: 19.
Significant differences between the two groups are indicated with * for $p < 0.05$ and ** for $p < 0.01$.

Bibliography

[1] Frederic H. Martini. *Fundamentals of Human Anatomy and Physiology*. 9th ed. 2012.

[2] Hjerteforeningen. *Fakta om Hjerte-Kar-Sygdom i Danmark*. 2019.

[3] Valentin Fuster et al. *Hurst's The Heart*. 12th ed. McGraw-Hill Companies, Inc., 2008, p. 2473.

[4] Sundhedsstyrelsen. *Sundhedsstyrelsens specialevejledning for Intern medicin: Kardiologi*. Tech. rep. november. 2018, pp. 1–14.

[5] Sundhedsstyrelsen. *Anbefalinger for tværsektorielle forløb for mennesker med hjertesygdom*. 2018, pp. 1–48.

[6] Linda E. Pinsky and Joyce E. Wipf. *Learning and Teaching at the Bedside*. 1998.

[7] Christiane A. Geluk et al. "Measurement of coronary calcium scores by electron beam computed tomography or exercise testing as initial diagnostic tool in low-risk patients with suspected coronary artery disease". In: *European Radiology* 18.2 (2008), pp. 244–252.

[8] Arthur S. Agatston et al. "Quantification of coronary artery calcium using ultrafast computed tomography". In: *Journal of the American College of Cardiology* 15.4 (1990), pp. 827–832.

[9] Louise Nissen et al. "Danish study of Non-Invasive testing in Coronary Artery Disease (Dan-NICAD): study protocol for a randomised controlled trial". In: *Trials* (2016), pp. 1–11.

[10] Andrew Cassar et al. "Chronic coronary artery disease: Diagnosis and management". In: *Mayo Clinic Proceedings* 84.12 (2009), pp. 1130–1146.

[11] Jesper Kjærgaard, Hans Christian Kjeldsen, and Ole Gøtzsche. *Mislyd over Hjertet hos Voksne*. 2016.

[12] Gari D. Clifford et al. "Recent advances in heart sound analysis". In: *Physiological Measurement* 38.8 (2017), E10–E25.

[13] Simon Winther et al. "Diagnosing coronary artery disease by sound analysis from coronary stenosis induced turbulent blood flow : diagnostic performance in patients with stable angina pectoris". In: *The International Journal of Cardiovascular Imaging* 32.2 (2016), pp. 235–245.

[14] Ana Castro and Tiago T V Vinhoza. "Automatic Heart Sound Segmentation and Murmur Detection in Pediatric Phonocardiograms". In: *IEEE Engineering in Medicine and Biology Society* (2014), pp. 2294–2297.

[15] Lillian S W Lai et al. "Computerized Automatic Diagnosis of Innocent and Pathologic Murmurs in Pediatrics: A Pilot Study". In: *Congenital Heart Disease* 11.5 (2016).

[16] Grzegorz Redlarski, Dawid Gradolewski, and Aleksander Palkowski. "A system for heart sounds classification". In: *PLoS ONE* 9.11 (2014).

[17] E. Sejdić and E. Veledar. "An algorithm for classification of opening snaps and third heart sounds based on wavelet decomposition". In: *IFMBE Proceedings* 37 (2011), pp. 404–407.

[18] Tatiana Mularek-Kubzdela et al. "First heart sound and opening snap in patients with mitral valve disease. Phonocardiographic and pathomorphologic study". In: *International Journal of Cardiology* 125.3 (2008), pp. 433–435.

[19] Simon Winther et al. "Diagnostic performance of an acoustic-based system for coronary artery disease risk stratification". In: *Heart* 104.11 (2018), pp. 928–935.

[20] Lin-peng Jin and Jun Dong. "Ensemble Deep Learning for Biomedical Time Series Classification". In: *Hindawi Publishing Corporation* (2016), pp. 1–13.

[21] Fazle Karim et al. "LSTM Fully Convolutional Networks for Time Series Classification". In: *IEEE Access* 6 (2017), pp. 1662–1669.

[22] Hoang Anh Dau et al. "The UCR Time Series Archive". In: (2018).

[23] Jürgen Schmidhuber. "Deep Learning in neural networks: An overview". In: *Neural Networks* 61 (2015), pp. 85–117.

[24] Zhiguang Wang, Weizhong Yan, and Tim Oates. "Time series classification from scratch with deep neural networks: A strong baseline". In: *Proceedings of the International Joint Conference on Neural Networks* 2017-May (2017), pp. 1578–1585.

[25] Christopher Olah. *Conv Nets: A Modular Perspective*. 2014.

[26] Andrej Karpathy, Justin Johnson, and Li Fei-Fei. "Visualizing and Understanding Recurrent Networks". In: *ICLR* (2015), pp. 1–12.

[27] Christopher Olah. *Understanding LSTM networks*. 2015.

[28] Umberto Michelucci. *Applied Deep Learning*. 2018.

[29] Kyunghyun Cho et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". In: (2014).

[30] Klaus Greff et al. "LSTM: A Search Space Odyssey". In: *IEEE Transactions on Neural Networks and Learning Systems* 28.10 (2017), pp. 2222–2232.

[31] MathWorks. *regressionLayer*. 2019.

[32] Andres Cicuttin et al. "A programmable System-on-Chip based digital pulse processing for high resolution X-ray spectroscopy". In: *2016 International Conference on Advances in Electrical, Electronic and Systems Engineering, ICAEES 2016* 15 (2017), pp. 520–525.

[33] MathWorks. *MatLab: trainingOptions for NN*. 2019.

[34] David B. Pryor et al. "Estimating the likelihood of significant coronary artery disease". In: *The American Journal of Medicine* 75.5 (1983), pp. 771–780.

[35] Anthony P. Morise, W. John Haddad, and David Beckner. "Development and validation of a clinical score to estimate the probability of coronary artery disease in men and women presenting with suspected coronary disease". In: *American Journal of Medicine* 102.4 (1997), pp. 350–356.