Answer Sheet for Assessment 1
L349 - Mobile Health

Tim Luka Horstmann
Computer Laboratory
University of Cambridge
tlh45@cam.ac.uk
983 words (excluding bibliography, headers, and captions)

Part 1: Audio Processing Basics [25 marks]

Task 1.2

Question 1: Discuss any differences between the two files in the time domain, giving possible reasons.

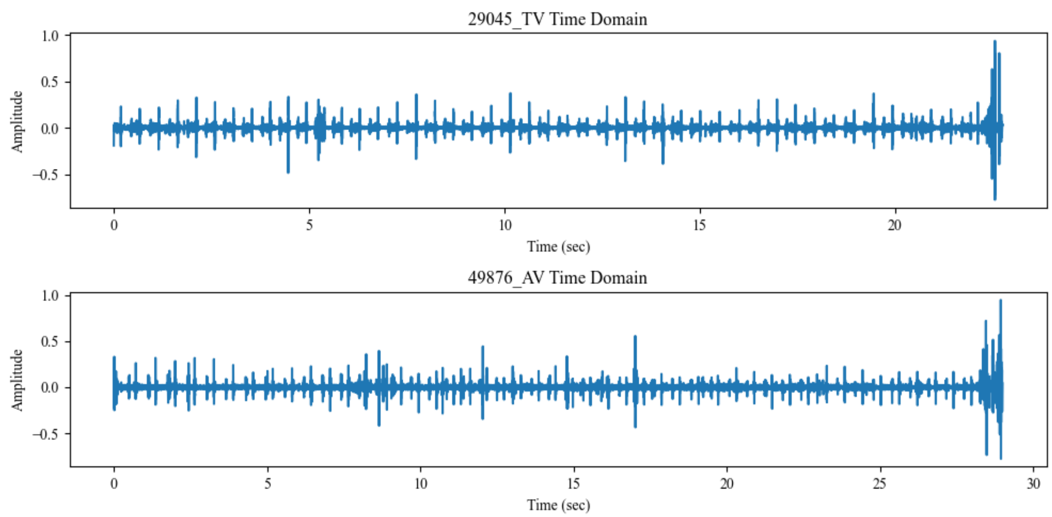

Figure 1: Time domain plot of the 29045_TV and 49876_AV audio files. Full length. Both audio signals show clear heartbeats with relatively regular patterns. The plots shown in this Figure do not allow a clear visual differentiation between murmur and non-murmur.

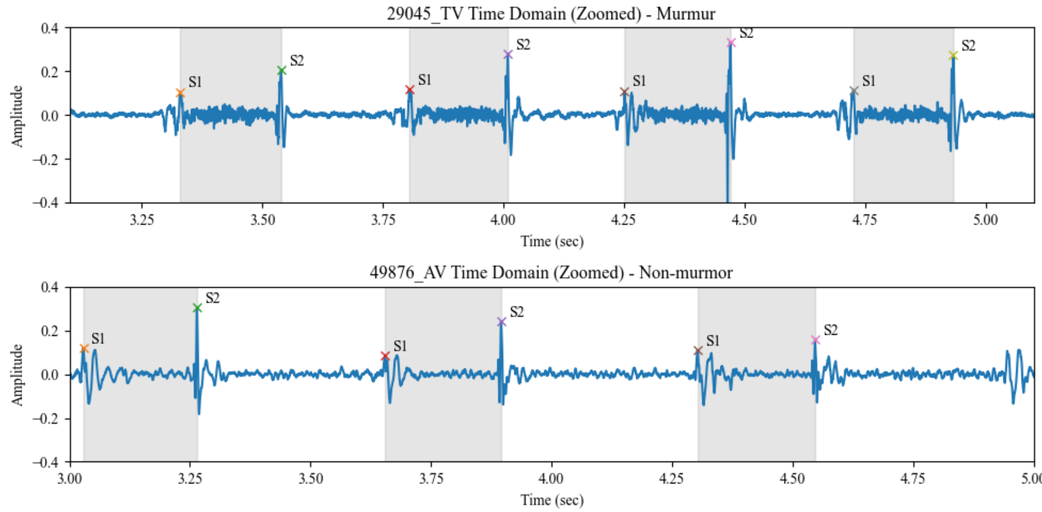

Figure 2: Time domain plot of the 29045_TV and 49876_AV audio files. Zoomed x-axis (time). Both plots show a clear, regular heartbeat with S1 and S2 mechanically annotated in the plot. As research suggests that the sound between S1 and S2 (area highlighted in grey) is particularly relevant for identifying murmurs, the significant amplitude in 29045_TV in this area indicates a murmur [37].

Table 1 provides an overview of the differences between the two files in the time domain:

Table 1: Comparison of heartbeat audio signals in time domain

File	Observations	Possible Reasons
29045_TV Figure 1 & Figure 2	- clearly recognisable heartbeats, especially when zoomed in - constantly noticeable amplitude between S1 and S2 (see grey area in Figure 2), otherwise relatively small amplitude - heart rate: ∼120 bpm, frequency: ∼2 Hz	- clear sound recording with little noise - noticeable amplitude between S1 & S2 might indicate murmur - higher heart rate potentially consequence of heart conditions that might also cause murmor [49]
49876_AV Figure 1 & Figure 2	- heartbeats (incl. S1 & S2) can also be identified, but with greater irregularity and variation in amplitude - difficult to accurately identify S1 - stronger fluctuations throughout - heart rate: ∼90 bpm, frequency: ∼1.5 Hz	- likely normal heart - less accurate recording with more noise

Question 2: Based on the above, can you visually differentiate between the murmur and non-murmur heart sounds? Predict which is the murmur and which is the non-murmur.

- research shows sound development between S1 ("lub") and S2 ("dub") is particularly relevant to identify murmurs [37]
- **29045_TV** shows pronounced amplitude here (Figure 2) -> could represent "blowing, whooshing, or rasping sound heard during a heartbeat", indicating murmur [47]
- possible murmur type: pattern similar to mitral regurgitation, but strongly audible at tricuspid valve (TV) -> might indicate tricuspid regurgitation [14, 21]
- -> **29045_TV = murmur**

Task 1.3

Question 1: Discuss any differences between the frequency domain representations of the murmur and non-murmur files.

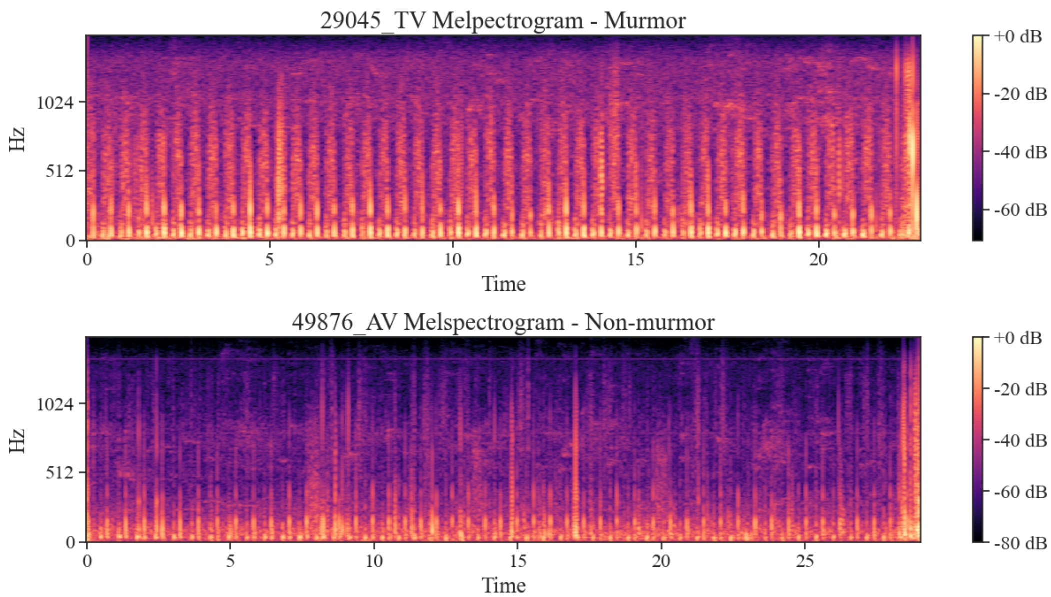

Figure 3: Melspectogram representation of 29045_TV and 49876_AV. Full length with parameters n_fft=1024, hop_length=128, and n_mels=256. Even without filtering and zooming, both melspectograms clearly show distinguishable heartbeats with a clear separation of S1 and S2.

- 29045_TV (Murmur):
  - concentration of peaks primarily in low-frequency range, up to $\sim$150Hz
  - two further peaks/plateaus in frequency distribution around $\sim$250Hz and $\sim$650Hz
  - generally more spread-out
  - melspectogram representation, especially when zoomed (Figure 4), clearly reveals S1 and S2 with energy in high frequencies –> corresponds to typical high(er) frequencies of murmurs [2]
- 49876_AV (Non-murmur):
  - almost exclusively concentration of peaks in low-frequency range, although slightly higher, up to $\sim$200Hz
  - exhibits some isolated high-frequency peaks, likely noise
  - melspectogram: short "bursts" for S1 & S2, indicating normal heartbeats

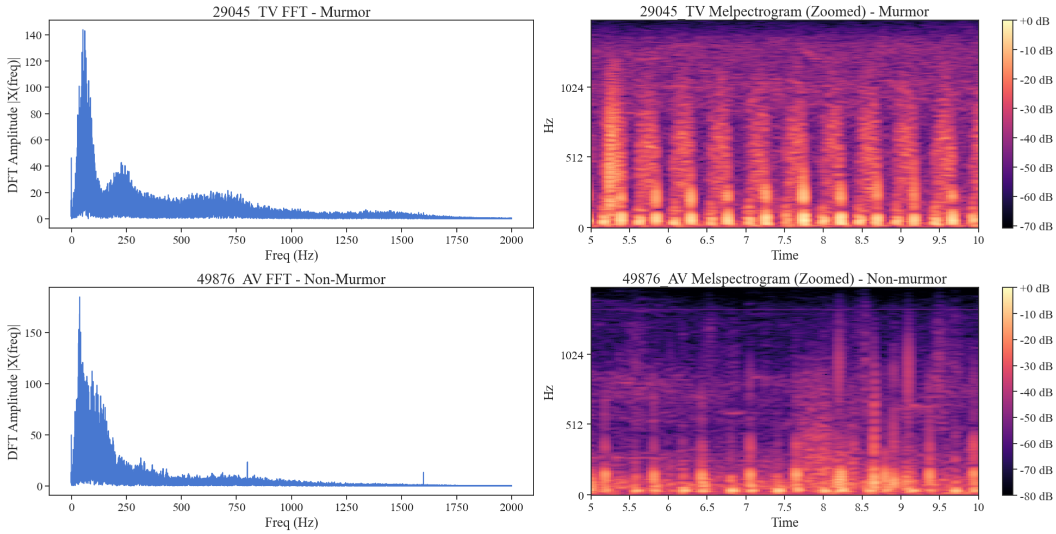

Figure 4: FFT and zoomed melspectogram representation of 29045_TV and 49876_AV (same parameters as in Figure 3). Both frequency domain representations allow a clear identification of 29045_TV as the murmur file (see Question 1).

Question 2: Are there any features that are evident in the frequency domain that you could not distinguish in the time domain?

- time domain cannot show which frequencies make up the signal –> frequency domain allows conclusions regarding involved frequencies and extent of their involvement
- reveals that 29045_TV is composed of much stronger higher frequencies than 49876_AV -> confirms, in line with scholarship [e.g. 2, 16, 39], classification of 29045_TV as murmur

Task 1.4

Question 1: Discuss and provide reasons for your choice of filter type and cutoffs.

Step 1 consider findings of initial visual analysis of (unfiltered) frequency domain graphs -> Figure 5 exhibits majority of peaks to a maximum of ∼1000Hz –> cutoff latest after 1000Hz seems reasonable

Step 2 confirm detailed cutoffs & filter types through relevant scientific research about the typical frequency range for heart sounds and murmurs & settings typically used in heart sound analyses -> Table 2 informed decision to use **6th-order Butterworth bandpass filter (BBF) with cutoff frequency from 20 to 600 Hz**

Element of Interest	Relevant Frequencies	Filter Specified?	Source
Cardiopulmonary auscultation	50-1200 Hz	N/A	Charbonneau and Sudraud [1]
Heart sound	40-1100 Hz	BFF	El-Segaier et al. [6] Markaki et al. [18]
Heart sound	25-900 Hz	6th-order BBF	Chakir et al. [23]
Heart beat	40-500 Hz	4th-order BBF	Shekhar et al. [39]
S1 and S2	50-500 Hz	N/A	Spencer and Pennington [22]
S3 and S4	20-200 Hz	N/A	Spencer and Pennington [22]
Critical heartbeat	70-120 Hz	N/A	Bankaitis [46]
Murmurs	80-500 Hz	N/A	Tomas et al. [16]
Murmurs	200-410 Hz	N/A	Donnerstein [2]
Murmurs	< 300 Hz	N/A	Spencer and Pennington [22]
Murmurs	20-150 Hz, < 500 Hz	N/A	Rennert et al. [5]
Most murmurs	< 200 Hz	N/A	Debbal and Bereksi-Reguig [11]

Table 2: Overview of selected frequency and filter settings for analysing heartbeat and murmur sounds in scientific literature

Question 2: Provide a discussion of the differences between the raw and filtered data, and thus on the importance of filtering in signal processing. Are there any potential disadvantages or tradeoffs of applying signal processing?

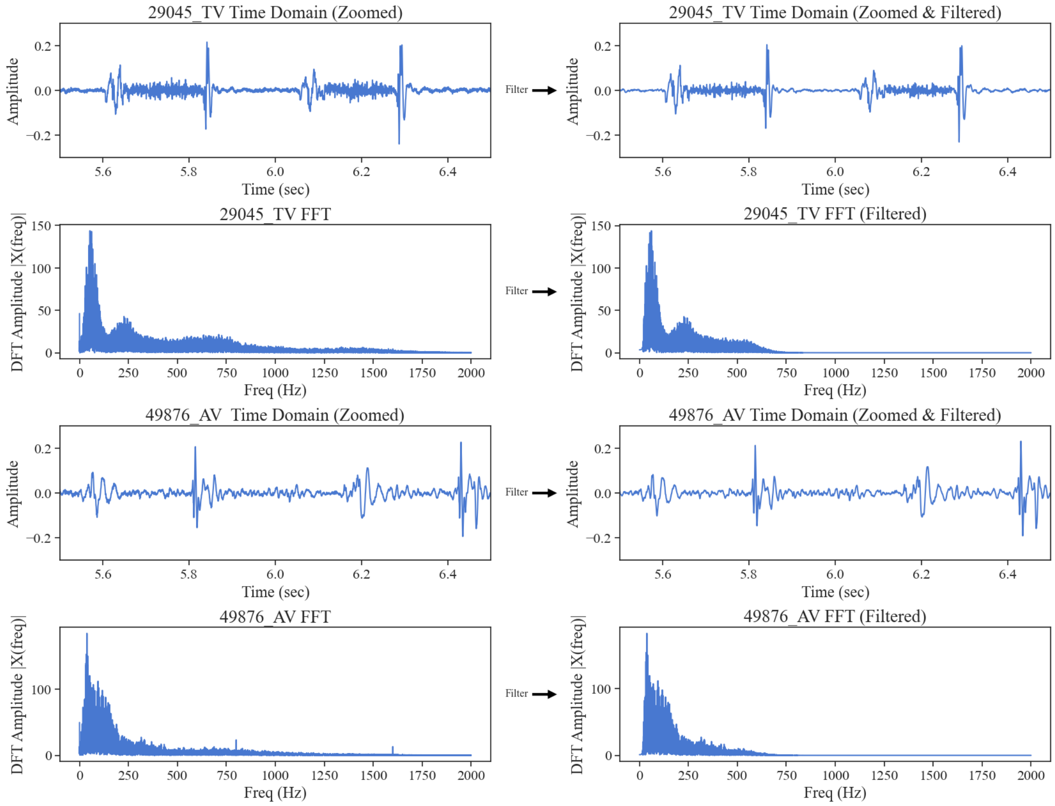

Figure 5: Comparison of the unfiltered and filtered time and frequency domain plots for 29045_TV (Murmur) and 49876_AV (Non-murmur) respectively. Filtering the audio signals facilitates the analysis of the relevant sections of the signals (i.e. the heartbeat and murmur frequencies).

Differences :

- filtered signal in time domain is visually "thinner" & exhibits less noise around regions of interest (i.e. heartbeat sound between S1 & S2) -> indicates reason for signal processing
- isolated and clear outliers (i.e. "noise") are removed
- filtered FFT plots reveal focus on strong frequencies and regions with relatively high amplitude

Importance:

- one of the most important and typical first steps in analysis of (heart) sounds [45, 27]
- goal: reveal information in measurements/signal [12, 17], remove undesired/unwanted signal components (e.g. noise), increase reliability, facilitate analysis of relevant parts [15]

Disadvantages & Trade-offs:

- filtering is a challenging task [8]
- facilitates analysis but risks removing relevant sounds, as noise often shares heart sound frequency range [45]

Task 1.5

Question 1: Discuss whether you can differentiate between the signals or not and if not, why not.

- both files show numerous irregularities in time domain graphs (unfiltered & filtered) -> heartbeats can be visually detected, but not as clearly analysed as in Task 1.2
- frequency domain plots below suggest murmur in AV_29045, due to stronger higher frequencies and two regions with peaks, but AV_39043 exhibits strong energy in these areas too
- -> overall no precise visual classification possible, at most hypothesis

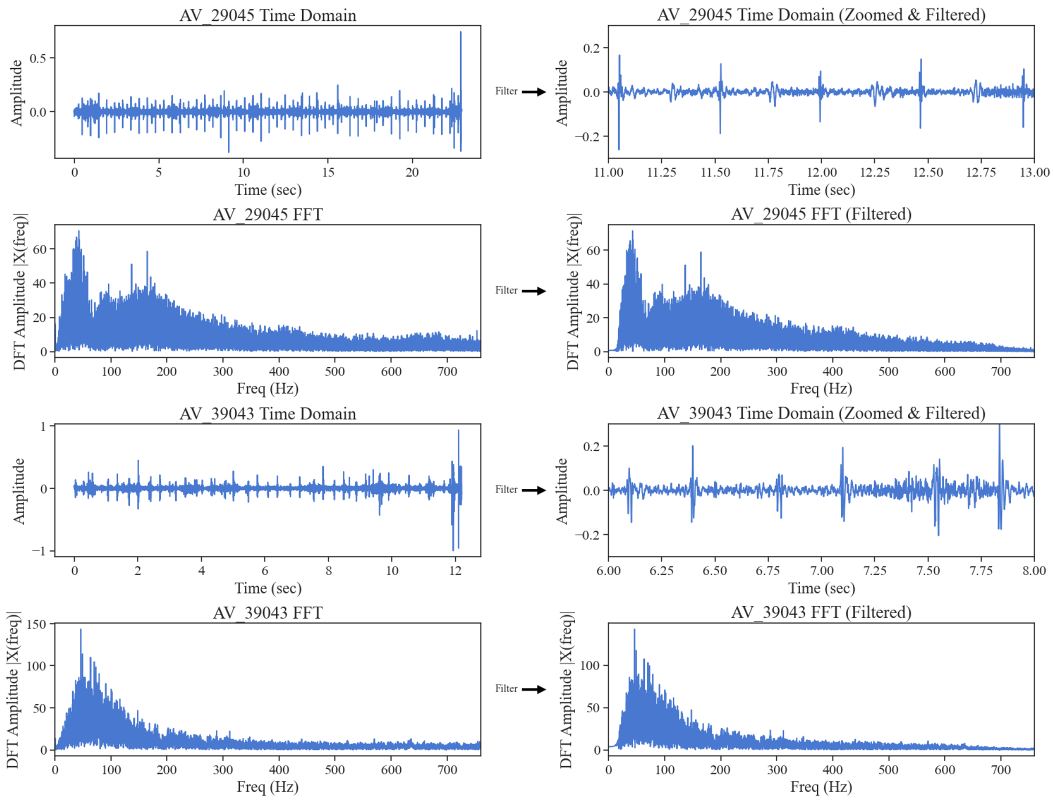

Figure 6: Comparison of the unfiltered and filtered time and frequency domain plots for AV_29045 (likely murmur) and AV_39043 (likely non-murmur) respectively. Although the classification of the audio signal still seems possible, it is very unreliable and, at best, an imprecise estimate.

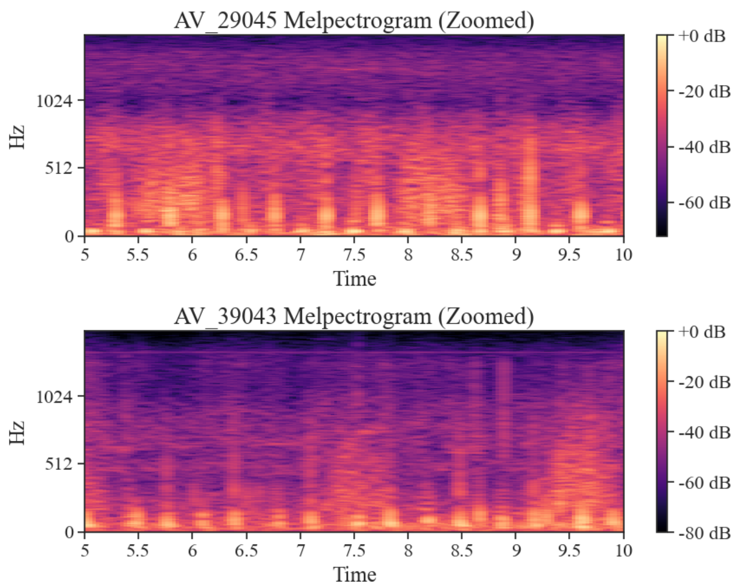

Figure 7: Melspectogram representation of AV_29045 and AV_39043 with the same parameters as in Task 1.3. The melspectograms still enable a visual classification of S1 & S2. However, a precise classification of murmur vs. non-murmur is aggravated. Since S1 & S2 in AV_29045 exhibit more power in higher frequencies, AV_29045 likely represents a murmur file. A more precise classification requires further analysis.

Part 2: Dataset processing [15 marks]

Task 2.2

Question 1: What is the ratio of normal to murmur patients? And what is the ratio of normal to murmur samples? Can you think of any implications of this?

- #Normal Patients: 135, #Murmur Patients: 56, **Ratio: 2.41**
- #Normal Samples: 584, #Murmur Samples: 180, **Ratio: 3.24**
- Implications: dataset is imbalanced -> likely strong negative impact on model performance:

Bias: models might be biased towards predicting majority class (i.e. normal diagnoses), due to higher frequency –> poor generalisation capabilities

Metrics: generally, but especially given class imbalance, accuracy can be inflated due to increased specificity (see Table 10)

Mitigation: data resampling needed

Domain: common problem in medical datasets [13, 35]

Question 2: Prepare some graphs representing basic demographic split across classes, such as sex, age, etc. Make sure you use the correct type of graph for your data to display the information intuitively.

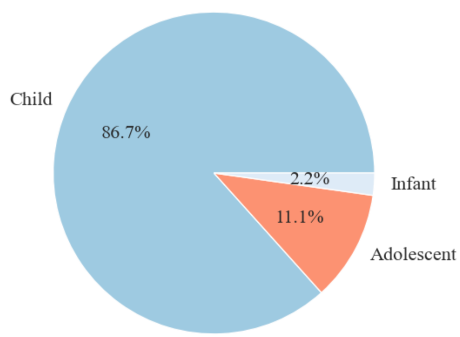

Figure 8: Age distribution of patients

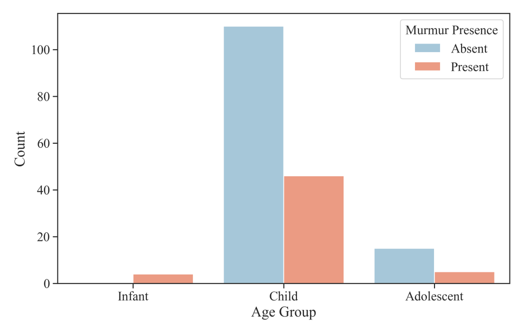

Figure 9: Age distribution of patients with and without murmur

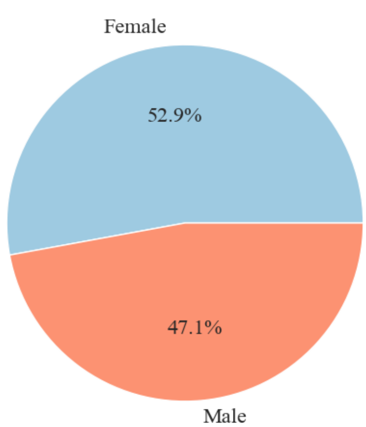

Figure 10: Sex distribution of patients

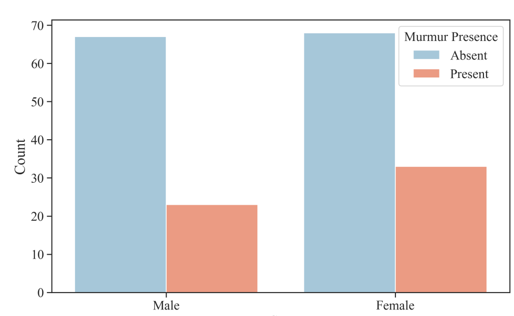

Figure 11: Sex distribution, patients with and without murmur

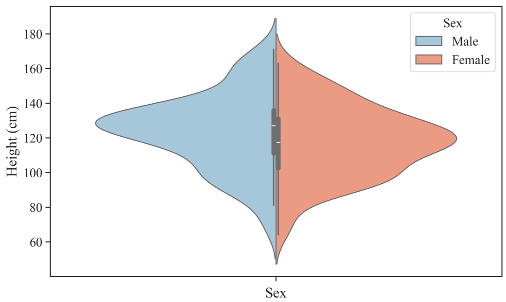

Figure 12: Height distribution by sex

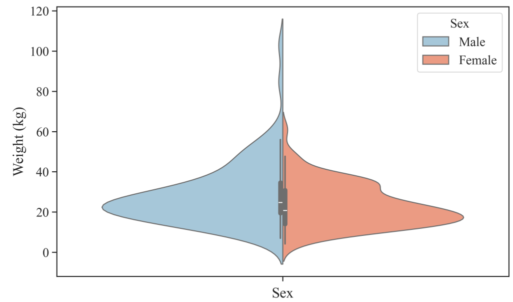

Figure 13: Weight distribution by sex

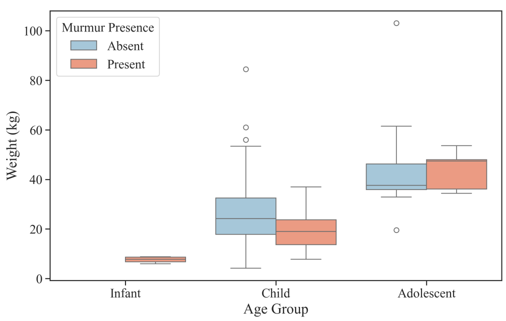

Figure 14: Weight across age groups

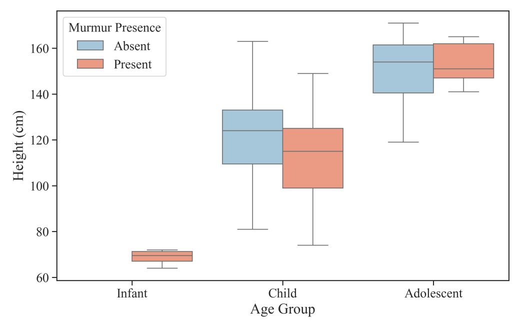

Figure 15: Height across age groups

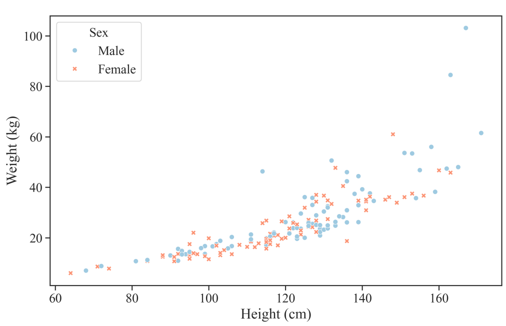

Figure 16: Height and weight development

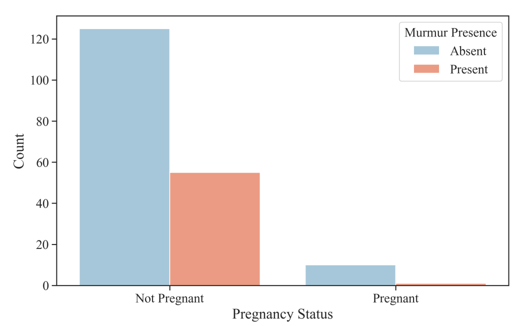

Figure 17: Pregnancy status distribution

Question 3: What significance does the demographic split carry in datasets used for ML?

- relevant for identifying potential biases in models and datasets -> if imbalanced, model might only perform well on majority class [24]
- diverse dataset is necessary for model to generalise well across gender, age, different populations etc., otherwise: bad performance for underrepresented groups [29] –> regulation and ethics: healthcare is sensitive domain and models should not discriminate against selected groups [38]
- can inform personalised treatment plans, drug development etc.

Task 2.4

Question 1: What is the effect of tackling the imbalance on the resulting classification performance? Give results to compare different methods of tackling imbalance.

- Challenges and risks to consider [24]:
  - upsampling: risk of overfitting as minority class samples are replicated
  - downsampling: potential loss of useful information as majority class is reduced
- comparison of the classification results using upsampling (Table 5) vs. downsampling (Table 6) reveals both methods allow for a balanced dataset leading to improved results
- upsampling: generally better Acc and Specificity
- downsampling: generallybetter MAcc and Sensitivity -> preferred in medical applications [30]

Part 3: Feature extraction [30 marks]

Task 3.1

Question 1: Which features did you choose and why? Use literature and/or performance assessments to inform your decisions.

Approach:

1. extensive literature review to identify relevant librosa features [51]: MFCC, Zero Crossing Rate, Chroma STFT, Spectral Centroid, Spectral Bandwith, Spectral Contrast, RMS Energy [7, 10, 25, 26, 31, 32]
2. model-based evaluation: extract feature importance using default settings (Figure 18 to Figure 20)
3. choose most promising features (Figure 21)
4. optimise feature parameters (Question 2) & re-calculate feature importance & SHAP [43] (Figure 23)
5. iteratively add features top-down, evaluate model, and pick best combination (note: risk of overfitting to test data!)

Table 3: Selection of Librosa [51] Features for Heart Sound Classification

Feature	Name	Function	Reason
RMS	Root Mean Square Energy	Measures signal energy	Reflects energy of heart sounds, helpful in detecting presence and intensity of heartbeats [7]
Chroma	Chromagram	Captures harmonic content [9]	Often used in music analysis [9]. Useful for identifying harmonic patterns within heart sounds -> might indicate pathologies [32]
ZCR	Zero Crossing Rate	Measures frequency of sign changes	Indicative of turbulence or irregularities in heart sounds, widely used [23]
Spectral Centroid	Spectral Centroid	Indicates "center of mass" of the spectrum [3]	Proven to be very successful in distinguishing between normal and murmur heart sounds [26]
Spectral Contrast	Spectral Contrast	Measures contrast in spectral peaks and valleys [4]	Distinguish between different phonological aspects of heart sounds, supporting identification of abnormal sounds [32]
Spectral Bandwidth	Spectral Bandwidth	Measures width of the spectrum (i.e. difference between upper and lower frequencies in a continuous band of frequencies)	Indication of spread of energy across frequencies, useful for detecting anomalies in heart sounds [26, 36].
MFCC	Mel Frequency Cepstrum Coefficients	Efficient representation of signal information, similar to human sound understanding on small scale	Widely used in heart sounds analysis [31, 33]

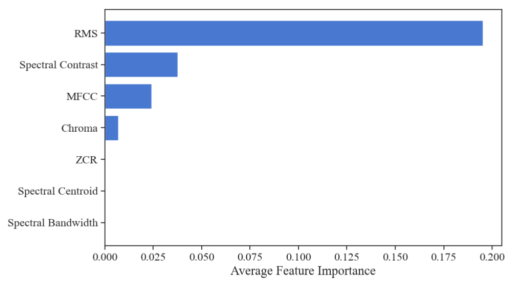

Figure 18: Decision Tree mean feature importances

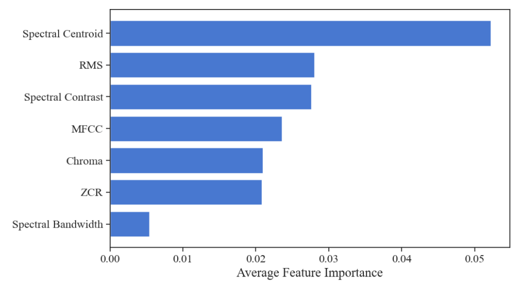

Figure 19: Random Forests mean feature importances

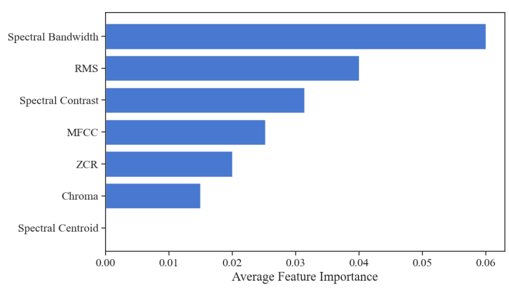

Figure 20: AdaBoost mean feature importances

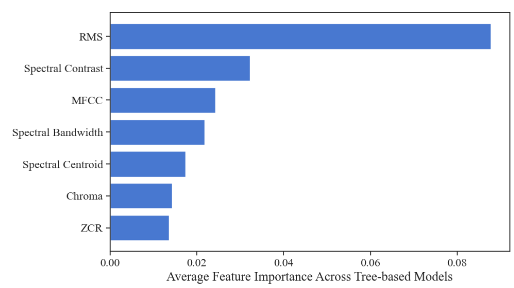

Figure 21: Mean aggregated feature importances

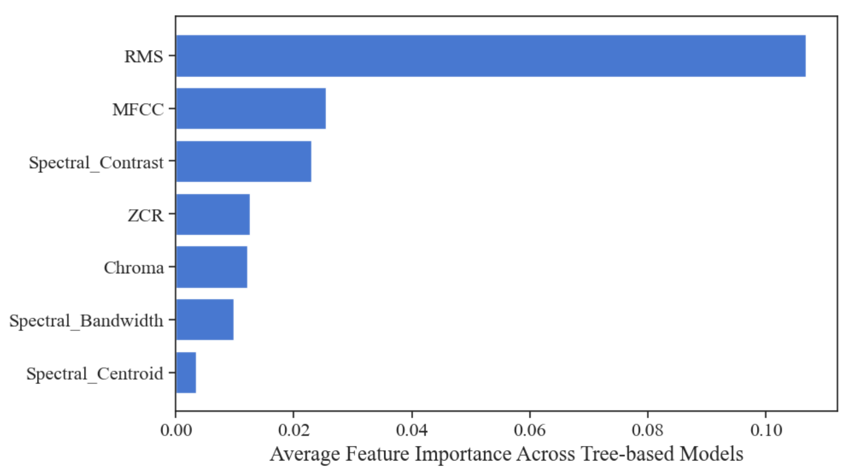

Figure 22: Mean aggregated feature importance of optimised features. In this setting, RMS is the most important feature for identifying murmurs, followed by MFCC, Spectral Contrast, Zero Crossing Rate (ZCR), Chromagram, Spectral Bandwith, and Spectral Centroid.

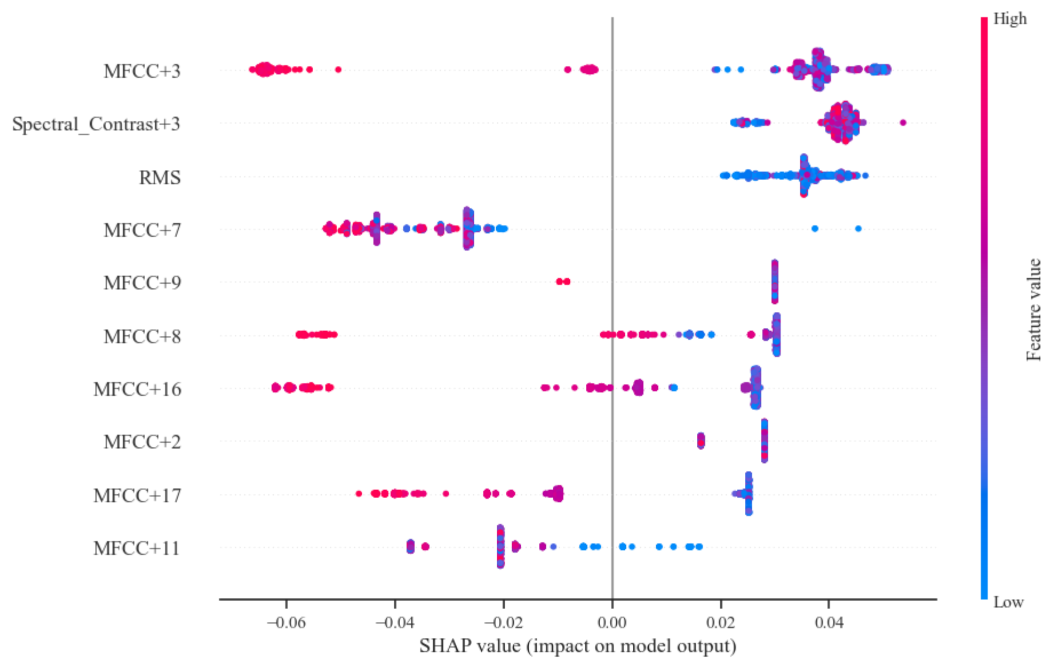

Figure 23: Top 10 features with highest SHAP (**SH**apley **A**dditive ex**P**lanations) scores. In line with Figure 22, RMS and MFCC should be considered the most relevant features for murmur classification.

Question 2: What parameters have you chosen for the features that you extracted (e.g. hop length, window size, etc.) and why?

Approach (Figure 24 to Figure 27):

1. grid search on parameters for above features
2. for each combination, calculate mean difference between distributions of feature calculated for both classes
3. pick parameter(s) generating the biggest difference
4. set best parameter for other features

Chosen parameters:

- hop_length=128 (Figure 24) & findings in Task 1.3
- n_fft=256 (Figure 25)
- fmin=50, n_bands=3 (Figure 26)
- n_chroma=24 (Figure 27)
- n_mfcc=19, informed by Yaseen et al. [25]

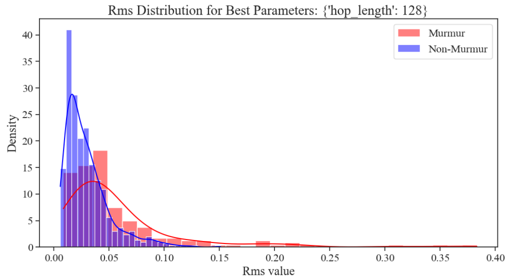

Figure 24: RMS distribution for **hop_length=128** (tested parameter values: 128, 256, 1024, 2048)

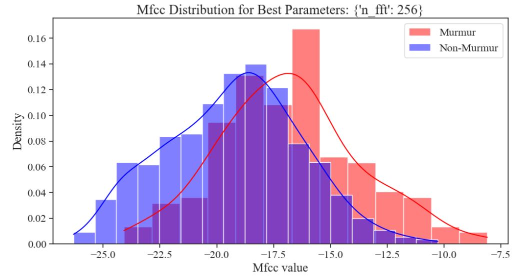

Figure 25: MFCC distribution for **n_fft=256** (tested parameter values: 128, 256, 512, 1024, 2048)

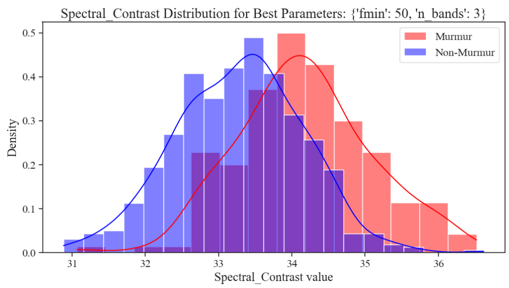

Figure 26: Spectral contrast distribution for **fmin=50, n_bands=3** (tested parameter values: fmin: [10, 20, 50], n_bands: [3, 4, 5, 6])

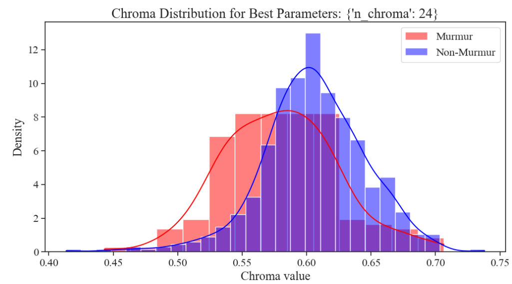

Figure 27: Chroma STFT distribution for **n_chroma=24** (tested parameter values: [12, 16, 20, 24])

Task 3.4

Question 1: Describe the full preprocessing pipeline that you used.

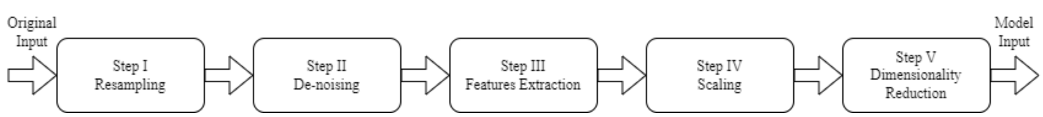

Figure 28: Full Preprocessing Pipeline, based on [33, 40]

1. Address class imbalance by upsampling or downsampling (Task 2.4)
2. De-noise data using the same BBF as in Task 1.4
3. Extract relevant features (see above)
4. Standardise features using scikit-learn's *StandardScaler* [50]
5. Reduce data dimension through 95%-Principal Component Analysis (PCA), proven to improve performance in cardiac analysis [40, 41]

Question 2: Which features or combination of features yield the best performance and why?

- Table 4 shows disadvantages and risks of using accuracy as (sole) performance metric, i.e. iterations 3 & 4 show high accuracy, but sensitivity of 0.0
- for clinical applications, such as murmur identification, sensitivity is of significant relevance -> goal: ensure patients with potential heart issues are identified (i.e. missing a true case (false negative) is more dangerous than vice-versa!) [20, 30]

Table 4: Performance metrics of best classifiers per iteration (based on Acc), **upsampled**

No.	Input Features	Best Classifier	Acc	MAcc	Se	Sp
1	rms	Naive Bayes	0.740	0.576	0.261	0.890
2	rms, mfcc	AdaBoost	0.781	0.692	**0.522**	0.863
3	rms, mfcc, spectral_contrast	RBF SVM	0.760	0.500	0.000	**1.000**
4	rms, mfcc, spectral_contrast, zero_crossing_rate	RBF SVM	0.760	0.500	0.000	**1.000**
5	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma	Neural Net	0.776	0.682	0.500	0.863
6	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth	Gaussian Process	0.776	0.674	0.478	0.870
7	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth, spectral_centroid	Gaussian Process	**0.786**	**0.696**	**0.522**	0.870

- –> evaluate model using MAcc (Mean Accuracy, arithmetic mean of sensitivity and specificity) instead [30]:

Table 5: Performance metrics of best classifiers per iteration (based on MAcc), **upsampled**

No.	Input Features	Best Classifier	Acc	MAcc	Se	Sp
1	rms	RBF SVM	0.651	0.607	0.522	0.692
2	rms, mfcc	Naive Bayes	0.714	0.693	0.652	0.733
3	rms, mfcc, spectral_contrast	Linear SVM	0.703	0.693	**0.674**	0.712
4	rms, mfcc, spectral_contrast, zero_crossing_rate	Linear SVM	0.698	0.682	0.652	0.712
5	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma	Nearest Neighbors	0.708	0.682	0.630	0.733
6	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth	Decision Tree	0.698	0.690	**0.674**	0.705
7	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth, spectral_centroid	Gaussian Process	**0.786**	**0.696**	0.522	**0.870**

Table 6: Performance metrics of best classifiers per iteration (based on **MAcc**), **downsampled**

No.	Input Features	Best Classifier	Acc	MAcc	Se	Sp
1	rms	RBF SVM	0.656	0.610	0.522	0.699
2	rms, mfcc	Neural Net	0.719	0.689	0.630	0.747
3	rms, mfcc, spectral_contrast	Naive Bayes	0.724	0.692	0.630	0.753
4	rms, mfcc, spectral_contrast, zero_crossing_rate	Naive Bayes	0.714	0.685	0.630	0.740
5	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma	Nearest Neighbors	0.651	0.689	**0.761**	0.616
6	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth	Neural Net	**0.740**	**0.702**	0.630	0.774
7	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth, spectral_centroid	Neural Net	**0.740**	0.695	0.609	**0.781**

Table 7: Features yielding the best performance in murmur classification and possible reasons for this. Based on the results summarised in Table 8, rms, mfcc, and spectral_contrast demonstrated strong performance. Broader explanations for all examined features are given in Table 3.

Feature	**Explanation for good performance**
rms	• proven to be strongly correlated with heart sound [19] • captures energy variations in heart sounds, which are indicative of abnormal heart functions [7] • sensitive to intensity differences between normal and murmur heartbeats, allowing to distinguish them [28]
mfcc	• extensively used in sound signal analysis [26] • efficient in capturing nuances in heart sounds that differentiate murmurs from normal heartbeats [25] • simulates human hearing capabilities and has performed well in a variety of tasks in sound event detection [34]
spectral_contrast	• highlights spectral peak valleys that likely separate murmurs from normal heart sounds in the frequency domain [44] • used for detecting the presence of murmurs, emphasizing the contrast between the dominant and less dominant frequencies [32, 42]

Question 3: Which classifier is yielding the best overall performance?

Table 8: Performance metrics of best classifier candidates, based on **MAcc**, **downsampled**. *AdaBoost* exhibits the best results for the models trained without PCA, while *Neural Net* performs best in the group of models trained with PCA. *Nearest Neighbors* with PCA shows the highest sensitivity (most important metric in clinical settings). Choosing *the* best model is downstream task-dependent and requires careful consideration of this tradeoff. Further models could/should be explored and optimised in more depth.

No.	Input Features	Best Classifier	Acc	**MAcc**	Se	Sp
3	rms, mfcc, spectral_contrast	**AdaBoost** without PCA	**0.755**	**0.727**	0.674	**0.781**
5	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma	**Nearest Neighbors** with PCA	0.651	0.689	**0.761**	0.616
6	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth	**Neural Net** with PCA	**0.740**	**0.702**	0.630	0.774

Question 4: What effect do individual preprocessing steps have on the final result?

Figure 29 provides an overview of the effect of individual preprocessing steps. Detailed results can be found in Appendix B.

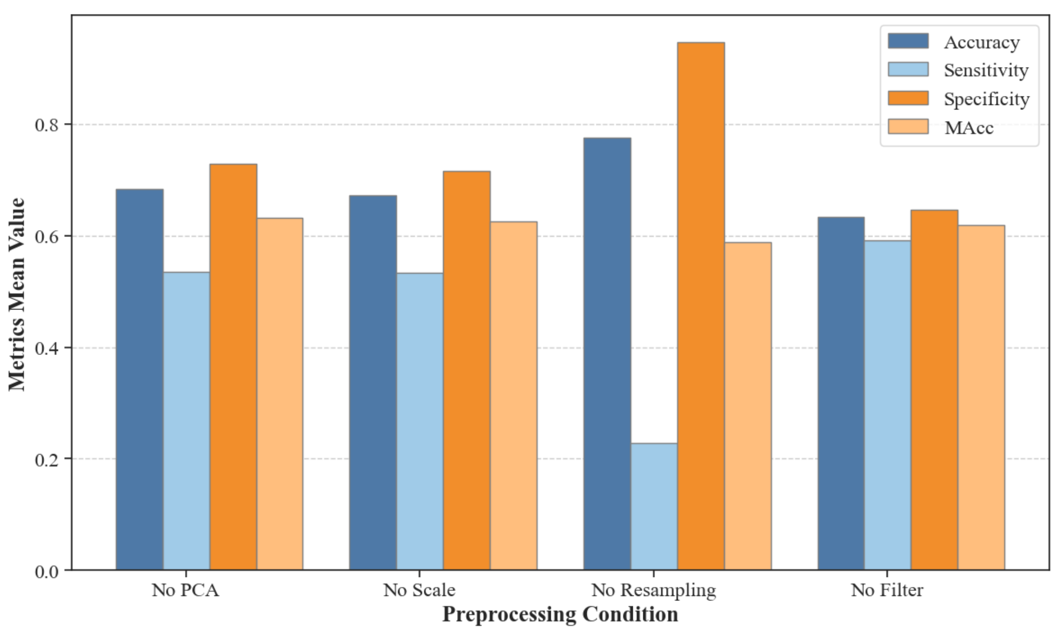

Figure 29: Comparison of model performance metrics across different preprocessing conditions. The results demonstrate the significance of tackling class imbalance. Without resampling, the models do not show high sensitivity. High accuracy in such cases stems from high specificity and the models' bias to predict the majority class. Not using PCA or scaling reduces the model performance only marginally, if at all. Detailed examination is therefore required for these preprocessing methods.Although not visible in this diagram, due to the value aggregation, not filtering the data has a negative effect on sensitivity. By omitting the filtering process, the metrics are additionally smoothed.

Part 4: Your Own Data [10 marks]

Task 4.1 (Note: two recordings were analysed to mitigate potential recording differences/errors)

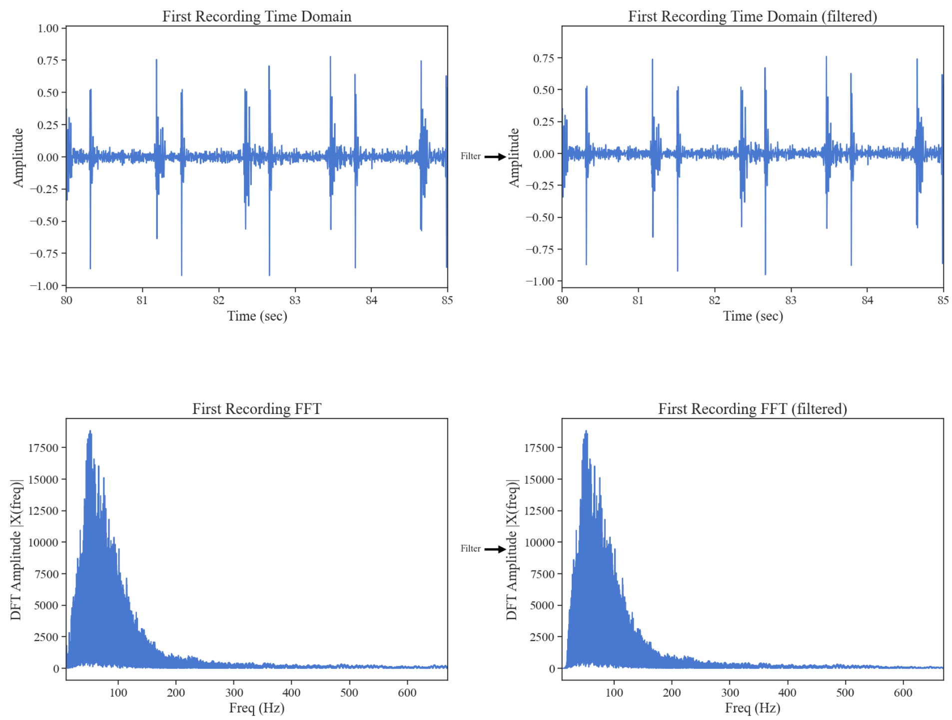

Figure 30: First Recording of Own Heart Sound Data (time domain and frequency domain). Zoomed x-axis.

Question 1: What differences are there between the frequency spectrums of your recording and the files we provided? Discuss why there might be differences.

Differences:

- own recordings: seem less noisy, exhibit relatively stronger concentration of lower frequencies -> majority of signal consists of frequencies < 200 Hz, contrasting provided files, especially with murmur
- visible differences in overall amplitude (own recordings generally stronger)
- own recordings: sampling rate=48,000, cover wide frequency range

Reasons: different...

- ...physiology of recorded individual
- ...recording setup (i.e. different locations)
- ...recording device (i.e. iPhone vs. digital stethoscope [48])
- ...environment -> noise

983 words (excluding bibliography, headers, and captions)

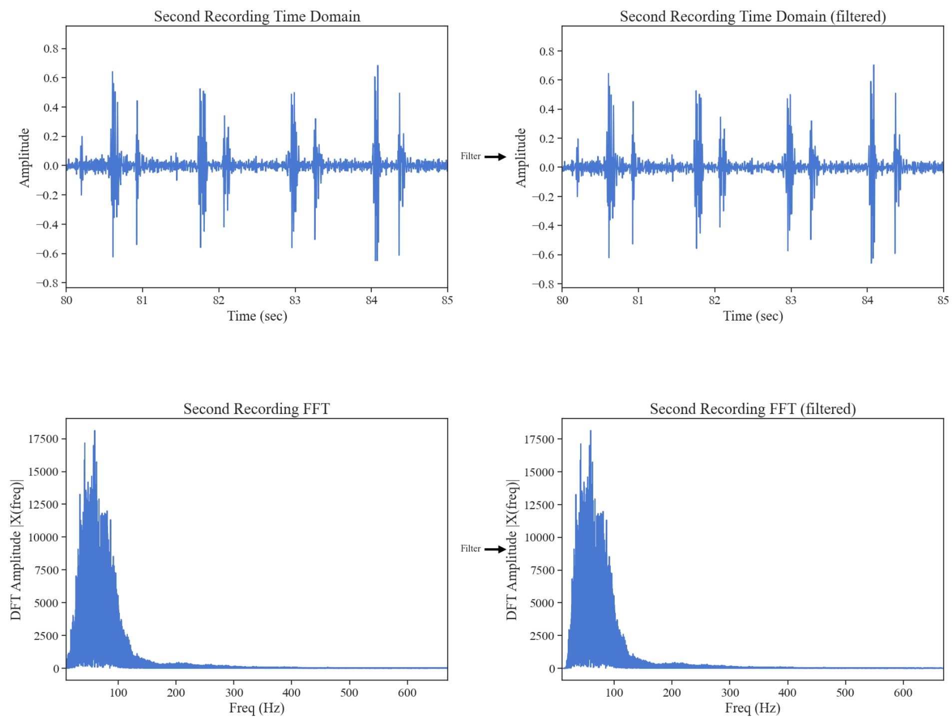

Figure 31: Second recording of own heart sound data (time domain and frequency domain). Zoomed x-axis.

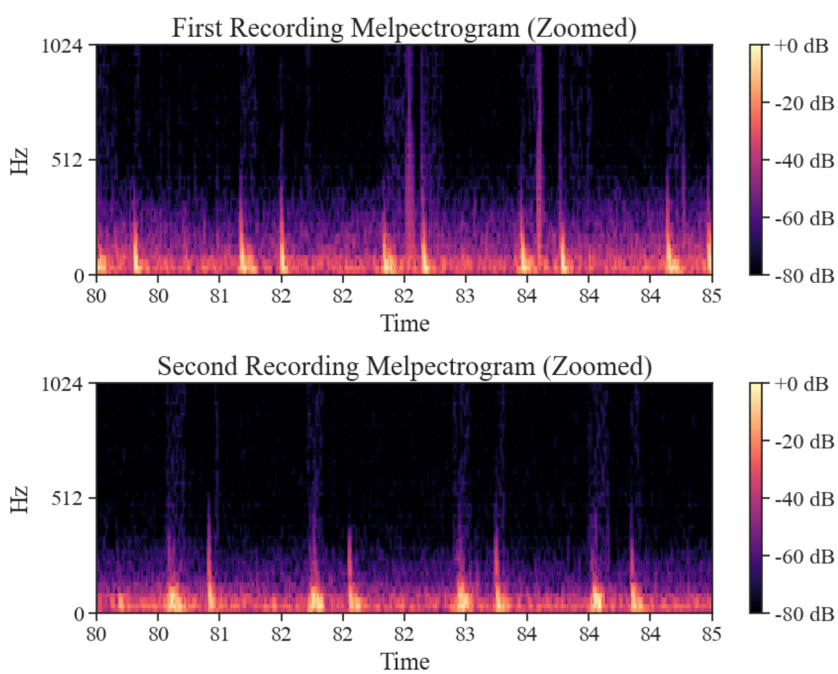

Figure 32: Melspectograms of heart sound recordings

References

[1] G. Charbonneau and M. Sudraud. "[Measurement of the frequency response of several stethoscopes in common use. Consequences for cardiac and pulmonary auscultation]". In: *Bulletin Europeen De Physiopathologie Respiratoire* 21.1 (1985), pp. 49–54. ISSN: 0395-3890.

[2] Richard L. Donnerstein. "Continuous spectral analysis of heart murmurs for evaluating stenotic cardiac lesions". In: *The American Journal of Cardiology* 64.10 (Sept. 1989), pp. 625–630. ISSN: 0002-9149. DOI: 10.1016/0002-9149(89)90491-8.

[3] K.K. Paliwal. "Spectral subband centroid features for speech recognition". In: *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*. Vol. 2. Seattle, WA, USA: IEEE, 1998, pp. 617–620. ISBN: 978-0-7803-4428-0. DOI: 10.1109/ICASSP.1998.675340.

[4] Dan-Ning Jiang, Lie Lu, Hong-Jiang Zhang, Jian-Hua Tao, and Lian-Hong Cai. "Music type classification by spectral contrast feature". In: *Proceedings. IEEE International Conference on Multimedia and Expo*. Vol. 1. 2002, 113–116 vol.1. DOI: 10.1109/ICME.2002.1035731.

[5] Nancy J. Rennert, Rebecca Morris, and Candice C. Barrere. *How to Cope with Scopes: Stethoscope Selection and Use with Hearing Aids and CIs*. Feb. 2004. URL: https://hearingreview.com/practice-building/practice-management/how-to-cope-with-scopes-stethoscope-selection-and-use-with-hearing-aids-and-cis (visited on 02/16/2024).

[6] M. El-Segaier, O. Lilja, S. Lukkarinen, L. Sörnmo, R. Sepponen, and E. Pesonen. "Computer-Based Detection and Analysis of Heart Sound and Murmur". In: *Annals of Biomedical Engineering* 33.7 (July 2005), pp. 937–942. ISSN: 1573-9686. DOI: 10.1007/s10439-005-4053-3.

[7] Christer Ahlstrom, Peter Hult, Peter Rask, Jan-Erik Karlsson, Eva Nylander, Ulf Dahlström, and Per Ask. "Feature Extraction for Systolic Heart Murmur Classification". In: *Annals of Biomedical Engineering* 34.11 (Nov. 2006), pp. 1666–1677. ISSN: 1573-9686. DOI: 10.1007/s10439-006-9187-4.

[8] Abdel-Ouahab Boudraa and Jean-Christophe Cexus. "EMD-Based Signal Filtering". In: *IEEE Transactions on Instrumentation and Measurement* 56.6 (Dec. 2007), pp. 2196–2202. ISSN: 0018-9456. DOI: 10.1109/TIM.2007.907967.

[9] Dan Ellis. *Chroma Feature Analysis and Synthesis*. Apr. 2007. URL: https://www.ee.columbia.edu/~dpwe/resources/matlab/chroma-ansyn/ (visited on 02/18/2024).

[10] Vladimir Kudriavtsev, Vladimir Polyshchuk, and Douglas L Roy. "Heart energy signature spectrogram for cardiovascular diagnosis". In: *BioMedical Engineering OnLine* 6 (May 2007), p. 16. ISSN: 1475-925X. DOI: 10.1186/1475-925X-6-16.

[11] S. M. Debbal and F. Bereksi-Reguig. "Frequency analysis of the heartbeat sounds(<Special Issue>Contribution to 21 Century Intelligent Technologies and Bioinformatics)". In: *International Journal of Biomedical Soft Computing and Human Sciences: the official journal of the Biomedical Fuzzy Systems Association* 13.1 (2008), pp. 85–90. DOI: 10.24466/ijbschs.13.1_85.

[12] Kihong Shin and Joseph Hammond. *Fundamentals of signal processing for sound and vibration engineers*. Wiley, Feb. 2008. ISBN: 978-0-470-51188-6. URL: https://eprints.soton.ac.uk/63688/.

[13] Der-Chiang Li, Chiao-Wen Liu, and Susan C. Hu. "A learning method for the class imbalance problem with medical data sets". In: *Computers in Biology and Medicine* 40.5 (2010), pp. 509–518. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2010.03.005.

[14] Steven McGee. "Etiology and Diagnosis of Systolic Murmurs in Adults". In: *The American Journal of Medicine* 123.10 (Oct. 2010), 913–921.e1. ISSN: 0002-9343. DOI: 10.1016/j.amjmed.2010.04.027.

[15] Véronique Millette and Natalie Baddour. "Signal processing of heart signals for the quantification of non-deterministic events". In: *BioMedical Engineering OnLine* 10.1 (Jan. 2011), p. 10. ISSN: 1475-925X. DOI: 10.1186/1475-925X-10-10.

[16] Božo Tomas, Darko Zelenika, Željko Rončević, and Antonija Krtalić. *Classification of Pathologic and Innocent Heart Murmur Based on Multimedia Presentations of Acoustic Heart Signals,* The Third International Conference on Creative Content Technologies. Sept. 2011. ISBN: 978-1-61208-157-1.

[17] Babatunde S. Emmanuel. "A review of signal processing techniques for heart sound analysis in clinical diagnosis". In: *Journal of Medical Engineering & Technology* 36.6 (Aug. 2012). Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/03091902.2012.684831, pp. 303–307. ISSN: 0309-1902. DOI: `10.3109/03091902.2012.684831`.

[18] M. Markaki, I. Germanakis, and Y. Stylianou. "Automatic classification of systolic heart murmurs". In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver, BC, Canada: IEEE, May 2013, pp. 1301–1305. ISBN: 978-1-4799-0356-6. DOI: `10.1109/ICASSP.2013.6637861`. (Visited on 02/16/2024).

[19] Young Duck Shin, Kyoung Hoon Yim, Sang Hi Park, Yong Wook Jeon, Jin Ho Bae, Tae Soo Lee, Myoung Hwan Kim, and Young Jin Choi. "The correlation between the first heart sound and cardiac output as measured by using digital esophageal stethoscope under anaesthesia". In: *Pakistan Journal of Medical Sciences* 30.2 (2014), pp. 276–281. ISSN: 1682-024X. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3998993/` (visited on 02/18/2024).

[20] Grzegorz Redlarski, Dawid Gradolewski, and Aleksander Palkowski. "A System for Heart Sounds Classification". In: *PLoS ONE* 9.11 (Nov. 2014). Ed. by Alena Talkachova, e112673. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0112673`.

[21] Mani Arsalan, Thomas Walther, Robert L. Smith, and Paul A. Grayburn. "Tricuspid regurgitation diagnosis and treatment". In: *European Heart Journal* (Sept. 2015), ehv487. ISSN: 0195-668X, 1522-9645. DOI: `10.1093/eurheartj/ehv487`.

[22] Cara S. Spencer and Karen Pennington. "Nurses with Undiagnosed Hearing Loss: Implications for Practice". In: *Online Journal of Issues in Nursing* 20.1 (Jan. 2015), p. 6. ISSN: 1091-3734.

[23] Fatima Chakir, Abdelilah Jilbab, Chafik Nacir, and Ahmed Hammouch. "Phonocardiogram signals processing approach for PASCAL Classifying Heart Sounds Challenge". In: *Signal, Image and Video Processing* 12.6 (Sept. 2018), pp. 1149–1155. ISSN: 1863-1711. DOI: `10.1007/s11760-018-1261-5`.

[24] Joffrey L. Leevy, Taghi M. Khoshgoftaar, Richard A. Bauder, and Naeem Seliya. "A survey on addressing high-class imbalance in big data". In: *Journal of Big Data* 5.1 (Nov. 2018), p. 42. ISSN: 2196-1115. DOI: `10.1186/s40537-018-0151-6`.

[25] Yaseen, Gui-Young Son, and Soonil Kwon. "Classification of Heart Sound Signal Using Multiple Features". In: *Applied Sciences* 8.12 (Dec. 2018). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 2344. ISSN: 2076-3417. DOI: `10.3390/app8122344`.

[26] Munia Ferdoushi, Mala Paul, and Shaikh Anowarul Fattah. "A Spectral Centroid Based Analysis of Heart sounds for Disease Detection Using Machine Learning". In: *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*. Nov. 2019, pp. 1–6. DOI: `10.1109/WIECON-ECE48653.2019.9019975`.

[27] Ali Raza, Arif Mehmood, Saleem Ullah, Maqsood Ahmad, Gyu Sang Choi, and Byung-Won On. "Heartbeat Sound Signal Classification Using Deep Learning". In: *Sensors* 19.21 (Jan. 2019). Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, p. 4819. ISSN: 1424-8220. DOI: `10.3390/s19214819`.

[28] J.K. Roy, T.S. Roy, and S.C. Mukhopadhyay. "Heart Sound: Detection and Analytical Approach Towards Diseases". English. In: *Smart Sensors, Measurement and Instrumentation* 29 (2019), pp. 103–145. ISSN: 2194-8402. DOI: `10.1007/978-3-319-99540-3_7`.

[29] Selen Bozkurt, Eli M Cahan, Martin G Seneviratne, Ran Sun, Juan A Lossio-Ventura, John P A Ioannidis, and Tina Hernandez-Boussard. "Reporting of demographic data and representativeness in machine learning models using electronic health records". In: *Journal of the American Medical Informatics Association : JAMIA* 27.12 (Sept. 2020), pp. 1878–1884. ISSN: 1067-5027. DOI: `10.1093/jamia/ocaa164`.

[30] Fan Li, Hong Tang, Shang Shang, Klaus Mathiak, and Fengyu Cong. "Classification of Heart Sounds Using Convolutional Neural Network". In: *Applied Sciences* 10.11 (Jan. 2020). Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, p. 3956. ISSN: 2076-3417. DOI: `10.3390/app10113956`.

[31] Khalid M.O. Nahar, Obaida M. Al-Hazaimeh, Ashraf Abu-Ein, and Nasr Gharaibeh. "Phonocardiogram Classification Based on Machine Learning with Multiple Sound Features". In: *Journal of Computer Science* 16.11 (Nov. 2020), pp. 1648–1656. ISSN: 1549-3636. DOI: 10.3844/jcssp.2020.1648.1656.

[32] Balagopal Unnikrishnan, Pranshu Ranjan Singh, Xulei Yang, and Matthew Chin Heng Chua. *Semi-supervised and Unsupervised Methods for Heart Sounds Classification in Restricted Data Environments*. arXiv:2006.02610 [cs]. June 2020. URL: http://arxiv.org/abs/2006.02610.

[33] Wei Chen, Qiang Sun, Xiaomin Chen, Gangcai Xie, Huiqun Wu, and Chen Xu. "Deep Learning Methods for Heart Sounds Classification: A Systematic Review". In: *Entropy* 23.6 (May 2021), p. 667. ISSN: 1099-4300. DOI: 10.3390/e23060667.

[34] Haoran Kui, Jiahua Pan, Rong Zong, Hongbo Yang, and Weilian Wang. "Heart sound classification based on log Mel-frequency spectral coefficients features and convolutional neural networks". In: *Biomedical Signal Processing and Control* 69 (Aug. 2021), p. 102893. ISSN: 1746-8094. DOI: 10.1016/j.bspc.2021.102893. URL: https://www.sciencedirect.com/science/article/pii/S1746809421004900 (visited on 02/18/2024).

[35] Uddipan Mukherjee and Sidharth Pancholi. *Heartbeat Sound Classification with Visual Domain Deep Neural Networks*. arXiv:2107.13237 [cs, eess]. Oct. 2021. URL: http://arxiv.org/abs/2107.13237 (visited on 02/16/2024).

[36] Ahmed Ali Dawud, Thamineni Bheema Lingaiah, and Towfik Jemal. "Classification of heart sounds associated with murmur for diagnosis of cardiac valve disorders". en. In: *Health Technology* 6.0 (July 2022). Number: 0 Publisher: AME Publishing Company. ISSN: 2616-2717. DOI: 10.21037/ht-21-23. URL: https://ht.amegroups.org/article/view/7549 (visited on 02/18/2024).

[37] Hassaan Malik, Umair Bashir, and Adnan Ahmad. "Multi-classification neural network model for detection of abnormal heartbeat audio signals". In: *Biomedical Engineering Advances* 4 (Dec. 2022), p. 100048. ISSN: 2667-0992. DOI: 10.1016/j.bea.2022.100048.

[38] Nithesh Naik, B. M. Zeeshan Hameed, Dasharathraj K. Shetty, Dishant Swain, Milap Shah, Rahul Paul, Kaivalya Aggarwal, Sufyan Ibrahim, Vathsala Patil, Komal Smriti, Suyog Shetty, Bhavan Prasad Rai, Piotr Chlosta, and Bhaskar K. Somani. "Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?" In: *Frontiers in Surgery* 9 (2022). ISSN: 2296-875X. URL: https://www.frontiersin.org/articles/10.3389/fsurg.2022.862322.

[39] Raj Shekhar, Ganesh Vanama, Titus John, James Issac, Youness Arjoune, and Robin Doroshow. "Automated Identification of Innocent Still's Murmur Using a Convolution Neural Network". In: *Frontiers in Pediatrics* (Sept. 2022). DOI: https://doi.org/10.3389/fped.2022.923956.

[40] Yasser Zeinali and Seyed Taghi Akhavan Niaki. "Heart sound classification using signal processing and machine learning algorithms". In: *Machine Learning with Applications* 7 (Mar. 2022), p. 100206. ISSN: 2666-8270. DOI: 10.1016/j.mlwa.2021.100206.

[41] George Zhou, Yunchan Chen, and Candace Chien. "On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks". In: *BMC Medical Informatics and Decision Making* 22 (Aug. 2022), p. 226. ISSN: 1472-6947. DOI: 10.1186/s12911-022-01942-2.

[42] Mihai-Andrei Costandache, Matei-Alexandru Cioată, and Adrian Iftene. "Automated Heart Murmur Detection using Sound Processing Techniques". In: *Procedia Computer Science*. 27th International Conference on Knowledge Based and Intelligent Information and Engineering Sytems (KES 2023) 225 (Jan. 2023), pp. 2961–2970. ISSN: 1877-0509. DOI: 10.1016/j.procs.2023.10.289. URL: https://www.sciencedirect.com/science/article/pii/S1877050923014473 (visited on 02/19/2024).

[43] Chenyang Xu, Xin Li, Xinyue Zhang, Ruilin Wu, Yuxi Zhou, Qinghao Zhao, Yong Zhang, Shijia Geng, Yue Gu, and Shenda Hong. "Cardiac murmur grading and risk analysis of cardiac diseases based on adaptable heterogeneous-modality multi-task learning". In: *Health Information Science and Systems* 12.1 (Dec. 2023), p. 2. ISSN: 2047-2501. DOI: 10.1007/s13755-023-00249-4.

[44] Sidra Abbas, Stephen Ojo, Abdullah Al Hejaili, Gabriel Avelino Sampedro, Ahmad Almadhor, Monji Mohamed Zaidi, and Natalia Kryvinska. "Artificial intelligence framework for heart disease classification from audio signals". en. In: *Scientific Reports* 14.1 (Feb. 2024), p. 3123. ISSN: 2045-2322. DOI: 10.1038/s41598-024-53778-7. URL: https://www.nature.com/articles/s41598-024-53778-7 (visited on 02/19/2024).

[45] Christer Ahlstrom. "Nonlinear Phonocardiographic Signal Processing". In: ().

[46] Bankaitis. *Amp Steth HA Programming*. URL: https://www.oaktreeproducts.com/amp-steth-ha-programming (visited on 02/16/2024).

[47] Michael A. Chen. *Heart Murmur - Symptoms and Causes*. URL: https://www.pennmedicine.org/for-patients-and-visitors/patient-information/conditions-treated-a-to-z/heart-murmur (visited on 02/15/2024).

[48] *George B. Moody PhysioNet Challenge*. en-US. URL: https://moody-challenge.physionet.org/2022/ (visited on 02/17/2024).

[49] *Heart murmurs - causes, symptoms & treatment*. en. URL: https://www.bhf.org.uk/informationsupport/conditions/heart-murmurs (visited on 02/18/2024).

[50] skicit learn. *sklearn.preprocessing.StandardScaler*. URL: https://scikit-learn/stable/modules/generated/sklearn.preprocessing.StandardScaler.html (visited on 02/17/2024).

[51] librosa. *Feature extraction — librosa 0.10.1 documentation*. URL: https://librosa.org/doc/main/feature.html (visited on 02/17/2024).

Appendices

A Best performing models based on an evaluation of Sensitivity measures

Table 9: Performance metrics of best classifiers per iteration (based on Sensitivity), **downsampled**. Although the sensitivity values are notably high, values for Accuracy, MAcc, and Specificity are not. This underscores the importance of a balanced performance assessment of classifiers against the requirements of the downstream task in the context of the application.

No.	Input Features	Best Classifier	Acc	MAcc	Se	Sp
1	rms	Nearest Neighbors	0.552	0.549	0.543	0.555
2	rms, mfcc	RBF SVM	0.385	0.536	0.826	0.247
3	rms, mfcc, spectral_contrast	RBF SVM	0.271	0.498	0.935	0.062
4	rms, mfcc, spectral_contrast, zero_crossing_rate	RBF SVM	0.281	0.505	0.935	0.075
5	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma	Nearest Neighbors	0.651	0.689	0.761	0.616
6	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth	Nearest Neighbors	0.641	0.667	0.717	0.616
7	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth, spectral_centroid	Nearest Neighbors	0.641	0.667	0.717	0.616

B Detailed results for the performance of models under different preprocessing conditions

Table 10: Performance metrics of best classifiers per iteration, based on MAcc, **without resampling**. The results clearly indicate the effect of the imbalanced dataset. Without resampling, the highest sensitivity stands at 50%, while a Specificity of up to 96% is reached. In this case, the models are heavily biased towards predicting the majority class (i.e. non-murmur). This underscores the significance of tackling class imbalance in medical datasets.

No.	Input Features	Best Classifier	Acc	MAcc	Se	Sp
1	rms	QDA	0.771	0.574	0.196	0.952
2	rms, mfcc	Nearest Neighbors	0.797	0.688	0.478	0.897
3	rms, mfcc, spectral_contrast	Decision Tree	0.807	0.642	0.326	0.959
4	rms, mfcc, spectral_contrast, zero_crossing_rate	AdaBoost	0.781	0.648	0.391	0.904
5	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma	Nearest Neighbors	0.807	0.702	0.500	0.904
6	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth	Nearest Neighbors	0.786	0.681	0.478	0.884
7	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth, spectral_centroid	Nearest Neighbors	0.792	0.684	0.478	0.890

Table 11: Performance metrics of best classifiers per iteration, based on MAcc, **downsampled, without filtering**. Not filtering the audio signal has a substantial negative effect on Sensitivity.

No.	Input Features	Best Classifier	Acc	MAcc	Se	Sp
1	rms	RBF SVM	0.688	0.616	0.478	0.753
2	rms, mfcc	Linear SVM	0.745	0.683	0.565	0.801
3	rms, mfcc, spectral_contrast	Linear SVM	0.771	0.715	0.609	0.822
4	rms, mfcc, spectral_contrast, zero_crossing_rate	Linear SVM	0.745	0.691	0.587	0.795
5	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma	Gaussian Process	0.750	0.709	0.630	0.788
6	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth	Gaussian Process	0.740	0.695	0.609	0.781
7	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth, spectral_centroid	Gaussian Process	0.750	0.709	0.630	0.788

Table 12: Performance metrics of best classifiers per iteration, based on MAcc, **downsampled, without Scaling, with PCA**. Without scaling the features, the performance slightly decreases when compared to the best-performing models. Still, the effect is relatively small, and the results are still comparably competitive.

No.	Input Features	Best Classifier	Acc	MAcc	Se	Sp
1	rms	Neural Net	0.693	0.597	0.413	0.781
2	rms, mfcc	Naive Bayes	0.708	0.704	0.696	0.712
3	rms, mfcc, spectral_contrast	QDA	0.719	0.711	0.696	0.726
4	rms, mfcc, spectral_contrast, zero_crossing_rate	QDA	0.719	0.711	0.696	0.726
5	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma	QDA	0.719	0.711	0.696	0.726
6	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth	Naive Bayes	0.693	0.694	0.696	0.692
7	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth, spectral_centroid	AdaBoost	0.698	0.697	0.696	0.699

Table 13: Performance metrics of best classifiers per iteration, based on MAcc, **downsampled, without PCA**. While not applying PCA seems to generally hurt accuracy, there seems to be no major decline in MAcc, Sensitivity and Specificity. In fact, the *AdaBoost* model in iteration 3 performs comparably well.

No.	Input Features	Best Classifier	Acc	MAcc	Se	Sp
1	rms	RBF SVM	0.661	0.614	0.522	0.705
2	rms, mfcc	Gaussian Process	0.745	0.698	0.609	0.788
3	rms, mfcc, spectral_contrast	AdaBoost	0.755	0.727	0.674	0.781
4	rms, mfcc, spectral_contrast, zero_crossing_rate	AdaBoost	0.755	0.727	0.674	0.781
5	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma	Naive Bayes	0.740	0.695	0.609	0.781
6	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth	AdaBoost	0.724	0.722	0.717	0.726
7	rms, mfcc, spectral_contrast, zero_crossing_rate, chroma, spectral_bandwidth, spectral_centroid	AdaBoost	0.724	0.722	0.717	0.726