Classification Of Systolic Heart Murmurs of Children Using Dynamic Time Warping and K-Nearest Neighbours

Ben Alexander, *Music Signal Analysis Student, University of York,*

CONTENTS

I. INTRODUCTION

HEART murmurs in children are incredibly common (up to 80% [1]) and can imply many heart conditions. Although most of these cases are completely harmless (innocent), some may be a sign of an underlying problem with a heart or blood flowing through the heart faster than usual. The former is in definite need of being checked, however, the latter can be explained by growing too quickly, an overactive thyroid, having a fever or a more serious condition like anaemia or untreated high blood pressure. Recognising the signs of a heart murmur is challenging, given their often generic symptoms including dizziness, shortness of breath, chest pain, and fatigue, which could go undiagnosed into or after their adolescent years [2]. Accompanying these less extreme symptoms, people who suffer from heart palpitations can experience swelling, continuous coughing, and growth problems. Although some of these patients may have innocent heart murmurs with no underlying heart problems, doctors can also prescribe treatments to some cases to manage the effect of the symptoms on the individual[3].

This model hopes to allow children to be routinely pre-screened for heart murmurs before seeking help from professionals on whether the murmur is a harmless sound or a sign of a more severe condition. Not only would this streamline the diagnostic process and optimise resource allocation on strained health services, but it could also catch cases of abnormal heart murmurs intervening before the potential condition can worsen, and allow doctors to give the right treatments to manage symptoms.

This report will introduce a novel approach to classifying systolic heart murmurs in children, by using a combination of dynamic time warping and K-Nearest-Neighbours. It will describe and evaluate the efficacy of the model based on current approaches to this classification task.

A. Background

Heart murmurs can be characterised as an extra sound in the heart caused by turbulent blood flow through a valve [4]. These extra sounds can be heard throughout the whole cardiac cycle, named a continuous murmur, or only in certain phases of the cardiac cycle.

Systolic heart murmurs can be heard in the contraction phase of the heart and can be subdivided into ejection murmurs and regurgitant murmurs. The systolic phase of the heart cycle is between the strong and weaker beats making up the "Lub-Dub"(*S1*-*S2*) sound heard (as shown in Fig. 1). Ejection murmurs are caused by narrow vessels or an irregular valve, and regurgitant murmurs are categorised by the backward flow of blood into chambers of the heart[5]

Diastolic heart murmurs, however, are found after the weaker beat of the heart or between the *S2* to the next cycle's *S1* (as shown in Fig. 1), when the heart is in relaxation. These can be categorized by a narrowing of the mitral or tricuspid valves, or regurgitation by the aortic or pulmonary valves[5]. Since this project is solely based on the classification of systolic heart murmurs, this report will not go into any more depth on the diastolic murmurs, as if this model is successful in classifying systolic murmurs, the diastolic classification should follow simply.

Along with being categorised by the timing, the murmurs can also be categorised by shape (as shown in Fig. 2) and grading from I to V[6]:

- Grade I - Not immediately heard
- Grade II - Soft, but immediately heard
- Grade III - Loud, but no thrill
- Grade IV - Associated with a thrill
- Grade V - Heard with the edge of the tilted stethoscope
- Grade VI - Heard with the stethoscope lifted away from the chest wall

All of these characteristics as well as pitch and quality are represented in the dataset chosen for this classification model.

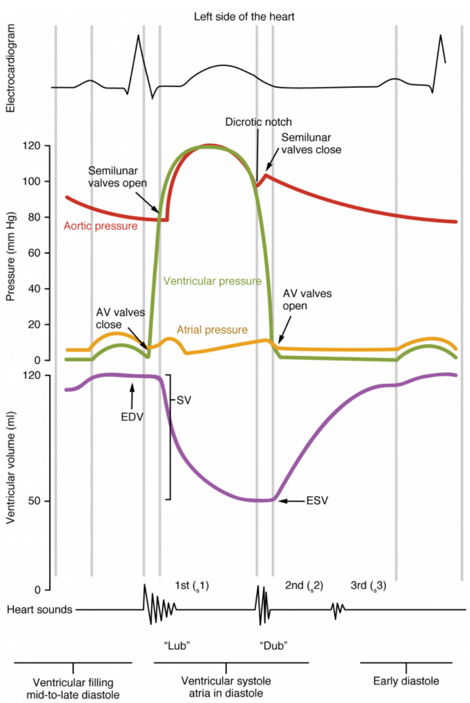

Fig. 1. The cardiac cycle[7].

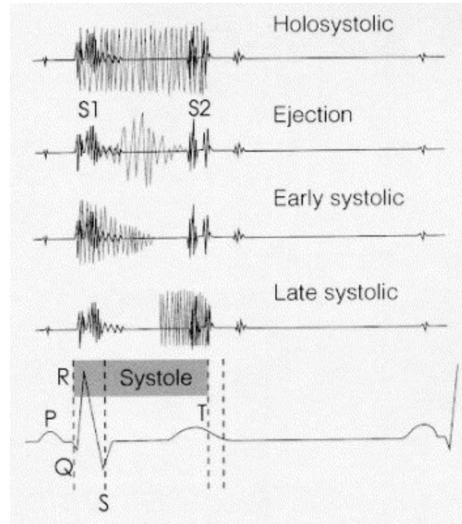

Fig. 2. Systolic murmur shapes[8].

B. Dataset

The dataset "The CirCor DigiScope Phonocardiogram Dataset" from PhysioNet is a vast dataset of 1568 patients from age 0 to 21. The dataset includes both healthy individuals and those with both systolic and diastolic heart murmurs. Audio recordings were taken of the patients at four areas located on the chest surrounding the heart. These four areas are the areas associated with the four valves of the heart (see Fig. 3). The subjects' data were recorded as part of two screening campaigns in Brazil in 2014 and 2015, and were examined by an expert paediatric cardiologist. When needed, the experts could request to see the patient in person, if there were any doubts before coming to a diagnosis [9].

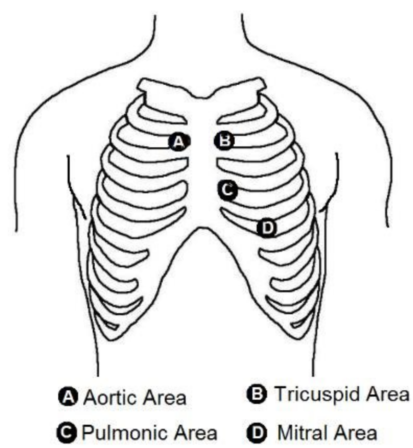

Fig. 3. Four areas for recording or listening to heart murmurs[10].

The data is represented in audio (wav) files and text files containing the patient data and diagnosis information. The data recorded from patients are age (in categories shown in Table I), sex, type of murmur (nan, systolic, or diastolic), locations of where this murmur is heard (the four valve positions), and the outcome of the diagnosis (Normal/Abnormal). Along with this, there is some other information about the characteristics of the murmur, which is less important for this project as this aims only to sense if there is a murmur present.

The vast size of the dataset is a huge benefit for this type of classification model, however, the audio

Age Label	Numerical Value
Neonate	0 to 27 days
Infant	28 days to 1 year
Child	1 to 11 years
Adolescent	12 to 18 years
Young Adult	19 to 21 years

TABLE I

AGE LABEL COMPARED TO NUMERIC VALUE OF DATASET [9]

files themselves are of varying quality. This is, however, real-world data, and for the model to work in real applications, there is an expectation for it to perform well under varying and overly noisy environments. The audio data too is standardised to have a sampling rate of 4kHz as most audio information above the Nyquist frequency [1] (2kHz) can be neglected for these recordings[8]. Despite the sampling rate being standardised, the length of the recorded data is vastly varied, and the noise of the signals is inconsistent even within samples. The noise originates from several sources noted in the dataset, including the stethoscope rubbing on skin, speech, crying and laughing sounds[9].

With this vast dataset of varying quality, the focus is now to make sense of the varying quality in terms of audio preparation and model design.

II. MODEL DESIGN

The design process of this model is split into two separate stages: audio preparation, and classification. The large amount of audio data (above 400MB) needs to be made comprehensible, by negating all of the overly noisy data that may influence the classification step, and standardise the length of the audio. Due to processing power constraints, there have been some necessary measures to improve efficiency, however, this introduces a trade-off in the model's accuracy against the speed at which it can classify the test set.

A. Data and Audio Preparation

Dataset preparation is extremely important, as although "all entries were screened for incorrectly entered or measured values, inconsistent data or outliers, and deleted as appropriate"[9], there were still some classifications by professionals, such as an "Unknown" diagnosis, that could hinder the performance of the model.

As mentioned in the previous section, the patient's data is stored in text files with references to the audio files accompanying them. The diagnosis of the patient is given as "Abnormal" or "Normal", and the murmurs are given as "Present" or "Absent". The categories of "Present-Normal" and "Absent-Abnormal" are too vast in the dataset to ignore, as these quantities are above the human error rate of 1% at 3.08% and 27.92% respectively [12]. Therefore, in this classifier, the model will group any "Present-Normal" or "Absent-Abnormal" with the "Present-Abnormal" as this is proposed as a preliminary screening before expert diagnosis. On top of that, there are 7.22% unclassified murmur types which have been removed altogether from the training and test set. The data also has mostly the child age group of 1 to 11 years at 70.49% and an almost even sex split of 51.59% to 48.41% (Female to Male).

Following the flowchart Fig. A.1, the text files for all the patients are collected and filtered for only the Children (aged 1-11 years) and the sex of the patients under test. This filtering also excludes any "Unknown" or unclassifiable records from the dataset. When the patient's text files are obtained, the bias of classified positive to negative is negated so that there is a 50/50 split of "Absent-Normal" diagnoses to an otherwise positive classification. This bias is removed at random and will only attempt to make the patient data equal in quantity, there will be a small amount of random bias left (as can be seen in Fig. A.5, and A.6). The dataset is then split randomly into 85% training samples and 15% testing samples. Note that this is the percentage of the patient data, not the audio files or the cardiac periods extracted from the files. For each of the patients, all relevant audio files are collected (in the case where a murmur is present, all audio files where a murmur can be heard are selected). The next step is the collate the audio files for the testing and training sets and prepare the training set in the audio preparation stage.

Due to the noisiness of the data, finding individual periods of audio using onset detection is relatively difficult. As seen in the flowchart Fig. A.2 this is still attempted with varying results, as will be explained in the evaluation section below.

The audio file to be separated into cardiac periods is gathered and the starting threshold is set (0.8) and the end threshold is also set (0.5). The minimum number of acceptable cardiac periods per audio sample to extract is set to 5, as during testing this seemed most reasonable (there is, of course, no upper limit). The audio then goes through a low-pass filter to attempt to isolate the heart's S1 and S2 sounds. The cut-off frequency on this filter is set to the maximum valid resting heart rate for the age of the patient. After re-scaling (-1 to 1) to normalise the audio, the model gathers the Mel Spectral

[1]This is the largest frequency able to be properly represented by the sampling rate due to sampling theory [11]

data to calculate the Spectral Flux[2] of the signal. This signal is highly noisy still, so a second low-pass filter is applied to this signal, and the outliers are filled with the previous values. After taking an absolute value and re-scaling between 0 to 1, the onsets are calculated using the starting threshold. If the time between subsequent onsets are valid cardiac period times for the age of the patient, this period is saved. After iterating through the whole audio file and finding which onsets are valid, the algorithm checks whether there are more periods than the minimum number required. If there are not enough periods, the algorithm lowers the threshold by 0.1 until it reaches the ending threshold or until the minimum required periods are attained. Finally, the audio is down-sampled to increase the speed of the Classification algorithm and to save storage space.

The training set is then made up of all of the periods and labelled accordingly with their diagnoses in the audio filenames. This is then where the model can use this database to classify and calculate the effectiveness using the previously split testing set.

B. *Classification*

The audio being incredibly noisy and hard to diagnose by professional doctors, preliminary metrics were considered before Dynamic Time (DTW). These included Root Mean Square Energy, Zero Crossing Rate, and Signal to Noise Ratio. However, when plotting these three metrics in any configuration against each other, the classification is clear to be highly inaccurate (see Fig. 4, 5, and 6) due to the high density and no apparent clustering or separation of the data. Hence, Dynamic Time Warping was the chosen metric for classification due to the possibility of some improvement to random guessing.

DTW, as defined by the following equation:

$$DTW_q(x, x') = \min_{\pi \in \mathcal{A}(x,x')} \langle A_\pi, D_q(x, x') \rangle^{\frac{1}{q}} \qquad (1)$$

can be described as the Euclidean distance with the use of time-domain elongation and compression. This is shown in Fig. 7, where the samples are compared in a non-linear way. The sample in either signal can be "held" (although not both at the same sample) to compare to the next sample of the signal in comparison. Iterating through all the combinations is an expensive operation, especially when then iterated through all the audio in the training set, for every test sample. DTW has a big-O of $O(N^2)$, hence the need to down-sample the audio for this much data[15]. Using DTW allows the model to mitigate

[2] The spectral change between frames[13]

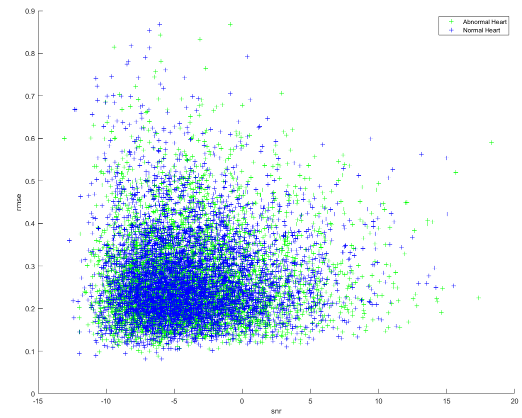

Fig. 4. Root Mean Square Energy against Signal to Noise Ratio for the dataset.

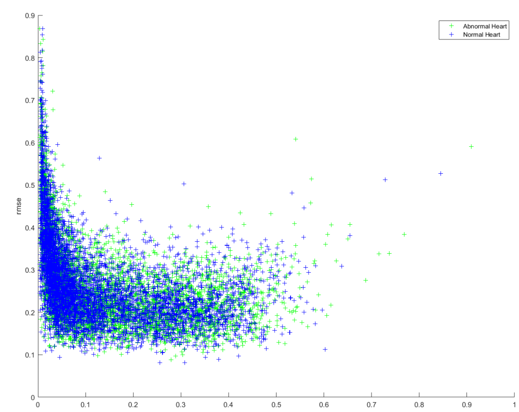

Fig. 5. Root Mean Square Energy against Zero Crossing Rate for the dataset.

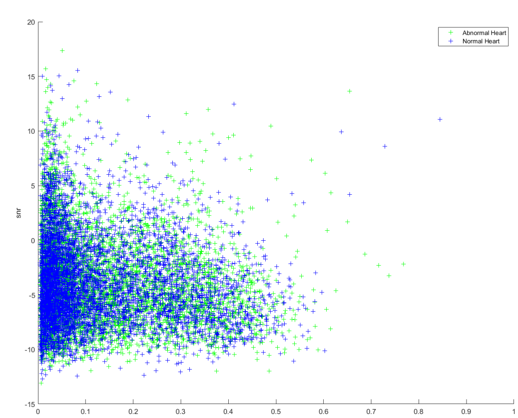

Fig. 6. Signal to Noise Ratio against Zero Crossing Rate for the dataset.

the heart rate period in comparison and subtle individual differences across patients when making a classification.

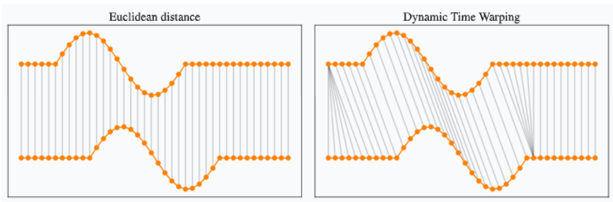

Fig. 7. Dynamic Time Warping Graphical Representation[15].

Classification by K-Nearest-Neighbours (KNN), seemed to best fit the way the data had been split into periods. KNN uses the closest K entries in the training set to the test sample, gathers these classification labels, and on a voting system predicts the class of the test sample. This can be seen in Fig. 8. The benefits of KNN rely on its ease of implementation, the lack of parameters to fine-tune and the adaptability of the model when new training data is added to the model. The disadvantages lie in the fact that this model does not scale well when adding extra dimensions to the input data, and the possibility of over-fitting to training data. The dimensionality problem does not affect this classifier, due to the model only using DTW as an input to the KNN algorithm, and this only has one metric[14].

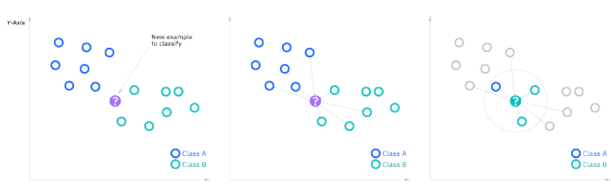

Fig. 8. K-Nearest-Neighbours Visualisation[14].

Without any clear indication of what K-Value this model would require to best perform, the model is trained with the same training data for four K-Values: 3,5,7,9. These values have to be odd so that there is a clear majority in the KNN voting.

Following the flowchart for testing the model (Fig. A.3), the test audio is separated into periods, and then checked if the number of acceptable periods is odd due to a need for a no-tie in the voting stage of the classifier. Once all the periods have been extracted from the patient's audio sample (as explained in more depth in the previous section), the DTW algorithm is used against all the entries of heart periods in the training set. For a given K-Value, the KNN algorithm finds the K minimum distances outputted from the DTW algorithm. For this cardiac period, a KNN classification is given by the voting of the nearest neighbour classes. Once all periods have been classified as either "Normal-Absent" or otherwise, there is a vote between all of the periods (hence the odd number) for the whole patient audio. After this is done for all K-Values and sexes (5 times to gain an average), the efficacy of the model can be examined.

III. Results

Added in the appendix, there are the full results shown in Fig. A.4, A.5, A.6, A.7, and A.8, but there is a summary of the overall performance.

After running the model 5 times for each of the 4 K-Values and the two sexes, the average accuracy, precision, recall and F1-Score of the model can be seen in Fig. 9. The separated averages for K-Values for each of the sexes can be seen in Fig. 10, and 11.

- The APRF scores for each sex/run/N

Average Test Results For Whole Model									
Accuracy	Precision	Recall	F-Score	Classification Confusion Matrix (%)			Per Period Confusion Matrix (%)		
50.25%	52.39%	30.00%	36.81%		Actual Pos	Actual Neg		Actual Pos	Actual Neg
				Predicted Pos	14.56%	14.51%	Predicted Pos	22.45%	21.93%
				Predicted Neg	35.24%	35.69%	Predicted Neg	27.79%	27.82%

Fig. 9. Average Results for all tests.

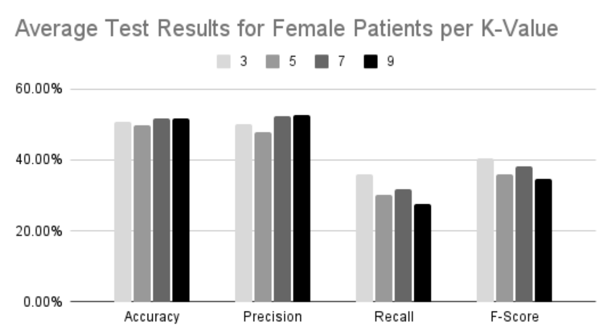

Fig. 10. Average Test Results for Female Patients per K-Value.

IV. Evaluation

A. Analysis

This model is not a good classifier, and it seems it would have a similar effectiveness to comparing: the Root Mean Square Energy, Zero Crossing Rate, and Signal to Noise Ratio. Delving into the metrics further, separating the sexes and K-Values (as seen in Fig. 10,

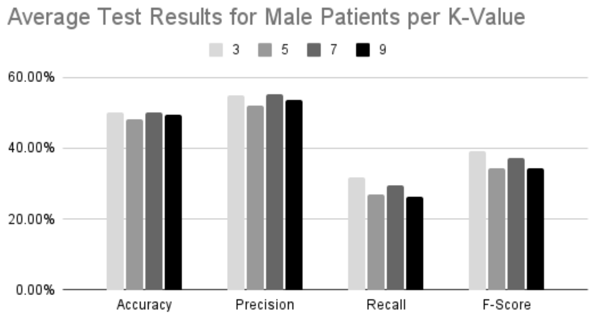

Fig. 11. Average Test Results for Male Patients per K-Value.

and 11), the K-value and different sex data makes a marginal change to the model.

To be used as a preliminary screening with the knowledge of its performance, using this current model may be unethical. Furthermore, the recall of the model is shown to be quite poor, and therefore not suitable for clinical contexts whatsoever, as it will misdiagnose those with potential underlying heart problems.

This shouldn't mean, however, that Dynamic Time Warping should not be used in the classification of heart murmurs, as there may be some improvements to be made with the cardiac period separation. As seen in Fig. 12, a lot of the periods extracted appear to be of single cardiac periods. Nonetheless, the cardiac sample in Fig. 13 extracted is two separate periods in sequence. Comparing this to a one-period audio file would have a great distance using DTW. Similarly, the noisiness of some of the extracted periods, such as Fig. 14, could be the source of completely incoherent results.

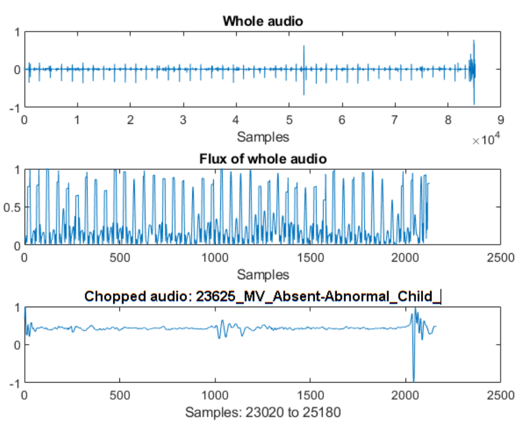

Fig. 12. Spectral flux, and extracted normal cardiac period from patient audio.

Some possible reasons for this could include, the cut-off frequency of the preliminary low-pass filter being too

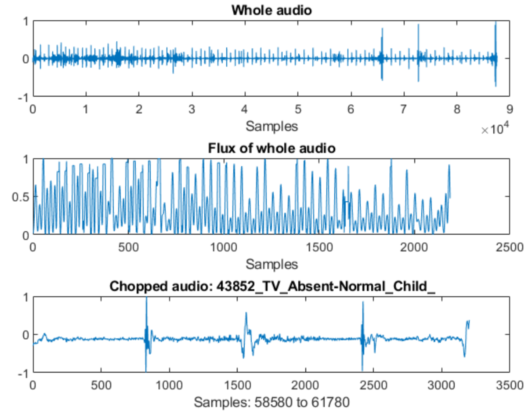

Fig. 13. Spectral flux, and extracted double-cardiac period from patient audio.

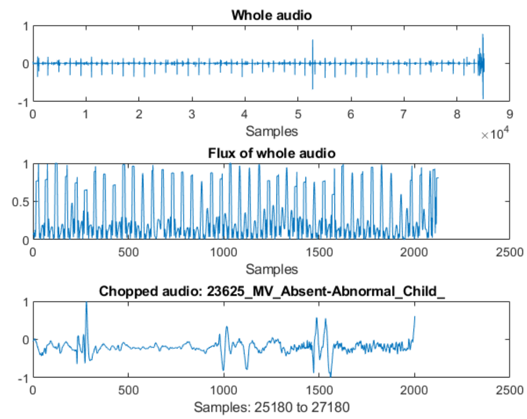

Fig. 14. Spectral flux, and extracted noisy segment from patient audio.

low, smoothing out the heart S1 and S2 sounds, making them indistinguishable from the noise of the signal. Although, this could introduce a trade-off between the strength of the S1 and S2 sounds, and the energy of the higher noise now being included in the signal. Another reason may be that not all patients are at resting heart rate during a clinical procedure, due to stress or anxiety. This could be mitigated, by finding the average heart rate for the period and segmenting the audio into cardiac periods related to the patient's heart rate in the recording, rather than the excepted resting heart rates depending on age. The down-sampling could also play a big role in identifying the murmurs, however, this was necessary for the speed of the model. A final reason could be that using dynamic time warping as the only KNN input variable is not the ideal way of categorising this overly noisy data. Using metrics of these cardiac periods could yield better results when used in conjunction with the DTW.

Looking at other methods in use for categorising heart murmurs, will be incredibly beneficial, and shed light on the potential shortcomings of this model in its current state.

B. Comparison to Other Methods

Sticking with the use of DTW and KNN, another method of classifying and segmenting the audio signal into cardiac periods could be done by using open-source algorithms[16]. There are four mentioned models for the segmentation of heart periods in this paper: Envelope (onset), Feature-Based, Machine Learning, and Hidden-Markov model. With the Hidden-Markov model being the state-of-the-art method.

In this paper, the model used to classify the abnormal from the normal heart recordings is based on feature extraction inputted into a logistic regression classifier. A benefit of using a logistic regression algorithm is the white-box nature, where the operator can extract the equation of how the model equates the results and make evaluations based on their predictions of which features would be the most beneficial[16].

Another method, described in [17], uses DEEP learning to classify the murmur. Although this has much better success rates of classification than the model proposed in this report and the previous logistic regression model, this type of model requires a large dataset and is also a black box, so the inner workings cannot be dissected with ease.

V. Conclusion

This report has given context to why heart murmur classification in children is important, not only for the health of the patients but for the time effectiveness of overly strained health services. Although this model has proved to be ineffective in its current state, the concept is not something to give up altogether, and in the future, the model will be expanded to increase its accuracy. Whether these improvements will lie in the audio preparation of the sample, the structure and feature selection for the classifier, or the classifying algorithm, they will only be implemented when more research and experimentation are conducted.

References

[1] B. Ford, S. Lara, and J. Park, “Heart murmurs in children: Evaluation and management,” American family physician, https://pubmed.ncbi.nlm.nih.gov/35289571/.

[2] “Signs of heart problems in children and teens,” Children’s Healthcare of Atlanta, https://www.choa.org/parent-resources/heart/signs-of-heart-problems-in-children-and-teens.

[3] “Heart murmurs - causes, symptoms treatment,” British Heart Foundation, https://www.bhf.org.uk/informationsupport/conditions/heart-murmurs.

[4] “Heart Murmur,” Penn Medicine, https://www.pennmedicine.org/for-patients-and-visitors/patient-information/conditions-treated-a-to-z/heart-murmur:∼:text=Definition,valves%20or%20near%20the%20heart.

[5] “Heart murmurs,” Johns Hopkins Medicine, https://www.hopkinsmedicine.org/health/conditions-and-diseases/heart-murmurs.

[6] P. Syamasundar Rao, “Diagnosis of cardiac murmurs in children,” Vessel Plus, https://www.oaepublish.com/articles/2574-1209.2021.105.

[7] L. M. Biga, S. Bronson, S. Dawson, and Et al, “19.3 cardiac cycle,” Oregon State University, https://open.oregonstate.education/aandp/chapter/19-3-cardiac-cycle/.

[8] A. N. Pelech, “THE CARDIAC MURMUR: When to Refer?,” Pediatric Clinics of North America, https://www.sciencedirect.com/science/article/pii/S003139550570585X.

[9] J. Oliveira, F. Renna, P. Costa, Et al, “The CIRCOR DigiScope Phonocardiogram Dataset,” The CirCor DigiScope Phonocardiogram Dataset v1.0.3, https://physionet.org/content/circor-heart-sound/1.0.3/training_data/#files-panel.

[10] H. Ren, H. Jin, C. Chen, and Et al, “A novel cardiac auscultation monitoring system based on wireless sensing for healthcare,” IEEE journal of translational engineering in health and medicine, https://pubmed.ncbi.nlm.nih.gov/30042903/.

[11] G. Wright, “What is the Nyquist theorem?,” WhatIs, https://www.techtarget.com/whatis/definition/Nyquist-Theorem:∼:text=The%20Nyquist%20theorem%20is%20also,and%20video%20to%20reduce%20aliasing.

[12] H. Laurila, “Manual Data Entry Errors,” Beamex Blog for Calibration Professionals, https://blog.beamex.com/manual-data-entry-errors:∼:text=This%20is%20despite%20the%20fact,data%20entry%20is%20about%201%20%25.

[13] “Spectral flux,” Spectral Flux - an overview — ScienceDirect Topics, https://www.sciencedirect.com/topics/engineering/spectral-flux.

[14] “What is the K-nearest neighbors algorithm?,” IBM, https://www.ibm.com/topics/knn:∼:text=The%20k%2Dnearest%20neighbors%20algorithm%2C%20also%20known%20as%20KNN%20or,of%20an%20individual%20data%20point.

[15] R. Tavenard, “An introduction to Dynamic Time Warping,” An introduction to dynamic time warping, https://rtavenar.github.io/blog/dtw.html.

[16] C. Liu, D. Springer, Q. Li, and Et Al, “An open access dataset for the evaluation of heart sound algorithms,” IOP Science, https://iopscience.iop.org/article/10.1088/0967-3334/37/12/2181.

[17] J. S. Chorba, A. M. Shapiro, L. Le, and Et al, “Deep learning algorithm for automated cardiac murmur detection via a Digital Sthethoscope Platform,” JAHA, https://www.ahajournals.org/doi/10.1161/JAHA.120.019905.

Appendix

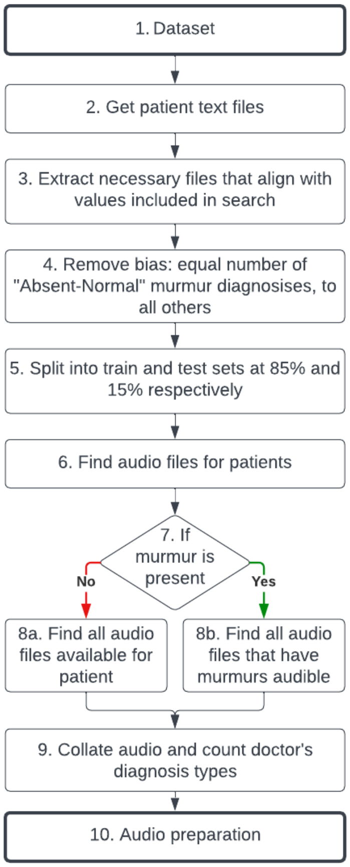

Fig. A.1. Dataset preparation flow chart.

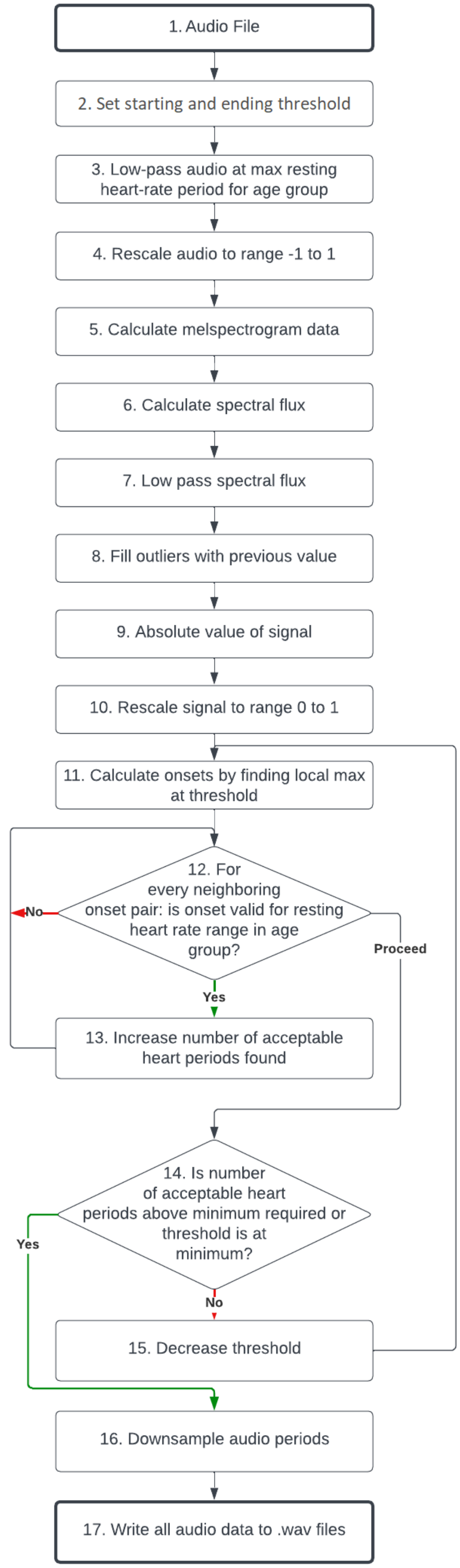

Fig. A.2. Audio preparation flow chart.

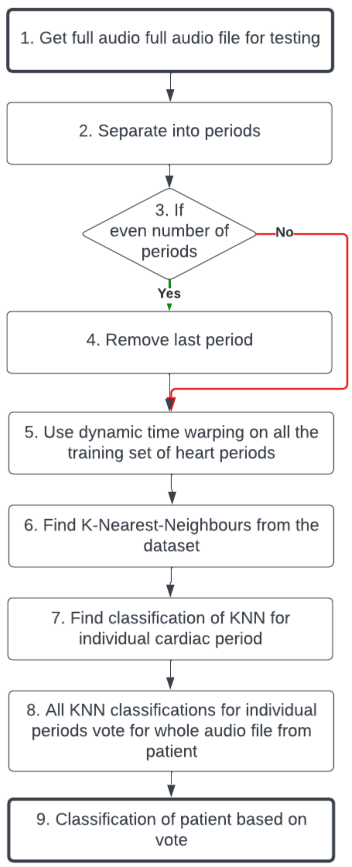

Fig. A.3. Testing model flow chart.

Sex	Average Test Results per Sex and K-Value										
	K-Value	Accuracy	Precision	Recall	F-Score	Classification Confusion Matrix (%)			Per Period Confusion Matrix (%)		
Female	3	50.85%	50.25%	35.88%	40.52%		Actual Pos	Actual Neg		Actual Pos	Actual Neg
						Predicted Pos	16.76%	17.90%	Predicted Pos	22.02%	23.28%
						Predicted Neg	31.25%	34.10%	Predicted Neg	26.01%	28.70%
	5	49.76%	48.02%	30.18%	35.86%		Actual Pos	Actual Neg		Actual Pos	Actual Neg
						Predicted Pos	14.07%	16.30%	Predicted Pos	21.78%	23.06%
						Predicted Neg	33.94%	35.69%	Predicted Neg	26.24%	28.91%
	7	51.80%	52.41%	31.74%	38.17%		Actual Pos	Actual Neg		Actual Pos	Actual Neg
						Predicted Pos	14.93%	15.13%	Predicted Pos	21.74%	22.98%
						Predicted Neg	33.07%	36.87%	Predicted Neg	26.29%	29.00%
	9	51.70%	52.68%	27.51%	34.69%		Actual Pos	Actual Neg		Actual Pos	Actual Neg
						Predicted Pos	12.86%	13.16%	Predicted Pos	21.61%	23.12%
						Predicted Neg	35.14%	38.84%	Predicted Neg	26.41%	28.86%
Male	3	50.05%	54.80%	31.83%	39.24%		Actual Pos	Actual Neg		Actual Pos	Actual Neg
						Predicted Pos	16.27%	14.63%	Predicted Pos	23.15%	20.57%
						Predicted Neg	35.32%	33.78%	Predicted Neg	29.31%	26.97%
	5	48.32%	51.95%	27.03%	34.52%		Actual Pos	Actual Neg		Actual Pos	Actual Neg
						Predicted Pos	13.63%	13.73%	Predicted Pos	22.92%	20.73%
						Predicted Neg	37.95%	34.68%	Predicted Neg	29.54%	26.81%
	7	50.21%	55.34%	29.43%	37.25%		Actual Pos	Actual Neg		Actual Pos	Actual Neg
						Predicted Pos	14.78%	12.99%	Predicted Pos	23.18%	20.87%
						Predicted Neg	36.81%	35.43%	Predicted Neg	29.28%	26.67%
	9	49.34%	53.65%	26.37%	34.20%		Actual Pos	Actual Neg		Actual Pos	Actual Neg
						Predicted Pos	13.18%	12.25%	Predicted Pos	23.22%	20.87%
						Predicted Neg	38.41%	36.16%	Predicted Neg	29.25%	26.66%

Fig. A.4. Average results for of datasets for each sex and K-Value.

Sex	Dataset Cleaning											
	Number	Valid Patient Files	Patient Files after Bias	Test Patient Number	Training Patient Number	Training Audio Number	Quantities For Full Audio			Quantities For Period Audio		
Female	1	304	279	41	238	823		Present	Absent		Present	Absent
							Abnormal	113	240	Abnormal	1706	3422
							Normal	16	453	Normal	203	6036
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								369	453		5331	6036
	2	304	289	43	246	857		Present	Absent		Present	Absent
							Abnormal	123	262	Abnormal	1729	3686
							Normal	19	452	Normal	257	6094
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								404	452		5672	6094
	3	304	281	53	228	793		Present	Absent		Present	Absent
							Abnormal	109	219	Abnormal	1530	3060
							Normal	14	450	Normal	186	6059
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								342	450		4776	6059
	4	304	288	47	241	833		Present	Absent		Present	Absent
							Abnormal	140	251	Abnormal	1929	3464
							Normal	14	427	Normal	204	5756
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								405	427		5597	5756
	5	304	292	47	245	855		Present	Absent		Present	Absent
							Abnormal	150	270	Abnormal	2116	3712
							Normal	17	417	Normal	219	5638
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								437	417		6047	5638

Fig. A.5. Dataset information for Female Runs of Model.

	Dataset Cleaning											
Sex	Number	Valid Patient Files	Patient Files after Bias	Test Patient Number	Training Patient Number	Training Audio Number	Quantities For Full Audio			Quantities For Period Audio		
Male	1	322	322	46	276	964		Present	Absent		Present	Absent
							Abnormal	128	316	Abnormal	1824	4560
							Normal	35	484	Normal	591	6857
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								479	484		6975	6857
	2	322	322	60	262	898		Present	Absent		Present	Absent
							Abnormal	110	285	Abnormal	1634	4115
							Normal	26	476	Normal	456	6736
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								421	476		6205	6736
	3	322	322	46	276	960		Present	Absent		Present	Absent
							Abnormal	134	315	Abnormal	1928	4565
							Normal	36	474	Normal	591	6776
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								485	474		7084	6776
	4	322	322	56	266	926		Present	Absent		Present	Absent
							Abnormal	109	293	Abnormal	1665	4275
							Normal	26	497	Normal	443	7030
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								428	497		6383	7030
	5	322	322	60	262	898		Present	Absent		Present	Absent
							Abnormal	110	285	Abnormal	1634	4115
							Normal	26	476	Normal	456	6736
							Categorised for Classifier			Categorised for Classifier		
								Present	Absent		Present	Absent
								421	476		6205	6736

Fig. A.6. Dataset information for Male Runs of Model.

Testing																
K-Value	Accuracy	Precision	Recall	F-Score	Classification Confusion Matrix (Num)			Classification Confusion Matrix (%)			Per Period Confusion Matrix (Num)			Per Period Confusion Matrix (%)		
3	45.95%	49.06%	32.91%	39.39%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	26	27	Predicted Pos	17.57%	18.24%	Predicted Pos	1217	1027	Predicted Pos	24.83%	20.95%
					Predicted Neg	53	42	Predicted Neg	35.81%	28.38%	Predicted Neg	1438	1220	Predicted Neg	29.33%	24.89%
5	45.95%	48.65%	22.78%	31.03%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	18	19	Predicted Pos	12.16%	12.84%	Predicted Pos	2001	1661	Predicted Pos	24.49%	20.33%
					Predicted Neg	61	50	Predicted Neg	41.22%	33.78%	Predicted Neg	2424	2084	Predicted Neg	29.67%	25.51%
7	52.70%	62.16%	29.11%	39.66%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	23	14	Predicted Pos	15.54%	9.46%	Predicted Pos	2814	2291	Predicted Pos	24.60%	20.03%
					Predicted Neg	56	55	Predicted Neg	37.84%	37.16%	Predicted Neg	3381	2952	Predicted Neg	29.56%	25.81%
9	50.68%	58.33%	26.58%	36.52%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	21	15	Predicted Pos	14.19%	10.14%	Predicted Pos	3613	2961	Predicted Pos	24.57%	20.13%
					Predicted Neg	58	54	Predicted Neg	39.19%	36.49%	Predicted Neg	4352	3780	Predicted Neg	29.59%	25.70%
3	49.65%	50.00%	30.99%	38.26%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	22	22	Predicted Pos	15.60%	15.60%	Predicted Pos	1025	943	Predicted Pos	23.64%	21.75%
					Predicted Neg	49	48	Predicted Neg	34.75%	34.04%	Predicted Neg	1156	1211	Predicted Neg	26.67%	27.94%
5	47.52%	46.34%	26.76%	33.93%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	19	22	Predicted Pos	13.48%	15.60%	Predicted Pos	1665	1579	Predicted Pos	23.04%	21.85%
					Predicted Neg	52	48	Predicted Neg	36.88%	34.04%	Predicted Neg	1970	2011	Predicted Neg	27.27%	27.83%
7	53.90%	56.52%	36.62%	44.44%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	26	20	Predicted Pos	18.44%	14.18%	Predicted Pos	2307	2212	Predicted Pos	22.81%	21.87%
					Predicted Neg	45	50	Predicted Neg	31.91%	35.46%	Predicted Neg	2782	2814	Predicted Neg	27.50%	27.82%
9	51.06%	52.94%	25.35%	34.29%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	18	16	Predicted Pos	12.77%	11.35%	Predicted Pos	2969	2838	Predicted Pos	22.83%	21.82%
					Predicted Neg	53	54	Predicted Neg	37.59%	38.30%	Predicted Neg	3574	3624	Predicted Neg	27.48%	27.87%
3	50.54%	72.00%	31.58%	43.90%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	36	14	Predicted Pos	19.35%	7.53%	Predicted Pos	1558	944	Predicted Pos	26.18%	15.86%
					Predicted Neg	78	58	Predicted Neg	41.94%	31.18%	Predicted Neg	2084	1366	Predicted Neg	35.01%	22.95%
5	47.85%	68.09%	28.07%	39.75%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	32	15	Predicted Pos	17.20%	8.06%	Predicted Pos	2609	1546	Predicted Pos	26.30%	15.58%
					Predicted Neg	82	57	Predicted Neg	44.09%	30.65%	Predicted Neg	3461	2304	Predicted Neg	34.89%	23.23%
7	46.77%	68.29%	24.56%	36.13%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	28	13	Predicted Pos	15.05%	6.99%	Predicted Pos	3607	2142	Predicted Pos	25.97%	15.42%
					Predicted Neg	86	59	Predicted Neg	46.24%	31.72%	Predicted Neg	4891	3248	Predicted Neg	35.22%	23.39%
9	46.77%	74.19%	20.18%	31.72%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	23	8	Predicted Pos	12.37%	4.30%	Predicted Pos	4605	2766	Predicted Pos	25.79%	15.49%
					Predicted Neg	91	64	Predicted Neg	48.92%	34.41%	Predicted Neg	6321	4164	Predicted Neg	35.40%	23.32%
3	56.25%	45.45%	38.46%	41.67%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	25	30	Predicted Pos	15.63%	18.75%	Predicted Pos	907	1343	Predicted Pos	18.19%	26.94%
					Predicted Neg	40	65	Predicted Neg	25.00%	40.63%	Predicted Neg	1022	1714	Predicted Neg	20.50%	34.38%
5	56.25%	45.28%	36.92%	40.68%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	24	29	Predicted Pos	15.00%	18.13%	Predicted Pos	1505	2250	Predicted Pos	18.11%	27.08%
					Predicted Neg	41	66	Predicted Neg	25.63%	41.25%	Predicted Neg	1710	2845	Predicted Neg	20.58%	34.24%
7	56.88%	45.83%	33.85%	38.94%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	22	26	Predicted Pos	13.75%	16.25%	Predicted Pos	2114	3150	Predicted Pos	18.17%	27.08%
					Predicted Neg	43	69	Predicted Neg	26.88%	43.13%	Predicted Neg	2387	3983	Predicted Neg	20.52%	34.24%
9	60.63%	52.00%	40.00%	45.22%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	26	24	Predicted Pos	16.25%	15.00%	Predicted Pos	2714	4112	Predicted Pos	18.14%	27.49%
					Predicted Neg	39	71	Predicted Neg	24.38%	44.38%	Predicted Neg	3073	5059	Predicted Neg	20.54%	33.82%
3	51.88%	34.72%	45.45%	39.37%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	25	47	Predicted Pos	15.63%	29.38%	Predicted Pos	875	1568	Predicted Pos	17.24%	30.89%
					Predicted Neg	30	58	Predicted Neg	18.75%	36.25%	Predicted Neg	940	1693	Predicted Neg	18.52%	33.35%
5	51.25%	31.75%	36.36%	33.90%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	20	43	Predicted Pos	12.50%	26.88%	Predicted Pos	1433	2578	Predicted Pos	16.94%	30.47%
					Predicted Neg	35	62	Predicted Neg	21.88%	38.75%	Predicted Neg	1592	2857	Predicted Neg	18.82%	33.77%
7	48.75%	29.23%	34.55%	31.67%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	19	46	Predicted Pos	11.88%	28.75%	Predicted Pos	2028	3613	Predicted Pos	17.12%	30.50%
					Predicted Neg	36	59	Predicted Neg	22.50%	36.88%	Predicted Neg	2207	3996	Predicted Neg	18.63%	33.74%
9	49.38%	25.93%	25.45%	25.69%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	14	40	Predicted Pos	8.75%	25.00%	Predicted Pos	2549	4668	Predicted Pos	16.74%	30.65%
					Predicted Neg	41	65	Predicted Neg	25.63%	40.63%	Predicted Neg	2896	5115	Predicted Neg	19.02%	33.59%

Fig. A.7. Testing information for Female Runs of Model.

Testing																
K-Value	Accuracy	Precision	Recall	F-Score	Classification Confusion Matrix (Num)			Classification Confusion Matrix (%)			Per Period Confusion Matrix (Num)			Per Period Confusion Matrix (%)		
3	50.00%	42.19%	39.13%	40.60%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	27	37	Predicted Pos	17.09%	23.42%	Predicted Pos	1112	1304	Predicted Pos	21.88%	25.66%
					Predicted Neg	42	52	Predicted Neg	26.58%	32.91%	Predicted Neg	1141	1525	Predicted Neg	22.45%	30.01%
5	48.73%	39.29%	31.88%	35.20%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	22	34	Predicted Pos	13.92%	21.52%	Predicted Pos	1784	2151	Predicted Pos	21.06%	25.40%
					Predicted Neg	47	55	Predicted Neg	29.75%	34.81%	Predicted Neg	1971	2564	Predicted Neg	23.27%	30.27%
7	54.43%	47.27%	37.68%	41.94%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	26	29	Predicted Pos	16.46%	18.35%	Predicted Pos	2531	3007	Predicted Pos	21.34%	25.36%
					Predicted Neg	43	60	Predicted Neg	27.22%	37.97%	Predicted Neg	2726	3594	Predicted Neg	22.99%	30.31%
9	59.49%	54.90%	40.58%	46.67%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	28	23	Predicted Pos	17.72%	14.56%	Predicted Pos	3255	3868	Predicted Pos	21.35%	25.37%
					Predicted Neg	41	66	Predicted Neg	25.95%	41.77%	Predicted Neg	3504	4619	Predicted Neg	22.98%	30.30%
3	50.22%	63.33%	29.92%	40.64%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	38	22	Predicted Pos	17.04%	9.87%	Predicted Pos	1761	1211	Predicted Pos	24.80%	17.05%
					Predicted Neg	89	74	Predicted Neg	39.91%	33.18%	Predicted Neg	2364	1765	Predicted Neg	33.29%	24.86%
5	45.74%	55.77%	22.83%	32.40%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	29	23	Predicted Pos	13.00%	10.31%	Predicted Pos	2930	2058	Predicted Pos	24.76%	17.39%
					Predicted Neg	98	73	Predicted Neg	43.95%	32.74%	Predicted Neg	3945	2902	Predicted Neg	33.33%	24.52%
7	48.43%	61.11%	25.98%	36.46%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	33	21	Predicted Pos	14.80%	9.42%	Predicted Pos	4173	2924	Predicted Pos	25.19%	17.65%
					Predicted Neg	94	75	Predicted Neg	42.15%	33.63%	Predicted Neg	5452	4020	Predicted Neg	32.90%	24.26%
9	45.29%	55.32%	20.47%	29.89%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	26	21	Predicted Pos	11.66%	9.42%	Predicted Pos	5356	3751	Predicted Pos	25.14%	17.61%
					Predicted Neg	101	75	Predicted Neg	45.29%	33.63%	Predicted Neg	7019	5177	Predicted Neg	32.95%	24.30%
3	50.31%	34.55%	30.16%	32.20%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	19	36	Predicted Pos	11.80%	22.36%	Predicted Pos	935	1431	Predicted Pos	18.30%	28.01%
					Predicted Neg	44	62	Predicted Neg	27.33%	38.51%	Predicted Neg	1078	1665	Predicted Neg	21.10%	32.59%
5	53.42%	38.46%	31.75%	34.78%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	20	32	Predicted Pos	12.42%	19.88%	Predicted Pos	1520	2413	Predicted Pos	17.85%	28.34%
					Predicted Neg	43	66	Predicted Neg	26.71%	40.99%	Predicted Neg	1835	2747	Predicted Neg	21.55%	32.26%
7	52.80%	38.18%	33.33%	35.59%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	21	34	Predicted Pos	13.04%	21.12%	Predicted Pos	2155	3382	Predicted Pos	18.08%	28.37%
					Predicted Neg	42	64	Predicted Neg	26.09%	39.75%	Predicted Neg	2542	3842	Predicted Neg	21.32%	32.23%
9	49.69%	32.69%	26.98%	29.57%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	17	35	Predicted Pos	10.56%	21.74%	Predicted Pos	2765	4369	Predicted Pos	18.04%	28.51%
					Predicted Neg	46	63	Predicted Neg	28.57%	39.13%	Predicted Neg	3274	4919	Predicted Neg	21.36%	32.09%
3	49.49%	70.59%	30.00%	42.11%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	36	15	Predicted Pos	18.37%	7.65%	Predicted Pos	1563	906	Predicted Pos	25.97%	15.05%
					Predicted Neg	84	61	Predicted Neg	42.86%	31.12%	Predicted Neg	2193	1356	Predicted Neg	36.44%	22.53%
5	47.96%	70.45%	25.83%	37.80%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	31	13	Predicted Pos	15.82%	6.63%	Predicted Pos	2626	1517	Predicted Pos	26.18%	15.12%
					Predicted Neg	89	63	Predicted Neg	45.41%	32.14%	Predicted Neg	3634	2253	Predicted Neg	36.23%	22.46%
7	46.94%	69.05%	24.17%	35.80%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	29	13	Predicted Pos	14.80%	6.63%	Predicted Pos	3668	2151	Predicted Pos	26.12%	15.32%
					Predicted Neg	91	63	Predicted Neg	46.43%	32.14%	Predicted Neg	5096	3127	Predicted Neg	36.29%	22.27%
9	46.94%	70.00%	23.33%	35.00%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	28	12	Predicted Pos	14.29%	6.12%	Predicted Pos	4768	2759	Predicted Pos	26.41%	15.28%
					Predicted Neg	92	64	Predicted Neg	46.94%	32.65%	Predicted Neg	6500	4027	Predicted Neg	36.00%	22.31%
3	50.22%	63.33%	29.92%	40.64%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	38	22	Predicted Pos	17.04%	9.87%	Predicted Pos	1761	1211	Predicted Pos	24.80%	17.05%
					Predicted Neg	89	74	Predicted Neg	39.91%	33.18%	Predicted Neg	2364	1765	Predicted Neg	33.29%	24.86%
5	45.74%	55.77%	22.83%	32.40%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	29	23	Predicted Pos	13.00%	10.31%	Predicted Pos	2930	2058	Predicted Pos	24.76%	17.39%
					Predicted Neg	98	73	Predicted Neg	43.95%	32.74%	Predicted Neg	3945	2902	Predicted Neg	33.33%	24.52%
7	48.43%	61.11%	25.98%	36.46%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	33	21	Predicted Pos	14.80%	9.42%	Predicted Pos	4173	2924	Predicted Pos	25.19%	17.65%
					Predicted Neg	94	75	Predicted Neg	42.15%	33.63%	Predicted Neg	5452	4020	Predicted Neg	32.90%	24.26%
9	45.29%	55.32%	20.47%	29.89%		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg		Actual Pos	Actual Neg
					Predicted Pos	26	21	Predicted Pos	11.66%	9.42%	Predicted Pos	5356	3751	Predicted Pos	25.14%	17.61%
					Predicted Neg	101	75	Predicted Neg	45.29%	33.63%	Predicted Neg	7019	5177	Predicted Neg	32.95%	24.30%

Fig. A.8. Testing information for Male Runs of Model.