



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

## Automatic Heart Diseases Detection Techniques using Musical Approaches

Ehsan Safar Khorasani, Shyamala Doraisamy and Azreen Azman  
Faculty of Computer Science and Information Technology, University Putra Malaysia,  
Selangor, Malaysia

---

**Abstract:** In this study, a musical approach to provide an automatic heart disease detection system is proposed. Heart sounds are recorded with audio format. Audio files are converted to semi-structured music files that can be represented textually. Samples were captured from different heart diseases and were stored in a database. Two different approaches which are information retrieval based on n-gram and longest common subsequence are used to retrieve the similarity of a given sample with existing heart diseases in the database. Since the frequency of heart sound is relative to age and physical characteristics of a patient, an important feature of using n-gram in this study is to retrieve diseases without respect to the different heart sounds frequencies. The effects of window sizes for n-gram approach on the accuracy of the information retrieval were tested and a proper window size was extracted. The results of the performed experiments showed that window size of 5 notes revealed a high performance in comparison with other window sizes. Hence, the proposed technique can detect and recognize a heart disease with a reliable accuracy. Average of precision values for around 85% in information retrieval and 55% in longest common subsequence technique were obtained for the retrieval of heart sound categories. Moreover, the results of string matching technique demonstrated that threshold level of 65% could appropriately detect heart disease.

**Key words:** Heart disease, signal preprocessing, heart sound, multimedia, music information retrieval

---

### INTRODUCTION

PCG (phonocardiogram) is the sonic vibration of heart beat. Commonly, PCG are utilized as an audio input of the system for studying different techniques of digital signal processing (Babaei and Geranmayeh, 2009). Over the past years, several approaches for automated and semi-automated detection and classification of various heart diseases have been proposed. These approaches mainly utilize the heart sound signals as basis in their underlying process for such classifications and detections. These levels can be ranged from processing/analyzing of heart sounds signals up to final classification of these signals. Numerous studies have addressed analyzing heart sound signals in order to generate the optimum representations of them (Arvin and Doraisamy, 2010a). Moreover, ECG (electrocardiogram) was previously studied in biomedical engineering and various researched on heart (Selva Vijila *et al.*, 2006; Amien and Lin, 2007; Swarnalatha and Prasad, 2010).

Various approaches have been employed by different scholars aiming at the heart sound signal segmentation.

Four heart sound components are: S1, systole phase, S2 and diastole phase which researches were commonly focused on: Power spectral density calculation based on autoregressive models, normalized average Shannon energy calculation, wavelet transform, homomorphic filtering and K-means clustering (Wang *et al.*, 2005; Gupta *et al.*, 2007; Arvin and Doraismay, 2010b). Hebden and Torry (1997) also proposed a method based on back-propagation neural network for recognizing S1 peaks from S2's in both normal and abnormal heart sound signals (Hendel *et al.*, 2010). Wavelet transform (known as wavelet analysis) has shown to have much interest among the researchers compared to other techniques (El-Asir and Mayyas, 2004; El-Ramsisi and Khalil, 2007; Dokur and Olmez, 2008; Babaei and Geranmayeh, 2009). Syeda-Mahmood and Wang (2008) have proposed a shape-based disease detection technique. Based on the proposed technique, similar heart sounds were retrieved from database using visualized heart sounds.

N-grams have been extensively used in text retrieval, where the strings of symbols are divided into overlapping constant-length subsequences. "Unigram" is an n-gram

of size 1 while bigram and trigram are referred to as size 2 and size 3, respectively. Finally size 4 and above which is merely called an n-gram. For instance, the word heart includes the bi-grams he, ea, ar and rt. According to (Roark *et al.*, 2007), a sequence of characters created from n adjoining characters within a text is called an n-gram. In addition, Longest Common Subsequence (LCS) is another technique which is extensively used in several research areas including DNA similarity measurement, video sequence measurement and web browsing pattern recognition (Bergroth *et al.*, 2000).

In this study, heart sounds are captured by electronic stethoscope and recorded in Pulse-code Modulation (PCM) audio file formats from different heart diseases. The recorded heart sounds are converted to semi-structured MIDI (Musical Instrument Digital Interface) format for applying future processes. After extraction of the onset time of notes, n-gram approach is employed to prepare text with intervals and pitch ratio of the selected sequences. For testing information retrieval, two different approaches are utilized and efficiency of both is investigated. Information retrieval is performed using n-gram and Longest Common Subsequence (LCS) approaches. Hence, the accuracy of approaches is tested for different heart diseases.

### HEART SOUND AND DISEASES

There are two major sounds in healthy adults that create normal heart sound like “lub dub”. The first one “lub” named first heart sound (S1) and another one is “dub” called second heart sound (S2) (Andreoli *et al.*, 2007; Crawford, 2009). In addition, there is a third heart sound (S3) and a fourth heart sound (S4). They can be related to pathological processes or happen in normal persons. Gallops are called based on of the rate or rhythmic timing of S3 and S4. First heart sound, S1, is generated by the closure of Tricuspid and Mitral valves at the start of systole. As shown Fig. 1, Systole is the time between S1 and S2 and Diastole is the time between S2 and S1. S1 occurs at the onset of ventricular systole and corresponds to closure of the atrioventricular valves. Although Mitral and Tricuspid are occasionally two components, it is normally heard as a single sound. The second heart sound is produced by the closure of Aortic valves and Pulmonic valves. Second heart sound ultimate of the aortic and Pulmonic valves generates the second heart sound (S2). The Pulmonic valve typically lags slightly behind while the two valves close concurrently. A third heart sound is low-pitched sound occurs throughout early quick filling of the left ventricle. Fourth heart sound is a low-pitched, presystolic sound due to a

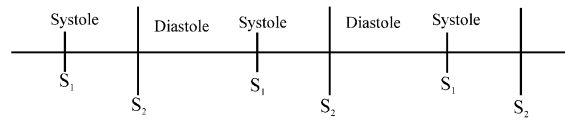


Fig. 1: Two cycles of normal heart sound (S1 and S2)

Table 1: Category of samples that are used in this study

Heart diseases	No. of samples
Normal Sound (NS)	15
Split Second Sound (SS)	12
Ejection Sound (ES)	10
Clicks (CL)	12
Gallop Rhythm (GR)	12
Systolic Murmurs (SM)	13
Diastolic Murmurs (DM)	14
Continues Murmurs (CM)	12

strong atrial reduction into a strengthened left ventricle that is generated in the ventricle during ventricular filling. Table 1 shows the samples and categories that are used in this study.

**Gallop sounds:** Gallops are low-frequency heart sounds which come in triplets and are similar to the sound of a galloping horse. The third heart sound (S3) is a gallop sound which occurs during the rapid blood filling in ventricle. S3 comes immediately after S2. Such a sound is associated with myocardial disease and also Heart Failure (HF) in which ventricular expel of the blood is failed during systole. The patient lying on the left side happens to be the best position for S3 to be heard. The fourth heart sounds (S4) are Gallops heard during atrial contraction and is heard immediately prior to the S1. An S4 often occurs when the ventricle enlarged or hypertrophied. S4 is associated with Coronary Artery Disease (CAD), hypertension and stenosis of the aortic valve. It is rare that all four sounds of the heart occur within one cardiac cycle. However, if happens, it results in a quadruple rhythm. Figure 2 shows S3 and S4 gallops.

**Snaps and clicks:** An opening snap is an unusual high-pitched sound occurring very early in diastole. This sound is associated with stenosis of the mitral valve which is a result of rheumatic heart disease. More specifically, the sudden displacement of a calcified mitral valve increases the pressure within the left atrium and leads to an opening snap sound. Along the left sternal border is where this sound is best heard. Opening snap is not mistaken for a part of S2 sound since it happens too long later than S2. It also occurs so early in diastole to be mistaken for a gallop. Similarly, an ejection click is a sharp, high-pitched and short sound, which is caused by stenosis of the aortic valve. It is happen in early systole

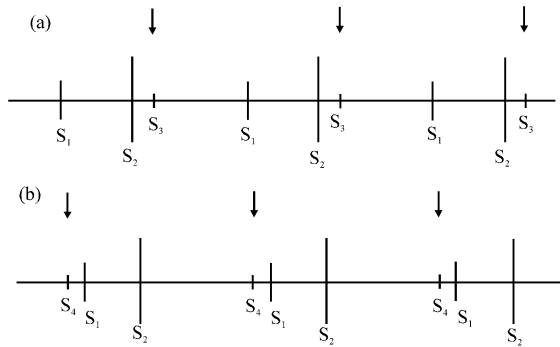


Fig. 2 (a-b): S3 Gallop and S4 gallop

and can be heard closely after S1 sound. The displacement of a rigid and inflexible aortic valve increases the pressure within the ventricle and leads to an ejection click. The left ventricular apex and also the second right intercostals space are where ejection click sound is best heard. Moreover, there is also non-ejection or mid-systolic clicks which denote the prolapsed of one or both leaflets of the mitral valve. They may occur with or without a late systolic murmur.

**Heart murmurs:** A murmur is a series of vibrations containing varied durations which comes from the heart or great vessels and can be heard through a stethoscope at the chest wall. In fact, murmurs are considered as additional heart sounds that are produced by a turbulent blood flow which is adequate to produce audible noise. Listening to the large number of murmurs can only be achieved through a stethoscope and auscultation. Murmurs are categorized in terms of their timing during a cardiac cycle. Systolic murmur is happening in systole time. Diastolic murmurs occur between the first and the second heart sounds and subside in mid or late diastole.

**TRANSCRIPTION OF HEART SOUNDS SIGNALS**

**Text-based transcription:** Transcription of the heart sound signals is another level that takes place after the preprocessing and prior to the classification and detection level. In this regard, as far as we are aware, there is one study (Modegi, 2001) which has applied the MIDI technology for encoding medical signals such as heart sounds and lung sounds for creating medical audio databases. This transcription method for heart sound is based on XML. In other words, this work automatically converts the heart sound data represented by MIDI into XML. It is worth emphasizing that in our proposed technique we have also used the MIDI format, as similar to the above presented work. However, in our work we transform the MIDI file into text-string and musical words

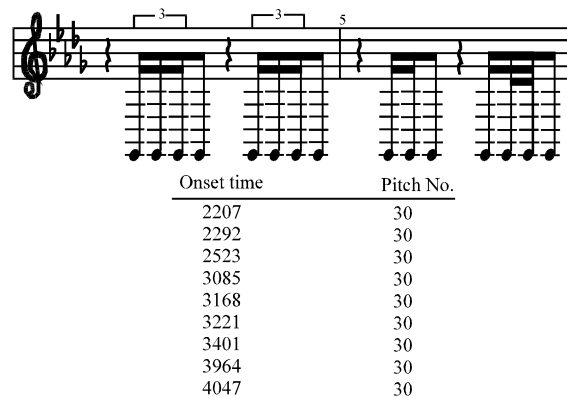


Fig. 3: Transcription of heart sound including murmur problem

instead of XML which is fully structured. This can have several possible advantages over the XML-based as it provides more flexibility as well as higher accuracy.

**Midi encoding:** MIDI is a popular format in computer-based music production such as composition, transcription, etc. Size of MIDI files is generally smaller than other music formats. With wave to MIDI converters, we will be able to generate MIDI data of PCM-based sounds rather easily. It extracts acoustical characteristics such as pitch, volume and duration for converting them into a sequence of notes for producing music scores. Thus, melodies will be translated into chromatic pitches without human intervention (Itou and Nishimoto, 2007). The quality of MIDI transcribed would also depend on the algorithm capabilities which are utilized. Since the heart sound frequency is between 25-50 Hz (Phua *et al.*, 2008) therefore, the encoding component just filtered heart sound with this frequency and ignore another sounds. Since the heart sound has low frequency and small duration hence the tools should be able of accurate transcription of these parameters. To achieve this objective we played 30 different pitches in C<sub>1</sub> channel and select Drums Program of YAMAHA Electronic synthesizer. The pitches were captured in Wave format using microphone and then encoded them to MIDI files by different tools. Figure 3 divides the sequence of events into overlapping subsequences of different adjacent events. All possible monophonic pitch sequences are extracted for each window that constructs the corresponding musical words.

**AUTOMATIC HEART DISEASES DETECTION**

Information Retrieval (IR) involves searching unstructured data available in a repository with the

purpose of reaching the desired information (Manning *et al.*, 2008). A relational database is an example of unstructured data. A query is a set of words and phrases representing the information a user wants to know. The field of IR is not only about retrieval of the information but also includes browsing and filtering the retrieved results of an IR system.

**Information retrieval process:** A general classic IR model is shown in Fig. 4 which briefly describes the IR process. Information items in a collection are preprocessed and indexed. Submitted queries are processed and then formulated to be in similar format of indexed items within the collection. The similarity of the formulated query and the indexed document collection is measured and items with the highest similarity are retrieved and presented to the user (Doraisamy and Ruger, 2003).

**Ranking:** An IR system must determine the similarity between a user query and a document. Via a closeness calculation mechanism, IR system decides to whether retrieve the document and specifies the ranking of the retrieved text. In order to determine the rank of a matched document, system makes an estimation of the relevance of the document. This estimate must be highly correlated to the user's perception of the relevance of the document (Meadow *et al.*, 1999).

**Music IR with n-gram:** Since there are high probabilities of errors in music queries, n-grams has been used in musical sequences towards addressing fault tolerance in music retrieval (Doraisamy and Ruger, 2003). The MIDI-to-text string is constructed from the interval sequences of heart sound pitches. For a sequence of n pitches, an interval sequence is derived with n-1:

$$I_i = p_{i+1} - p_i \tag{1}$$

where, I is interval among two pitches, p. The onset values are shown in Fig. 3. For a sequence of n onset times (t), a rhythmic ratio, R, sequence is derived with n-2 ratios obtained by the following equation:

$$R_i = \frac{t_{i+2} - t_{i+1}}{t_{i+2} - t_i} \tag{2}$$

Therefore, MIDI-to-text string that incorporates interval and rhythmic ratio sequences using n onset times and pitches, would be constructed in the following pattern:

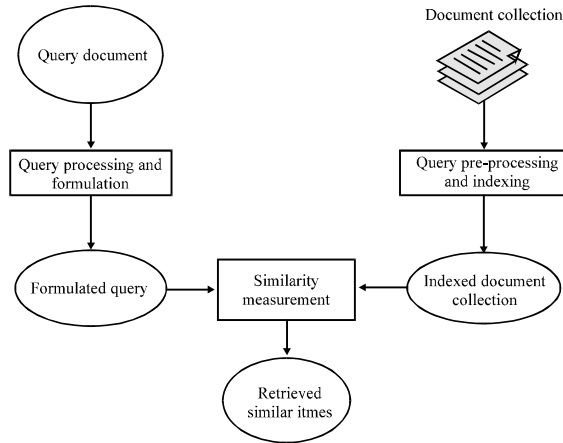


Fig. 4: A general classic IR models

$$[I_1R_1]...[I_{n-2}R_{n-2}] \tag{3}$$

**Music IR with LCS:** A key class of string algorithms is String matching algorithms. These algorithms try to locate one or several strings within a larger string or text. Uncovering the maximum number of matching symbols in two strings is a clear evaluation of their closeness while maintaining the symbol order. Longest Common Subsequence (LCS) is defined as the longest common subsequence between two sequences.

## RESULTS AND DISCUSSION

**Experimental metric:** Recall and precision are two basic measures of IR effectiveness. Recall is the total number of relevant documents retrieved in response to a search query, divided by the total number of relevant documents in the entire index (Syeda-Mahmood and Wang, 2008):

$$\eta = \frac{N_r}{N_e} \tag{4}$$

where,  $\eta$  is the recall value,  $N_r$  is the number of correct retrieved heart diseases and  $N_e$  is the number of correct records that exist in database. Precision (P) is the fraction of retrieved documents that are relevant. If the IR system returns all the relevant texts in its index, calculated recall equals to 100%. Nevertheless, recall by itself is not a sufficient measurement because return of the whole collection of documents simply results in a recall of 100%. The percentage of precision balances the recall. Precision is the number of relevant heart sounds retrieved divided by the total number of heart sounds retrieved.

Table 2: The measured recall and precision for different window sizes.  $\zeta$  is the recall and P is the precision

Heart sample	n = 3		n = 4		n = 5		n = 6	
	$\eta$	P	$\eta$	P	$\eta$	P	$\eta$	P
NS	0.63	0.10	0.63	0.63	0.50	1.00	0.50	1.00
SS	1.00	0.10	0.80	0.80	0.60	1.00	0.20	0.50
ES	0.58	0.08	0.21	0.26	0.44	0.63	0.18	0.41
CL	0.43	0.07	0.14	0.25	0.43	0.60	0.14	0.33
GR	0.85	0.24	0.31	0.27	0.23	1.00	0.15	0.87
SM	1.00	0.25	0.57	0.44	0.29	1.00	0.14	1.00
DM	1.00	0.12	0.86	0.19	0.43	0.75	0.14	1.00
CM	1.00	0.07	1.00	0.11	1.00	0.57	0.50	0.40
Average	0.81	0.13	0.56	0.37	0.50	0.82	0.24	0.69

$$P = \frac{N_r}{N_t} \tag{5}$$

where  $N_r$  is the number of total retrieved heart sound records.

In order to search the given textual query in the database in this experiment we used two types of ways: 1) in the first way, we used *Lemur* search engine in order to retrieve the relevant data according to the submitted query. The Lemur Toolkit (version 4.12) is designed to facilitate research in language modeling and information retrieval and 2) in the next step, retrieval run was executed investigating a variety of experimental factors for the retrieval run.

**N-gram window size:** The effects of the window size in accuracy of the retrieval were investigated in different window sizes. Figure 5 reveals the obtained accuracy depends on window sizes. In case of the heart sound, the window size of 5, showed higher performance in comparison with other sizes. The window size of 4 was selected for music files.

**Intervals of heart sounds:** Figure 6 shows the distribution of intervals for the heart sound samples that were used in this study. It reveals the most intervals are less than 5. Hence, the characters that are mapped for each interval can be selected between A and K for positive values and the characters between L and V are utilized for negative values.

Table 2 shows four run of the musical n-gram variants were used in indexing the collection in all categories.

**IR experiments:** The IR measure was an average of Recall and Precision for each run between all categories. The best retrieval recall is 0.84 was obtained by using  $n = 3$  but for diagnosing purpose the precision is more important, that the best precision was obtain in  $n = 5$  that is 0.85. As shown in Table 2, best precision was obtained in Systolic Murmurs category with precision 0.25 and best recall was in Split S2 and Murmurs Category. The lowest recall and precision was in Clicks category. After we run this

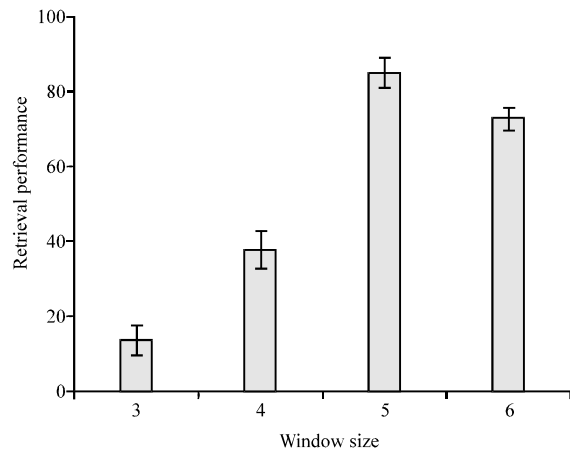


Fig. 5: Average of retrieval accuracy for different window sizes

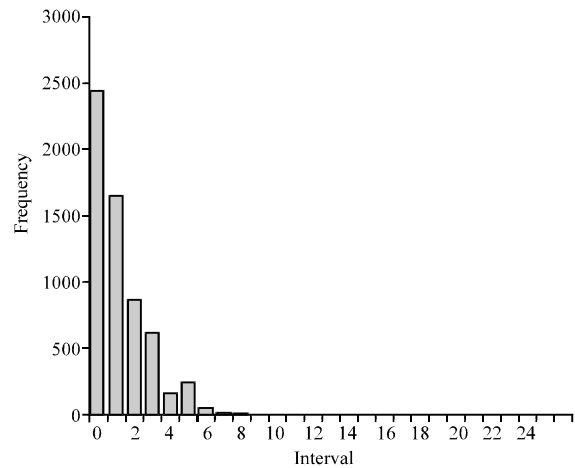


Fig. 6: Distribution of intervals for heart sound samples studied

experiment with  $n = 4, 5$  and  $6$ , thus the best performance was obtained in  $n = 5$  for window size. In this run, the best precision was obtained in these categories: Normal, Split S2, Gallop and Systolic Murmur with 1.00 for precision but recall are decrease against run with  $n = 3$ .

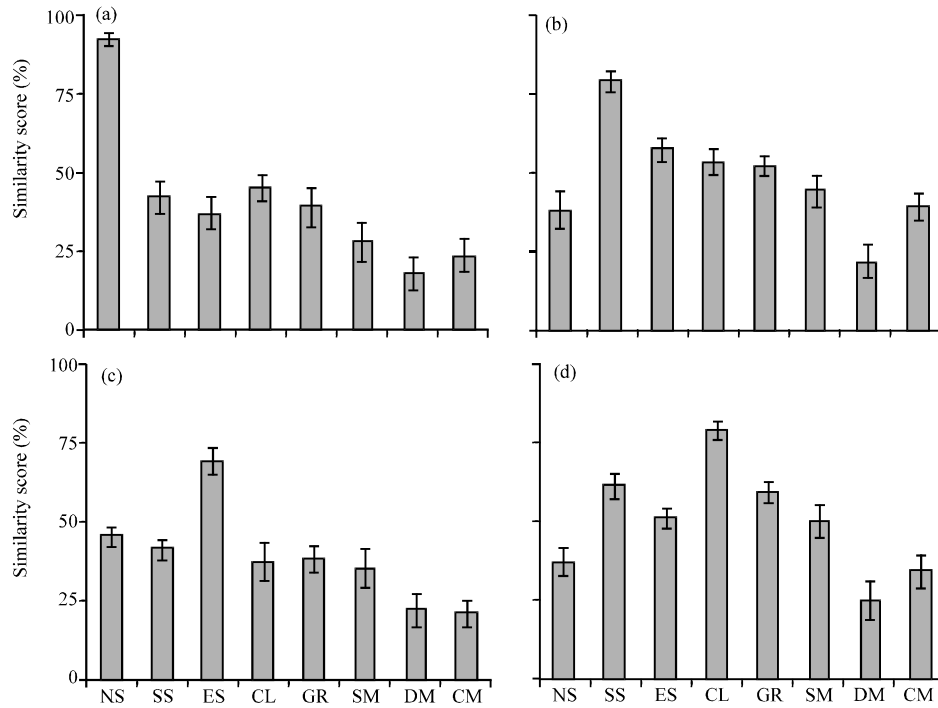


Fig. 7 (a-d): Average similarity rate for (a) normal sound, (b) split second sound, (c) ejection sound and (d) clicks

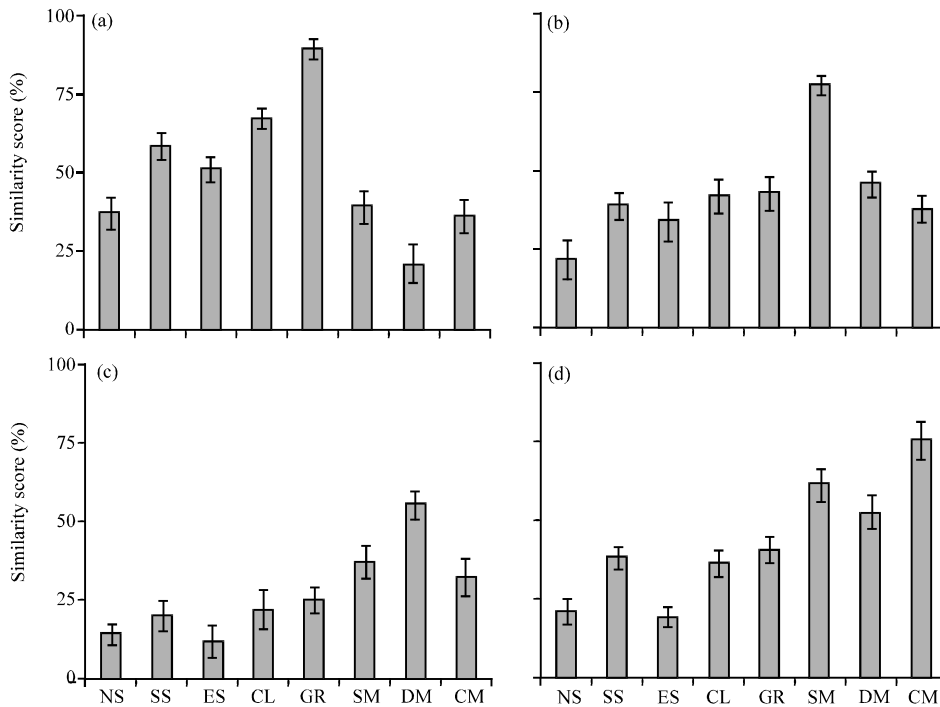


Fig. 8 (a-d): Average similarity rate for (a) gallop, (b) systolic murmurs, (c) diastolic murmurs and (d) continues murmurs

**String-matching experiments:** In the process of saving records in database in order to easily distinguish relevant results, we assign relevant category by using relation

between Category table and Auscultation Area table of the database to show the category of the ranked disease in addition to the name of the disease. This helps us to

find relevant result more conveniently. For instance, Systolic Murmur is the relevant category for Aortic Stenosis problem. In order to compare the similarity of two strings LCS algorithm has been used. Similarities more than 65% is relevant for our method. If result categories are similar with query categories they will be considered as correct retrieved heart diseases. It should be considered that the queries that we used have two important properties, first it should not exist in the pre-created database and second categorization of the query is known.

In order to evaluate the string matching method, 10 samples (out of existing samples in database) were selected and the average of similarity rates was calculated for each category. Figure 7, 8 show the average similarity scores for different categories. Figure 7a is the matching results for normal short queries. It shows a clear difference in comparison with other categories. According to similarity of some categories in case of pulses and rhythms, the similarity of the extracted categories for these groups was shown in the results (Fig. 7d, 8a). In case of diastolic murmurs as shown in Fig. 8c the accuracy of the detection was lower than other categories which because of complexity of this type of disease. As shown in the results, the proposed method could totally retrieve the similar samples with threshold level of 65%. Hence, this threshold is used for performance metric approach.

In contrast with previous study on disease detection (Syeda-Mahmood and Wang, 2008) which heart sounds were analyzed with visual processing, the proposed system uses the heart sounds that are directly converted to the text format. Due to n-gram approach that extracts the interval between a set of n notes, the converted texts do not rely on frequency of different heart samples (children heart sounds are in higher frequency than old person). Also, this technique gives us the ability of searching and indexing.

For the future work, the dynamic value of threshold to automatic detection will be used. In case of real-time system, performing a dynamic threshold is not a simple task. Hence, based on the obtained results, the threshold level of 65% can be selected for the real-time system.

## CONCLUSION

In this study, an automatic heart diseases detection and classification was proposed. Heart sounds which were recorded from various diseases are converted into semi-structural audio formats. It allows applying several text-based techniques in order to search or index the samples. N-gram approach was utilized to provide text-based transcription. Several parameters such as window sizes and samples cycles were investigated. Results

showed that precision with window sizes of 5 notes were better than others. However, it has to be noted that recall with window size of 3 notes were better than another categories.

## REFERENCES

- Amien, M.B.M. and J. Lin, 2007. Dual-mode continuous arrhythmias telemonitoring system. *J. Applied Sci.*, 7: 965-971.
- Andreoli, T.E., C.C.J. Carpenter, R.C. Griggs and I. Benjamin, 2007. *Andreoli and Carpenter's Cecil Essentials of Medicine*. 7th Edn., WB Saunders, Philadelphia, ISBN-13: 978-1416029335.
- Arvin, F. and S. Doraisamy, 2010a. Heart sound musical transcription technique using multi-level preparation. *Int. Rev. Comp. Soft.*, 5: 595-600.
- Arvin, F. and S. Doraisamy, 2010b. Real-time segmentation of heart sound pattern with amplitude reconstruction. *Proceedings of the Biomedical Engineering and Sciences (IECBES)*, Nov. 30-Dec. 2, Kuala Lumpur, Malaysia, pp: 130-133.
- Babaei, S. and A. Geranmayeh, 2009. Heart sound reproduction based on neural network classification of cardiac valve disorders using wavelet transforms of PCG signals. *Comp. Biol. Med.*, 39: 8-15.
- Bergroth, L., H. Hakonen and T. Raita, 2000. A survey of longest common subsequence algorithms. *Proceedings of the 7th International Symposium on String Processing and Information Retrieval*, Sept. 27-29, Spain, pp: 39-48.
- Crawford, M.H., 2009. *Current Diagnosis and Treatment Cardiology*. 3rd Edn., McGraw-Hill Medical, New York, USA., ISBN-13: 978-0071442114, pp: 556.
- Dokur, Z. and T. Olmez, 2008. Heart sound classification using wavelet transform and incremental self-organizing map. *Digital Signal Proc.*, 18: 951-959.
- Doraisamy, S. and S. Ruger, 2003. Robust polyphonic music retrieval with N-grams. *J. Intell. Inf. Sys.*, 21: 53-70.
- El-Asir, B. and K. Mayyas, 2004. Multiresolution analysis of heart sounds using filter banks. *Inform. Technol. J.*, 3: 36-43.
- El-Ramsisi, A.M. and H.A. Khalil, 2007. Diagnosis system based on wavelet transform, fractal dimension and neural network. *J. Applied Sci.*, 7: 3971-3976.
- Gupta, C.N., R. Palaniappan, S. Swaminathan and S.M. Krishnan, 2007. Neural network classification of homomorphic segmented heart sounds. *Applied Soft Comp.*, 7: 286-297.
- Hebden, J.E. and J.N. Torry, 1997. Identification of aortic stenosis and mitral regurgitation by heart sound analysis. *Comput. Cardiol.*, 24: 109-112.



- Hendel, M., A. Benyettou, F. Hendel and H. Khelil, 2010. Automatic heartbeats classification based on discrete wavelet transform and on a fusion of probabilistic neural networks. *J. Applied Sci.*, 10: 1554-1562.
- Itou, N. and K. Nishimoto, 2007. A voice-to-MIDI system for singing melodies with lyrics. Proceedings of the International Conference on Advances in Computer Entertainment Technology, June 13-15, Salzburg, Austria, pp: 183-189.
- Manning, C.D., P. Raghavan and H. Schutze, 2008. An Introduction to Information Retrieval. Cambridge University Press, Cambridge, UK., ISBN-13: 9780521865715, pp: 482.
- Meadow, C.T., B.R. Boyce and D.H. Kraft, 1999. Text Information Retrieval Systems. 2nd Edn., Academic Press, Inc., New York, USA., ISBN-13: 9780124874053, pp: 364.
- Modegi, T., 2001. XML transcription method for biomedical acoustic signals. *Stud. Health Technol. Inform.*, 84: 366-370.
- Phua, K., J. Chen, T.H. Dat and L. Shue, 2008. Heart sound as a biometric. *Pattern Recog.*, 41: 906-919.
- Roark, B., M. Saraclar and M. Collins, 2007. Discriminative n-gram language modeling. *Comput. Speech Lang.*, 21: 373-392.
- Selva Vijila, C.K., M.E.P. Kanagasabapathy and S. Johnson, 2006. Fetal ECG extraction using softcomputing technique. *J. Applied Sci.*, 6: 251-256.
- Swamalatha, R. and D.V. Prasad, 2010. A novel technique for extraction of FECG using multi stage adaptive filtering. *J. Applied Sci.*, 10: 319-324.
- Syeda-Mahmood, T. and F. Wang, 2008. Shape-based retrieval of heart sounds for disease similarity detection. *Comput. Vision*, 5303: 568-581.
- Wang, P., Y. Kim, L.H. Ling and C.B. Soh, 2005. First heart sound detection for phonocardiogram segmentation. Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society, Jan. 17-18, Shanghai, China, pp: 5519-5522.