Relevant Multi Domain Features Selection Based on Mutual Information for Heart Sound Classification

Rima Touahria[1], Abdenour Hacine-Gharbi[1][a], Philippe Ravier[2][b] and Messaoud Mostefai[1]

[1]*LMSE Laboratory, University of Bordj Bou Arréridj, Elanasser, 34030 Bordj Bou Arréridj, Algeria*

[2]*PRISME Laboratory, University of Orleans, 12 rue de Blois, 45067 Orleans, France*

Keywords: Heart Sound, Multidomain Features, Feature Extraction, Feature Selection, Mutual Information, Classification.

Abstract: Many classification systems of the heart sound signals use a combination of features from different domains. In a former reference paper, 324 multidomain features were used for classifying segmented phonocardiogram signals. However, the large feature dimension requires high memory space, high calculus and probably reduces the classification accuracy caused by the curse of dimensionality. In the present work, we propose to reduce the dimensionality of features vectors by selecting the relevant features using six heuristic strategies of feature selection based on mutual information maximisation criterion. In order to validate the selected subset of features, a k-NN model based-classifier was used and evaluated on the PhysioNet/Computing in Cardiology Challenge2016 dataset using the same features sets described in the reference paper. The results demonstrate that the Joint Mutual Information (JMI) selection strategy increases the classification rate from 85. 57% to 89.28% and simultaneously reduces dimension from 324 to 46. Furthermore, this work demonstrates that systolic segment features are the most relevant for murmur/normal classification. It also demonstrates the capability of feature selection algorithms to emphasize specific key areas in signals, which is helpful for aided diagnostic systems and fundamental research.

1 INTRODUCTION

The human heart provides the phonocardiogram (PCG) signal, which can be captured by a traditional or electronic stethoscope. PCG signal processing has mainly two goals. The first goal is to divide the PCG signal into heart cycles and to detect the successive components that make up each cardiac cycle: first heart sound (S1), systolic period (Sys), second heart sound (S2) and diastolic period (Dias). Heart sounds (S1, S2) are audible signals associated with the closing of valves. The time duration of them is approximately 150 ms and 120 ms respectively with a corresponding frequency between 20 Hz to 150 Hz. The second goal consists to classify the heartbeats in a PCG signal into normal and abnormal heart sounds for diagnostic of cardiovascular diseases.

The heart sound classes can be identified by a feature extraction step followed by a classification step. Techniques for feature extraction may use Discrete Wavelet Transform, mel-frequency cepstral coefficient (MFCC). Classification may use algorithms such as k-Nearest Neigbors (k-NN), Artificial Neural Network (ANN), Support Vector Machine (SVM) (Ortiz, Phoo, & Wiens, 2016), (Jinghui, Li Ke, & Qiang Du, 2019) . In (Rubin, et al., 2016), the authors have used the MFCC with a deep convolutional neural network algorithm. (Tang, Chen, Li, & Zhong, 2016) presented a method using multi-domain features. In (Touahria, Hacine-Gharbi, & Ravier, 2021), the authors have proposed the use of LWE (Log Wavelet Energy) features to automatically classify the PCG signal in a class label "normal" (N) or "abnormal" (AN). A survey paper on heart sound classification methods is published by (Liu, et al., 2016). Particularly, in (Tang, Chen, Li, & Zhong, 2016), the authors proposed a classifier applied on multi-domains features for PCG classification based on Back Propagation Neural Network. In this work, a set of 324 multi-domains features (domain of heart sound intervals, energy domain, frequency spectrum,

[a] https://orcid.org/0000-0002-7045-4759
[b] https://orcid.org/0000-0002-0925-6905

Touahria, R., Hacine-Gharbi, A., Ravier, P. and Mostefai, M.
Relevant Multi Domain Features Selection Based on Mutual Information for Heart Sound Classification.
DOI: 10.5220/0012565300003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 918-923
ISBN: 978-989-758-684-2; ISSN: 2184-4313

heart rate sequence, frequency spectrum of heart rate sequence, kurtosis, cyclostationarity, power spectral density and power spectral density of heart rate sequence) has been used, that lead to classification accuracy of 83.6%. However, this high dimension of features vectors can reduce the performance of the classification system of PCG signals in terms of complexity (memory space, calculus time) and probably in accuracy. The principal aim of the present work is to reduce this high dimension using feature selection algorithms for a lower complex system with possible higher classification accuracy. Two principal features selection approaches have been used in the states of arts. The first "wrapper" approach is applied in low dimension and uses the classifiers to measure the relevancy of features. Conversely, the second 'filter' approach, which is independent of classifiers, is generally applied in high dimension and uses the information provided by the features to explain the classes. Hence, in our work, we use a "filter" approach based on the criterion of mutual information maximisation of the selected features. Several heuristic strategies of feature selection based on mutual information will be applied to select the relevant local and global features (S1, Systole, S2 and Diastole) extracted from the previous set of multi-domains features. In order to validate the importance of this features selection, we propose to evaluate the performance of the classification system using k-NN classifier using the cross-validation with 5 folds applied on the same database used in (Tang, Chen, Li, & Zhong, 2016).

The organization of this study is as follows. Section 2 shows the suggested approaches for PCG signals classification and the related work, *i.e.* extracting feature vectors and choosing the relevant features for the classification task. Section 3 describes the proposed classification system, which adds the feature selection step. The experiments and their findings are presented in Section 4. Section 5 concludes the paper with some ideas for further research.

2 RELATED WORK

Several studies on heart sound classification have used the pattern recognition approach for the task of cardiovascular diseases diagnosis (Barschdorff, Bothe, & Rengshausen, 1989) (Ali, et al., 2019) (Whitaker, Suresha, Liu, Clifford, & Anderson, 2017). This approach requires three steps: pre-processing and segmentation, features extraction with optional step of features selection, and classification. Firstly, the phonocardiogram (PCG) is pre-processed and segmented in local regions (S1, Sys, S2, Dias). Then, the features of each PCG recording are extracted. Finally, the features are fed into the designed model to classify normal and abnormal heart sound. As a result, traditional classification system for heart sounds includes the steps listed below. The conception of this system requires training phase for building the model of each class and testing phase for evaluating performance of the classification system using training and testing databases.

In (Tang, Chen, Li, & Zhong, 2016), the authors have proposed a system of PCG signals classification based on Back Propagation Neural Network classifier applied on sequences of feature vectors extracted from several domains. In this system, the segmentation step is based on the hidden semi-Markov model (HSMM) method that uses the ECG information to locate the different local regions of the heartbeat sound (Springer, Tarassenko, & Clifford, 2016). Then, the global and the local regions (S1, Systole, S2 and Diastole) of the heartbeat sound have been used to extract local and global of several multi-domains features in (Springer, Tarassenko, & Clifford, 2016) (Tang, Chen, Li, & Zhong, 2016). Hence, each feature vector representing the heartbeat sound is composed of the concatenated features vectors extracted on each local region and the global region. However, this concatenation increases the vectors dimension which augments the space memory, computing time and probably reduces the accuracy caused by the curse of dimensionality phenomenon.

3 PROPOSED CLASSIFICATION SYSTEM

3.1 Descriptions of the Heart Sound Databases

The dataset described in (Tang, Chen, Li, & Zhong, 2016) is used in the present work. This includes six databases (labeled A through F), collecting a total of 3153 phonocardiogram (PCG) recordings. These recordings were gathered from diverse settings, including clinical and non-clinical environments, and involve subjects ranging from healthy individuals to those with pathological conditions. Each PCG recording has undergone manual labeling, indicating whether it is categorized as normal (-1) or abnormal (1). The database is constituted of 2500 PCG recordings of normal class and 653 of PCG recordings of abnormal class. In the present work, this database is partitioned into five folds for cross-validation evaluation.

3.2 Flowchart of the System

In this work, we propose to reduce the dimensionality of the feature vectors described previously using the feature selection approach based on the mutual information for the task of heartbeat sounds classification. Particularly, the k-NN classifier is used for its simplicity (k is the number of neighbors), with applying the same segmentation step and the same features dataset used in (Tang, Chen, Li, & Zhong, 2016). The flowchart of the proposed classification system of the heartbeat sounds is illustrated on Figure 1.

The two steps of segmentation and multidomain feature extraction are carried out by following the procedure given in the reference paper in (Tang, Chen, Li, & Zhong, 2016). First, PCG recordings are segmented into primary heart sounds, including S1, Systole, S2, and Diastole, through the application of the Hidden Semi-Markov Model (HSMM) method originally introduced by Springer (Springer, Tarassenko, & Clifford, 2016). Second, features are extracted from each segment or between segments in multidomain which gives a set of 324 features. The domains and the number of features per domain are those presented in (Tang, Chen, Li, & Zhong, 2016): 22 features in domain of heart sound intervals, 10 features in energy domain, 82 features in frequency spectrum, 2 features in heart rate sequence, 57 features in frequency spectrum of heart rate sequence, 8 features in Kurtosis, 4 features in cyclostationarity, 82 features in power spectral density and 57 of features in power spectral density of heart rate sequence.

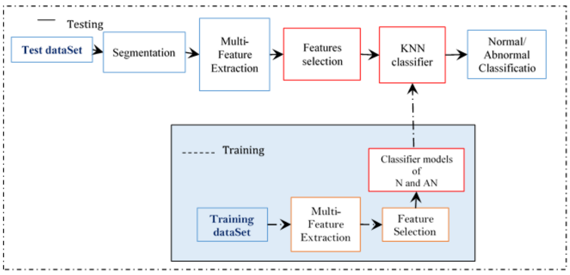

Figure 1: flowchart of the proposed classification system of the PCG heart sounds.

The feature selection step considered in this work consists to select the relevant features using the mutual information maximization. This step is carried out using several feature selection strategies such as JMI (Yang & Moody, 1999), ICAP (Jakulin, 2005), CIFE (Kojadinovic, 2005), MRMR (Peng, 2005) and CMI (Fleuret, 2004). These strategies are implemented using the feast Matlab toolbox (Brown, Pocock, Zhao, & Lujan, 2012). This step will be described in the next section.

Performance measures of the classification system are evaluated using the classification rate (CR), defined as follow:

$$CR = \frac{\text{number of recognised testing signals}}{\text{total number of testing signals}} \tag{1}$$

3.3 Feature Selection Based on Mutual Information

The feature selection consists to select a small subset of N relevant features $\mathcal{S}_{opt} = \{p_{s_1}, p_{s_2}, \dots, p_{s_N}\}$ that explains the different classes of signals, from an initial set of M features $F = \{p_1, p_2, \dots, p_M\}$ that produces the maximal mutual information with the following class variables:

$$\mathcal{S}_{opt} = arg \max_{S \subset F} I(C; S) \tag{2}$$

This can be performed using the forward 'greedy' algorithm, which selects at each iteration j one feature p_{s_j} that verifies the following equation:

$$p_{s_j} = arg \max_{p_i \in F - \mathcal{S}_{j-1}} [I(C; p_i, \mathcal{S}_{j-1})] \tag{3}$$

Since we have $I(C; p_i, \mathcal{S}_{j-1}) = I(C; \mathcal{S}_{j-1}) + I(C; p_i \backslash \mathcal{S}_{j-1})$ (Cover & Thomas, 1991), Equation (3) can be reduced to:

$$p_{s_j} = arg \max_{p_i \in F - \mathcal{S}_{j-1}} [I(C; p_i \backslash \mathcal{S}_{j-1})] \tag{4}$$

Two feature selection approaches are considered in the state of arts. The first approach 'wrappers' uses the accuracy of the classification system as measure of features relevancy. This approach is applied in low dimension cases (Kohavi & John, 1997). The second approach 'Filters' is independent of the classification system. This approach is adapted for the high dimension cases. In our work, we use the 'Filters' approach because the classification system is based on extraction of features vectors of high dimension. Next, five strategies of feature selection will be described.

- JMI (Joint Mutual Information) (Yang & Moody, 1999)

$$p_{s_j} = arg \max_{p_i \in F - \mathcal{S}_{j-1}} \left[I(C; p_i) - \frac{1}{j-1} \sum_{k=1}^{j-1} I(C; p_i; p_{s_k}) \right] \tag{5}$$

- ICAP (Interaction Capping) (Jakulin, 2005)

$$p_{s_j} = arg \max_{p_i \in F - S_{j-1}} \left[I(C;p_i) - \sum_{k=1}^{j-1} max\,[I(C;p_i;p_{s_k}),0] \right] \quad (6)$$

- CIFE (Conditional Infomax Feature Extraction) (Kojadinovic, 2005) (Hacine-Gharbi, Ravier, & Mohamadi, 2009)

$$p_{s_j} = arg \max_{p_i \in F - S_{j-1}} \left[I(C;p_i) - \sum_{k=1}^{j-1} I(C;p_i;p_k) \right] \quad (7)$$

- MRMR (Maximum-Relevance Minimum Redundancy) (Peng, 2005)

$$p_{s_j} = arg \max_{p_i \in F - S_{j-1}} \left[I(C;p_i) - \frac{1}{j-1} \sum_{k=1}^{j-1} I(p_i;p_{s_k}) \right] \quad (8)$$

- CMI (Conditional Mutual Information) (Fleuret, 2004)

$$p_{s_j} = arg \max_{p_i \in F - S_{j-1}} \left[I(C;p_i) - \max_{p_{s_k} \in S_{j-1}} I(C;p_i;p_{s_k}) \right] \quad (9)$$

4 EXPERIMENTAL RESULTS

The aim of this paper is to select the relevant features from a set of combined features used in (Tang, Chen, Li, & Zhong, 2016) for the task of PCG signals classification. This is realized using these steps:

- Apply the feature selection method based on the MI with different strategies JMI, ICAP, CIFE, MRMR and CMI;
- Use the k-NN classifier to validate the relevancy of the features selected at the iteration j;
- Estimate the optimal number of features with the two stopping criterions considered (Touahria, Hacine-Gharbi, & Ravier, 2023): in the first criterion (CRT1), the optimal feature subset is the selected feature subset that yields a CR higher than or equal to the CR obtained using the set of all features (324 features); in the second criterion (CRT2), the optimal feature subset yields to the maximal CR. Note that the second criterion requires more time than the first criterion.

4.1 Optimal Nearest Neighbors (K Considering Two Stopping Criterions) Parameterization

This experiment has the objective to search for the optimal number of nearest neighbors (K) that gives he best performance by varying K from 1 to 30 and by changing the distance function (“Correlation”, “Cosines”, “Euclidean”, “Cityblock”). Table 1 gives the results. The best result is obtained when K is set to 8 and the function is set to “Cityblock”. It reaches a CR of **85. 57**%.

Table 1: CR as a function of optimal number of nearest neighbors K with four distance functions.

	Euclidean	Correlation	Cityblock	Cosine
K optimal	6	4	**8**	4
CR (%)	83.34	83.38	**85. 57**	83.00

In the next sections, the value of nearest neighbor’s K will be set to 8 and the “Cityblock” function will be chosen.

4.2 Performance Study for Feature Selection Using Different Strategies

This section focuses on the selection of the most relevant features that explain the normal and abnormal classes using the JMI, ICAP, CIFE, MRMR and CMI strategies.

Table 2: CR and number of relevant features by using JMI, ICAP, CIFE, MRMR and CMI strategies with selection criteria CRT1 and CRT2.

	CRT1 CR ≥ CR(end)		**CRT2 CR==max(CR)**	
	# of relevant features	***CR (%)***	***# of relevant features***	***CR (%)***
JMI	**7**	**86.71**	**46**	**89.28**
ICAP	13	86.97	22	87.95
CIFE	72	85.95	94	86.90
MRMR	**13**	86.11	38	86.81
CMI	6	85.63	23	89.06

The outcomes of this selection for the different feature selection strategies in terms of CR and relevant feature number using the two stopping criterions CRT1and CRT2 are shown in Table 2. From this table, we can give the following points:

- In the case of the first criterion, we notice a strong convergence in the values of CR for all strategies, which is around 86%. Particularly, the JMI strategy gives the best trade-off between accuracy of 86.71% and optimal number of 7 features.
- In the case of the second criterion, the highest CR of 89.28% is achieved using the JMI strategy. However, the CMI gives the best

trade-off between accuracy of 89.06% and optimal number of 23 features, which represents a dimension reduction of 92.90%. In addition, the ICAP gives the smallest feature subsets with CR of 87.95%.

- The CIFE strategy selects the largest subset of features whatever the considered criterion.

Figure 2 shows the results obtained by the five strategies (JMI, ICAP, CIFE, MRMR and CMI) applied on vectors of multidomain and with different feature domains. From these curves, it can be seen that approximately **46** selected features obtained by JMI strategy are adequate for clearly explaining the classes. Note that we observe that all strategies identified the same first feature, which is the systolic (“sd_energy_SysCycle”) energy as previously demonstrated in (Touahria, Hacine-Gharbi, & Ravier, 2023).

Further, this figure shows clearly the curse dimensionality phenomenon explained by the great peak in the CR curves corresponding to JMI, CIFE, MRMR. Generally, the most features selection strategies improve the performances in terms of complexity (space memory and computing time) and accuracy.

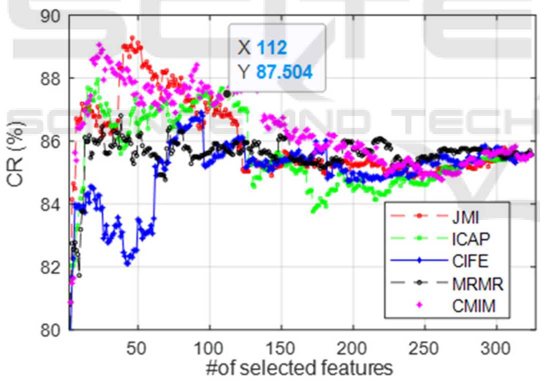

Figure 2: CR (%) as a function of the number of selected features with the five feature selection strategies.

5 CONCLUSIONS

In this study, we have proposed the use of several feature selection strategies based on the criterion of mutual information maximization for reducing high dimensionality vectors composed of 324 features of multi-domains types extracted from database of PCG recordings used in a previous system of PCG signals classification. The dataset of these features is used for evaluating the performance of the proposed classification system of PCG signals based on k-NN classifier combined with feature selection algorithms, using five folds cross-validation strategy.

The obtained results demonstrate that including the feature selection step in the classification system of PCG signals improves the performance in terms of accuracy and complexity with high dimension reduction of features vectors. We report a high reduction of the dimension number (from 324 to **46**) and an increased **89.28%** CR value using feature selection procedure based on JMI strategy by applying the second criterion. These results demonstrate the efficiency of the feature selection step for reducing the complexity and increasing the accuracy of the classification system of PCG signals.

ACKNOWLEDGEMENTS

If any, should be placed before the references section without numbering.

REFERENCES

Ali, R., Arif, M., Saleem, U., Maqsood, A., Gyu, S., & Byung-Won, O. (2019). Heart sound classification based on temporal alignment techniques. Sensors 19; 4819.

Barschdorff, B., Bothe, A., & Rengshausen, U. (1989). Heart sound analysis using neural and statistical classifiers a comparison. Comput. Cardiol. 415–418.

Brown, G., Pocock, A., Zhao, M.-J., & Lujan, M. (2012). Conditional Likelihood Maximisation A UnifyingFramework for information theoretic feature selection. Journal of Machine Learning Research, 13(1), pp.27-66.

Cover, T., & Thomas, J. (1991). Elements of Information Theory. . New York: Wiley Series In Telecommunications.

Fleuret, F. (2004). Fast binary feature selection with conditional mutual information. Journal of Machine Learning Research, 5:1531–1555.

Hacine-Gharbi, A., Ravier, P., & Mohamadi, T. (2009). Une nouvelle méthode de sélection des paramètres pertinents: application en reconnaissance de la parole. Proceedings of the Conference Traitement et Analyse de l’Information:Méthodes et Applications (TAIMA), pp.399–407.

Jakulin, A. (2005). Learning based on attribute interactions. hD thesis, University of Ljubljana, Slovenia.

Jinghui, L., Li Ke, & Qiang Du. (2019). Classification of Heart Sounds Based on the Wavelet Fractal and Twin Support Vector Machine. Entropy , 21, 472;doi:10.3390/e21050472www.mdpi.com/journal/entropy.

Kohavi, J., & John, G. (1997). Wrappers for feature subset selection. Artificial Intelligence, Vol. 97, Nos. 1/2, pp.273–324.

Kojadinovic, I. (2005). Relevance measures for subset variable selection in regression problems based on k-additive mutualinformation. Computational Statistics and Data Analysis,Vol. 49, No. 4, pp.1205–1227.

Li, F., Tang, H., Shang, S., Mathiak, K., & Cong, F. (2020). Classification of Heart Sounds Using Convolutional Neural Network. Appl. Sci. 10 (11)3956.

Liu, C., Springer, D., Li, Q., Moody, B., Juan, R., Chorro, F., . . . Clifford, G. (2016). An open access database for the evaluation of heart sound algorithms. Physiological Mea-surement.

Ortiz, J., Phoo, C., & Wiens, J. (2016). Heart sound classification based on temporal alignment techniques. IEEE.

Peng, H. L. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 8, pp. 1226–1238.

Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., & Sricharan, K. (2016). Classifying heart sound recordings using deep convolutional neural and mel-frequency cepstral coefficients. Computing in Cardiology Conference (CinC), IEEE, pp. 813–816.

Springer, D., Tarassenko, L., & Clifford, G. (2016). Logistic regression-HSMM-based heart sound segmentation. *Transactions on Biomedical Engineering (IEEE)*, 822-32.

Tang, H., Chen, H., Li, T., & Zhong, M. (2016). Classification of Normal/Abnormal Heart Sound Recordings Based on Multi-Domain Features and Back Propagation Neural Network. (pp. 593–596). In Proceedings of the 2016 Computing inCardiology Conference (CinC).

Touahria, R., Hacine-Gharbi, A., & Ravier, P. (2021). Discrete Wavelet based Features for PCG Signal Classification using Hidden Markov Models. International Conference on Pattern Recognition Applications and Methods.

Touahria, R., Hacine-Gharbi, A., & Ravier, P. (2023). Feature selection algorithms highlight the importance of the systolic segment for normal/murmur PCG beat classification. Biomedical Signal Processing and Control.

Whitaker, B., Suresha, P., Liu, C., Clifford, G., & Anderson, D. (2017). Combining sparse coding and time domain features for heart sound classification. Physiol. Meas.1701 .

Yang, H., & Moody, J. (1999). Feature selection based on joint mutual information. Intelligent Data Analysis (AIDA) and Computational Intelligent Methods and Application (CIMA).