EFFECTIVE NOISE REMOVAL AND UNIFIED MODEL OF HYBRID FEATURE SPACE OPTIMIZATION FOR AUTOMATED CARDIAC ANOMALY DETECTION USING PHONOCARDIOGARM SIGNALS

Arijit Ukil, Soma Bandyopadhyay, Chetanya Puri, Rituraj Singh, Arpan Pal

Research and Innovation, Tata Consultancy Services, Kolkata, India
{arijit.ukil, soma.bandyopadhyay, chetanya.puri, singh.rituraj, arpan.pal}@tcs.com

ABSTRACT

In this paper, we present completely automated cardiac anomaly detection for remote screening of cardio-vascular abnormality using Phonocardiogram (PCG) or heart sound signal. Even though PCG contains significant and vital cardiac health information and cardiac abnormality signature, the presence of substantial noise does not guarantee highly effective analysis of cardiac condition. Our proposed method intelligently identifies and eliminates noisy PCG signal and consequently detects pathological abnormality condition. We further present a unified model of hybrid feature selection method. Our feature selection model is diversity optimized and cost-sensitive over conditional likelihood of the training and validation examples that maximizes classification model performance. We employ multi-stage hybrid feature selection process involving first level filter method and second level wrapper method. We achieve 85% detection accuracy by using publicly available MIT-Physionet challenge 2016 datasets consisting of more than 3000 annotated PCG signals.

***Index Terms*—**PCG; anomaly; noise filtering; feature optimization; sensor analytics

1. INTRODUCTION

One of the game changers in IoT would be proactive, round-the-clock, remote health monitoring without human intervention for the health condition related alert generation. In this work, we focus on effective cardiac management that is affordable and preferably does not require additional hardware other than smartphone and simple enough to carry out in-house monitoring. Our proposed scheme RONUN (RObust Noise filtering and UNified feature selection Method) is primarily an automated, preventive cardiac management solution that uses smartphone or other wearable sensor-captured heart sound, PCG signal. PCG detects cardiac auscultation, which is a fundamental cardiac condition analysis that is performed by doctors through listening the heart sound signal or PCG using stethoscope.

PCG is prone to various corruption, distortion and different artifacts. Such corrupted PCG signals are to be eliminated for ensuring meaningful and less erroneous subsequent decision making process. One of our primary contributions is to develop effective noise identification algorithm that is capable in differentiating noisy and clean PCG signals. Secondly, we derive a unified model of hybrid feature space optimization from a diverse set of criteria functions from the conditional likelihood of the training class labels given the features. Traditionally, we encounter two types of feature selection approaches [10]. Filter method: where selection rule is independent of underlying classifier; Wrapper method: where feature selection rule is dictated by the classifier performance. The proposed model is a novel hybrid approach, which is wrapper over filter method of feature selection such that learning process capability is maximized. Formally, let $\mathcal{A}$ be the learning method (classifier), Π be the wrapper method, $\mathcal{L}$ be the cost function, the performance function be $\mathcal{P}$, $\mathbb{S}$ be the set of diverse feature selection criteria (For example, $\mathbb{S}$ may consist of minimum Redundancy Maximum Relevance (mRMR) [1], Conditional Mutual Information Maximization (CMIM) [2], Interaction Capping (ICAP) [3], and other filter feature selection methods) and $\mathcal{F}_{select}$ be the selected top K features from the full feature set $\mathbb{F}$. Our goal is to find $\mathcal{F}_{select}$; $\mathcal{F}_{select} \subseteq \mathbb{F}$:

$$\mathcal{F}_{select} = \underset{\mathbb{F}}{argmin}\, \mathcal{L}\left(\Pi\left(\mathcal{A}\big(\mathbb{S}(\mathbb{F})\big)\right)\right)$$
$$\mathcal{F}_{select} = \underset{\mathbb{F}}{argmax}\, \mathcal{P}\left(\Pi\left(\mathcal{A}\big(\mathbb{S}(\mathbb{F})\big)\right)\right) \quad (1)$$

Filter method finds the feature set that maximizes over some criteria on the training set distribution. Different filter methods with different selection criteria would provide different set of selected feature list. Wrapper function tries to find the feature set that maximizes classifier performance (over-fitting problem). Our main novelty is proposing a hybrid feature selection method that optimally selects the feature subset $\mathcal{F}_{select}$ over diverse filter-based feature selection methods to maximize the learning capability of the classifier model with reduced over-fitting error. Thus we can capture the training set distribution as well as maximize the learner model performance.

2. RELATED WORKS AND CONTRIBUTION

RONUN addresses two important aspects of accurate cardiac event detection: 1. Identification of noisy PCG signal, 2.

Detection of cardiac abnormality from PCG signal with high accuracy. We propose optimal feature selection through robust learning method. Researchers have made attempts to identify corruption in physiological signals like PCG. In [4, 16], authors aim to filter out the high frequency noise in PCG signal. But, this is an easier problem to solve than to identify noisy PCG segments consisting of temporal and spectral disturbances. For cardiac anomaly detection, different classification methods are proposed, like hidden Markov model [5], support vector machine [6]. It is often argued that presence of noise or failure to identify noisy PCG signal is one of the reasons of low accuracy of automated PCG based cardiac event detection algorithms [7]. Another challenge is to find the optimal feature set for robust learner model construction [15]. Different filter methods optimize on the training data distribution [1– 3], wrapper or hybrid method maximizes classifier model performance [8, 13]. We propose a unified model of diverse filter methods with wrapper functions which attempts to provide training data distribution-centric classifier performance maximization.

3. FUNCTIONAL FLOW

RONUN consists of three main segments: 1. Noisy PCG signal identification and rejection, 2. Optimal feature selection, and 3. Classification of clean PCG signal to determine the presence of cardiac abnormality. In Figure 1, we show the detailed functional flow.

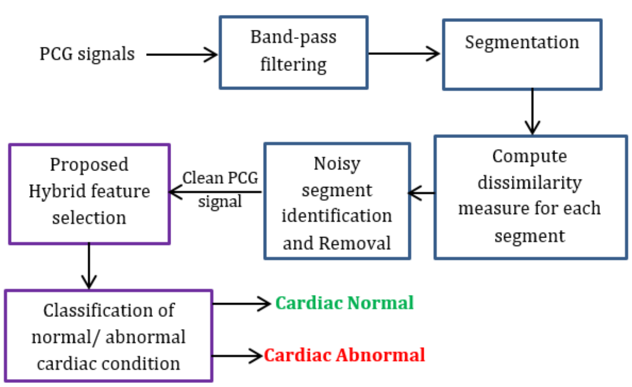

Fig. 2. RONUN functional flow

4. RONUN SCHEME AND ALGORITHMS

RONUN successfully implements and automates noisy PCG signal elimination and subsequently detects clinical abnormality condition from the clean PCG signals.

4.1. Corrupt (Noisy) Non-stationary Physiological (PCG) Signal Identification and Filtering

Our first task is to develop a method that identifies and isolates clean segments ($\Psi_i^{C\prime}$) from noisy segments ($\Psi_i^{N\prime}$), $\Psi = [\Psi^{C\prime}, \Psi^{N\prime}]$, which is a non-trivial task. We apply Dynamic Time Warping (DTW) feature to identify the noisy segments. The amount of noisy segments in a PCG signal helps us to identify whether that PCG signal is noisy. It measures the intrinsic dissimilarity between $\Psi_i^{C\prime}$ and $\Psi_i^{N\prime}$. It is a semi-supervised technique. First, we consider a template $\mathbb{T}$, which is an ideal representation of a segment of $\mathcal{P}_f$, $\mathbb{T} := \{\mathrm{t}_1, \mathrm{t}_2, \ldots, \mathrm{t}_M\}$, $M = \frac{60 \times f_s}{HR_{ideal}}$, where f_s be the sampling frequency and HR_{ideal} be the ideal human heart rate, $HR_{ideal} = 72$ beats per minute. In order to counter non-linearity in segment lengths, we compute the most probable segment length l_p of $\mathcal{P}_f$ form the segment series Ψ_i , $i = 1,2, \ldots. K$. We consider DBSCAN method to find the most likely segment distance set l_p from the set of segment length set $l_i, i = 1,2, \ldots. K, \Psi_i \longrightarrow \Psi'_k$, where each of Ψ'_k is of length l_p [14]. DBSCAN finds the outliers in the segment lengths l_i. Next, we compute DTW distance between the template of length M and each of the segments normalized to length l_p. DTW distance $\delta_{\Psi'_k \mathbb{T}}$ is computed between template $\mathbb{T}$ and each of the segments $\Psi'_k = \left\{\omega_1, \omega_2, \ldots, \omega_{l_p}\right\}_k$ of the extracted PCG signal of lengths M, l_p respectively.

4.2. Unified Model of Hybrid Feature Space Optimization

Classification methods are significantly dependent on the selected features. Optimality of feature selection process is always relative to some objective function. We choose the objective function as the maximization conditional likelihood of the class labels from the target class given the features [10], where the conditional likelihood function is mutual information. Different filter-selection criteria are proposed in the literature [1 – 3, 10] to achieve the objective function.

Let, training data $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{M}$, where $x_i \in \mathbb{R}^d$ be the complete training instances, $y_i \in \mathcal{Y} = \{+1, -1\}$ be the corresponding class labels. Let, testing data $\mathcal{T} = \{(t_i, y_i)\}_{i=1}^{N}$. Let, validation dataset $\mathcal{V} = \{(v_i, y_i)\}_{i=1}^{P}$, and classifier training set $\mathcal{U} = \{(u_i, y_i)\}_{i=1}^{Q}$, $\mathcal{U}, \mathcal{V} \subseteq \mathcal{X}, P + Q = M$. Classifier is trained with $\mathcal{U}$ and the generated model and feature space are validated with $\mathcal{V}$; $\mathcal{U}$ is used for filter method selection and $\mathcal{V}$ is used for wrapper method evaluation. Let, $\mathbb{S}$ be the set of diverse feature selection criteria. For example $\mathbb{S} = \{mRMR, CMIM, ICAP, JMI, ..\}$.

Let $\mathbb{F} = \{f_1, f_2, \ldots, f_z\}$ be the complete set of features of the given data set $\mathcal{X}$ and $\mathcal{F}_{select} = \{(f_j)\}_{j=1}^{o}$ be the selected feature set by the feature selection method $\mathcal{W}$, where $\mathcal{F}_{select} \subseteq \mathbb{F}$, where o be the cardinality of $\mathcal{F}_{select}$, $o \leq z$, and $\mathcal{F}_{select}$ unifies the feature space stability among $\mathbb{S}$ such that learner performance function $\mathcal{P}$ is maximized and diversity of choice $\mathcal{D}$ among $\mathbb{S}$ is minimized. We assume $\mathcal{P} = \mathcal{L}^{-1}$ and hence, maximizing $\mathcal{P}$ is sufficient. Informally, diversity of choice minimization means that feature is not selected as a matter of random chance from the training set distribution of $\mathcal{U}$ or by the arbitrary performance of the learner over $\mathcal{V}$.

Definition: Diversity of choice $\mathcal{D}$ among $\mathbb{S}$ minimization through unification:

$$\min \mathbb{U}\,(\mathbb{F}, \mathbb{C}|\mathbb{S}),\ \ \mathbb{U} = \lambda_{\{f_1, f_2, \ldots, f_o\}}(\mathbb{C}, \mathcal{U}) \quad (2)$$

Where, $\lambda_{\{f_1, f_2, \ldots, f_o\}}(\mathbb{C})$ is a diversity criteria between feature subset $\mathcal{F}_{select}$ and the target class label $\mathbb{C}$ on the classifier training dataset $\mathcal{U}$, over the conditional likelihood probability $\mathbb{I}(x, y)$ for a large set of selection criteria $\mathbb{S}$ = (ICAP, JMI, DISR, CONDRED, mRMR, MIFS, CIFE, CMIM).

Definition: Performance function $\mathcal{P}$ maximization:

$$\max \mathbb{V}\,(\mathbb{F}, \mathbb{C}|\mathbb{S}),\ \ \mathbb{V} = \beta_{\{f_1, f_2, \ldots, f_o\}}(\mathbb{C}, \mathcal{V}|\mathcal{A}) \quad (3)$$

Where, $\beta_{\{f_1, f_2, \ldots, f_o\}}(\mathbb{C})$ is a classification performance criteria between feature subset $\mathcal{F}_{select}$ and the target class label $\mathbb{C}$ on the classifier performance validation dataset $\mathcal{V}$, by the classifier method $\mathcal{A}$ for a large set of selection criteria $\mathbb{S}$ = (ICAP, JMI, DISR, CONDRED, mRMR, MIFS, CIFE, CMIM).

The selected feature set $\mathcal{F}_{select}$ is the top $\mathcal{K}$ features from complete feature set $\mathbb{F}$ satisfying at least equation (3). It is to be noted equation (3) is a sufficient condition to achieve the goal of equation (1), whereas (2) is necessary to ensure diversity of choice among diverse set of feature selection criteria to reduce the random guessing error. Here, we illustrate a solution for equation (2) and (3) in exemplary form. For each $\mathbb{S}$, on the labeled classifier training set $\mathcal{U}$, each of the features in $\mathbb{F}$ has unique ranks up to z, i.e. $\mathbb{F}$ transforms to a vector of dimension j, where j be the total number of selection criteria in $\mathbb{S}, \mathbb{F} \rightarrow \mathbb{F}^j, f_i \rightarrow f_i^j, i = 1,2, \ldots. z$. Let, the feature selection set $\mathbb{S}$ consists of h number of selection criteria, which would be mRMR, JMI, MIFS, etc.., $\mathbb{S} = \{\mathfrak{s}_1, \mathfrak{s}_2, \ldots, \mathfrak{s}_h\}$ and we require to find the top $\mathcal{K}$ feature set from $\mathbb{F} = \{f_1, f_2, \ldots, f_z\}, \mathcal{K} < z$. Let, $f_i^h, i = 1,2, .., z$ be the vector of dimension h of i^{th} feature $\mathbb{F}$ consisting of its rank by each of the h from the selection criteria set $\mathbb{S}$. We compute the centroid ς_i of the denser cluster by *k-means* (*k*= 2) clustering on f_i^h ; $f_i \rightarrow \varsigma_i$. The rank of f_i is function of ς_i, lower the value of ς_i, higher is the rank. The top ranked feature among the feature set $\mathbb{F}$ is the one with $\min_i \varsigma_i$, $i = 1,2, .., z$. In case of conflict, unbiased tossing would be done to uniquely rank each of the feature elements. Thus, we satisfy equation (2). Let's call the above stated method of feature ranking method our filter method (FM) and consequently, we get $h + 1$ feature ranking method. Performance maximization criterion (equation (3)) that formulates the wrapper method is hard to implement, alternatively we relax the condition such that those features among the complete set $\mathbb{F}$ that is capable of maximization of certain performance function $\mathcal{P}$. For example, $\mathcal{P}$ can be F1-score or geometric mean of sensitivity and specificity so that performance on validation data is balanced on both type-I and type-II errors. In order to find top $\mathcal{K}$ feature set from $\mathbb{F}$, we evaluate performance on validation data $\mathcal{V}$ by the learner $\mathcal{A}$ from the top feature of each of the $h + 1$ methods. The feature is selected if the next best feature has not less that Δ times of the previous best, where Δ can be typically 0.95. Strictly, it should be Δ =1. The next best feature from filter method would be discarded if its performance in combination of previous best method is less than Δ times of the previous best performance. This iteration would stop either when top K features are found. The above mentioned method would ensure that features are not only discretely best; it is also equally good in combination. Further numerical example would clarify the proposed method.

Numerical example. Let, there be total *z*=4 features, 7 different feature selection criteria in $\mathbb{S}$ and top 2 features, are to be chosen, $z = 4, h = 7, \mathcal{K} = 2$ and the rank of the four features from the seven criteria over training data $\mathcal{U}$ be as: $f_1 \rightarrow \{3,1,4,2,2,1,3\}, f_2 \rightarrow \{2,3,1,4,3,4,2\}, f_3 \rightarrow \{4,4,3,3,4,3,4\}$, $f_4 \rightarrow \{1,2,2,1,1,2,1\}$; the centroid of denser cluster after performing 2-means clustering on the above vectors be $\varsigma_1 = 2.63, \varsigma_2 = 2.59, \varsigma_3 = 3.87, \varsigma_4 = 1.21$. The top 2 features by diverse FM method is f_2 and f_4. Let, Δ = 0.95, F1-score $\{f_2\}$ by learner $\mathcal{A}$ (say SVM-RBF) be 0.81 on validation dataset $\mathcal{V}$, $\{f_1, f_2\} \rightarrow 0.78, \{f_1, f_2, f_3\} \rightarrow 0.71$, drop f_3; $\{f_2, f_4\} \rightarrow 0.85$, combine f_4, $\{f_1, f_2, f_4\} \rightarrow 0.89, \{f_1, f_4\} \rightarrow 0.91$, the top 2 features by the proposed Hybrid Method (HM) is: f_1, f_4. We can observe that HM method eliminates the need of exhaustive searching (a property of wrapper method) for finding the optimal feature set combination.

5. RESULTS

There are three novel components in this work: 1. Effective noise filtering of PCG signals, 2. Optimal hybrid feature selection, 3. Classification of pathological normal and abnormal PCG signals. For demonstrating the efficacy of our scheme, we have experimented with publicly available large annotated PCG database from [7, 11]. In Table 1, we depict the data distribution for training and testing partitioning as well as the distribution of normal/ abnormal PCG signals, noisy/ clean PCG signals.

Table 1. Data description: Training/ Testing, Noisy/ Clean dataset distribution

Training		**Testing**	
Normal	716	Normal	1772
Abnormal	312	Abnormal	353
Noisy	92	Noisy	187
Clean	936	Clean	1938

We achieve 84.24% detection accuracy for detecting noisy PCG segments by our DTW-based method.

In Figure 3, we observe the performance score of RONUN when noise filtering is performed and overall, it outperforms in detecting the abnormality condition, i.e. high sensitivity value.

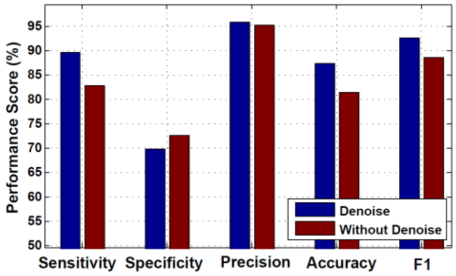

Fig. 3. Performance measures of RONUN with and without noise filtering component.

Next, we depict the efficacy of our proposed as a unified hybrid feature selection method HM. One of the crucial issues among the feature selection methods is ensuring the stability property among the component feature selection criteria $\mathbb{S} = \left\{\begin{matrix} mRMR, CMIM, ICAP, JMI, Betagamma, Condred, DISR, \\ MIFS, Relief, CIFE, MIM \end{matrix}\right\}$ over selected top feature $\mathcal{K}$. The stability property needs to be quantified to understand the similarity between the feature set selected among the different criteria. Let, there be h number of different selection criteria: $\mathbb{S} = \{\mathfrak{s}_1, \mathfrak{s}_2, \dots, \mathfrak{s}_h\}$, z number of total features and top $\mathcal{K}$ features be chosen and $\mathcal{K} < z$. We consider Kuncheva's consistency index $(\varpi \in (0,1))$ which is based on hypergeometric distribution. When, $\varpi > 0.5$, feature selection stability is high [12]. Let, K_a, K_b be the top $\mathcal{K}$ features selected by feature selection criteria $\mathfrak{s}_a, \mathfrak{s}_b$ respectively; $K_a, K_b \subset \mathbb{F}$, $|K_a| = |K_b| = \mathcal{K}$, and $|K_a \cap K_b| = d$, $\varpi(K_a, K_b) = \frac{dz - \mathcal{K}^2}{\mathcal{K}(z - \mathcal{K})}$, where $z = |\mathbb{F}|$, total number of features. Here, $\mathfrak{s}_a = HM$, $\mathfrak{s}_b \subset \mathbb{S}$. Figure 4 shows the stability in terms of ϖ of our proposed HM method with $\mathbb{S}$, over common dataset (Table 1). We observe that that $\varpi(HM, \mathbb{S})$ is high for majority of the selection criteria (like mRMR, JMI, CMIM and others for different top feature set selected as $\mathcal{K} = 3, 8$. High stability in HM indicates that the top $\mathcal{K}$ feature set selected by HM is in fact supported by the feature selection distribution pattern of the component feature selection set $\mathbb{S}$. We observe that consistency index of component feature selection functions with our proposed HM is stable at different values of top selected features $\mathcal{K}$ ($\mathcal{K} = 3, 8$) and confirms the invariance to the number of selected top feature setting ($\mathcal{K}$). In fact, for different feature sets and different $\mathcal{K}$, HM ensures stable selection.

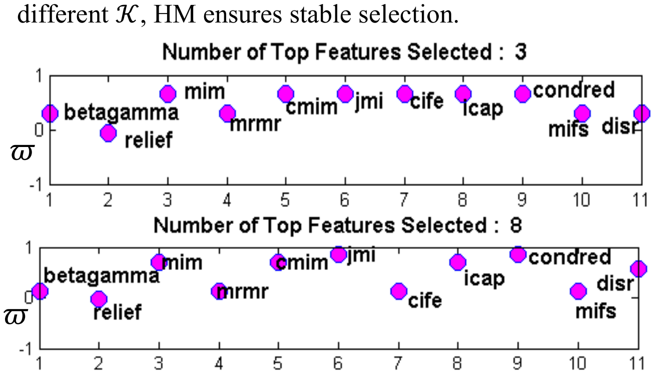

Fig. 4. Stability of our proposed hybrid feature selection method (HM) from Kuncheva's consistency index, ϖ.

The cardiac abnormality classification performance, is depicted in figure 5. This experimental results is a major contribution. We consider Radial Basis Function kernel Support Vector Machine with (SVM-RBF) as our classifier method. We show the performance of our RONUN scheme with our proposed noise filtering (in Figure 3) and hybrid feature selection optimization algorithms at different settings of number of top features $\mathcal{K}$. We observe that closest-performing method is FM, which provides balanced sensitivity and specificity, however underperforms in both the performance measures. We can safely claim that RONUN outperforms the state-of-the-art feature selection algorithms. RONUN consistently performs well with balanced sensitivity and specificity scores. The most salient aspect is the high sensitivity of the RONUN scheme, which means that RONUN detects cardiac abnormal condition with high probability. Indeed, minimization of undetected abnormal condition is the most vital condition in clinical analytics: high false negatives would be fatal to the patients.

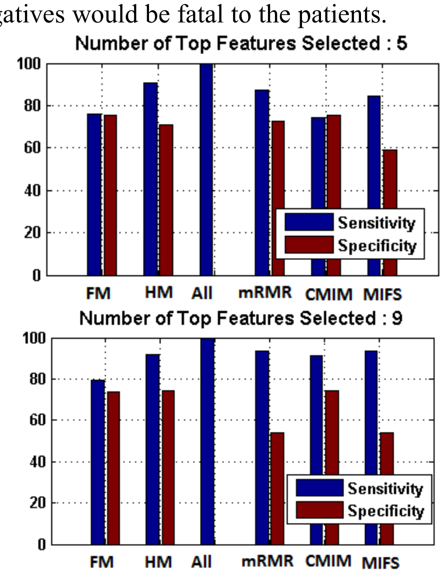

Fig. 5. Performance of RONUN in varying number of top features for cardiac abnormality detection (5 and 9 respectively): FM: Filter method; HM: Our proposed hybrid method; mRMR, CMIM, MIFS: state-of-the-art methods.

6. CONCLUSION

In this paper, we have established that automated computational analysis on PCG signals can potentially indicate the presence of cardiac anomaly condition. We have shown that effective noise filtering and optimal hybrid feature selection would significantly improve the cardiac anomaly condition detection capability. We have demonstrated that hybrid feature space optimization, which is a classification learner performance optimization based wrapper method over diverse filter-based feature selection techniques outperform state-of-the-art approaches. We have validated that the proposed scheme ensures high cardiac anomaly detection efficacy through our experiment with publicly available expert-annotated MIT-Physionet database.

8. REFERENCES

[1] H. Peng, F. Long, and C. Ding, “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, 2005.

[2] F. Fleuret, “Fast binary feature selection with conditional mutual information,” Journal of Machine Learning Research, pp. 1531–1555, 2004.

[3] A. Jakulin, “Machine learning based on attribute interactions,” Ph.D. thesis, Univerza v Ljubljani, 2005.

[4] D. Gradolewski and G. Redlarski, “Waveletbased denoising method for real phonocardiography signal recorded by mobile devices in noisy environment,” Computers in biology and medicine, vol. 52, pp. 119–129, 2014.

[5] R. Saracoglu, "Hidden markov model-based classification of heart valve dis-ease with pca for dimension reduction," Elsevier Engineering Applications of Applied Artificial Intelligence, pp. 1523-1528, 2012.

[6] I. Maglogiannis, E. Loukis, E. Zafiropoulos, and A. Stasis, “Support vectors machine based identification of heart valve diseases using heart sounds,” Computer methods and programs in biomedicine, vol. 95, no. 1, pp. 47–61, 2009.

[7] “Physionet challenge 2016,” https://physionet. org/pn3/challenge/2016/, 2016.

[8] G. Chen and J. Chen, “A novel wrapper method for feature selection and its applications,” Neurocomputing. pp. 219-226, 2015.

[9] S. E Schmidt, C. Holst-Hansen, J. Hansen, E. Toft, and J. J Struijk, “Acoustic features for the identification of coronary artery disease,” IEEE Transactions on Biomedical Engineering, vol. 62, no. 11, pp. 2611–2619, 2015.

[10] G. Brown, A. Pocock, M. Zhao, and M. Luján, "Conditional likelihood maxi-misation: a unifying framework for information theoretic feature selection," Journal of machine Learning Research, pp. 27 - 66, 2012.

[11] A.L. Goldberger et al. PhysioBank, PhysioToolkit, and PhysioNet: Compo-nents of a New Research Resource for Complex Physiologic Signals," Circulation, 2010.

[12] L. I. Kuncheva, “A stability index for feature selection,” International Multi-Conference: Artificial Intelligence and Applications, pp. 390–395, 2007.

[13] S. Solorio-Fernández, J. Ariel Carrasco-Ochoa, and J. Martínez-Trinidad, "A new hybrid filter-wrapper feature selection method for clustering based on ranking," Neurocomputing, pp. 866-880, 2016.

[14] M. Ester, Ha. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise," SIGKDD, pp. 226-231, 1996.

[15] G. Redlarski, D. Gradolewski, A. Palkowski, "A System for Heart Sounds Classification," PLoS ONE vol. 9, issue. 11, 2014.

[16] A. S. Kyzdarbekova, D. M. Dutbaeva, K. B. Kasymbekova and U. S. Kyzdarbek, "Adaptive noise reduction phonocardiograms based on wavelet transformation," International Conference "Quality Management,Transport and Information Security, Information Technologies" pp. 391-394, 2017.