TATA CONSULTANCY SERVICES
Experience certainty.

TATA

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

ICASSP 2018

Arijit Ukil,
Soma Bandyopadhyay,
Rituraj Singh,
Arpan Pal

Chetanya Puri
Researcher
Research and Innovation
TATA Consultancy Services, India

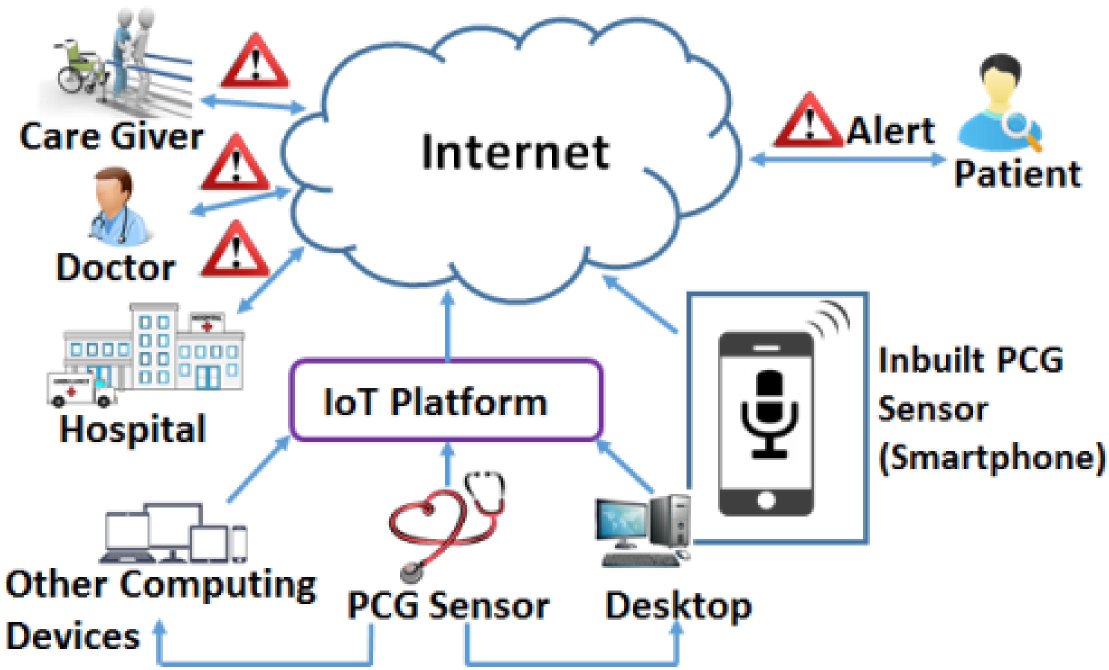
20-APR-2018

Affordable Healthcare Analytics in Internet of Things (IoT)

- **Cardiac problem is a serious issue** : An estimated 17.7 million people died from CVDs in 2015, representing 31% of all global deaths.

- "Over three quarters of CVD deaths take place in low- and middle-income countries"

- IoT revolutionizing healthcare
  - Low cost wearable sensors
  - Early screening and detection



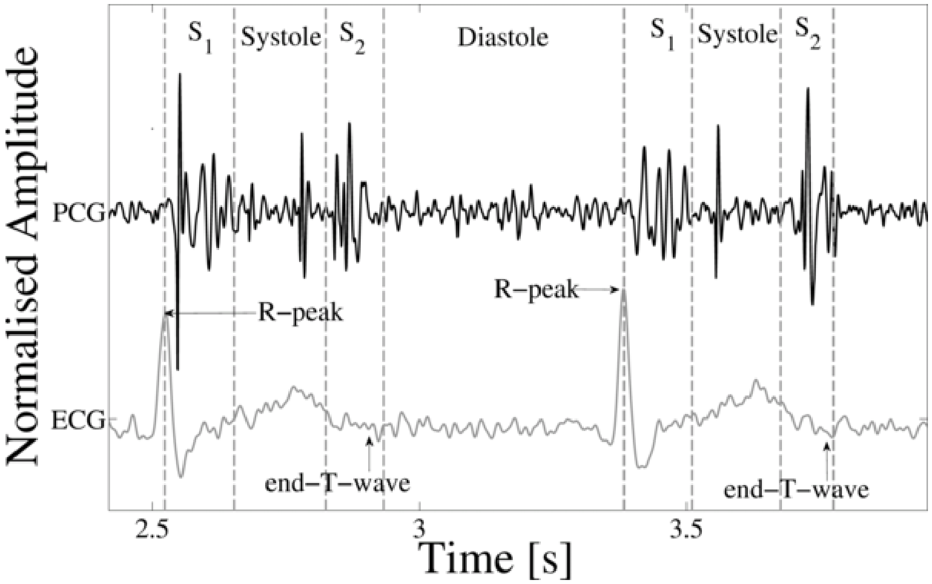
*http://www.who.int/mediacentre/factsheets/fs317/en/

Which signal?

- Phonocardiogram/Heart Sound/PCG signals
  - Transient sounds produced when heart valve opens and closes.
  - Healthy adults have two normal heart sounds S1 (‘*lub*’) - S2 (‘*dub*’) in a beat.
  - Pathological condition/Disease induces murmurs and other sounds.

- PCG signals <u>prone to noise.</u>

- <u>**Challenges**</u>
  - Accurate diagnosable quality data collection
  - Reliable Disease Prediction



https://physionet.org/challenge/2016/

What we propose?

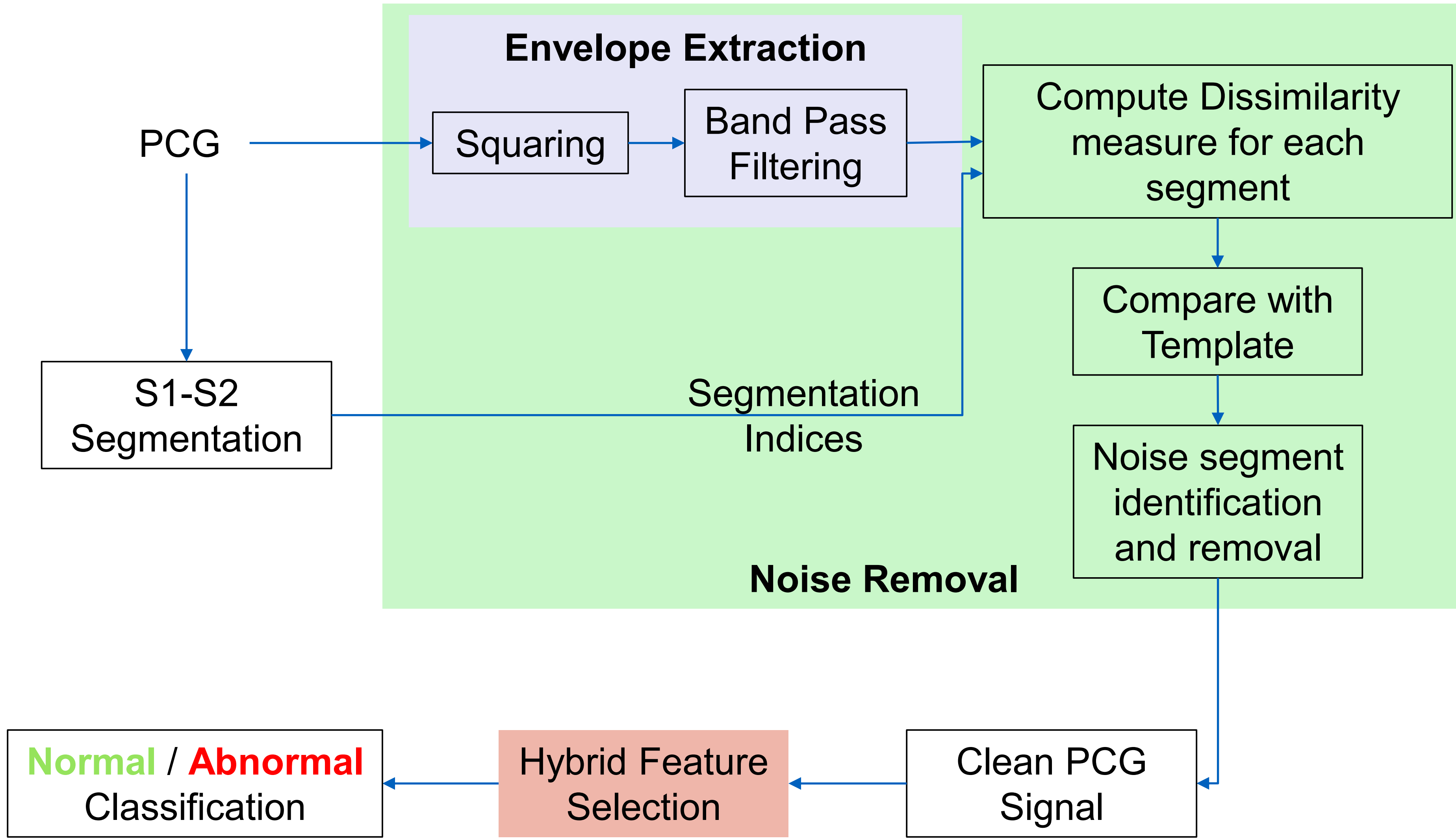
RONUN

- **Denoising**
  - Noisy PCG signal identification and rejection
- **Optimal Features**
  - Optimal feature selection through Hybrid approach
- **Classification**
  - Classification of clean PCG signal to determine the presence of cardiac abnormality.

- RONUN is an automated, preventive cardiac management solution that uses smartphone or other wearable sensor-captured heart sound, PCG signal.

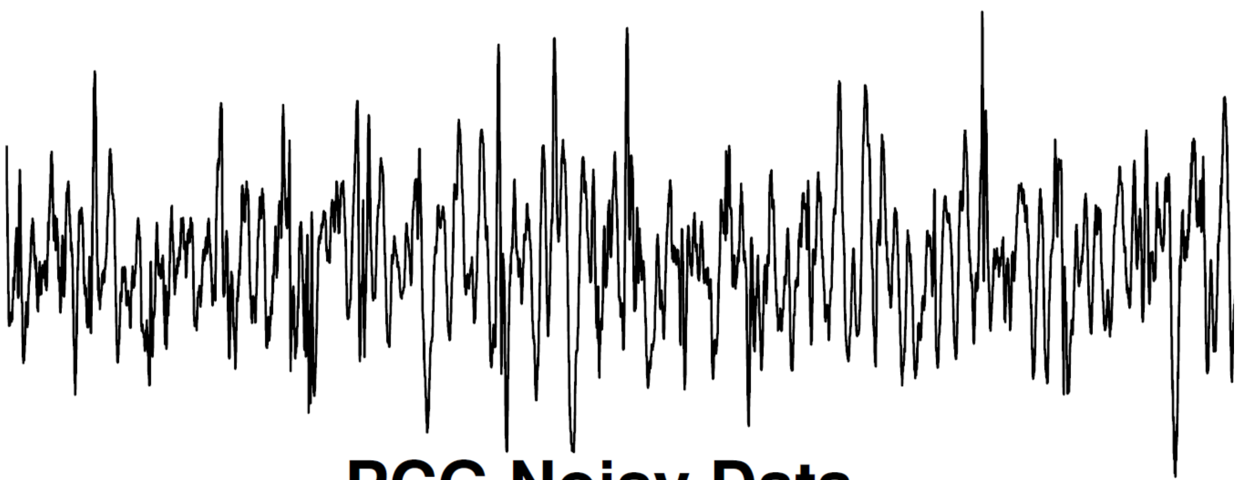
- Generates alerts by finding the pathological condition of the user.

Functional Flow



Glimpse of PCG Clean and Noisy Data

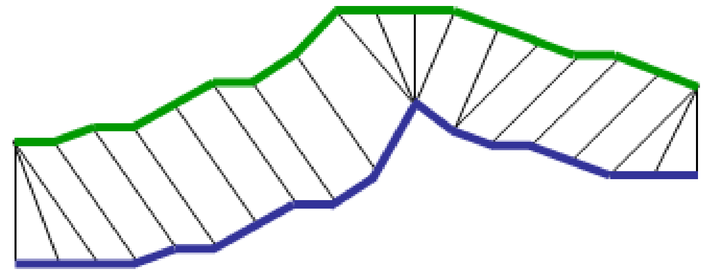
PCG Clean Data

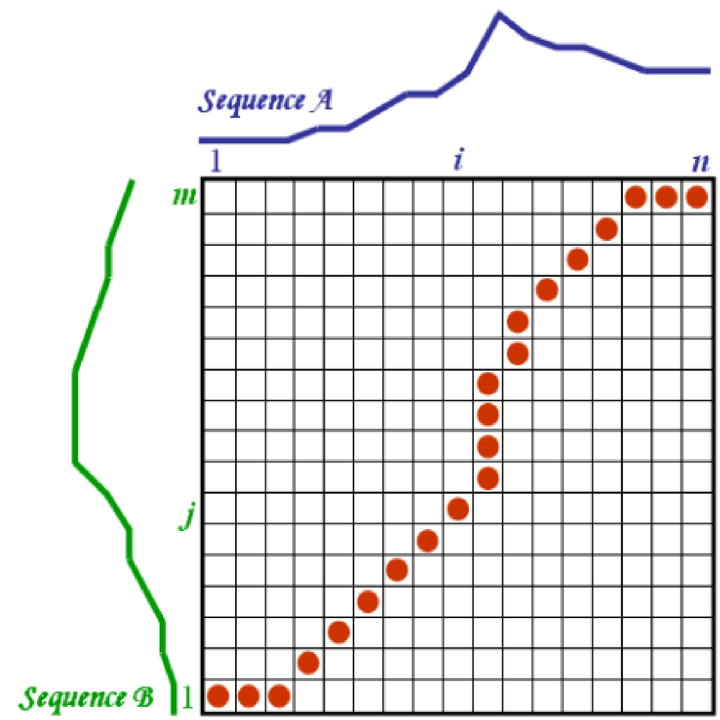


PCG Noisy Data

Semi-supervised Learning Based Noise Identification

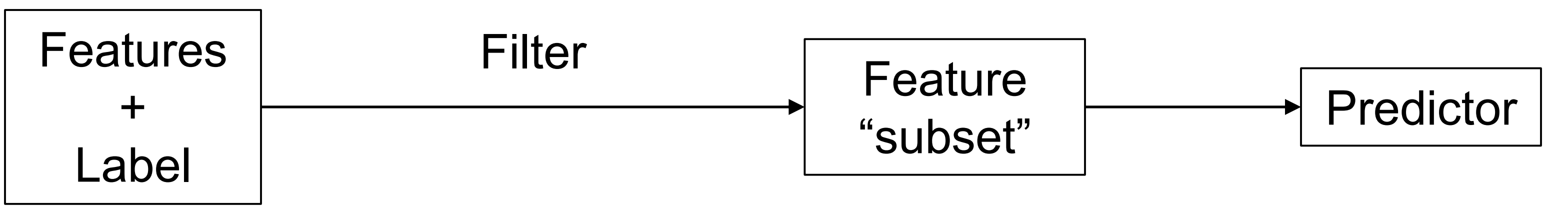
- Use a representative template
- Find most probable segment length L_p
  - Through density based clustering.
- Segmentize the signal based on heart beat
- Normalize each segment length to L_p
- Compute Dynamic Time warping (DTW) distance of each PCG segment
- DTW distances of corrupt portions are more → higher dissimilarity
- Accuracy → 84.24%



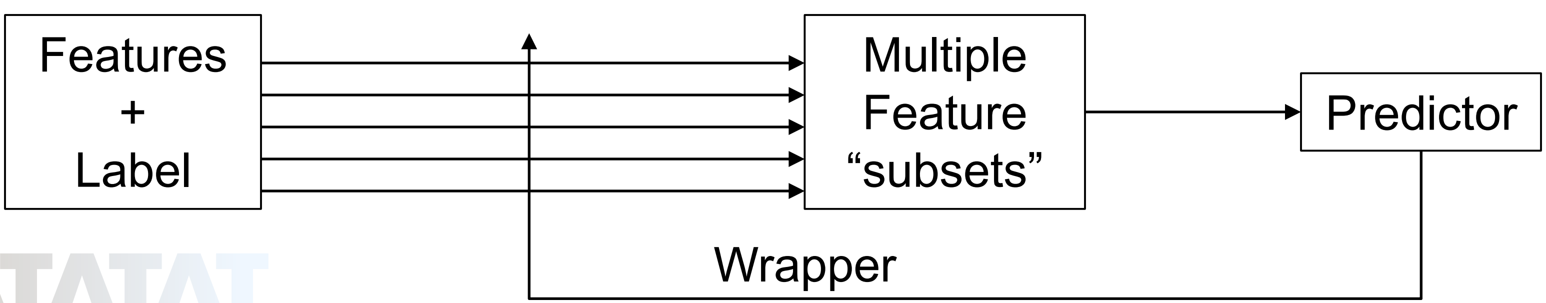


Hybrid Feature Space Optimization

- **Feature Selection and reduction** [2]
  - Improve Prediction performance
  - Provide fast and cost-efficient models
  - Better understanding of the underlying processes that generated data
- **Methods**
  - Filter



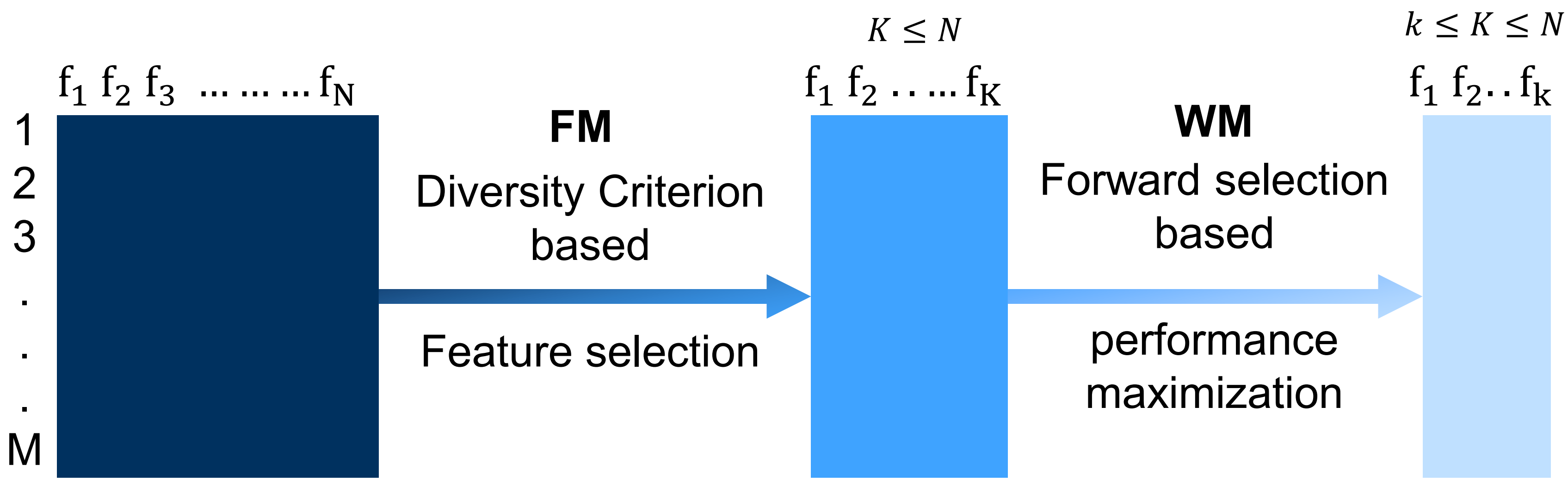
  - Wrapper



Hybrid Feature Space Optimization

- **Hybrid Feature Selection**

Given a feature set of M instances $F = \{f_1, f_2, f_3, \ldots \ldots f_N\}$ consisting N features



FM : Diversifying Features

- J different feature selection methods.
  - For example, mRMR: $R = \frac{1}{|F|^2} \sum_{f_i, f_j} I\left(f_i, f_j\right)$; $Relevance, D(F, c) = \frac{1}{|F|} \sum_{f_i} I\left(f_i, c\right)$

$$max_F \ \phi(D, R), \phi = D - R$$

- Obtain J different ranks for each feature in the feature set.
- 2-means clustering to get centroids C_1^i, C_2^i. $(C_1 < C_2)$ for i^{th} feature .
- Features are ordered in increasing order of their C_1^i values $\forall\ i = 1,2,3, \dots N$.
  - Best feature is $f_{best}\ if\ C_1^{best} < C_1^i\ \forall\ i = 1,2,3,4, \dots N$

FM : Diversifying Features

- Example:
  - 4 features $\{f_1, f_2, f_3, f_4\}$
  - K = 2
- J different feature selection methods.
  - 7 diverse selection criteria {ICAP, JMI, DISR, mRMR, MIFS, CIFE, CMIM}
- Obtain J different ranks for each feature in the feature set.

$$
\begin{aligned}
f_1 &\rightarrow [3,1,4,2,2,1,3] \\
f_2 &\rightarrow [2,3,1,4,3,4,2] \\
f_3 &\rightarrow [4,4,3,3,4,3,4] \\
f_4 &\rightarrow [1,2,2,1,1,2,1]
\end{aligned}
$$

- 2-means clustering to get centroids C_1^i, C_2^i. $(C_1 < C_2)$ for i^{th} feature .
  - $C_1^1 = 2.63,\ C_1^2 = 2.59,\ C_1^3 = 3.87, C_1^4 = 1.21$
- Features are ordered in increasing order of their C_1^i values $\forall\ i = 1,2,3, \dots N$.
  - Best feature is $f_{best}\ if\ C_1^{best} < C_1^i\ \forall\ i = 1,2,3,4, \dots N$
  - $\{f_4, f_1\}$

WM : Forward Selection Performance Maximization

- Feature ranked from filter methods are given as input.
- Choose a performance metric. For example, accuracy, F1-score, $\sqrt{Sensitivity \times Specificity}$
- Select the feature if validation performance increases else drop the feature.

Typical Features used for PCG classification

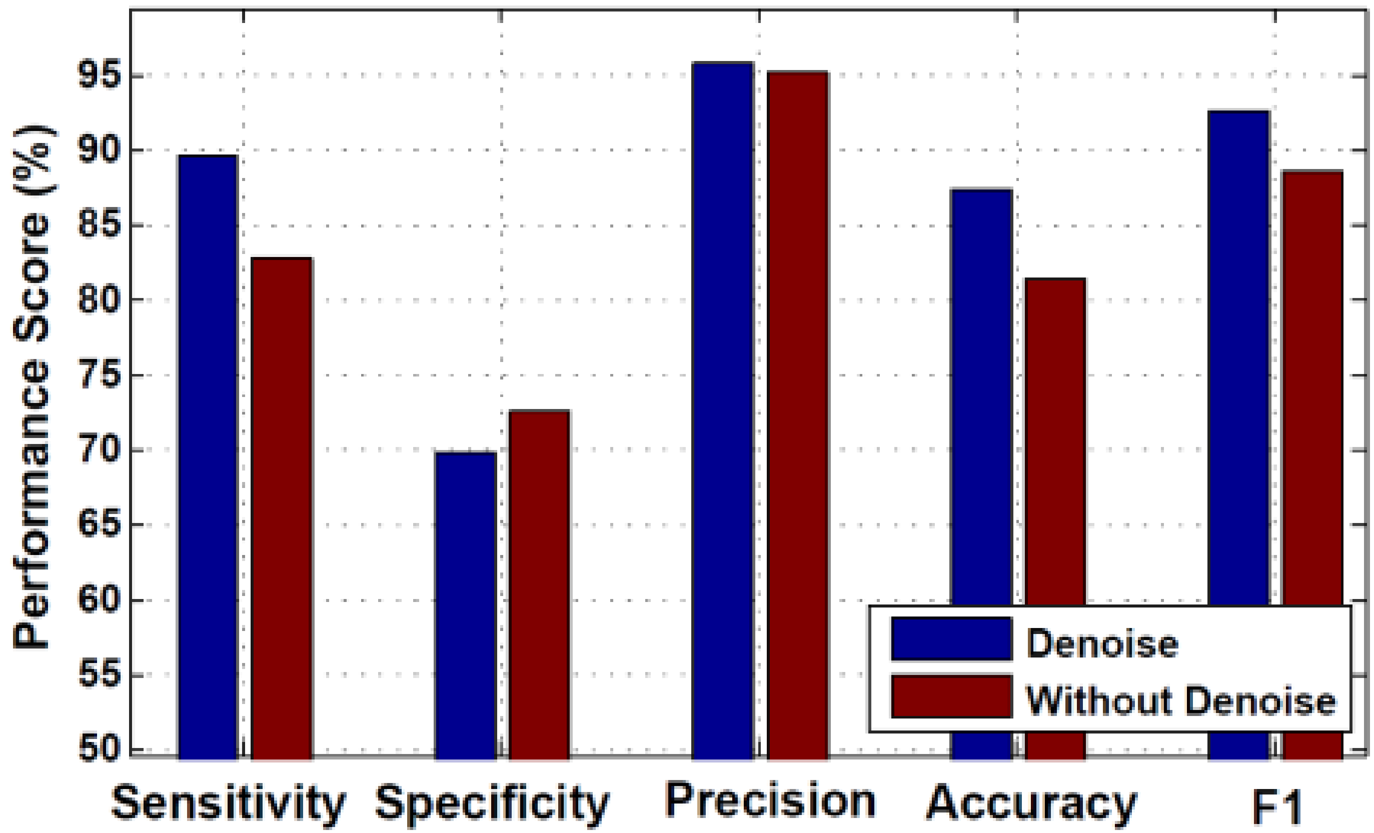
- Time domain
  - Mean, std of the RR intervals, Systole intervals, Diastole intervals
  - Mean, std ratios of interval between systole and RR in each beat
  - Mean absolute amplitude ratios between diastole period and S2 period in each heart beat.
- Frequency Domain
  - Skewness, Kurtosis of the FFT.
  - Total Normalized power in different frequency bands.
- Wavelet domain
  - 'db3' mother wavelet. 3 levels of decomposition
  - Mean and standard deviation over S1, S2, Systole diastole intervals.

Dataset Description [5]

- Digital Stethoscope
- 8 seconds to 5 minutes down-sampled at 2000 Hz
- 3153 recordings
  - 2488 Normal
  - 665 Abnormal
- Variety
  - 9 heart sound databases collected independently.
  - 7 different research teams from 7 countries and 3 continents

Performance with and without Noise Removal

- Increased abnormality detection



Stability of the features selected

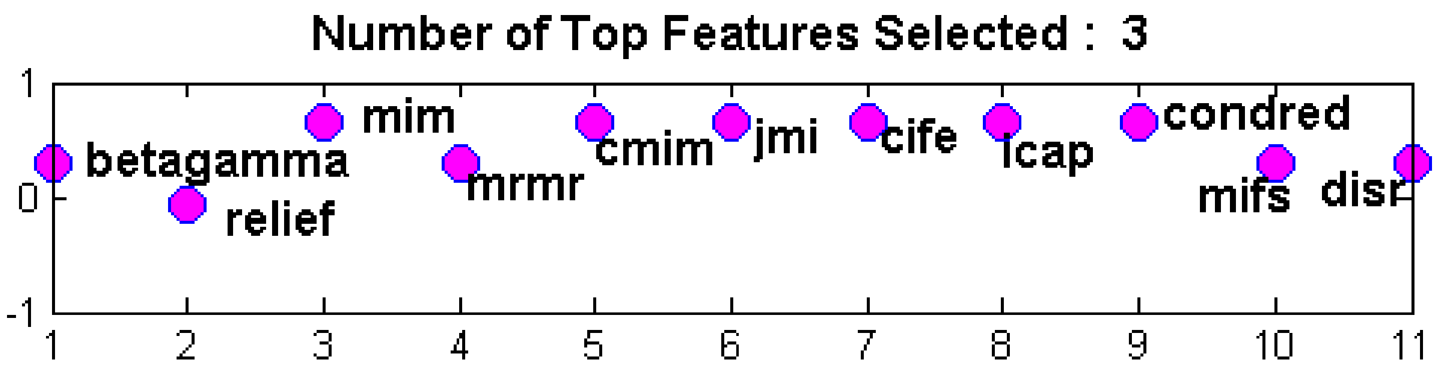
- Kuncheva's consistency index
  - The Consistency Index for two subsets $A \subset X$ and $B \subset X$, such that $|A| = |B| = k$, where $0 < k < |X| = n$, is

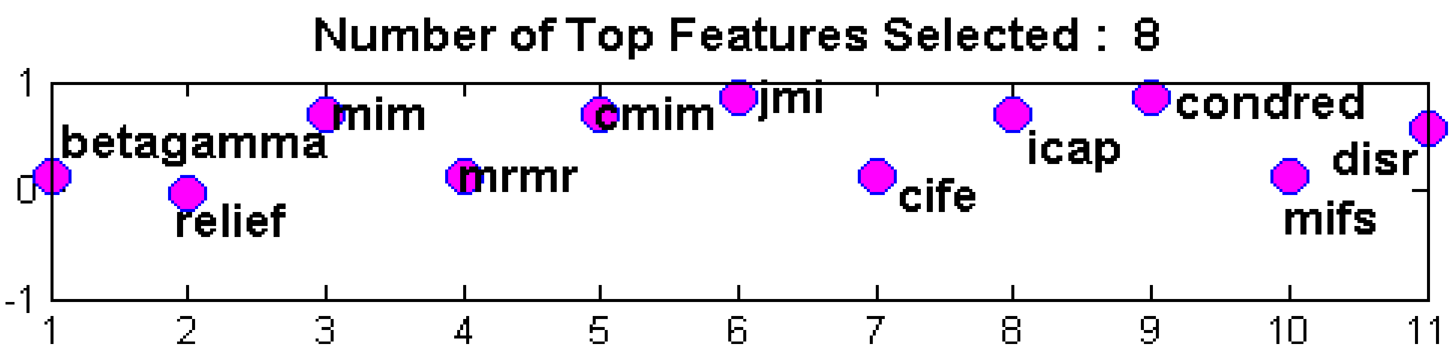
$$\omega(A,B) = \frac{\left(r - \frac{k^2}{n}\right)}{\left(k - \frac{k^2}{n}\right)} = \frac{rn - k^2}{k(n - k)}$$

k: Cardianality of A and B, r : $|A \cap B|$, n : Original number of features

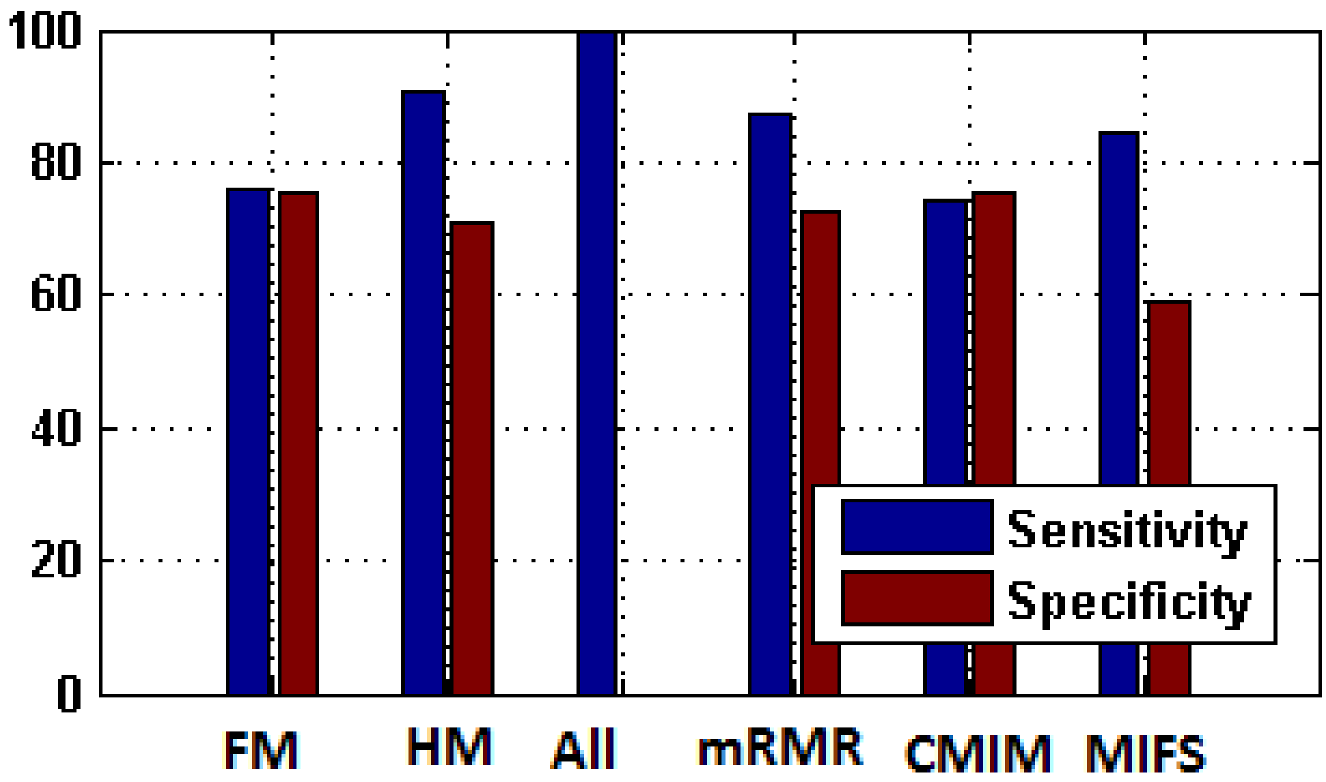
Stability of the features selected

- $\omega \geq 0.5$ ensures stability.





Ronun Abnormality Detection Performance



Concluding remarks

- Automated cardiovascular disease detection based on PCG signals.
- Appropriately pre-processed PCG signal would return higher clinical utility
- Hybrid feature space optimization

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) 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.
3) G. Chen and J. Chen, “A novel wrapper method for feature selection and its applications,” Neurocomputing. pp. 219-226, 2015.
4) 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.
5) A.L. Goldberger et al. PhysioBank, PhysioToolkit, and PhysioNet: Compo-nents of a New Research Resource for Complex Physiologic Signals," Circulation, 2010.
6) L. I. Kuncheva, “A stability index for feature selection,” International Multi-Conference: Artificial Intelligence and Applications, pp. 390–395, 2007.
7) 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.

TATA CONSULTANCY SERVICES
Experience certainty.

TATA

Thank You

IT Services
Business Solutions
Consulting