



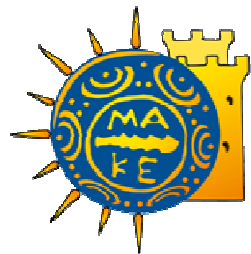
# Ανίχνευση Παροξυσμικής Κολπικής Μαρμαρυγής μέσω Μηχανικής Μάθησης απο ΗΚΓ ηρεμίας

Διπλωματική Εργασία  
Ευαγγελία Μυρόβαλη

Επιβλέπων Καθηγητής : Δημήτριος Χρήστου-Βαρσακέλης

Πανεπιστήμιο Μακεδονίας, Σχολή Επιστημών Πληροφορίας,  
Τμήμα Εφαρμοσμένης Πληροφορικής

Φεβρουάριος 2021



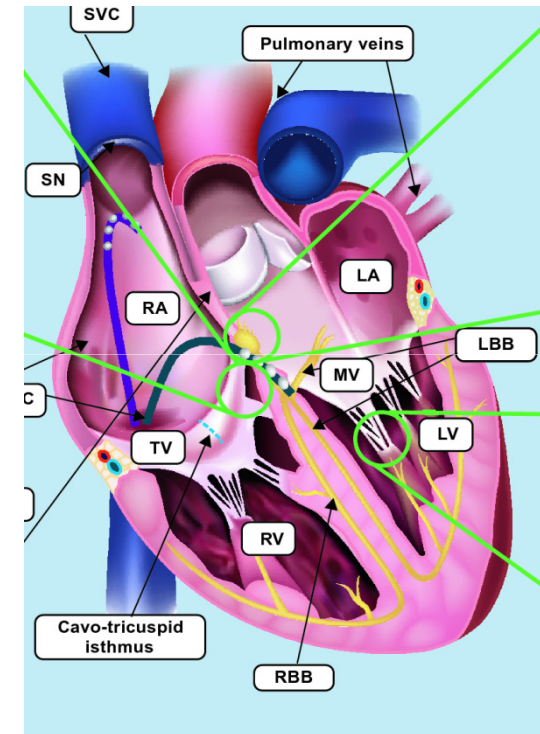
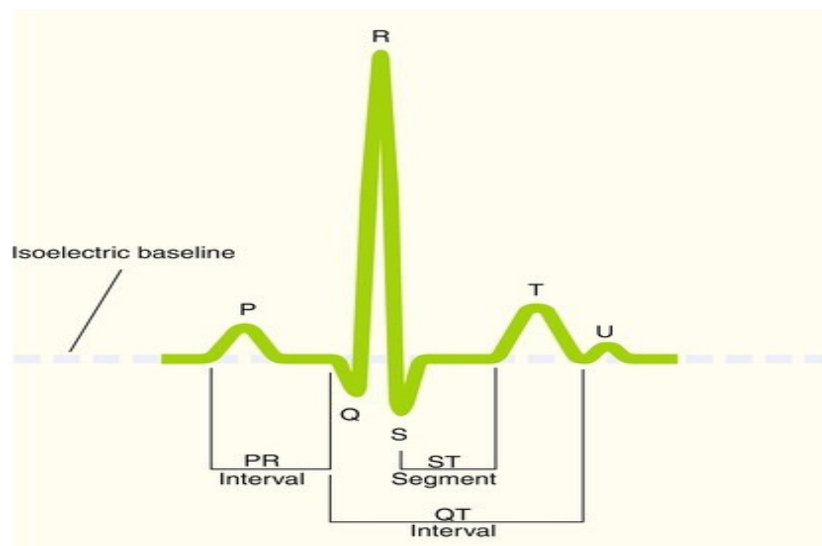
## Σκοπός

- Ανίχνευση Παροξυσμικής κολπικής μαρμαρυγής (ΠΚΜ) απο ΗΚΓ ηρεμίας
- Εξαγωγή και ανάλυση P-κυμάτων σε δυο ομάδες ασθενών Υγιείς και ΠΚΜ
- Υπολογισμός μετρικών beat-to-beat πριν και μετα την εφαρμογή χρονικής κλίμακας
- Στατιστική ανάλυση των χαρακτηριστικών των ομάδων
- Εφαρμογή μεθόδων επιλογής χαρακτηριστικών
- Εφαρμογή ενός ταξινομητή Random Forests



# Φυσιολογία της καρδιάς

- Η καρδιά αποτελείται από αριστερούς και δεξιούς κόλπους και κοιλίες
- Το ηλεκτρικό ερέθισμα ξεκινά από τον φλεβοκόμβο SN, διαδίδεται στους κόλπους, AV node, Purkinje fibers και στις κοιλίες
- Το P-κύμα συμβολίζει την κολπική δραστηριότητα ενώ το QRST την κοιλιακή και το U την επαναπόλωση





# Κολπική Μαρμαρυγή

- Ακανόνιστος καρδιακός ρυθμός
- Απουσία P-κυμάτων
- Ακανόνιστη κολπική δραστηριότητα



- Παροξυσμική κολπική μαρμαρυγή (ΚΜ)
- Εμμένουσα ΚΜ
- Μόνιμη ΚΜ



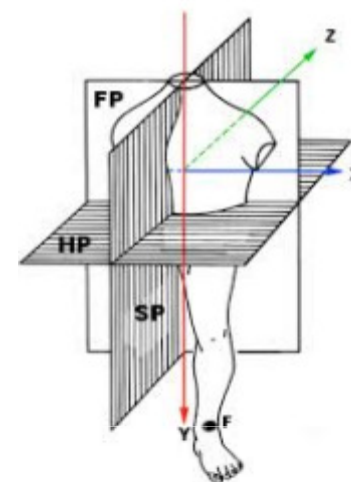
## Κολπική Μαρμαρυγή

- Ανίχνευση σε πραγματικό χρόνο μέσω Smartwatches και photoplethysmography (PPG) SE= 67 % - 99.6%
  - ❖ Apple watch
  - ❖ FibriCheck
  - ❖ Alive Core
- Ανίχνευση του συμβάντος και διαχωρισμός απο τον sinus rhythm ACC= 89.2% - 98.1%
- Ανίχνευση κατά την διάρκεια SR χωρίς εμφανείς αλλαγές στο ΗΚΓ



## Συλλογή Δεδομένων

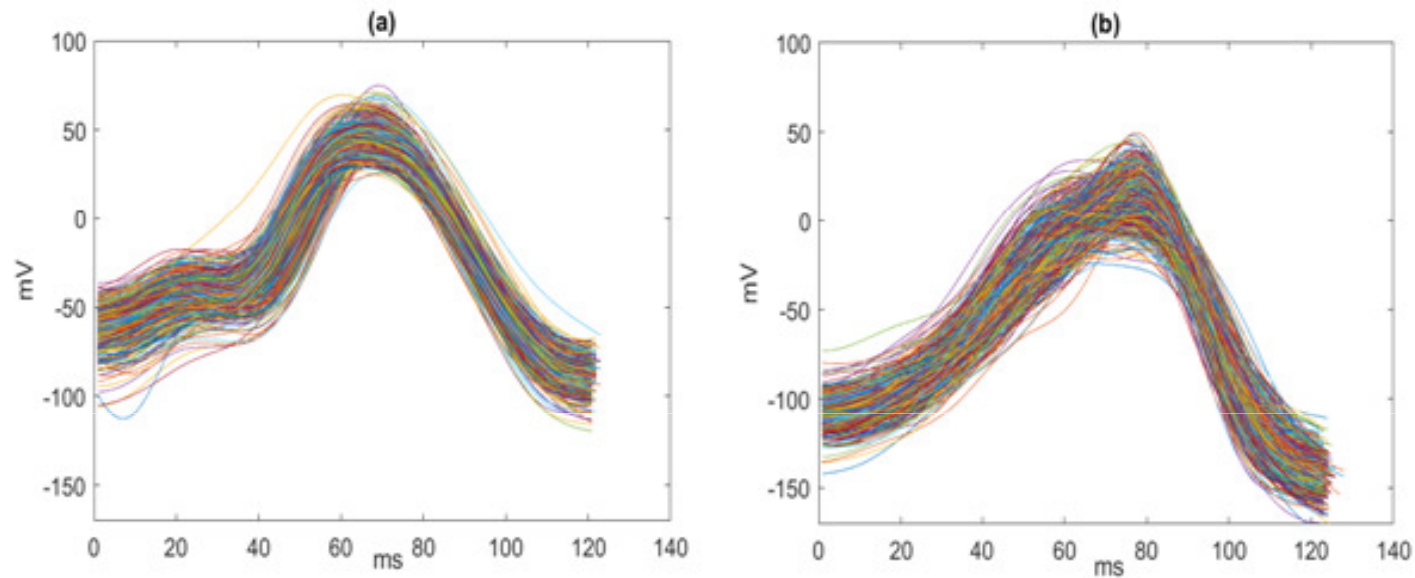
- ECG Data of 69 ΠΚΜ ασθενών και 59 Υγιών
- Orthogonal ECG καταγραφές,  $F_s = 1000$  Hz
- ECG διάρκειας 10 λεπτών σε ηρεμία



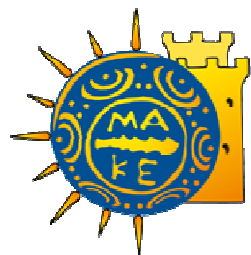
	PAF n=69	Healthy n=59	P-value
Age	$57.89 \pm 10.03$	$55.55 \pm 5.88$	<b>0.0176</b>
HR	$61.6 \pm 11.7$	$69.7 \pm 11.3$	<b>9.2196e-05</b>
P-waves	$510 \pm 102$	$540 \pm 103$	-
Male	52 (75.36%)	38 (64.4%)	0.1763
Female	17 (24.64%)	21 (35.6%)	0.1763



## Εξαγωγή Ρ-κύματος



- Inverse Euclidean Distance  $d(t, t_{ref}) = ||X(t) - X(t_{ref})||$
- tonset & toffset ενός Ρ-κύματος ως οδηγό



# Εξαγωγή Ρ-κύματος

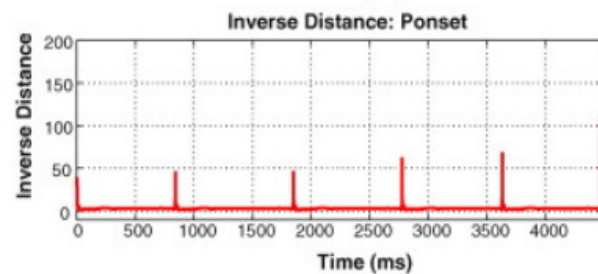
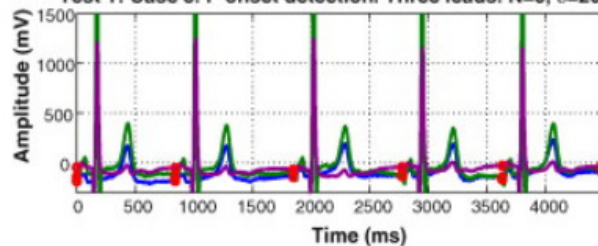
$$X(t) = \begin{bmatrix} s(t + b_1 \delta) - s(t + b_2 \delta) \\ \vdots \\ s(t + b_N \delta) - s(t + b_{N+1} \delta) \end{bmatrix}$$

$$b_{on} = [0, \dots, N]^T$$

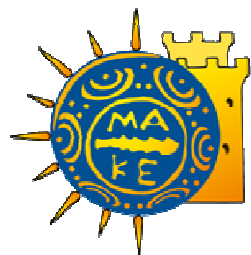
$$b_{of} = [-N, \dots, 0]^T$$

$$N=6, \delta=30$$

Test 1: Case 3. P onset detection. Three leads. N=8,  $\delta=20$ .





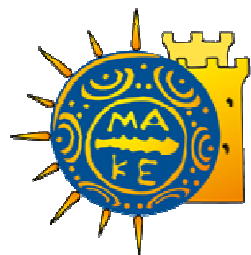


## Μέθοδοι

- Time domain features
- Euclidean Distance
- P-wave amplitude
- Area features
- Slope features
- Time scaling for HR adjustment at 60 bpm (1000 ms)

Feature	Description
$P_d$	P-wave duration
$P_{on}R$	time from P onset to R peak
$P_{of}R$	time from P offset to R peak
$P_a$	P-wave amplitude
$D$	Euclidean distance between consecutive P-waves
$A$	P-wave area
$A_0$	P-wave area after offsetting P-wave to zero value at onset.
$S_{max}$	Max Slope
$S_{5\%,10\%,20\%}$	Slope at 5%,10%,20% of $P_d$

$$a_{coef}(i) = \frac{1000}{RR_i}$$



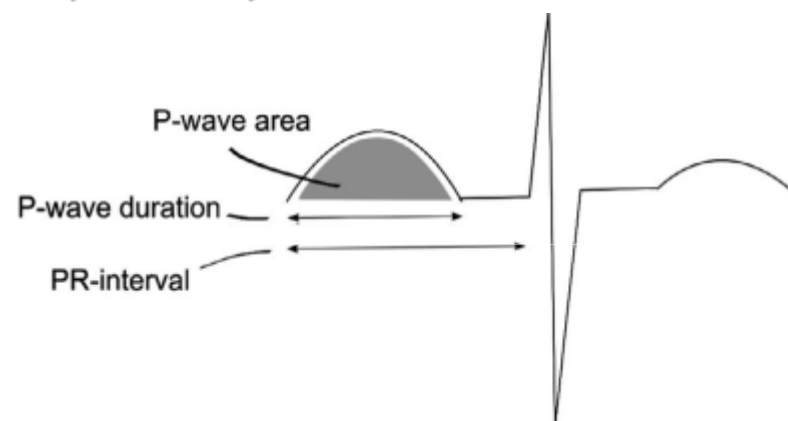
# Μετρικές Σήματος

$$P_d = t_{of} - t_{on}$$

$$P_{on}R = t_R - t_{on}$$

$$P_{of}R = t_R - t_{of}$$

$$S_c = \frac{P(t_c) - P(t_{on})}{t_c - t_{on}}$$



$$A = \int_{\tau_{on}}^{\tau_{of}} P(\tau) d\tau$$

$$A_0 = \int_{\tau_{on}}^{\tau_{of}} (P(\tau) - P(\tau_{on})) d\tau = A - P(\tau_{on})P_d$$

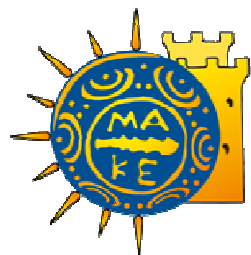
$$\Delta A = A_0 - A$$

$$D(i) = \frac{\sqrt{\sum_{t=0}^{l-1} (P_{i+1}(t_{i+1}^{on} + t) - P_i(t_i^{on} + t))^2}}{\sqrt{\sum_{t=0}^{l-1} (P_{i+1}(t_{i+1}^{on} + t))^2}}$$



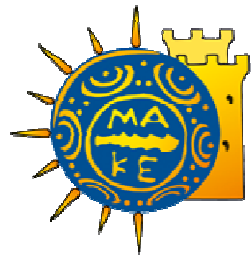
# Lead X ΠΚΜ vs. Υγιείς

<i>Features</i>	$\mu$	$\sigma$	<i>cv</i>	<i>max</i>	<i>min</i>
$P_d$	*	1.4853e-05	1.0088e-07	<b>4.9257e-04</b>	7.5171e-07
$P_{on}R$	*	<b>1.3464e-06</b>	<b>1.6943e-08</b>	<b>3.8277e-04</b>	9.4417e-08
$P_{of}R$	2.62e-02	<b>2.0963e-05</b>	<b>3.0620e-07</b>	<b>0.0051</b>	<b>9.5982e-07</b>
$P_a$	*	*	*	<b>2.47e-02</b>	*
$D$	<b>4.7527e-04</b>	<b>9.3217e-08</b>	<b>4.7527e-04</b>	<b>5.6415e-09</b>	*
$A$	<b>6.7020e-06</b>	*	0.0334	4.4834e-09	*
$A_0$	<b>3.4569e-08</b>	*	1.4853e-05	*	1.4929e-07
$\Delta A$	<b>9.4114e-09</b>	*	<b>1.0947e-10</b>	<b>0.0398</b>	<b>7.6897e-15</b>
$S_{max}$	<b>4.5521e-05</b>	<b>3.6910e-04</b>	7.6e-03	<b>1.48e-02</b>	<b>2.2829e-05</b>
$S_{5\%}$	*	<b>1.68e-02</b>	*	*	*
$S_{10\%}$	*	<b>0.62e-02</b>	*	*	*
$S_{20\%}$	*	<b>0.13e-02</b>	*	*	*



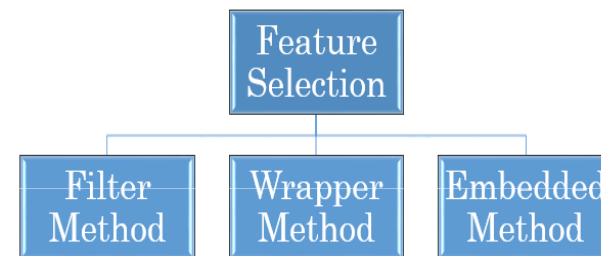
## Lead Υ ΠΚΜ vs. Υγιείς

<i>Features</i>	$\mu$	$\sigma$	<i>cv</i>	<i>max</i>	<i>min</i>
$P_d$	*	*	4.5521e-05	0.0484	3.6910e-04
$P_{on}R$	0.0334	5.9870e-06	1.4173e-07	<b>0.0042</b>	2.7046e-05
$P_{of}R$	*	0.0032	4.1155e-04	<b>0.0092</b>	2.3255e-04
$D$	*	*	<b>4.4540e-06</b>	<b>0.0075</b>	*
$A$	*	*	*	5.1047e-04	*
$A_0$	<b>0.0031</b>	*	*	*	<b>0.0012</b>
$S_{max}$	<b>0.0427</b>	<b>0.0126</b>	<b>0.0447</b>	<b>1.48e-02</b>	<b>0.0359</b>
$S_{5\%}$	*	0.0253*	*	*	*
$S_{10\%}$	*	*	*	*	*
$S_{20\%}$	*	<b>0.0484</b>	*	0.0354*	*



# Feature Selection Methods

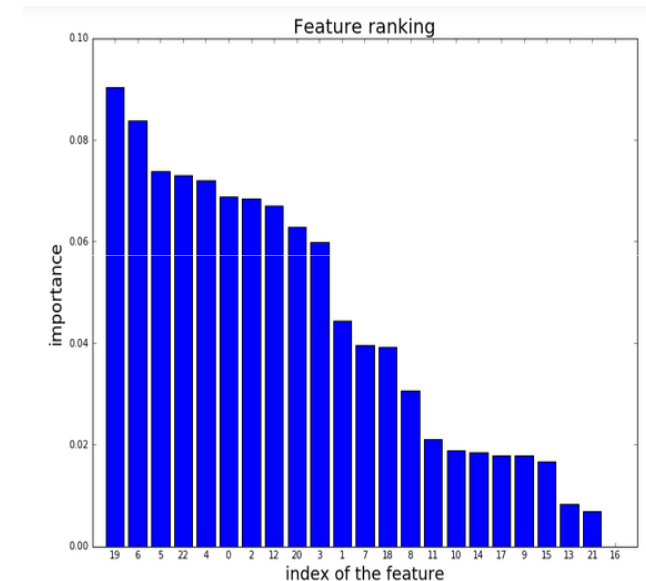
- 60 features με 36 features  
ΠΚΜ vs. Υγιείς  $p < 0.05$
- **Feature Selection ε έθοδοι:**
  - Max Volume (MV)
  - Overlap of Confidence Interval (OCI)
  - Max Determinant
  - Singular Value Decomposition (SVD)
- OCI, MV, Max Det με 14 (εκτός time domain +ΔA) features  
ΠΚΜ vs. Υγιείς  $p < 0.05$  πριν και μετα το **time scaling**





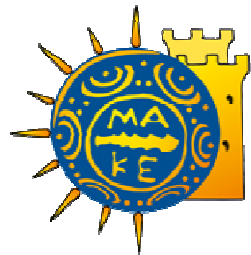
# Feature Selection Methods

- 34 features (εκτός ΔΑ features) με HR & age  
FV 34 features ΠΚΜ vs. Υγιείς  $p < 0.05$
- FV 34 features με αντικατάσταση Αο απο ΔΑ
- Random Forests Feature Importance (FI) can be used to rank the importance of variables
- Trained model and test data (OOB sample) for prediction and calculation of error value
- Randomly shuffle values of j-th feature prediction and calculation of error value OOB
- FI score for the j-th feature is computed by averaging the difference in OOB error before and after the permutation



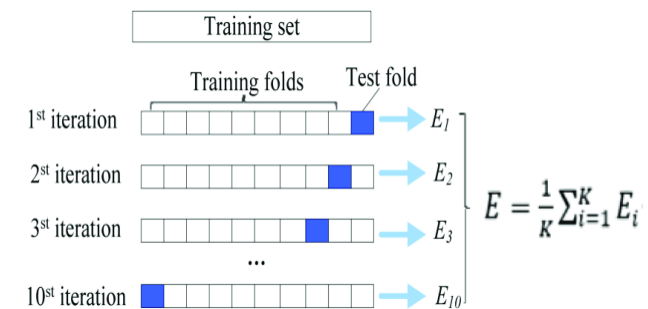
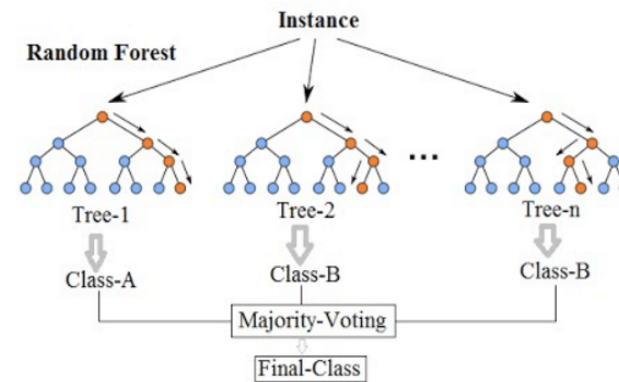


- 
- ```
graph TD; TS[Training Set] --> TS1[Training Sample 1]; TS --> TS2[Training Sample 2]; TS --> TSn[Training Sample n]; TS1 --> DT1[Decision Tree 1]; TS2 --> DT2[Decision Tree 2]; TSn --> DTn[Decision Tree n]; DT1 --> V((Voting)); DT2 --> V; DTn --> V; V --> P[Prediction]; TSet[Test Set] --> P;
```

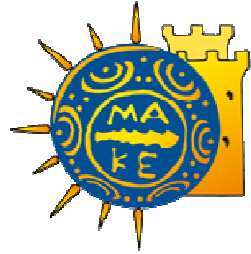


# Random Forests Classifier

- Randomly select “k” features from total “m” features where  $k \ll m$
- Among the “k” features, calculate the node “d” using the best split point
- Split the node into daughter nodes using the best split
- Repeat the 1 to 3 steps until “l” number of nodes has been reached
- Build forest by repeating steps 1 to 4 for “n” number times to create “n” number of trees
- 10 repetitions of 10-fold cross validation
- Classification metrics ACC, SE, SP, AUC







## Classification results Lead X

| Feature Sets                                                                                  | ACC                                | SE                                 | SP                                 |
|-----------------------------------------------------------------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| $[\mu A, \mu A_0, cvS_{max}, cvD, stdS_{20\%}, stdS_{max}, maxA, minS_{max}, cvP_d]$          | <b><math>93.70 \pm 1.10</math></b> | <b><math>95.26 \pm 2.00</math></b> | <b><math>91.86 \pm 3.82</math></b> |
| $[\mu A_0, stdS_{10\%}, minA_0, maxA, maxS_{max}, \mu A, maxD, cvA, cvA_0]$                   | $91.06 \pm 1.58$                   | $89.90 \pm 3.02$                   | $92.37 \pm 3.03$                   |
| $[\mu A_0, \mu A, stdS_{max}, cvA, stdS_{20\%}, stdS_{5\%}, stdS_{10\%}, maxS_{max}, maxA_0]$ | $92.4 \pm 1.16$                    | $93.10 \pm 2.52$                   | $91.57 \pm 2.94$                   |
| $[\mu A_0, cvS_{max}, cvD, stdS_{20\%}, stdS_{max}, cvP_d]$                                   | <b><math>93.07 \pm 2.00</math></b> | <b><math>94.81 \pm 4.66</math></b> | <b><math>91.03 \pm 3.48</math></b> |
| $[maxA, minA_0, \mu A, minS_{max}, maxD, cvA]$                                                | $86.47 \pm 2.69$                   | $86.36 \pm 3.68$                   | $86.47 \pm 3.44$                   |
| $[cvA, stdS_{max}, maxA_0, maxS_{max}, stdS_{5\%}, stdS_{10\%}]$                              | $87.86 \pm 1.73$                   | $87.07 \pm 2.98$                   | $88.90 \pm 3.06$                   |
| $[\mu A_0, cvD, stdS_{20\%}, stdS_{max}]$                                                     | <b><math>92.34 \pm 1.49</math></b> | <b><math>94.47 \pm 2.94</math></b> | <b><math>89.83 \pm 3.10</math></b> |
| $[\mu A, minA_0, maxA, cvA]$                                                                  | $87.01 \pm 2.01$                   | $86.52 \pm 3.78$                   | $87.47 \pm 3.98$                   |
| $[stdS_{max}, maxS_{max}, cvA, maxA_0]$                                                       | $86.19 \pm 2.08$                   | $85.64 \pm 3.00$                   | $86.80 \pm 3.10$                   |

Ranked Feature sets of length l=9,6,4, 1st row OOB, 2nd row MV, 3rd row OCI



## Classification results Lead X

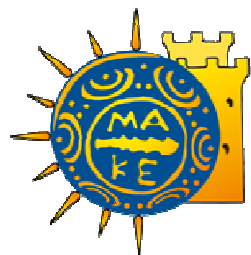
| Method | Feature Sets                                                              | ACC                                | SE                                 | SP                                 |
|--------|---------------------------------------------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| OOB    | $[\mu\Delta A, \sigma S_{20\%}, \sigma S_{max}, min\Delta A, minS_{max}]$ | <b><math>95.01 \pm 0.91</math></b> | <b><math>95.47 \pm 2.03</math></b> | <b><math>94.43 \pm 3.56</math></b> |
| OCI    | $[\sigma S_{max}, cv\Delta A, max\Delta A, maxS_{max}, cvA]$              | $87.87 \pm 2.04$                   | $90.05 \pm 2.25$                   | $85.50 \pm 4.38$                   |
| MV     | $[cv\Delta A, \mu A, max P_a, maxD, maxA]$                                | $81.92 \pm 2.65$                   | $83.19 \pm 2.21$                   | $80.50 \pm 3.36$                   |

| Feature Sets                                                                           | ACC                                | SE                                 | SP                                 |
|----------------------------------------------------------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| $[\mu\Delta A, \mu S_{max}, \sigma S_{20\%}, \sigma S_{max}, min\Delta A, minS_{max}]$ | $94.14 \pm 1.04$                   | $94.64 \pm 1.96$                   | $93.56 \pm 2.28$                   |
| $[\mu\Delta A, \sigma S_{20\%}, \sigma S_{max}, min\Delta A, minS_{max}]$              | <b><math>95.01 \pm 0.91</math></b> | <b><math>95.47 \pm 2.03</math></b> | <b><math>94.43 \pm 3.56</math></b> |
| $[\mu\Delta A, \sigma S_{max}, min\Delta A, minS_{max}]$                               | $93.42 \pm 1.92$                   | $93.71 \pm 2.48$                   | $93.06 \pm 3.96$                   |



## Classification results Lead Y

| Feature Sets                                                                               | ACC                                | SE                                | SP                                | Method    |
|--------------------------------------------------------------------------------------------|------------------------------------|-----------------------------------|-----------------------------------|-----------|
| $[\mu A_0, \mu S_{max}, cvS_{max},$<br>$stdS_{20\%}, stdS_{max}, stdS_{10\%}, minS_{max}]$ | $77.69 \pm 2.26$                   | $78.43 \pm 3.46$                  | $76.77 \pm 2.97$                  | min cond. |
| $[cvD, stdS_{20\%}, stdS_{5\%}, stdD,$<br>$maxS_{20\%}, maxA, maxS_{max}]$                 | $76.54 \pm 2.6$                    | $78.4 \pm 2.83$                   | $74.53 \pm 2.86$                  | max det   |
| $[\mu S_{max}, cvS_{max}, stdS_{5\%},$<br>$stdS_{max}, stdD, maxS_{20\%}, maxS_{max}]$     | $76.15 \pm 2.59$                   | $79.76 \pm 2.9$                   | $71.8 \pm 4.84$                   | MV        |
| $[cvD, stdS_{10\%}, stdD, maxS_{20\%},$<br>$maxA, maxS_{max}, minA_0]$                     | <b><math>78.49 \pm 2.16</math></b> | <b><math>80.79 \pm 3.3</math></b> | <b><math>75.43 \pm 3.8</math></b> | OCI       |



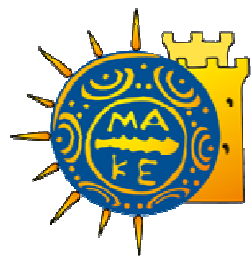
## Σύγκριση με State-of-the-art

| Dataset                       | ACC    | SE     | SP     | AUC  | Features                  | PAF    | Healthy |
|-------------------------------|--------|--------|--------|------|---------------------------|--------|---------|
| Martínez <i>et al.</i> [26]   | 86.33% | -      | -      | -    | V1 lead, 5min             | 46     | 53      |
| Conte <i>et al.</i> [25]      | 88%    | 92%    | 83%    | -    | lead II, 5min             | 76     | 40      |
| Vassilikos <i>et al.</i> [28] | -      | -      | -      | 85%  | X and Z lead, 10 min      | 50     | 50      |
| Filos <i>et al.</i> [27]      | 93.75% | 100%   | 87.5%  | -    | X, Y, Z lead, 10 min      | 29     | 25      |
| Attia <i>et al.</i> [23]      | 83.3%  | 82.3%  | 83.4%  | 0.90 | leads I,II, V1-V6, 10 sec | 180922 | N/D     |
| Chen <i>et al.</i> [24]       | 79.9%  | -      | -      | 0.90 | V5 lead, 10 sec           | 59     | 38      |
| Proposed method               | 95.01% | 95.47% | 94.43% | 0.98 | lead X, 10 min            | 69     | 59      |



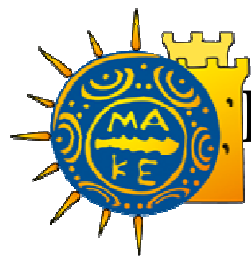
## Συμπεράσματα

- Με ακρίβεια  $ACC = 95.01\%$  ταξινομούνται οι υγιείς vs. ασθενείς με ιστορικό ΠΚΜ
- 5 νέα features του P-κύματος area  $\Delta A$  and slope  $S_{max}$ ,  $S_{20\%}$  επιτυγχάνουν state of the art αποτελέσματα
- Τα features της lead Y λόγω του ότι η δεξιά κολπική εκπόλωση συγκαλύπτεται με την αριστερή αποδίδουν χειρότερα
- Ο καρδιακός παλμός φαίνεται να έχει μικρή επιρροή στα features του P-κύματος
- Το καλύτερο FV επιλέχθηκε από την RF βάσει του Feature Importance



## Συμπεράσματα

- Έγκαιρη ανίχνευση της ΠΚΜ χωρίς εμφανείς ενδείξεις στο ΗΚΓ ή συμπτώματα
- Αποφυγή εγκεφαλικού επεισοδίου ή της μόνιμης ΚΜ μέσω άμεσης φαρμακευτικής αγωγής
- Αποφυγή επεμβατικών μεθόδων θεραπείας και κόστους υγειονομικής περίθαλψης



# Περιορισμοί και Μελλοντική Εργασία

- Ανάλυση της Z απαγωγής
- Στατιστική ανάλυση των χαρακτηριστικών πριν και μετά την εφαρμογή χρονικής κλίμακας
- Η σύγκριση με άλλες εργασίες ίσως οδηγεί σε ασάφειες
- Μεγαλύτερο δείγμα
- Προοπτική μελέτη