Comparison of Stacking-based Classifier Ensembles using Euclidean and Riemannian Geometries

Vitaliy Tayanov, Adam Krzyżak, Ching Y. Suen

Concordia University
Montreal, Canada

vtayanov@gmail.com

Overview

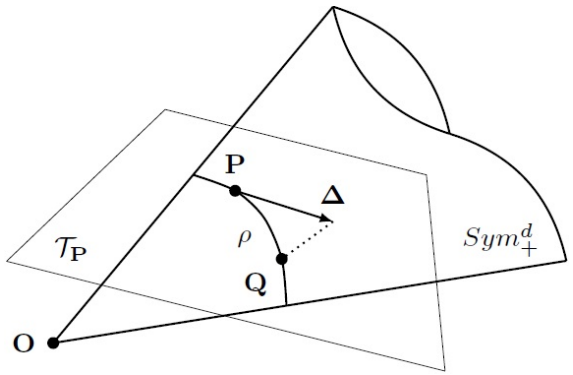
Objectives

- Build a nonlinear version of classifier stacking using classifier interactions or predictions from classifier subensembles (R manifolds of SPD matrices).
- Build a nonlinear version of cascades of classifier ensembles
- Carry out preliminary experiments on the motion-based data set such as Gesture Classification Data Set from UCI repository
- Compare nonlinear version of classifier stacking build using Riemannian manifolds of SPD matrices and CNN as meta-learner versus cascades of classifier ensembles in both Euclidean and Riemannian geometries.
- Compare SVM and stacking-based SVM to all aforementioned classification algorithms. This is important because SVMs are often used as classifiers in both conventional and representation learning.
- Compare all stacking-based algorithms on the general data sets of different scales from UCI repository

Riemannian manifolds of SPD matrices: geometrical view

The space of $d \times d$ SPD matrices Sym_+^d is an open convex cone

$$Sym_+^d = \bigcap_{x \in R^d} \{P \in Sym^d : x^T P x > 0\}$$



$$Sym_+^d(2) = \left\{ \begin{bmatrix} a & c \\ c & b \end{bmatrix}, a > 0, ab - c^2 > 0 \right\},$$

$$u = \frac{1}{2}(a+b), v = \frac{1}{2}(a-b),$$

$$c^2 + v^2 < u^2, u > 0.$$

Riemannian manifolds: Gauss maps

A pair of Gauss maps (logarithmic and exponential maps)

$$\begin{aligned} \exp_{\mathrm{P}}(\Delta) &= \exp(\log(\mathrm{P}) + \Delta) = \mathrm{Q} \\ \log_{\mathrm{P}}(\mathrm{Q}) &= \log(\mathrm{Q}) - \log(\mathrm{P}) = \Delta \end{aligned} \tag{1}$$

Geodesic on Sym_{+}^{d} in case of log Euclidean metric

$$d(\mathrm{P}, \mathrm{Q}) = ||\log(\mathrm{Q}) - \log(\mathrm{P})|| \tag{2}$$

Homotopy diagram for data transformation and learning on Riemannian manifolds of SPD matrices

$$
\begin{tikzcd}
X \arrow[r, "f_1"] & Z_1^X \arrow[r, "f_2"] \arrow[d, "\pi_1"] & Z_2^X \arrow[r] \arrow[d, "\pi_2"] & \cdots \arrow[r, "f_i"] & Z_i^X \arrow[r] \arrow[d, "\pi_i"] & \cdots \arrow[r, "f_n"] & Z_n^X \arrow[r] \arrow[d, "\pi_n"] & Y \\
& S_d^+(1) \arrow[r, "\phi_1"] \arrow[d, "\psi_1"] & S_d^+(2) \arrow[r] \arrow[d, "\psi_2"] & \cdots \arrow[r, "\phi_{i-1}"] & S_d^+(i) \arrow[r] \arrow[d, "\psi_{i-1}"] & \cdots \arrow[r, "\phi_{n-1}"] & S_d^+(n) \arrow[r] \arrow[d, "\psi_n"] & Y \\
& U(1) \arrow[r, "p_1"] \arrow[d, "h_1"] & U(2) \arrow[r] \arrow[d, "h_2"] & \cdots \arrow[r, "p_{i-1}"] & U(i) \arrow[r] \arrow[d, "h_{i-1}"] & \cdots \arrow[r, "p_{n-1}"] & U(n) \arrow[r] \arrow[d, "h_n"] & Y \\
& Z^U(1) \arrow[r, "t_1"] \arrow[d] & Z^U(2) \arrow[r] \arrow[d] & \cdots \arrow[r, "t_{i-1}"] & Z^U(i) \arrow[r] \arrow[d] & \cdots \arrow[r, "t_{n-1}"] & Z^U(n) \arrow[r] \arrow[d] & Y \\
& Y & Y & & Y & & Y &
\end{tikzcd}
$$

Computing classifier prediction pairwise matrix (CPPM)

We compose a tensor T of size $T \times T \times L$, where for each class $C_\ell, \ell = 1, ..., L$ we have a CPPM $\mathrm{A}^\ell(x)$ $T \times T$ with elements $a^\ell_{ij}, \{i, j\} = 1, ..., T$:

$$a^\ell_{ij}(x) = p_i(y = c_\ell|X)p_j(y = c_\ell|X) = h^\ell_i(x)h^\ell_j(x), i \neq j;$$
$$a^\ell_{ij} = p_i(y = c_\ell|X) = h^\ell_i, i = j, \tag{3}$$

Using matrix form we can write $A^\ell(x)$ as

$$\mathrm{A}^\ell(x) = \begin{bmatrix} h^\ell_1(x) & \dots & h^\ell_1(x)h^\ell_j(x) & \dots & h^\ell_1(x)h^\ell_T(x) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h^\ell_1(x)h^\ell_i(x) & \dots & h^\ell_i(x) & \dots & h^\ell_i(x)h^\ell_T(x) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h^\ell_1(x)h^\ell_T(x) & \dots & h^\ell_T(x)h^\ell_j(x) & \dots & h^\ell_T(x) \end{bmatrix} \tag{4}$$

Learning on Riemannian manifolds

The final distance can be computed as the sum of distances on all $L-1$ R manifolds:

$$d(x,y) = \sum_{i=1}^{L-1} d(x_i, y_i), \tag{5}$$

where $d(x_i, y_i)$ is a distance between two points on $\mathcal{M}_i$ manifold.
The projection matrix is computed as

$$Proj(\mathrm{A}^{\ell}) = \mathrm{U} \log(\Lambda) \mathrm{U}^T, \tag{6}$$

where U is the matrix of eigenvectors of A^{ℓ} and Λ is the diagonal matrix containing eigenvalues of A^{ℓ}.

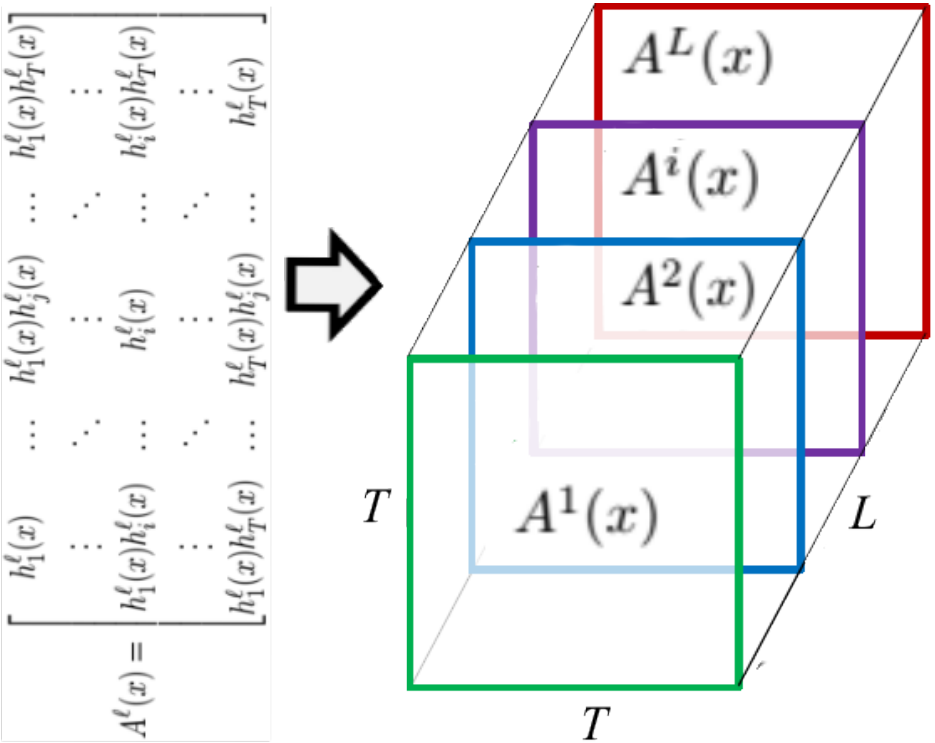
Experimental part: general machine learning data sets of different scales

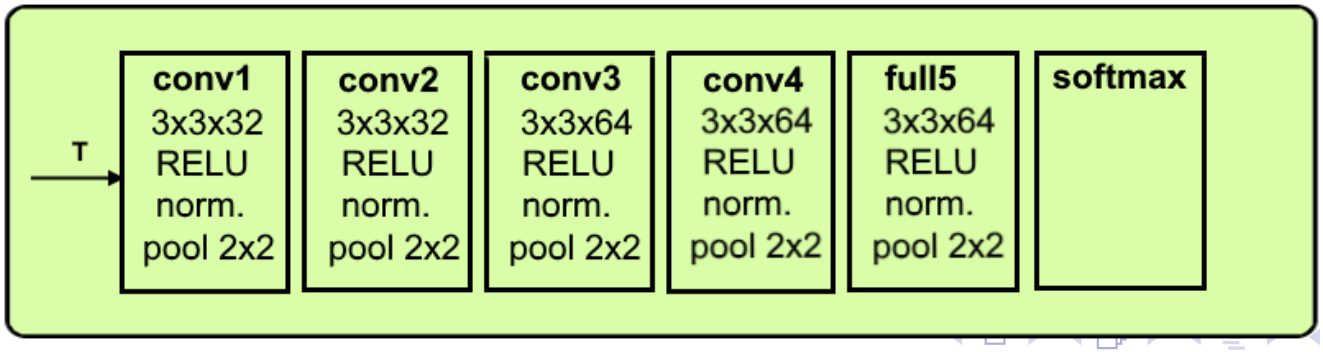
dataset	size	features	classes	Tr,%	Ts,%
balance	625	4	3	50	50
bupa	345	6	2	50	50
gamma	19200	10	2	50	50
german	1000	24	2	50	50
heart	270	13	2	50	50
mfeat-mor	2000	6	10	50	50
mfeat-zer	2000	47	10	50	50
pima	768	8	2	50	50
segment	2310	19	7	50	50
sonar	208	60	2	50	50
spambase	4601	57	2	50	50

Summary of characteristics of Gesture Phase Segmentation data set: raw data

data set	size	features	classes	Tr, %	Ts, %
gesture-raw1	1747	18	5	50	50
gesture-raw2	1264	18	5	50	50
gesture-raw3	1834	18	5	50	50
gesture-raw4	1073	18	5	50	50
gesture-raw5	1424	18	5	50	50
gesture-raw6	1111	18	5	50	50
gesture-raw7	1448	18	5	50	50

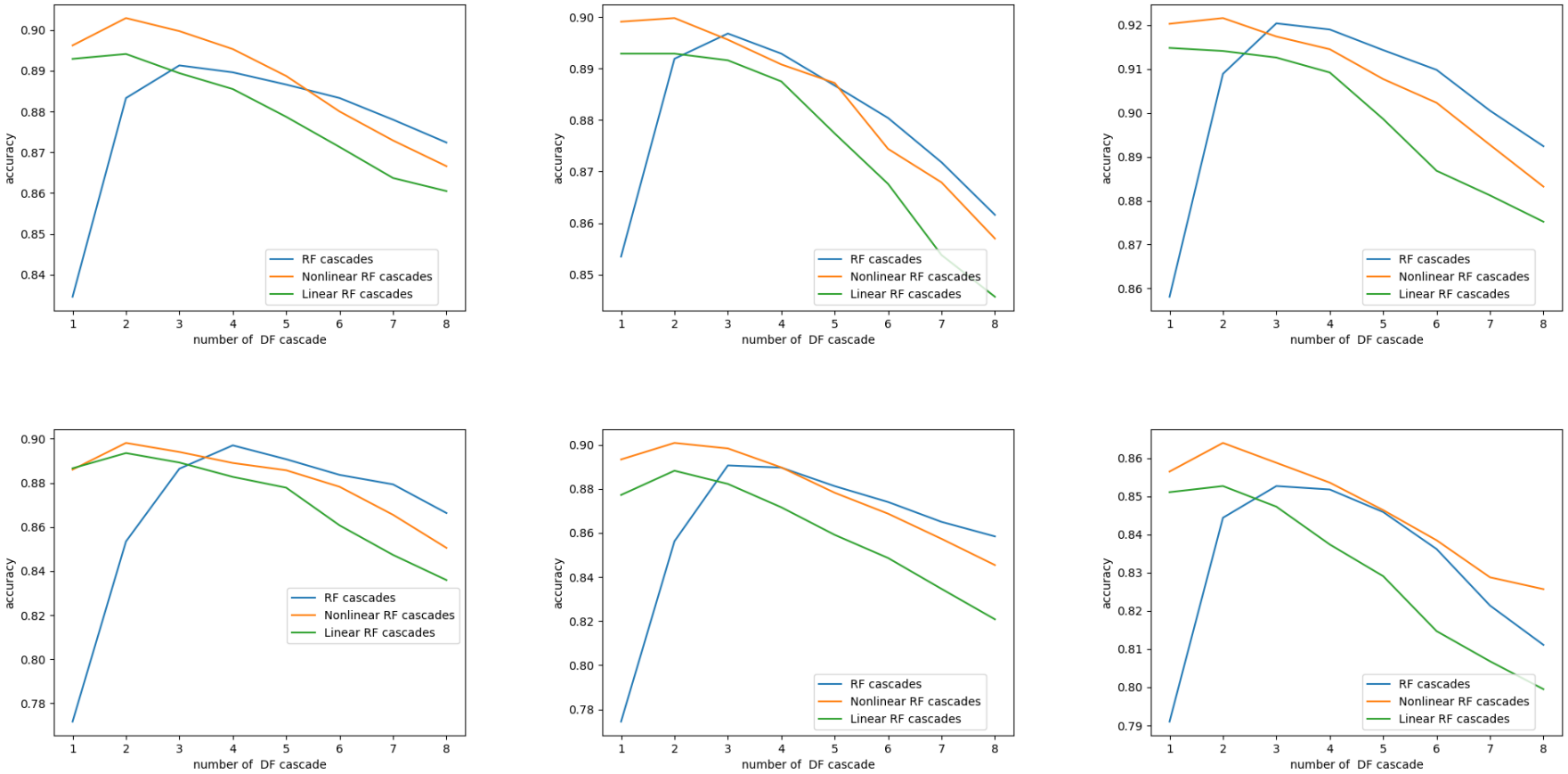
Formation of 3D tensors of SPD matrices and architecture of CNN to learn these tensors



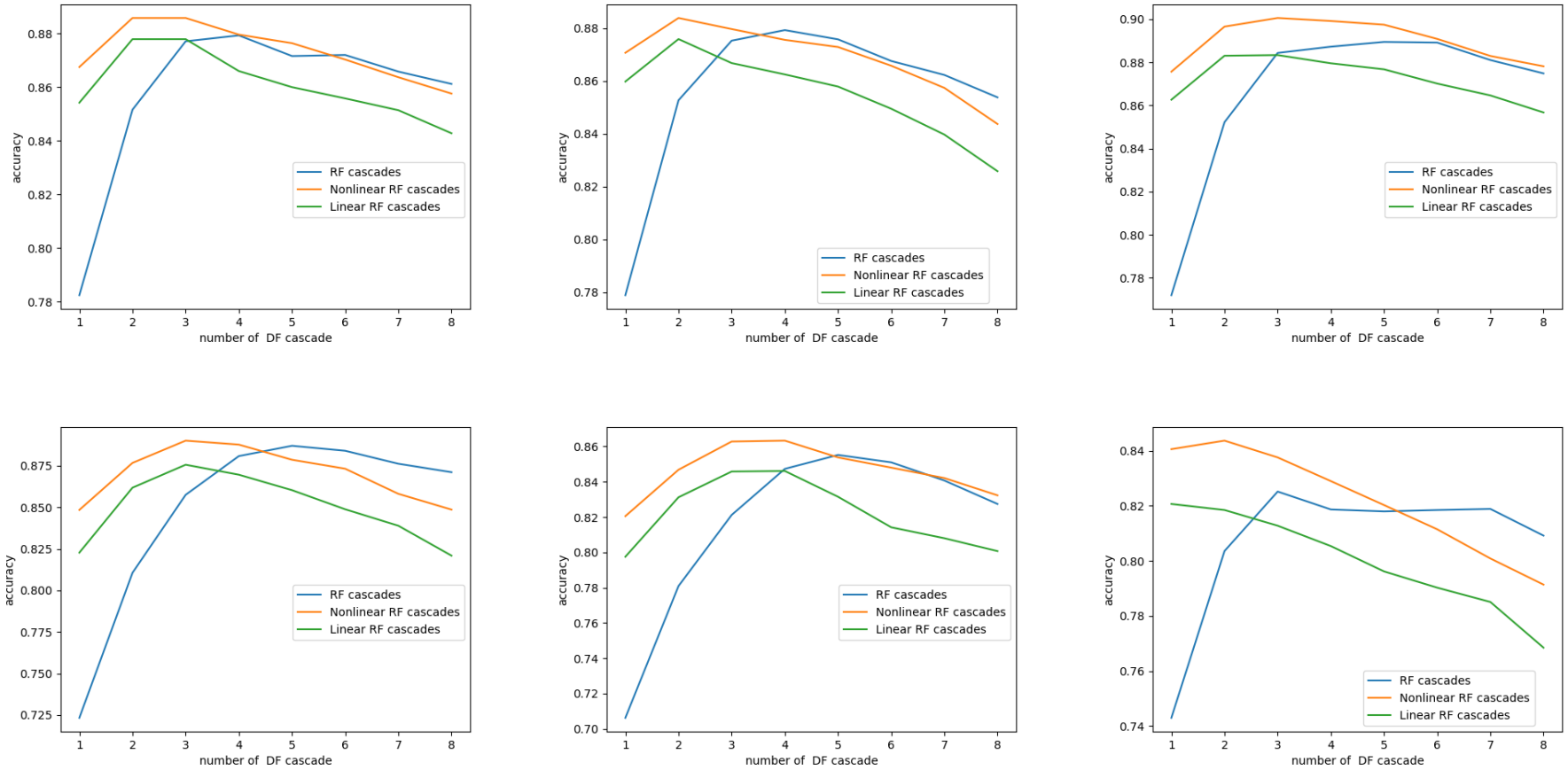


Method/Dataset	gesture-raw1	gesture-raw2	gesture-raw3	gesture-raw4	gesture-raw5	gesture-raw6	gesture-raw7
A number of decision trees in a DF is equal to 50 and its depth is equal to 2							
RF_1	$74,45 \pm 1,36$	$69,07 \pm 1,44$	$64,25 \pm 2,01$	$49,94 \pm 3,75$	$54,48 \pm 1,48$	$62,52 \pm 3,48$	$58,49 \pm 2,37$
RF_{max}	$76,31 \pm 1,50$	$73,38 \pm 1,44$	$68,12 \pm 1,39$	$55,25 \pm 5,96$	–	–	–
RF-nonlinear$_{max}$	$77,94 \pm 1,42$	$75,31 \pm 1,41$	$69,65 \pm 1,98$	$61,34 \pm 3,55$	$57,35 \pm 2,25$	$64,62 \pm 2,93$	$60,33 \pm 1,94$
RF-linear$_{max}$	$77,62 \pm 1,75$	$75,01 \pm 2,14$	$69,28 \pm 1,96$	$65,72 \pm 2,73$	$57,51 \pm 2,02$	$63,42 \pm 3,85$	$60,55 \pm 2,57$
CNN-R-nonlinear	$85,46 \pm 1,21$	$84,29 \pm 1,73$	$86,93 \pm 0,91$	$80,88 \pm 2,94$	$80,62 \pm 1,78$	$79,59 \pm 2,16$	$81,81 \pm 1,55$
CNN-R-linear	$85,69 \pm 1,23$	$84,94 \pm 1,53$	$87,19 \pm 1,02$	$80,99 \pm 3,25$	$80,46 \pm 2,00$	$80,54 \pm 2,15$	$82,17 \pm 1,66$
SVM(RBF)	$74,19 \pm 1,24$	$68,05 \pm 1,46$	$70,12 \pm 1,53$	$46,16 \pm 1,79$	$41,19 \pm 2,15$	$48,15 \pm 1,77$	$49,28 \pm 1,27$
SVM(RBF)-stacking	$80,41 \pm 1,49$	$77,91 \pm 2,23$	$81,25 \pm 1,69$	$69,78 \pm 3,07$	$69,44 \pm 3,12$	$73,33 \pm 2,01$	$74,59 \pm 1,76$
A number of decision trees in a DF are equal to 50 and its depth is equal to 5							
RF_1	$83,46 \pm 0,93$	$85,35 \pm 1,12$	$85,81 \pm 0,71$	$77,17 \pm 2,66$	$77,44 \pm 2,11$	$79,10 \pm 1,98$	$79,16 \pm 1,29$
RF_{max}	$89,13 \pm 1,88$	$89,68 \pm 1,04$	$92,04 \pm 1,12$	$89,70 \pm 0,94$	$89,05 \pm 1,75$	$85,27 \pm 1,06$	$86,20 \pm 1,41$
RF-nonlinear$_{max}$	$90,29 \pm 1,52$	$89,98 \pm 0,86$	$92,16 \pm 1,14$	$89,81 \pm 1,10$	$90,08 \pm 1,45$	$86,40 \pm 1,54$	$87,00 \pm 1,08$
RF-linear$_{max}$	$89,41 \pm 1,22$	$89,29 \pm 0,92$	$91,48 \pm 1,18$	$89,35 \pm 1,25$	$88,82 \pm 1,69$	$85,27 \pm 1,20$	$85,88 \pm 1,23$
CNN-R-nonlinear	$89,91 \pm 1,39$	$89,46 \pm 0,83$	$92,34 \pm 1,00$	$89,29 \pm 1,02$	$89,31 \pm 1,27$	$86,29 \pm 1,31$	$86,34 \pm 0,84$
CNN-R-linear	$90,67 \pm 1,14$	$90,33 \pm 1,16$	$93,04 \pm 1,06$	$90,73 \pm 1,31$	$90,45 \pm 1,22$	$87,27 \pm 1,53$	$87,87 \pm 1,14$
SVM(RBF)-stacking	$89,15 \pm 1,54$	$89,95 \pm 1,15$	$91,55 \pm 0,83$	$87,09 \pm 1,50$	$89,41 \pm 1,73$	$86,98 \pm 1,60$	$86,71 \pm 1,33$
A number of decision trees in ETs are equal to 50 and its depth is equal to 2							
RF_1	$74,98 \pm 0,75$	$68,97 \pm 1,00$	$62,99 \pm 0,92$	$45,90 \pm 2,21$	$49,90 \pm 2,95$	$58,62 \pm 3,01$	$55,40 \pm 3,11$
RF_{max}	$76,99 \pm 1,53$	$71,90 \pm 1,51$	$68,21 \pm 2,36$	$50,00 \pm 3,03$	$54,23 \pm 2,51$	$60,81 \pm 2,37$	$59,92 \pm 2,62$
RF-nonlinear$_{max}$	$77,87 \pm 1,21$	$73,96 \pm 1,64$	$70,92 \pm 1,58$	$55,51 \pm 4,56$	$56,76 \pm 1,59$	$63,21 \pm 2,55$	$61,15 \pm 2,46$
RF-linear$_{max}$	$77,84 \pm 0,78$	$73,58 \pm 1,07$	$68,78 \pm 1,48$	$61,15 \pm 3,75$	$55,74 \pm 2,59$	$63,41 \pm 1,86$	$60,69 \pm 2,38$
CNN-R-nonlinear	$87,14 \pm 1,66$	$85,92 \pm 1,43$	$87,21 \pm 0,68$	$81,73 \pm 2,34$	$83,89 \pm 2,03$	$80,40 \pm 2,43$	$84,03 \pm 1,51$
CNN-R-linear	$88,07 \pm 1,67$	$87,15 \pm 1,65$	$87,75 \pm 0,54$	$83,48 \pm 2,95$	$83,89 \pm 1,69$	$81,55 \pm 1,94$	$83,49 \pm 1,39$
SVM(RBF)-stacking	$81,69 \pm 1,00$	$78,51 \pm 2,20$	$82,88 \pm 1,00$	$71,34 \pm 2,61$	$73,19 \pm 2,74$	$77,28 \pm 2,97$	$79,41 \pm 1,69$
A number of decision trees in the ETs are equal to 50 and its depth is equal to 5							
RF_1	$78,24 \pm 1,40$	$77,88 \pm 1,79$	$77,20 \pm 2,01$	$72,33 \pm 3,65$	$70,62 \pm 1,87$	$74,30 \pm 2,06$	$72,75 \pm 2,38$
RF_{max}	$87,93 \pm 0,95$	$87,93 \pm 0,86$	$88,94 \pm 1,39$	$88,70 \pm 1,36$	$85,52 \pm 1,53$	$82,52 \pm 1,73$	$84,46 \pm 1,83$
RF-nonlinear$_{max}$	$88,58 \pm 1,42$	$88,39 \pm 0,80$	$90,05 \pm 0,81$	$89,01 \pm 1,32$	$86,33 \pm 1,33$	$84,37 \pm 2,10$	$84,85 \pm 1,73$
RF-linear$_{max}$	$87,79 \pm 0,87$	$87,59 \pm 1,26$	$88,33 \pm 1,62$	$87,56 \pm 1,18$	$84,61 \pm 1,47$	$82,07 \pm 1,72$	$83,20 \pm 2,22$
CNN-R-nonlinear	$89,28 \pm 0,98$	$89,41 \pm 1,13$	$90,88 \pm 0,66$	$89,39 \pm 2,14$	$89,14 \pm 1,04$	$86,22 \pm 1,79$	$86,93 \pm 1,52$
CNN-R-linear	$90,90 \pm 0,86$	$89,98 \pm 0,93$	$92,06 \pm 1,05$	$90,15 \pm 0,92$	$90,00 \pm 0,98$	$87,43 \pm 1,47$	$87,89 \pm 1,35$
SVM(RBF)-stacking	$87,49 \pm 1,41$	$87,63 \pm 0,93$	$88,17 \pm 1,22$	$85,49 \pm 1,85$	$86,18 \pm 1,19$	$85,00 \pm 1,65$	$85,46 \pm 1,60$

Classification accuracy as a function of the number of cascades of RFs plotted for six experiments from Gesture Phase Segmentation data set. Depth of decision trees in random forests is equal to 5



Classification accuracy as a function of the number of cascades of ETs plotted for six experiments from Gesture Phase Segmentation data set. Depth of decision trees in extratrees is equal to 5



Extratrees: number of trees=100, depth={2, 5}

Table: Learning classifier predictions using different classifier stacking techniques: means and standard deviations of prediction accuracy (shown in %) for each method on different data sets from UCI repository.

Method/Dataset	Balance	Bupa	Gamma	German	Heart	Mfeat-mor	Mfeat-zer	Pima	Segment	Sonar	Spambase
A number of decision trees in the ETs are equal to 100 and its depth is equal to 2											
ET	$83,75 \pm 3,81$	$58,95 \pm 2,33$	$71,03 \pm 0,69$	$70,06 \pm 1,40$	$79,02 \pm 2,61$	$64,58 \pm 3,44$	$62,18 \pm 2,90$	$66,12 \pm 1,04$	$79,32 \pm 5,76$	$75,96 \pm 3,13$	$66,35 \pm 2,68$
ET_{max}	–	$66,10 \pm 2,80$	$80,79 \pm 0,61$	$73,42 \pm 0,90$	$80,00 \pm 3,10$	–	–	$75,25 \pm 1,44$	–	–	$89,07 \pm 1,05$
ET-nonlinear$_{max}$	$83,40 \pm 1,83$	$67,33 \pm 3,17$	$81,73 \pm 0,64$	$74,22 \pm 0,97$	$79,77 \pm 3,14$	$33,73 \pm 6,62$	$43,52 \pm 7,18$	$75,44 \pm 1,57$	$50,31 \pm 7,39$	$75,96 \pm 4,85$	$89,63 \pm 1,02$
ET-linear$_{max}$	$82,82 \pm 2,46$	$66,34 \pm 2,58$	$80,71 \pm 1,16$	$74,24 \pm 1,97$	$80,30 \pm 3,17$	$34,76 \pm 6,55$	$50,05 \pm 5,43$	$75,29 \pm 2,52$	$56,27 \pm 9,25$	$74,42 \pm 3,11$	$90,50 \pm 0,85$
SVM-stacking	$85,54 \pm 2,89$	$58,26 \pm 2,01$	$80,64 \pm 0,90$	$70,52 \pm 1,56$	$79,02 \pm 2,61$	$57,21 \pm 2,16$	$37,83 \pm 3,57$	$75,42 \pm 2,00$	$82,56 \pm 4,85$	$76,54 \pm 3,47$	$89,69 \pm 1,31$
SVM-R-nonlinear	$85,87 \pm 2,93$	$58,31 \pm 1,91$	$80,81 \pm 0,87$	$71,80 \pm 1,97$	$79,02 \pm 2,61$	$29,12 \pm 5,85$	$15,2 \pm 2,99$	$75,86 \pm 1,42$	$61,59 \pm 7,20$	$76,44 \pm 2,33$	$90,18 \pm 0,95$
SVM-R-linear	$63,91 \pm 7,59$	$58,31 \pm 1,95$	$76,77 \pm 1,57$	$70,02 \pm 1,32$	$79,02 \pm 2,61$	$68,98 \pm 1,14$	$71,00 \pm 1,23$	$65,73 \pm 0,73$	$89,71 \pm 1,01$	$62,31 \pm 7,51$	$79,59 \pm 2,21$
kNN-stacking	$78,97 \pm 1,69$	$62,97 \pm 3,81$	$81,24 \pm 0,52$	$71,14 \pm 2,22$	$77,82 \pm 4,50$	$67,26 \pm 1,43$	$72,13 \pm 1,16$	$71,80 \pm 1,82$	$93,29 \pm 1,19$	$78,56 \pm 2,36$	$86,29 \pm 1,38$
kNN-R-nonlinear	$79,81 \pm 1,54$	$63,43 \pm 4,27$	$81,21 \pm 0,54$	$71,04 \pm 2,35$	$78,12 \pm 4,19$	$67,16 \pm 1,79$	$70,50 \pm 1,54$	$71,77 \pm 1,36$	$92,84 \pm 1,39$	$78,46 \pm 1,88$	$86,60 \pm 1,39$
kNN-R-linear	$80,16 \pm 1,02$	$64,53 \pm 4,31$	$81,23 \pm 0,44$	$70,44 \pm 1,37$	$73,98 \pm 4,04$	$67,08 \pm 1,35$	$71,15 \pm 1,66$	$71,67 \pm 1,01$	$93,07 \pm 0,86$	$73,94 \pm 4,22$	$91,10 \pm 0,70$
MLP-stacking	$53,56 \pm 1,37$	$57,44 \pm 10,50$	$80,43 \pm 1,31$	$73,20 \pm 2,68$	$59,55 \pm 26,30$	$17,01 \pm 4,18$	$16,23 \pm 3,79$	$74,69 \pm 1,32$	$19,35 \pm 6,32$	$50,58 \pm 12,37$	$90,50 \pm 0,86$
MLP-R-nonlinear	$46,99 \pm 2,27$	$54,07 \pm 10,13$	$84,25 \pm 0,61$	$74,68 \pm 1,21$	$28,65 \pm 14,89$	$10,48 \pm 0,96$	$10,68 \pm 1,47$	$75,36 \pm 1,31$	$18,02 \pm 4,13$	$52,12 \pm 6,96$	$91,88 \pm 0,61$
MLP-R-linear	$46,41 \pm 1,93$	$44,48 \pm 7,23$	$84,32 \pm 0,38$	$72,20 \pm 1,54$	$35,94 \pm 23,72$	$11,61 \pm 2,08$	$10,66 \pm 0,85$	$73,05 \pm 2,18$	$19,43 \pm 5,51$	$47,79 \pm 2,36$	$91,87 \pm 0,77$
CNN-R-nonlinear	$89,17 \pm 2,03$	$70,81 \pm 2,62$	$84,30 \pm 0,46$	$76,32 \pm 1,54$	$82,78 \pm 2,36$	$72,90 \pm 1,01$	$74,58 \pm 0,92$	$77,45 \pm 1,40$	$95,42 \pm 0,69$	$80,58 \pm 2,54$	$92,49 \pm 0,63$
CNN-R-linear	$85,83 \pm 1,10$	$69,13 \pm 3,94$	$84,58 \pm 0,48$	$75,04 \pm 0,99$	$81,35 \pm 2,58$	$73,05 \pm 0,89$	$76,93 \pm 1,09$	$76,46 \pm 1,45$	$95,66 \pm 0,68$	$74,04 \pm 3,85$	$92,45 \pm 0,75$
A number of decision trees in the ETs are equal to 100 and its depth is equal to 5											
ET	$87,08 \pm 1,83$	$64,01 \pm 5,08$	$78,77 \pm 0,38$	$72,10 \pm 1,57$	$78,95 \pm 2,60$	$69,50 \pm 0,66$	$74,61 \pm 1,30$	$73,78 \pm 1,45$	$89,79 \pm 1,14$	$78,85 \pm 3,10$	$82,36 \pm 1,66$
ET_{max}	–	$69,24 \pm 3,01$	$85,70 \pm 0,42$	$74,92 \pm 1,32$	$80,75 \pm 1,85$	–	–	$76,20 \pm 1,51$	$91,66 \pm 1,05$	$79,71 \pm 3,43$	$93,10 \pm 0,53$
ET-nonlinear$_{max}$	$82,82 \pm 1,64$	$69,07 \pm 3,73$	$85,90 \pm 0,40$	$75,34 \pm 1,29$	$80,98 \pm 1,75$	$62,65 \pm 4,95$	$72,54 \pm 1,39$	$75,36 \pm 1,53$	$86,05 \pm 4,16$	$79,33 \pm 4,33$	$93,58 \pm 0,61$
ET-linear$_{max}$	$83,81 \pm 1,42$	$68,31 \pm 2,10$	$85,39 \pm 0,29$	$73,64 \pm 1,85$	$80,98 \pm 2,18$	$62,01 \pm 3,46$	$73,64 \pm 0,93$	$75,57 \pm 1,72$	$90,82 \pm 3,18$	$78,17 \pm 2,82$	$93,30 \pm 0,63$
SVM-stacking	$87,15 \pm 1,06$	$67,73 \pm 2,84$	$84,49 \pm 0,45$	$74,70 \pm 1,64$	$80,83 \pm 1,85$	$69,71 \pm 0,57$	$72,83 \pm 0,69$	$76,09 \pm 1,35$	$91,55 \pm 1,21$	$79,90 \pm 3,14$	$92,43 \pm 0,67$
SVM-R-nonlinear	$86,38 \pm 1,78$	$68,43 \pm 3,53$	$84,62 \pm 0,44$	$74,82 \pm 1,32$	$81,05 \pm 1,84$	$65,40 \pm 1,99$	$38,65 \pm 3,56$	$75,94 \pm 1,22$	$89,84 \pm 1,32$	$79,52 \pm 2,85$	$92,48 \pm 0,78$
SVM-R-linear	$85,03 \pm 1,78$	$59,30 \pm 3,04$	$79,32 \pm 0,79$	$71,22 \pm 1,86$	$80,15 \pm 2,13$	$69,88 \pm 0,69$	$78,40 \pm 0,61$	$69,84 \pm 2,16$	$92,85 \pm 1,15$	$76,35 \pm 3,31$	$85,83 \pm 1,84$
kNN-stacking	$80,93 \pm 1,53$	$66,86 \pm 3,41$	$83,58 \pm 0,538$	$72,76 \pm 1,71$	$79,77 \pm 1,32$	$67,93 \pm 1,00$	$75,89 \pm 0,83$	$72,53 \pm 1,74$	$95,46 \pm 0,97$	$79,33 \pm 3,73$	$90,21 \pm 1,94$
kNN-R-nonlinear	$80,93 \pm 1,79$	$67,03 \pm 2,97$	$83,63 \pm 0,37$	$72,72 \pm 2,00$	$80,60 \pm 2,17$	$67,43 \pm 1,04$	$70,95 \pm 1,80$	$72,03 \pm 2,09$	$95,28 \pm 0,68$	$79,33 \pm 2,99$	$90,95 \pm 1,71$
kNN-R-linear	$78,75 \pm 1,51$	$63,31 \pm 1,48$	$83,36 \pm 0,42$	$70,26 \pm 2,09$	$76,62 \pm 2,78$	$67,71 \pm 1,41$	$76,02 \pm 0,98$	$68,85 \pm 2,02$	$95,30 \pm 1,17$	$76,83 \pm 3,83$	$91,67 \pm 0,68$
MLP-stacking	$61,25 \pm 1,32$	$40,06 \pm 7,72$	$84,34 \pm 0,52$	$69,06 \pm 3,83$	$33,45 \pm 7,44$	$14,08 \pm 4,75$	$14,46 \pm 4,32$	$48,98 \pm 1,89$	$19,04 \pm 6,39$	$48,75 \pm 6,79$	$92,07 \pm 0,81$
MLP-R-nonlinear	$49,90 \pm 6,79$	$59,77 \pm 6,28$	$85,65 \pm 0,41$	$74,88 \pm 2,54$	$44,89 \pm 24,41$	$11,26 \pm 2,20$	$10,31 \pm 0,69$	$75,21 \pm 1,16$	$20,44 \pm 5,34$	$50,77 \pm 6,84$	$93,33 \pm 0,63$
MLP-R-linear	$47,79 \pm 1,93$	$58,37 \pm 8,77$	$85,48 \pm 0,65$	$74,54 \pm 1,49$	$40,60 \pm 21,61$	$12,76 \pm 2,49$	$11,62 \pm 2,30$	$73,98 \pm 1,67$	$16,73 \pm 3,90$	$57,79 \pm 1,40$	$93,36 \pm 0,48$
CNN-R-nonlinear	$88,14 \pm 2,01$	$71,10 \pm 2,47$	$85,71 \pm 0,31$	$76,32 \pm 1,54$	$82,85 \pm 2,20$	$73,50 \pm 0,84$	$78,52 \pm 0,89$	$77,19 \pm 1,16$	$95,42 \pm 0,69$	$81,25 \pm 3,14$	$93,80 \pm 0,64$
CNN-R-linear	$88,21 \pm 1,45$	$67,97 \pm 2,09$	$85,53 \pm 0,31$	$75,04 \pm 0,99$	$81,95 \pm 1,93$	$73,53 \pm 0,97$	$79,14 \pm 0,99$	$75,39 \pm 1,19$	$95,66 \pm 0,68$	$81,64 \pm 2,90$	$93,20 \pm 0,51$

Random Forests: number of trees=100, depth={2, 5}

Method/Dataset	Balance	Bupa	Gamma	German	Heart	Mfeat-mor	Mfeat-zer	Pima	Segment	Sonar	Spambase
A number of decision trees in a RF is equal to 100 and its depth is equal to 2											
RF	$80,06 \pm 4,41$	$68,72 \pm 4,03$	$75,84 \pm 0,67$	$70,26 \pm 1,34$	$79,25 \pm 2,54$	$62,02 \pm 3,90$	$58,37 \pm 3,42$	$75,16 \pm 1,07$	$77,50 \pm 5,24$	$75,29 \pm 2,98$	$89,10 \pm 0,75$
RF_{max}	$81,92 \pm 1,90$	$68,77 \pm 3,29$	$80,89 \pm 0,79$	$74,16 \pm 1,56$	$80,82 \pm 2,54$	–	–	–	–	$76,44 \pm 3,71$	$92,12 \pm 0,36$
RF-nonlinear$_{max}$	$83,27 \pm 1,95$	$69,59 \pm 3,80$	$82,47 \pm 0,60$	$74,60 \pm 1,51$	$80,30 \pm 2,72$	$45,98 \pm 3,19$	$43,52 \pm 6,99$	$76,12 \pm 1,47$	$70,26 \pm 7,12$	$77,69 \pm 3,82$	$92,63 \pm 0,29$
RF-linear$_{max}$	$83,43 \pm 2,15$	$69,94 \pm 3,15$	$82,95 \pm 0,63$	$74,34 \pm 1,95$	$80,83 \pm 3,23$	$46,66 \pm 4,44$	$47,35 \pm 5,85$	$76,12 \pm 1,02$	$71,54 \pm 8,01$	$76,63 \pm 4,17$	$92,51 \pm 0,28$
SVM-stacking	$82,50 \pm 4,37$	$67,62 \pm 3,69$	$81,31 \pm 0,57$	$72,16 \pm 2,21$	$80,53 \pm 2,98$	$53,10 \pm 3,64$	$45,65 \pm 7,03$	$76,09 \pm 1,32$	$71,12 \pm 5,70$	$76,73 \pm 3,87$	$92,54 \pm 0,42$
SVM-R-nonlinear	$82,53 \pm 4,46$	$68,60 \pm 3,01$	$81,44 \pm 0,56$	$73,22 \pm 1,93$	$80,83 \pm 2,80$	$44,28 \pm 3,50$	$27,38 \pm 2,15$	$76,25 \pm 1,16$	$65,77 \pm 6,46$	$77,21 \pm 3,60$	$92,61 \pm 0,43$
SVM-R-linear	$71,03 \pm 5,87$	$59,88 \pm 2,16$	$79,73 \pm 0,34$	$70,04 \pm 1,34$	$79,32 \pm 3,09$	$65,60 \pm 2,43$	$65,34 \pm 1,89$	$75,03 \pm 1,29$	$83,37 \pm 4,23$	$74,13 \pm 2,93$	$89,97 \pm 1,29$
kNN-stacking	$79,17 \pm 1,35$	$66,86 \pm 2,66$	$81,65 \pm 0,47$	$70,44 \pm 1,81$	$80,53 \pm 2,09$	$58,90 \pm 3,24$	$65,22 \pm 1,70$	$72,97 \pm 1,09$	$83,40 \pm 6,24$	$77,98 \pm 3,69$	$89,52 \pm 2,29$
kNN-R-nonlinear	$79,49 \pm 1,75$	$68,02 \pm 2,59$	$82,25 \pm 0,48$	$70,40 \pm 1,92$	$80,68 \pm 2,31$	$59,36 \pm 2,64$	$64,29 \pm 1,70$	$72,89 \pm 1,32$	$83,37 \pm 5,91$	$78,08 \pm 3,64$	$90,39 \pm 1,81$
kNN-R-linear	$77,95 \pm 2,13$	$63,43 \pm 3,82$	$81,46 \pm 0,91$	$71,04 \pm 1,13$	$78,12 \pm 2,23$	$62,85 \pm 3,85$	$64,53 \pm 1,26$	$73,88 \pm 1,56$	$87,31 \pm 2,67$	$75,48 \pm 3,28$	$92,50 \pm 0,38$
MLP-stacking	$50,16 \pm 1,01$	$58,49 \pm 9,05$	$83,15 \pm 0,60$	$74,32 \pm 1,30$	$51,58 \pm 26,81$	$13,43 \pm 4,62$	$11,05 \pm 11,20$	$74,38 \pm 1,56$	$20,36 \pm 7,29$	$53,65 \pm 11,62$	$92,76 \pm 0,37$
MLP-R-nonlinear	$47,08 \pm 2,29$	$60,23 \pm 7,29$	$84,15 \pm 0,50$	$74,06 \pm 2,33$	$30,30 \pm 14,37$	$12,03 \pm 3,28$	$11,40 \pm 2,35$	$76,85 \pm 1,30$	$21,63 \pm 6,74$	$57,88 \pm 10,51$	$93,56 \pm 0,40$
MLP-R-linear	$46,63 \pm 2,20$	$51,92 \pm 10,58$	$84,14 \pm 0,61$	$72,52 \pm 2,05$	$30,30 \pm 14,37$	$13,84 \pm 3,49$	$12,20 \pm 2,34$	$74,24 \pm 2,89$	$21,63 \pm 6,74$	$58,75 \pm 10,87$	$93,34 \pm 0,54$
CNN-R-nonlinear	$86,35 \pm 1,53$	$73,31 \pm 3,56$	$84,35 \pm 0,54$	$75,74 \pm 1,11$	$83,38 \pm 2,27$	$70,40 \pm 0,85$	$71,99 \pm 1,48$	$77,81 \pm 1,40$	$89,88 \pm 2,73$	$80,00 \pm 3,21$	$93,63 \pm 0,21$
CNN-R-linear	$85,90 \pm 1,93$	$69,01 \pm 3,05$	$84,31 \pm 0,51$	$74,50 \pm 1,48$	$82,11 \pm 2,42$	$70,68 \pm 0,78$	$73,44 \pm 1,42$	$76,54 \pm 0,84$	$90,10 \pm 2,78$	$77,12 \pm 3,84$	$93,70 \pm 0,30$
A number of decision trees in a RF are equal to 100 and its depth is equal to 5											
RF	$79,87 \pm 4,41$	$71,74 \pm 3,48$	$83,89 \pm 0,29$	$73,48 \pm 1,79$	$80,60 \pm 2,93$	$69,32 \pm 0,84$	$73,94 \pm 0,90$	$76,54 \pm 1,36$	$92,74 \pm 1,12$	$77,50 \pm 2,48$	$92,74 \pm 0,53$
RF_{max}	$82,08 \pm 2,43$	–	$86,49 \pm 0,34$	$75,30 \pm 1,23$	$81,28 \pm 2,46$	–	–	–	$93,05 \pm 3,84$	$78,94 \pm 3,76$	$94,22 \pm 0,38$
RF-nonlinear$_{max}$	$82,92 \pm 2,99$	$70,35 \pm 3,46$	$86,60 \pm 0,30$	$75,12 \pm 1,67$	$81,28 \pm 2,50$	$68,49 \pm 2,15$	$72,54 \pm 1,66$	$75,31 \pm 1,12$	$94,90 \pm 1,24$	$79,52 \pm 2,98$	$94,31 \pm 0,34$
RF-linear$_{max}$	$83,49 \pm 2,32$	$69,71 \pm 2,82$	$86,48 \pm 0,34$	$74,98 \pm 1,26$	$81,20 \pm 3,29$	$67,83 \pm 2,26$	$72,42 \pm 1,12$	$75,49 \pm 1,14$	$94,37 \pm 1,00$	$79,71 \pm 2,21$	$94,28 \pm 0,36$
SVM-stacking	$81,54 \pm 4,89$	$70,99 \pm 3,16$	$85,83 \pm 0,34$	$74,62 \pm 1,17$	$80,83 \pm 2,92$	$69,98 \pm 0,76$	$73,88 \pm 0,93$	$75,55 \pm 1,26$	$93,80 \pm 1,09$	$79,42 \pm 2,16$	$93,61 \pm 0,39$
SVM-R-nonlinear	$81,70 \pm 4,88$	$70,81 \pm 2,94$	$85,95 \pm 0,34$	$75,02 \pm 1,35$	$80,90 \pm 2,84$	$69,36 \pm 0,77$	$61,86 \pm 3,33$	$75,99 \pm 1,25$	$92,63 \pm 1,32$	$78,94 \pm 2,60$	$93,63 \pm 0,33$
SVM-R-linear	$71,73 \pm 6,28$	$69,07 \pm 3,71$	$83,69 \pm 0,48$	$73,20 \pm 0,77$	$79,62 \pm 3,23$	$69,95 \pm 0,73$	$76,39 \pm 0,67$	$75,34 \pm 1,42$	$94,05 \pm 1,36$	$76,92 \pm 3,44$	$92,95 \pm 0,40$
kNN-stacking	$78,85 \pm 2,47$	$70,70 \pm 3,72$	$85,07 \pm 0,35$	$74,14 \pm 1,76$	$80,75 \pm 2,92$	$68,58 \pm 0,72$	$74,08 \pm 0,98$	$74,82 \pm 1,66$	$94,88 \pm 0,88$	$78,37 \pm 2,33$	$93,46 \pm 0,28$
kNN-R-nonlinear	$79,04 \pm 2,34$	$70,76 \pm 3,94$	$85,07 \pm 0,35$	$74,00 \pm 1,50$	$81,50 \pm 3,05$	$68,22 \pm 0,70$	$69,16 \pm 1,04$	$75,16 \pm 1,46$	$94,84 \pm 0,95$	$78,37 \pm 1,98$	$93,35 \pm 0,38$
kNN-R-linear	$78,21 \pm 1,46$	$65,41 \pm 2,62$	$84,86 \pm 0,34$	$70,30 \pm 1,47$	$80,60 \pm 2,59$	$68,74 \pm 0,93$	$73,62 \pm 0,76$	$73,54 \pm 1,77$	$94,67 \pm 0,84$	$76,54 \pm 4,28$	$92,99 \pm 0,40$
MLP-stacking	$47,40 \pm 1,98$	$43,37 \pm 8,90$	$85,52 \pm 0,44$	$74,32 \pm 1,30$	$50,68 \pm 18,98$	$11,99 \pm 3,46$	$12,43 \pm 3,02$	$66,59 \pm 6,93$	$22,96 \pm 6,57$	$52,88 \pm 11,43$	$93,87 \pm 0,58$
MLP-R-nonlinear	$48,33 \pm 4,58$	$58,26 \pm 9,59$	$86,35 \pm 0,41$	$74,06 \pm 2,33$	$26,62 \pm 11,58$	$11,70 \pm 2,85$	$10,68 \pm 0,99$	$76,56 \pm 1,29$	$18,50 \pm 5,21$	$52,69 \pm 10,94$	$94,24 \pm 0,37$
MLP-R-linear	$46,28 \pm 2,00$	$61,69 \pm 8,19$	$86,31 \pm 0,42$	$72,52 \pm 2,05$	$33,16 \pm 24,00$	$17,38 \pm 3,26$	$12,28 \pm 2,32$	$75,63 \pm 1,83$	$24,65 \pm 4,95$	$50,38 \pm 9,65$	$94,20 \pm 0,38$
CNN-R-nonlinear	$86,51 \pm 2,12$	$73,78 \pm 3,57$	$86,39 \pm 0,35$	$75,74 \pm 1,11$	$82,63 \pm 2,70$	$73,00 \pm 0,48$	$77,28 \pm 0,97$	$77,71 \pm 1,24$	$95,97 \pm 0,49$	$81,44 \pm 2,43$	$94,20 \pm 0,26$
CNN-R-linear	$85,87 \pm 1,58$	$73,02 \pm 3,66$	$86,18 \pm 0,26$	$74,50 \pm 1,48$	$82,41 \pm 3,25$	$73,44 \pm 0,68$	$77,71 \pm 0,75$	$77,32 \pm 0,98$	$96,38 \pm 0,62$	$80,67 \pm 1,90$	$94,33 \pm 0,32$

Conclusions

- Using classifier interactions as some sort of a classifier stacking is advantageous in comparison to a classical both simple classifier stacking or cascades of classifier ensembles which can be considered as recursive stacking. Because interactions between classifiers can not be presented in Euclidean geometry we need to use Riemannian manifolds of Symmetric Positive Definite matrices.
- Interactions between classifiers depend on predictions made by those individual classifiers which are dependent on properties and parameters of those classifiers. This means that properties of a Riemannian manifold (such as for instance local curvature) depend on which classifiers we use in a classifier ensemble.
- Using cascades of classifier ensembles such as Random Forests or Extratrees is advantageous for most of the general problems from UCI repository and problems from Gesture Phase Segmentation data set. We need less cascades to achieve the maximum of prediction accuracy, when applying Riemannian manifolds to cascaded classifier stacking, in comparison with what we need in Euclidean geometry. The maximum of prediction accuracy using cascaded classifier stacking is higher if one uses Riemannian manifolds.