Audio Signal Classification Using Linear Predictive Coding and Random Forests

Lăcrimioara GRAMA, Corneliu RUSU

Technical University of Cluj-Napoca
Faculty of Electronics, Telecommunications and Information Technology
Basis of Electronics Department
Signal Processing Group

SpeD 2017
Bucharest, Romania
July 6-9, 2017

Outline

- Research aim
- Acoustic Wildlife Intruder Detection System
- Wildlife Database
- Linear Predictive Coding
- Random Forests
- Stratified 10-fold cross validation
- Results
- Conclusion

Research Aim

- Audio signal classification system based on Linear Predictive Coding and Random Forests
  - Acoustic wildlife intruder detection system (WIDS)
- Sound classification has been the focus of intensive research and several approaches have been proposed in different domains
  - Medical applications: hearing aids and remote monitoring
  - Identification of the musical instruments from an audio recording
  - Environmental sound classification
  - Classification of the kitchen sounds
  - Vehicle identification

Why this research?

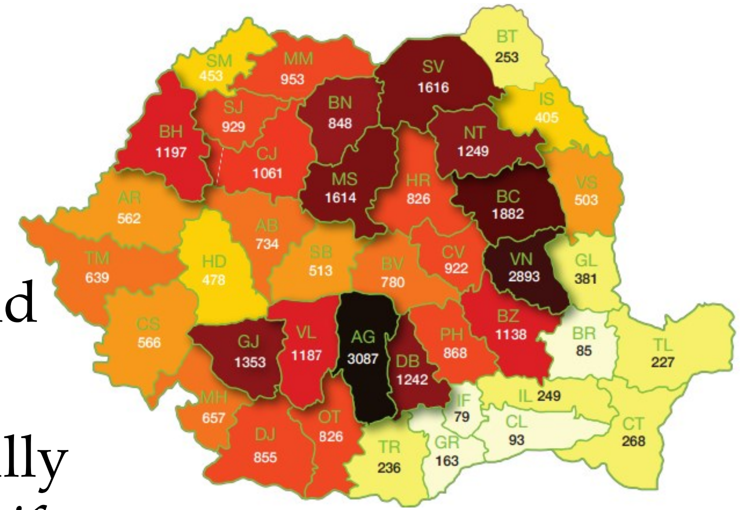
- The number of events that imply
  - Illegal logging, hunting,
  - Trespassing of natural reservations, parks, forests

  increased so much in the past decade

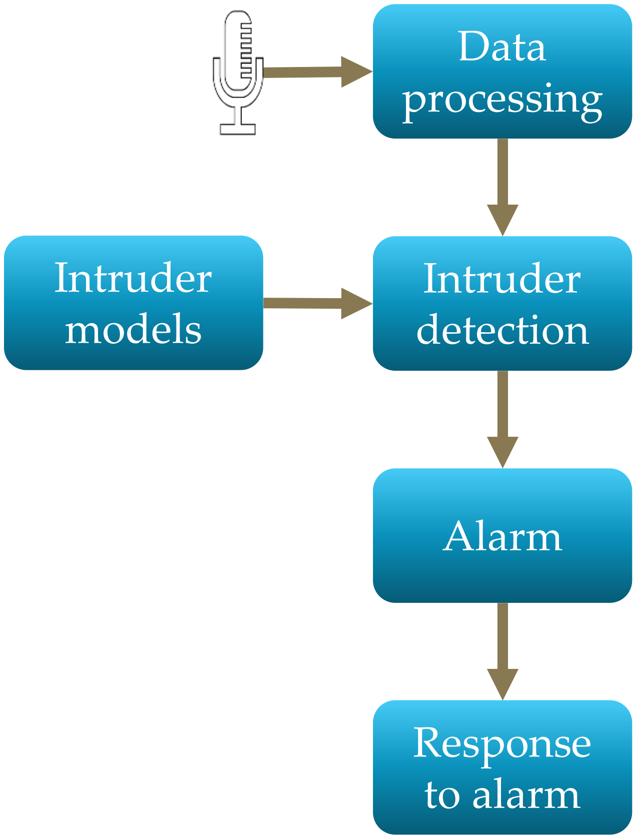
  ⇒ On a high demand became the design of WIDS
- To detect in time unwanted activities within the protected areas + help the authorities to take an action

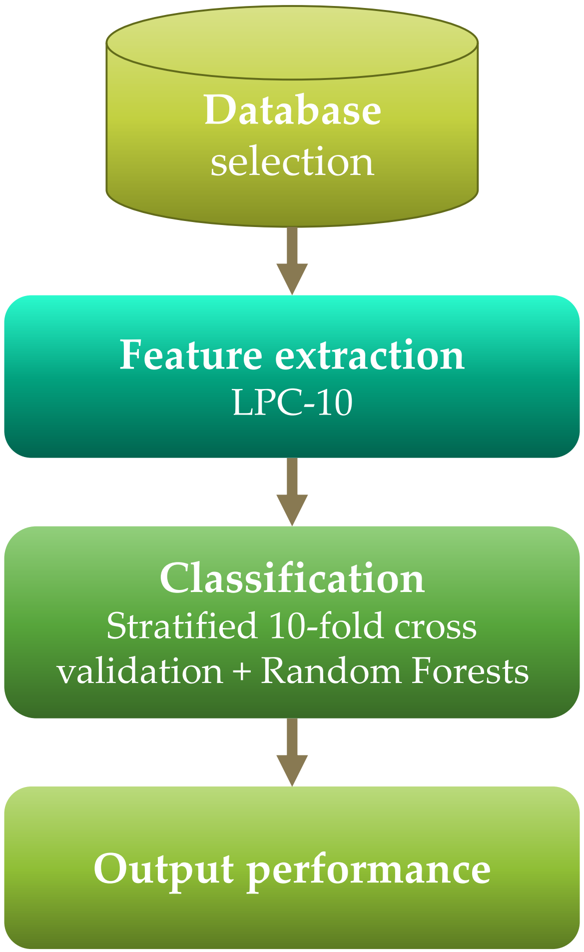
Why this research?

- Over 25 environmental agencies and organizations world wide, are being proactive in tracking illegal logging and hunting
- About 25 million birds are killed illegally in the Mediterranean every year [*BirdLife International 2017*]
- Romania: in 2015 the authorities registered 34 870 cases of illegal logging, which means 96 cases/day [*Greenpeace 2015*]
  - Regarding the gravity of the deeds, of all cases of illegal logging recorded in 2015, 32% of them were classified as criminal offences, while 68% were contraventions



Acoustic Wildlife Intruder Detection System





Wildlife Database

Birds dataset – 654 audio files originated from 70 different species of birds (Internet)

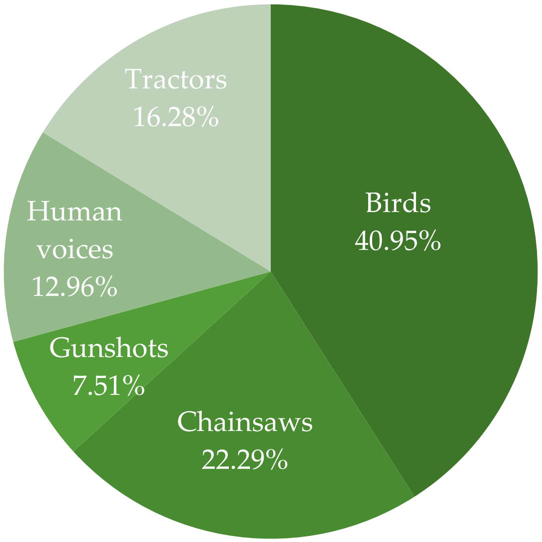
Chainsaws dataset – 356 audio files originated from 18 different types of chainsaws (SPG)

Gunshots dataset – 120 audio files originated from 40 different types of guns (Internet)

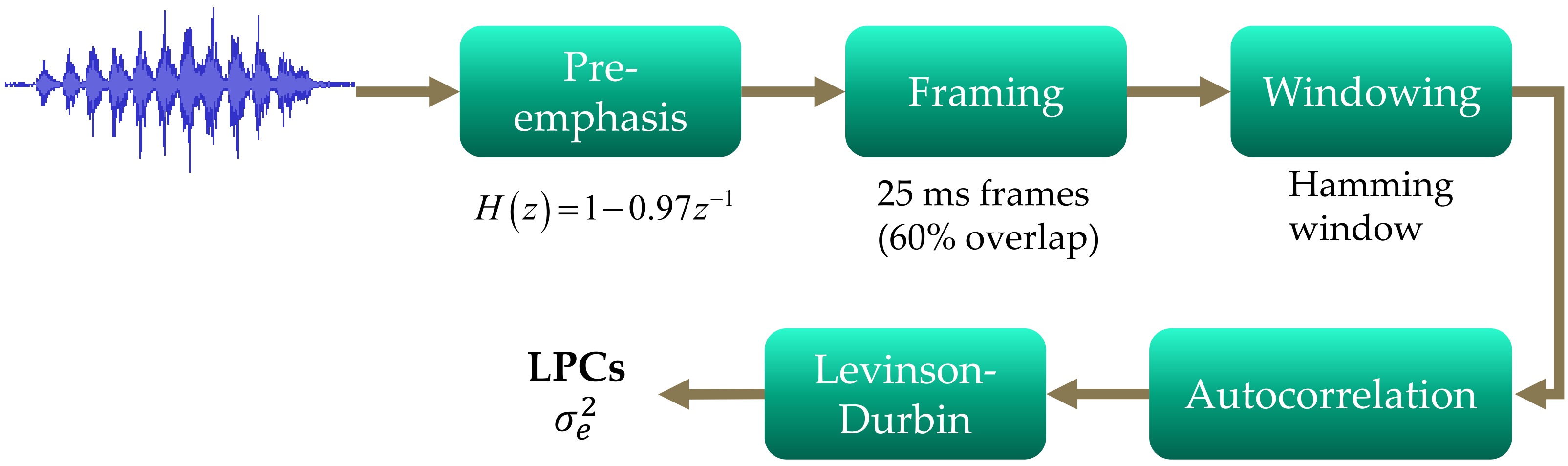
Human voice dataset – 207 speech sounds originated from 50 different former students from the TUCN

Tractors dataset – 260 audio files originated from 17 different types of tractors (SPG)

- 16 kHz, 16-bit
- None of the audio signals are studio recordings ⇒ they are subject to some additive noise from surroundings



Linear Predictive Coding Coefficients



- Fetures vector $F_k = \begin{bmatrix}\sigma_k^2 & a_{k,1} & a_{k,2} & \dots & a_{k,10}\end{bmatrix}$
- Features matrix $F_{Nx11} = \begin{bmatrix} \sigma_1^2 & a_{1,1} & \cdots & a_{1,10} \\ \sigma_2^2 & a_{2,1} & \cdots & a_{2,10} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_N^2 & a_{N,1} & \cdots & a_{N,10} \end{bmatrix}$

- σ_k^2 – prediction error variance
- $a_{k,i}$ – last 10 LPC coefficients
- $N = 1\,597$ – number of audio files

Why Random Forests?

- Acoustic WIDS – look for suspiciouss sound signals
  - Attack/unauthorized access to the natural environment
  - At an abstract level – WIDS purpose – to classify the input correctly as non-intruders or intruders
- Tradition systems can detect known intruders but cannot identify unknown ones

  ⇒ Nowadays machine learning techniques are attempting to be apply to this area of cybersecurity
- Many industries use machine learning techniques to better automate
  - Security screening
  - Border entry
  - College applicant selection
  - Loan analytics
  - Health care
- Almost all kind of stuffs can be tackled with machine learning in order to take good decisions

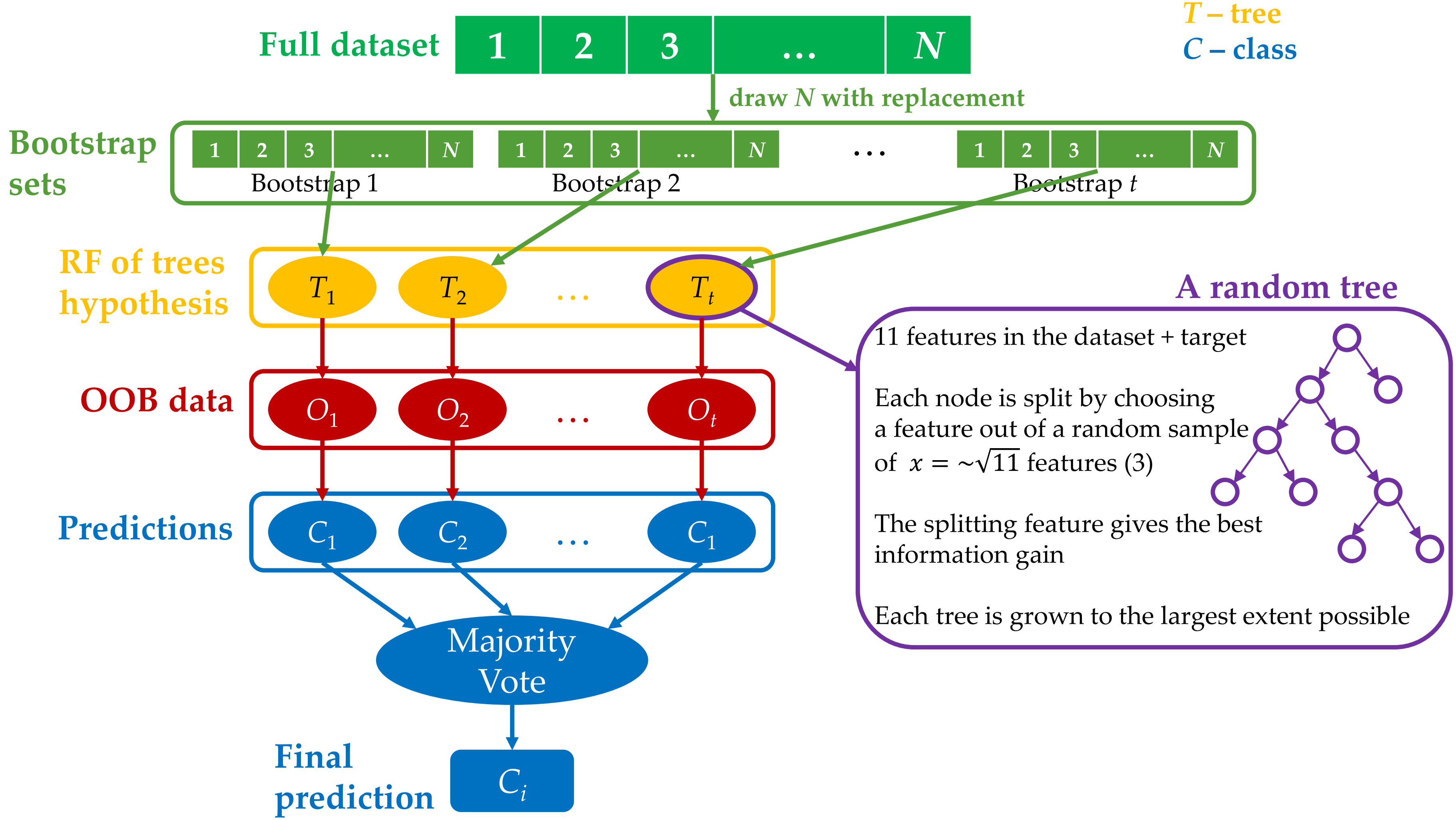
Why Random Forests?

- IBM – machine learning techniques
  - Applied to historical alert data
    - Can significantly improve classification accuracy
    - Can decrease research time for analysts
    - Can supplement analysts with additional data and insights to make better judgments
  - Very effective
    - In the elimination of white noise
    - Classification of benign data with a high degree of accuracy
      - For our framework, the benign data are the non-intruders
  - Machine learning and security are old friends

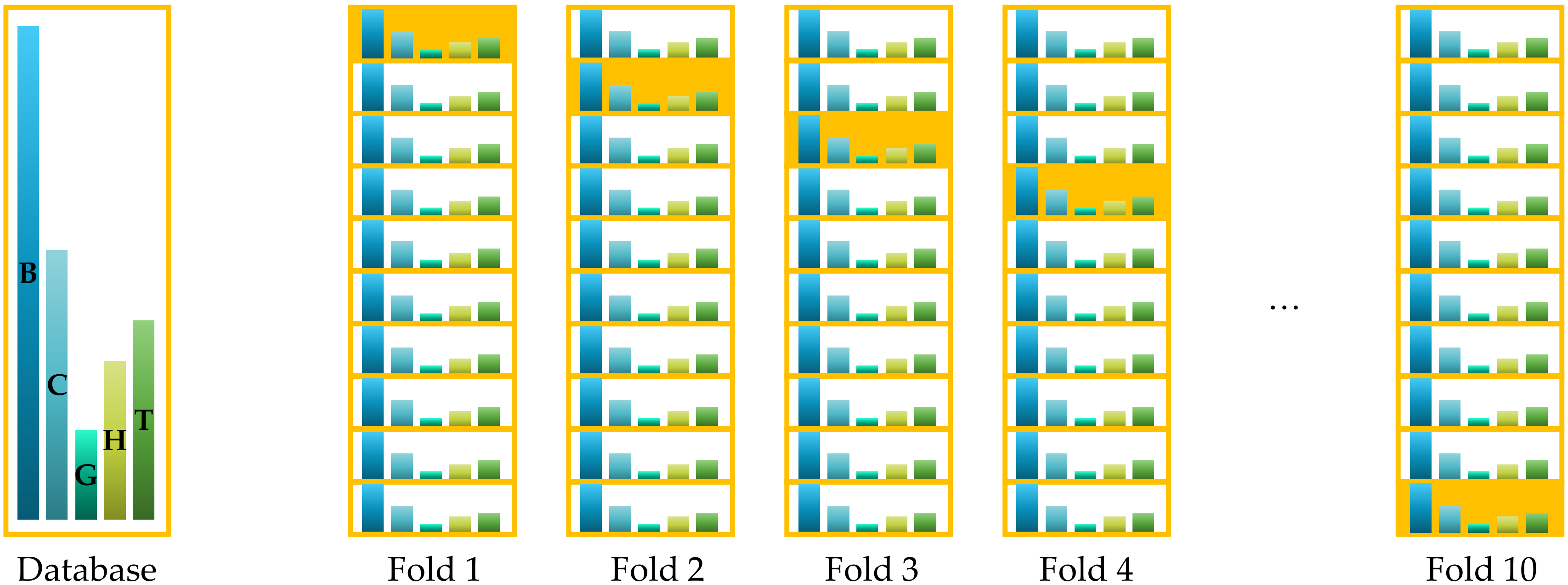
$\Rightarrow$ We should use for classification Random Forests

Random Forests

2/3rd – training
1/3rd – testing (OOB)
***O* – OOB data**
***T* – tree**
***C* – class**



Stratified 10-fold cross validation



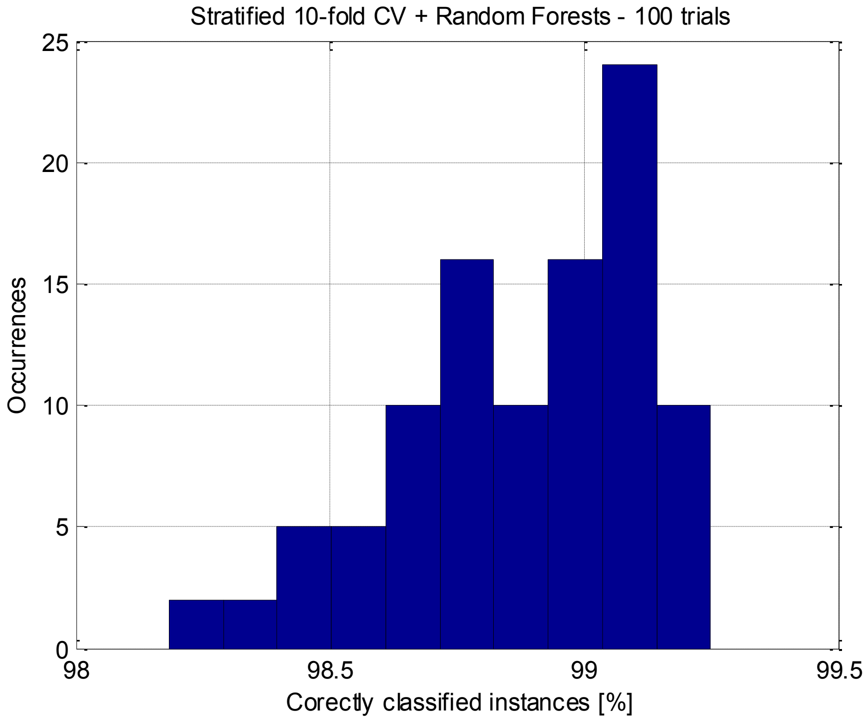
- Stratification
  - Is important for classification problems involving imbalanced datasets
  - Preserves classes distributions during training and testing
  - Reduces the estimate's variance

Results

- 49 classifiers
  - Open source software issued under the GNU General Public License
  - A collection of machine learning algorithms for data mining tasks
  - Tools for data pre-processing, classification, regression, clustering, association rules, and even visualization
- 10 times stratified 10-fold cross validation
  - 27 classifiers out of 49 average CCR >90%
  - Random Forests

Classifier	Average CCR [%] (St.Dev.)
Bagging	94.88 (1.73)
Logistic	92.77 (1.57)
Multilayer Perceptron	93.35 (1.72)
SVM (linear kernel)	97.64 (1.14)
SVM (radial basis kernel)	**98.90 (0.81)**
lazy.IBk	98.52 (0.98)
lazy.IBkLG	98.52 (0.98)
lazy.KStar	**98.70 (1.04)**
Logit Boost	92.60 (1.95)
CHIRP	92.68 (1.92)
JRip	92.75 (2.10)
PART	94.84 (1.64)
J48	94.70 (1.94)
Logistic Model Tree	96.96 (1.61)
Random Forest	**98.95 (0.91)**
Random Tree	95.97 (1.58)
REP Tree	92.13 (2.37)

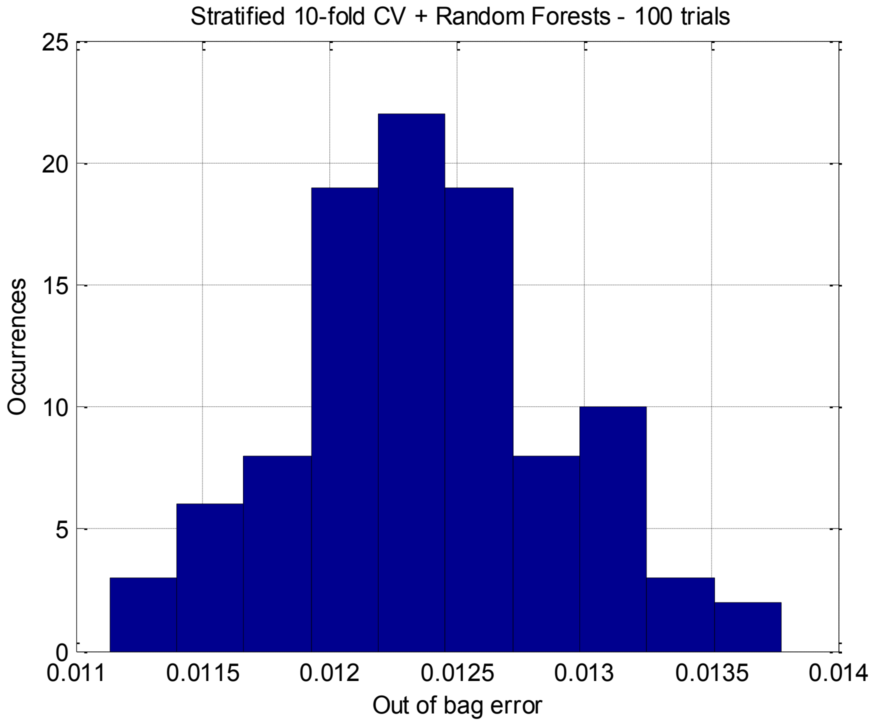
Results – Random Forests



Histogram of the correct classification rate

- 100 times stratified 10-fold cross validation
- Test phase
- Averaged CCR of each run
  - Minimum: 98.183% (frequency of apparition 1)
  - Maximum: 99.249% (frequency of apparition 10)
  - Mean value: 98.879%; Std.Dev.: 0.246

Results – Random Forests

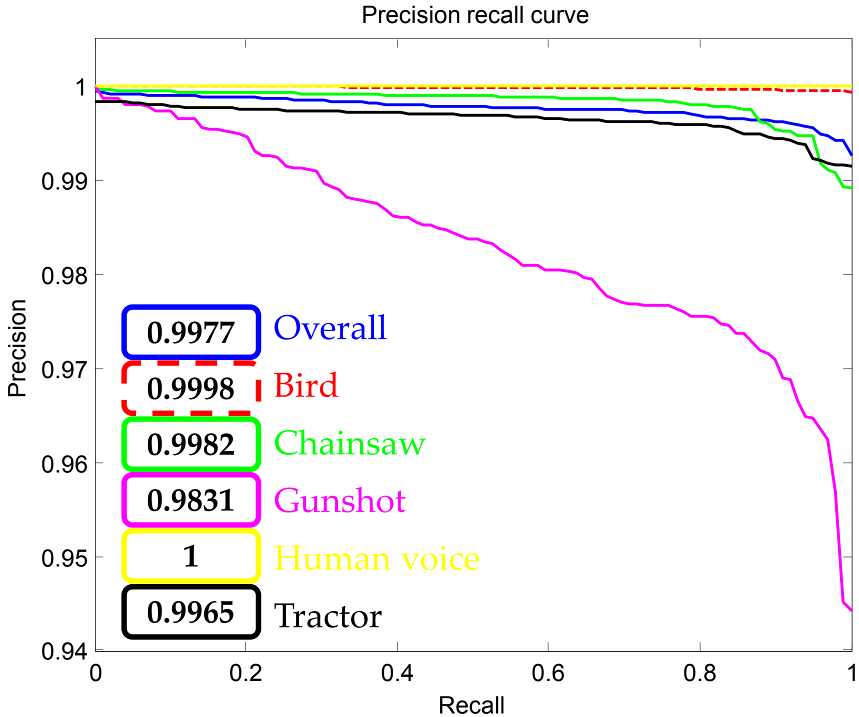


Histogram of the out-of-bag error

⇒ good model for classification

- OOB error is evaluated by computing the error rate for each class and then averaging over all classes (the misclassification probability)
- Averaged OOB of each run
  - Minimum: 0.01113
  - Maximum: 0.01378
  - Mean value: 0.01239; Std.Dev.: 0.00053

Results – Random Forests



Comparison of classes performance in PR space

- Precision vs recall curve – insensitive to classes distribution
- One-vs-all approach
  - I.e., the dotted red line labeled 'Bird' means that the positive class is the class of birds, while the negative class consists of chainsaws, gunshots, human voices and tractors
    - All five possible variations are illustrated

Results – Random Forests

Confusion matrix

No. of correctly classified instances

Prevalence (BR)

Probability of detection (TPR)

Miss rate (FNR)

Precision (PPV)

False discovery rate

Target class \ Output class	B	C	G	H	T	
B	**652** 40.83%	**2** 0.13%	**0** 0.00%	**0** 0.00%	**0** 0.00%	**99.69%** 0.31%
C	**0** 0.00%	**351** 21.98%	**1** 0.06%	**0** 0.00%	**4** 0.25%	**98.60%** 1.40%
G	**0** 0.00%	**2** 0.13%	**118** 7.39%	**0** 0.00%	**0** 0.00%	**98.33%** 1.67%
H	**0** 0.00%	**0** 0.00%	**0** 0.00%	**207** 12.96%	**0** 0.00%	**100.00%** 0.00%
T	**0** 0.00%	**0** 0.00%	**3** 0.19%	**0** 0.00%	**257** 16.09%	**98.85%** 1.15%
	100.00% 0.00%	**98.87%** 1.13%	**96.72%** 3.28%	**100.00%** 0.00%	**98.47%** 1.53%	**99.25%** **0.75%**

- 12 audio signals out of 1 597 are misclassified

	FAR	FOR
B	0.00%	0.21%
C	0.32%	0.40%
G	0.27%	0.14%
H	0.00%	0.00%
T	0.30%	0.22%

Conclusion

- A model for audio signal classification: LPC + RF
- The signals under classification belong to the class of sounds from WID applications
- The step by step model building was illustrated
- Evaluation of the proposed classification system: 100 x stratified 10-fold CV
- Multiclass classification – average CCR: 99.25%
  - There is no probability of false alarms: birds + human voices
  - For the other three classes the probability is low (~0.3%)
  - The false omission rate is also low: ~0.2% for birds and tractors, a little bit higher for chainsaws (0.4%), lower for gunshots (0.14%) and zero for human voices

⇒ Proposed audio classification system can be used as a good detection system, i.e. for WID problems

Audio Signal Classification Using Linear Predictive Coding and Random Forests

Lăcrimioara GRAMA, Corneliu RUSU

Technical University of Cluj-Napoca
Faculty of Electronics, Telecommunications and Information Technology
Basis of Electronics Department
Signal Processing Group

SpeD 2017
Bucharest, Romania
July 6-9, 2017