ENRAM
EUROPEAN NETWORK FOR THE RADAR SURVEILLANCE OF ANIMAL MOVEMENT

vogelwarte.ch

TOBB UNIVERSITY OF ECONOMICS AND TECHNOLOGY

IDENTIFICATION OF BIOLOGICAL SIGNALS WITH RADAR

SHORT TERM SCIENTIFIC MISSION REPORT

Host Scientist / Institution:

- Felix Liechti, Swiss Ornithological Institute, Sempach, Switzerland

Visiting Scientist / Institution:

- Sevgi Zübeyde Gürbüz, TOBB University of Economics and Technology, Ankara, Turkey
- Also accompanied by a student, Mehmet Saygın Seyfioğlu

Research Visit Period: 18 October – 24 October 2015

Introduction

The main goal of this joint work is to develop machine learning algorithms for the automatic classification of bird data as collected by the BirdScan Radar developed by the Swiss Ornithological Institute (Vogelwarte). Because validation of any algorithms developed requires data that has been accompanied by observations of birds by expert ornithologists, a database of measurements recorded from at Grenchenberg, which included a high percentage of observations, was used to develop and test algorithms. The Grenchenberg data set consists of the following ornithologist validated measurements:

- Insects : 482
- Birds : 213
- Waders : 68
- Waders (big) : 8
- Passerines : 328
- Passerines (big) : 7
- Flock (small) : 2
- Flock (passerine) : 1
- Other : 58

To ensure that the classes are more equally distributed in numbers, this data was re-partitioned into the following classes:

- Insects : 482
- Birds : 213
- Passerines : 336
- Waders : 77
- Other : 59

The similarities and differences between examples of signatures from the main biological classes may be seen in Figure 1. A number of important observations may be made. First, at initial glance the insect and wader signatures may appear to be highly similar as they both exhibit a downward concavity; however, the frequency content of the oscillations superimposed upon this trend is visually different. Insect oscillation frequencies are higher than that of the oscillations caused by the wing flapping of waders. A similar difference in frequency may be observed between birds and passerines. Not unexpectedly, therefore, the wing flapping frequency and derivatives relating to flapping pattern are critical for discriminating differing classes of biological targets.

Features for Class Discrimination

In particular, two types of features were extracted to aide in class discrimination: physical features and transform based features. Physical features refer to those parameters that may have physical or biological significance; they are features that humans may easily relate to the flight characteristics of the birds or insects. Physical features include:

- Distance (d)
- Wing flapping frequency (wff)
- The derivative of the wing flapping frequency (wff_2)
- The number of times there is a period of pausing or no flapping (nPause)
- The average duration of flapping (Avg_pulseL)

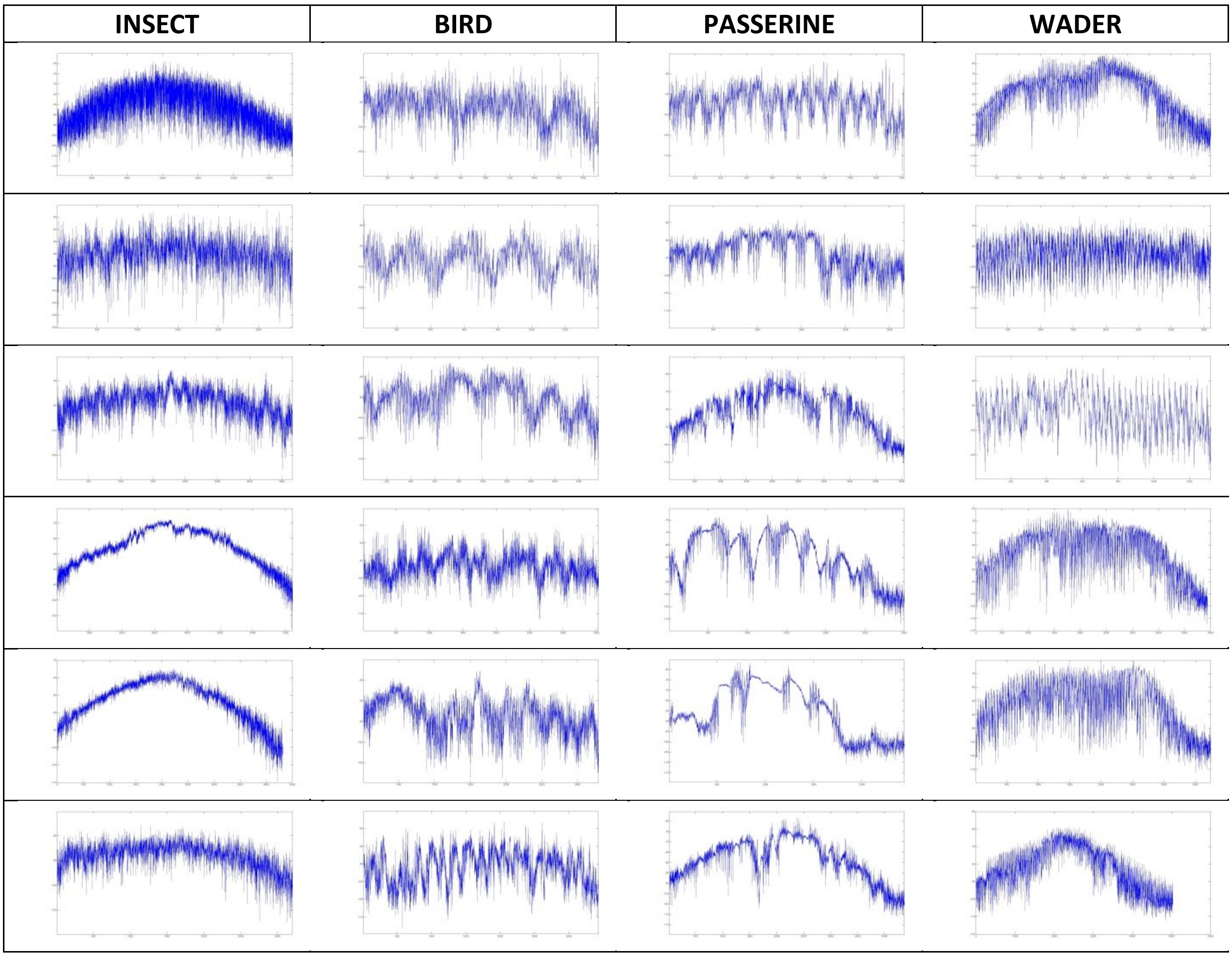

Figure 1. Example signatures for the main biological classes.

- The average duration of pausing (Avg_pauseL)
- The standard deviation of the duration of flapping (Dev_pulseL)
- The standard deviation of the duration of pausing (Dev_pauseL)
- The ratio of the flapping duration to the pausing duration (PulsePauseR)
- The maximum level of the signal (pegel)
- Polarization ratio
- Radar cross section (RCS)
- Square root of the RCS

Physical features, while intuitive and directly relevent to the classification problem, have the disadvantage of now always being easily estimable. For example, how accurately can the pulsing and pausing periods really be distinguished? Based on these estimates, how accurately can wing flapping frequency be estimated?

Indeed, within the Grenchenberg database there were many instances when wing flapping frequency or other parameters could not be estimated, and where hence simply empty boxes within the database, as shown in Figure 2. However, the classifier needs to have all features

	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	459,77	6,45	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,879706	10,5043	1	NULL	0,15125
2	462,663	559,2	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,157198	-78,1926	1	NULL	0,011535
3	462,62	317,7	NULL	NULL	18,184	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,413228	-95,2427	1	NULL	2,37E-0
4	462,577	154,95	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	1,00099	-95,1303	1	NULL	1,38E-0
5	462,62	109,2	NULL	NULL	9,21626	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,0759503	-69,1135	1	NULL	0,000135
6	462,577	379,2	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,584038	-92,1965	1	NULL	9,70E-0
7	462,577	154,2	NULL	NULL	15,359	NULL	0,8	5	0,179429	0,352373	0,344755	0,476802	1,96386	0,204214	-90,3705	1	NULL	4,04E-0
8	462,577	410,7	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,563451	-96,1398	1	NULL	5,38E-0
9	462,62	482,7	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,446214	-96,1178	1	NULL	0,0001032
10	462,62	71,7	NULL	NULL	NULL	15,3604	0,8	0	NULL	NULL	NULL	NULL	NULL	0,169661	-93,411	1	NULL	9,38E-0
11	462,577	104,7	NULL	NULL	10,5405	10,3899	0,8	0	NULL	NULL	NULL	NULL	NULL	0,0375043	-77,3592	1	NULL	1,72E-0
12	462,62	176,7	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,746886	-92,5126	1	NULL	4,25E-0
13	462,62	88,2	NULL	NULL	NULL	47,4366	0,8	1	NULL	NULL	NULL	NULL	NULL	0,423074	-75,845	1	NULL	1,23E-0
14	462,577	92,7	NULL	NULL	10,5405	10,8417	0,8	1	NULL	NULL	NULL	NULL	NULL	0,0679847	-82,2576	1	NULL	3,42E-0
15	459,77	74,7	NULL	NULL	13,7692	NULL	0,8	14	0,11745	0,358875	0,0785674	0,190943	3,05556	0,285791	-70,1178	1	NULL	2,36E-0
16	462,577	101,7	NULL	NULL	9,93819	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,275571	-68,2573	1	NULL	0,0001243
17	462,577	59,7	NULL	NULL	NULL	17,6177	0,8	0	NULL	NULL	NULL	NULL	NULL	0,0195047	-82,958	1	NULL	5,00E-0
18	459,77	92,7	NULL	NULL	15,8944	NULL	0,8	6	0,4089	0,41325	0,16443	0,252577	1,01064	0,0967451	-78,2703	1	NULL	8,56E-0
19	462,577	566,7	NULL	NULL	10,0888	10,0888	0,8	0	NULL	NULL	NULL	NULL	NULL	0,252546	-95,4526	1	NULL	0,0002286
20	459,728	73,2	NULL	NULL	17,5092	NULL	0,8	5	0,374134	0,282776	0,426207	0,560663	0,755814	0,193281	-74,7095	1	NULL	7,55E-0
21	462,62	143,7	NULL	NULL	17,544	NULL	0,8	11	0,207514	0,436643	0,0894864	0,099553	2,10417	0,0924655	-73,723	1	NULL	0,0001407
22	462,62	368,7	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,288095	-93,7912	1	NULL	6,01E-0
23	462,62	76,2	NULL	NULL	19,065	NULL	0,8	7	0,166443	0,477714	0,140836	0,146399	2,87013	0,18147	-69,5257	1	NULL	2,93E-0
24	462,577	97,2	NULL	NULL	NULL	18,9729	0,8	1	NULL	NULL	NULL	NULL	NULL	0,0600929	-68,3535	1	NULL	0,0001014
25	462,62	68,7	NULL	NULL	18,5229	NULL	0,8	9	0,166443	0,350179	0,358451	0,332979	2,1039	0,268024	-62,5406	1	NULL	9,66E-0
26	462,577	283,2	NULL	NULL	15,9613	NULL	0,8	5	NULL	NULL	NULL	NULL	NULL	0,336053	-89,1609	1	NULL	6,07E-0
27	459,728	94,2	NULL	NULL	NULL	NULL	0,8	8	0,174016	0,596005	0,223257	0,0739895	3,425	0,670316	-67,9647	1	NULL	9,79E-0
28	459,77	61,2	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,0898915	-82,3416	1	NULL	6,37E-0
29	459,77	89,7	NULL	NULL	NULL	NULL	0,8	5	0,30885	0,2871	0,283533	0,758552	0,929577	0,0940964	-83,7792	1	NULL	2,11E-0
30	462,62	73,2	NULL	NULL	17,6193	NULL	0,8	15	0,16212	0,261554	0,101893	0,126422	1,61333	0,131376	-64,6493	1	NULL	7,66E-0
31	462,62	137,7	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,493941	-80,6588	1	NULL	2,40E-0
32	462,577	158,7	NULL	NULL	9,30576	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,168844	-88,4377	1	NULL	7,07E-0
33	462,577	548,7	NULL	NULL	NULL	10,8417	0,8	0	NULL	NULL	NULL	NULL	NULL	0,192301	-94,7802	1	NULL	0,0002346
34	462,577	163,2	NULL	NULL	NULL	NULL	0,8	1	NULL	NULL	NULL	NULL	NULL	0,189934	-91,3768	1	NULL	4,02E-0
35	462,577	181,2	NULL	NULL	NULL	NULL	0,8	0	NULL	NULL	NULL	NULL	NULL	0,970374	-92,6025	1	NULL	4,61E-0

Figure 2. Snap shot of database showing features with NULL entries.

extracted in all cases to successfully discriminate the data: some number needs to be filled in; it can't just be left blank. Two approaches for dealing with null features were tried:

1. Assign a numberical value to the NULL that is clearly different from all possible values, i.e. assign an outlier. In this option, the value of 8888 was replaced for all NULL entries.

2. Assign a mean value of the class based upon other measurements to the NULL entries. Thus, however many, for example, passerine data had a measurement of wing flapping frequency, these measurements would be averaged and used in place of NULL values for passerine wing flapping frequency.

Each of these approaches has advantages and disadvantages. Assigning an outlier gives a numerical distinction representative of that feature not being attainable; however, it also introduces dependences that are not normally present in the data when varying features both are recorded NULL values. Assigning the mean class value to a NULL prevents statistical distortion of results, and precludes introduction of abnormal inter-class dependencies, but, it is dependent upon the accuracy of human classification results. And since the human observations serve as "ground truth" to test the algorithms, average may actually lead to over-optimistic results by improving the accuracy of feature value.

In addition to physical features, three types of transform based features were also extracted:

- 1st – 4th Cepstrum Coefficients
- 1st – 4th Linear Predictive Coding Coefficients
- 1st – 5th Discrete Cosine Coefficients

The reason for including these features is to have a certain number of features that are guaranteed NOT to have any NULL values. Moreover, these features have been found in the literature to yield good discrimination results in processing time-frequency distributions attained in fields such as speech processing and micro-Doppler based human activity recognition. Thus, in this work, we wanted to determine whether they could also be useful in discriminating biological signals. Next, the technical definitions of these features are described.

Cepstrum Coefficients: The cepstrum, *c[n],* is defined as the inverse Discrete Fourier Transform (DFT) of the log magnitude of the DFT of the input, *x[n]:*

$$c[n] = \Im^{-1}\{\log|\Im\{x[n]\}|\}\,, \tag{1}$$

where $\Im\{\cdot\}$ is the Fourier transform. Any number of the coefficients of the cepstrum may be extracted as features. As with the Fourier Transform, the cepstrum contains harmonic information; however, the log spectrum enables compression of the dynamic range and thus reduction of amplitude differences in the harmonics.

Linear Predictive Coding Coefficients: LPC's are typically computed from the I/Q output of the radar, *x[n]*, by representing this signal as the linear combination of past values:

$$\hat{x}[n] = \sum_{k=1}^{p} a[k]x[n-k]\,, \tag{2}$$

where *a[k]* are the LPC's and *p* is the total number of LPC's. To compute the LPC's, the difference between the model in (2) and the true signal – the error, $e[n] = x[n] - \hat{x}[n]$ – is sought to be minimized. Many methods may be employed for this minimization, such as computing the autocorrelation followed by a Levinson-Durbin recursion.

Discrete Cosine Coefficients: The DCT is computed for a one-dimensional sequence of length N as

$$C(u) = \alpha(u)\sum_{k=0}^{N-1} f(k)cos\left[\frac{\pi(2k+1)u}{2N}\right], \tag{3}$$

for $\mathrm{u} = 0,1,2,\ldots,\mathrm{N}-1$. The coefficients of this transform may be used as features for classification.

Feature Ranking with Mutual Information[1]

Feature ranking is accomplished by assessing the degree of relevancy of a feature to a given classification problem. This is done by first quantifying the information content of features. Formally, the entropy of a discrete random variable is defined as

$$H(X) = -\sum_{x} p(x) \cdot \log_2 p(x) \tag{4}$$

The entropy of a random variable is the measure of uncertainty about that random variable. The mutual information between two discrete random variables is defined as

$$I(X;Y) = \sum_{x}\sum_{y} p(x,y) \cdot \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \tag{5}$$

where *p(x,y)* is the joint probability mass function (PMF), and *p(x)* and *p(y)* are the marginal PMFs of the random variables *X* and *Y*, respectively. Mutual information between two random variables is the measure of information one random variable gives about the other. Random variables that are independent of each other have zero mutual information, while mutual information reaches its maximum value when the two random variables are fully correlated.

The concept of mutual information can be used to order features based on their contribution of information. Let's define f_1 and f_2 as random variables referring two features, while *C* is a random variable referring to class. Then, the mutual information between f_1, f_2 and *C* is defined as $I(f_1, f_2; C)$ and can be written in two ways:

$$I(f_1, f_2; C) = I(f_1; C) + I(f_2; C | f_1) \tag{6}$$

$$I(f_1, f_2; C) = I(f_2; C) + I(f_1; C | f_2) \tag{7}$$

where, *I(f1;C)* and *I(f2;C)* represent the mutual information between the class variable *C* and the first feature f_1 and second feature f_2, respectively. The term *I(f2;C|f1)* represents the mutual information between the second feature and *C*, given the first feature while *I(f1;C|f2)* is defined similarly. Then, to determine which of the two features (f_1 or f_2) provides more information about the class variable *C*, *I(f1;C)* and *I(f2;C)* are compared and the feature that results in a larger value is selected. Let the set of all features be denoted by F, for F = $\{f_1,f_2,\ldots,f_N\}$, where f_i $(i=1,2,\ldots,N)$ indicate individual features.

Then, to find a *M*-element subset S of F, S = $\{f_1',f_2',\ldots,f_M'\}$, such that *I(S;C)* is the largest among all such subsets, the above approach may be generalized by comparing all *I(S;C)* for all *M*-element subsets.

Ideally, computing the conditional and joint probability mass functions used to calculate *I(S;C)* requires the calculation of multi-dimensional histograms. However, due to the "curse of dimensionality" approaches that yield high performance for low volumes of data become

[1] Exerpted from prior relevent work by Dr. Gürbüz, namely: B. Tekeli, S.Z. Gürbüz, and M. Yüksel, "Information theoretic features selection for human micro-Doppler classification," IEEE Transactions on Geoscience and Remote Sensing, Vol. 54, Iss. 5, pp. 2749 – 2762.

increasingly unworkable as the dimensionality of the data (number of features and classes) increases. Therefore, practical implementation of $I(S;C)$ requires reducing the dimensionality by either eliminating features with low information content or high redundancy with respect to other features for the classification problem at hand. Much research has been done on how to optimally compute mutual information with the minimum computational load.

The Mutual Information Based Feature Selection (MIFS) Algorithm proposed by Battiti in 1994 implements greedy selection of features to arrive at a more computationally tractable approach to mutual information calculations. In MIFS, features are selected one by one. In each iteration, the feature that maximizes

$$\max\left(I(C;f_i)-\beta\sum_{f_s\in S}I(f_s;f_i)\right) \tag{8}$$

is sought. The first term in (8) is the mutual information between feature f_i in F and the class variable C, and $I(f_s;f_i)$ is the mutual information between f_i and the already selected feature f_s in the selected feature set S.

Initially, the set of selected features is empty. Therefore, in the first round of the algorithm, simply, the feature f_i' in F that maximizes $I(C;f_i)$ is calculated. Then f_i' is excluded from F, and is included in S. In the second round, the feature in F that maximizes $I(C;f_i)-\beta\sum_{f_s\in S}I(f_s;f_i)$ is searched. The algorithm repeats itself until all k features are selected. The computation in (7) accounts for selecting a feature that is highly informative about the class variable, but at the same time, is not very similar to previously selected features.

In MIFS, the parameter β controls the relative importance of relevance and redundancy. If $\beta=0$, the algorithm chooses features that more or less provide the same information. As β increases, maximization of the expression in (8) requires that the features selected in each round of the algorithm should be such that they are increasingly independent from each other ($I(f_s;f_i)$ should be small). Moreover, for large β, the relationship between the selected features and the class variable, as given by $I(C;f_i)$, is valued less, so that the relationship between features has a more significant effect on the expression in (8), which is maximized during feature selection. This is a problem because we would like to find the minimal feature set that has the greatest relevance to the class variable, not simply a set of independent features. To alleviate this problem, Kwak and Choi suggest an adjustment to MIFS as follows

$$\max\left(I(C;f_i)-\beta\sum_{f_s\in S}\frac{I(C;f_s)}{H(f_s)}I(f_s;f_i)\right) \tag{9}$$

Here, $H(f_s)$ is the entropy of a feature in the selected feature set S. As $I(C;f_s) \leq H(f_s)$, the ratio in (9) indicates the importance of f_s in S, and this ratio is used as a coefficient for $I(f_s;f_i)$. Thus, f_i needs to be less relevant to more informative features; yet, redundancy is tolerable for less informative features. Moreover, at the beginning of the selection process, relevance to the class variable is ensured, while as the selected feature set grows, the new selections are required to be increasingly less redundant. In this way, the ratio $I(C;f_s)$ / $H(f_s)$ facilitates efficient de-selection of features.

Classification Performance Achieved With All Features

Once all 25 features were extracted, these were supplied to four different types of classifiers:
1. Naive Bayesian
2. Multi-Class Support Vector Machine (SVM)
3. Multilayer Perceptron
4. Random Forest

Results were obtained for both possible ways of dealing with NULL values. These results are now presented in turn.

REPLACE NULL WITH 8888 & USE NAIVE BAYESIAN CLASSIFIER

- Correctly Classified Instances 480 68.5714 %
- Incorrectly Classified Instances 220 31.4286 %

Table 1. Confusion matrix for Naive Bayesian classifier with 8888 replacing nulls.

	0	**1**	**4000**	**5000**	**9999**
0	**281**	0	3	0	1
1	0	**57**	18	0	52
4000	0	75	**69**	0	55
5000	0	0	0	**51**	0
9999	0	9	6	1	**21**

REPLACE NULL WITH 8888 & USE MULTI-CLASS SVM CLASSIFIER

- Correctly Classified Instances 566 80.8571 %
- Incorrectly Classified Instances 134 19.1429 %

Table 2. Confusion matrix for Multi-Class SVM classifier with 8888 replacing nulls.

	0	**1**	**4000**	**5000**	**9999**
0	**285**	0	0	0	0
1	0	**63**	64	0	0
4000	2	31	**166**	0	0
5000	0	0	0	**52**	0
9999	1	22	13	1	**0**

REPLACE NULL WITH 8888 & USE MULTILAYER PERCEPTRON CLASSIFIER

- Correctly Classified Instances 581 83 %
- Incorrectly Classified Instances 119 17 %

Table 3. Confusion matrix for Multilayer Perceptron classifier with 8888 replacing nulls.

	0	1	4000	5000	9999
0	**285**	0	0	0	0
1	0	**76**	47	0	4
4000	2	36	**160**	0	1
5000	0	0	0	**52**	0
9999	1	18	10	0	**8**

REPLACE NULL WITH 8888 & USE RANDOM FOREST CLASSIFIER

- Correctly Classified Instances 693 99 %
- Incorrectly Classified Instances 7 1 %

Table 4. Confusion matrix for Random Forest classifier with 8888 replacing nulls.

	0	1	4000	5000	9999
0	**285**	0	0	0	0
1	0	**125**	2	0	0
4000	0	1	**198**	0	0
5000	0	0	0	**51**	0
9999	0	0	3	1	**33**

This same procedure was then repeated for the case when the NULL values were replaced with the class average. A summary of results is provided in Table 5.

Table 5. Summary of classification results when all 25 features are utilized.

%	NULL → 8888	NULL → Average
Naive Bayesian	69	90
Multi-Class SVM	81	91
Multilayer Perceptron	83	93
Random Forests	99	98

Classification Performance versus Number of Features

As mentioned in the section on ranking features, usually using all possible features does not result in the optimal classification performance due to the "curse of dimensionality." In this work, one of our goals was not just to determine the best possible classifier, but also to ascertain the minimum number of required features and which features were the most critical for class discrimination. To accomplish this goal, the mutual information metric was first used to rank all features according to their importance. Below, each of the features is listed by order of the importance ranking as given by their mutual information.

Ranking of Features:

1. Distance
2. 2nd LPC Coefficient
3. Number of Pauses
4. Pegel
5. 2nd Cepstrum Coefficient
6. Wing Flapping Frequency
7. 3rd Cepstrum Coefficient
8. 1st Cepstrum Coefficient
9. Average Pulse Length
10. Average Pause Length
11. Standard Deviation of the Pulse Length
12. Standard Deviation of the Pause Length
13. Ratio of Pulse to Pause Length
14. Derivative of Wing Flapping Frequency
15. RCS
16. Polarization Ratio
17. Square Root of the RCS
18. 3rd DCT Coefficient
19. 1st DCT Coefficient
20. 2nd DCT Coefficient
21. 4th DCT Coefficient
22. 4th LPC Coefficient
23. 1st LPC Coefficient
24. 3rd LPC Coefficient
25. 4th Cepstrum Coefficient

In regards to this ranking, a number of important comments and observations should be made. First, this is not an absolute ranking; use of a different metric would result in a different ranking. Furthermore, results could change as additional data is included in the study. Second, while some parameters seem statistically significant, they are known to be more or less independent of the biological class, e.g. distance. Third, as expected several physical features, especially, wing flapping frequency scored very high; happily, some of the newly explored features, such as the 2nd LPC coefficient and cepstral coefficients seem to be hopeful as novel features that could improve discriminitivity.

Table 6 summarizes the performance variation as a function of feature number, when the above listed importance ranking is utilized. In other words, when just 24 features are used, the 4th Cepstral Coefficient (ranked 25) is discarded from the feature set. Figure 3 graphically shows the dependency of classification performance on number of features used. Notice that beyond 5-8 features, comparably little performance improvement is achieved.

Table 6. Classification performance versus number of features.

%	CLASSES			
# of features	**Naive Bayesian**	**Multi-Class SVM**	**Multilayer Perceptron**	**Random Forest**
25	69.14	81	82.43	99.14
24	69.14	81.14	84	99.14
23	69.14	81.14	81.86	98.86
22	69.71	81.14	82.43	99
21	68.86	81.43	84.43	99.29
20	68.71	81.14	84.86	99.14
19	68.86	80.71	84.14	98.86
18	68.29	80.57	84.57	99.14
17	68.29	81.14	82.86	99.14
16	68.29	80.86	81.57	99.14
15	68.29	80.86	82.57	99.14
14	68.14	80.86	81.86	99.43
13	70	80.86	81.14	97.14
12	70	80.86	81.86	97.57
11	70	81	81.29	97
10	70	80.86	81.43	98.14
9	69.86	80.86	82.14	98.57
8	62.71	73.43	76.43	94
7	62.71	73.43	76.43	92.86
6	62.71	74	76	94
5	59.14	62	67.86	66.86
4	49.71	53.57	56	53.54
3	50.14	48.29	50.26	45.14
2	46.57	40.71	46.86	39.57

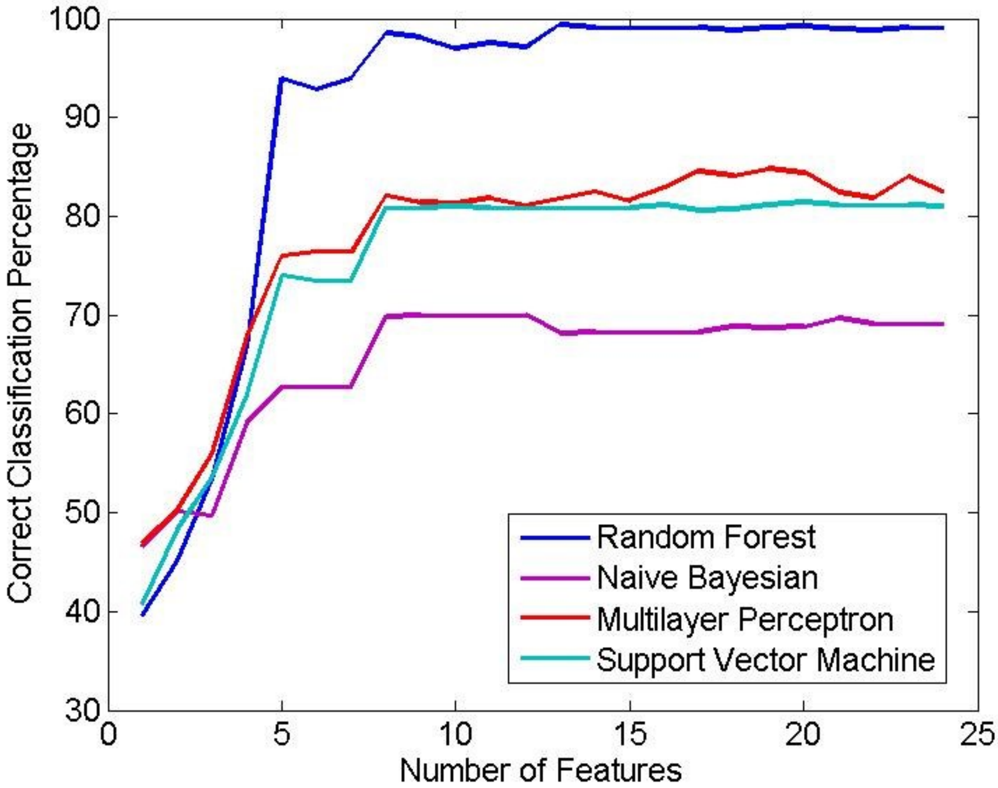

Figure 3. Graphic showing dependence of classification performance of number of features.

Classification Performance for a Single Feature

The significance of single features may also be assessed by looking at the impact of classification directly, as opposed to relying on a independent metric. In this case, the results are dependent upon the classifier. Note that the wing flapping frequency (wff) and average duration of flapping (avgPulseL) distinguish themselves as critical features when using a random forest classifier.

Table 7. Single feature classification results.

%	distance	2nd lpc	npause	wff	2nd cepstrum	avgPulseL
Random Forest	39.29	33.14	64.29	**93.14**	36.26	**91**
Naive Bayesian	45.14	46.86	61	58.86	40.71	48.14
Multilayer Perceptron	47	48.29	64.14	68.43	40.71	48.14
Multi-Class SVM	40.71	46.57	62.14	68.43	40.71	48.14

Conclusions

From these studies a number of important conclusions may be made:

- Random forest appears to be a superior classifier for biological signal classification.

- Wing flapping frequency and related time parameters pertinent to durations of flapping and gliding are important. Algorithms that could be developed to improve estimates of these parameters would contribute significantly to improving classification performance.

- Transform based coefficients, such as the 2nd LPC coefficient and cepstral coefficients, at least at first sight, appear to have the potential to improve biological signal classification.

- Carefully selecting a handful of features is more critical than applying by brute force a large number of features.

Acknowledgements

We would like to acknowledge the contributions and assistance of all the wonderful scientists at the Swiss Ornithological Institute, especially Felix Liechti and Baptiste Schmidt, as well as Dominik Kleger from the Applied University of Zurich (ZWAW), School of Ingenieurs.

Emulating Nature for Better Engineering