Signal Analysis with Application of k-Nearest Neighbors Method

Ivelina Balabanova[1], Georgi Georgiev[2] and Stela Kostadinova[3]

Abstract – **The paper presents the results of synthesizing k-Nearest Neighbors (k-NN) models for identification the presence of noises. Signals with Gaussian White and Periodic Random noises are analyzed. The models are evaluated by resubstitution and cross-validation procedures. Classifiers at Euclidean, Minkowski, Cityblock and Chebychev metric distances are examined. Classification model with 97.375% accuracy, parameter k = 5 and Minkowski distance is selected.**

Keywords – **noise identification, k-NN models, metric distance, resubstitution, cross-validation.**

I. Introduction

The wide use of the k-Nearest Neighbors (k-NN) method in the processing of biomedical signals is confirmed by a number of studies.

The method is particularly useful in assessing the quality levels of ECG data from wireless sensor systems for continuous monitoring the conditions of patients, where the possibility of interference from the human body movement is especially large [1].

The k-NN is used for diagnosing the condition of the human cognitive system, for detecting disturbances in the normal brain function as well as exploration of the human motor system on the basis of EEG signals [2-4].

Voice recognition in background noise environment or speech segmentation for identification of the minimal degraded speech fragments in the transmission through a co-channel interface, are investigated in some studies [5, 6].

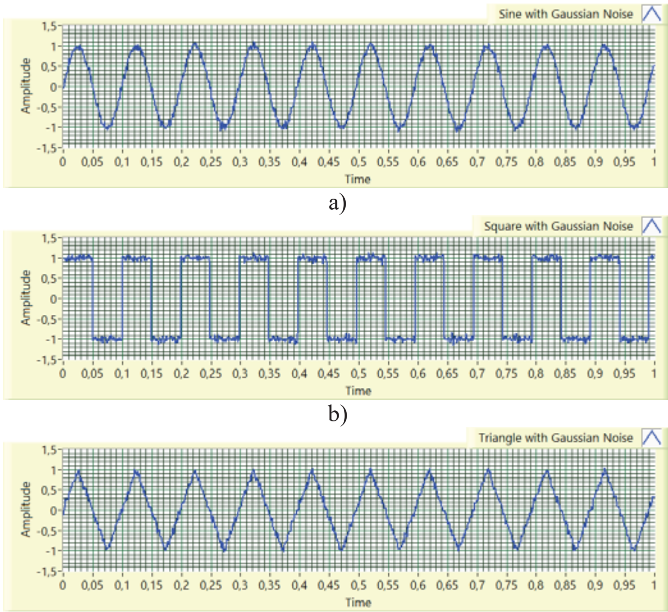
There is interest in detecting anomalies in the echo distribution of meteorological radar data, as well as the classification of objects based on 2D scans of separate radar routes [7-9].

In the report applicability and method efficiency are studied for identification of electrical signals with Gaussian White Noise (GWN) and Periodic Random Noise (PRN).

II. Experimental data

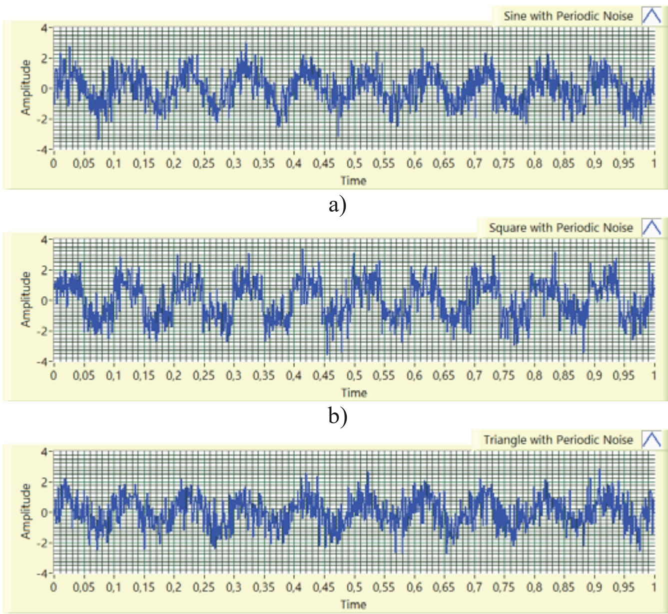
Sine, square and triangle waveforms with GWN and PRN are simulated in LabVIEW environment. The simulation is conducted at an amplitude level "one unit" and a frequency of the signals "10.1 Hz", as well as specified noise parameters - "0.05" for Standard deviation and Spectral amplitude, respectively for GWN and PRN.

The taken oscillograms are shown on Fig. 1 and Fig. 2.



a) b) c)

Fig. 1. Sine a), square b) and triangle c) waveforms with GWN



a) b) c)

Fig. 2. Sine a), square b) and triangle c) waveforms with PRN

[1]Ivelina Balabanova is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 Hadji Dimitar Str., Gabrovo 5300, Bulgaria, E-mail: ivstoeva@abv.bg.

[2]Georgi Georgiev is with the Faculty of Electrical Engineering and Electronics at Technical University of Gabrovo, 4 Hadji Dimitar Str., Gabrovo 5300, Bulgaria.

[3]Stela Kostadinova is with the Faculty of Computer Sciences and Automation at Technical University of Varna, 1 Studentska Str., Varna 9000, Bulgaria, E-mail: stela.kostadinova@tu-varna.bg.

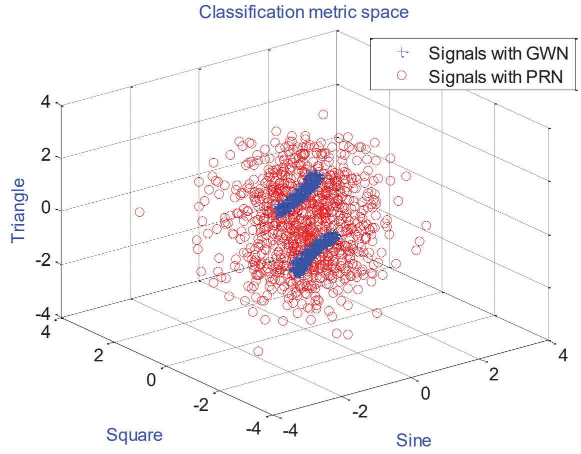


Fig. 3. Sine a), square b) and triangle c) waveforms with PRN

Sample from 2000 information samples for GWN (Class №1) and PRN (Class №2) signals defined as target identification groups is acquired based on data from experimental signals. Graphical interpretation of the investigated signals as points in a three-dimensional classification space is given in Fig. 3.

III. SYNTHESIS OF THE MODELS FOR NOISE IDENTIFICATION BY K-NN METHOD

Classification models based on the k-Nearest Neighbors method are created for the following metric distances:

- Euclidean (№1);
- Minkowski (№2);
- Cityblock (№3);
- Chebychev (№4).

The study is performed with a successive increase of k-neighbors (points in the classification space) in the range of 5 to 100 neighbors. Key indicators defining the quality of classifiers are the errors in the technical approaches "resubstitution" and "cross-validation".

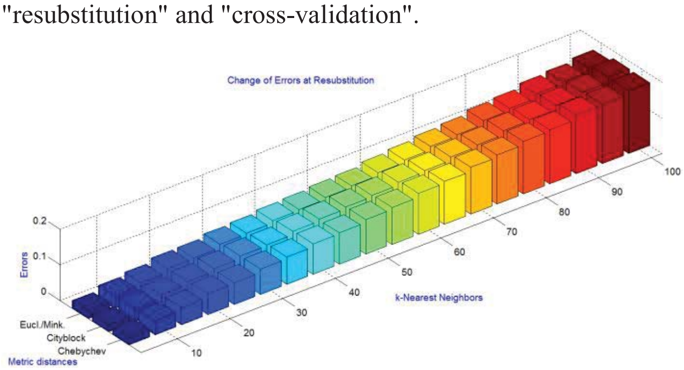


Fig. 4. Resubstitution errors at a) Euclidean/Minkowski, b) Cityblock and c) Chebychev distances

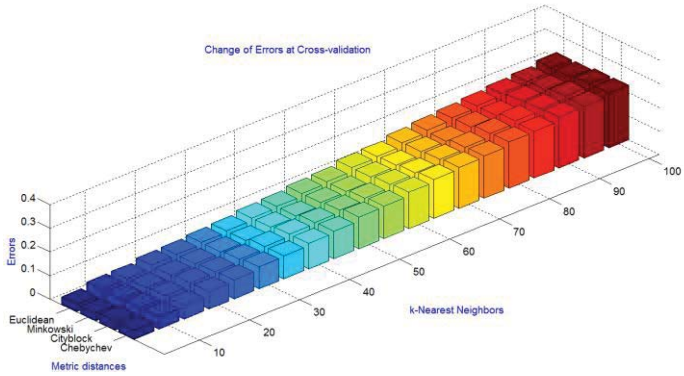


Fig. 5. Cross-validation errors at a) Euclidean/Minkowski, b) Cityblock and c) Chebychev distances

3D representations of the resulting errors of model analysis are shown in Fig. 4 and Fig. 5. The following variational limits were obtained:

For resubstitution:

- Euclidean / Minkowski: from 0.0210 to 0.1665;
- Cityblock: from 0.0195 to 0.1785;
- Chebychev: from 0.0215 to 0.1775.

at cross-validation:

- Euclidean: from 0.0325 to 0.2225;
- Minkowski: from 0.0315 to 0.2225;
- Cityblock: from 0.0375 to 0.2310;
- Chebychev: from 0.0365 to 0.2315.

Their accuracy levels are calculated according to faults (error "0" is equal to 100.00 and "1" to zero precisions).The accuracy results for signal identification are contained in Table 1 and Table 2.

TABLE I
ACCURACIES AT RESUBSTITUTION, %

k-neighbors	Distances №1 and №2	Distance №3	Distance №4
5	97.900	98.050	97.850
10	96.450	96.300	95.950
15	95.550	95.350	95.250
20	94.400	94.350	94.150
25	94.200	93.600	93.800
30	93.300	92.800	92.800
35	92.900	92.600	92.200
40	92.200	91.600	91.350
45	91.950	90.950	91.150
50	91.400	90.400	90.250
55	91.000	89.900	89.600
60	90.100	89.100	88.450
65	89.350	88.600	88.000
70	88.400	87.650	86.850
75	87.900	87.000	86.250
80	86.750	85.900	85.200
85	86.150	85.550	84.700
90	84.900	84.550	83.350
95	84.350	83.550	83.250
100	83.350	82.150	82.250

TABLE II
ACCURACIES AT CROSS-VALIDATION, %

k-neigh.	Dist. №1	Dist. №2	Dist. №3	Dist. №4
5	96.750	96.850	96.250	96.350
10	94.850	94.950	94.900	94.700
15	94.550	94.400	94.100	94.200
20	93.150	93.400	92.850	92.700
25	92.800	92.750	92.450	92.000
30	92.000	92.100	91.350	90.650
35	91.300	91.550	90.150	90.250
40	90.150		89.250	
45	89.500	89.650	88.550	88.400
50	88.300	88.250	87.650	
55	87.550	87.750	86.750	86.250
60	85.900		85.200	84.650
65	85.400	85.250	84.600	83.550
70	83.850	83.700	83.100	82.600
75	82.800	83.600	82.200	82.050
80	81.900	81.800	80.700	80.450
85	81.000	81.250	79.650	80.200
90	80.000	79.400	78.850	78.450
95	79.300	79.000	77.750	78.100
100	77.750		76.900	76.850

There are achieved, regarding the assessment of the quality of classification in the procedures of resubstitution:

- ✓ minimum accuracy 83.350%, 85.150% and 82.250 at Euclidean/Minkowski, Cityblock and Chebychev distances;
- ✓ Highest value of the 98.050% for criterion for Cityblock, followed by 97.900% for Euclidean/Minkowski and 97.850% for Chebychev.

In test validation are established:

- ✓ Highest criteria of success rate in a sequential order of 96.850% for Minkowski, 96.750% for Euclidean, 96.350% for Chebychev and 96.250% for Cityblock distances;
- ✓ Lowest accuracy, as follows: 77.750% for Euclidean/Minkowski distances, 76.900% for Cityblock and 76.850% for Chebichev.

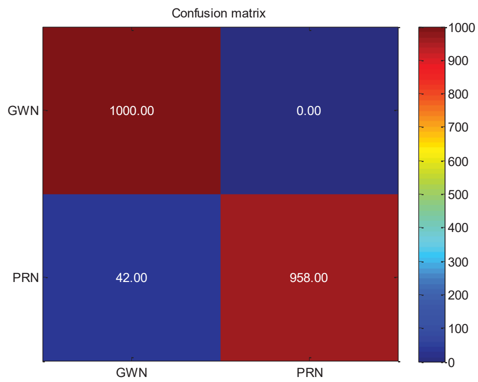


Fig. 6. Confusion matrix at Resubstitution Euclidean/Minkowski distances for k = 5

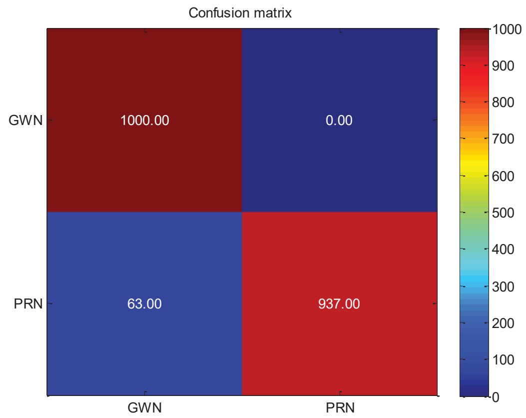


Fig. 7. Confusion matrix at Cross-validation at Minkowski distance for k = 5

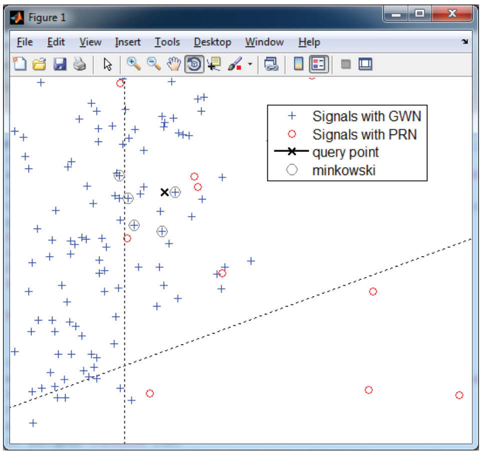
In terms of results from an analysis of classifiers concerned, matrices of correct and incorrect classifications are composed. Matrices with minimum parameter k = 5 in Fig. 6 and Fig. 7 informs about the distribution of the prototypes by groups, for Minkowski distance respectively:

- the first element of the first matrix line represents those with a properly defined affiliation of Class №1;
- The second element of the second matrix line represents the correct patterns of Class №2.

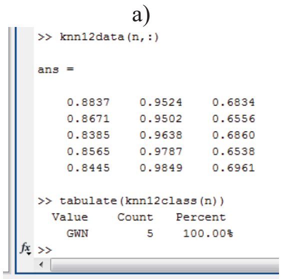
TABLE III
APPROXIMATELY EXPECTED ACCURACIES, %

k-neigh.	Dist. №1	Dist. №2	Dist. №3	Dist. №4
5	97.325	97.375	97.150	97.100
10	95.650	95.700	95.600	95.325
15	95.020	94.975	94.725	
20	93.775	93.900	93.575	93.425
25	93.500	93.475	93.025	92.900
30	92.650	92.700	92.075	91.725
35	92.100	92.225	91.375	91.225
40	91.175		90.425	90.300
45	90.725	90.800	89.750	89.775
50	89.850	89.825	89.025	88.950
55	89.275	89.375	88.325	87.925
60	88.000		87.150	86.550
65	87.375	87.300	86.600	85.775
70	86.125	86.050	85.375	84.725
75	85.350	85.800	84.600	84.150
80	84.325	84.275	83.300	82.825
85	83.575	83.700	82.600	82.450
90	82.450	82.150	81.700	80.900
95	81.825	81.675	80.650	80.675
100	80.550		79.525	79.550

The forecasts for approximate estimates of accuracy values in connection with identification of GWN or PRN impacts to signals that did not participate in k-NN patterns are given in Table 4.3. Models with the most appropriate and lowest potential application for noise identification are synthesized with 97.375% accuracy at k = 5 and 79.525% k = 100 at Minkowski and Cityblock metric distances.



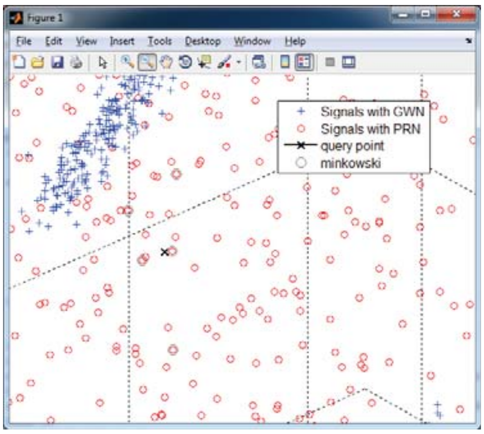
a)



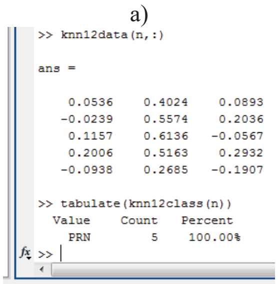
```
>> knn12data(n,:)

ans =

    0.8837    0.9524    0.6834
    0.8671    0.9502    0.6556
    0.8385    0.9638    0.6860
    0.8565    0.9787    0.6538
    0.8445    0.9849    0.6961

>> tabulate(knn12class(n))
  Value    Count   Percent
    GWN        5   100.00%
>>
```

b)

Fig. 8. k-NN search for class №1 at Minkowski distance for k = 5 in a) graphical and b) numeric types



a)



```
>> knn12data(n,:)

ans =

    0.0536    0.4024    0.0893
   -0.0239    0.5574    0.2036
    0.1157    0.6136   -0.0567
    0.2006    0.5163    0.2932
   -0.0938    0.2685   -0.1907

>> tabulate(knn12class(n))
  Value    Count   Percent
    PRN        5   100.00%
>>
```

b)

Fig. 9. k-NN search for class №2 at Minkowski distance for k = 5 in a) graphical and b) numeric types

New points have been claimed in the formed classification space to illustrate the k-neighbors search process through the selected best model. The new patterns are marked with "x", whereas the k-neighbors found in the search areas k = 5 are surrounded by "o" (Fig. 8.a) and Fig. 9.a)). In addition to the graphical, the nearest found prototypes at a defined Minkovski distance are also shown in numerical form (Fig. 8.b) and Fig. 9.b)). The identification process shows 100.00% accuracy in both test groups.

IV. Conclusion

The analysis of the background noise to analogue and digital signals using a statistical method k - Nearest Neighbors is assessed with a high degree of efficiency. A drop in accuracy of less than 90.00% is observed when searching for more than 40 points in the vicinity of new patterns. The synthesized classification model can be integrated into various signal processing systems in communications and electronics.

References

[1] T. Tanantong , "A kNN Approach for ECG Signal Quality Classification", International Journal of Information and Electronics Engineering, vol. 66, no. 4, pp. 269-272, 2016.

[2] K. Sabanci, M. Koklu, "The Classification of Eye State by Using kNN and MLP Classification Models According to the EEG Signals", International Journal of Intelligent Systems and Applications in Engineering, vol. 3, no. 3, pp. 127-130, 2015.

[3] V. Kuandanya, A. Patil, A. Panat, "Classification of Emotions from EEG Using k-NN Classifier", International Journal of Advances in Science Engineering and Technology, vol. 3, no. 2, pp. 103-106, 2015.

[4] N. Isa, A. Amir, A. Panat, M. Iyas, M. Razalli, "The Performance Analysis of K-Nearest Neighbors (K-NN) Algorithm for Motor Imagery Classification Based on EEG Signal", MATEC Web of Sciences 140, International Conference, on Emerging Electronic, Solutions for IoT, 01024, pp. 1-6, 2017.

[5] Ch. Prakash, V. Gaikwad, R. Singh, Om Prakash, "Analysis of Emotion Recognition System through Speech Signal Using KNN & GMM Classifier", International Journal of Advances in Science Engineering and Technology, vol. 10, no. 2, pp. 55-61, 2015.

[6] T. Priya, N. Raajan, N. Radju, P. Preethi, S. Mathini, "Speech and Non-speech Identification and Classification Using KNN Algorithm", ELSEVIER, International Conference on Modeling Optimization and Computing, vol. 38, pp. 952-958, 2012.

[7] H. Lee, J. Kim, S. Wibowo and S. Kim, "Bagged Fuzzy k-nearest Neighbors for Identifying Anomalous Propagation in Radar Images", INTELLI, The Sixth International Conference on Intelligent Systems and Applications, pp. 62-67, 2017.

[8] L. Lochumsen, "Radar Target Classification using Recursive Knowledge-Based Methods - Dissertation", Aalborg Universitetsforlag. Ph.d.-serien for Det Teknisk-Naturvidenskabelige Fakultet, Aalborg Universitet, pp. 1-152, 2016.

[9] Rankovska, V., S. Rankovski. Multi-functional Demo Module for Microprocessor Development Systems for Educational Purposes. Proc. XXVII International Scientific Conference Electronics - ET2018, September 13 - 15, 2018, Sozopol, Bulgaria, 2018, pp. 55-58. ISBN: 978-1-5386-6691-3. IEEE Catalog Number CFP18H39-CDR.