

Genetic Wrapper Approach for Automatic Diagnosis of Speech Disorders related to Autism

E. M. Albornoz^{1,2*}, L. D. Vignolo^{1,2}, C. E. Martínez^{1,3} and D. H. Milone^{1,2}

¹Research Center for Signals, Systems and Computational Intelligence (SINC(i))
Dpto. Informática, Facultad de Ingeniería y Cs. Hídricas, Universidad Nacional del Litoral
CC217, Ciudad Universitaria, Paraje El Pozo, S3000, Santa Fe, Argentina

²Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

³Laboratorio de Cibernética, Facultad de Ingeniería, Universidad Nacional de Entre Ríos

* emailbornoz@fich.unl.edu.ar

Abstract—The pervasive development disorders in autism condition lead to impairments in language and social communication. They are evidenced as atypical prosody production, emotion recognition and apraxia, among others communication deficits. This work tackle with the problem of the recognition of pathologies derived from these disorders in children, based on the acoustic analysis of speech. Specifically, the task consists of the diagnosis of normality (typically developing children) or three different pathologies. We propose an evolutionary approach to the feature selection stage. It relies on the use of genetic algorithm to find the set of features that optimally represent the speech data for this classification task. The genetic algorithm uses a support vector machine in order to evaluate the solutions (each individual) during the search. The results showed that our methodology improves the baseline provided for the task. The obtained unweighted classification accuracy was 54.80% on the development set, which represents a relative improvement of 6%, and 55.41% on test set. On the related task of binary classification between typical versus atypical developing condition, our approach achieved an unweighted classification accuracy of 92.66% on the test set.

Index Terms—evolutionary wrapper, speech feature selection, autism disorders

I. INTRODUCTION

In the last years, there was a big impulse to the study speech and research of language impairments in individuals with autism. This pervasive developmental disorder leads to atypical development of social communication skills such as unusual or odd-sounding prosody [1], words and content repetition [2] and impediments to appreciate and produce emotions [3]. Efforts are now focused on the study of verbal (and non verbal) behavioral features and language abilities of children, from three years old to adolescence. At early ages, the presence and characterization of voice are the strongest predictors of autism spectrum disorders [4]. Different aspects of communication were studied, including childhood apraxia [5], intonation recognition [6], interpretation and recognition of prosody [7], proficiency in spontaneous conversations [8] and vocalization patterns [9], [10].

In this work, we target the problem of detecting the presence

and type of pathology of children speakers. The problem is divided in two related tasks: a binary 'typicality' classification (typically vs. atypically developing children), and a four-classes 'diagnosis' task (typically developing and three pathologies related to autism). The classes presented are:

- TYP: typically developing children.
- PDD: pervasive development disorder.
- NOS: PDD Non-Otherwise Specified.
- DYS: DYSphasia, a specific language impairment.

The last three items comprise the 'atypically developing children' condition for the binary task.

Here we propose an evolutionary approach to optimize the selection of acoustic features measured on speech utterances. In the recent past, many efforts have been invested in the optimization of feature sets, rather than the development of more complex classifiers [11], [12], [13], [14].

Our approach relies on a genetic algorithm (GA) for finding the best combination of features. In order to evaluate the solutions during the search, the algorithm uses a support vector machine as classifier. A similar methodology was previously applied to related tasks: the search for the optimal speech parameterization for speech recognition through cepstral filterbanks and wavelet packet decomposition [15], [16], and facial features for face recognition based on active shape models [17]. They rely on the benefits provided by evolutionary computation to find better signal representations.

The organization of this paper is as follows. In Section II, brief descriptions of the speech corpora and baseline system are given. Next, our method for the automatic selection of acoustic features is described. Section III presents the results along with a discussion about the natural class grouping and further ideas on classification. Finally, Section IV gives the general conclusions and outlines future work.

II. MATERIALS AND METHODS

A. Speech data and baseline system

The dataset used consists of three sets of the *Child Pathological Speech Database* (CPDS): train, development and test

partition [18]. It provides 2.5 k instances of speech recordings from 99 children aged 6 to 18 years. 35 of these children show Pervasive Development Disorders either of autism spectrum condition (PDD, 10 male, 2 female), specific language impairment such as dysphasia (DYS, 10 male, 3 female) or PDD Non-Otherwise Specified (NOS, 9 male, 1 female) according to the DSM-IV criteria [19]. A monolingual control group consists of 64 further children (TYP, 52 male, 12 female). The French speech includes prompted sentence imitation of 26 sentences representing different modalities (declarative, exclamatory, interrogative, and imperative) and four types of intonations (descending, falling, floating, and rising). Recordings were collected in two university departments of child and adolescent psychiatry, located in Paris, France (Université Pierre et Marie Curie/Pitié- Salpêtrière Hospital and Université René Descartes/Necker Hospital).

This database contains a total of 6373 acoustic features calculated with the openSMILE feature extractor [20]. The set includes energy, spectral, cepstral (MFCC) and voicing related low-level descriptors (LLDs) as well as a few LLDs including logarithmic harmonic-to-noise ratio, spectral harmonicity, and psychoacoustic spectral sharpness [21].

The baseline system provided for both classification tasks consists of linear kernel Support Vector Machines (SVM) [18]. The system was trained with the Sequential Minimal Optimisation algorithm [22].

B. Evolutionary approach

This work presents an evolutionary method, based on evolutionary algorithms, for selecting the best features of speech [23]. Genetic algorithms are meta-heuristic optimization methods motivated by the process of natural evolution.

A classic GA consists of three kinds of operators: selection, variation and replacement. They operate on a population of individuals, which are represented as chromosomes (possible solutions) [24]. The selection operator, as in natural evolution, assigns more chance to reproduce to the fittest individuals in some task. Variation is carried out by crossover and mutation of information from different individuals, maintaining the population diversity. The replacement strategy defines the number of individuals that are substituted in the next generation. Each individual is assigned a value of fitness, measured by an objective function which is specific to the given problem. In each epoch of the learning process, the population is evolved until a desired criterion is reached. Then, the best individual in the population is taken as the solution for the problem.

The proposed method uses the 6373 features provided. Due to the high dimensionality of the data, if each feature is represented in the chromosome, the evolution process to find a near optimal solution would be excessively expensive. Thus, in order to reduce the size of the search space, the features were selected by groups. This means that each gene within the chromosome represents a group of features. The groups of features were defined based on the given hierarchical structure of the provided labels, which denote the sequence of processes applied to the signals. In this way, we grouped all the

	@attribute name string			
	L1	L2	L3	L4
1	@attribute	mfcc_sma	de[1]	_maxPos numeric
2	@attribute	mfcc_sma	de[1]	_minPos numeric
3	@attribute	mfcc_sma	de[2]	_lpc0 numeric
4	@attribute	mfcc_sma	de[2]	_lpc1 numeric
5	@attribute	pcm_zcr	sma_de	_maxPos numeric
6	@attribute	pcm_zcr	sma_de	_minPos numeric
7	@attribute	pcm_zcr	sma_de	_lpc0 numeric
8	@attribute	pcm_zcr	sma_de	_lpc1 numeric
9	@attribute	pcm_RMSenergy	sma	_maxPos numeric
10	@attribute	pcm_RMSenergy	sma	_minPos numeric
11	@attribute	pcm_RMSenergy	sma	_lpc0 numeric
12	@attribute	pcm_RMSenergy	sma	_lpc1 numeric

Fig. 1. Example illustrating the four levels feature groups (L1, L2, L3 and L4).

TABLE I
EXAMPLE ILLUSTRATING THE FEATURE GROUPS AT DIFFERENT LEVELS (SEE FIGURE 1).

L1	L2	L3	L4
1-4	1-4	1-2	1-1
5-12	5-8	3-4	2-2
	9-12	5-8	3-4
		9-12	5-8
			9-9
			10-10
			11-12

features that match the first processing stage (MFCC, auditory spectrum, etc.) in a single gene to define the chromosomes of level 1 (*GAlevel1*). Likewise, we defined the chromosome of level 2 by grouping all the features that match the first and second processing stages (first derivative of MFCC, etc.). In addition, we defined the levels 3 and 4 in the same way. As a result, the length of the chromosomes (according to the number of groups) is 37 for level 1, 67 for level 2, 975 for level 3 and 2223 for level 4. Fig. 1 shows an example of the level grouping in WEKA format (line numbers on left side) [25], while Table I shows the line numbers of the attributes selected for each group.

The GA uses the tournament selection and the classic mutation and one-point crossover operators [26]. Also, an elitist replacement strategy was applied, which maintains the best individual to the next generation and accelerates the convergence of the algorithm. The chromosomes are binary strings and the value of each gene indicates whether the corresponding group of features is used for classification or not.

The fitness function evaluates the *unweighted classification accuracy* (UAR) obtained by a SVM. The UAR is defined as the accuracy per class divided by the number of classes without considerations of instances per class [27]. A validation set to evaluate the GA evolution was extracted from the training set and so, the development set was preserved for evaluating the system performance after the evolution.

The whole approach, also known as *wrapper* [28], [29], is

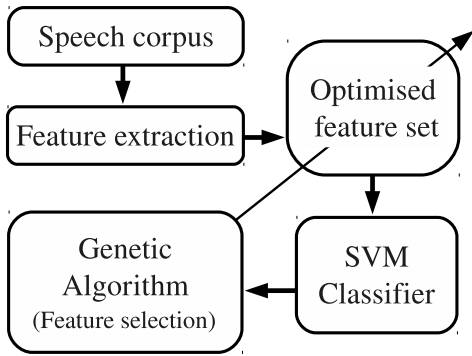


Fig. 2. Representation of the evolutionary approach proposed for feature selection.

widely used as it allows to obtain good solutions in comparison with other techniques [30]. Fig. 2 shows a schematic representation of the proposed method. It illustrates a representation optimization by genetic wrapper. The 'Genetic Algorithm' block performs the operations described before, based on the results obtained by the 'SVM Classifier' block on the complete population. Finally, in the 'Optimised feature set' block the individual with the best solution is obtained.

III. EXPERIMENTS AND RESULTS

This section describes the details of the experiments carried out in this framework. We first present the tuning and evaluation of the system for the diagnosis task, along with a discussion of the obtained results. Then, the best individual given by the GA was tested on the test set. Finally, the results on the typology task are also presented.

In order to obtain the best performance in the diagnosis task, four experiments considering the described chromosomes were done. For each case, the evolution of GA was stopped after a proper number of iterations and the best performance was kept for the validation set. The number of feature-groups selected by the best chromosomes were 20 of 37, 34 of 67, 465 of 975 and 1115 of 2223, for levels 1, 2, 3 and 4 respectively. As result, the actual number of features considered in these chromosomes were 3886 (arranged in 20 groups), 2311 (arranged in 34 groups), 3835 (arranged in 465 groups) and 3592 (arranged in 1115 groups) for levels 1, 2, 3 and 4 respectively. These chromosomes were then used for evaluating the system on the development set.

Table II shows, on one side, the best performance reached for each level on the development set, and on the other side the results on the test set. The two figures of merit used were the *Accuracy* (ACC) calculated as the recognition rate and the UAR. As can be observed, the results on the development partition with the optimised feature sets are always better than baseline for ACC, while *GAlevel1* and *GAlevel2* improved also the UAR rate. The chromosome of level 2, which is the best optimization found for this task, provided a relative improvement of 6.3% in ACC and 3.1% in UAR, which demonstrates the good learning performed by the GA. The test partition was evaluated and the method achieved 78.66%

TABLE II
CLASSIFICATION RESULTS OBTAINED FOR THE DIAGNOSIS TASK (RATES IN [%]), WITH THE BEST PERFORMANCE IN BOLD FACE.

	development partition		test partition	
	ACC	UAR	ACC	UAR
baseline	69.80	51.70	—	—
<i>GAlevel1</i>	75.60	54.30	77.68	54.88
<i>GAlevel2</i>	76.10	54.80	78.66	55.41
<i>GAlevel3</i>	73.90	46.20	77.19	51.31
<i>GAlevel4</i>	72.50	45.00	76.95	51.72

TABLE III
RECALL RESULTS FOR DIAGNOSIS TASK ON THE DEVELOPMENT SET (RATES IN [%]). THE BEST RATES FOR EACH CLASS ARE UNDERLINED.

	TYP	PDD	NOS	DYS
baseline	85.45	19.23	<u>50.00</u>	51.92
<i>GAlevel1</i>	92.82	30.77	39.71	53.85
<i>GAlevel2</i>	92.08	<u>33.65</u>	25.00	<u>68.27</u>
<i>GAlevel3</i>	<u>95.95</u>	24.04	23.53	41.35
<i>GAlevel4</i>	94.47	21.15	22.06	42.31

TABLE IV
CONFUSION MATRIX OF BASELINE FOR THE DEVELOPMENT PARTITION.

	TYP	PDD	NOS	DYS
TYP	464	44	31	4
PDD	3	20	46	35
NOS	8	16	34	10
DYS	2	33	15	54

for ACC and a 55.41% for UAR, also using the *GAlevel2* chromosome. These results, together with the previous ones, reveal a great generalization ability of our approach.

The Recall rates on the development set, exhibited in Table III, show that *GAlevel2* improves the baseline for all classes except for NOS class. Tables IV and V introduce the confusion matrices using the development partition for baseline and *GAlevel2*, respectively. These matrices give a good representation of the results per each class allowing to make a detailed analysis of performance and to find the main classification errors. The rows symbolize the actual class labels and the columns have the predicted labels; therefore, the main diagonal shows the classes correctly recognized. Comparing both tables, it can be observed the increment in the number of well-classified patterns for classes TYP (464 → 500), PDD (20 → 35) and DYS (54 → 71), and the reduction for class NOS (34 → 17). This reduction in the performance could be attributed to the fact that class NOS is the least numerous in the database, which affect the learning capabilities of the system. Almost all cases reflect less confusion than baseline except the increment of cases classified as TYP (numbers in first column of Table V for classes PDD, NOS and DYS). Nevertheless, it is important to mention that the length of this chromosome is only near to 36% (2311 features) of the original size.

The corresponding confusion matrix for test set is presented in Table VI. As can be seen, some errors arise due to patterns classified as TYP (gray shaded) and confusions between two classes: PDD-NOS and PDD-DYS (underlined). These results could guide the exploration of specific features for different

TABLE V

CONFUSION MATRIX OF GALEVEL2 FOR THE DEVELOPMENT PARTITION.

	TYP	PDD	NOS	DYS
TYP	500	36	0	7
PDD	12	35	25	32
NOS	15	22	17	14
DYS	7	21	5	71

TABLE VI

CONFUSION MATRIX OF GALEVEL2 FOR TEST PARTITION.

	TYP	PDD	NOS	DYS
TYP	527	9	3	3
PDD	14	42	16	27
NOS	24	25	24	2
DYS	8	43	1	52

TABLE VII

CLASSIFICATION RESULTS FOR TYPICALITY TASK (RATES IN [%]), WITH THE BEST PERFORMANCE IN BOLD FACE.

	development partition		test partition	
	ACC	UAR	ACC	UAR
baseline	92.60	92.80	—	—
GAlevel1	92.40	92.30	93.41	91.78
GAlevel2	91.50	92.20	93.66	92.66
GAlevel3	93.30	93.30	93.17	91.68
GAlevel4	91.60	91.20	91.32	90.51

groups of classes in a scheme of hierarchical classifiers structured from non-supervised clustering [31]. We consider our results on the development and the test set very encouraging. This is especially true given the increase in the rates obtained on the test set with respect to the development set, shown in Table II.

In the *typicality* task, the experiments were carried out using the same optimised chromosomes for the *diagnosis* task. The performance reached for the four levels on the development and test set are presented in Table VII. The best result on the development set was achieved with the chromosome of level 3, for which the obtained ACC and UAR rate are better than baseline. For the test set, the best performance was obtained –as in the diagnosis task– with the same chromosome of level 2. In this case, the ACC and UAR were, respectively, 93.66% and 92.66%. As can be seen, the results are even better than those obtained with the development set, which is also pointing out the good generalization abilities of our system.

IV. CONCLUSIONS

In this paper we presented an evolutionary method for automatic selection of features of speech in a classification task. It is based on a genetic algorithm that selects the best combination of acoustic features using support vector machines as classifier. The approach is applied to the diagnosis task of typically developing children or the presence of speech disorders in autism.

Results showed an unweighted classification accuracy (UAR) of 54.80% on the development set, which represent a relative improvement of 6% with respect to the baseline. On the test set, the UAR obtained was 55.41%. The pro-

posed method optimizes the number of features, reducing the complexity of the classifiers. Our approach outperform the baseline system and also results in an improvement of the generalization capabilities.

The analysis of the confusion matrices suggests further work in order to find and model natural groupings between classes. They could lead to the proposal of hierarchical classifiers, with specialized features and classifiers at each stage.

ACKNOWLEDGEMENTS

The authors wish to thank the support of the *Agencia Nacional de Promoción Científica y Tecnológica*, the *Universidad Nacional de Litoral* (with PAE-PICT-2007-00052 and CAI+D 012-72), the *Universidad Tecnológica Nacional-FRSF* (with PID O/0153), and the *Consejo Nacional de Investigaciones Científicas y Técnicas* (CONICET) from Argentina.

REFERENCES

- [1] J. McCann and S. Peppe, "Prosody in autism spectrum disorders: a critical review," *International Journal of Language & Communication Disorders*, vol. 38, no. 4, pp. 325–350, 2003.
- [2] H. Green and Y. Tobin, "Repetitive extralinguistic, prosodic and linguistic behavior in autism spectrum disorders-high functioning (asd-hf)," *A Comprehensive Book on Autism Spectrum Disorders*, p. 49, 2011.
- [3] H. Macdonald, M. Rutter, P. Howlin, P. Rios, A. Conteur, C. Evered, and S. Folstein, "Recognition and expression of emotional cues by autistic and normal adults," *Journal of Child Psychology and Psychiatry*, vol. 30, no. 6, pp. 865–877, 2006.
- [4] M. Mody and J. Belliveau, "Speech and Language Impairments in Autism: Insights from Behavior and Neuroimaging," *North American Journal of Medicine and Science*, vol. 5, no. 3, pp. 157 – 162, 2012.
- [5] L. Shriberg, R. Paul, L. Black, and J. van Santen, "The hypothesis of apraxia of speech in children with autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 41, no. 4, pp. 405–426, 2011.
- [6] F. Ringeval, J. Demouy, G. Szaszak, M. Chetouani, L. Robel, J. Xavier, D. Cohen, and M. Plaza, "Automatic intonation recognition for the prosodic assessment of language-impaired children," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 5, pp. 1328–1342, 2011.
- [7] J. van Santen, E. Prudhommeaux, and L. Black, "Automated assessment of prosody production," *Speech Communication*, vol. 51, no. 11, pp. 1082–1097, 2009.
- [8] M. Black, D. Bone, M. Williams, P. Gorrindo, P. Levitt, and S. Narayanan, "The usc care corpus: Child-psychologist interactions of children with autism spectrum disorders," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
- [9] M. Sharda, T. Subhadra, S. Sahay, C. Nagaraja, L. Singh, R. Mishra, A. Sen, N. Singhal, D. Erickson, and N. Singh, "Sounds of melodypitch patterns of speech in autism," *Neuroscience Letters*, vol. 478, no. 1, pp. 42–45, 2010.
- [10] D. Bone, M. P. Black, C. Lee, M. Williams, P. Levitt, S. Lee, and S. Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist," in *Proc. Interspeech*, 2012.
- [11] H. Liu and H. Motoda, *Feature selection for knowledge discovery and data mining*. Springer, 1998.
- [12] I. Vatolkin, M. Preu, G. Rudolph, M. Eichhoff, and C. Weihs, "Multi-objective evolutionary feature selection for instrument recognition in polyphonic audio mixtures," *Soft Computing*, vol. 16, no. 12, pp. 2027–2047, 2012.
- [13] I. Mierswa and K. Morik, "Automatic feature extraction for classifying audio data," *Machine Learning*, vol. 58, no. 2-3, pp. 127–149, 2005.
- [14] A. Atyabi, M. Luerssen, S. Fitzgibbon, and D. M. W. Powers, "Evolutionary feature selection and electrode reduction for eeg classification," in *Evolutionary Computation (CEC), 2012 IEEE Congress on*, 2012, pp. 1–8.

- [15] L. D. Vignolo, H. L. Rufiner, D. H. Milone, and J. C. Goddard, "Evolutionary Cepstral Coefficients," *Applied Soft Computing*, vol. 11, no. 4, pp. 3419 – 3428, 2011.
- [16] L. D. Vignolo, D. H. Milone, and H. L. Rufiner, "Genetic wavelet packets for speech recognition," *Expert Systems with Applications*, vol. 40, no. 6, pp. 2350 – 2359, 2013.
- [17] L. D. Vignolo, D. H. Milone, and J. Scharcanski, "Feature selection for face recognition based on multi-objective evolutionary wrappers," *Expert Systems with Applications*, vol. 40, no. 13, pp. 5077 – 5084, 2013.
- [18] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism," in *Proc. Interspeech 2013, ISCA, Lyon, France*, 2013.
- [19] M. B. First, "Diagnostic and statistical manual of mental disorders (dsm-iv)," *American Psychiatric Association, Washington, DC*, 1994.
- [20] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE: The Munich Versatile and Fast open-source Audio Feature Extractor," in *Proceedings of the international conference on Multimedia*, ser. MM '10. New York, NY, USA: ACM, 2010, pp. 1459–1462. [Online]. Available: <http://doi.acm.org/10.1145/1873951.1874246>
- [21] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. van Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, and B. Weiss, "The INTERSPEECH 2012 Speaker Trait Challenge," in *Proc. Interspeech*, 2012.
- [22] J. Platt *et al.*, "Sequential minimal optimization: A fast algorithm for training support vector machines," *Microsoft Research*, 1998.
- [23] S. N. Sivanandam and S. N. Deepa, *Introduction to Genetic Algorithms*. Springer London, Limited, jan 2008.
- [24] H. Youssef, S. M. Sait, and H. Adiche, "Evolutionary algorithms, simulated annealing and tabu search: a comparative study," *Engineering Applications of Artificial Intelligence*, vol. 14, no. 2, pp. 167 – 181, 2001.
- [25] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [26] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, 1st ed. Addison-Wesley Professional, Jan. 1989.
- [27] B. Schuller, B. Vlasenko, F. Eyben, M. Wollmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll, "Cross-corpus acoustic emotion recognition: Variances and strategies," *Affective Computing, IEEE Transactions on*, vol. 1, no. 2, pp. 119–131, July-Dec.
- [28] S. Durbha, R. King, and N. Younan, "Wrapper-based feature subset selection for rapid image information mining," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, no. 1, pp. 43 –47, January 2010.
- [29] H.-H. Hsu, C.-W. Hsieh, and M.-D. Lu, "Hybrid feature selection by combining filters and wrappers," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8144 – 8150, 2011.
- [30] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273 – 324, 1997.
- [31] E. M. Albornoz, D. H. Milone, and H. L. Rufiner, "Spoken emotion recognition using hierarchical classifiers," *Computer Speech and Language*, vol. 25, pp. 556–570, 2011.