INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

# Detecting Mild Cognitive Impairment from Spontaneous Speech by Correlation-Based Phonetic Feature Selection

*Gábor Gosztolya* [1,2], *László Tóth* [1], *Tamás Grósz* [2], *Veronika Vincze* [1],
*Ildikó Hoffmann* [3,4], *Gréta Szatlóczki* [5], *Magdolna Pákáski* [5], *János Kálmán* [5]

[1] MTA-SZTE Research Group on Artificial Intelligence, Szeged, Hungary
[2] University of Szeged, Department of Informatics, Szeged, Hungary
[3] University of Szeged, Department of Linguistics, Szeged, Hungary
[4] Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary
[5] University of Szeged, Department of Psychiatry, Szeged, Hungary
{ ggabor, tothl, groszt, vinczev } @ inf.u-szeged.hu

## Abstract

Mild Cognitive Impairment (MCI), sometimes regarded as a prodromal stage of Alzheimer's disease, is a mental disorder that is difficult to diagnose. Recent studies reported that MCI causes slight changes in the speech of the patient. Our previous studies showed that MCI can be efficiently classified by machine learning methods such as Support-Vector Machines and Random Forest, using features describing the amount of pause in the spontaneous speech of the subject. Furthermore, as hesitation is the most important indicator of MCI, we took special care when handling filled pauses, which usually correspond to hesitation. In contrast to our previous studies which employed manually constructed feature sets, we now employ (automatic) correlation-based feature selection methods to find the relevant feature subset for MCI classification. By analyzing the selected feature subsets we also show that features related to filled pauses are useful for MCI detection from speech samples.

**Index Terms**: mild cognitive impairment, machine learning, temporal parameters of speech, feature selection

## 1. Introduction

Alzheimer's disease (AD) is a very distinct neurodegenerative disorder that may develop for years before clinical manifestation. It is estimated that over 7% of the population aged 60 and over suffer from AD or some other kind of dementia in Europe, and similar figures are estimated for the U.S. as well [1]. However, the symptoms of Mild Cognitive Impairment (MCI) might be detected years before the actual diagnosis of AD [2]. This tells us that the clinical appearance of AD is preceded by a prolonged, preclinical phase. Therefore, timely diagnosis and treatment are very important, as its progression can be decelerated and occurrence of new symptoms can be delayed [3].

MCI is known to influence the verbal fluency of the spontaneous speech of the patient [4], which manifests itself in longer hesitations and a lower speech rate [5, 6, 7]. Therefore, many studies performed MCI or AD detection by extracting features related to articulation/speech rate and pause (e.g. [7, 8, 9]). These studies concentrated only on silent pauses, as these can be detected easily by using simple signal processing tools. However, filled pauses (sounds like "er", "hmm" etc.) also indicate hesitations, and can take up a significant amount of speech time. For example, Tóth et al. found that about 10% of the hesitations in a Hungarian speech database appear as filled pauses [10]. However, signal processing-based pause detection methods usually carry out a voiced-unvoiced split of the utterance, and this approach is unable to find filled pauses. The only study we found besides ours which deals with filled pauses is that of Roark et al. [6]; they, however, created a manual annotation of the utterances, and did not deal with the automatic detection of filled pauses (or automatic feature extraction at all).

Focusing on filled pauses, the study by Hoffmann et al. suggested the calculation of eight features (or *biomarkers*) from utterances containing spontaneous speech [5]. These features either described the speech rate of the speaker (i.e. phonemes per second) or the amount of pausing in the speech of the subjects. The novelty in this study was that it proved that besides silent pauses, the amount of filled pauses also display a statistically significant difference between the speech of the two groups of speakers (MCI and control). In a more recent study [10] we found that the extraction of these features can be reliably automated by applying standard Automatic Speech Recognition (ASR) techniques, and that extending the features with biomarkers related to other phonemes (commonly confused by certain types of filled pauses) can be beneficial for MCI detection performance. For example, the most frequent sound uttered during hesitation is a schwa, which is easily confused with the vowel [ø]. By adding further such features to the set we were able to outperform the results achieved with the basic feature set of Hoffmann et al. [5].

Notice that both feature sets were manually constructed ones. This is typical in this area; for example, López-de-Ipiña et al. tested four such feature sets consisting of acoustic, voice quality and duration features to detect AD [8], while Sztahó et al. used jitter, shimmer, articulation rate and speech intensity to detect Parkinson's disease [11]. Still, applying automatic feature selection methods over a broader set of input features can be expected to lead to a better dementia identification performance. This approach is also used in studies dealing with the automatic identification of various types of dementia from speech samples. For example, both Satt et al. [12] and Fraser et al. [13] calculated single-tailed $p$-values for the features and used only those that had lower $p$ scores. Sidorov et al. used a genetic algorithm to construct the final feature set [14]. Yu et al. selected their features by using the sequential forward feature selection algorithm [15].

In this study, we extend our previous investigations by performing automatic feature selection for detecting Mild Cogni-

http://dx.doi.org/10.21437/Interspeech.2016-384

tive Impairment. For this, we first extract a highly redundant feature set based on ASR output, then perform feature selection to get the most suitable feature subset. To do this, we will apply two correlation-based feature selection algorithms, used for conflict intensity estimation before [16], and show the superiority of the selected feature subsets compared to the earlier, manually constructed ones by Hoffmann et al. [5] and Tóth et al. [10]. A further novelty of this study is that we also compare the efficiency of these correlation-based methods with other feature selection ones applied in this area before: Sequential Forward Selection utilized by Yu et al. [15], and the filtering of attributes according to statistical significance (applied previously by Satt et al. [12] and by Fraser et al. [13]). We also analyze the time requirement of each feature selection algorithm and the feature subsets selected. In the end we find that both automatic feature selection and the detection of filled pauses are important in order to achieve efficient automatic MCI detection.

## 2. Automatic Feature Selection for MCI Detection

### 2.1. Utterance Recording Setup

We recorded our utterances containing Hungarian spontaneous speech from our patients in the following way [5]. After the presentation of a specially designed one-minute-long animated film, the subjects were asked to talk about the events seen on the film (immediate recall). After the presentation of a second film, the subjects were asked to talk about their previous day (spontaneous speech). For the last task, the subjects were asked to talk about the second film (delayed recall). Each task made the subjects produce spontaneous speech, but in a different way, hence their speech can be expected to be different as well. Of course, it may turn out that some tasks are less useful for detecting MCI than others, but we cannot know this in advance.

This set-up is special in the sense that we will have three utterances for each speaker. Although we performed our experiments on utterances recorded this way, we think that the techniques applied here and the proposed methods can be readily carried over to other tasks and databases in similar areas.

### 2.2. Basic Feature Sets

The basic feature set defined by Hoffmann et al. consisted of eight features for each utterance [5]. Firstly, it contained the duration of the utterance, the articulation rate (the number of phones per second during speech, excluding hesitations) and the speech rate (number of phones per second during speech, including hesitations). Four further features were the number of occurrences and the total duration of both silent and filled pauses. Besides these, it included the total duration of pauses (both silent and filled) divided by the duration of the utterance. Some personal attributes of the speaker were also added to the feature set: age, gender, and level of education (expressed in terms of school years), resulting in a 27-item set for each speaker. We will refer to this feature set as the *basic* feature set.

Tóth et al. proposed an extended version of this feature set [10]. This, like the basic feature set, first consisted of the duration of the utterance, articulation rate and speech rate. Next, it included four descriptors for each of some specific phonemes, listed in Table 1. These descriptors were calculated for the silent pauses, the filled pauses, the silent and filled pauses together, and the phonemes [m], [n] and [ø]. This gave 27 attributes for each utterance, which, along with the speaker-related attributes, gave 84 features for each speaker. Tóth et al. found this feature set to be superior to the basic feature set of Hoffmann et al. Here, we will refer to this set as the *extended* feature set.

(1) The number of occurrences of the given phoneme divided by the total number of phoneme occurrences.

(2) The total duration of occurrences of the given phoneme divided by the duration of the utterance.

(3) The mean length of the occurrences of the given phoneme.

(4) The standard deviation of the length of the occurrences of the given phoneme.

Table 1: *The four descriptors, following the work of Tóth et al. [10].*

### 2.3. Overcomplete Feature Set

The feature sets we have used so far were all manually constructed ones. The extended feature set was constructed in the hope that by adding some redundant or irrelevant features, the machine learning methods applied would ignore these unnecessary extra features. However, this is not always the case, and in the machine learning literature an enormous number of feature selection methods have been proposed, even in speech technologies (e.g. [16, 17, 18, 19, 20]). Therefore, next we will extract a highly redundant feature set, which will serve as the basis of our feature selection experiments (*overcomplete* feature set).

Firstly, we add the three general features (utterance duration, articulation rate and speech rate) to the feature set. Next, we add the descriptors listed in Table 1 for *each* phoneme to this feature set. The only exception is the case of hesitations, where, following the studies of Hoffmann et al. and Tóth et al., we calculate these descriptors for silent pauses and filled pauses separately and treat them as one phoneme as well, resulting in 12 pause-related attributes overall. With 57 phonemes (including filled pauses, breathing noises, laughter and coughs), this resulted in 235 features for each utterance and 708 for each speaker. Although most of these features can be expected to be irrelevant or redundant, we will filter them out in the next step.

### 2.4. Automatic Feature Selection

A large number of automatic feature selection methods have been described in the literature. Following our previous study [16], where we performed conflict intensity estimation, we will combine two feature selection approaches. The first one relies on the correlation of the target and each feature, and seeks to utilize the more correlated features [20, 21, 22]. Although this approach is designed for regression tasks where we have to match a continuous annotation, it is quite straightforward to adapt to a binary classification task like MCI detection.

The weakness of performing feature selection only on the basis of correlation is, however, that after feature selection, some kind of machine learning method (SVM, ANN etc.) is trained using the restricted feature set, but the special aspects of this method are ignored during feature selection. This may be improved if we incorporate the machine learning method in the feature selection process [22, 23]. This approach has the advantage that we will more likely pick those features which are relevant *for the given machine learning method*.

We used *forward* methods, which commence with an empty (or very restricted) feature set, and expand it step-by-step [23]. They tend to converge to the final feature set quite efficiently [17]. Perhaps the most well-known forward method is the Sequential Forward Selection algorithm (SFS, [24]): this, for each step, adds each feature to the set, and keeps the one which resulted in the biggest improvement in accuracy.

Besides applying SFS, we also utilized two correlation-based methods. Firstly, the features were sorted according to the absolute value of their correlation score with the class (now MCI or control group) in *descending* order. As we add the features to our selected feature subset in this order, more correlating features will be used first. In the first algorithm each feature is examined only once: if using the actual feature improves the classification performance, we permanently add this feature to our set of selected features, otherwise we permanently discard it. In the second correlation-based algorithm, in the $n$th iteration, we use all the $n$ most correlated features. These two-step feature selection methods, despite their relative simplicity, were able to efficiently improve the classification scores, while also reducing the computational requirements of MCI classification.

## 3. Experimental Setup

### 3.1. ASR-based Feature Extraction

The speech recognizer was trained on the BEA Hungarian Spoken Language Database [25]. This database contains spontaneous speech, like the recordings collected from our MCI patients. We utilized roughly seven hours of speech data from the BEA corpus – mainly recordings from elderly persons, in order to match the age group of the targeted MCI audience. The annotation of the dataset included filled pauses, breath intakes and exhales, laughter, coughs and gasps in the transcriptions.

The ASR system was trained to recognize the phones in the utterances, where the phone set included the special non-verbal labels listed above. For acoustic modeling we applied a special convolutional deep neural network-based technology. With this approach we managed to achieve one of the lowest phone recognition error rates on the TIMIT database [26, 27]. We employed a simple phone bigram language model (once again, including all the above-mentioned non-verbal audio tags). The output of the ASR system is the phonetic segmentation and labeling of the input signal, including filled pauses. Based on this output, the acoustic biomarkers listed in Section 2 can be readily extracted.

### 3.2. MCI Classification

Our database of MCI patients is continuously growing; at the time of writing we had recordings taken from over 100 persons. For various reasons (poor sound quality, controversial diagnosis, etc.) we had to filter out some patients, so in our experiments we used the recordings of 84 subjects. From these, 48 had MCI and 36 were control subjects. For each subject we had three recordings for the three different tasks. From a machine learning perspective, this is an extremely small dataset, but the number of diagnosed MCI patients is limited, and collecting recordings of their speech is tedious. Perhaps this is why in all the similar studies we found involved fewer than 100 patients [6, 7, 9, 13, 28]. The only exception we know of is the study by Yu et al. [15], which had 139 speakers; however, from these, only 20 had MCI.

Having so few examples, we did not create separate training and test sets, but applied the common solution of 10-fold cross validation. We used Support Vector Machines (SVM, [29]) with the LibSVM library [30]; the $C$ parameter of the linear kernel was tested in the range $10^{\{-5,\dots,1\}}$. To be able to perform comparisons with previous results (Tóth et al. had only 51 speakers and used the Weka toolkit [31]), we evaluated the feature sets proposed earlier using the set-up outlined above.

### 3.3. Evaluation Metrics

In the past, many studies relied on standard classification accuracy (e.g. [13, 32]). However, while in our actual dataset the distribution of subjects is quite balanced, this is not so for most datasets of this area. The distribution of the two groups in the elderly populations is not balanced either. For such an unbalanced class distribution, though, accuracy is not a reliable metric. For this reason, we opted for the standard Information Retrieval metrics of *precision*, *recall* and their harmonic mean, *F-measure* (or $F_1$-score). We will also use the metric called Unweighted Average Recall (UAR), being the mean value of the recall scores for all the classes. This is a popular metric in the area of computational paralinguistics (see e.g. [33, 34]), and its main advantage is that, unlike the traditional accuracy and F-measure metrics, it is unaffected by a change in class frequency.

During feature selection we seek to select the feature subset by which we can achieve the most accurate classification. However, we can use three of the above-listed metrics to measure this "accurateness" (precision and recall are clearly not suitable for this). So we tested the methods using accuracy, UAR and $F_1$.

## 4. Results and Discussion

Table 2 lists the scores got for the different feature selection strategies when optimized for different metrics. It can be seen that using the manually constructed feature sets led to quite low accuracy scores. The worst-performing one was clearly the overcomplete feature set, where all patients were assumed to have MCI. Notice that, out of the three accuracy metrics optimized for, only UAR was capable of detecting this phenomenon, as it was able to counter the imbalanced class distribution; $F_1$, however, was quite high due to the recall value being 100%. Still, it is clear that this approach cannot be used in practice, which indicates a weakness of $F_1$.

Perhaps this is the reason why the Sequential Forward Selection method produced bad scores when optimized for $F_1$: a relatively high $F_1$ value could be achieved by using only one attribute (and classifying each speaker as one having MCI), but then the optimization quickly got stuck in a local maximum. Using the most correlated $n$ features ("Top-Ranked" strategy) was slightly worse than the other two feature selection methods ("Greedy" and SFS), while the latter two produced very similar accuracies. However, SFS produced quite compact feature sets.

Table 3 shows the number of SVM models trained during the feature selection process. (Note that we halted SFS after it was unable to improve the accuracy score in an iteration, while the correlation-based approaches examined each attribute, leading to 708 iterations overall.) We can see that SFS required over 7 times more SVM trainings (except when optimized for $F_1$, but then it yielded an unusable model). Of course, as SFS used fairly compact feature sets, on which the SVM models could be trained much more quickly, this 7-fold difference does not fully reflect the relation between the actual training times. Using the rough estimation that, when the number of examples and classes is the same, the time required for training an SVM model is linearly proportional to the number of attributes used, we still find

|  |  | F. set | Measured Metrics |  |  |  |  |
|---|---|---|---|---|---|---|---|
| Feature set / selection method | Opt. metric | size | Acc. | UAR | Prec. | Rec. | $F_1$ |
| Sequential Forward | Acc. | 6 | 86.9% | 86.5% | 87.8% | 89.6% | **88.7%** |
|  | UAR | 6 | **88.1%** | **88.5%** | **93.2%** | 85.4% | **89.1%** |
|  | $F_1$ | 3 | 63.1% | 56.9% | 60.8% | **100.0%** | 75.6% |
| Correlation-Based (Greedy) | Acc. | 10 | 86.9% | 87.5% | **93.0%** | 83.3% | 87.9% |
|  | UAR | 11 | 86.9% | 87.5% | **93.0%** | 83.3% | 87.9% |
|  | $F_1$ | 13 | **88.1%** | **88.5%** | **93.2%** | 85.4% | **89.1%** |
| Correlation-Based (Top-Ranked) | Acc. | 40 | 85.7% | 86.1% | 90.9% | 83.3% | 87.0% |
|  | UAR | 40 | 85.7% | 86.1% | 90.9% | 83.3% | 87.0% |
|  | $F_1$ | 42 | 85.7% | 85.8% | 89.1% | 85.4% | 87.2% |
| Feature selection by $t$-test ($p \leq 0.05$) [12, 13] |  | 46 | 81.0% | 81.3% | 86.4% | 79.2% | 82.6% |
| Basic feature set [5] |  | 27 | 57.1% | 56.3% | 62.5% | 62.5% | 62.5% |
| Extended feature set [10] |  | 84 | 63.1% | 62.5% | 68.1% | 66.7% | 67.4% |
| Overcomplete feature set |  | 708 | 57.1% | 50.0% | 57.1% | **100.0%** | 72.7% |

Table 2: The accuracy scores achieved using the different feature selection methods.

|  | Opt. | No. of | Avg. feat. |
|---|---|---|---|
| Selection method | metric | SVMs | count |
| Sequential Forward | Acc. | 408 240 | 4.0 |
|  | UAR | 408 240 | 4.0 |
|  | $F_1$ | 233 760 | 2.5 |
| Correlation (Greedy) | Acc. | 56 640 | 10.4 |
|  | UAR | 56 640 | 11.3 |
|  | $F_1$ | 56 640 | 13.3 |
| Correlation (Top-Ranked) | Acc. | 56 640 | 354.5 |
|  | UAR | 56 640 | 354.5 |
|  | $F_1$ | 56 640 | 354.5 |

Table 3: The number and average size of SVM models trained during feature selection.

that the greedy correlation-based feature selection method required 54-64% less time than SFS. The top-ranked correlation-based method, however, turns out to be much slower, due to the large feature vectors employed in the later iterations.

Figure 1 shows what kinds of features make up the feature subsets. We can see that the attributes describing the speaker (age, gender and years of education) take up only a small fraction of the more successful feature sets, and the same is true for speech rate, articulation rate and utterance length. Nevertheless, silence-related attributes are very common, and so are the biomarkers related to filled pauses. In our opinion this result justifies our efforts related to detecting filled pauses in order to detect MCI. What is pretty surprising, though, is that attributes related to the miscellaneous phonemes are also quite common: they take up at least 40% of each automatically constructed feature set. However, they are completely ignored in the manual feature sets, which may be the reason why these performed quite poorly. Many of these features corresponded to phonemes commonly used in another form of signaling hesitation. In spontaneous speech speakers tend to lengthen certain sounds within a content word; for instance, the word *kész* [ke:s] (meaning “ready”) is sometimes pronounced as [ke:ss]. These cases are also indicators of hesitation and may have also influenced the frequency of each phoneme. They were found by the automatic feature selection methods, while they were missing from the manually constructed feature sets.

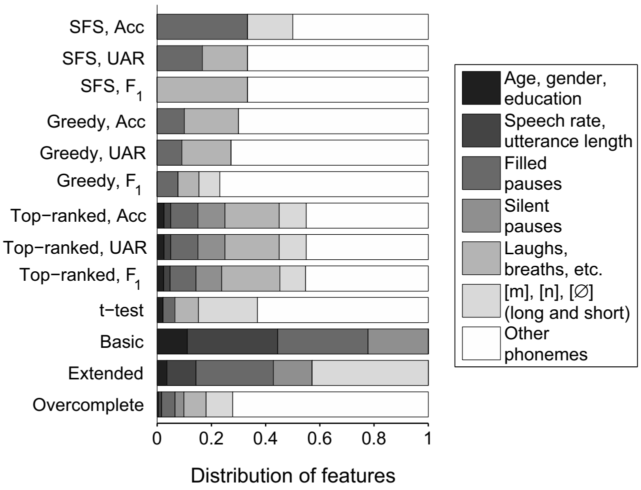

Figure 1: The composition of the selected feature subsets.

## 5. Conclusions

Mild cognitive impairment (MCI) is known to cause slight changes in the spontaneous speech of the patient. Our starting point was a study that created feature sets describing the amount of hesitations present in spontaneous speech. In this study, we utilized automatic feature selection methods to automate the construction of the set of phonetic level biomarkers used for MCI detection. We found the automatically selected feature sets to be superior to the manually constructed ones, presumably because they contained information about phonemes frequently used for another form of hesitation: lengthening. We also compared various feature selection methods, and found the proposed correlation-based technique to be just as accurate, but much faster than the sequential forward selection method.

## 6. Acknowledgements

The Titan X graphics card used for this research was donated by the NVIDIA Corporation. Ildikó Hoffmann was supported by the Bolyai János Research Scholarship.

# 7. References

[1] B. Duthey, *Background Paper 6.11: Alzheimer Disease and other Dementias*. WHO, 2013.

[2] K. L. de Ipiña, J.-B. Alonso, C. M. Travieso, J. Sol-Casals, H. Egiraun, M. Faundez-Zanuy, A. Ezeiza, N. Barroso, M. Ecay-Torres, P. Martinez-Lage, and U. M. de Lizardui, "On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis," *Sensors*, vol. 13, no. 5, pp. 6730–6745, 2013.

[3] J. Kálmán, M. Pákáski, I. Hoffmann, G. Drótos, G. Darvas, K. Boda, T. Bencsik, A. Gyimesi, Z. Gulyás, M. Bálint *et al.*, "Early mental test – developing a screening test for mild cognitive impairment," *Ideggyógyászati szemle*, vol. 66, no. 1-2, pp. 43–52, 2013.

[4] C. Laske, H. R. Sohrabi, S. M. Frost, K. L. de Ipiña, P. Garrard, M. Buscema, J. Dauwels, S. R. Soekadar, S. Mueller, C. Linnemann, S. A. Bridenbaugh, Y. Kanagasingam, R. N. Martins, and S. E. O'Bryant, "Innovative diagnostic tools for early detection of Alzheimer's disease (in press)," *Alzheimer's & Dementia*, 2015.

[5] I. Hoffmann, D. Németh, C. D. Dye, M. Pákáski, T. Irinyi, and J. Kálmán, "Temporal parameters of spontaneous speech in Alzheimer's disease," *International Journal of Speech-Language Pathology*, vol. 12, no. 1, pp. 29–34, 2010.

[6] B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye, "Spoken language derived measures for detecting mild cognitive impairment," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 7, pp. 2081–2090, 2011.

[7] W. Jarrold, B. Peintner, D. Wilkins, D. Vergryi, C. Richey, M. L. Gorno-Tempini, and J. Ogar, "Aided diagnosis of dementia type through computer-based analysis of spontaneous speech," in *Proceedings of CLPsych*, Baltimore, Maryland, USA, 2014, pp. 27–37.

[8] K. L. de Ipiña, J. B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C. M. Travieso, M. Ecay-Torres, P. Martínez-Lage, and H. Eguiraun, "On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature," *Cognitive Computation*, vol. 7, no. 1, pp. 44–55, 2015.

[9] A. Satt, R. Hoory, A. König, P. Aalten, and P. H. Robert, "Speech-based automatic and robust detection of very early dementia," in *Proceedings of Interspeech*, Singapore, 2014, pp. 2538–2542.

[10] L. Tóth, G. Gosztolya, V. Vincze, I. Hoffmann, G. Szatlóczki, E. Biró, F. Zsura, M. Pákáski, and J. Kálmán, "Automatic detection of mild cognitive impairment from spontaneous speech using ASR," in *Proceedings of Interspeech*, Dresden, Germany, Sep 2015, pp. 2694–2698.

[11] D. Sztahó, G. Kiss, and K. Vicsi, "Estimating the severity of Parkinson's disease from speech using linear regression and database partitioning," in *Proceedings of Interspeech*, Dresden, Germany, 2015, pp. 498–502.

[12] A. Satt, A. Sorin, O. Toledo-Ronen, O. Barkan, I. Kompatsiaris, A. Kokonozi, and M. Tsolaki, "Evaluation of speech-based protocol for detection of early-stage dementia," in *Proceedings of Interspeech*, Lyon, France, 2013, pp. 1692–1696.

[13] K. C. Fraser, F. Rudzicz, and E. Rochon, "Using text and acoustic features to diagnose progressive aphasia and its subtypes," in *Proceedings of Interspeech*, Lyon, France, 2013, pp. 25–29.

[14] M. Sidorov, C. Brester, and A. Schmitt, "Contemporary stochastic feature selection algorithm for speech-based emotion recognition," in *Proceedings of Interspeech*, Dresden, Germany, 2015, pp. 2699–2703.

[15] B. Yu, T. F. Quatieri, J. R. Williamson, and J. C. Mundt, "Cognitive impairment prediction in the elderly based on vocal biomarkers," in *Proceedings of Interspeech*, Dresden, Germany, 2015, pp. 3734–3738.

[16] G. Gosztolya, "Conflict intensity estimation from speech using greedy forward-backward feature selection," in *Proceedings of Interspeech*, Dresden, Germany, Sep 2015, pp. 1339–1343.

[17] M. Brendel, R. Zaccarelli, and L. Devillers, "A quick sequential forward floating feature selection algorithm for emotion detection from speech," in *Proceedings of Interspeech*, Makuhari, Japan, 2010, pp. 1157–1160.

[18] K. Kirchhoff, Y. Liu, and J. Bilmes, "Classification of developmental disorders from speech signals using Submodular Feature Selection," in *Proceedings of Interspeech*, Lyon, France, Sep 2013, pp. 187–190.

[19] O. Räsänen and J. Pohjalainen, "Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech," in *Proceedings of Interspeech*, Lyon, France, Sep 2013, pp. 210–214.

[20] H. Kaya, F. Eyben, A. A. Salah, and B. Schuller, "CCA based feature selection with application to continuous depression recognition from acoustic speech features," in *Proceedings of ICASSP*, Florence, Italy, 2014, pp. 3757–3761.

[21] H. Kaya, T. Özkaptan, A. A. Salah, and F. Gürgen, "Random discriminative projection based feature selection with application to conflict recognition," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 671–675, 2015.

[22] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, University of Waikato, 1999.

[23] L. Molina, "Feature selection algorithms: a survey and experimental evaluation," in *Proceedings of ICDM*, 2002, pp. 306–313.

[24] P. Devijver and J. Kittler, *Pattern Recognition, a Statistical Approach*. Prentice Hall, 1982.

[25] M. Gósy, "BEA a multifunctional Hungarian spoken language database," *The Phonetician*, vol. 105, no. 106, pp. 50–61, 2012.

[26] L. Tóth, "Convolutional deep maxout networks for phone recognition," in *Proceedings of Interspeech*, 2014, pp. 1078–1082.

[27] ——, "Phone recognition with hierarchical Convolutional Deep Maxout Networks," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 25, pp. 1–13, 2015.

[28] M. Lehr, E. Prudhommeaux, I. Shafran, and B. Roark, "Fully automated neuropsychological assessment for detecting Mild Cognitive Impairment," in *Proceedings of Interspeech*, Portland, OR, USA, 2012.

[29] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[30] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011.

[31] 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.

[32] P. Garrard, V. Rentoumi, B. Gesierich, B. Miller, and M. L. Gorno-Tempini, "Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse," *Cortex*, vol. 55, pp. 122–129, 2014.

[33] 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 *Proceedings of Interspeech*, 2013.

[34] B. Schuller, S. Steidl, A. Batliner, S. Hantke, F. Hnig, J. R. Orozco-Arroyave, E. Nth, Y. Zhang, and F. Weninger, "The INTERSPEECH 2015 computational paralinguistics challenge: Nativeness, Parkinson's & eating condition," in *Proceedings of Interspeech*, 2015.