Random Discriminative Projection Based Feature Selection with Application to Conflict Recognition

Heysem Kaya, *Student Member, IEEE*, Tuğçe Özkaptan, Albert Ali Salah, *Member, IEEE*, and Fikret Gürgen

***Abstract*—Computational paralinguistics deals with underlying meaning of the verbal messages, which is of interest in manifold applications ranging from intelligent tutoring systems to affect sensitive robots. The state-of-the-art pipeline of paralinguistic speech analysis utilizes brute-force feature extraction, and the features need to be tailored according to the relevant task. In this work, we extend a recent discriminative projection based feature selection method using the power of stochasticity to overcome local minima and to reduce the computational complexity. The proposed approach assigns weights both to groups and to features individually in many randomly selected contexts and then combines them for a final ranking. The efficacy of the proposed method is shown in a recent paralinguistic challenge corpus to detect level of conflict in dyadic and group conversations. We advance the state-of-the-art in this corpus using the INTERSPEECH 2013 Challenge protocol.**

***Index Terms*—CCA, computational paralinguistics, discriminative projection, feature selection, random projection.**

I. Introduction

COMPUTATIONAL paralinguistics is the study of the underlying message(s) from speech apart from the linguistic content. Partly due to maturity of automatic speech recognition (ASR) technology, but mostly due to the richness of the application domain, the field is evolving rapidly.

Applications of the research area include interactive and communicative robots [1], [2]; diagnosis, monitoring, and screening of diseases and speech disorders [3] such as Parkinson's disease [4], [5]. Recent paralinguistic challenges introduce biomedical corpora e. g. for autism detection/diagnosis [6], and depression level prediction [7], [8] in order to help advance the field by providing transparency and comparability with state-of-the-art studies.

State-of-the-art computational paralinguistics applications are built using suprasegmental features obtained from functionals (e. g. moments, extremes) operating on frame-level Low Level Descriptors (LLD) e. g. Fundamental Frequency (F_0), Mel-Frequency Cepstral Coefficients (MFCC), jitter, shimmer [9], [10]. Brute-force extraction of high-dimensional potent features is commonly encountered in competitive baseline feature sets of the most recent computational paralinguistics challenges [6], [7]. In such systems, feature extraction is not a bottleneck, since brute-forcing yields an over-complete feature set. However, this approach requires an elaborate task-relevant pruning of features. This issue comprises the research problem of this study.

This study extends a recent work that uses Canonical Correlation Analysis (CCA) as a ranking feature selector [11]. CCA is a statistical method to find linear bases that maximize mutual correlation between two sets of variables [12], [13]. It has been used for a variety purposes ranging from multi-view feature extraction [14], [15], to feature selection [11], [16] and regression [17]. Motivated by the success of [11] as well as its limitations, we propose the use of stochasticity to extend [11] by applying CCA between a random subset of features and then by aggregating the feature importance weighted with the canonical correlation value (feature group saliency). The approach is validated on a recent challenge corpus: INTERSPEECH 2013 Conflict sub-challenge, where we advance the state-of-the-art using only the audio modality.

The remainder of this work is organized as follows. In the next section, background on CCA and a brief review of relevant literature are given. In Section III we introduce the proposed method and the baseline approach. Section IV details the challenge corpus. In Section V experimental results are presented. Finally, Section VI concludes the work.

II. Background

A. Canonical Correlation Analysis

Proposed by Hotelling [12], CCA seeks to maximize the mutual correlation between two sets of variables by finding linear projections for each set. Mathematically, CCA seeks to maximize the mutual correlation between two views of the same semantic phenomenon (e. g. audio and video of a speech) denoted $X \in \mathbb{R}^{n \times d}$ and $Y \in \mathbb{R}^{n \times p}$, where n denotes the number of paired samples, via:

$$\rho(X,Y) = \max_{w,v} corr(Xw, Yv). \quad (1)$$

Here, "corr" corresponds to Pearson's correlation, and w and v correspond to the projection vectors of X and Y, respectively. Let C_{XY} denote the cross-set covariance between the sets X

Manuscript received September 04, 2014; revised October 14, 2014; accepted October 23, 2014. Date of publication October 27, 2014; date of current version November 10, 2014. This work was supported in part by Boğaziçi University Project BAP 6531. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mathew Magimai Doss.

H. Kaya, A. A. Salah, and F. Gürgen are with the Department of Computer Engineering, Boğaziçi University, Bebek, Istanbul 34342, Turkey (e-mail: heysem@boun.edu.tr; salah@boun.edu.tr; gurgen@boun.edu.tr).

T. Özkaptan is with UEKAE, TÜBİTAK, Gebze, Kocaeli, Turkey (e-mail: tugce.ozkaptan@tubitak.gov.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2365393

TABLE I
SUMMARY OF RECENT WORKS IN COMPUTATIONAL PARALINGUISTICS EMPLOYING/PROPOSING FEATURE SELECTION (FS) METHODS

Work	Paralinguistic Task	Method
Espinosa et al. (2011) [22]	Emotion	Bilingual Acoustic FS
Giannoulis and Potamianos (2012) [23]	Emotion	mRMR + SBE
Räsänen et al. (2013) [21]	Autism, Emotion and Level of conflict	Random Subset FS
Kirchhoff et al. (2013) [24]	Autism	Submodular FS
Moore et al. (2014) [19]	Emotion	Correlation FS
Kaya et al. (2014) [11]	Depression	CCA based FS
Bejani et al. (2014) [25]	Emotion	ANOVA based FS
Kim et al. (2014) [20]	Level of Conflict	Automatic Relevance Determination

and Y, and similarly let C_{XX} denote within set covariance for X. The problem given in eq. (1) can be re-formulated as:

$$\rho(X,Y) = \max_{w,v} \frac{w^T C_{XY} v}{\sqrt{w^T C_{XX} w \cdot v^T C_{YY} v}}. \tag{2}$$

The formulation in Eq. (2) can be converted into a generalized eigenproblem for both projections (i. e. w and v), the solution can be shown [13] to have the form of:

$$C_{XX}^{-1} C_{XY} C_{YY}^{-1} C_{YX} w = \lambda w, \tag{3}$$

where the correlation appears to be the square root of eigenvalue:

$$\rho(X,Y) = \sqrt{\lambda}. \tag{4}$$

To attain maximal correlation, the eigenvector corresponding to the largest eigenvalue in Eq. (3) should be selected. Similarly, by restricting the new vectors to be uncorrelated with the previous ones, it can be shown that the projection matrices for each set are spanned by the k eigenvectors corresponding to the k largest eigenvalues. It is important to note that the maximum number of covariates m in U_X and U_Y are limited with the matrix rank of X and Y:

$$m = min(rank(X), rank(Y)) \tag{5}$$

B. Literature Review

Since the primary focus of this study is computational paralinguistics, we provide a brief summary of recent paralinguistic works that utilize feature selection in Table I. These papers use different feature selection (FS) methods, like Correlation based Feature Selection [18], [19] and Automatic Relevance Determination (ARD) [20]. While some utilize existing methods, majority of the works listed in the table propose new feature selection methods that better target the specific domain.

Of particular relevance is Random Subset Feature Selection (RSFS), a method that is introduced in [21]. At each iteration, the algorithm selects a random feature set and then measures relevance of each feature based on the performance of the subset that the feature participates in. To compute the relevance, the authors increase the weights of features participating in a set providing higher than average performance by a predefined value p, and similarly reduce the weight by the same amount for the features performing lower than the average. Despite its success, feature and group level weighting are not handled very well, and the method is not scalable as it relies on thousands of simulations.

III. PROPOSED METHOD

A. Discriminative Projection Based Filters

The proposed method in this paper extends [11] by applying discriminative projection based ranking to random subsets of a large feature set. The main idea behind the Samples versus Labels CCA Filter (SLCCA-Filter) algorithm in [11] is as follows. When all features in one view are subjected to CCA against the labels on the other view, the absolute value of the projection matrix W can be used to rank the features. The application to regression is straightforward, since the resulting matrix is $n \times 1$, therefore a vector. It can be applied in the same way to binary classification, where the classes can be denoted with 0 and 1 in the target vector. For $K > 2$, we can use the canonical correlation value (ρ^i) to weight the corresponding projection column (eigenvector W^i). In short, the SLCCA-Filter algorithm, which takes as inputs a dataset $X \in \mathbb{R}^{n \times d}$ and a label matrix $T \in \{0,1\}^{n \times (K)}$ is given as:

$$[W, V, \rho] = CCA(X, T), \tag{6}$$

$$H = \sum_{i=1}^{m} abs(W^i)\rho^i, \tag{7}$$

$$R = \mathrm{sort}(\mathrm{H},' \mathrm{descend}',), \tag{8}$$

where the 1-of-K coded label matrix T is defined as

$$T_{j,k} = \begin{cases} 1 & \text{if } y_j = k, \\ 0 & \text{if } y_j \neq k, \end{cases} \tag{9}$$

and R is the output of feature ranking. Here, j and k index the instances and the classes, respectively. Since K classes have $K-1$ degrees of freedom, the rank of matrix T is $K-1$. Therefore it is possible to remove any of the columns from the 1-of-K coded matrix. The filter can be applied to Fisher Discriminant Analysis (FDA) or to its localized version LFDA [26] in a similar manner as we have shown in our recent study [16]. In FDA variants, instead of ρ^i, the square root of the corresponding eigenvalue λ^i is used to weight the projection matrix. Note that the approach can also be used for multi-task learning, both in classification and regression.

B. Random Discriminative Projection Based Filters

Though it looks efficient in suppressing redundant features, SLCCA-Filter has an important drawback, which gives the motivation to this study: the number of non-zero weight features in the projection is upper-bounded by the rank of the data matrix. When $d > n$, a pseudo-inverse operation takes the place of the inverse for the singular covariance matrix, and subsequently valuable information is lost.

By means of random sampling of features, it is possible to evaluate feature relevance/redundancy in different conditions and aggregate them to obtain a final ranking. While the absolute

Input:
X: n x d dimensional data matrix
T: target matrix n x (K-1) in classification, n x 1 in regression
L : Number of iterations
Require X, **T** and L as input.
p $\Leftarrow$ d / 2;
W $\Leftarrow$ zeros(d,1);
for $i = 1$ to L **do**
RandFeats $\Leftarrow$ randperm(d);
Feats $\Leftarrow$ RandFeats(1:p);
$\mathbf{X_{rand}} \Leftarrow \mathbf{X}$(:, Feats);
$\overline{\mathbf{X_{rand}}} \Leftarrow \mathbf{X}$(:, Complement Set of Feats);
Apply CCA between feature set **X-rand** and **T** using Eq. (6);
Apply CCA between complement set $\overline{\mathbf{X_{rand}}}$ and **T** using Eq. (6);
Compute the weight vectors $\mathbf{H_i}$ and $\overline{\mathbf{H_i}}$ for each projection using Eq. (7);
Combine $\mathbf{H_i}$ and $\overline{\mathbf{H_i}}$ to obtain d dimensional importance vector $\mathbf{w_i}$;
Accumulate weighted features in vector $\mathbf{W}$: $\mathbf{W} \Leftarrow \mathbf{W} + \mathbf{w_i}$;
end for
Obtain the sorted importance, $\mathbf{W_s}$ and feature ranking $\mathbf{R}$ by applying Eq. (8) on $\mathbf{W}$;

Fig. 1. Random SLCCA Algorithm.

value of feature projection matrix provides information about feature level importance (driven to zero if the feature is redundant or irrelevant), the square root of the eigenvalue in a discriminative projection can be used to weight how good the feature group collectively performs.

In our approach, at each iteration we project a random subset of $d/2$ features and its complement set, where d is dataset dimensionality. We then aggregate the absolute value of projection weights multiplied with corresponding eigenvalues. After L iterations, the accumulated feature importance vector is sorted in descending order to provide the ranking. With this approach, we can both access all features at each iteration, and also obtain compatible feature weights in the projection matrix. The proposed algorithm is given in Fig. 1.

If we only select $p << d$ for the projection without the complement set (as in the case of Random Forests [27]) the algorithm needs hundreds of iterations to include the majority of features. Thus, the proposed algorithm is expected to give better performance with much fewer iterations compared to sampling in Random Forests fashion. In our preliminary experiments, we verified that selecting $p << d$ results in poor performance when small values for p and L (in the range [10–100]) are used. On the other hand, if there is a great discrepancy between the dimensionality of the random set and its complement, the weights are incompatible, and the ranking would be misleading.

We expect the proposed algorithm to outperform SLCCA-Filter, since at each iteration: i) the covariance inversion becomes numerically more stable, ii) SLCCA-Rand finds $2 \times rank(X)$ non-zero-weight features as opposed to a maximum of $rank(X)$ features output by SLCCA-Filter, and iii) the feature saliencies are weighted by the canonical correlation values (collective goodness), which are not considered in SLCCA-Filter.

TABLE II
STATISTICS OF THE CONFLICT CORPUS

Property	Statistic
# of Clips	1430
# of Subjects	138
# of Females	23 (1 moderator , 22 participants)
# of Males	133 (3 moderator, 120 participants)
# of Political Debates	45
Mean Clip Duration	30 seconds
Conflict Score Range	$(-10, +10)$

TABLE III
PARTITIONING OF THE SSPNET CONFLICT CORPUS INTO TRAIN, DEVELOPMENT, AND TEST SETS FOR BINARY CLASSIFICATION [6]

#	train	dev	test	total
low	471	127	226	824
high	322	113	171	606
total	793	240	397	1430

IV. INTERSPEECH 2013 CONFLICT CORPUS

The INTERSPEECH 2013 Conflict Sub-Challenge [6] aims at automatically analyzing group discussions with the purpose of recognizing conflict. The subject is important since it involves dyadic speech and speaker group analysis in realistic everyday communication. The Conflict Sub-Challenge uses the "SSPNet Conflict Corpus" [28]. It contains political debates televised in Switzerland[1]. The statistics of the corpus are summarized in Table II.

The clips have been annotated following the process illustrated in [29] with respect to conflict level by roughly 550 assessors recruited via Amazon Mechanical Turk. Each clip is assigned a continuous conflict score in the range $[10, +10]$, giving rise to a regression task. For the challenge, a binary classification task is created based on these labels, namely to classify into 'high' (> 0) or 'low' (< 0) level of conflict. The distribution of instances among partitions (the challenge protocol) is given in Table III.

The challenge baseline acoustic feature set contains 6 373 features extracted via openSMILE [30] using 54 statistical functionals (e.g. moments, extremes, percentiles, polinomial regression coefficients) operating on 65 low-level descriptors (LLD). LLDs cover a wide range of popular Spectral (e.g. Rasta-style auditory spectrum bands 1-26, skewness, variance, spectral flux, centroid, and slope), Cepstral (MFCC 1-14), energy related (e.g. Root Mean Square Energy, Zero Crossing Rate) and voicing related (e.g. F0, jitter, shimmer) descriptors. The full list of LLDs and functionals can be found in [31] and in the supplementary material of this paper.

V. EXPERIMENTAL RESULTS

For the classification task in the Conflict corpus, we use Support Vector Machines with Linear Kernel, to maximize comparability with previous work on the same corpus. We use Random Forests (RF) to provide an independent classifier benchmark. RF is a combination of decision tree predictors, where each tree is grown with a random (sampled with replacement) set of N instances and a random subset of features [27].

[1]The clips are in French. The data are publicly available and can be accessed from http://sspnet.eu/2013/09/sspnet-conflict-corpus/

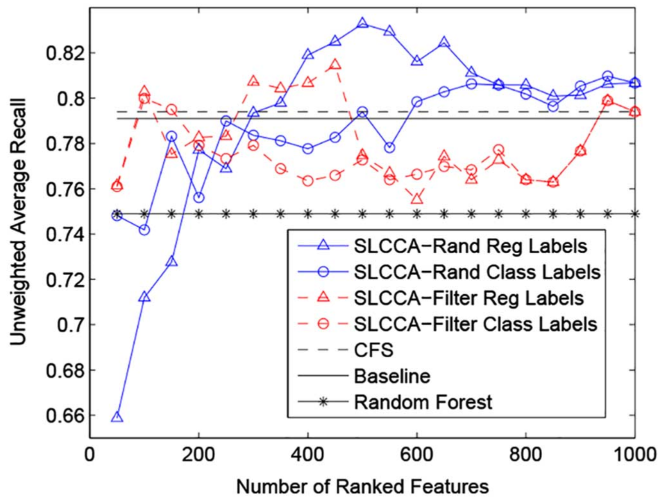

Fig. 2. Comparison of feature ranking learned from regression and classification labels in relation to challenge baseline [6] and best performance of CFS [18] as an independent feature selector and Random Forest [27] as an independent classifier.

TABLE IV
COMPARISON OF THE HIGHEST TEST SET UAR PERFORMANCES USING CONFLICT CORPUS WITH IS 2013 CHALLENGE PROTOCOL

Work	Best UAR (%)
Challenge baseline [6]	80.8
Grèzes et al. [34]	83.1
Räsänen et al. [21]	83.9
Proposed method	84.6
Random Forest [27]	78.6

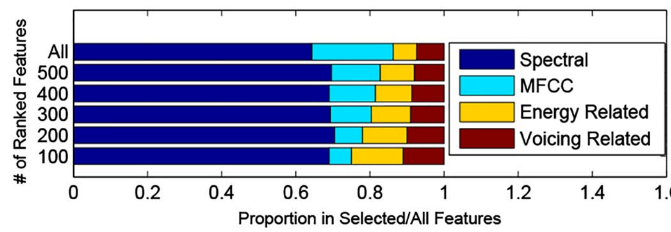

Fig. 3. Distribution of LLD categories w. r. t. number of ranked features.

RFs are known to generalize well and are successfully employed in high dimensional pattern recognition. We train SVM models with Platt's Sequential Minimal Optimization (SMO) algorithm [32]. We choose the SVM complexity parameter $\in 10^{\{-5,-4,-3,\ldots,2\}}$. For RF simulations, we use {10, 20, 30} trees each with a random feature dimensionality sampled in the range of [50, 1000] with steps of 50. For reproducibility, we set the seed of random number generator to one before simulations.

We use the WEKA [33] implementation of Correlation based Feature Selection (CFS) [18] with "Best First" search and SLCCA-Filter methods as independent benchmarks for the Conflict challenge. We employ Unweighted Average Recall (UAR), which is the mean of individual recalls, as primary evaluation measure:

$$UAR = \frac{1}{K}\sum_{k=1}^{K} TP(k)/P(k), \qquad (10)$$

where K is the number of classes; $TP(k)$ and $P(k)$ denote the number of true positive instances and total positive instances for class k, respectively. We carry out classification on selected features ranked using both continuous and discretized labels.

Fig. 2 summarizes the experiments on the training and development sets. The figure shows UAR performances of discretized (class) labels based versus continuous (regression) labels based ranking using SLCCA-Rand and SLCCA-Filter methods in relation to two other baselines. On the overall, we see that the best results are obtained with SLCCA-Rand (blue solid lines) using continuous labels. In the same vein, classification with features ranked by continuous labels are observed to provide better UAR scores than ranking by class labels. We observe that using continuous labels gives a smoother performance contour, improving feature selection for the test set. Moreover, in both types of labels, SLCCA-Rand achieves better performance than other benchmarks.

Finally, we evaluate the proposed method on the challenge test set using the setting that gives the best development set UAR performance. We restrict our test set trials to four: we use the first 500 features that yield the best development set results learned from the training set and the same number of features ranked by training and development sets, together with the two best SVM complexity parameters (0.01 and 0.001). Using the features learned from the training set, a UAR test set performance of 83.2% is reached. The UAR results improve to 84.6% when the proposed filter method is applied to the combined (training and development) set. The results achieved advance the state-of-the-art UAR (c. f. Table IV) on this corpus [21], without resorting to thousands of classification iterations used in [21].

When we analyze the distribution of SLCCA-Rand features yielding the best test set performance with respect to LLD categories, we observe higher proportion of energy- and voicing-related features among the top ranks, compared to MFCC features (c. f. Fig. 3).

VI. CONCLUSION

In this work, we proposed a novel feature selection approach that extends a recently introduced discriminative projection based filter. The proposed approach uses the power of stochasticity to overcome the curse of dimensionality by learning feature level and feature group level weights in a variety of random contexts. To maximize the comparison we use the baseline acoustic feature set with SVM Linear Kernel. Ranking features with the proposed method, we advance the state-of-the-art on the Conflict corpus using the INTERSPEECH 2013 challenge protocol. We observe that learning ranking using regression labels provides better results than using class labels both in SLCCA-Filter and in SLCCA-Rand. The decrease observed in performance with feature selection using class labels is attributed to loss of information during discretization. With regression labels, the continuity in feature space is better mapped to the continuous target variable. Utilizing other methods giving discriminative projections (such as SVM discriminant) and extension to kernel methods constitute our future works.

REFERENCES

[1] C. Martinez and A. Cruz, “Emotion recognition in non-structured utterances for human-robot interaction,” in *IEEE Int. Workshop Robot and Human Interactive Communication, 2005 ROMAN*, Aug. 2005, pp. 19–23.

[2] A. Batliner, S. Steidl, and E. Noth, “Associating children’s non-verbal and verbal behaviour: Body movements, emotions, and laughter in a human-robot interaction,” in *2011 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 5828–5831.

[3] J. Schoentgen, “Vocal cues of disordered voices: An overview,” *Acta Acustica United with Acustica*, vol. 92, no. 5, pp. 667–680, 2006.

[4] I. Rektorova, J. Barrett, M. Mikl, I. Rektor, and T. Paus, “Functional abnormalities in the primary orofacial sensorimotor cortex during speech in Parkinson’s disease,” *Movement Disorders*, vol. 22, no. 14, pp. 2043–2051, 2007.

[5] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, “Accurate telemonitoring of Parkinson’s disease progression by noninvasive speech tests,” *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 884–893, 2010.

[6] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, “The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism,” in *Proc. INTERSPEECH 2013* , Lyon, France, Aug. 2013, pp. 148–152.

[7] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, “AVEC 2013–the continuous audio/visual emotion and depression recognition challenge,” in *Proc. 3rd ACM Int. Workshop on Audio/Visual Emotion Challenge, ser. AVEC ’13*, 2013, pp. 3–10.

[8] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic, “AVEC 2014–3D dimensional affect and depression recognition challenge,” in *Proc. 4rd ACM Int. Workshop on Audio/Visual Emotion Challenge, ser. AVEC ’14*, 2014, pp. 3–10.

[9] B. Schuller, “Voice and Speech Analysis in Search of States and Traits,” in *Computer Analysis of Human Behavior*, A.A. Salah and T. Gevers, Eds. Berlin, Germany: Springer, 2011, pp. 227–253.

[10] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S. Narayanan, “Paralinguistics in speech and language–state-of-the-art and the challenge,” *Comput. Speech Lang., Special Issue on Paralinguistics in Naturalistic Speech and Language*, vol. 27, no. 1, pp. 4–39, 2013.

[11] H. Kaya, F. Eyben, A. A. Salah, and B. W. Schuller, “CCA based feature selection with application to continuous depression recognition from acoustic speech features,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 2014)*, 2014, pp. 3757–3761.

[12] H. Hotelling, “Relations between two sets of variates,” *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[13] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, “Canonical correlation analysis: An overview with application to learning methods,” *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, 2004.

[14] C. O. Sakar, “Multi-view feature extraction based on canonical correlation analysis,” Ph.D. dissertation, Inst. Grad. Studies Sci. Eng., Boğaziçi Univ., Istanbul, Turkey, Jan. 2014.

[15] H. Kaya and A. A. Salah, “Eyes whisper depression: A CCA based multimodal approach,” in *Proc. ACM Multimedia 2014*, Orlando, FL, USA, Nov. 2014, pp. 961–964, ACM.

[16] H. Kaya, T. Özkaptan, A. A. Salah, and S. F. Gürgen, “Canonical correlation analysis and local Fisher discriminant analysis based multi-view acoustic feature reduction for physical load prediction,” in *Proc. INTERSPEECH 2014*, Singapore, Sep. 2014, pp. 442–446, ISCA.

[17] H. Kaya, F. Çilli, and A. A. Salah, “Ensemble CCA for continuous emotion prediction,” in *Proc. 4th ACM Int. Workshop on Audio/Visual Emotion Challenge*, Orlando, FL, USA, Nov. 2014, pp. 19–26.

[18] M. A. Hall, “Correlation-based feature selection for machine learning,” Ph.D. dissertation., Univ. Waikato, Hamilton, New Zealand, 1999.

[19] J. Moore, L. Tian, C. Lai, and A. Gelbukh, “Word-level emotion recognition using high-level features,” in *Computational Linguistics and Intelligent Text Processing, ser. Lecture Notes in Computer Science*. Berlin/Heidelberg, Germany: Berlin/Heidelberg, 2014, vol. 8404, pp. 17–31.

[20] S. Kim, M. Filippone, F. Valente, and A. Vinciarelli, “Predicting continuous conflict perception with Bayesian Gaussian processes,” *IEEE Trans. Affect. Comput.*, vol. 5, no. 2, pp. 187–200, 2014.

[21] O. Räsänen and J. Pohjalainen, “Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech,” in *Proc. INTERSPEECH 2013*, Lyon, France, Aug. 2013, pp. 210–214, ISCA.

[22] H. Espinosa, J. Garcia, and L. Pineda, “Bilingual acoustic feature selection for emotion estimation using a 3d continuous model,” in *2011 IEEE Int. Conf. Automatic Face Gesture Recognition and Workshops (FG 2011)*, Mar. 2011, pp. 786–791.

[23] P. Giannoulis and G. Potamianos, “A hierarchical approach with feature selection for emotion recognition from speech,” in *Proc. Eighth Int. Conf. Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey, May 2012, pp. 1203–1206, ELRA.

[24] K. Kirchhoff, Y. Liu, and J. Bilmes, “Classification of developmental disorders from speech signals using submodular feature selection,” in *Proc. Interspeech 2013*, Lyon, France, Aug. 2013, pp. 187–190, ISCA.

[25] M. Bejani, D. Gharavian, and N. Charkari, “Audiovisual emotion recognition using ANOVA feature selection method and multi-classifier neural networks,” *Neural Comput. Applicat.*, vol. 24, no. 2, pp. 399–412, 2014.

[26] M. Sugiyama, “Local Fisher discriminant analysis for supervised dimensionality reduction,” in *Proc. 23rd Int. Conf. Machine Learning, ser. ICML ’06*, New York, NY, USA, 2006, pp. 905–912, ACM.

[27] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[28] S. Kim, M. Filippone, F. Valente, and A. Vinciarelli, “Predicting the conflict level in television political debates: An approach based on crowdsourcing, nonverbal communication and Gaussian processes,” in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 793–796.

[29] A. Vinciarelli, S. Kim, F. Valente, and H. Salamin, “Collecting data for socially intelligent surveillance and monitoring approaches: The case of conflict in competitive conversations,” in *Int. Symp. Communications, Control, and Signal Processing*, 2012.

[30] F. Eyben, M. Wöllmer, and B. Schuller, “Opensmile: The Munich versatile and fast open-source audio feature extractor,” in *Proc. Int. Conf. ACM Multimedia*, 2010, pp. 1459–1462.

[31] F. Eyben, F. Weninger, and B. Schuller, “Affect recognition in real-life acoustic conditions-a new perspective on feature selection,” in *Proc. Interspeech 2013*, Lyon, France, Aug. 2013, pp. 2044–2048, ISCA.

[32] J. Platt *et al.*, “Sequential minimal optimization: A fast algorithm for training support vector machines,” Tech. Rep. MSR-TR-98-14, 1998 [Online]. Available: http://web.iitd.ac.in/~sumeet/tr-98-14.pdf

[33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: An update,” *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.

[34] F. Grèzes, J. Richards, and A. Rosenberg, “Let me finish: Automatic conflict detection using speaker overlap,” in *Proc. Interspeech 2013*, Lyon, France, Aug. 2013, pp. 200–204, ISCA.