Snore Sound Recognition: On Wavelets and Classifiers from Deep Nets to Kernels

Kun Qian[1,2], Christoph Janott[3], Jun Deng[1,4], Clemens Heiser[5], Winfried Hohenhorst[6], Michael Herzog[7], Nicholas Cummins[1], and Björn Schuller[1,8]

***Abstract*— In this paper, we present a comprehensive comparison of wavelet features for the classification of snore sounds. Wavelet features have proven to be efficient in our previous work; however, the benefits of wavelet transform energy (WTE) and wavelet packet transform energy (WPTE) features were not clearly established. In this study, we firstly present our updated snore sounds database, expanded from 24 patients (collected by one medical centre) to 40 patients (collected by three medical centres). We then study the effects of varying frame sizes and overlaps for extraction of the wavelet low-level descriptors, the effect of which have yet to be fully established. We also compare the performance of the WTE and WPTE features when fed into multiple classifiers, namely, Support Vector Machines (SVM), *K*-Nearest Neighbours, Linear Discriminant Analysis, Random Forests, Extreme Learning Machines, Kernel Extreme Learning Machines, Multilayer Perceptron, and Deep Neural Networks. Key results presented indicate that, when fed into a SVM, WTE outperforms WPTE (one-tailed z-test, $p < 0.002$). Further, WPTE can achieve a significant improvement when trained by a *k*-nearest neighbours classifier (one-tailed z-test, $p < 0.001$).**

I. INTRODUCTION

Affecting 13 % of men and 6 % of women in the US population [1], *Obstructive Sleep Apnoea* (OSA) is a chronic sleep-related disorder which increases the risk for cardiovascular diseases [2], hypertension [3], and stroke [4]. Snoring, as a common symptom of OSA (reported in more than 80 % of OSA patients [5]), has been studied to find an acoustic-based, non-invasive method for diagnosing OSA [6], [7].

On the other hand, studies which aim to identify the *site of vibration and obstruction* in the upper airways during snoring are quite limited. A recent literature review on this subject identified a total of eight publications [8]. In medical practice, understanding the mechanisms of snore generation is helpful for *Ear, Nose, and Throat* (ENT) surgeons when planning a targeted surgical intervention.

In terms of automatic acoustic analysis, frequency features [9], [10], amplitude features [11], statistical time series features [12], and psychoacoustic features [13] have been analysed for their potential to distinguish different *snore sound* (SnS). However, state-of-the-art machine learning techniques were not employed in these studies.

II. RELATION TO PRIOR WORK

Wavelet features have been explored for the classification of different SnS types [14]. The authors found that wavelets had a superior performance for this task over other frequently-used features such as *Mel-frequency cepstral coefcients* (MFCCs), formants and power ratios. However, the data used was small (only 24 patients, without a development set). A *bag-of-audio-words* approach, which combined MFCC, formant and wavelet features, was recently shown to improve on the performance of SnS classification when compared to (non-bagged) wavelets [15]. Again, these results are limited by the size of the associated data set.

In this study, we firstly expand the database used in [14], [15] to a larger size (from 24 patients' data collected by one medical centre to 40 patients' data provided by three medical centres). Secondly, we investigate the effects of varying wavelet frame sizes and overlaps on classification performance of two kinds of wavelet features. Thirdly, in [14], only *Support Vector Machines* (SVMs) were employed for the classification task; the performance of other machine learning models was not tested. Therefore, in this work, we compare the two feature sets' classification capacity by feeding them into a range of different classifiers.

The rest of this paper is organised as follows: The SnS database and methods used are introduced in Section III. The experimental results and discussion follow in Section IV, and Section V, respectively. Section VI contains the conclusions.

III. MATEARIALS AND METHODS

A. Snore Sounds Database

We collected SnS data of 40 patients, from 3 clinical sites: Klinikum rechts der Isar, Munich, Germany; Alfried Krupp Hospital, Essen, Germany; and, University Hospital Halle (Saale), Germany, during a drug-induced sleep endoscopy [16]. The demographic information of the patients is shown in Table I.

An ENT expert labelled the SnS data according to their excitation location using the classes *the velum* (V), *the*

[1]Kun Qian, Jun Deng, Nicholas Cummins and Björn Schuller are with the Chair of Complex & Intelligent Systems, Universität Passau, Germany {nicholas.cummins}@uni-passau.de

[2]Kun Qian is also with the Machine Intelligence & Signal Processing group, Technische Universität München, Germany andykun.qian@tum.de

[3]Christoph Janott is with the Institute for Medical Engineering, Technische Universität München, Germany c.janott@gmx.net

[4]Jun Deng is also with audEERING GmbH, Gilching, Germany jdeng@audeering.com

[5]Clemens Heiser is with the Department of Otorhinolaryngology/Head and Neck Surgery, Technische Universität München, Germany clemens.heiser@tum.de

[6]Winfried Hohenhorst is with the Clinic for ENT Medicine, Head and Neck Surgery, Alfried Krupp Krankenhaus, Essen, Germany mail@hohenhorst.com

[7]Michael Herzog is with the Clinic for ENT Medicine, Head and Neck Surgery, Cottbus, Germany m.herzog@ctk.de

[8]Björn Schuller is also with the Department of Computing, Imperial College London, U.K. schuller@ieee.org

TABLE I

THE DEMOGRAPHIC INFORMATION OF THE 40 PATIENTS WHICH COMPRISE OUR SNS DATASET. BMI: BODY MASS INDEX; AHI: APNEA HYPOPNEA INDEX.

	MEAN	STD	RANGE
Age (*years*)	47.4	11.50	26–71
BMI (kg/m^2)	26.9	3.06	21.2–38.4
AHI (*events/h*)	21.7	12.77	1.3–59.1

TABLE II

NUMBER OF SUBJECT INDEPENDENT SEGMENTS FOR EACH SNORE-TYPE IN OUR SNS DATASET.

	train	dev	test	Σ
V	363/[7]	104/[2]	152/[2]	619/[11]
O	326/[7]	125/[2]	122/[2]	573/[11]
T	289/[4]	90/[2]	78/[2]	457/[8]
E	323/[6]	96/[2]	148/[2]	567/[10]
Σ	1 301/[24]	415/[8]	500/[8]	2 216/[40]

oropharyngeal area including the palatine tonsils (O), *the tongue base* (T) and the *epiglottis* (E). Details on the data collection process and the anatomical positions of V, O, T, and E in the upper airway can be found in [17]. We segmented the original snoring events into 200 ms units with an overlap of 50 %. We partitioned the collected data into subject independent *train*, *development* (dev), and *test* sets as displayed in Table II.

B. The WTE and WPTE Feature Sets

This section gives a brief overview of the *wavelet transform energy* (WTE), and the *wavelet packet transform energy* (WPTE) features; for a more detailed description the reader is referred to [17]. Khushaba et al. first employed WPTE for classification of drowsiness levels based on EEG, EOG, and ECG signals [18]. WTE is the energy related features generated by the *wavelet transform* (WT), whereas WPTE is the coefficients generated by *wavelet packet transform* (WPT). Compared with the WT, the WPT decomposes both the approximation and the detailed part of the original analysed signal [19]; i. e., WT uses low pass filters only, while WPT uses both low pass filters and high pass filters.

For the WTE features, we generate $4 \times (J_{max} + 1)$ features as low-level descriptors (LLDs), where J_{max} is the maximum decomposition level of any particular wavelet function ('sym3' in this study[1]). For the WPTE features, we generate $2^{J_{max}+1} - 1$ LLDs. In addition, we apply confined functionals to the LLDs, i. e., the maximum, mean and minimum values, and the bias of the estimated linear regression of the frame-level features in one segment, which have proven to be efficient in [17]. Thus, in total we extract $16 \times (J_{max} + 1)$ WTE features and $4 \times (2^{J_{max}+1} - 1)$ WPTE features. Before being fed into a classifier, all the original extracted feature are normalised into a scale of [0, 1].

[1]The wavelet function names and the decomposition scripts are based on the Wavelet Toolbox of Matlab by MathWorks® (http://www.mathworks.com/products/wavelet/).

TABLE III

INFORMATION OF EACH FEATURE SET.

	16 ms	32 ms	64 ms
J_{max}	5	6	7
WTE Dimension	96	112	128
WPTE Dimension	252	508	1 020
WTE+WPTE Dimension	348	620	1 148

TABLE IV

THE MAIN OPERATING PARAMETERS AND GRID SEARCH SPACE FOR EACH CLASSIFIER TESTED.

Classifiers	**Main Parameters**
SVM	kernels: 'linear', 'polynomial', 'radial basis function', 'sigmoid'; C-value: 10^{-5}, 10^{-4}, $\cdots$, 10^4, 10^5
K-NN	K-value: 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100; distance metrics: 'euclidean', 'cityblock', 'chebychev', 'correlation', 'cosine', 'hamming', 'jaccard', 'minkowski', 'seuclidean', 'spearman'
LDA	discriminant type: 'linear','diaglinear','pseudolinear'; gamma: 0:0.05:1.00
RF	number of trees: 2^1, 2^2, $\cdots$, 2^9, 2^{10}; fraction for the treebagger: 0.1:0.1:1.00
ELM	activation functions: 'signmoidal', 'sine', 'hardlim', 'tribas', 'radbas'; number of hidden neurons: 2^1, 2^2, $\cdots$, 2^{14}
KELM	kernels: 'radial basis function', 'linear', 'polynomial', 'wavelet'; regularization coefficients: 10^{-5}, 10^{-4}, $\cdots$, 10^4, 10^5
MLP	two hidden layers; neurons: 2^1, 2^2, $\cdots$, 2^9, 2^{10}
DNN	structured by two-layer stacked auto-encoders, neurons: [64 64], L_2: 10^{-3}, $\cdots$, 10^3; Sparsity Proportion: 0.1:0.1:0.9; Sparsity Regularization: 2

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

We conduct two comparative tasks on the WTE and WPTE features to investigate the effects on classification performance by (i) varying frame sizes and overlaps when extracting the LLDs, and (ii) by varying the classifier used. For the frame sizes and overlaps test we conducted a grid search using frame sizes of 16 ms, 32 ms, and 64 ms and overlaps of 25 %, 50 %, and 75 % resulting in nine different combinations.

The decomposition levels and the dimensions of each kind of wavelet feature is given in Table III. It can be seen that the WPTE feature set has a much larger dimensionality than the WTE set. In particular, at a decomposition level of 7, the dimensionality of the WTE features is approximately 10 % of the WPTE features.

For classifiers, we test and compare: Support Vector Machines (SVM) [20], k-Nearest Neighbours (k-NN) [21], Linear Discriminant Analysis (LDA) [21], Random Forests (RF) [22], Extreme Learning Machines (ELM) [23], Kernel Extreme Learning Machines (KELM) [24], Multilayer Perceptrons (MLP) [25], and Deep Neural Networks (DNN) [26]. Table IV shows the main operating parameters and grid search space for each classifier. All the parameters are optimised by the dev data set, with all stated dev data set results being the best performance from among these settings. For the comparison of frame sizes and overlaps, a SVM (implemented by LIBSVM [27]) is employed as the

TABLE V

THE UNWEIGHTED AVERAGE RECALLS [%] ACHIEVED BY WTE AND WPTE FEATURE SETS WITHIN VARIED FRAME SIZES AND OVERLAPS. THE CLASSIFIER IS A SVM AND PARAMETRES WERE OPTIMISED BY THE DEV SET.

	Frame Size Overlap	16 ms 25 %	16 ms 50 %	16 ms 75 %	32 ms 25 %	32 ms 50 %	32 ms 75 %	64 ms 25 %	64 ms 50 %	64 ms 75 %
WTE	train vs dev	50.0	51.3	50.6	46.2	46.8	46.6	45.0	44.8	45.8
	train vs test	61.4	55.5	65.2	65.2	60.1	57.8	**66.3**	65.8	64.7
	train+dev vs test	58.7	55.3	65.3	61.3	51.8	54.4	54.5	61.9	59.9
	mean	56.7	54.0	**60.4**	57.6	52.9	52.9	55.3	57.5	56.8
	std	±5.96	±2.37	±8.46	±10.04	±6.72	±5.74	±10.67	±11.17	±9.82
WPTE	train vs dev	50.8	49.2	48.6	54.9	53.3	60.1	52.2	53.6	50.2
	train vs test	44.9	49.8	35.0	**50.1**	49.2	45.4	48.7	48.9	39.3
	train+dev vs test	48.4	43.0	35.1	45.2	46.2	47.0	49.6	47.7	47.8
	mean	48.0	47.3	39.6	50.1	49.6	**50.8**	50.2	50.1	45.8
	std	±2.97	±3.76	±7.82	±4.85	±3.56	±8.06	±1.82	±3.12	±5.73
WTE+WPTE	train vs dev	50.1	50.2	50.0	49.7	50.8	51.0	50.7	49.9	50.4
	train vs test	55.5	49.3	54.5	46.9	51.6	50.3	51.9	47.5	44.8
	train+dev vs test	62.7	59.3	**63.7**	48.1	47.5	49.0	48.8	49.7	47.4
	mean	**56.1**	52.9	**56.1**	48.2	50.0	50.1	50.5	49.0	47.5
	std	±6.32	±5.53	±6.98	±1.40	±2.17	±1.01	±1.56	±1.33	±2.80

TABLE VI

UNWEIGHTED AVERAGE RECALLS [%] ACHIEVED BY WTE AND WPTE FEATURE SETS WITHIN VARIED CLASSIFIERS. WTE FRAME SIZE: 16 MS, OVERLAP: 75 %; WPTE FRAME SIZE: 32 MS, OVERLAP: 75 %.

	Classifiers	SVM	*K*-NN	LDA	RF	ELM	KELM	MLP	DNN
WTE	train vs dev	50.6	51.7	53.1	47.0	51.9	52.5	54.1	50.1
	train vs test	65.2	52.4	57.0	58.3	56.6	52.8	47.4	41.6
	train+dev vs test	**65.3**	53.7	54.7	58.5	46.5	52.1	48.3	44.4
	mean	**60.4**	52.6	54.9	54.6	51.7	52.5	49.9	45.4
	std	±8.46	±1.01	±1.96	±6.58	±5.05	±0.35	±3.64	±4.33
WPTE	train vs dev	60.1	50.0	50.9	49.4	52.7	51.2	61.2	59.4
	train vs test	45.4	**61.9**	58.9	43.0	43.1	43.6	43.8	48.3
	train+dev vs test	47.0	**61.9**	52.4	42.5	43.1	42.7	44.6	44.1
	mean	50.8	**57.9**	54.1	45.0	46.3	45.8	49.9	50.6
	std	±8.06	±6.87	±4.25	±3.85	±5.54	±4.67	±9.82	±7.91
WTE+WPTE	train vs dev	49.9	55.7	52.0	49.8	57.3	55.2	61.4	57.0
	train vs test	52.0	61.1	**64.6**	48.7	33.3	40.5	52.0	40.4
	train+dev vs test	52.1	62.5	64.0	60.6	43.3	36.0	50.7	25.0
	mean	51.3	59.8	**60.2**	53.0	44.6	43.9	54.7	40.8
	std	±1.24	±3.59	±7.11	±6.58	±12.06	±10.04	±5.84	±16.00

classifier because of its robust classification performance. Due to the imbalanced distribution of the SnS data, we use the unweighted average recall (UAR), i. e., the averaged accuracy by each individual class, as the evaluation metric. To test the significance level between results from different experimental configuration, a one-tailed z-test [28] is performed.

B. Results

As can be seen in Table V, the frame size and overlap used when exacting the wavelet LLDs impacts SVM classification performance. Considering the mean performance, the optimal configuration for the WTE and WPTE features are 16 ms (75 % overlap), and 32 ms (75 % overlap), respectively. When comparing the strongest test set performances the WTE features achieved a UAR of 16.2 percent points higher than the WPTE features ($p < 0.001$). The fusion of the two feature sets results in a slight decrease compared with the WTE features alone, which achieved a best UAR of 63.7 %, and a mean UAR of 56.1 %.

As can be seen in Table VI, the best results were gained either with a SVM or a k-NN classifier. When compared with SVM, k-NN significantly improves the performance of the WPTE features (mean UAR of 50.8 % vs 57.9 %, $p < 0.05$). Further, even though the results are close to each other, the WTE features achieved the strongest performance (mean UAR at 60.4 % – SVM), followed by the fusion of the two features (mean UAR at 60.2 % – LDA), and then the WPTE features (mean UAR at 57.9 % – k-NN). Taking feature dimensionality into account (see Table III), these results indicate that WTE uses less features to capture more SnS information when compared with WPTE.

V. DISCUSSION

Results gained in this study indicate that changing the frame size and overlap when extracting the underlying wavelet LLDs impacts the performance of a SVM classifier (see Table V). For the WTE features, the UARs vary from 44.8 % to 66.3 % ($p < 0.001$), while for the WPTE features, the UARs vary from 35.0 % to 60.1 % ($p < 0.001$).

Further results of interest are that, when combined with a SVM classifier, the WTE feature set outperformed the WPTE features on the test set (maximum UAR: 66.3 % vs 50.1 %, $p < 0.001$), and the fusion of the two feature sets achieves no significant improvement (maximum UAR is 63.7 %). For the WPTE features, the use of a k-NN classifier increases test set performance by 11.8 percent points compared with the SVM ($p < 0.001$).

When looking at the mean UARs (see Table VI), SVM, k-NN, and LDA are the best suited classifiers for the WTE features, the WPTE features, and their fusion respectively. It is worth noting that for the WTE features, the SVM's mean UAR (60.4 %) is higher than that achieved by the LDA's mean UAR (54.9 %) at a significant level ($p < 0.05$). Finally, most likely limited by the small size of the SnS corpus, the DNN and MLP classifiers did not prove to be efficient.

VI. Conclusion

In this paper we conducted a comparative work to evaluate two kinds of wavelet features; wavelet transform energy (WTE), and wavelet packet transform energy (WPTE), for their performance when classifying snore sounds. Key results indicate that, the frame size, and overlaps, can effect the final classification performance for both of the two wavelet features. When used in combination with a Support Vector Machine, the WTE features can outperform the WPTE features at a significant level of $p < 0.002$ (one-tailed z-test). Further, the WPTE features can reach up to an UAR at 61.9 % when combined with a k-Nearest Neighbour classifier.

We observed that state-of-the-art machine learning methods including Extreme Learning Machines, and Deep Neural Networks were not as efficient when compared to conventional classifiers (e. g., SVM); however, given the the small size of the database used, this may be less surprising. In future work, we will collect more snore sounds data from a larger population of patients, and re-investigate the potential of wavelet features combined with deep neural networks.

Acknowledgment

This work was partly supported by the China Scholarship Council (CSC), and the European Unions's Seventh Framework under grant agreement No. 338164 (ERC StG iHEARu).

References

[1] P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults," *American Journal of Epidemiology*, vol. 177, no. 9, pp. 1006–1014, 2013.

[2] J. M. Marin, S. J. Carrizo, E. Vicente, and A. G. Agusti, "Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study," *The Lancet*, vol. 365, no. 9464, pp. 1046–1053, 2005.

[3] P. E. Peppard, T. Young, M. Palta, and J. Skatrud, "Prospective study of the association between sleep-disordered breathing and hypertension," *New England Journal of Medicine*, vol. 342, no. 19, pp. 1378–1384, 2000.

[4] H. K. Yaggi, J. Concato, W. N. Kernan, J. H. Lichtman, L. M. Brass, and V. Mohsenin, "Obstructive sleep apnea as a risk factor for stroke and death," *New England Journal of Medicine*, vol. 353, no. 19, pp. 2034–2041, 2005.

[5] M. S. Aldrich, *Sleep Medicine*. New York, NY, USA: Oxford University Press, 1999.

[6] D. Pevernagie, R. M. Aarts, and M. De Meyer, "The acoustics of snoring," *Sleep Medicine Reviews*, vol. 14, no. 2, pp. 131–144, 2010.

[7] A. Roebuck, V. Monasterio, E. Gederi, M. Osipov, J. Behar, A. Malhotra, T. Penzel, and G. Clifford, "A review of signals used in sleep analysis," *Physiological Measurement*, vol. 35, no. 1, pp. R1–R57, 2014.

[8] C. Janott, B. Schuller, and C. Heiser, "Acoustic information in snoring noise," *HNO*, vol. 65, no. 2, pp. 107–116, 2017.

[9] S. Miyazaki, Y. Itasaka, K. Ishikawa, and K. Togawa, "Acoustic analysis of snoring and the site of airway obstruction in sleep related respiratory disorders," *Acta Oto-Laryngologica*, vol. 118, no. 537, pp. 47–51, 1998.

[10] S. Agrawal, P. Stone, K. McGuinness, J. Morris, and A. Camilleri, "Sound frequency analysis and the site of snoring in natural and induced sleep," *Clinical Otolaryngology & Allied Sciences*, vol. 27, no. 3, pp. 162–166, 2002.

[11] P. Hill, B. Lee, J. Osborne, and E. Osman, "Palatal snoring identified by acoustic crest factor analysis," *Physiological Measurement*, vol. 20, no. 2, pp. 167–174, 1999.

[12] R. J. Beeton, I. Wells, P. Ebden, H. Whittet, and J. Clarke, "Snore site discrimination using statistical moments of free field snoring sounds recorded during sleep nasendoscopy," *Physiological Measurement*, vol. 28, no. 10, pp. 1225–1236, 2007.

[13] M. Herzog, S. Plößl, A. Glien, B. Herzog, C. Rohrmeier, T. Kühnel, S. Plontke, and P. Kellner, "Evaluation of acoustic characteristics of snoring sounds obtained during drug-induced sleep endoscopy," *Sleep and Breathing*, pp. 1–9, 2014.

[14] K. Qian, C. Janott, Z. Zhang, C. Heiser, and B. Schuller, "Wavelet features for classification of vote snore sounds," in *Proc. of the IEEE ICASSP*. Shanghai, China: IEEE, 2016, pp. 221–225.

[15] M. Schmitt, C. Janott, V. Pandit, K. Qian, C. Heiser, W. Hemmert, and B. Schuller, "A bag-of-audio-words approach for snore sounds excitation localisation," in *Proc. of the ITG*. Paderborn, Germany: VDE, 2016, pp. 230–234.

[16] E. J. Kezirian, W. Hohenhorst, and N. de Vries, "Drug-induced sleep endoscopy: the vote classification," *European Archives of Oto-Rhino-Laryngology*, vol. 268, no. 8, pp. 1233–1236, 2011.

[17] K. Qian, C. Janott, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, "Classification of the excitation location of snore sounds in the upper airway by acoustic multi-feature analysis," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, 2017, in press.

[18] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 121–131, 2011.

[19] R. R. Coifman, Y. Meyer, and V. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and their Applications*. Sudbury: MA: Jones and Barlett, 1992, pp. 153–178.

[20] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[21] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[22] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[23] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[24] G.-B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels," *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.

[25] I. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3–31, 2000.

[26] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[27] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[28] M. R. Spiegel, J. J. Schiller, R. A. Srinivasan, and M. LeVan, *Probability and Statistics*. New York, NY, USA: McGraw-Hill, 2009.