AI-POWERED SPEECH ANALYSIS FOR ALCOHOL INTOXICATION DETECTION

G. Costantini[1], V. Cesarini[1], F.Amato[2], G. Olmo[2], G. Saggio[1]

[1] Department of Electronic Engineering, University of Rome Tor Vergata
Via del Politecnico 1, 00133 - Rome, Italy

[2] Department of Control and Computer Engineering, Polytechnic University of Turin
Corso Castelfidardo, 34/D, 10138 - Turin, Italy

This study investigated the feasibility of detecting alcohol intoxication through automatic speech analysis, addressing a crucial need for social prevention and safety.

We employed the Alcohol Language Corpus (ALC) by the Bavarian archive of Speech Signals (BAS) [1], comprising recordings from 162 individuals captured both sober and intoxicated in a realistic vehicle setting using head-mounted microphones. We opted to use a tongue-twister as it was proven to be the most effective task for sober-vs-drunk classification.
We conducted stratified statistical analyses to examine how alcohol alters speech production and how factors like gender, age, and drinking habits influence speech patterns. We used two AI approaches to classify, one based on a feature extraction + selection + traditional Machine Learning classifier pipeline, and the other based on a custom-made speaker-independent Discriminative Adversarial Neural Network (DANN) architecture, trained to optimize drunk-vs-sober classification while minimizing the effect of the single speakers (individual differences). The latter was also used on an extended dataset comprised of all sentences uttered by the speakers, and not just the tonguetwister.

Our pre-processing involved silence removal, normalization, and extraction of 6385 acoustic features using OpenSmile and Praat algorithms [2]. Through differential analysis, we identified a subset of features, notably RASTA-filtered auditory spectrum components, that showed systematic variation after intoxication across subjects.

The initial classification experiment using tongue-twister tasks and classical machine learning models like SVM, Gradient Boosting, and Random Forest yielded balanced accuracies around 70%, with SVM achieving the highest at 70.7%. However, we recognized the limitations of task-specific models and moved toward building a task- and subject-independent classifier. By applying DANN with a domain-adversarial training strategy, we minimized inter-speaker variability, improving generalization across unseen individuals and speech prompts. Our DANN model [3] achieved a balanced accuracy of 70.9% on an independent test set, slightly outperforming previous best studies like Bone et al. (2011) which had achieved 70.5%. We observed that features extracted from RASTA-filtered spectra and low-level spectral properties were consistently among the most discriminative, confirming that alcohol significantly affects speech timing, rhythm, and voice stability. Compared to earlier works, our approach reduced complexity, avoided intensive feature normalization steps, and created a general-purpose model without requiring prior knowledge about the speaker or spoken content [4].

The results show that speech can effectively be used for automatic alcohol detection, providing a low-cost, non-invasive method suitable for integration into automotive, law enforcement, or personal monitoring systems. In conclusion, traditional ML yielded good performances reaching the state of the art of almost 71% using a SVM applied to a selected features subset, but our custom DANN allowed to reach similar performances on a much larger dataset, namely the whole ALC corpus.

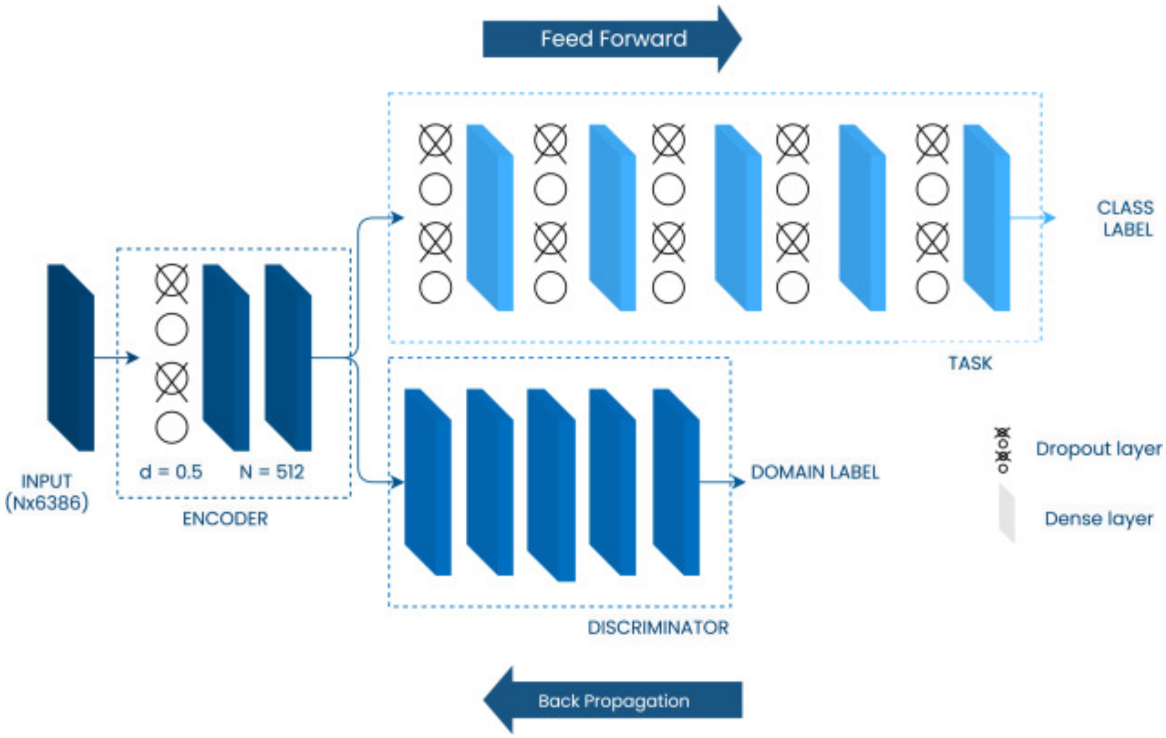

Fig. 1 - Domain-Adversary Neural Network architecture

TABLE I – Classification results of traditional ML classifiers applied on the tongue-twister task.

	SVM	**GB**	**RF**
Balanced accuracy	0.707	0.682	0.682
Precision	0.718	0.692	0.674
F1 score	0.7	0.675	0.69
Sensitivity	0.683	0.658	0.707
Specificity	0.732	0.707	0.658
AUC	0.707	0.682	0.682

References

[1] F. Schiel, C. Heinrich, S. Barfüßer, and T. Gilg, 'ALC: Alcohol Language Corpus', in *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco: European Language Resources Association (ELRA), May 2008.

[2] B. Schuller *et al.*, 'The INTERSPEECH 2016 computational paralinguistics challenge: 17th Annual Conference of the International Speech Communication Association, INTERSPEECH 2016', *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 08-12-September-2016, pp. 2001–2005, 2016, doi: 10.21437/Interspeech.2016-129.

[3] Y. Ganin *et al.*, 'Domain-Adversarial Training of Neural Networks', May 26, 2016, *arXiv*: arXiv:1505.07818. doi: 10.48550/arXiv.1505.07818.

[4] F. Amato, V. Cesarini, G. Olmo, G. Saggio, and G. Costantini, 'Beyond breathalyzers: AI-powered speech analysis for alcohol intoxication detection', *Expert Systems with Applications*, vol. 262, p. 125656, Mar. 2025, doi: 10.1016/j.eswa.2024.125656