A Deep Ensemble Learning Method for Automatic Classification of Multiplets in 1D NMR Spectra

G. Fischetti[1], N. Schmid[2,3], S. Bruderer[4], F. Paruzzo[4], G. Toscano[4], D. Graf[4], M. Fey[4], A. Henrici[2], A. Scarso[1], G. Caldarelli[1], B. Heitmann[4], and D. Wilhelm[2]

[1]Ca' Foscari University of Venice, Italy
[2]Zurich University of Applied Sciences (ZHAW), Switzerland
[3]University of Zurich (UZH), Switzerland
[4]Bruker BioSpin, Switzerland

Introduction

In the last decades, NMR scientists have turned to deep learning to automatize the resonances characterization process of NMR spectra (*1*) to obtain a higher level of robustness and reproducibility of the results, while speeding up the workflow. Here, we present a novel supervised deep learning method to perform automatic detection and classification of multiplets in 1H NMR spectra. We show that an ensemble of deep convolutional neural networks is able to effectively discriminate between non-overlapping and overlapping resonances while predicting the class of non-overlapping resonances.

Input

- **Training:** 100000 synthetic spectra with non-overlapping multiplets
- **Testing:** 10 experimental 1H NMR spectra of small molecules

Networks Ensemble

- **Networks' architecture:**
  - ⟶ Inception-like module with 1D Convolutional layers
  - ⟶ Long Short Term Memory layer
  - ⟶ Softmax output
- **Networks' initialization:**
  - ⟶ different weights from the same distribution (**He Normal**)
- **Networks' training:**
  - ⟶ different synthetic training set, with same statistical properties

Variance Computation

The **variance Δ** across networks' outputs was measured with the following expression (*3*):

$$\Delta = -\sum_{c=1}^{n_{classes}} p_c \log(p_c),$$

$$p_c = \frac{1}{M}\sum_{i=1}^{M} \mathrm{Softmax}(f^{\hat{W}_t}(x)).$$

An optimal threshold t was found so that:

- if $\Delta < t$, the outputs are aggregated with the **majority voting criteria**;
- if $\Delta \geq t$, an **out-of-distribution resonance** was detected.

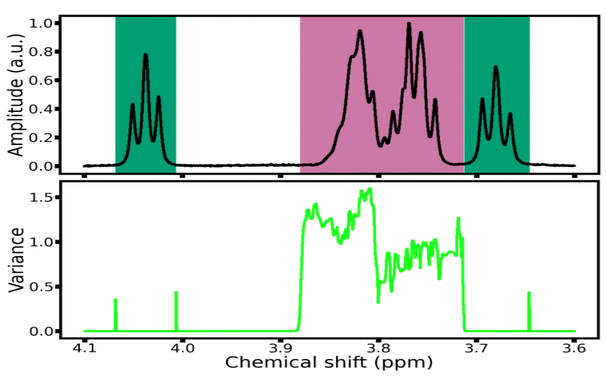

Figure 2: **Variance computation**: segment of a spectrum with class prediction (*top*) and the corresponding variance values (*bottom*).

In collaboration with

zhaw School of Engineering

BRUKER

Ensemble Learning Framework

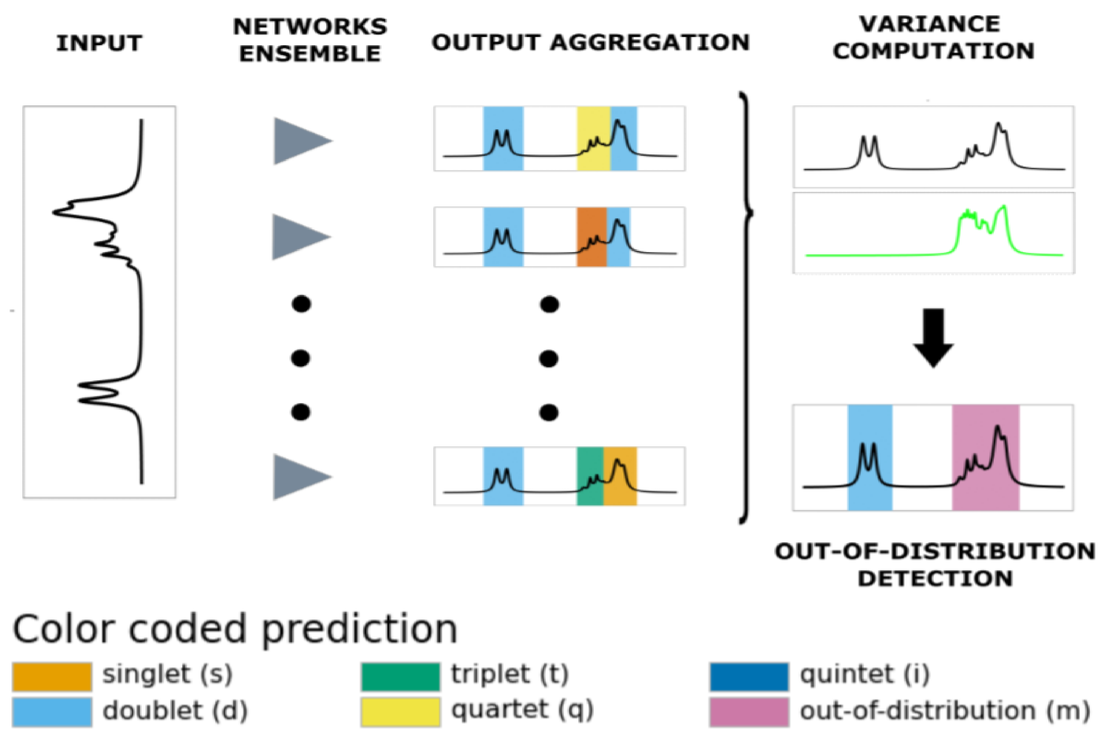

Figure 1: **The framework**: the input spectrum is fed to **10 networks** (*2*); the networks will give the same output on multiplets represented in the training set, while they will produce arbitrary errors on unseen multiplets due to the **epistemic uncertainty** of the model.

Results

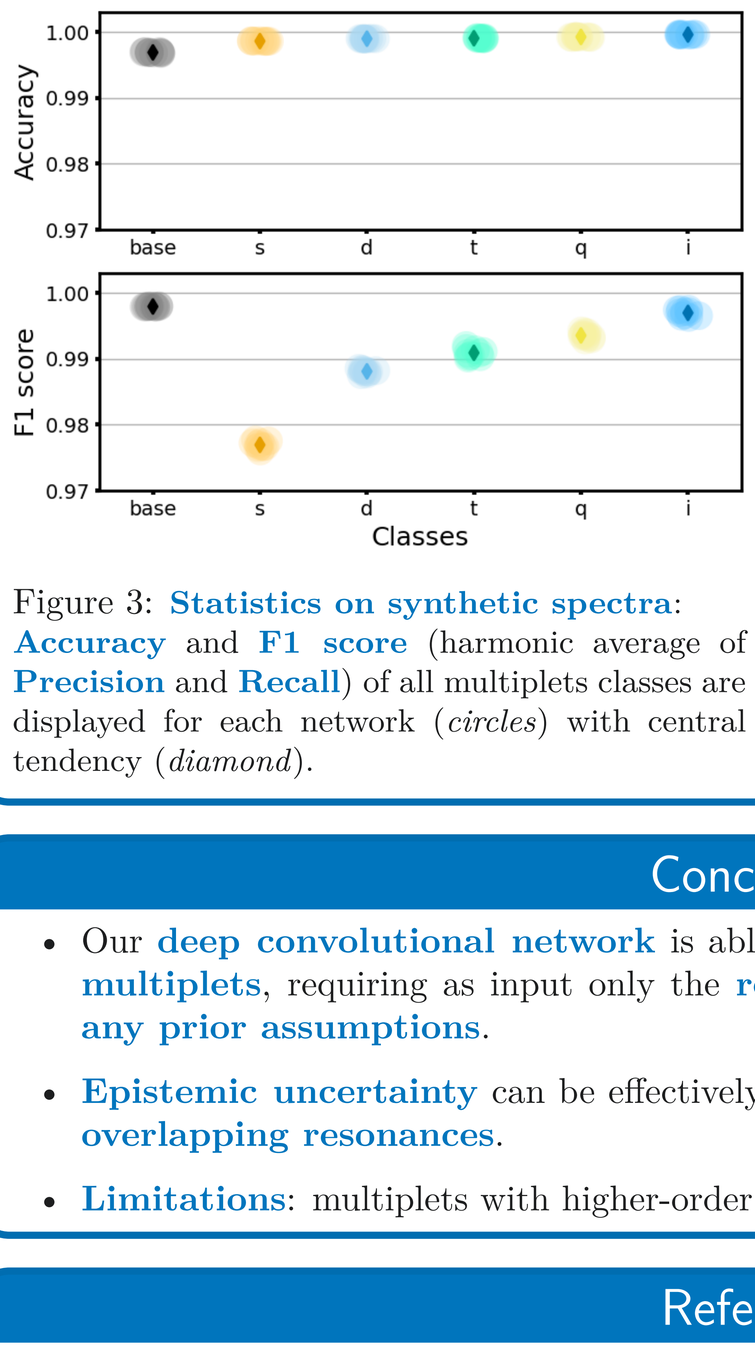

Figure 3: **Statistics on synthetic spectra**: **Accuracy** and **F1 score** (harmonic average of **Precision** and **Recall**) of all multiplets classes are displayed for each network (*circles*) with central tendency (*diamond*).

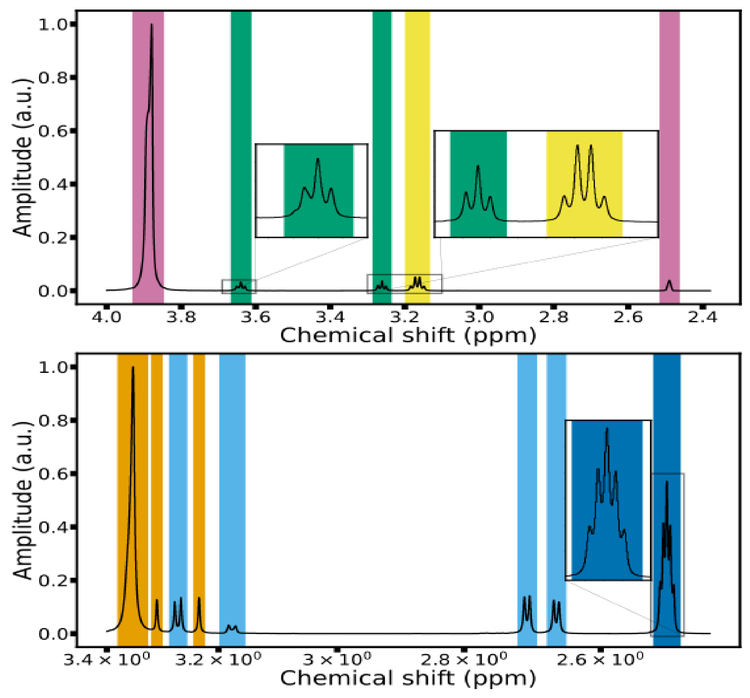

Figure 4: **Prediction** on segments of **experimental spectra**.

Conclusions

- Our **deep convolutional network** is able to classify the **multiplicity of basics isolated multiplets**, requiring as input only the **real amplitudes** of the NMR spectrum, **without any prior assumptions**.
- **Epistemic uncertainty** can be effectively applied to **discriminate non-overlapping and overlapping resonances**.
- **Limitations**: multiplets with higher-order couplings, low SINO.

References

1. D. Chen, Z. Wang, V. Orekhov, X. Qu, *Chem. Eur. J.* **26**, 10391 (2020).
2. O. Sagi, L. Rokach **8**, e1249 (2018).
3. A. Kendall, Y. Gal, *NIPS* (Association for Computing Machinery, 2017).