Multi-label Classification of Ground-based Auroral Emission Spectra Using 1D Vision Transformers

Matthieu Le Lain [1] Gaël Cessateur [2] Sébastien Lefèvre [1,3]

[1]IRISA, UMR CNRS 6074, UBS, Vannes, France [2]Royal Belgian Institute for Space Aeronomy, Brussels, Belgium
[3]UiT – The Arctic University of Norway, Tromsø, Norway

Introduction

- Auroral emissions arise from energetic particles precipitating into the upper atmosphere.
- Classifying spectra characterises energy input into the ionosphere–thermosphere system.
- Needs subjective expert visual inspection, which does not scale.

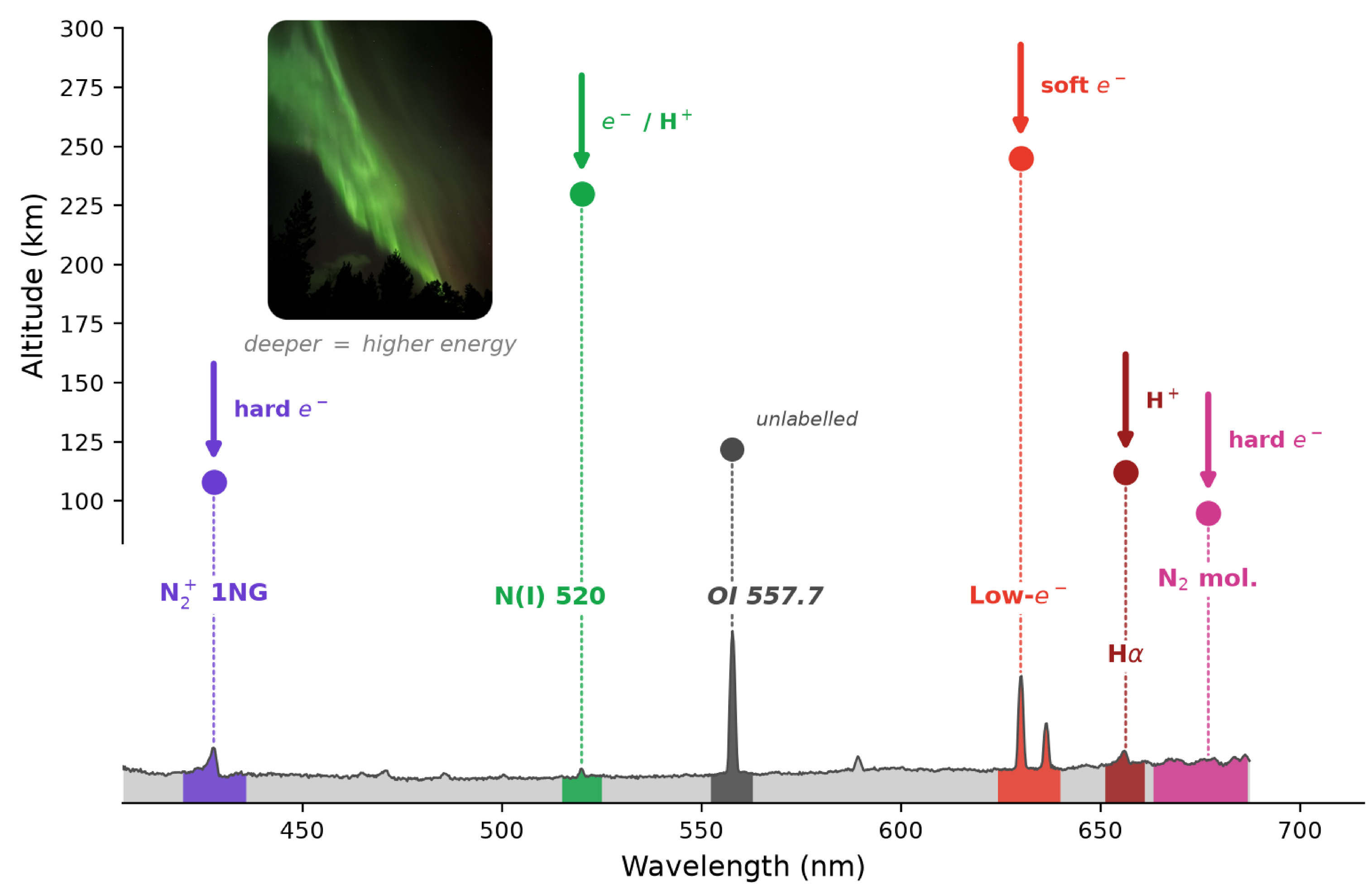

Figure 1. **From precipitation to class label.** Particle energy sets the emission *altitude* and hence the *colour* (top); each line appears in the mean spectrum (bottom) and defines one expert *class*. Soft e^- stay high (red OI 630, Low-e^-); harder e^- go deep (green OI 557.7; violet N_2^+ 427.8); protons give Hα. The strong OI 557.7 nm line (dark grey) is present in all spectra and is *not* a label.

Motivation

*Can Vision Transformer attention act as a **built-in interpretability** mechanism on ordered 1D scientific signals, where physical ground truth enables direct validation?*

Data

- **ASIS** spectrograph, Skibotn (Norway): fibre-fed Czerny–Turner, **405–687 nm**, **1024 bins**.
- **719** min–max-normalised spectra, expert-annotated with **5 binary labels**.

Table 1. Emission classes, physical tracer and sample count ($N = 719$).

Class	Physical tracer	n
Hα	Proton precipitation	346
N_2 mol.	High-energy e^- precipitation	114
Low-e^-	Low-energy e^- precipitation	214
N(I) 520	N(^{2}D $\to$ ^{4}S)	307
N_2^+ 1NG	N_2 ionisation	267

Method

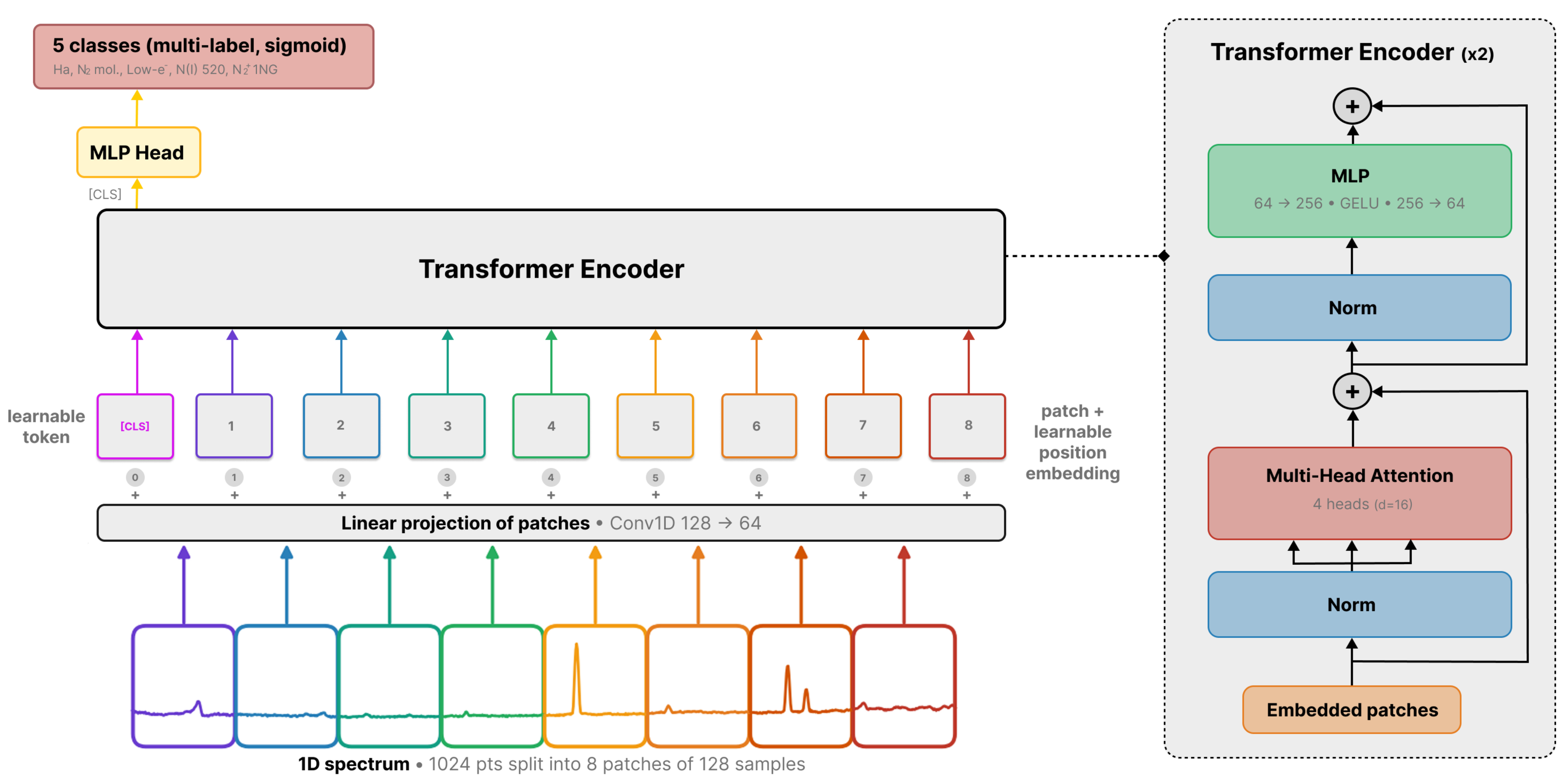

Figure 2. **ViT-1D architecture** - patch tokenisation, learnable `[CLS]`, Transformer encoder blocks, `[CLS]` $\to$ head.

- We train a **1D Vision Transformer (ViT-1D)** on the raw spectra: patch self-attention with a `[CLS]` classifier (architecture above).
- We **compare it to an MLP baseline**, the same backbone with a *single full-spectrum patch* (1 layer, 1 head), so attention collapses to a **learned spectral projection**; this isolates the contribution of attention.
- **Training & evaluation**: augmentation (noise, scale jitter, channel dropout, Mixup); ViT-1D BCE/AdamW (300 ep.), MLP focal loss (200 ep.); **stratified 5-fold CV** $\times 3$ (15 folds), macro F1 & mAP.

Resources & Demo

Scan for the **paper & poster**, supplementary material, and a **web app** to run the trained ViT-1D, and inspect its attention maps, on auroral spectra.

Label Co-occurrence

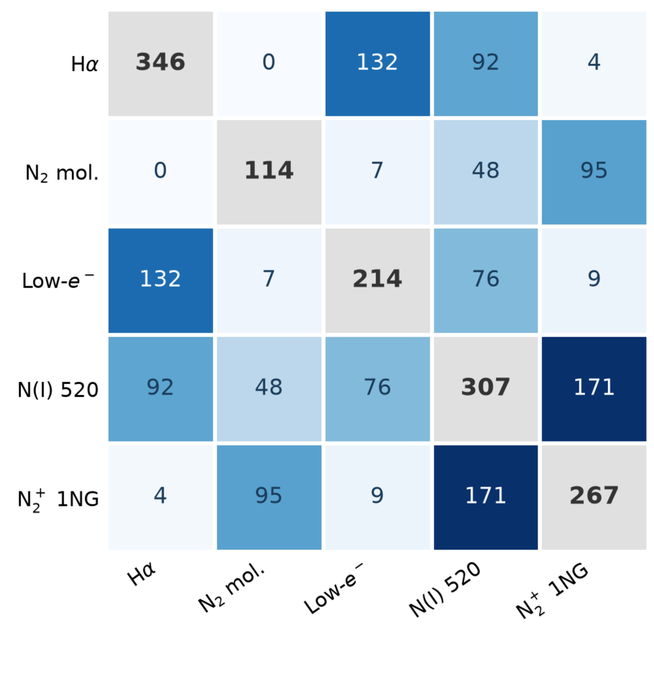

Figure 3. Label Co-occurrence (data).

Co-occurrence is **set by the precipitating particles**:

- **Hα & N_2 mol. : 0%**, proton vs. high-energy-electron events, *disjoint in this dataset*.
- **N(I) 520 & N_2^+ : 64%.** Same particles dissociate *and* ionise N_2 (co-production).
- **N_2 mol. & N_2^+ : 83%.** High-energy electrons make N_2 bands *and* N_2^+.

Ground truth

Does the model's **attention recover this same structure ?**

Results

Both architectures reach **strong performance** (macro AP: MLP 81.1%, ViT-1D 77.8%).

Table 2. Per-class F1 and AP (%), mean over 15 folds. Best in **bold**.

	F1 (%)		AP (%)	
Class	MLP	ViT-1D	MLP	ViT-1D
Hα	**81.8**	80.6	**84.2**	82.7
N_2 mol.	47.8	**52.3**	**57.4**	51.6
Low-e^-	**81.1**	72.5	**86.6**	81.9
N(I) 520	75.2	**75.4**	**84.2**	80.9
N_2^+ 1NG	**86.0**	84.1	**93.3**	91.9
Macro	**74.4**	73.0	**81.1**	77.8

- **N_2^+ 1NG** is easiest for both (F1 $\geq$ 84%), a sharp, isolated spectral band.
- **N_2 mol.** is hardest (F1 $\leq$ 53%): diffuse band structure and only 114 samples. The ViT-1D wins here by **+4.5%** F1, suggesting patch-level attention better captures diffuse band patterns.
- The MLP's largest advantage is on **Low-e^-** (+8.6% F1), whose broad signature benefits from a global receptive field.

Attention as Built-in Interpretability

The `[CLS]` token's attention is a **saliency map over wavelength**, *where the model looks* to decide. Per class we contrast positive vs. negative samples:

$$\Delta A_c(\lambda) = \bar{A}_{c=1}(\lambda) - \bar{A}_{c=0}(\lambda),$$

red ($\Delta A_c > 0$) marks where attention *increases* for the class, blue where it decreases.

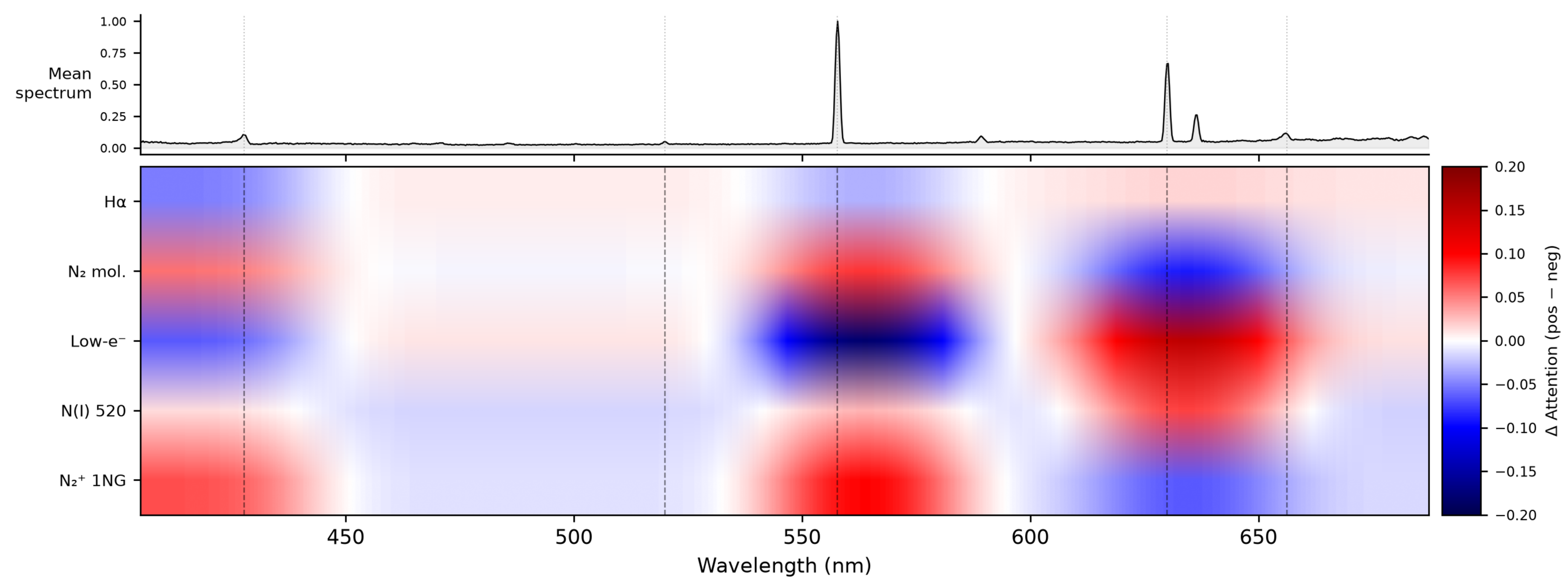

Figure 4. Differential attention $\Delta A_c(\lambda)$ per class, averaged over the 15 cross-validation models. Red: attention *increased* on positive samples; blue: *decreased*. Dashed lines mark known emission wavelengths; top: mean spectrum.

For the trained ViT-1D, the maps recover **known physical patterns** from *classification supervision alone*:

- ✓ N_2^+ 1NG peaks at **427.8 nm**; Hα rises toward **656 nm**; Low-e^- is broad **above 630 nm** (red OI).
- ✓ N_2 mol. is read from **co-occurring** emissions (427.8/557.7 nm), consistent with co-production by **energetic particle impact on N_2**.
- ✓ The *unlabelled* 557.7 nm line is attended by N_2 mol. and N_2^+ 1NG, **co-excitation recovered without supervision**.
- ? N(I) 520: read via co-occurrence (OI 630/636 nm), never its own line, **needs further study**.

Contributions & Conclusion

- **New 1D task**: multi-label classification on 719 expert-annotated auroral spectra (no baseline).
- Patch attention **transfers from images to 1D signals**; ViT-1D matches the MLP.
- Attention = **built-in attribution**: recovers emission lines **without priors**.
- Validated by physical ground truth, a path to **self-supervised** spectral analysis.

References

[1] A. Dosovitskiy *et al.*, "An image is worth 16x16 words," *ICLR*, 2021.
[2] A. Vaswani *et al.*, "Attention is all you need," *NeurIPS*, 2017.
[3] M. H. Rees, "Auroral ionization and excitation by incident energetic electrons," *Planet. Space Sci.*, 1963.
[4] L. B. N. Clausen & H. Nickisch, "Automatic classification of auroral images," *JGR Space Physics*, 2018.
[5] G. Tsoumakas & I. Katakis, "Multi-label classification: an overview," *Int. J. Data Warehousing & Mining*, 2007.
[6] BIRA-IASB, "ASIS: Auroral Spectrograph In Skibotn," 2024.