Improving Ames prediction with graph transformer neural networks

Luke Thompson, Slade Matthews

Sydney Pharmacy School, Faculty of Medicine and Health, The University of Sydney, NSW, Australia

An Introduction to the Ames test and computational modelling

The Ames Test 🧪

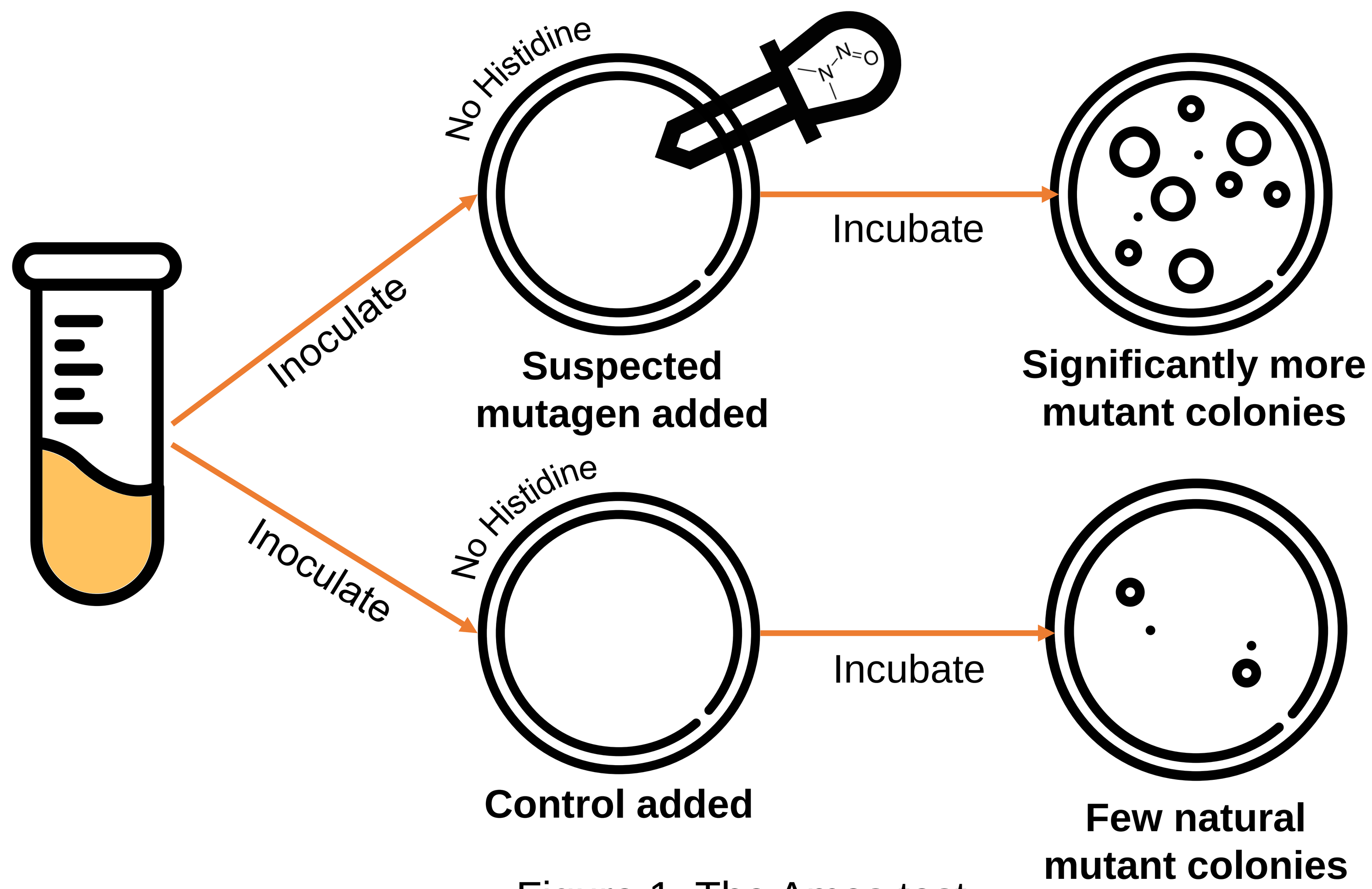


Figure 1. The Ames test

How does the Ames test work? [1]

1. Inoculate an empty plate with histidine-dependent Salmonella
2. Add a control to one plate, and the suspected mutagen to another
3. **If the plate shows more, a mutagen's for sure; if it's scant and bare, no mutagen there.**

Why *In silico*? 🤔

The Ames test is great…
But it's *expensive at scale*
15 000 new molecules daily * $2000 each = **Too expensive to test everything** [3]

However, we can't just take Ames *in silico*. It's too ***unreliable:***

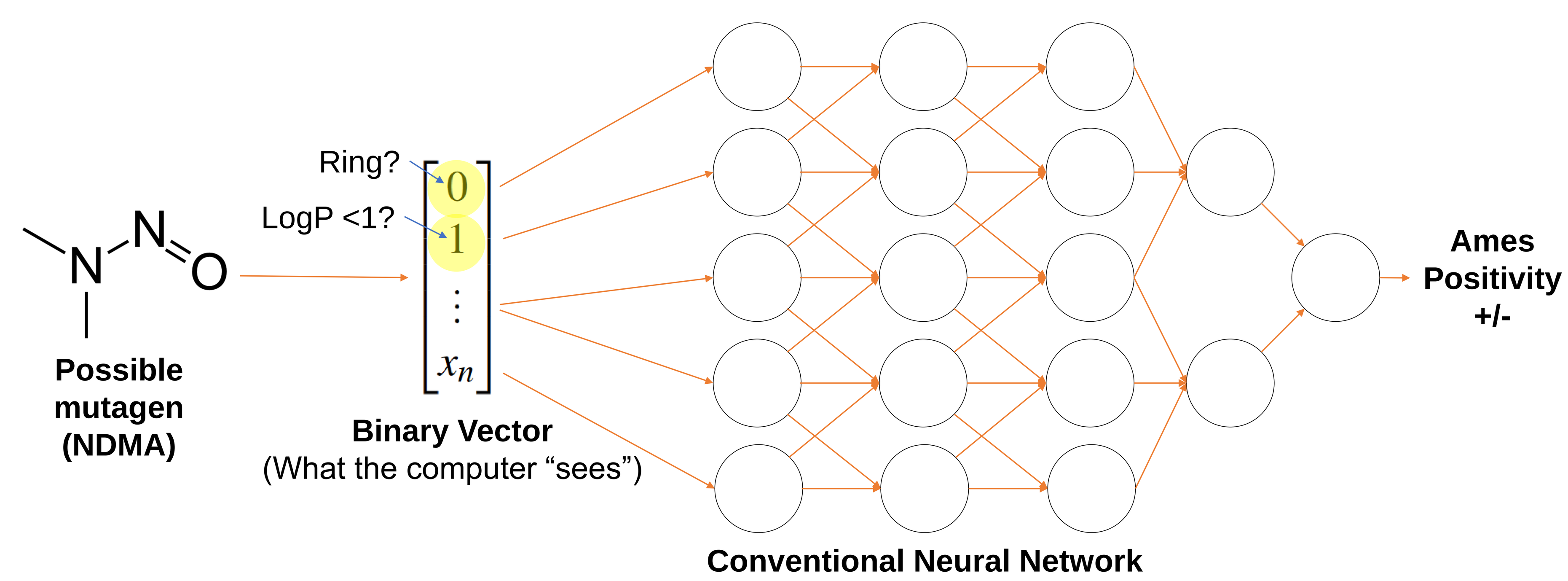


Figure 2. Classical neural network for predicting Ames mutagenicity

Classical methods poorly capture chemical structures [2]. We lose:
Structural info – The molecule is just a vector
Quantity info – The vector is binary, we don't know the quantity of each atom

How can we solve these problems? – *Graph Transformers*

We aim to:

1. **Improve Ames models by incorporating chemical structure into the network**
2. Leverage a cutting-edge graph transformer approach

We hypothesise that:

1. **Our graph transformer will achieve near-state-of-the-art performance**
2. A working model is trainable with current hardware & Ames data availability

So, our approach is:

- Give the network extra positional information via new encodings
- Make the network "pay attention" to each atom's local neighbourhood

All enabled by the transformer!

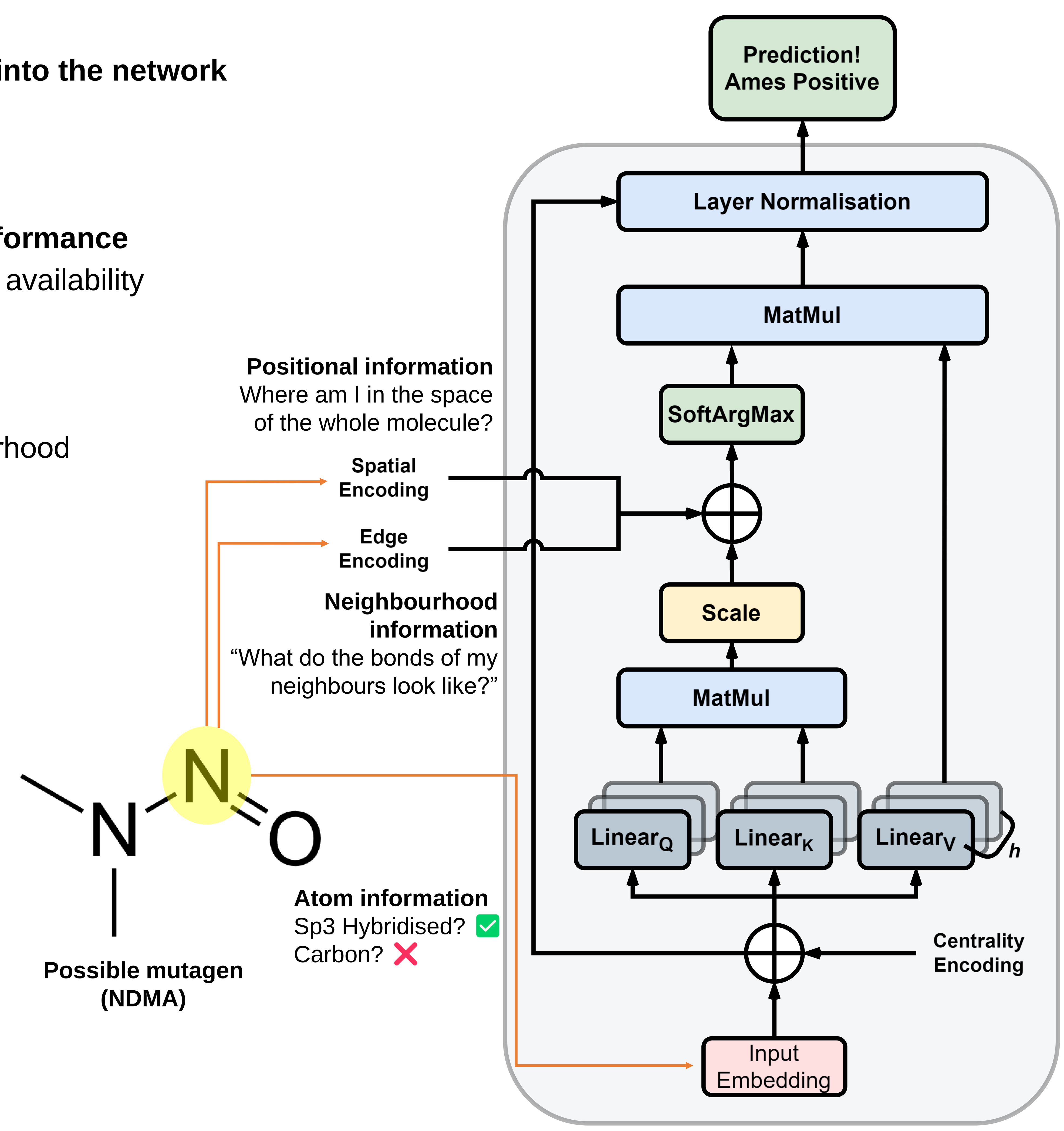
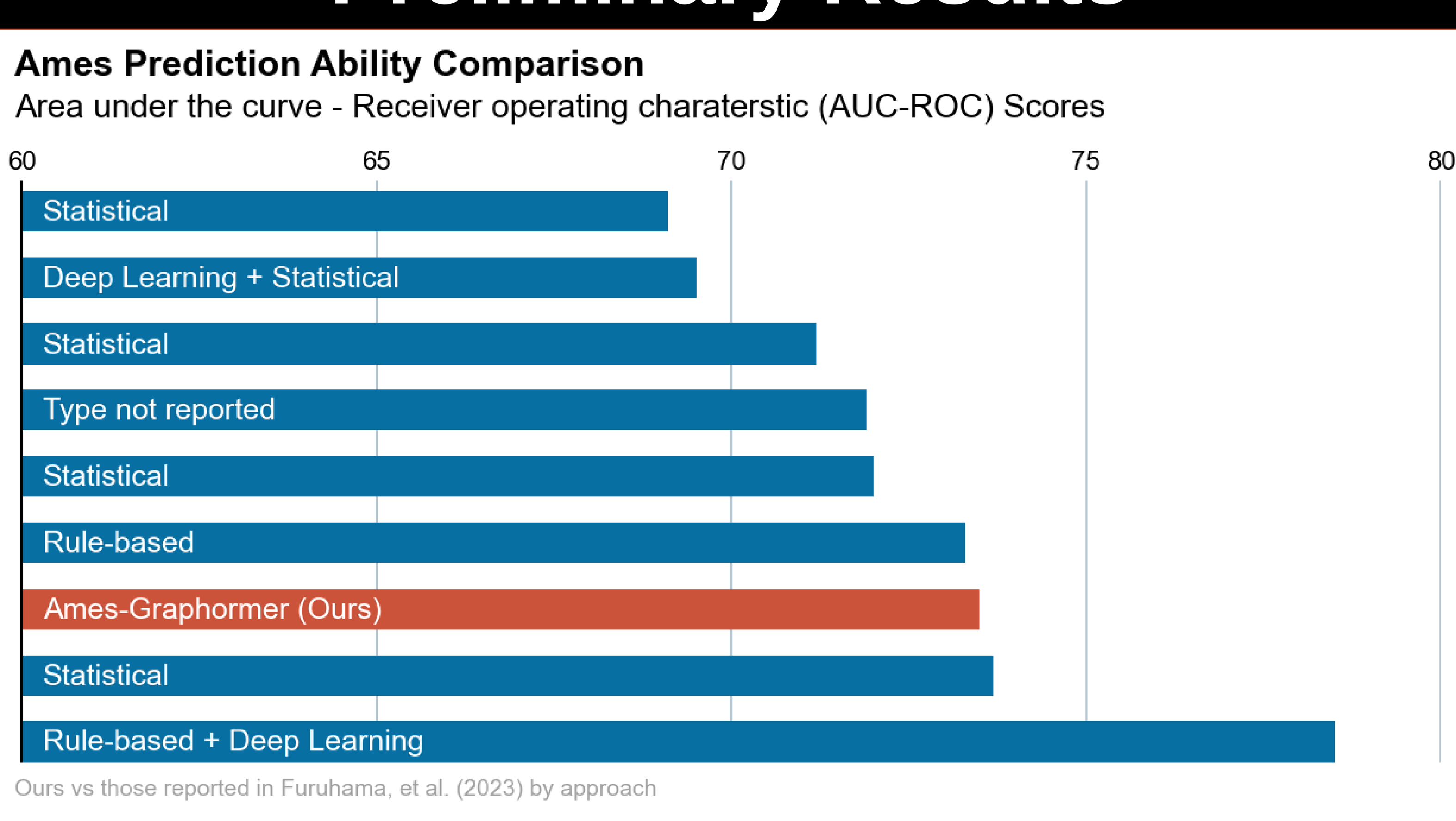


Figure 3. The input and architecture of our model – The Ames graph transformer

Preliminary Results



$$\text{AUC} = \int_0^1 \text{TPR}(x)\,dx$$

Equation 1. The ROC-AUC formula

Conclusions & Future Directions

We explore the cutting edge of neural network architectures for Ames prediction

We show that transformer-based graph neural networks *achieve near-state-of-the-art performance* for Ames mutagenicity prediction

Our method is *extensible* - The Python code is written in a modular form allowing future architectural developments, such as FlashAttention2, to be incorporated and improve performance without the need to re-write the whole code-base.

With the addition of uncertainty estimation, our model has direct regulatory application and fulfills OECD requirements

References: [1] Ames, et al. (1973) Procedures of the National Academy of Sciences. 70, DOI: 10.1073/pnas.70.3.782 [2] Boldini, et al. (2023), Chemrxiv, DOI: 10.26434/chemrxiv-2023-0m355 [3] Honma, et al. (2019), Mutagenesis, 34, DOI: 10.1093/mutage/gey031