



INSTITUTE OF COMPUTER SCIENCE

Academy of Sciences of the Czech Republic

Martin Holeňa

Pod Vodárenskou věží 2, 182 07 Praha 8, phone: +420 266052921, fax: +420 286585789, e-mail: martin@cs.cas.cz
web: www.cs.cas.cz/~martin

Review of the master thesis

**Comparative Markov state analysis of
APOE protein dynamics by neural networks**

by Jakub Kopko

The reviewed master thesis deals with simulation-based analysis of the APOE protein based on advanced artificial neural networks of the kind VAMPnet, in particular on its comparative extension CoVAMPnet. The thesis seems to be the first application of those kinds of neural networks to simulation-based protein analysis. Basically, it validates and extends results obtained using more traditional methods by researchers from the Loschmidt Laboratories at the Masaryk University Brno, which have been recently published in the journal *Molecular Neurodegeneration*.

I especially appreciate the high interdisciplinary competences of Jakub Kopko that his thesis shows. His participation in the reported research would not be possible without a sufficiently deep understanding of the underlying biological and biochemical problems. At the same time, the employed models need above-average mathematical knowledge.

After reading the thesis, I had a number of questions to the methodology of the reported research. I really appreciate that the author was so kind and visited me to answer in detail all of them.

From the formal point of view, the thesis is well written, but the presentation in some parts lacks the information that the reader expects. In particular:

- A chapter called “Related work” is expected to survey, in its individual sections, main works concerning the topic of the respective section, especially recent publications. In the thesis, however, each section rather explains background knowledge concerning its topic and the relation of its topic to the topic of the thesis.
- In Section 2.3, a kind of hierarchical clustering is described without that term being at least once mentioned.
- Theorem 1 refers to the Hilbert-Schmidt norm without recalling what it is.
- The description of the experiment reported in Section 3.3 states that for each random split, three models have been prepared. However, it does not explain how those three models differed from each other.

- In the description of the experiment reported in Section 3.3, no justification of choosing the values 0.001 and 20 for the parameters of early stopping has been provided.
- The caption of Figure 3.1 does not explain what each of the four curves in that figure denotes.
- In Section 3.4, it has not been explained how weighting according to the probability of assignment to a given state is performed to avoid choosing frames from a single trajectory.

Despite the above-mentioned presentation flaws, I like the submitted thesis, it definitely belongs to the most noteworthy master theses at the four faculties where I am supervising master and PhD theses. Therefore, I recommend assigning it the grade 1.

Prague, September 4, 2023.

Martin Holeňa