JOHANNES KEPLER
UNIVERSITY LINZ

Univ.-Prof.
Dr. Sepp Hochreiter
Head of Institute for
Machine Learning

P +43 732 2468 4521
F +43 732 2468 4539
hochreit@ml.jku.at

Office:
Birgit Hauer
Ext.4520
birgit@ml.jku.at

Linz, March 17, 2022

Supervisor Statement about the Bachelor's thesis of Markus Promberger

This work is about designing and implementing a sequence-to-sequence auto-encoder (AE) pipeline on the basis of Long Short-Term Memory (LSTM) in the domain of bioinformatics. As the downstream task, the learned latent representation of the AE is clustered and analyzed. The provided dataset contains hemoglobin sequences from three different kingdoms of animals – mammals, birds and fish. The total amount of datapoints is extraordinarily small for deep learning methods, and the kingdoms themselves are rather unbalanced. These characteristics render the tasks at hand considerably difficult, especially for the downstream analysis (clustering) of the latent representation. Hence, to make the best use of this datasets, it is of utmost importance to structure it in an adequate way. For this very reason, Markus applied his combined knowledge about sequence analysis methods and machine learning evaluation procedures to tackle this challenging task. The design and implementation of the AE architecture as well as the interpretation of the results were carried out thoughtfully and without the need of extensive input from our side. Markus always put a special emphasis on finding biological parallels when interpreting his results, e.g. when analyzing reconstruction errors on the basis of single amino acids.

In the course of his thesis, Markus showed a dedicated and goal-driven motivation to not only work on the tasks assigned to him, but to also delve into the topic itself. If faced with a problem, he autonomously looked for additional sources, analyzed and adapted their respective methods and incorporated the newly gathered knowledge into his own work. This is especially noticeable in the data preparation section of his

JOHANNES KEPLER
UNIVERSITY LINZ
Altenberger Str. 69
4040 Linz, Austria
www.jku.at
DVR 0093696

JOHANNES KEPLER
UNIVERSITY LINZ

thesis, where intrinsic knowledge about methods from the domain of bioinformatics is applied to structure the challenging dataset. Moreover, Markus acted conscientiously on comments from our side and embraced every opportunity to improve his thesis – be it with respect to the methodology, or the writing style and form of the thesis itself.

The scope and quality of this work definitely meets what is commonly expected from a Bachelor's student. Hence, we can confidently state that Markus deserves the degree Bachelor of Science.

Johannes Kepler Universität Linz
Institut für Machine Learning
Altenberger Straße 69
A-4040 Linz, Austria

Sepp Hochreiter

JOHANNES KEPLER
UNIVERSITY LINZ
Altenberger Str. 69
4040 Linz, Austria
www.jku.at
DVR 0093696