UNIVERSITY OF SOUTHAMPTON

<u>ABSTRACT</u>

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
ELECTRONICS AND COMPUTER SCIENCE

<u>MSc in Artificial Intelligence</u>

by Guillermo Romero Moreno

The relatively new field of deep learning is slowly being transferred to the biomedical field and pushing its state-of-the-art achievements. However, improvements in performance with deep learning machines come with the drawback of opaqueness, which is especially detrimental to biomedical research. A few authors have already started applying deep network interpretability techniques on biological problems to overcome this issue, but none of them has approached problems like structural tagging, where every element in the sequence has a classification output.

This work aims to develop interpretability techniques for deep networks that have been trained to solve the secondary-structure prediction problem as one of the best studied structural tagging problems. For doing so, a state-of-the-art-similar convolutional neural network is first trained, and then saliency maps are applied to it. Since a saliency map gets produced per position (instead of by sequence, as in previous studies), new ways of aggregating them for gathering meaningful insights have been construed. These preliminary techniques could be of double value: on one side, they may help biologists to get a better understanding on the underlying protein structural forming process; on the other, machine learning researchers can better understand their machines and spot their previously uncovered flaws.