A DEEP LEARNING PLATFORM FOR DIABETES BIG DATA ANALYSIS

K. Li[1], F. Liu[2], H. Dong[2], P. Herrero[1], Y. Guo[2], P. Georgiou[1]

[1]Centre for Bio-Inspired Technologies, Institute of Biomedical Engineering, Imperial College London, UK
[2]Data Science Institute, Imperial College London, UK

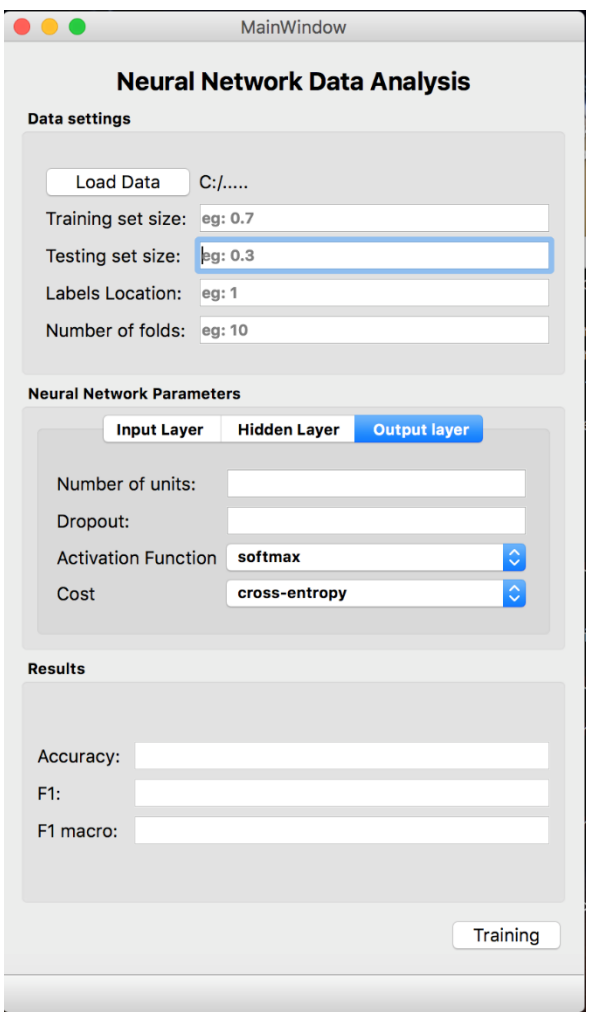
Introduction

- Deep learning has been proven to achieve state-of-art performance in modelling complex behaviour and the prediction accuracy.
- The extensive data collected in hospitals opens the door to such powerful machine learning technology to improve healthcare management.
- So far, few artificial intelligence (AI) tools have enabled non-experts in data science to analyse clinical data intuitively.
- This work aims at building a bridge to link this demand.

Methods

- We develop a deep-learning platform for biomedical data analysis, and demonstrate its applicability using a diabetes dataset.
- All components are built using open-source software libraries, including popular TensorFlow [1] (by *@google*), TensorLayer [2] and TensorDB [3] (by *@Imperial College London*).
- It provides healthcare professionals with an easy-to-use system to manage the heterogeneous clinical data with non-relational databases (e.g. MongoDB).

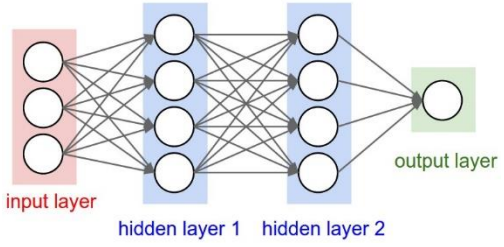
Figure 1: The UI of the learning platform.



Settings

- The user can set various hyper-parameters in the deep neural network on the UI manually, such as *number of layers*, *number of neurons* in each layer, *activation functions*, *dropout*, *cost function*, *training iterations*, etc.
- Alternatively, the user can train the system automatically and let the platform choose the best parameters for the specific task.

- *Figure 2: An example of neural network structure*



Results

- We have tested the platform by processing the *Pima Indians Diabetes Dataset [4]* in a problem of predicting diabetes in a population of 700 subjects. A 10-fold cross-validation model validation technique was used for this purpose.
- On the platform, a plainest 2-layer neuron network was constructed within 10 minutes. It achieved an accuracy of 78.6% in identifying diabetes patients, ranking 3rd amongst the existing 45 published methods.

- *Table 1: Accuracy comparisons to other works.*

	FES[5]	ANN-FNN[6]	proposed	Logdisc	Other 42[6]
Accuracy	85.9	84.2	78.6	77.7	<77.6

Conclusions

- A user-friendly deep learning platform for clinical big data analysis was developed.
- With little specification, medical and clinical scientists with minimal expertise in data science can build their own deep neural network models, and the system can train, update and deploy deep neural network autonomously.
- It was tested in a diabetes dataset and achieved acceptable results when compared with published methods.

Acknowledgement and References

This work is funded by EPSRC under grant agreement EP/P00993X/1.

[1] https://www.tensorflow.org/
[2] https://tensorlayer.readthedocs.io/en/latest/
[3] http://tensorlayer.readthedocs.io/en/latest/ modules/db.html
[4] https://archive.ics.uci.edu/ml/datasets/ima+indians+diabetes
[5] Chang-S. Lee, et. al. IEEE Trans. on Sys. Man. And Cyber., vol. 41, no. 1, pp. 139-153, 2011.
[6] H. Kahramanli, et. al. Expert Systems with Applications, vol. 35, pp. 82–89, 2008.