Electronic Supplementary Material (ESI) for Environmental Science: Advances.

Application of deep learning to support peak picking during non-target high resolution mass spectrometry in environmental research

Kate Mottershead[a] *& Thomas H Miller*[b]*

[a]Department of Analytical, Environmental & Forensic Sciences, School of Population Health & Environmental Sciences, Faculty of Life Sciences and Medicine, King's College London, 150 Stamford Street, London SE1 9NH, UK

[b]Centre for Pollution Research & Policy, Department of Life Sciences, Brunel University London, Kingston Lane, Uxbridge, UB8 3PH, UK

Table of Contents

List of Tables

List of Figures

S1.0 Manual labelling of ROI files

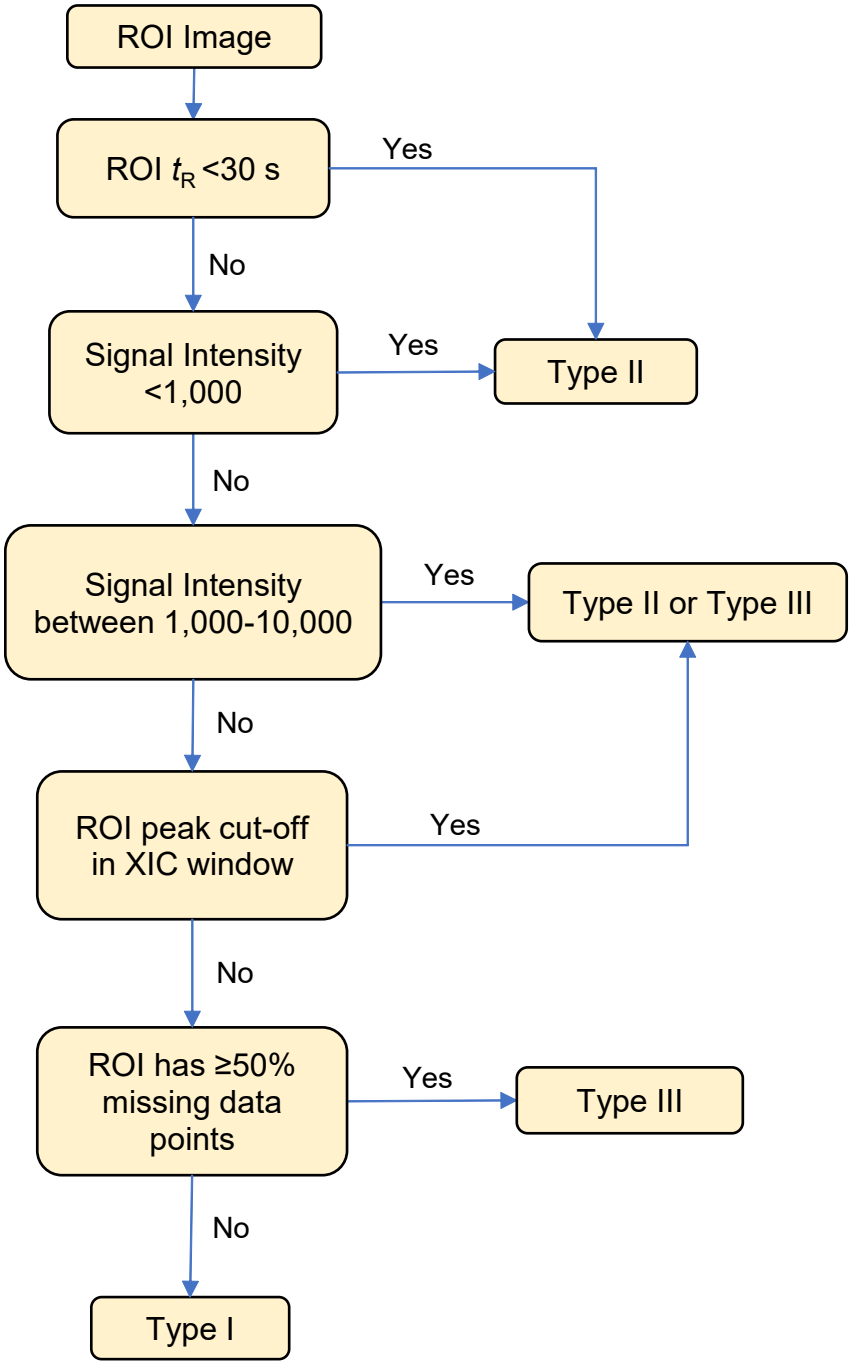

Figure S1. Several criteria followed during the manual labelling of ROIs into respective class labels; Type I, Type II or Type III. Low intensity peaks (1,000 to 10,000) were mainly classified at Type III, unless there were very clear cases that there was no signal (Type II).

S2.0 ROI image feature extraction

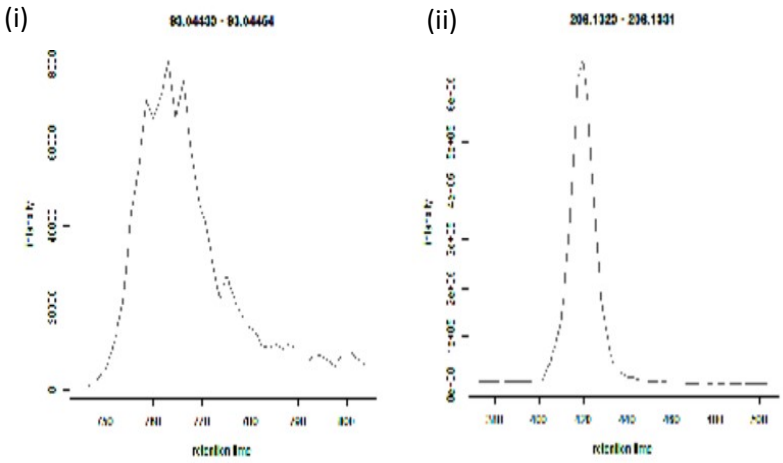

Figure S2. Example ROIs after feature extraction by the Keras image processing applications; (i) Mobile Net (ii) Xception.

S3.0 CNN architecture

Table S1. CNN architecture of the pre-trained models downloaded from ImageNet

Model	Type	No. Layers	No. Parameters (trainable)	Keras Top-5 Accuracy
VGG16	Sequential	23	138,357,544 (138,357,544)	0.901
Xception	Functional	126	22,910,489 (20,813,099)	0.945
MobileNet	Functional	88	4,253,864 (4,234,979)	0.895

S4.0 CNN prediction examples

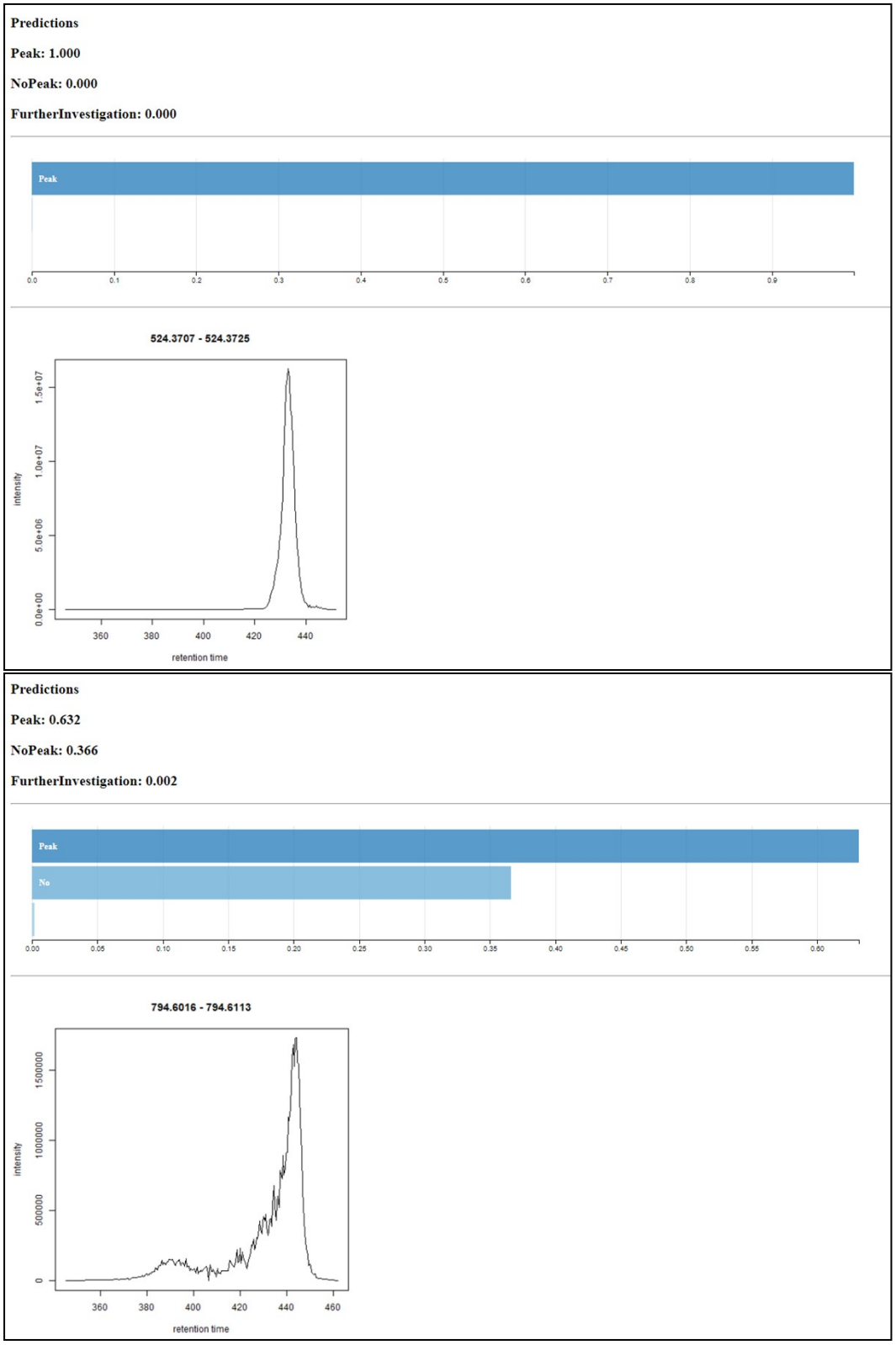

Figure S3. ROI example files classified by the CNN and the respective probability of the prediction.

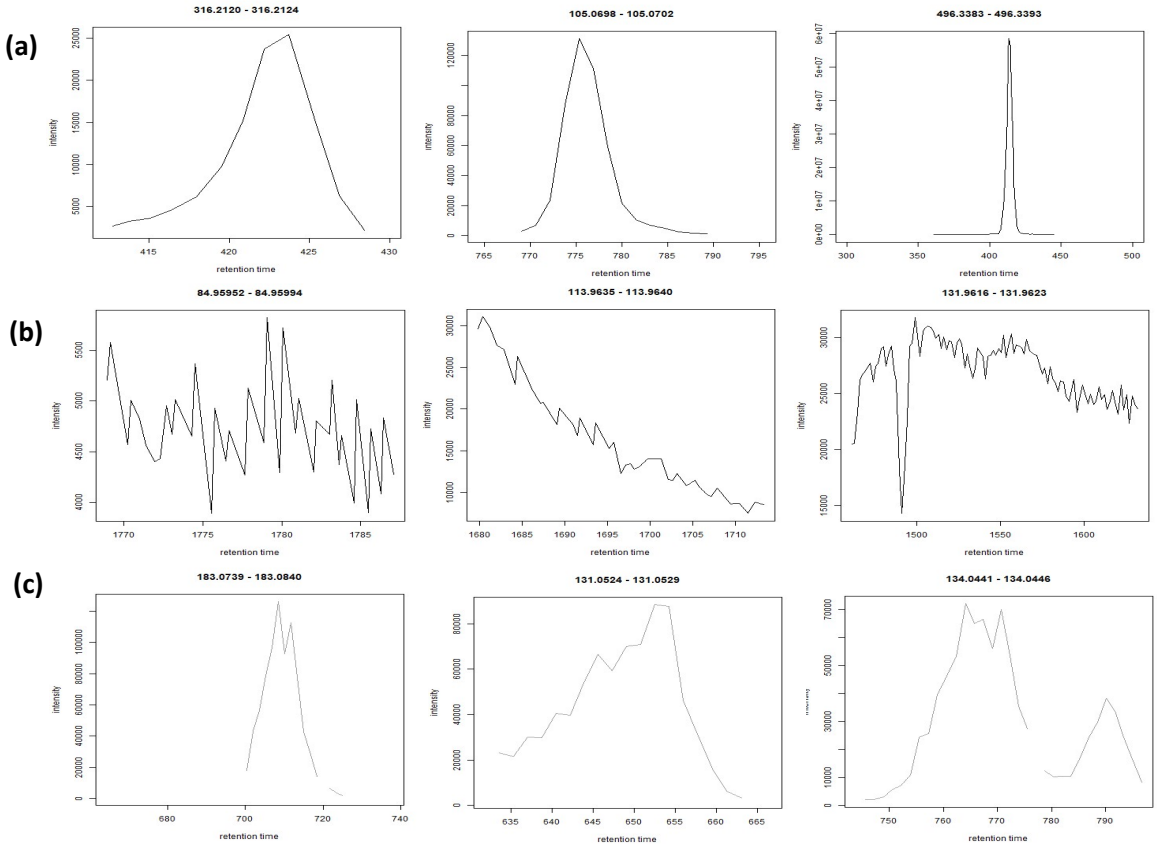

Figure S4: ROI example files for each class (a) type I, (b) type II and (c) type III

S5.0 CNN external test set misclassification

Table S2. Misclassification of ROIs extracted from Fiorini et al., 2020 by the optimised CNN model.

		Probability			
True class	Feature	Type I	Type II	Type III	Predicted Class
Type II	1FT0582	0.562	0.142	0.295	Type I
Type I	1FT0302	0.418	0.579	0.003	Type II
Type I	1FT0427	0.333	0.014	0.653	Type III
Type I	1FT0520	0.336	0.661	0.003	Type II
Type I	1FT0531	0.289	0.711	0	Type II
Type I	1FT0604	0.201	0.029	0.77	Type III

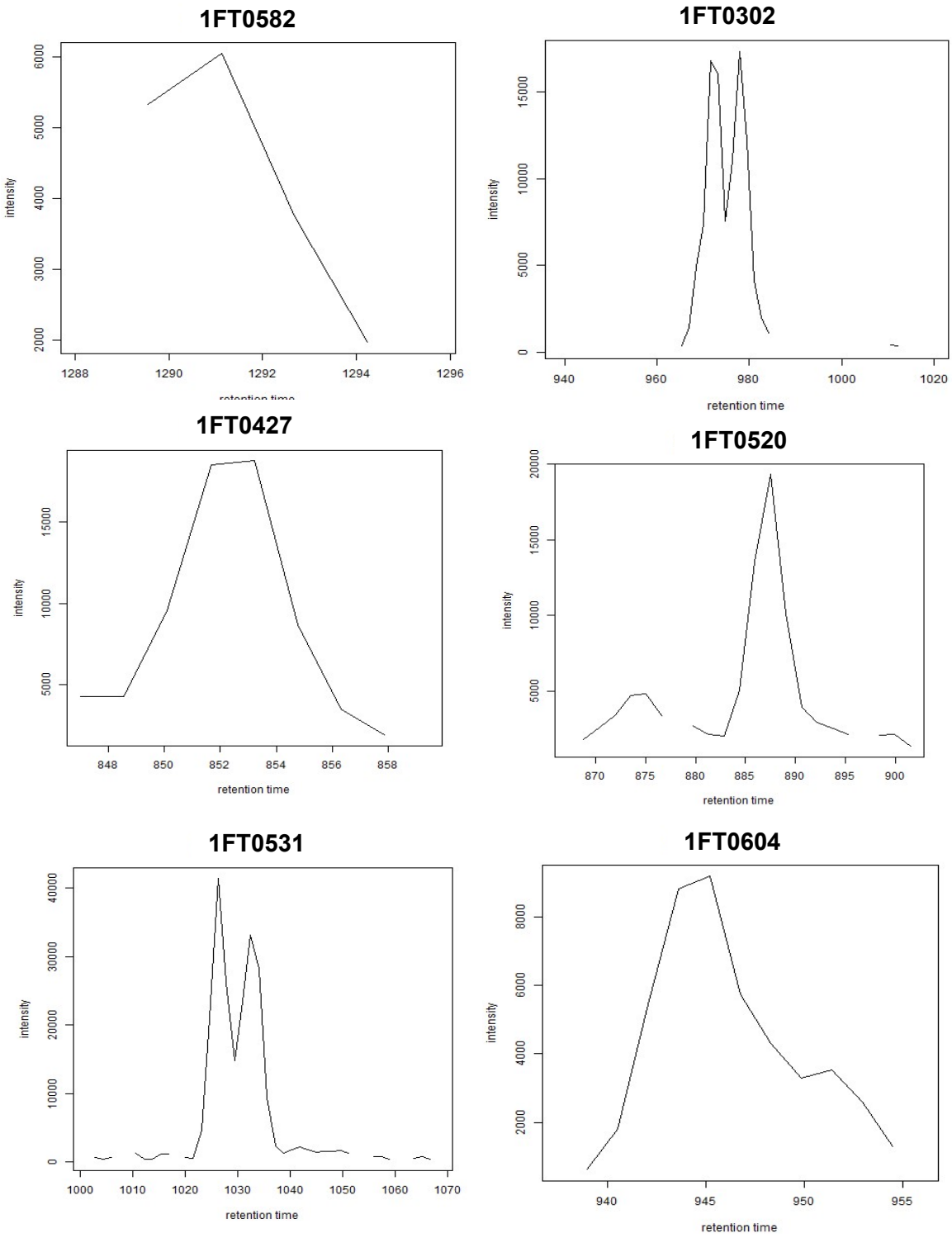

Figure S5. Misclassified ROIs from Fiorini et al., 2020.