

## Recurrent Neural Networks for Time Domain Modelling of Vibrational Spectra: Application to Brain Tumour Detection

**Authors:** Georgios Antoniou, Justin J.A. Conn, Benjamin R. Smith, Paul M. Brennan, Matthew J. Baker, David S. Palmer

### Supplementary Information

**Table S1.** Age (deciles) and sex information about cohort A (size 1438). The corresponding dataset was used as a training set for all models. Where data was unavailable or not recorded, 'NA' (Not available) was noted. Where no samples correspond to a combination of age or sex values, '-' was noted

		Cancer (M/F)	Non-cancer (M/F/NA)
Age	10+	3/2	1/0
	20+	21/18	62/38
	30+	43/47	105/30
	40+	65/56	41/16
	50+	118/149	79/27
	60+	155/114	31/3
	70+	99/79	-
	80+	13/17	-
	90+	1/0	-
	NA	-	1/3/1
Total		1000	438

**Table S2.** Age and sex information about cohort B (385). The corresponding dataset was used as an external test set for the final model. Where no samples correspond to a combination of age or sex values, '-' was noted

		Cancer (M/F)	Non-cancer (M/F)
Age	20+	-	5/14
	30+	2/3	16/20
	40+	4/5	12/18
	50+	2/6	23/28
	60+	12/9	20/44
	70+	9/5	34/36
	80+	4/5	23/21
	90+	0/1	3/1
Total		67	318

**Table S3.** Patient cohort A breakdown with tumour classification details

	Tumour type	Total
<b>Tumour</b>		
	Metastatic	315
	Glioblastoma	265
	Meningioma	137
	Astrocytoma	81
	Pituitary adenoma	78
	Schwannoma	46
	Oligodendroglioma	31
	Ependymoma	8
	Glioma	8
	Haemangioblastoma	6
	Haemangiopericytoma	6
	Lymphoma	4
	Subependymoma	3
	Neurocytoma	2
	Embryonal tumour	1
	Ganglioglioma	1
	Germinoma	1
	Medulloblastoma	1
	Neurofibroma	1
	Orbital tumour	1
	PPTID	1
	Papillary tumour	1
	Pineocytoma	1
	Malignant spindle cell tumour	1
<b>Control</b>		438
	<b>Total</b>	1438

\*PPTID – *Pineal parenchymal tumour of intermediate differentiation*

**Table S4.** Patient cohort B breakdown with tumour classification details

	Tumour type	Total
<b>Tumour</b>		
	Glioblastoma	35
	Metastatic	16
	Meningioma	5
	Astrocytoma	3
	Oligodendroglioma	3
	Ependymoma	1
	Pituitary adenoma	1
	Schwannoma	1
	Subependymoma	1
	Medulloblastoma	1
<b>Control</b>		318
	<b>Total</b>	385

**Table S5.** Distributions of patients against number of spectra measured in each dataset

Dataset	Number of spectra	Percentage of patients (%)
<b>A</b>	2	0.07
	3	1.88
	4	0.07
	5	0.42
	6	5.35
	7	0.83
	8	2.85
	9	88.53
<b>B</b>	9	100

**Table S6.** Additional patient metadata for Cohort B including symptomatology. Where data was unavailable or not recorded, 'NA' (Not available) was noted. BMI, body mass index; NSAID, non-steroidal anti-inflammatory drugs

<b>Metadata</b>		<b>Cancer</b>	<b>Non-cancer</b>
<b>Symptomatology</b>			
Headache	Yes	34	102
	No	32	209
	NA	3	5
Headache duration	< 1month	10	49
	<6 months	14	31
	> 6 months	8	115
	NA	37	121
Headache worse in morning	Yes	9	61
	No	18	146
	NA	42	109
Headache with nausea	Yes	9	76
	No	17	133
	NA	43	107
Memory change	Yes	13	124
	No	47	185
	NA	9	7
Personality change	Yes	11	80
	No	50	230
	NA	8	6
Verbal fluency	15+	26	47
	<15	16	91
	NA	27	178
<b>Drugs</b>			
Steroids	Yes	35	50
	No	4	105
	NA	30	161
Anti-epileptic drugs	Yes	20	17
	No	49	299
Statins	Yes	17	60
	No	52	256
Anti-hypertensives	Yes	18	93
	No	51	223
Anti-thrombotics	Yes	8	45
	No	61	271
NSAIDs	Yes	6	53
	No	63	263
<b>BMI</b>			
	Underweight	5	37
	Normal	37	145
	Overweight	18	91
	Obese	6	21
	NA	3	22

**Fig S1.** Diagrammatic representation of data representations and corresponding transformations. *IDFT* denotes Inverse Discrete Fourier Transform,  $P_{EMSC}$  denotes two sequential pre-processing steps applied to the raw frequency domain spectra, namely cut to the wavenumber region  $3500 - 1000\text{cm}^{-1}$  and EMSC transformation (up to degree 2 corrections).

