Automated Labelling using an Attention model for Radiology reports of MRI scans (ALARM)

David A. Wood[1], Jeremy Lynch[2], Sina Kafiabadi[2], Emily Guilhem[2], Aisha Al busaidi[2], Antanas Montvila[2], Thomas Varsavsky[1], Juveria Siddiqui[2], Naveen Gadapa[2], Matthew Townend[3], Martin Kiik[1], Keena Patel[1], Gareth Barker[4], Sebastian Ourselin[1], James H. Cole[4,5], Thomas C. Booth[1,2]

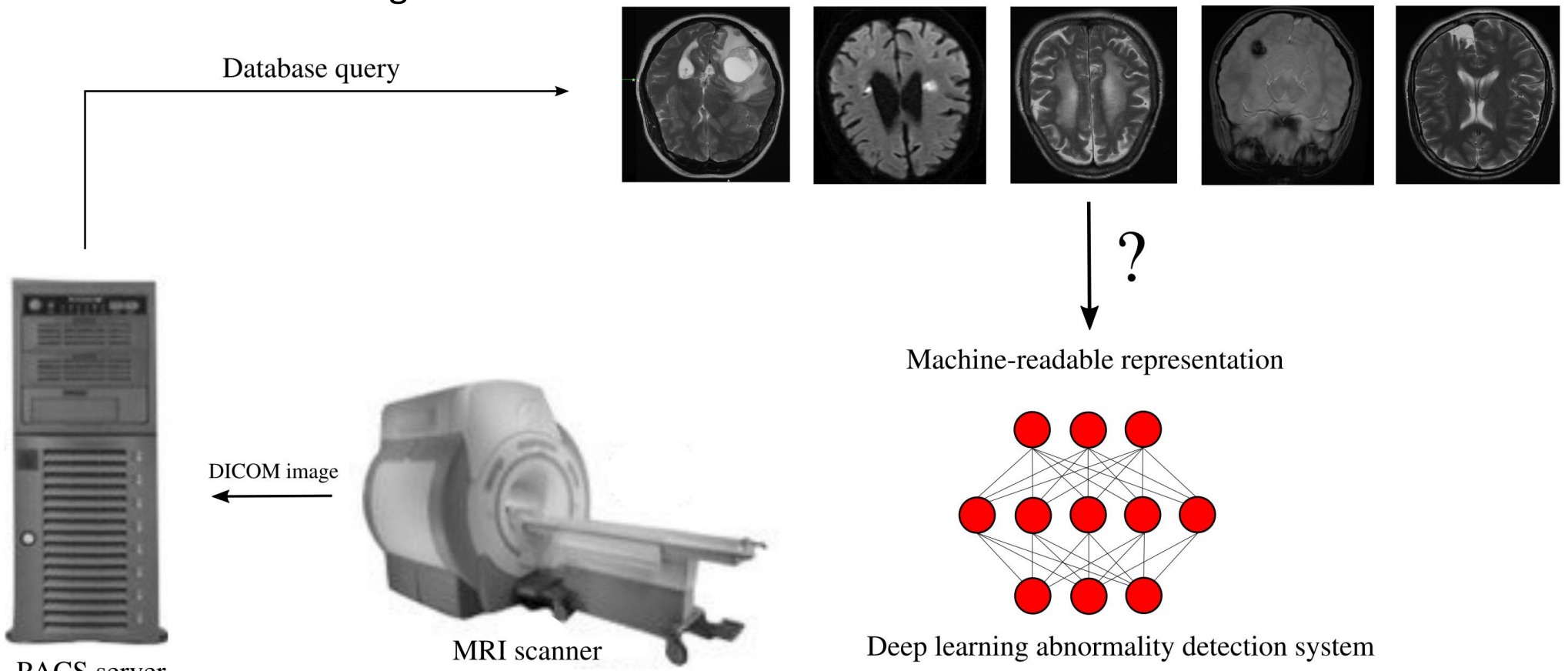
MIDI consortium

[1] School of Biomedical Engineering, King's College London
[2] King's College Hospital
[3] Wrightington, Wigan & Leigh NHSFT
[3] Institute of Psychiatry, Psychology & Neuroscience, King's College London
[4]Centre for Medical Image Computing, Dementia Research, University College London

Background

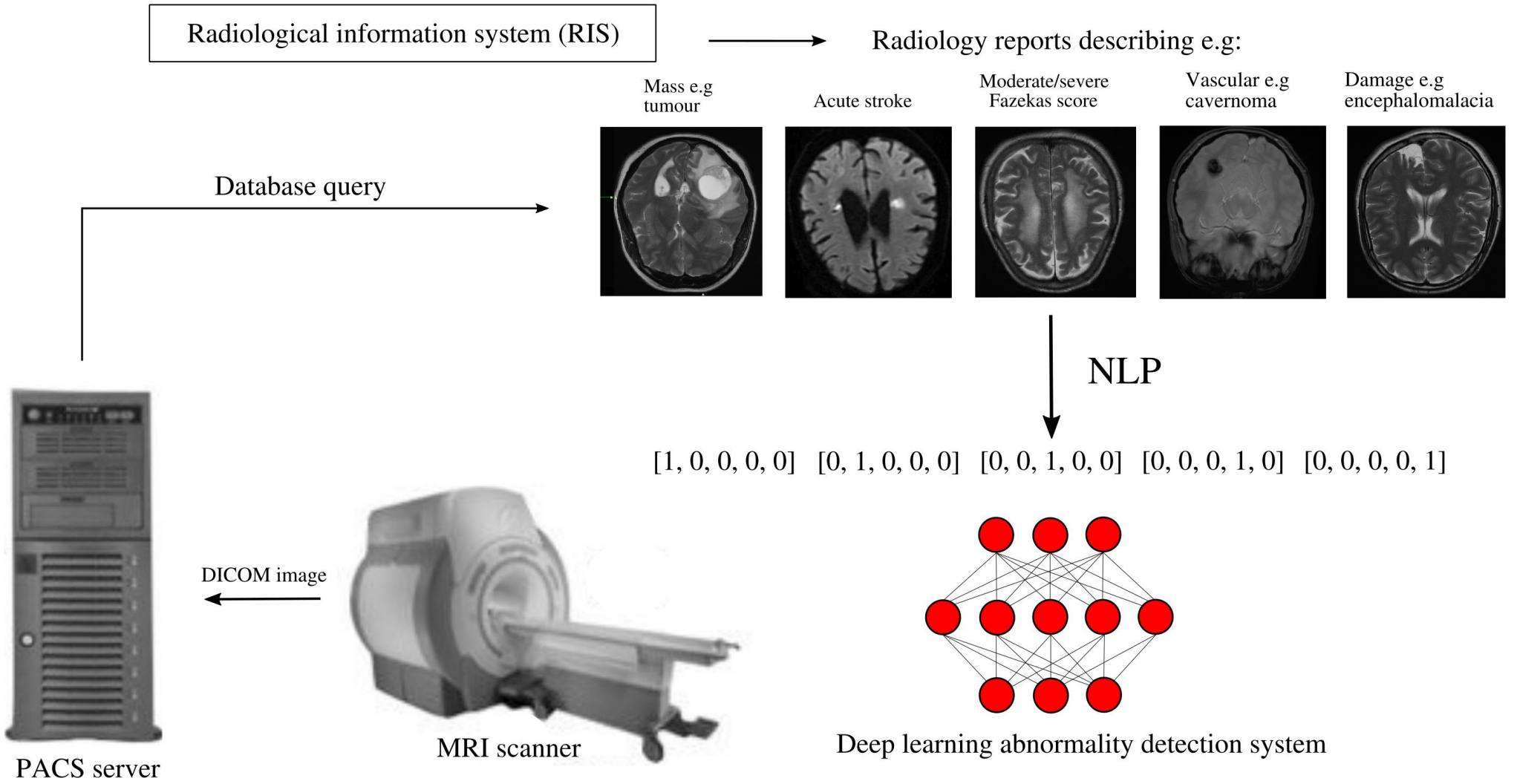
- Labelling training datasets is a rate-limiting step for clinical deep learning applications

- Laborious task requiring considerable domain knowledge and experience
  i.e. neuroradiologist



Automatic labelling with NLP

- Promising alternative – derive labels from radiology reports using natural language processing



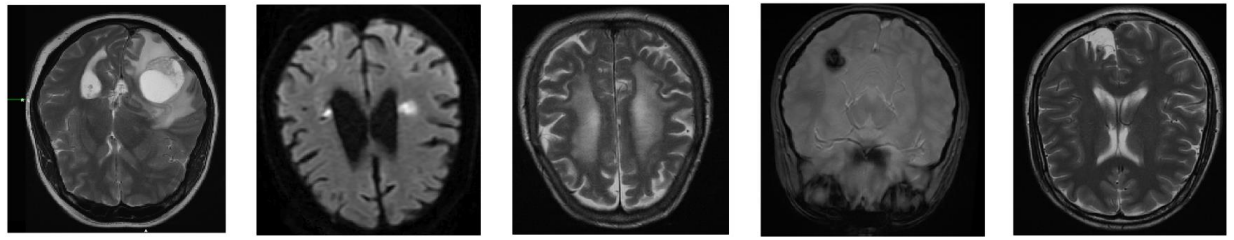
Automatic labelling with NLP

- Promising alternative – derive labels from radiology reports using natural language processing

Radiological information system (RIS)

Radiology reports describing e.g:

Mass e.g tumour | Acute stroke | Moderate/severe Fazekas score | Vascular e.g cavernoma | Damage e.g encephalomalacia



Database query

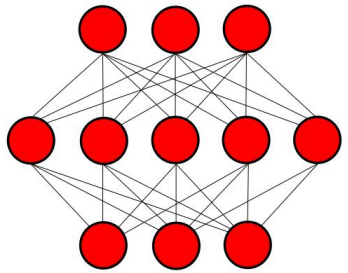
NLP

[1, 0, 0, 0, 0] [0, 1, 0, 0, 0] [0, 0, 1, 0, 0] [0, 0, 0, 1, 0] [0, 0, 0, 0, 1]

PACS server

DICOM image

MRI scanner



Deep learning abnormality detection system

Labelling imaging datasets on the basis of neuroradiology reports: a clinical perspective

David A. Wood[1], Sina Kafiabadi[2], Aisha Al Busaidi[2], Emily Guilhem[2], Jeremy Lynch[2], Matthew Townend[3], Antanas Montvila[2], Juveria Siddiqui[2], Naveen Gadapa[2], Matthew Benger[2], Gareth Barker[4], Sebastian Ourselin[1], James H. Cole[4,5], and Thomas C. Booth[1,2]

MIDI consortium

Automatic labelling with NLP

- Previously demonstrated for head computed tomography reports (Zech et al. 2018)
- No dedicated MRI neuroradiology report classifier
- MRI higher soft tissue contrast, so more detailed descriptions – difficult NLP task
- Reports contain abbreviations, list of absent abnormalities, abnormalities considered insignificant

Radiology

Natural Language–based Machine Learning Models for the Annotation of Clinical Radiology Reports

John Zech, Margaret Pain, Joseph Titano, Marcus Badgeley, Javin Schefflein, Andres Su, Anthony Costa, Joshua Bederson, Joseph Lehar, Eric Karl Oermann ✉

Example reports

Normal study

Clinical Details: Left sided headache, weakness, numbness. initially aphasic but no receptive deficit, now verbalising as normal. OE subtle drift left side and reduced sensation, no visual deficit, no speech deficit now Specific question to be answered: acute ischaemia
MRI Head: Correlation is made to the CT of the same date. No focus of restricted diffusion is demonstrated to indicate an acute infarct. No mass or other focal parenchymal abnormality is identified. No evidence of vascular malformations or areas of signal abnormality. The major intracranial vessels demonstratenormal flow related voids.

Abnormal study

Clinical Details: balance problem, slurring of words, atactic lacunar syndrome vs cerebellar stroke Specific question to be answered: acute ischaemia/posterior circulation infarct.
MRI Head: There is focus of restricted diffusion within the right thalamus indicating an acute infarct. There is no micro or macro-haemorrhages identified. There are two foci of high T2 signal without restricted diffusion or SWI signal change within the right cingulate gyrus which appears longstanding and although in an unusual location is most likely ischaemic. The remaining intracranial appearances are normal. Normal flow voids are noted within the intracranial vessels.

Example reports

Normal study

MRI Head:There is minor generalised prominence of sulci and ventricles. Thereis no lobar predominance to this minor generalised volume loss. There are scattered punctate patches of high signal on T2/FLAIR within the white matter of both cerebral hemispheres. These number approximately 12. Conclusion: There is minor generalised volume loss without lobar predominance. The extent of this volume loss is probably within normal limits for a patient of this age. Scattered punctate foci of high signal within the white matter of both cerebral hemispheres have a non specific appearance and are of doubtful clinical significance

Abnormal study

Clinical Details: RRMS, on Natalizumab therapy, JCV low positive. Interin DMT monitoring
Specific question to be answered: PML, New lesions
MRI Head: Comparison is made with the study of 04/02/2018. No contrast has been administered on this occasion. There has been no increase in the number of demyelinating lesions. There are no lesions demonstrating restriction of diffusion. There are stable appearances, with no imaging features of PML.

Example reports

Normal study

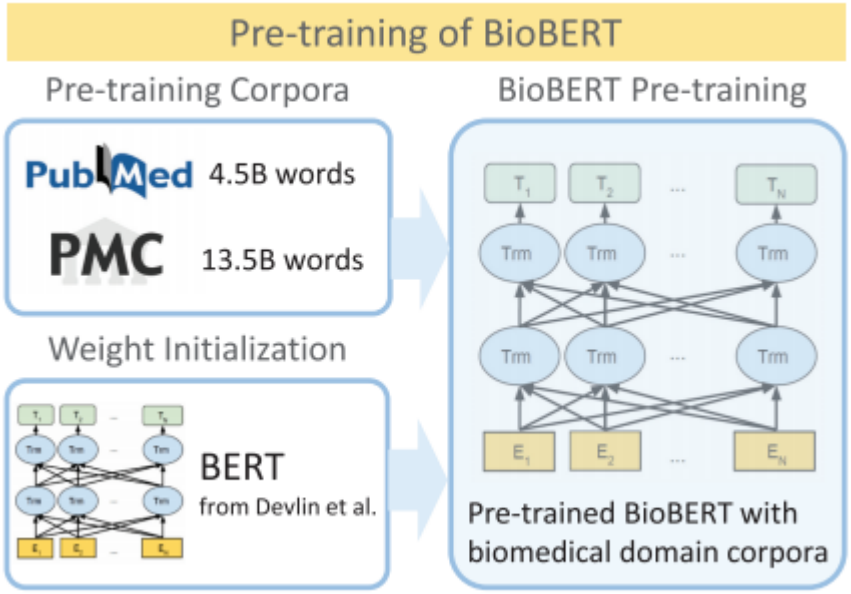
MRI Head:There is minor generalised prominence of sulci and ventricles. Thereis no lobar predominance to this minor generalised volume loss. There are scattered punctate patches of high signal on T2/FLAIR within the white matter of both cerebral hemispheres. These number approximately 12. Conclusion: There is minor generalised volume loss without lobar predominance. The extent of this volume loss is probably within normal limits for a patient of this age. Scattered punctate foci of high signal within the white matter of both cerebral hemispheres have a non specific appearance and are of doubtful clinical significance

Abnormal study

Clinical Details: RRMS, on Natalizumab therapy, JCV low positive. Interin DMT monitoring
Specific question to be answered: PML, New lesions
MRI Head: Comparison is made with the study of 04/02/2018. No contrast has been administered on this occasion. There has been no increase in the number of demyelinating lesions. There are no lesions demonstrating restriction of diffusion. There are stable appearances, with no imaging features of PML.

BioBERT

- Need sophisticated language model trained on relatively few labelled reports
- Fine-tune BioBERT, transformer-based biomedical language model
- Inherit low level language comprehension i.e. transfer learning
- See “The illustrated Transformer” by Jay Alammar for introduction to transformers

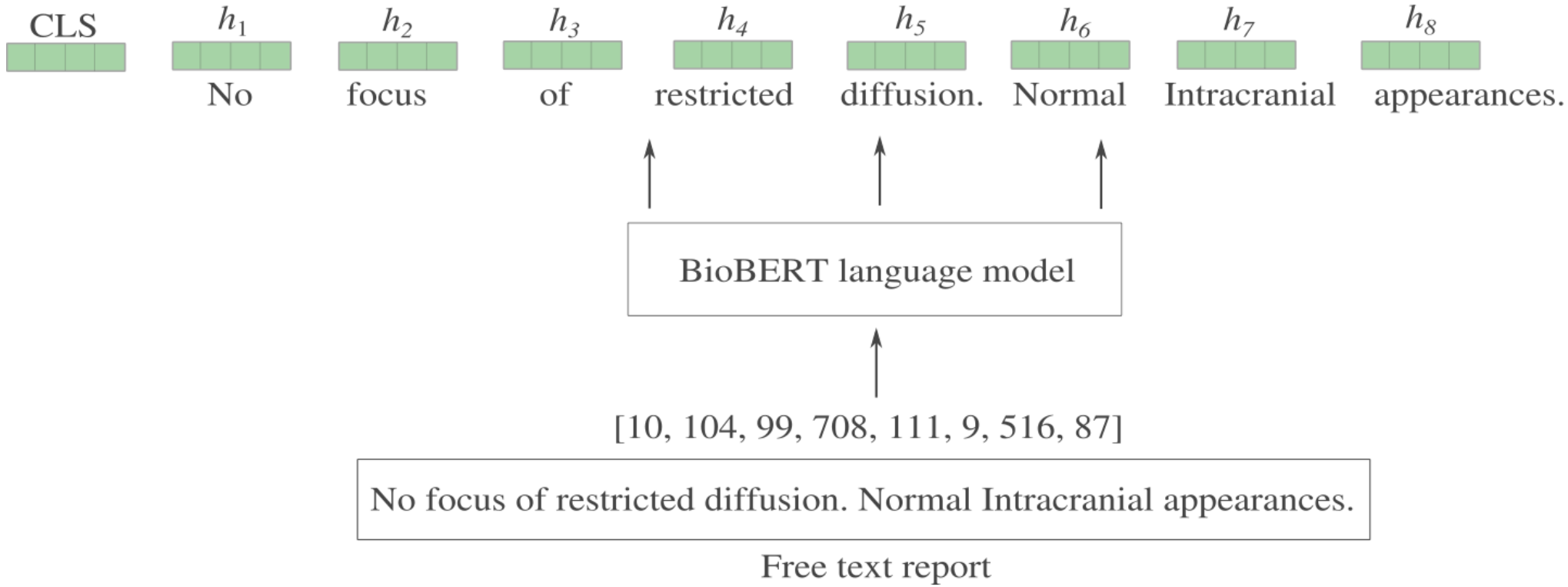


Corpus	Number of words	Domain
English Wikipedia	2.5B	General
BooksCorpus	0.8B	General
PubMed Abstracts	4.5B	Biomedical
PMC Full-text articles	13.5B	Biomedical

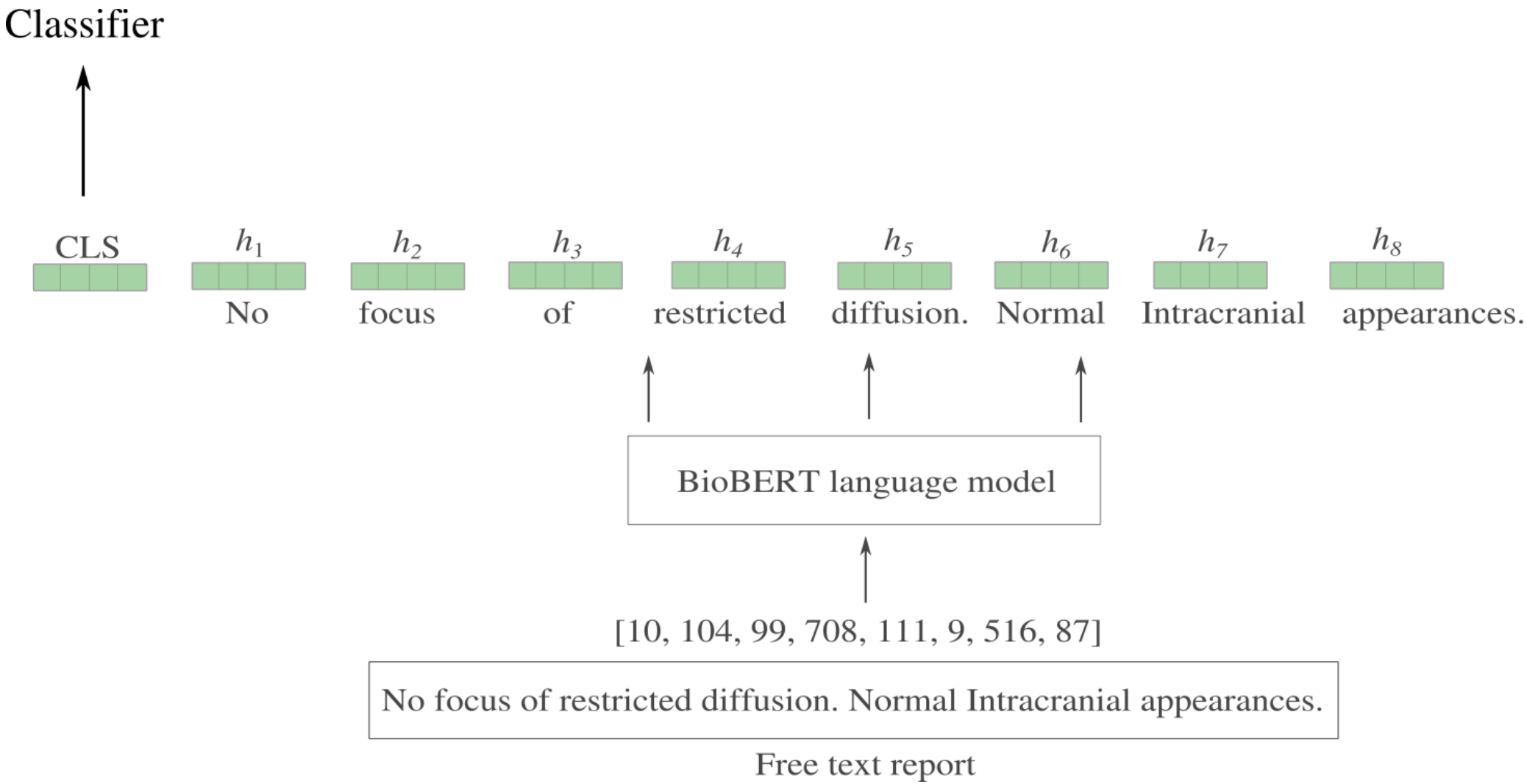
From Lee et al., 2019

Model

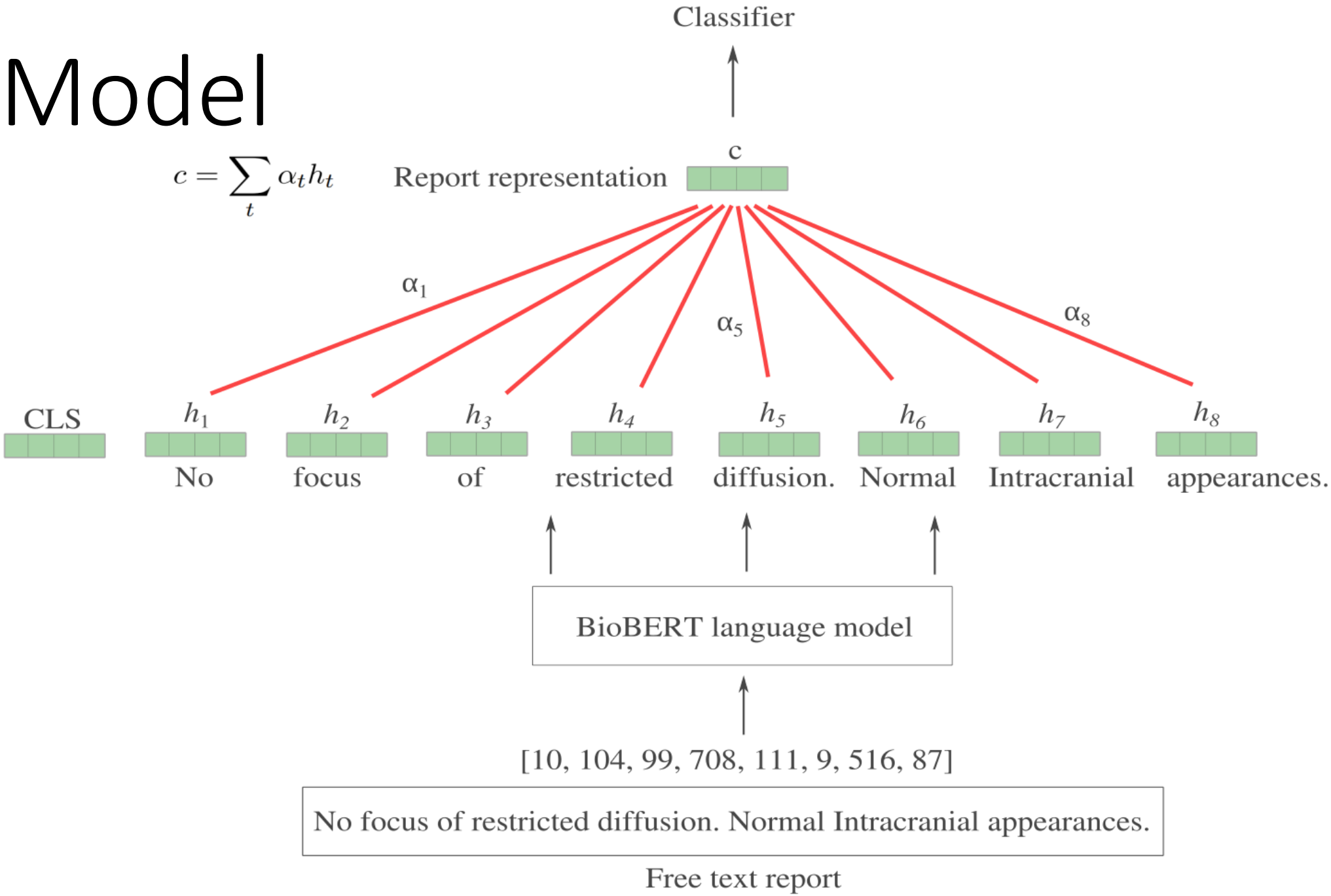
- BioBERT converts text to contextualised word embeddings
- Downstream classification can be performed by aggregation of embeddings
- CLS, max, average, attention weighted



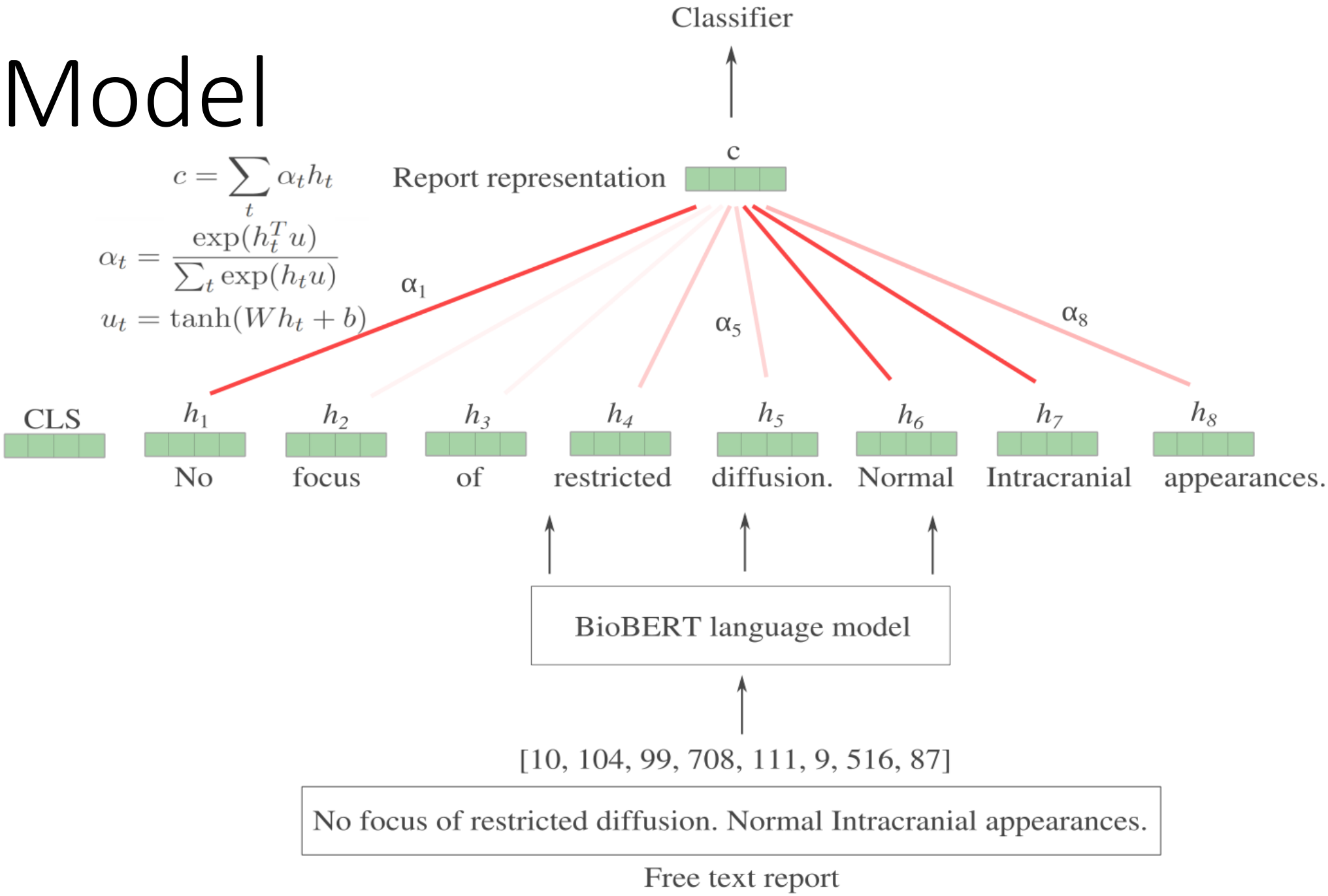
Model


Classifier
CLS
h_1
h_2
h_3
h_4
h_5
h_6
h_7
h_8
No
focus
of
restricted
diffusion.
Normal
Intracranial
appearances.
BioBERT language model
[10, 104, 99, 708, 111, 9, 516, 87]
No focus of restricted diffusion. Normal Intracranial appearances.
Free text report

Model


Classifier
$c = \sum_t \alpha_t h_t$
Report representation
c
α_1
α_5
α_8
CLS
h_1
h_2
h_3
h_4
h_5
h_6
h_7
h_8
No
focus
of
restricted
diffusion.
Normal
Intracranial
appearances.
BioBERT language model
[10, 104, 99, 708, 111, 9, 516, 87]
No focus of restricted diffusion. Normal Intracranial appearances.
Free text report

Model


Classifier
c
Report representation
$c = \sum_t \alpha_t h_t$
$\alpha_t = \frac{\exp(h_t^T u)}{\sum_t \exp(h_t u)}$
$u_t = \tanh(W h_t + b)$
α1
α5
α8
CLS
h1
h2
h3
h4
h5
h6
h7
h8
No
focus
of
restricted
diffusion.
Normal
Intracranial
appearances.
BioBERT language model
[10, 104, 99, 708, 111, 9, 516, 87]
No focus of restricted diffusion. Normal Intracranial appearances.
Free text report

Data and report labelling

- > 120, 000 radiology reports and corresponding MRI scans obtained

- 3000 randomly selected for labelling by team of neuroradiologists for model training and validation

- 1000 reports labelled into 5 clinically relevant granular categories:
    - Mass e.g. tumour
    - Vascular abnormality e.g. aneurysm
    - Damage e.g. previous brain injury
    - Acute stroke
    - Fazekas small vessel disease score

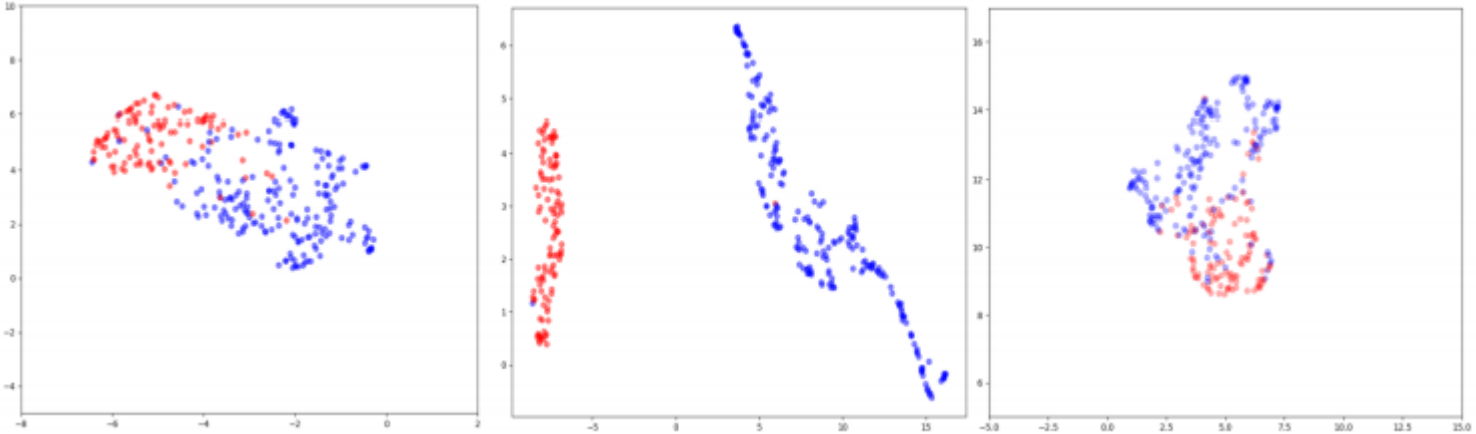
- 2000 reports labelled for presence or absence of any abnormality (on the basis of criteria defined by team over the course of 6 months of practice experiments)

Results

- Binary classification i.e. normal/abnormal

Model	accuracy (%)	sensitivity(%)	specificity(%)
Our model	**99.4**	**99.1**	**99.6**
base-BioBERT	97.7	97.3	97.9
frozen-BioBERT	96.5	96.4	96.6
word2vec (Zech et al.)	91.5	79.2	97.1
Expert physician	92.7	77.2	98.9

t-SNE visualisation of test set report embeddings



Frozen BioBERT Our model word2vec

Results

- Granular classification

Model	damage			vascular			mass			acute stroke			Fazekas		
	acc	sens.	spec.	acc	sens.	spec.	acc	sens.	spec.	acc	sens.	spec.	acc	sens.	spec.
Our model	93.8	92.6	94.3	95.8	**96.1**	95.7	95.8	**92.6**	96.4	98.8	94.5	100	**99.4**	**100**	99.3
frozen-BioBERT	86.3	68.5	95.2	88.5	57.7	94.3	92.3	59.3	98.6	88.7	52.7	96.2	97.1	95.5	98.2
Neuroradiologist	**96.8**	**96.2**	**97.1**	**96.9**	84.6	**99.3**	**96.4**	77	**100**	**99.4**	**97.2**	**100**	99.3	96.1	**100**

- NLP labelling on basis of reports comparable to expert neuroradiologist
- Do reports agree with images? normal/abnormal - yes, granular - mostly (see Wood et al. 2020)
- 120, 000 MRI images labelled in < 0.5 hours

Results

- Granular classification

Model	damage			vascular			mass			acute stroke			Fazekas		
	acc	sens.	spec.	acc	sens.	spec.	acc	sens.	spec.	acc	sens.	spec.	acc	sens.	spec.
Our model	93.8	92.6	94.3	95.8	**96.1**	95.7	95.8	**92.6**	96.4	98.8	94.5	100	**99.4**	**100**	99.3
frozen-BioBERT	86.3	68.5	95.2	88.5	57.7	94.3	92.3	59.3	98.6	88.7	52.7	96.2	97.1	95.5	98.2
Neuroradiologist	**96.8**	**96.2**	**97.1**	**96.9**	84.6	**99.3**	**96.4**	77	**100**	**99.4**	**97.2**	**100**	99.3	96.1	**100**

- NLP labelling on basis of reports comparable to expert neuroradiologist
- Do reports agree with images? normal/abnormal - yes, granular - mostly (see Wood et al. 2020)
- 120, 000 MRI images labelled in < 0.5 hours

Labelling imaging datasets on the basis of neuroradiology reports: a clinical perspective

David A. Wood[1], Sina Kafiabadi[2], Aisha Al Busaidi[2], Emily Guilhem[2], Jeremy Lynch[2], Matthew Townend[3], Antanas Montvila[2], Juveria Siddiqui[2], Naveen Gadapa[2], Matthew Benger[2], Gareth Barker[4], Sebastian Ourselin[1], James H. Cole[4,5], and Thomas C. Booth[1,2]

MIDI consortium

Interpretability

- Inspection of attention weights allows form of model interpretability

Label: abnormal study

Clinical Details: Sudden onset dysarthria following AVR 5 days ago. CT head NAD. Specific question to be answered: is there evidence of embolic stroke on DWI?
MRI Head: There is an acute cortical infarct involving the left inferior frontal gyrus. There are no mass lesions or intracranial collections. The remainder of the intracranial appearances are normal

Label: normal study

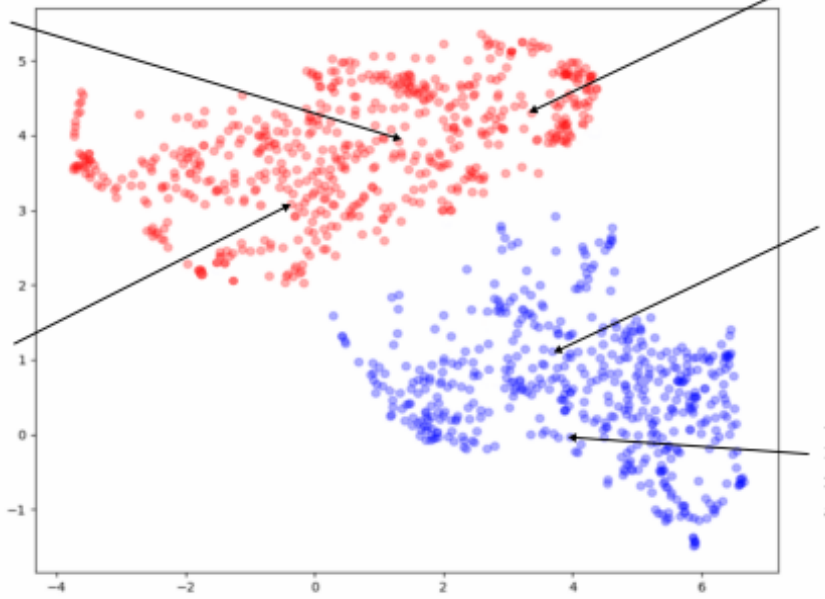
Clinical Details: Left sided mild facial droop with drooling of saliva. Specific question to be answered: stroke?
MRI Head: Normal intracranial appearances. In particular there are no sites of diffusion restriction or abnormal parenchymal susceptibility. There are no sites of established infarction.

Semi-supervised labelling

- Pathology-dependent clustering in predicted binary labels allows semi-supervised labelling of granular datasets (e.g. Alzheimer's, high grade glioma etc.)
- "Lasso" too available at https://github.com/tomvars/sifter

MRI Head with contrast : There is a mixed T2 signal, heterogeneously enhancing, (peripheral predominance) lesion centred on the genu and the body of the corpus callosum. It extends into the frontal white matter of both cerebral hemispheres. It measures 53mm (TR) x 36mm (AP) x 40mm (CC). There is mild surrounding peri-lesional white matter signal change. A right molar periapical abscess is noted. This is most in keeping with a high grade glioma.

MRI Head : There is a 4.1 x 3.6cm partly solid with heterogenous T2w signal and partly fluid mass which is exophytic from the floor of the anterior part of the body of the right lateral ventricle. The medial wall abuts the septum pellucidum and causes deviation to the left. There is also an intraparenchymal component of the mass involving the posterior part of the inferior right frontal lobe with surrounding T2w hyperintensity also spreading into the right temporal lobe and through the genu of the corpus callosum into the left frontal lobe. The appearances are suggestive of a high grade glioma.



There is a 44 x 18mm solitary heterogeneously enhancing irregular left cingulate intra-axial mass with subfalcine extension to the right. There is positive mass effect of 6mm to right of the midline. No evidence of hydrocephalus seen. There is extensive T2 hyperintensity in the left cerebral hemisphere and the corpus callosum. No haemorrhage or recent infarct demonstrated. There is evidence of gliosis in the right cerebellar hemisphere/bilateral vermis likely of ischaemic origin. There is an incidental 4 mm right MCA aneurysm. CONCLUSION: Appearance would be consistent with a high grade glioma.

Clinical Details: 4th nerve palsy in context of migraine.
MRI Head : Axial T2/DWI, Coronal FLAIR, Sagittal T1 Brain. Coronal T1/STIR, axial and Coronal T1 FS post contrast orbits. There are normal intracranial and orbital appearances with no obvious structural cause found for symptoms.

Axial T2, T2 GE and DWI sequences acquired. There are normal intracranial appearances. Normal flow voids are present. No foci of restricted diffusion. Both CPA angles have normal appearances. Normal appearances of the posterior fossa. Conclusion: Normal intracranial

Conclusion

- Dedicated MRI neuroradiology report classifier for automatic image labelling
- Binary classification performance outperforms trained neurologist
- Granular classification performance comparable to experienced neuroradiologist
- 120,000 radiology reports and corresponding MRI scans labelled in < 0.5 hours

Acknowledgements

This work was supported by The Royal College of Radiologists, King's College Hospital Research and Innovation, King's Health Partners Challenge Fund and the Wellcome/Engineering and Physical Sciences Research Council Center for Medical Engineering (WT 203148/Z/16/Z). We also thank Joe Harper, Justin Sutton, Mark Allin and Sean Hannah at KCH for their informatics and IT support, Ann-Marie Murtagh at KHP for research process support, and KCL administrative support, particularly from Denise Barton and Patrick Wong.