

Project Title

Evaluating the Compliance of MRI Brain Scans to American College of Radiology (ACR)

Ordering Guidelines Using Text Mining

Project Lead and Members

Tang PH, Jing Xuan, Liu Nan, Lam Shaowei, Joshua Lim, Louis McAdory, Chan Ling Ling

Organisation(s) Involved

KK Women's and Children's Hospital, Singapore General Hospital, SingHealth

Healthcare Family Group(s) Involved in this Project

Allied Health

Applicable Specialty or Discipline

Radiology

Aims

To develop a predictive model using text mining to automate this classification process

Background

See poster appended/ below

Methods

See poster appended/ below

Results

See poster appended/ below

Conclusion

See poster appended/ below



Project Category

Training & Education

Education Research, Analytics

Keywords

Indications for Mri Scan, Compliance with Guidelines, Text Mining, Machine Learning Algorithm

Name and Email of Project Contact Person(s)

Name: Tang PH

Email: singaporehealthcaremanagement@singhealth.com.sg

Evaluating the Compliance of MRI Brain Scans to American College of Radiology (ACR) **Ordering Guidelines Using Text Mining.**

Singapore Healthcare Management 2018

Tang PH, Jing Xuan, Liu Nan, Lam Shaowei, Joshua Lim, Louis McAdory, Chan Ling Ling



KK Women's and Children's Hospital





Introduction and Aim

Our previous research showed that MRI brain scans with indications



SingHealth

There were 2911 MRI brain scans done in KKH and SGH of which 2262 scans were ordered according to the given 16 ACR guidelines.

following the American College of Radiology (ACR) 2013 guidelines show a higher percentage of brain abnormalities compared to scans not following guidelines.

Process of manually classifying these indications according to whether or not they are following the ACR guidelines is labor intensive and time consuming.

Aim of our study is to develop a predictive model using text mining to automate this classification process.

Methods

Anonymized reports of patients who had MRI brain scans in KK Women's and Children's Hospital (KKH) and Singapore General Hospital (SGH) were collected.

Reports were manually categorized by human operators in each hospital as to whether or not they conformed to the ACR ordering guidelines and key words identified.

80% of all subjects were randomly selected for model derivation and the rest of the 20% subjects for model validation.

ACR	ACR criteria		cases	cases
category			in KKH	in SGH
no.			(2006 to	o (2014 to
			2013)	2015)
1	Neoplastic conditions		2	22 64
2	Vascular		1	18 200
3	Congenital disorders and anatomical		1	12 1
	abnormalities			
4	Congenital or acquired neurodegenerat	tive		2 8
	disorders			
5	Congenital or acquired hydrocephalus		· · · · · · · · · · · · · · · · · · ·	45 2
6	Metabolic disorders			15 0
7	Trauma			70 3
8	Hemorrhage		· · · · · · · · · · · · · · · · · · ·	45 9
9	Inflammatory and autoimmune disorde	ers		77 5
10	Infectious disorders		13	37 12
11	Endocrine disorders			34 0
12	Evaluation of cranial nerves			79 23
13	Epilepsy and movement disorders		4	14 23
14	Psychiatric disorders			10 5
15	Follow-up of treatment		4	60 13
16	Image guidance			17 7
ACR	ACR criteria	Top ke	ey words	Top key
ACR category	ACR criteria	Top ke in KKH	ey words I dataset	Top key words in SGH
ACR category no.	ACR criteria	Top ke in KKH	ey words I dataset	Top key words in SGH dataset
ACR category no. 1	ACR criteria Neoplastic conditions	Top ke	ey words I dataset tumour	Top key words in SGH dataset metastasis
ACR category no. 1 2	ACR criteria Neoplastic conditions Vascular	Top ke	ey words dataset tumour stroke	Top key words in SGH dataset metastasis stroke
ACR category no. 1 2 3	ACR criteriaNeoplastic conditionsVascularCongenital disorders and anatomical	Top ke in KKH	ey words dataset dataset tumour stroke lopmental	Top key words in SGH dataset metastasis stroke
ACR category no. 1 2 3	ACR criteria ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities	Top ke in KKH deve	ey words dataset dataset tumour stroke lopmental delay	Top key words in SGH dataset metastasis stroke
ACR category no. 1 2 3 3	ACR criteria ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired	Top ke in KKH deve neuro	ey words dataset dataset tumour stroke lopmental delay regression	Top key words in SGH dataset metastasis stroke degeneration
ACR category no. 1 2 3 3	ACR criteria ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders	Top ke in KKH deve neuro	ey words dataset tumour stroke lopmental delay regression	Top key words in SGH datasetMatasetmetastasis strokedegeneration
ACR category no. 1 2 3 3 4 4	ACR criteria ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression	Top key words in SGH datasetMatasetmetastasis strokedegenerationhydrocephalus
ACR category no. 1 2 3 3 4 4 5 5 6	ACR criteria ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression cocephalus MELAS	Top key words in SGH datasetMatasetmetastasis strokedegenerationhydrocephalus
ACR category no. 1 2 3 3 4 4 5 5 5 6 5 6 7	ACR criteria ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders Trauma	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression cocephalus MELAS injury	Top key words in SGH datasetmetastasis strokedegenerationhydrocephalusconcussion
ACR category no. 1 2 3 3 4 5 5 5 6 5 6 5 6 7 8	ACR criteria ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders Trauma Hemorrhage	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression regression sinjury emorrhage	Top key words in SGH datasetmetastasis strokedegenerationhydrocephalusconcussion bleed
ACR category no. 1 2 3 3 4 4 5 5 6 5 6 5 6 7 8 8 9	ACR criteria ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders Trauma Hemorrhage Inflammatory and autoimmune	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression delay regression	Top key words in SGH dataset metastasis stroke degeneration hydrocephalus concussion bleed lupus
ACR category no. 1 2 3 3 4 4 5 5 6 5 6 7 8 8 9	ACR criteria ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders Trauma Hemorrhage Inflammatory and autoimmune disorders	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression cocephalus MELAS injury emorrhage vasculitis	Top key words in SGH dataset metastasis stroke degeneration hydrocephalus concussion bleed lupus
ACR category no. 1 2 3 3 4 4 5 6 5 6 5 6 7 8 8 9 9	ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders Trauma Hemorrhage Inflammatory and autoimmune disorders Infectious disorders	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression delay regression sinjury emorrhage vasculitis	Top key words in SGH dataset metastasis stroke degeneration hydrocephalus concussion bleed lupus encephalitis
ACR category no. 1 1 2 3 3 4 4 4 5 6 5 6 5 6 7 5 6 7 8 9 9 9 10 11	ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders Trauma Hemorrhage Inflammatory and autoimmune disorders Infectious disorders Endocrine disorders	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression regression socephalus MELAS injury morrhage vasculitis	Top key words in SGH dataset metastasis stroke degeneration hydrocephalus concussion bleed lupus encephalitis
ACR category no. 1 2 2 3 3 4 4 5 5 6 5 6 5 6 7 8 9 9 10 11 10 11 12	ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders Trauma Hemorrhage Inflammatory and autoimmune disorders Infectious disorders Endocrine disorders	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression ocephalus MELAS injury emorrhage vasculitis	Top key words in SGH dataset metastasis stroke degeneration hydrocephalus concussion bleed lupus encephalitis
ACR category no. 1 2 3 3 4 4 5 6 5 6 5 6 5 6 7 8 9 9 10 10 11 10 11 12 12	ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders Trauma Hemorrhage Inflammatory and autoimmune disorders Infectious disorders Endocrine disorders Evaluation of cranial nerves	Top ke in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression ocephalus MELAS injury emorrhage vasculitis meningitis meningitis	Top key words in SGH dataset metastasis stroke degeneration hydrocephalus concussion bleed lupus encephalitis
ACR category no. 1 2 3 3 3 4 4 5 6 5 6 5 6 7 5 6 7 8 9 9 10 10 11 11 12 12 13 13	ACR criteria Neoplastic conditions Vascular Congenital disorders and anatomical abnormalities Congenital or acquired neurodegenerative disorders Congenital or acquired hydrocephalus Metabolic disorders Congenital or acquired hydrocephalus Metabolic disorders Trauma Hemorrhage Inflammatory and autoimmune disorders Infectious disorders Endocrine disorders Evaluation of cranial nerves Epilepsy and movement disorders	Top kei in KKH deve neuro hydr	ey words dataset dataset tumour stroke lopmental delay regression meringitis meningitis meningitis diplopia seizures dyskinesis	Top key words in SGH dataset metastasis stroke degeneration hydrocephalus concussion bleed lupus encephalitis diplopia seizure psychosis

Both term frequency and term frequency-inverse document frequency (TF-IDF) were used as the weighting factor in text mining. Term frequency-inverse document frequency is a weight widely used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the collection.

3 machine learning algorithms using the key words identified to predict whether or not the MRI brain scan follows ACR guidelines. Algorithms used were -generalized linear model, -random forest and

-support vector machines (SVM).

Area under the receiver operating characteristic curve (AUC) was used to compare the predictive models.

Model Development Overview

	Data Cleaning		> Combined dataset		
KKH-ACR →	-Lowercase		K	N	
	-Remove numbers	-	Fraining	Test set fo	
KKH-non ACR 🛶	-Remove whitespace	9	set (80%)	validation	
	-Remove stopwords	/		(20%)	
SGH-ACR →	-Word stemming				
		K	V	A	
SGH-non ACR →	Convert to term	Genera	Random	Support	
	frequency-inverse	Linear	Forest	Vector	
	document frequency	Model		Machine	
	format.				

Image guidance 16

surgery decompression

Based on term frequency-inverse document frequency weighting factor, performance was random forest model (AUC=0.84) generalised linear model (AUC=0.73) support vector machines (AUC=0.72).



Conclusion

t for

The discriminatory ability of the predictive models is best using random forest for automated classification of MRI brain scans indications.