Singapore Healthcare Management 2018

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

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

KK Women's and Children's Hospital
SingHealth

Singapore General Hospital

SingHealth
Defining Tomorrow's Medicine

Introduction and Aim

Our previous research showed that MRI brain scans with indications 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.

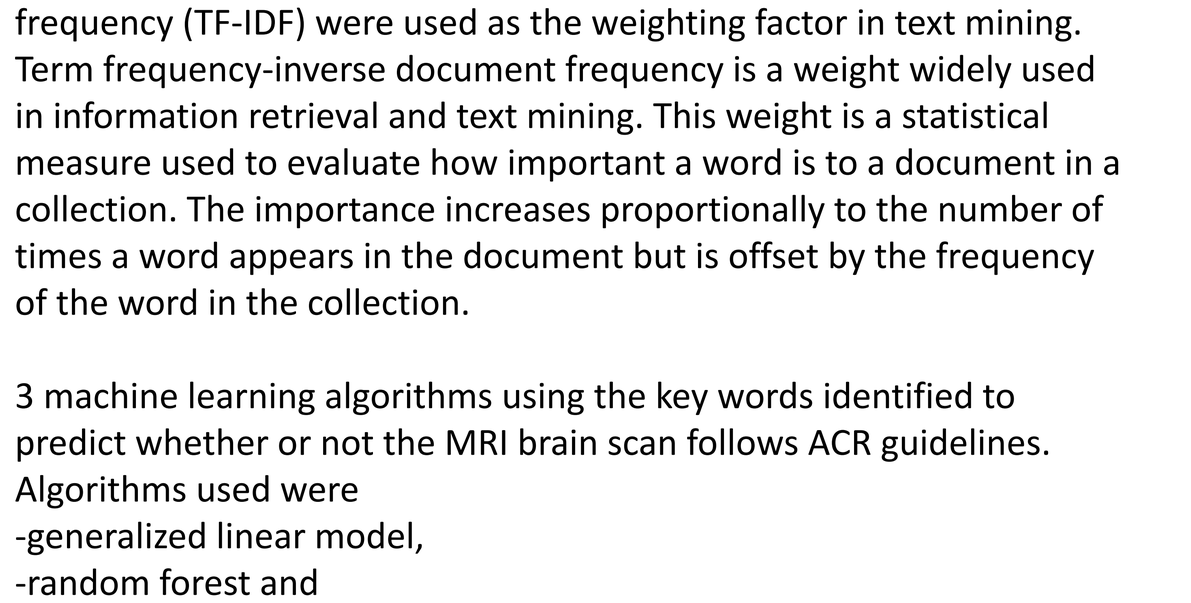
80% of all subjects were randomly selected for model derivation and the rest of the 20% subjects for model validation.

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



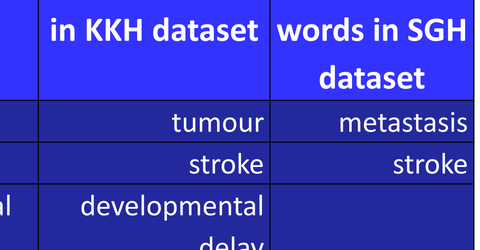
Results

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

ACR category no.	ACR criteria	cases in KKH (2006 to 2013)	cases in SGH (2014 to 2015)
1	Neoplastic conditions	222	64
2	Vascular	118	200
3	Congenital disorders and anatomical abnormalities	142	1
4	Congenital or acquired neurodegenerative disorders	2	8
5	Congenital or acquired hydrocephalus	45	2
6	Metabolic disorders	15	0
7	Trauma	70	3
8	Hemorrhage	45	9
9	Inflammatory and autoimmune disorders	77	5
10	Infectious disorders	137	12
11	Endocrine disorders	34	0
12	Evaluation of cranial nerves	79	23
13	Epilepsy and movement disorders	414	23
14	Psychiatric disorders	10	5
15	Follow-up of treatment	460	13
16	Image guidance	17	7

ACR category no.	ACR criteria	Top key words in KKH dataset	Top key words in SGH dataset
1	Neoplastic conditions	tumour	metastasis
2	Vascular	stroke	stroke
3	Congenital disorders and anatomical abnormalities	developmental delay	
4	Congenital or acquired neurodegenerative disorders	neuroregression	degeneration
5	Congenital or acquired hydrocephalus	hydrocephalus	hydrocephalus
6	Metabolic disorders	MELAS	
7	Trauma	injury	concussion
8	Hemorrhage	haemorrhage	bleed
9	Inflammatory and autoimmune disorders	vasculitis	lupus
10	Infectious disorders	meningitis	encephalitis
11	Endocrine disorders	pituitary	
12	Evaluation of cranial nerves	diplopia	diplopia
13	Epilepsy and movement disorders	seizures	seizure
14	Psychiatric disorders	dyskinesis	psychosis
15	Follow-up of treatment	medulloblastoma	glioma
16	Image guidance	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

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