Iowa Initiative for Artificial Intelligence

Final Report

Project title:	Machine Learning for Predicting the Development of Damage and Lesion Distribution Patterns on Whole Body MRI in Children with Chronic Recurrent Multifocal Osteomyelitis
Principal Investigator:	Takashi Shawn P. Sato, Aleksander Lenert
Prepared by (IIAI):	Yanan Liu
Other investigators:	
Date:	

Were specific aims fulfilled:	Yes
Readiness for extramural proposal?	No
If yes … Planned submission date	
Funding agency	
Grant mechanism	
If no … Why not? What went wrong?	Accuracy was not sufficiently good due to a small data set.

Brief summary of accomplished results:

We have developed and validated a Random Forest model to accurately predict diagnosis (damage /non-damage) of chronic recurrent multifocal osteomyelitis (CMRO) using bone lesion patterns. The achieved best prediction accuracy was 0.74 using combined bone lesion locations.

Research report:

Aims (provided by PI):

The overarching goal of our proposal is to use a novel data-driven machine learning (ML) approach to characterize MRI bone lesion patterns in a well-characterized longitudinal cohort of children with CRMO. Specifically, we propose to use a ML approach applied to bone lesion patterns that predict future bone damage (high-risk patient group).

Data:

Following an internal UIHC data search, we have access to a longitudinal cohort of children with CRMO (n=106 who have undergone serial WB-MRI's over the last 6 years (n=~300scans). 106 children with CRMO were identified with the following distribution of brain pathologies (damage: 42; non-damage: 64). Among those 106 children, only 74 had the second scan (damage:30; non-damage: 44).

AI/ML Approach:

Supervised machine learning algorithm was implemented for classification using Python. As many extracted features may be noisy, or highly correlated with each other, Random Forest (RF) algorithm was selected to classify damage/non-damage group. Due to a small data set, 5-fold cross validation was used for the RF model training and performance evaluation.

Experimental methods, validation approach:

All WB-MRI scans have been interpreted and annotated in detail with unique lesions characterized by location, temporal change, and associated findings. Specifically, each lesion is localized to one of 220 different locations. For each bone lesion, we have the following information: presence or absence of lesion on prior imaging (if available), and changes in edema pattern since immediate prior imaging (if available). Additionally, for each bone lesion, we have recorded presence or absence of soft tissue swelling. For each scan, we have 3 types of features: 8 lesion evolution features, 220 lesion location features and 220 lesion inflammation features.

Each subject has a different number of available scans. Total number of patients is 106, only 74 has second scans and only 49 has third scans. Considering number of subjects, we focused on cases who had 1 or 2 scans – and thus used data from the first two scans. Due to sparsity of data, we reduced the input feature by combining lesion locations. 4 different grouping methods (provided by the PI based on lesion locations) were used: 22 groups for method 1, 10 groups for method 2, 5 groups for method 3 and 17 groups for method 4. The detailed group information is in Table 1-4.

Table 1 Grouping method 1

Group number	Group name
1	Mandible
2	Clavicle
3	Scapula
4	Humerus
5	Radius
6	Ulna
7	Wrist
8	Metacarpals
9	Ribs
10	sternum/manubrium
11	cervical spine
12	thoracic spine
13	lumbar spine
14	Sacrum
15	Pelvis
16	Femur
17	Patella
18	Tibia
19	Fibula
20	hind/midfoot
21	Forefoot
22	Toes

Table 2 Grouping method 2

Group number	Group name
1	mandible
2	shoulder girdle
3	Arm
4	wrist
5	hand
6	ribs/ant chest
7	spine
8	pelvis
9	Leg
10	Foot

Table 3 Grouping method 3

Group number	Group name
1	mandible
2	upper limb
3	ribs/ant chest
4	spine
5	lower limb

Table 4 Grouping method 4

Group number	Group name
1	mandible
2	Anterior chest
3	shoulder
4	Upper arm
5	elbow
6	forearm
7	wrist
8	hand
9	Upper back
10	Neck
11	Lower back
12	Hip
13	upper leg
14	knee
15	lower leg
16	ankle

17	Foot

Among 106 patients, only 74 patients have second scan. Missing data was replaced by default -9999 for the remaining of 32 out of the total 106 patients. As already mentioned above, due to a small data set, 5-fold cross validation was performed to evaluate the predictive performance of each model. Different combinations of features were tested to build competing RF models. The mean accuracies are shown in table 5.

Table 5 Accuracy

Input features	Total number of features	Mean accuracy for 106 patients	Mean accuracy for 74 patients
loc+evo (8+220)*2	456	0.53	0.68
loc+evo+inf(8+220+220)*2	896	0.55	0.61
loc+evo+ inf+combined infl1(8+220+220+22)*2	940	0.53	0.66
loc+evo+ inf+combined infl2(8+220+220+10)*2	916	0.57	0.61
loc+evo+ inf+combined infl3(8+220+220+5)*2	906	0.58	0.61
loc+evo+ inf+combined infl4(8+220+220+17)*2	930	0.54	0.68
loc+evo+combined infl1(8+220+22)*2	500	0.55	0.62
loc+evo+combined infl2(8+220+10)*2	476	0.53	0.64
loc+evo+combined infl3(8+220+5)*2	466	0.57	0.69
loc+evo+combined infl4(8+220+17)*2	490	0.49	0.66
loc+evo+combined loc1(8+220+22)*2	500	0.59	0.68
loc+evo+combined loc2(8+220+10)*2	476	0.6	0.68
loc+evo+combined loc3(8+220+5)*2	466	0.58	0.64
loc+evo+combined loc4(8+220+17)*2	490	0.59	0.68
combined loc1 (22*2)	44	**0.7**	**0.74**
combined loc2 (10*2)	20	0.64	0.67
combined loc3 (5*2)	10	0.62	0.64
combined loc4 (17*2)	34	0.65	0.72
combined loc1+combined infl1(22*4)	88	0.71	0.7
combined loc2+combined infl2(10*4)	40	0.64	0.65
combined loc3+combined inf3l(5*4)	20	0.65	0.65
combined loc4+combined inf4l(17*4)	68	0.61	0.73
combined loc1+evo (22+8)*2	58	0.62	0.66
combined loc2+evo (10+8)*2	34	0.65	0.66

combined loc3+evo (5+8)*2	24	0.55	0.58
combined loc4+evo (17+8)*2	48	0.58	0.62
combined loc1+combined infl1+evo(22+22+8)*2	102	0.6	0.64
combined loc2+combined inf2+evo(10+10+8)*2	54	0.6	0.64
combined loc3+combined inf13+evo(5+5+7)*2	34	0.58	0.6
combined loc4+combined inf14+evo(17+17+7)*2	82	0.58	0.68

Loc : location features; evo: evolution features; infl: inflammation features. Combined loc 1-4: location features with grouping method 1-4; combined infl 1-4: inflammation features with grouping method 1-4.

Results:

For the task of damage group prediction, the best mean accuracy was 0.7 using location features with grouping method 1 for all 106 subjects. After removing missing data, the accuracy increased to 0.74 with 74 subjects.

Ideas/aims for future extramural project:

The project suffers from a small dataset and lesions located at 220 locations (too sparse).

More data is needed and will be collected – but that will likely be a slow process.

Publications resulting from project:

None.