Neurocomputing 000 (2017) 1–8

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Random forest and Self Organizing Maps application for analysis of pediatric fracture healing time of the lower limb

Sorayya Malek[a,*], R. Gunalan[b], S.Y. Kedija[a], C.F. Lau[a], Mogeeb A.A. Mosleh[a], Pozi Milow[a], S.A. Lee[b], A. Saw[b]

[a] *Institute of Biological Sciences, Faculty of Science, University of Malaya, Malaysia*
[b] *Department of Orthopaedics / NOCERAL, Faculty of Medicine, University of Malaya, Malaysia*

ARTICLE INFO

Article history:
Received 16 October 2016
Revised 8 March 2017
Accepted 10 May 2017
Available online xxx

Keywords:
Artificial neural networks
Lower limb fractures
Random forest
Immature skeleton
Kohonen Self Organizing Maps

ABSTRACT

In this study, we examined the lower limb fracture healing time in children using random forest (RF) and Self Organizing feature Maps (SOM) methods. The study sample was obtained from the pediatric orthopedic unit in University Malaya Medical Centre. Radiographs of long bones of lower limb fractures involving the femur, tibia and fibula from children ages 0–12 years, with ages recorded from the date and time of initial injury. Inputs assessment extracted from radiographic images included the following features: type of fracture, angulation of the fracture, contact area percentage of the fracture, age, gender, bone type, type of fracture, and number of bone involved. RF is initially used to rank the most important variables that effecting bone healing time. Then, SOM was applied for analysis of the relationship between the selected variables with fracture healing time. Due to the limitation of available dataset, leave one out technique was applied to enhance the reliability of RF. Results showed that age and contact area percentage of fracture were identified as the most important variables in explaining the fracture healing time. RF and SOM applications have not been reported in the field of pediatric orthopedics. We concluded that the combination of RF and SOM techniques can be used to assist in the analysis of pediatric fracture healing time efficiently.

1. Introduction

A fracture is defined as a break in the continuity of the bone involving the cortex of the bone. A fracture is a partial or complete break in the bone [1]. Fractures in children (aged 0–12 years) have considerably different features as opposed to fractures in adults. Skeletal trauma accounts for 15% of all injuries in children [2]. There are several types of fracture such as a transverse, spiral and torus. Transverse fracture occurs as the fracture passes at right angles to the shaft of the long bone. Spiral type of fractures passes at an angle oblique to the shaft of the long bone. Torus (or buckle) fracture is a unique type of incomplete fracture resulting because of children's bones has a thick periosteal layer.

Lower limb fracture in children usually takes half the time of the corresponding fracture in adults [3]. In pediatric cases, it is important for assessment of skeletal trauma as trauma may be a sign of non-accidental injury or abnormal healing, indicating an underlying medical condition affecting bone healing. While rates have been published for a normal bone healing process in adults, very little is known about healing rates in the pediatric population. Pediatric bone physiology indicates that younger individuals heal at a faster rate as compared to adults [4].

Lower limb long bones can be divided into three sections; the femur, tibia and fibula. The femur is the longest bone in the body. The main function of the femur is to transmit forces from the hip joint to the tibia. The tibia is the second largest bone in the body. It expands at the proximal and distal ends, articulating at the knee and ankle joints respectively. The fibula, along with the tibia, makes up the bones of the leg. Long bone fractures (taking into account the diaphyseal region of the bone) are described with reference to the direction of the fracture line in relation to the shaft of the bone. Few articles reported the evaluation of classification on pediatric fracture healing on the basis of radiographic fracture and statistical approach to determine healing rates. Correlating healing time with the chronologic history of injury, may aid in injuries that are healing abnormally or may indicate non-accidental injury [5–7]. The system for predicting healing time required should serve as a tool in the process of treatment for general practitioners and medical officers and in the follow-up period.

* Corresponding author.
E-mail addresses: sorayya@um.edu.my (S. Malek), roshan@um.edu.my (R. Gunalan).

http://dx.doi.org/10.1016/j.neucom.2017.05.094
0925-2312/

In recent years many machine learning algorithms have been developed that can be used as efficient tools for the analysis of databases and for extracting classification knowledge that can be used to solve new problems in the given problem domain [8–10]. Machine learning methods have been applied in various orthopedic fields [11–14]. Application of machine learning techniques such as K-nearest neighbors algorithm, the semi-naive Bayesian classifier, and back propagation with weight elimination learning of the multilayered neural in fracture recovery have been proven feasible [15–17].

However applications of machine learning methods have not been reported in the pediatric orthopedic field. The current study on prediction of fracture healing time in pediatric population is an extension of previous study by Sorayya et al. [18]. Previous study has reported pediatric fracture healing time (categorized into low and high healing time) using ANNs. Back-propagation multilayer perceptron (MLP) was used for supervised ANN. Kohonen self-organizing feature map (SOM) was used for the unsupervised learning [18]. Multilayer perceptron (MLP) is a supervised ANN learning method; this technique provides little insight to the significance of variables against the predictor [19]. Transparency is important in medical decision support system. This can be achieved by using different machine learning techniques such as RF and SOM to enhance predication and classification [20, 21].

SOM has been proven in its suitability to classify the dataset for osteoporosis classification problem of high and low osteoporosis risk, and it's usage in investigation of osteoporosis dataset [22]. The SOM was reported as an excellent tool in the visualization of high dimensional data. SOM reduces the dimensions of data of a high level of complexity and plots the data similarities through clustering technique [23].

RF method is developed by Breiman as classification and regression method based on the aggregation of large number of decision trees built by using several bootstrap samples [24]. Random forests (RF) is a machine learning method that is a specific instance of bagging. The two byproducts of RF method are out-of-bag (OOB) estimates of generalization error and variable importance measurements [24–27]. RF method has reported better accuracy compared to other supervised learning methods such as MLP and support vector machines (SVM) [24, 26]. RF has been applied in a variety of applications in computational biology and medicine where the relationship between response and predictors is complex and the predictors are strongly correlated [28–30]. However, application of RF in orthopedics, especially in pediatric orthopedic field has not been reported yet.

This study attempts to estimate pediatric fracture healing times of the lower limbs by employing different machine learning methods such as combination of SOM and RF methods. In this study, new methods have been applied, and extra variables have been selected according to its high importance with the predictor (healing weeks) as the quantification of variable importance is a crucial issue for understanding data in applied problems.

This article is organized as follows. Section 2 provides the material and methods used in this study, including application development techniques, and data analysis. Section 3 explores the study results. Section 4 discusses the finding of this study. Section 5 presents our conclusion within the expected scope and the limitations of this research.

2. Material and methods

2.1. Data

A collection of four years of patient data and radiographs from the years 2009, 2010, 2011 and 2014, respectively, were obtained from the University Malaya Medical Centre pediatric orthopedic unit, Orthopaedic Department in Kuala Lumpur, Malaysia. Radiographs of fractured bones (femur, tibia and fibula) from infants and young children of ages less than 12 years were included, with ages recorded from the time of initial injury. The individuals in the 57 samples of children age 12 and below from the time of injury are from the surrounding Selangor population. The radiographs examined consisted of images from each individual. The radiograph images were analyzed by a Pediatric Orthopedic surgeon. Any individuals demonstrating comorbidity or any systemic disorder, which may affect the bone healing rate, were excluded from the study. Data was retrieved based on radiography and patient records. Through radiograph examinations, variables such as bone involve, region of bone, type of fracture and measurement parameters such as angulation of the fracture (in coronal and sagittal planes) and contact area of the fracture were obtained. Diameter of the fractured bone, in two views, anterior and lateral were also analyzed. The time interval between injury and the union of the bone, age and sex of the patient and other demographic factors are also identified. Healing time was defined as the time in which the bone achieved union based on radiographic evidence. Healing was defined at the time the radiograph showed that the fracture line was absent, and that the cortices between the fracture sides were well formed. The remodeling of the bone, thereafter, was not taken into consideration for this study.

Table 1
Summary statistic of continous variable used in the RF and SOM developement.

Variable	Min.	Max.	Median	Standard deviation
Age (years)	0.16	13	8.5	3.92273
Lateral contact area (%)	0	100	100	32.0798
Lateral diameter (mm)	6.9	41.6	15.3	7.30651
Lateral contact area (mm)	0	41.6	12.6	8.95033
Lateral angulation (deg)	0	35	2.5	6.61140
Anterior diameter (mm)	5.9	42.3	14.8	7.81365
Anterior contact area (mm)	0	42.3	12	9.77492
Anterior angulation (deg)	0	46	3	7.86286
Anterior contact area (%)	0	100	91.3	33.6084
Healing weeks (output)	3	12	8	2.816

2.1.1. Data analysis

The parameters used in this study, were lateral (sagittal plane) and anterior (coronal plane) angle and contact area and age. The measurement is taken based on radiograph features [31]. Angulation describes the direction of the distal bone and degree of angulation in relation to the proximal bone. Loss of alignment or displacement is usually accompanied by some degree of angulation, rotation or change in bone length. Contact area is described as how much the bone is in contact with each other, taking into account the amount of contact in two radiograph views, in anterior and lateral views. Angulation and contact are interrelated to each other [32]. Besides continuous variables categorical variables used in this study are type of fracture, bone involved, race, gender, bone segment, fracture segment. Summary statistics of the continuous variable used in this study given in Table 1, where Table 2 displays categorical variables used in this study.

In the absence of a very large designated dataset in this study. We proposed the usage of Leave-one-out method. This method has a couple of major advantages such as; it has far less bias and provides an approximately unbiased estimate for the test error. In this method a single observation $(x1, y1)$ is used for the testing set, and the remaining observations $\{(x2, y2), \ldots, (xn, yn)\}$ is made up the training set. The RF model is built on the $n-1$ training observations and a prediction $\hat{y}1$ is made for the excluded observation, using its value $x1$. Mean square error value (MSE) is calculated as