AN INTELLIGENT METHOD FOR THE PROGNOSIS OF FEMORAL HEAD OSTEONECROSIS

A.E. Bassounas, F.I. Fotiadis and K.N. Malizos

12–2002

Preprint, no 12–02/2002

Department of Computer Science
University of Ioannina
45110 Ioannina, Greece
An Intelligent Method for the Prognosis of Femoral Head Osteonecrosis

AthanasiouS. BassounaS(1), Dimitrios I. Fotiadis(2)* and Konstantinos N. Maligos(3)

(1) Dept of Medical Physics, School of Medicine, University of Ioannina, GR 451 10 Ioannina, GREECE

(2) Unit of Medical Technology and Intelligent Information Systems, Dept of Computer Science, University of Ioannina, and Biomedical Research Institute – FORTH, GR 451 10 Ioannina, GREECE

(3) Orthopaedic Department, Medical School, University of Thessalia, GR 412 22 Larissa, GREECE

*Corresponding author

Dr D. I. Fotiadis

Unit of Medical Technology and Intelligent Information Systems

Dept of Computer Science

University of Ioannina

GR 451 10 Ioannina, GREECE

Tel.: +30 6510 98803

Fax: +30 6510 97092

e-mail: fotiadis@cs.uci.gr
Abstract

An automated method for the prognosis of femoral head osteonecrosis is introduced. The method consists of a module for the quantitative assessment of osteonecrosis and a classification module based on an artificial neural network. The semi-automatic volumetric module is used for the computation of the size and distribution of the necrotic lesion within the femoral head using magnetic resonance images. The necrotic lesion features extracted from the first module are used in the classification module to provide with a decision on the disease prognosis. The system was developed and evaluated using data from 57 hips with osteonecrosis. They were treated using a hip preserving method (free vascularized graft) and had a long-term follow-up. The system presents high performance in classifying the hips according to their prognosis (good or poor). The area under the ROC curve using the leave-one-out cross-validation method is 0.916. The achieved sensitivity of the method is 0.85 and the corresponding specificity and accuracy are 0.89 and 0.88, respectively.

Keywords: Osteonecrosis; MRI; Automated Prognosis
1. Introduction

Femoral head necrosis is a multifactorial and potentially disabling condition of the hip. It is responsible for a large number of hip arthroplasties, especially in young patients. Various operative methods have been utilized for the treatment of osteonecrosis. The major treatment objective is to preserve the femoral head, whereas, in advanced stages of the disease, total hip replacement is inevitable. Prognosis of osteonecrosis, i.e. knowledge about the probable course and outcome, is very important for the selection of the appropriate treatment. The problem for the orthopaedic surgeons is that prognosis of the disease depends on a number of diverse factors (radiographic stage, subchondral collapse, size and location of the lesion, presence of some risk factor, duration of symptoms), making the task of predicting the outcome of a particular patient difficult and ineffective [1-4].

Many studies have shown that the fate of femoral head necrosis is associated with the size, location and distribution of the necrotic lesion in the affected head [5-7]. A number of researchers have proposed systems for the evaluation of the extent of the necrotic lesion based on magnetic resonance imaging (MRI), which is the most suitable imaging technique for early diagnosis and for the volumetric description and positioning of the lesion. Most methods estimate the necrotic area or are in a single slice (or a limited number of slices) and consider this assessment as an index of the lesion extent. Only few methods attempt to evaluate the volume and location of the lesion, but they are characterized by their limited accuracy and repeatability since they use manual tracing of the necrotic lesion [6,8].

Femoral head prognosis is based on the utilization of a classification system. Several classification systems have been proposed, but their prognostic value, in predicting the fate of a particular patient has not been proven [9,10]. Although prognostic factors have been identified, orthopaedic surgeons are not able to perform reliable prediction of osteonecrosis outcome.
In this work we introduce an intelligent system for the prognosis of femoral head osteonecrosis. The system consists of a module for the quantitative assessment of osteonecrosis and a classification module based on an artificial neural network. The purpose of the first module is to extract volumetric features that describe the size and distribution of the necrotic lesion within the femoral head. Feature extraction is performed using a semiautomatic method for the processing and analysis of magnetic resonance images. The purpose of the second module is to classify hips for their prognosis. It is based on an artificial neural network, which discriminates hips with poor from hips with favorable prognosis, based on the volumetric features. The system can be a useful tool for the orthopaedic surgeon, supporting decision-making on the treatment of femoral head necrosis. To our knowledge no such system is mentioned in the literature.

2. Materials and Methods

2.1. Patients

We have developed a database containing text and image data for patients with osteonecrosis of the femoral head. The patient's record includes personal data, clinical examinations and laboratory examinations. Most of the patients of the database have been treated with free vascularized fibular grafting, a method that preserves the femoral head. Data of the operation and the postoperative examinations are also included in the patient's record. Moreover, the osteonecrosis database stores image data including pre-operative radiographs, MRI, computed tomography (CT) and bone scan and also post-operative images from the follow-up series.

Fifty-seven hips (42 patients) with femoral head necrosis were used for the development and evaluation of the proposed system. The patients were treated with free vascularized fibular grafting and complete pre-operative data and post-operative information were available. Pre-operative magnetic resonance images of the affected hips were taken and used for the evaluation
of the size and distribution of the necrotic lesion (the volumetric processing method is described on the next section). The hips included in the present study had a long-term follow-up. The mean follow-up period was 6.1 (2 to 11) years. Twenty hips failed within a period of 2 to 4 years after the operation. The femoral head collapsed and total hip replacement (THR) had to be applied. The femoral head was removed and replaced by an artificial joint. Those 20 hips are considered to have poor outcome. The remaining 37 hips had good clinical and radiological results at least 5 years after the operation (good outcome). The aim of the proposed system is to predict the outcome of the disease based on the features extracted using the volumetric method described below.

2.2. The volumetric method

A semi-automatic method has been developed for the volumetric description of the affected femoral head [11,12]. The method processes magnetic resonance images and extracts a number of features that describe the size and distribution of the lesion. T1-weighted coronal MRI sequences were used, with distance between slices ranging from 3.3 mm to 5 mm. The MRI films were digitized by a Vidar VXR 12 Film Digitizer (Vidar Systems Corporation, Va., USA).

The sphere equivalent model of the femoral head is considered [12]. The MRI slice with the largest projection of the femoral head is used for the definition of the radius of the sphere. The sphere is divided into eight octants by three planes (coronal, sagittal, and transverse) that intersect at the center of the sphere. The method calculates the percentage of the affected volume within each octant and also the overall necrotic percentage of the femoral head. Moreover, the method calculates the percentage of the femoral surface (superior hemisphere) affected by necrosis. The superior hemisphere is divided into a lateral and a medial quadrant and the necrotic area portion of each is calculated.
The calculation of the volumetric features is performed using area and arc measurements from the MRI slices. Lesion segmentation is performed using a semiautomatic image processing technique that requires limited user interaction [12]. In most cases, the system operator only needs to define the circle that circumscribes the femoral head in each MRI slice. Image processing methods such as thresholding and morphological operations are utilized to detect the lesion borders (Fig. 1). The necrotic area in each segment (quadrant) of the circle and the necrotic arc on the superior semicircle are measured. After performing this procedure for all the MRI slices, the method automatically calculates the corresponding volumetric features. Table 1 presents the 11 features extracted from the MRI images by the volumetric method. All features are expressed as percentages of necrotic volume or surface (values 0.0 to 1.0).

The accuracy of the method has been evaluated using 9 femoral heads removed from patients during hip hemi-arthroplasty. The specimens were trimmed and rounded-up with wax, to take the shape of a sphere. In order to simulate the affected segment, part of the core of the femoral head was curetted out and was then filled with a measured volume of surgical wax. MRI images of the specimens were processed using the presented method. The measurements of the method were on average 5% larger than the actual measurements. However, the percentages are not affected since this error affects both the lesion volume and the sphere volume by the same proportion.

2.3. The classification module

Not all eleven features extracted using the volumetric method present the same discriminant power for the prognosis of osteonecrosis. In fact, some of them are unrelated to the final outcome of the disease. We evaluated the relevance of each single feature to the outcome in order to find the input combination that will be utilized for the classification of the hips. The neural network input consists of five features: the overall necrotic volume percentage of the
femoral head (HeadNeer), the necrotic surface percentage of the lateral quadrant (LatSurf), the necrotic volume percentages of the anterior superior lateral (ASL), the anterior superior medial (ASM) and the posterior superior lateral (PSL) octants. The features we selected emphasize on the superior lateral segment of the femoral head. A number of studies have shown that femoral heads whose the superior lateral segment presents small necrotic lesions, retain a lateral supportive pillar of intact bone that acts as a stress shield for the affected segment and protects it from excessive loads. The absence of this pillar in hips with wide lesions extending in the lateral areas of the head makes those hips susceptible to collapse.

Several feed-forward back-propagation neural network architectures [13] were tested seeking the best classification performance. The network with the best results consists of three layers (Fig. 2). The input layer has 5 nodes, one for each of the 5 input features. Input data are normalized (mean value 0 and standard deviation 1) prior to the presentation to the neural network. The hidden layer consists of 7 nodes with the sigmoid hyperbolic tangent as the transfer function. The output layer has a single node with a logarithmic sigmoid transfer function, producing an output value between 0.0 and 1.0. The output values are thresholded giving a final output of 0 (good prognosis) or 1 (poor prognosis).

The leave-one-out cross-validation method [14,15] was used for the performance evaluation of the neural network. This method uses all N input samples both for training and for testing. N cross-validations are performed. Each time N-1 samples are used as the training set and the single remaining sample is used for simulation and testing. This procedure is repeated until all samples are used for testing. The evaluation of the network performance is carried out on the aggregate output. This cross-validation method is suitable for small data sets such as ours. The major advantage is that it uses all available data both for training and for testing. The main drawback is the time it requires, since 37 networks need to be trained and tested for a single evaluation procedure.
In each training session the input was the vector of the volumetric features and the target to be learned was the vector with the corresponding outcome results (0 for good outcome and 1 for poor outcome). Bayesian regularization [16] was used for training the neural network. It provided with the best classification results not only for the proposed network, but also for the majority of network architectures tested. The mean square error goal was 0.001 and the maximum number of training epochs was 5000. The mean number of epochs needed for the training of the 57 datasets during the leave-one-out evaluation was 1370.

3. Results

Table 2 displays the mean values and standard deviations for the volumetric features for the 37 hips of class 0 (good outcome) and the 20 hips of class 1 (poor outcome). The main observation is the high variability of the values of most features. Nevertheless, the mean values of all features of class 1 are larger than the corresponding mean values of class 0.

The outputs (predictions) of the neural network are displayed on the scatter graph of Fig. 3. The horizontal axis shows the outcome of the disease (correct prognosis). The vertical axis shows the outputs of the neural network simulation (leave-one-out cross-validation). These outputs are real numbers with values ranging from 0.0 to 1.0.

The performance of the neural network is evaluated in terms of the classification sensitivity (Se), specificity (Sp) and accuracy (Ac). Receiver operating characteristic (ROC) analysis is also performed and the area under the ROC plot (AUC) is calculated. AUC is an objective and acceptable measure of the network classification (prognostic) performance [17-20].

The network output is a continuous variable ranging from 0.0 to 1.0. A cutoff (threshold) value transforms the output into a dichotomous variable with value 0 or 1 corresponding to good and
poor prognosis, respectively. The true positives are the hips with poor outcome (head collapsed within 2 to 4 years after the operation) and positive test result (prediction value greater than the cutoff value). True negatives are the hips with good outcome (good clinical and radiological results for at least 5 years after the operation) and negative test result (prediction value less than the cutoff value). Accordingly, false positives are the hips with good outcome and positive test result, and false negatives are those with poor outcome and negative test result.

The ROC plot depicts the true positive rate (sensitivity) against the false positive rate (1-specificity), calculated for all possible threshold values (Fig. 4). The calculated area under the ROC curve is 0.916.

For the threshold value of 0.3 the sensitivity is 0.85 and the corresponding specificity is 0.89 and the classification accuracy is 0.88. The network failed to recognize 3 of the 20 hips with poor prognosis (3 false negatives) and 4 of the 37 hips with good prognosis (4 false positives).

Table 3 displays the obtained AUC, sensitivity, specificity and accuracy (using leave-one-out cross-validation) for the chosen and for some other training methods. The experiments were performed using the Matlab development environment on a personal computer with an Intel Pentium 4 1.7 GH processor with 512 MB memory.

4. Discussion and Conclusions

We have presented an automated intelligent system for the prognosis of femoral head necrosis. The system is composed of two basic modules: a volumetric feature extraction module and a classification module for the necrotic hips, according to their prognosis. The volumetric method uses MRI to extract information about the size and distribution of the lesion in the femoral head and the affected surface of the superior area of the head. Those features are the input of an
artificial neural network (being the classification module), which is trained to predict the fate of a
given hip (good or poor prognosis).

Artificial neural networks have been applied in solving a variety of medical diagnosis and
prognosis problems [21-27]. They have been used to predict the course of diseases, the
efficiency of treatment methods and to evaluate the risk of various pathologic conditions. Neural
networks can be trained to recognize complicated patterns in complex and noisy data sets. This
task is, in most cases, infeasible even for the most experienced physicians. The major strength of
neural networks that makes them valuable for the medical prognosis is their ability of
generalization. They can generate conclusions about patterns not previously encountered, e.g. the
neural network can predict the fate of a hip for which it has not been trained.

Magnetic resonance imaging is the most suitable technique for the diagnosis of osteonecrosis. It
can detect the disease in its early stages, when x-rays are still negative. MRI reveals anomalies at
the bone marrow, when the bone is still intact. Another major advantage of MRI is that it can be
used to extract important characteristics such as the necrotic tissue volume and precise
topographical distribution. As mentioned before, the significance of the extent of the necrotic
lesion for the prognosis of femoral head osteonecrosis is widely recognized. Many researchers
attempted to evaluate the size of the necrotic lesion [6,28,29] using MRI. Most methods perform
2-dimensional measurements (areas or areas) on single MR slices. Only few methods attempt to
estimate the volume of the necrotic lesion. These methods present a major drawback: they
perform manual segmentation (tracing) of the lesion and detection of the head border. This is
time-consuming and with low repeatability and accuracy.

A number of classification systems have been used for staging femoral head osteonecrosis [30-
32]. Most of them consider the extent of the lesion as a factor that affects the fate of the disease.
The major disadvantage of those systems is that they try to evaluate the extent of the lesion based
only on radiographs. Even system that utilize MRI, calculate only 2D measurements, which
provide only an inaccurate and limited description of the actual extent. Classification systems provide general knowledge about the course of osteonecrosis. They do not provide specific individualized information that can be used for the prognosis of a given hip. For example, a hip is classified to a specific stage for which it is known that there is 50% probability of collapse. This statistical measurement does not say whether the present hip will fail or not. The system we proposed is based on an objective, reliable and easy to use volumetric method for the description of the lesion. Additionally, the neural network, due to its generalization capability, can predict the fate of the specific hip, even thought its has not been trained using the data of the given hip.

The lesion size and distribution is not the only factor that determines the fate of femoral head osteonecrosis. A number of physical, clinical and laboratory factors may affect the outcome of the disease. Such factors are age, sex, aetiology, presence of the disease (or total hip replacement) on the opposite hip and various biochemical features. Future work should focus on the incorporation of such features to the intelligent system, in order to improve the prognostic performance.

Acknowledgements

The authors are grateful to Prof. A. Likas for various comments and suggestions.
References


Table 1

Features extracted by the volumetric method

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASL</td>
<td>Necrotic volume percentage of the Anterior Superior Lateral octant</td>
</tr>
<tr>
<td>ASM</td>
<td>Necrotic volume percentage of the Anterior Superior Medial octant</td>
</tr>
<tr>
<td>AIL</td>
<td>Necrotic volume percentage of the Anterior Inferior Lateral octant</td>
</tr>
<tr>
<td>AIM</td>
<td>Necrotic volume percentage of the Anterior Inferior Medial octant</td>
</tr>
<tr>
<td>PSL</td>
<td>Necrotic volume percentage of the Posterior Superior Lateral octant</td>
</tr>
<tr>
<td>PSM</td>
<td>Necrotic volume percentage of the Posterior Superior Medial octant</td>
</tr>
<tr>
<td>PIL</td>
<td>Necrotic volume percentage of the Posterior Inferior Lateral octant</td>
</tr>
<tr>
<td>PIM</td>
<td>Necrotic volume percentage of the Posterior Inferior Medial octant</td>
</tr>
<tr>
<td>HeadNecr</td>
<td>Overall necrotic percentage of the femoral head</td>
</tr>
<tr>
<td>LatSurf</td>
<td>Necrotic percentage of outer surface of the lateral quadrant</td>
</tr>
<tr>
<td>MedSurf</td>
<td>Necrotic percentage of outer surface of the medial quadrant</td>
</tr>
</tbody>
</table>
Table 2
Mean value (and standard deviation) of the volumetric features for the hips of class 0 and for the hips of class 1

<table>
<thead>
<tr>
<th></th>
<th>ASL</th>
<th>ASM</th>
<th>PSL</th>
<th>HeadNecr</th>
<th>LatSurf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 0 Mean</td>
<td>30.1</td>
<td>54.6</td>
<td>26.2</td>
<td>23.4</td>
<td>38.7</td>
</tr>
<tr>
<td>Std</td>
<td>17.2</td>
<td>25.8</td>
<td>21.7</td>
<td>11.1</td>
<td>17.5</td>
</tr>
<tr>
<td>Class 1 Mean</td>
<td>62.3</td>
<td>78.4</td>
<td>47.4</td>
<td>38.8</td>
<td>67.4</td>
</tr>
<tr>
<td>Std</td>
<td>15.5</td>
<td>15.5</td>
<td>20.2</td>
<td>8.7</td>
<td>12.4</td>
</tr>
</tbody>
</table>
Table 3

Network performance using various training methods: Area under the ROC curve (AUC), sensitivity (Se), Specificity (Sp) and Accuracy (Ac)

<table>
<thead>
<tr>
<th>Training method</th>
<th>AUC</th>
<th>Se</th>
<th>Sp</th>
<th>Ac</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian regularization</td>
<td>0.916</td>
<td>0.85</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>Resilient backpropagation</td>
<td>0.815</td>
<td>0.85</td>
<td>0.76</td>
<td>0.79</td>
</tr>
<tr>
<td>BFGS</td>
<td>0.791</td>
<td>0.80</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>One step secant</td>
<td>0.862</td>
<td>0.85</td>
<td>0.76</td>
<td>0.79</td>
</tr>
<tr>
<td>Levenberg-Marquardt</td>
<td>0.768</td>
<td>0.85</td>
<td>0.68</td>
<td>0.74</td>
</tr>
<tr>
<td>Scaled conjugate gradient</td>
<td>0.851</td>
<td>0.85</td>
<td>0.73</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Figure legends

Fig. 1. The volumetric method: (a) the MRI slice (b) drawing of the circle that circumscribes the femoral head, (c) the thresholded binary image (the calculation of the threshold value is performed using automatic sampling of the necrotic and the normal portion of the head), (d) application of morphological operations to remove “islands”, (e) the final trace of the necrotic area and (f) measurement of the affected area on each quadrant.

Fig. 2. The artificial neural network architecture.

Fig. 3. A scatter graph showing the output of the neural network against the class of the input samples (0 – good outcome, 1 – poor outcome).

Fig. 4. Receiver operating characteristic (ROC) plot.
Fig. 1. The volumetric method: (a) the MRI slice (b) drawing of the circle that circumscribes the femoral head, (c) the thresholded binary image (the calculation of the threshold value is performed using automatic sampling of the necrotic and the normal portion of the head), (d) application of morphological operations to remove "islands", (e) the final trace of the necrotic area and (f) measurement of the affected area on each quadrant.
Fig. 2. The artificial neural network architecture.
Fig. 3. A scatter graph showing the output of the neural network against the class of the input samples (0 – good outcome, 1 – poor outcome).
Fig. 4. Receiver operating characteristic (ROC) plot.