Abstract—Osteoarthritis is a degenerative joint disease, which causes the degradation of articular cartilage and subchondral bone. The disease may result in mechanical abnormalities of the joints, including weight bearing joints such as the knees and hips. In this work, we analyze gait biomechanical data using neural network models to predict the level of joint deterioration and the level of pain in participants suffering from knee osteoarthritis. The results of the analyses demonstrate strong correlation between gait kinetics and joint deterioration and level of pain in osteoarthritic individuals.

I. INTRODUCTION

OSTEOARTHITIS (OA) is a debilitating disease with a high prevalence in elderly individuals. OA affects nearly 27 million people in the United States, with OA of the knee accounting for an estimated 10% of visits to primary care physicians (Núñez, 2008). Better understanding of the nature of the disease and early detection and treatment are essential in improving and preserving the functionality of patients’ joints and in minimizing pain. In this study, gait biomechanical data from 30 participants were analyzed using pattern recognition neural networks in order to automatically classify the patients’ levels of joint deterioration and pain.

II. METHODS

A. Data Collection

The participants (60 to 85 years of age) with knee OA were screened by a rheumatologist (co-author) for knee OA. Participants were excluded if they had had arthroscopic surgery or an intra-articular injection within the past three months, neurological disorders, or had participated in a structured strength training or Tai Ji program in the past six months. A participant was included if he/she met the Classification Criteria for Knee OA of the American College of Rheumatology (Altman, 1986). After determining the severity of OA in each knee using the Kellgren/Lawrence Scale (Kellgren, 1957) on knee X-rays, participants underwent a biomechanical gait data collection session in the Biomechanics/Sports Medicine Lab at the University of Tennessee. The participants were instructed to walk at a speed of 1.1 m/s (± 5%), while simultaneous ground reaction forces (1200 Hz, Advanced Medical Technologies, Inc, Waterford, MA) and three-dimensional (3D) kinematic data (240 Hz, Vicon Motion Analysis System) were collected. The 3D kinematics and kinetics gait data were computed using a 3D biomechanics software suite (Visual 3D, 4.75, C-Motion, Inc.). The participants were divided into three groups. Group 1 are the control group, who did not take any form of training during the process. Group 2 took strength training whereas group 3 took Tai Ji program. Nevertheless, in this work, we investigated the gait dynamics before training took place. Thus, the form of training is irrelevant for the current study. This has been clearly stated in the revised manuscript. Prior to participating in the study, all participants provided an informed consent document approved by the Institutional Review Boards of the University of Tennessee and the UT Medical Center.

Fig. 1. Example of gait biomechanical data: ankle angle in degree (top left), knee angle in degree (top right), hip angle in degree (bottom left), and ground reaction force in body weight (bottom right).

A typical example of gait biomechanical data is shown in
Fig. 1. Conventional studies based on statistics of biomechanical properties such as mean, min, or max of a curve often fail to distinguish between patients from different classes, since these statistics are only samples of a few features of the gait data. Here, we aim to classify patients of knee OA based on analyses of the whole gait curves using artificial neural networks.

### B. Neural Networks

Neural networks are highly nonlinear functions designed to mimic the way the human brain evaluates stimuli. A neural network consists of nodes, arranged in layers. There are many different types of neural networks which vary based on their training algorithms and architecture. The network used in the current analyses was a pattern recognition neural network. A pattern recognition neural network is a highly nonlinear classifier used to automatically group data into categories. The network has as many outputs as the inputs have categories, with each output related to the tendency of the network to group the input into the individual categories. The largest output is selected as the final determined output. Training is required to correctly adjust the weights and biases of the network in order to generate the correct final output. An algorithm is used to adjust the weights and biases using training data in an iterative process. Once the network has been trained, it can be used to analyze new data.

The results from neural network prediction can be further improved by feeding the resulting predictions of several neural networks to a committee machine. Here, a common input is used to train a number of different neural networks (experts). Then, outputs from individual experts are combined to produce an overall output through ensemble averaging. An alternative approach is to replace the committee machine with a single neural network. However, this will lead to a large number of adjustable parameters (i.e., number of layers, number of nodes in layers, etc.), significantly increase the computational time required to train the network, and may lead to poor predictions to due exaggerations of minor fluctuations in data. For more detailed discussion of neural networks and committee machines, we refer interested readers to the book of Haykin (1994) and that of Hastie et al. (2001).

In this study, a committee machine was used to determine the mode of 20 experts’ predictions. The individual neural networks (experts) had three layers of 100, 50, and 25 neurons, respectively.

### III. RESULTS

Gait biomechanical data were collected on 30 individuals. Data for each individual were recorded for five trials of the individual’s full gait cycle. Seven 3D measurements resulting in a total of 21 kinetic and kinematic vector measurements were selected for each leg during each of the trials, including ankle angle, ankle angular velocity, knee angle, knee angular velocity, hip angle, hip angular velocity, and ground reaction force. Thirty participants participated in the experiment and each conducted five trials with both legs; thus we have a total of 300 records of an independent gait cycle. All 300 records were treated as independent records; however, care was taken to ensure that the five records from the same leg were all included in either the train set or the test set in order to avoid self-correlation of the records.

In order to predict Kellgren/Lawrence Scale scores (K/L scores), which ranged from 1 to 5, 200 records from 20 of the 30 participants were randomly selected as the train set. The remaining 100 records were used as the test set. This random selection of train and test records was repeated for several trials and the results presented are typical of those for all trials.

We first attempted to classify the K/L scores using all 21 vector measurements; see prediction results in Fig. 2. The average accuracy of 20 neural network predictions was 50%. The committee machine achieved a moderate level of accuracy, correctly predicting the K/L score of 67 of the 100 records (67%).

![Fig. 2. Experts’ and Committee Machine’s Results of K/L Score Predictions.](image)

Further analyses were then conducted to determine which of the 21 vector measurements contributed most to the accuracy of the predictions of the neural network. These analyses were performed by using different combinations of directional measurements’ data as input to the network. The results of these analyses demonstrated that using angles and angular velocities in the sagittal plane and ground reaction force data (in all three dimensions) produces the most accurate predictions; thus, these measures are more closely related to the K/L scores than the remaining 12 measurements. Using only these data as input, the committee machine achieved a greater level of accuracy than using data from all of the 21 vector measurements as input, with a minimal improvement of 2%.

Fig. 3 presents the confusion matrices for the committee machine results obtained using all vector measurements’ data (top) and using sagittal plane and ground reaction force data (bottom). The confusion matrix consists of two regions: the inner matrix and the outer region, which includes the last column and bottom row of the matrix. The diagonal squares of the inner matrix contain the number of correct matches (top) and their corresponding percentage (bottom) relative to the number of inputs. The off diagonal squares of the inner matrix contain mismatches in the same manner. The last column of the confusion matrix contains the true positive rate (top) and true negative rate (bottom) for individual categories of results. True positive rate is the probability that
an input classified by the neural network as belonging to a
certain category actually belongs to that category. True
negative rate is the probability that the neural network’s
predicted class for an input is incorrect. The bottom row of
the matrix presents the accuracy of the network’s predictions
for the individual categories, where the top percentage is the
percentage of a given category of input correctly classified
by the network and the bottom percentage is the percentage
misclassified. The bottom right corner square contains the
network’s overall accuracy for all of the categories, with top
percentage being the percentage correctly classified and the
bottom percentage being the percentage misclassified.

Fig. 3. Confusion Matrices of K/L Score Predictions: results using all (top)
and sagittal plane and ground reaction force (bottom) vector measurements.
K/L scores 1-5 correspond to K/L grades 0-4. The bottom right corner
squares contain the overall accuracy of the results.

The accuracy in predicting K/L scores 1 and 5 were very
low (0%) because K/L scores 1 and 5 were underrepresented
in the data and therefore in both the train and test sets of
records (two legs in category 1 and five legs in category 5),
causing an inherent bias in the network. Therefore, analyses
using fewer categories were also performed in attempt to
further improve the accuracy of the predictions. Records
with K/L scores of 1, 2, or 3 were assigned to the “low”
class (Class 1) and those with K/L scores of 4 or 5 were
assigned to the “high” class (Class 2). This grouping was
also chosen in order to balance the number of records in each
class and remove biases from the network. 150 records were
arbitrarily chosen to train the network and the remaining 150
records were used to test the network. Again, while records
were chosen randomly to be included in the train and test
sets, care was taken to ensure all records from the same leg
were in either the train or test set in order to avoid self-
correlation.

Fig. 4. Confusion Matrices for K/L High-Low Class Prediction: Class 1=
“low” class; Class 2= “high” class. The predictions attained using data from
all 21 vector measurements is presented on top; predictions attained using
only sagittal plane and ground reaction force data are presented on bottom.

As before, initially data for all 21 vector measurements
were used as input for the network before examining the
relevance of the individual directions. The predictions using
all available data achieved an accuracy of 62% from the
committee machine, with 97 out of 150 records correctly
being assigned to the “high” or “low” classes. Several
combinations of measurements and dimensions were also
tested, and again the sagittal plane variables and ground
reaction force proved to be the best input for the network.
Using these data as input, the neural network again achieved
a greater level of accuracy than that achieved using all
available data, correctly predicting “high” or “low” assignment of 114 of the 150 test records (76%), an
improvement of 14%. The confusion matrices of the
committee machine results are presented in Fig. 4.

We also explored the classification of pain scores.
Because the pain scores are numerical values ranging from 0
to 403.5, the scores were arbitrarily assigned to “high” and
“low” classes, with scores greater than 100 assigned to the
“low” class (Class 1) and scores of 100 or less being
assigned to the “high” class (Class 2). This grouping was
also chosen in order to create classes with equal numbers of
records and prevent biases in the network. Again, 150
records from 15 participants were used to train the network and 150 records were used as the test set with care taken to ensure all records from the same leg were in either the train or test set.

Initially, all data were used as input to the network. The committee machine’s predictions attained were relatively high, with 112 of the 150 correctly assigned to their “high” and “low” classes (74.7%). Further analyses of different measurements and dimensions as inputs revealed that sagittal plane and ground reaction force were the most relevant data for attaining higher levels of accuracy in predicting pain score classes. Using these data, the committee machine correctly assigned 113 of the 150 test records to their corresponding “high” and “low” classes (75.3%). Thus, the accuracy of the predictions remained essentially the same when only the sagittal plane and ground reaction force data were used; see Fig. 5.

IV. CONCLUSION

Pattern recognition neural networks have been used in other studies to differentiate between normal and pathological gait due to aging and disease. One of the first such studies was conducted by Holzreiter and Kohle (Holzreiter, 1993) who attempted to categorize gait pathology based on ground reaction forces. Since then, more complex neural networks have been used to differentiate between different gait pathologies, including foot pressure, arthrosis, and Parkinson’s and post-stroke gait (Chau, 2001).

The results of the current K/L scores and pain scores analyses reveal that gait data can be used to classify structure (K/L scores) and function (pain) in osteoarthritic patients. Of the measurements studied, sagittal plane and ground reaction force measurements provided the most accurate classification for X-ray and pain classes, with committee results attaining accuracies of 76% and 75.3%. Random selection of train and test sets was repeated for several trials and the results presented are typical of all trials. Future work will include more rigorous cross validation studies. The accuracy of the results suggest that OA affects patients’ gait to a degree in which it may be possible to analyze a patient’s gait in order to determine the condition of the patient’s OA. Further research and analyses of gait dynamics may lead to a biomarker capable of characterizing the disease.

REFERENCES