Research Article

iMedPub Journals
http://www.imedpub.com

American Journal of Computer Science and Information Technology

2021
Vol. 9 No. 5: 88

A Novel Method of Screening for Rheumatoid Arthritis Using Machine Learning

Priyanka Jakka* and Abhitha Vegi

Department of Computer Science, Mission San Jose District of Fremont, California, United States

Corresponding author:
Priyanka Jakka, Department of Computer Science, Mission San Jose District of Fremont, California, United States

priyankajakka@gmail.com

Citation: Jakka P, Vegi A (2021) A Novel Method of Screening for Rheumatoid Arthritis Using Machine Learning. Am J Compt Sci InformTechnol Vol.9 No.5: 88.

Abstract

Arthritis is a common ailment that affects gait and pressure distribution across patients' feet. While current treatments include physical therapy, medication, and surgery, these solutions are only possible after the detection of arthritis. A machine learning model can be utilized to predict arthritis in its early stages and also to monitor patient improvement during rehabilitation. A smart shoe involves a combination of sensors - force sensitive resistors, accelerometer, and magnetometer to capture data such as orientation, displacement, and pressure distribution using pressure sensors. All of the sensors are built into the soles of the shoe. The raw data is saved on to an SD card and sent to a laptop via Bluetooth for analysis. The data is sent through a duration program, and a machine learning algorithm then compares this data to the data of users (both arthritic and nonarthritic) to classify the walking pattern as arthritic or not.

Some common markers of arthritis include larger values for acceleration in the z-axis (tilt of the foot inwards) and also an increase in pressure on the first metatarsophalangeal joint. Using the algorithm's results, a physical therapist would be able to view a 3D rendering of the walking motion to create a more complete diagnosis of the arthritis.

Keywords: Arthritis; Machine learning; Biological factors

Received: April 12, 2021; **Accepted:** April 26, 2021; **Published:** May 02, 2021

Introduction

Background

Arthritis is a broad term that encompasses all joint diseases and affects more than 350 million patients worldwide. There are many types of arthritis, of which osteoarthritis (OA) and rheumatoid arthritis (RA) are the most prevalent. OA is a disease in which the cartilage in the joint slowly wears away, causing the joint to become rough. As the space between bones decreases, bone begins to rub on bone. OA can result from injury and old age, and doctors commonly use joint aspiration, x-rays, and MRIs for diagnosis. This paper primarily focuses on RA, a disease that causes the body's immune system to attack the joint's synovium. Unfortunately, RA is difficult to diagnose, because its symptoms can be mild and some tests (such as a blood test) may point to false positives.

Gait

Gait is a person's way of walking and is important for daily life. Every person has a unique gait that can even be used as a method of identification. A problem with one's gait can lead to a number of conditions ranging from a mild sprain to arthritis. In this paper a user's gait is weighed as an important factor in determining whether a person has arthritis.

In the time period between when a person's foot starts touching the ground to when it reaches the ground again, they take one step. This period of time is known as a gait cycle. A gait cycle has seven parts: heel strike, foot flat, mid stance, heel off, toe off, mid swing, and late swing. First, the heel of the first foot touches the ground in the heel strike stage. Second, the sole of the first foot lies flat on the ground during the foot flat stage. In the mid stance stage, the leg of the first foot is instantaneously at rest and perpendicular to the ground as the second leg moves forward. Next, in the heel off stage, the heel starts to leave the ground. The first foot's anterior toe lifts off from the ground as the sole of the second foot is touching the ground in the toe off stage. Then, in the mid swing stage, the first foot is off the ground and the second foot is completely flat. Finally, the late swing stage is the one right before the start of the next gait cycle. In addition, the gait cycle can be described as having a stance phase and a swing phase. The stance phase comprises the first five steps (heel strike, foot flat, mid stance, heel off, and toe off) since both feet are touching the ground. On the other hand, only one foot touches the ground at a time during the swing phase, which is the last two steps (mid swing and late swing) **(Figure 1)**.

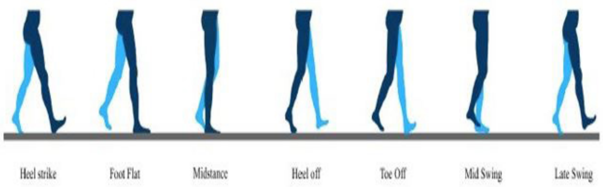

Figure 1: The gait cycle.

Anatomy of the foot

The joints that are most commonly affected by RA include the interphalangeal joints, which help to make up the toes, metatarsophalangeal joints, which connect the phalanges to the metatarsals, and the subtalar joint, which connects the talus and calcaneus bones. As arthritis affects these joints, movement along them becomes more restricted because of the increasing pain [1]. The patient often changes their gait to alleviate the pain caused from the inflammation while walking. Such changes can include intoeing, where the feet are oriented to point inwards, and out-toeing, where the feet point outwards. The shape of the foot can also experience changes due to RA. This can include the collapse of the medial longitudinal arch, lateral longitudinal arch, or anterior transverse arch of the foot, which can result in a flat foot. Patients may additionally shift their weight onto different parts of their feet such as their toes due to the excessive pain in the tibial, saphenous, and sural regions of the foot. This can result in bunions, calluses, or claw toes [2]. While early stages of recovery can include mild foot exercises and specialized shoes, later stages can require major treatment, such as fusing bones **(Figure 2)**.

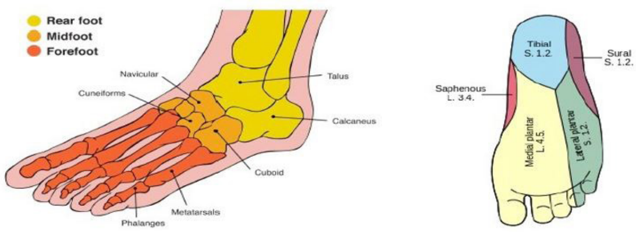

Figure 2: Anatomy of the foot.

Engineering goal: Although there is currently no permanent cure for RA, "studies show that people who receive early treatment feel better sooner and more often [3]. Our goal was to create an overall solution that would aid in an earlier diagnosis of RA. Our solution consists of three components: a smart shoe, a machine learning model, and a visual representation. The smart shoe is designed to collect data while the user is walking. It captures readings about the amount of pressure users are placing on different parts of their feet and the orientation of their feet while walking. The machine learning model attempts to solve the binary classification problem of predicting RA. The visual representation is a 3D model that replays how the patient walked for additional review by a doctor or physical therapist. This will aid medical experts in giving patients a more reliable diagnosis.

Existing research

Yoo et al. they collected data from 60 anonymous Korean patients that showed various degrees of RA. The factors they considered were Anti-CCP, Rheumatoid factor, SJC, and ESR. Anti-CCP, or anti-cyclic citrullinated peptides, are commonly found in patients that have RA. Rheumatoid factor, also known as RF, is a group of antibodies that attack someone's own tissue instead of invaders like viruses. SJC is swollen joint count and ESR is the erythrocyte sedimentation rate, which measures the degree of inflammation in the joints. They utilized the k-means algorithm for k = 4 to achieve an accuracy of 84.1%. Researchers developed a model to assess the destruction of finger joints in RA [4]. They used a cascade classifier using Haar-like features to detect the finger joints. They then trained a CNN (convolutional neural network) to provide a joint space narrowing (JSN) and erosion score to each joint. Their final overall accuracy was 95.3%. Butte et al published a paper in March 2019 forecasting clinical outcomes in patients with RA. The researchers obtained data from 820 patients from 2 hospitals that had various degrees of RA. They created a TDD GRU (time distributed gated recurrent unit) and achieved an AUROC (area under the receiver operating characteristic curve) of 0.91 at the university hospital and an AUROC of 0.74 at the public safety net hospital. All of the studies mentioned above attempted to predict arthritis using computer vision techniques or biological factors such as Anti-CCP. However, our research is unique in that it predicts RA using a smart shoe to collect data. This approach requires significantly less equipment and resources while still giving an accurate prediction.

Methodology

Proposed methodology

Adafruit adalogger: The Adafruit adalogger is placed in the middle of the shoe. The Adalogger stores the orientation, displacement, and pressure data in the SD card before it is transmitted to an external device via Bluetooth. By using an SD card instead of directly transmitting values at time of capture, all data is stored in a central location. This prevents the loss of data in case the shoe is not within range of the external device (Ex: laptop). All of the sensors are connected to the Adafruit adalogger through jumper cables and wires.

Adafruit Bluefruit: The Adafruit Bluefruit is also located in the middle of the shoe and is placed directly above the Adafruit adalogger. This microcontroller is used to capture the motion data from the force sensitive resistors, accelerometer, and magnetometer and transmits it via a built-in Bluetooth interface to an external device [5].

Force sensitive resistors: Six force sensitive resistors (FSRs) are placed in the shoe as shown in **Figures 3 and 4** below. Three are placed at the metatarsophalangeal joint (right under the toes), one is placed on the large toe, one is in the midfoot region, and one is on the tibial region (the heel). These locations are optimal since those areas are where most of the pressure is placed on the foot. Two types of FSRs are used; one is a larger sensor which is placed on the tibial and midfoot regions to account for

American Journal of Computer Science and Information Technology
2021
Vol. 9 No. 5: 88

flatfoot). This sensor is used to detect any changes in pressure over a larger area. The other FSRs are smaller sensors that detect precise pressure points and are placed in the forefoot region. The sensors receive pressure readings which range from 0 to 1023, thus requiring analog pins. The microcontroller has a limited number of analog pins, out of which several are used for the MPU-9250 sensor. Therefore, we are limited to a maximum of six FSRs per shoe.

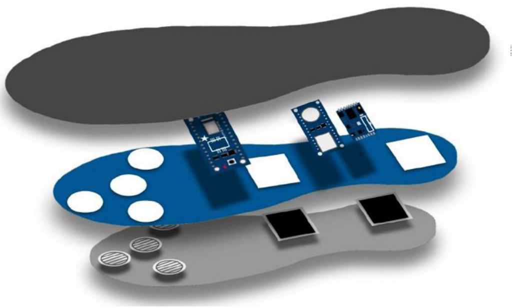

Figure 3: A layout of the sensors in the sole of the shoe.

Figure 4: Pictures of the actual soles used in the smart shoe.

Accelerometer and magnetometer (MPU 9250): The accelerometer is in the center of the sole of the smart shoe. The values for acceleration in the x, y, and z axes are stored in separate variables which are updated at 15 milliseconds intervals for accurate data [6]. The magnetometer is also part of the MPU-9250 device. This sensor is used to calculate the average stride length of the user. One shoe has a magnet embedded in the side of the shoe facing inwards. The magnetometer in the other shoe is used to sense the strength of the magnetic field as the magnet moves closer to and farther from the magnetometer during the user's stride **(Figures 3-5)**.

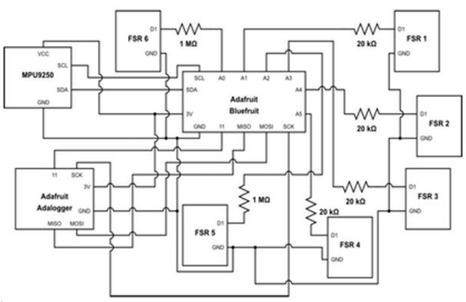

Figure 5: A schematic of the smart shoe (FSR represents a force sensitive resistor).

Procedures

Data collection: Every person has a different stride length and speed while walking. To make all the data collected as uniform as possible, subjects were asked to walk 10 steps in one direction. However, this would still result in users taking different amounts of time to walk. To accommodate this, we resized the data based on the shortest amount of time it took a person to walk 10 steps, which is about 8 seconds.

Data to SD card: The Adafruit Bluefruit (the microcontroller embedded in the smart shoe) receives data from the FSRs, accelerometer, and magnetometer at regular 15 millisecond intervals. A total of 12 values from all the sensors are received during each interval. These values are concatenated into a single line output and appended to a file on an SD card (attached to the Adafruit Adalogger). By the end of the data collection program, 2 files are created on each of the SD cards on the left and right shoes [7].

SD Card to external device via Bluetooth: A Bluetooth program was run to pull values from the two separate files. The external device discovers the microcontrollers (which are also Bluetooth devices) on each of the shoes. Once the laptop has established a connection with both of the devices, it pulls values from both of the files. A string parsing method in the program removes the extra characters that are recorded in front of the desired values as a result of the Bluetooth transmission. These parsed values are subsequently appended and saved to two separate files on the laptop, one file per shoe. When all the values have been received, the Bluetooth connection is terminated. The two files are then used for the 3D visualization of the user's movements [8].

Data analysis

3D rendering: The 3D model was created as an additional reference tool for analysis of the user's gait by a medical expert. The program receives data from the two separate files created for both shoes at the time of Bluetooth transmission. Each string of values is parsed and used in different parts of the 3D rendering. Two soles are drawn using many geometric shapes to render the visualization. The soles additionally have gradient circles placed at the locations of the six FSRs [9]. Based on the acceleration readings; the soles change their orientation to mimic the orientation of the user's feet in motion. The soles additionally show vertical displacement based on the values received by the magnetometer. The gradient circles on the soles change color from red to yellow to show the fluctuations in pressure as the user steps onto and off the ground. Each string of readings is retrieved at the same rate at which they are recorded to recreate the same walking pattern of the user. The soles are redrawn in their new positions and the colors of the gradients are updated after each reading is fetched from the file. By providing visualization for orientation, stride, and pressure distribution data, medical experts can easily identify the stance that the user is in at any particular moment [10].

Figure 6 is a screenshot of the 3D model. The orange and red gradient circles on the left foot show pressure placed on the sensors. This indicates that the left foot is on the ground. The other foot has light yellow circles, meaning very little pressure, which indicates that the foot is in the air. The user appears to be in a foot flat phase of the gait cycle. One of their feet is planted on the ground while the other is raised in the air and is in front of the grounded foot. The user doesn't have flat feet, since the pressure sensor in the midfoot region of the ground food shows no pressure readings (the gradient is light yellow indicating no pressure) **(Figure 6)**.

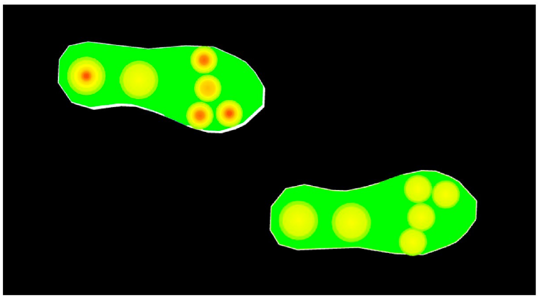

Figure 6: The gradient is light yellow indicating no pressure.

User input: Each user fills out a Google form describing their height, weight, age, biological sex, and history with arthritis. This information is later taken as input in the machine learning model.

Data concatenation: The information from the Google form is combined with raw data received during Bluetooth transmission for both shoes to create a single consolidated file that will be used by the machine learning algorithm [11]. Responses on the Google form are converted to 0's and 1's. Responses for age, height, and weight are further subcategorized into specific ranges and assigned boolean values. At the end of the concatenation process, the completed file is truncated to 501 lines. The first line has the names of the attributes, and the following 500 lines have all the values collected by both shoes.

Data filtration: The data from the smart shoe is run through a filtration program. To make the data more suitable for the machine learning algorithm, each column has its data normalized so that the values lie between 0 and 1. **Figure 7** is a screenshot of a concatenated csv file for the machine learning algorithm. The attributes with 'l' in front of them indicate values from the left shoe. The attributes with 'r' in front indicate values from the right shoe. The variables age, weight, and height are divided into 3 categories each as shown above by the attributes 'age1', 'age2', and 'age3' and so on [12]. The data corresponds to a user without arthritis as is shown by the 1 in the 'art' (arthritis) output variable. Since 1 is usually associated with the rarer class, the value of 'art' is 1 if the subject has arthritis and 0 if they do not **(Figure 7)**.

Machine learning model: Our machine learning approach provides a prediction for whether the user shows patterns in their gait and pressure distribution that would indicate arthritis. We use a binary classification model to determine whether the test data can successfully be attributed to its correct class arthritis or no arthritis. After testing several models, including decision trees, logistic regression, k-nearest neighbors, and Gaussian Naive Bayes, we decided on convolutional neural networks, otherwise commonly known as CNN.

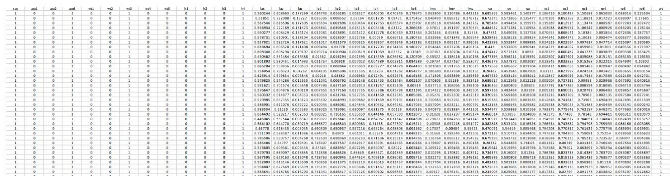

Figure 7: Screenshot of concatenated csv file used for ML algorithm.

CNNs can have three different layers: convolutional, pooling, and fully connected. The convolutional layer has matrices that slide across an image and multiplies with each section of the data to produce an output. In our model trained to detect arthritis, there are 32 matrices of dimensions 5 by 5. The pooling layer decreases the overall size of the input data. It performs aggregation operations to reduce the output to one number. Two common types of aggregation operations are max pooling and average pooling, which find the maximum and average of all the values, respectively. Similar to the convolutional layer, the input matrix is traversed and blocks of values are aggregated into single values. In our model, we utilize max pooling since it additionally, acts as a noise suppressor. The fully connected layer takes the results from the previous layers and classifies the data as arthritis or not. After randomly assigning weights to the features, it goes through a back propagation process to deduce the most accurate weights.

Overall, each CNN has a feature extractor (or featurizer) and a classifier. The featurizer decreases the size of an input to a smaller representation of the same matrix which is then used in the classifier. The featurizer is simply a pattern of convolutional and pooling layers. In our model, we used one convolutional layer, one pooling layer, and one fully connected layer before the final output [13]. For our convolutional layer and fully connected layer, we used the ReLU (Rectified Linear Unit) activation function. Since we are solving a binary classification problem, we used a sigmoid function for the output layer **(Figure 8)**.

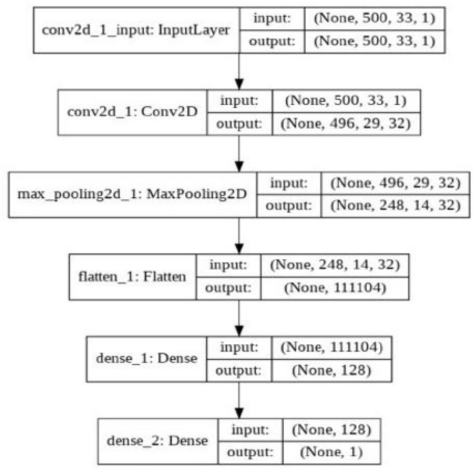

Figure 8: A diagram of the convolutional neural network discussed in this paper.

A total of 60 samples of data were collected, with 30 subjects falling in the category of arthritic patients and 30 having no history of arthritis. A 70-30 split was implemented, with 70% of the data being used for training and 30% for testing. The normalized confusion matrix in **Figure 9** shows the results for the test data. The final accuracy of our model was 94.44% **(Figure 9)**.

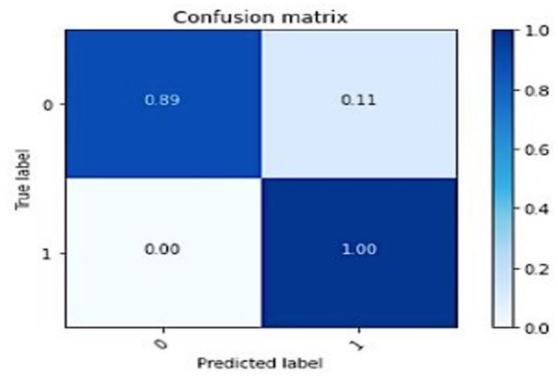

Figure 9: The normalized confusion matrix for the test data.

In this paper, we proposed a method for predicting whether or not a person has RA using CNN with sensor input from a smart shoe. 60 subjects (30 with RA and 30 without) were asked to take 10 steps wearing these smart shoes (which contain accelerometers, magnetometers, and force sensitive resistors) to collect data. The data was then sent to an SD card and consequently transmitted to an external device via Bluetooth. Subjects were asked about their biological sex, age, weight, and height; this data was concatenated to the filtered and curated data collected from the smart shoe. The data could then be viewed through our 3D model that visualizes the walking motion and pressure distribution to help professionals provide a better diagnosis. The data was also sent to the CNN for the binary classification task of predicting whether or not a person has rheumatoid arthritis [14]. The final accuracy achieved was 94.44% **(Figure 10)**.

$$\text{precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} = \frac{1.00}{1.00 + 0.00} = 1.00$$

$$\text{recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} = \frac{1.00}{1.00 + 0.11} = 0.89$$

$$\text{f1 score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = 2 \cdot \frac{1.00 \cdot 0.89}{1.00 + 0.89} = 0.94$$

Figure 10: Additional evaluation metrics for the convolutional neural network.

There is little scope for false positives and false negatives. A Type 1 error, or a false positive, can result in unnecessary medical expenditure and emotional trauma. Type 2 error, alternatively known as a false negative can also have numerous consequences including worsening of symptoms, increased pain, permanent deformities, and the loss of a limb. While it is not possible to completely eliminate these errors, our aim is that the smart shoe, 3D model, and CNN can be used to aid medical experts to provide a more detailed and accurate diagnosis **(Figure 11)**.

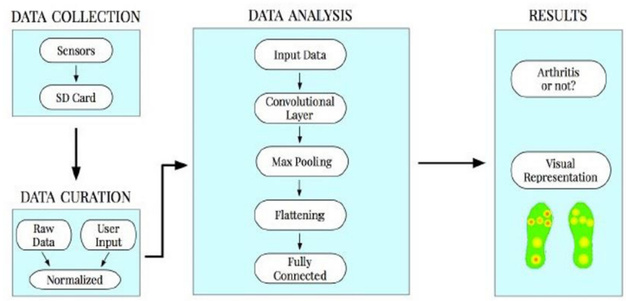

Figure 11: The overall work flow of our solution.

Results

The machine learning approach can additionally be changed from a binary classification problem to a multi-class classification task. This device could be used to classify the degree of arthritis in the user based on the severity of limping, for example. In addition, it was only possible to collect data from 60 unique subjects because of restrictions during the COVID-19 pandemic. We aim to collect data from at least 1000 individuals to improve our model. We will also include more layers to prevent over fitting in the CNN. This device can additionally be applied to other fields such as fashion and sports. In fashion, models can use the device to perfect their catwalk for the runway. In sports, athletes (Ex: runners) could use the smart shoe to improve their gait and achieve faster times.

Conclusion

This device can currently act as a supplementary resource for doctors to determine whether their patients have arthritis. While the 3D rendering can show the more subtle aspects of the user's motion that may escape the naked eye and help medical experts in the diagnosis of arthritis, it can additionally be used to monitor the progress of already known arthritic patients during physical therapy or doctor visits.

References

1 Mobasheri A, Kapoor M, Ali SA, Lang A, Madry H (2021) The future of deep phenotyping in osteoarthritis: How can high throughput omics technologies advance our understanding of the cellular and molecular taxonomy of the disease?. Osteoarthr cartil. 100144.

2 Arthritis Foundation (2017) Beyond joints: How RA affects the body.

3 Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, et al. (2019) Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. JAMA network. 2: e190606.

4 Ali DG (2019) Rheumatoid arthritis. American college of rheumatology.

5 Hirano T, Nishide M, Nonaka N, Seita J, Ebina K, et al. (2019) Development and validation of a deep-learning model for scoring of radiographic finger joint destruction in rheumatoid arthritis. Rheumato adv practi. 3: 47-49.

6 Jacsin Y, Kim AD (2020) Arches of the foot. Physiopedia. 2: 90-103.

7 Laliberte C, Marissa AL (2019) What Is Anti-CCP? what it can tell you about your autoimmune disease. CreakyJoints.

8 Laliberte C, Marissa AL (2019) What is rheumatoid factor? What this blood test can reveal about your autoimmune disease. CreakyJoints.

9 National rheumatoid arthritis society, laboratory tests used in the diagnosis and monitoring of rheumatoid arthritis (2019) NRAS.

10 Saha S (2018) A comprehensive guide to convolutional neural networks-the ELI5 way. Towards Data Sci.

11 Team, the healthline editorial (2015) What does rheumatoid arthritis look like healthline, healthline media?

12 Potluri S, Ahmed S, Diedrich C (2018) Convolutional neural networks for multi-class intrusion detection system. InInternational conference on mining intelligence and knowledge exploration.

13 Weatherford, Brian M (2019) Arthritis of the foot and ankle AAOS. Ort Am Aca Orthop Sur.

14 Yoo J, Lim MK, Ihm C, Choi ES, Kang MS (2017) A study on prediction of rheumatoid arthritis using machine learning. Int J Appl Eng Res. 7:9858-9862.