International Journal of Emerging Trends in Health Sciences

Volume 8, Issue 2, (2024) 26-35

www.ij-hs.eu

Machine learning techniques to predict and manage knee injury in sports medicine

Ravi Kumar [a*] , Sreenidhi Institute of Science and Technology, Hyderabad, India.
C. Sunil Kumar [b]**,** Dept. of CSE, The Apollo University, Chittoor, AP, India.
Subhani Shaik [c]**,** Dept. of IT, Sreenidhi Institute of Science and Technology, Hyderabad, India.
K. R. Praneeth [d]**,** School of Management, The Apollo University, Chittoor, AP, India.

Suggested Citation:
Kumar, R., Kumar, C.S., Shaik, S. & Praneeth, K.R. (2024). Machine learning techniques to predict and manage knee injury in sports medicine. *International Journal of Emerging Trends in Health Sciences, 8*(2), 26-35. https://doi.org/10.18844/ijeths.v8i2.9611

Received August 7, 2024; revised September 8, 2024; accepted December 3, 2024.
Selection and peer review under the responsibility of Prof. Dr. Nilgun Sarp, Uskudar University, İstanbul.

Abstract

The aim of this study is to conduct a complete review of the current state of Machine Learning (ML) in injury prediction and prevention. In recent years, there has been a growing importance in the application of ML techniques to find out and reduce risks associated with injuries, particularly in high-risk industries such as sports, healthcare, and manufacturing. The essential part of our body is the knee, sports persons commonly injuries during play games. Sports injuries result in stress & strain connected with athletic events. Sports wounds can affect soft tissue (ligaments, muscles, cartilage, and tendons). Injuries are common in sports and can have significant physical, psychological, and financial consequences. The aim of our study was therefore to perform a systematic review of Machine learning (ML) techniques that could be used to improve injury prediction and prevention in sports. ML algorithms play a crucial role in extracting accurate information from given images and they also handle the complex pattern of MRI knee-related clarifications. In this paper, discuss a real-life imagery rule, ML design used to recognize meniscus tears, bone marrow edema, and general abnormalities on knee MRI tests accessible. The final evaluation demonstrated the highest accuracy achieved by the support vector machine, closely followed by the KNN model and the RF Tree method, all yielding comparable performance levels.

Keywords: MRI images, SVM, Random Forest, KNN, Automated analysis, ACL Injury, Machine Learning

* ADDRESS FOR CORRESPONDENCE: Ravi Kumar, Sreenidhi Institute of Science and Technology, Hyderabad, India.
E-mail address: ravikumarsnist@gmail.com

1. INTRODUCTION

The knee is a body part of the body that has an important role, from walking and running to jumping and climbing. Unfortunately, the knee is also prone to sports injuries, which can have a significant impact on an athlete's performance and overall well-being. Here's why the knee is so important and why sports injuries to the knee are a major concern.

Improving medical care in modern society offers a broad range of digital health. Different types of technologies for sports knee issues (Buvik, Bugge, Knutsen, Småbrekke & Wilsgaard, 2019; Higgins, 2016). Digitalization among sports activity, orthopedics, and workout knowledge existed previously (Rigamonti, Albrecht, Lutter, Tempel, Wolfarth & Back, 2020; Bini, Schilling, Patel, Kalore, Ast, Maratt & Steele, 2020). Machine learning algorithms provide more accuracy on health-related issues (Weng, Reps, Kai, Garibaldi & Qureshi, 2017). Robotic technology is also used for these issues but all are in the trial stage (O'Sullivan, Nevejans, Allen, Blyth, Leonard, Pagallo,... & Ashrafian, 2019). Radiological images were used in previous studies and they reached in peak for human body parts (Haenssle, Fink, Schneiderbauer, Toberer, Buhl, Blum, & Zalaudek, 2018; Olczak, Fahlberg, Maki, Razavian, Jilert, Stark & Gordon, 2017). Patients regularly check their symptoms on Google and find the reasons for their unhealthy self-determination (Armstrong-Heimsoth, Johnson, McCulley, Basinger, Maki & Davison, 2017; Bickmore, Trinh, Olafsson, O'Leary, Asadi, Rickles & Cruz, 2018). Due to this their emotional pain on their health. Injury risk prediction is an emerging field in which more research is needed to recognize the best practices for accurate injury risk assessment. Important issues related to predictive ML need to be considered, for example, to avoid over-interpreting the observed prediction performance. Few artificial applications are also used in the field of sports injuries (Sandal, Stochkendahl, Svendsen, Wood, Øverås, Nordstoga & Søgaard, 2019). To carefully investigate the predictive potential of multiple predictive machine learning methods on a large set of risk factor data for anterior Cruciate ligament (ACL) injury; the proposed approach takes into account the effect of chance and random variations in prediction performance.

1.1. Background analysis: knee joint injury

Background analysis for knee joint injury presents the case study of the knee joint (ACL rupture at Knee Joint) discussed (Subhani Shaik, 2017; Jauhiainen, Kauppi, Krosshaug, Bahr, Bartsch, & Äyrämö, 2022). The following points about collagen and the function of the cartilage.

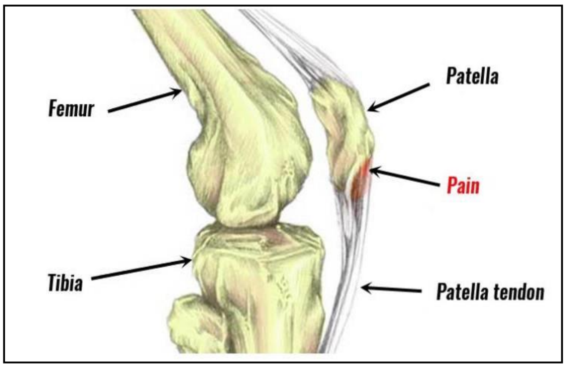
- Collagen is a type of protein fiber set up plentifully throughout our body. It supplies power and cushion too many dissimilar fields of the human body, together with the skin. collagen originates in our different types of connective tissues such as bones, cartilage, tendons, and ligaments Collagen is a structural protein made of amino acids enclosed together to form triple- Helices of elongated fibrils it is mostly found in fibrous tissue such as tendons, ligaments and skin.
- Cartilage characteristic resilience (= The Capacity to get well rapidly from difficulties, toughness, = flexibility, elasticity). However, the outcome from its elevated Proteoglycan satisfies. The application of pressure squeezes water away from these charge regions until charge-chare repulsion prevents further compression, when pressure is out, the water returns. Indeed, cartilage in joints lacks blood vessels; it is nourished by this flow of liquid brought about by the motion of the human body. Thus, if there is no forward, there is no sustenance of the cartilage layer and it becomes thin and fragile

1.1.1. Jumper's knee

Jumper's knee also known as patellar tendonitis is an injury of the patellar tendon, the cord-like tissue that joins the patella to the tibia. This problem, due to excess injury, recurrently happens to the one who plays sports that engage a lot of recurring jumping in volleyball, spiking, blocking, and spike service can place a lot of damage on their knees, and it also in football games. The following figure 1 shows the jumper knee.

Figure 1.

Jumper's knee injury-inflamed patellar tendon



A. Bio-Chemical Changes of Knee Injury

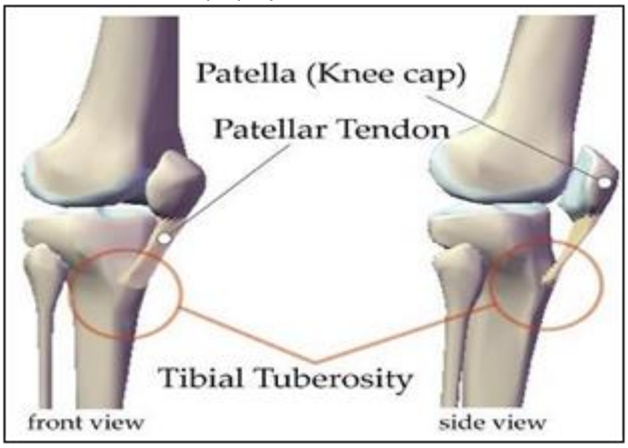
Develop of OA (Osteoarthritis) due to traumatic joint loading for the hip and knee are two major points. Straight sensitive traumas with an Osteochondral force encourage enhanced cartilage damage and eventually guide to OA. Due to matrix deprivation, chondrocyte death, and abolishment of anabolic functions happen after strength to articular cartilage. Chondrocyte death leads to biochemical changes in the extracellular matrix and declining in its biomechanical properties. In vitro, studies tell that human cartilage explants have also exposed considerably enlarged glycosaminoglycan (GAG) break and leave go off from wounded against unload unbroken cartilages all through zero to three days post-injury.

B. Biomechanical changes at the injured

Not only factors, but inflammatory cytokines from synovitis also affect the bio-mechanical unsteadiness of the joint does not come into sight. Volleyball and soccer players suffer from Meniscus injury and osteoarthritis. Acute wounds to the menisci are in general sports-related injuries. However, these injuries in volleyball/soccer are not clear as evidence of ACL injuries. But in 75% of cases meniscus tear rather than ACL injuries. Few researchers say that soccer players affected by menisci than the ligaments because a bluster to the knee while weight bearing is more likely to damage. An average of 30% players had menisci issues resulting in meniscectomy from past studies.

Ascertained and pre-arranged treatment not always winning, every time but also observes the remaining factors regarding sport type and health of the player. Chondromalacia patella has infrequently been used in treating such a knee issue. In anterior knee pain, Chandra changes are not possible always in anterior knee tenderness. Idiopathic anterior knee tenderness, no operative interference offers a superior long-term result than non-operative dealing. Tumors should be considered as a cause of enduring knee pain, particularly when night pain and localized swelling and skin change.

Figure 2.
Shows Traction apophysitis of the knee



2. METHOD AND MATERIALS

2.1. Pre-Processing of Data

Pre-processing of data is a basic process after raw data collection. In Python programming, pre-processing code is available to drop the empty values automatically. In my research, three different algorithms were used for predicting the results. A similar procedure applies to all algorithms (Keerthan, Nagasai & Shaik, 2019).

2.2. Feature assortment

Feature selection is the most important action for discovering the subset of unique characteristics by dissimilar strategies, all about data for this. They afford correctness and forecast errors. Features outcome is conversion of high-dimensional space data to a lower dimensional space of few attributes. Both linear and nonlinear diminution techniques can be used by the type of relations among the features in the dataset.

2.3. Model selection

Machine Learning (ML) has revolutionized the field of medical imaging, particularly in the analysis of Magnetic Resonance Imaging (MRI) scans. In the context of knee MRI images, ML algorithms have shown remarkable capabilities in extracting accurate information and handling complex patterns. Here are some key aspects:

1. Image Segmentation.
2. Feature Extraction
3. Pattern Recognition
4. Attention Mechanisms.

After completion of the feature selection and projection phases, the next crucial phase is model building based on chosen machine learning algorithms. Based on data size and data type, we chose different types of machine learning algorithms. These algorithms are classified as supervised and unsupervised learning. A supervised learning algorithm contains labeled input and output. To use training data to predict the future results. Unsupervised learning that is neither classified nor labeled allows the algorithm to analyze data without giving any direction.

2.2. Proposed method

The proposed methodology uses classification algorithms like Support Vector Machine (SVM), K- Nearest Neighbor (k-NN), and Random Forest (RF). Some real-life examples of Machine Learning (ML) design for identifying knee injuries.

KneeXpert: BME Detection using Gradient Boosting Automated Knee Injury Diagnosis.

2.2.1. K-nearest neighbor (KNN)

K-Nearest Neighbor is a machine learning algorithm, considered a supervised learning algorithm. It is a non-parametric method because it can't consider parameters (dimensions) and the classification of test data spots relies upon the near training data points of the dataset. It is solving classification and regression jobs both. The working procedure behind KNN is it presumes that data points recline in similar surrounds. It catches the thought of proximity based on a mathematical principle called Euclidean distance, it calculates distance between two points in a plane. For example, if the two points in a plane are G (p0, q0) and H (p1, q1) then the Euclidean distance between them is calculated as follows (Shravya, Pravalika & Subhani, 2019).

$$\sqrt{(p0-q1)^2 + (p0-q1)^2} \tag{1}$$

Based on the respective class identify the greater number of objects of its near neighbors. Now k values is 1, and then the novel input data point, and calculate the distances in the training dataset are computed. The shortest distance is considered as the nearest neighbor of our test data. Finally, the test data point is classified by its nearest neighbor. The value of k varies based on each dataset. K represents the proximity of a given test data point and the test data point is allotted to the class, it contains a greater number of nearest neighbors (Jauhiainen, et. al, 2022).

2.2.2. Support vector machine

SVM is a supervised machine learning algorithm that does well with pattern recognition issues and is used to study classification and regression rules from given data. This algorithm smoothly handles the outlier data compared to other algorithms. It uses the hyperplane for a binary classifier. The hyperplane of the support vector machine is built on mathematical equations. The hyperplane of the mathematical equation is y= ax + b. More number of features are handled by support vector machines and this training algorithm builds a design that plots novel data items to another category. The hyperplane that divides the two classes expertly. After that same data is mapped into a similar space and its category. Support vector machines handle linear data directly without any hazards. But it can't handle nonlinear data, so the support of the kernel overcomes the issue of nonlinear data. The following statistical equations represent linear and nonlinear data (Jauhiainen, et. al, 2022).

$$g(y) = W_i^{\cdot} \cdot Y + w_{i0} = 0 \tag{2}$$

Equation 2 represents the linear data handling. The following equation represents nonlinear data.

$$Y_i(W_i^{\cdot} y_i) \geq 1, \text{ i=1,2…………N} \tag{3}$$

$$0 \leq Y_i(W_i^{\cdot} \cdot Y + w_{i0}) < 1 \tag{4}$$

$$Y_i(W_i^{\cdot} y_i) < 0 \tag{5}$$

Equation 3 for vector falls outside of the circle and is correctly classified. Equations4 the data point vector falling inside the circle and are correctly classified. Equation 5 is used for vector misclassified. They are enclosed in a circle, which indicates inequality.

2.2.3. Random forest tree

Random forest is also a machine learning algorithm, it works on bagging methodology to make a bunch of small trees. This model is trained with a lot of random samples from the dataset to predict the performance. All trees combined and then create optimal prediction. The following equation 6 represents the random forest tree calculation (Subhani Shaik, 20.

RFfisub (j) = the significance of feature j measured from entire trees in the Random Forest design. (6)

2.2.4. Dataset description

KneeMRI dataset was collected from a Siemens Avanto 1.5T MR scanner, and obtained by proton density-weighted fat suppression method at the Clinical Hospital, India. The sample dataset is shown in Figure 3, which contains 91812-bit grayscale volumes of knees. Every record was assigned a diagnosis concerning the state of the anterior cruciate ligament in a double-blind manner; each record was labeled according to the ligament state: healthy, partially injured, and completely cracked. A broader rectangular region of interest was manually mined from the unique volumes and is also marked.

In this work, An Anterior Cruciate Ligament (ACL) technique addresses common knee injuries among top athletes. ACL tears happen when the anterior cruciate ligament is either stretched, partially torn, or completely torn. The most common injury is a complete tear. Symptoms include pain, a popping sound during injury, instability of the knee, and joint swelling. There are around 300,000 ACL tears every year in India, with over 125,000 ACL reconstruction surgeries per year. Magnetic Resonance Imaging is a medical imaging method used in radiology to form an image of the anatomy and the physiological process of the body. MRI is used to diagnose how well you responded to treatment as well as detecting tears and structural problems such as heart attacks, brain injury, and blood vessel damage, among others.

Figure 3.
Sample dataset for research

	kneeLR	roiX	roiY	roiZ	roiHeight	roiWidth	roiDepth
0	1	139	184	14	74	72	3
1	0	113	105	10	83	98	6
2	1	120	117	15	101	115	2
3	0	117	124	12	91	80	3
4	1	122	105	14	83	98	4
...	...	...	...	...	...	...	...
912	1	113	127	16	101	99	3
913	1	105	102	14	95	100	3
914	0	118	84	15	100	100	2
915	0	105	97	15	103	106	4
916	0	113	108	14	103	110	4

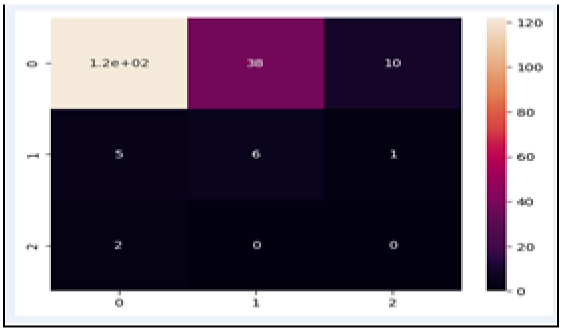
917 rows × 7 columns

3. RESULTS

3.1. Confusion Matrix

The confusion matrix is used to measure the performance of classification models, which intend to forecast a categorical label for every input instance. Multiple models are used to predict the results accurately. The following Figure 4 shows the confusion matrix for each input.

Figure 4.
Representation of Confusion Matrix

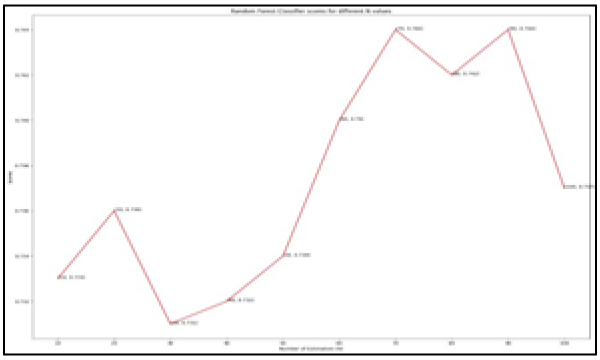


3.2. Random Forest Tree Reorientation

The following Figure 5 represents the graph of random forest tree classifier scores for different values.

Figure 5.

Representation Random Forest Tree

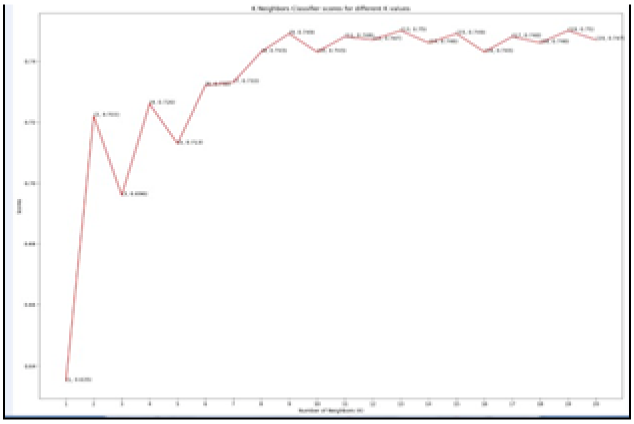


3.3. KNN Representation

The following Figure 6 represents the graph of KNN classifier scores for different values.

Figure 6.

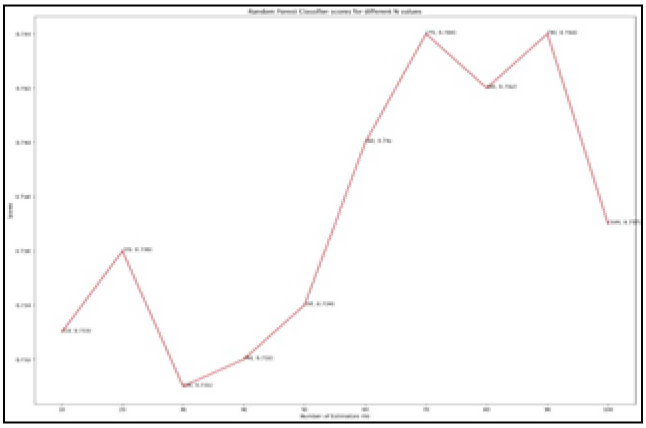
Representation of KNN



3.4. Support Vector Machine Representation

The following Figure 7 represents a graph for the SVM classifier scores for different values

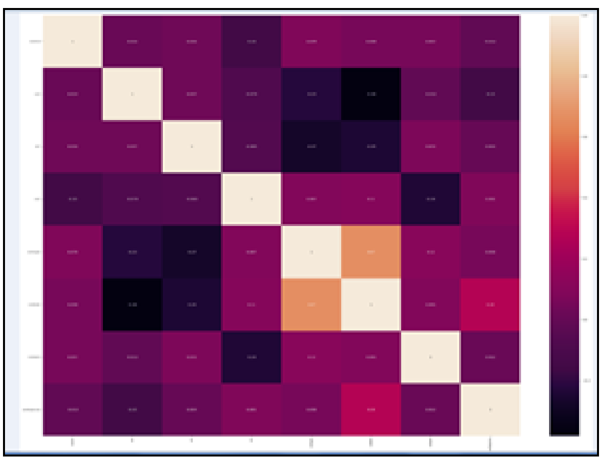
Figure 7.
Representation of SVM



3.5. HEAT Map

The purpose of Heat Map is to improve the visualization of the volume of locations within a dataset and assist in directing viewers towards areas on data visualizations shown in Figure 8.

Figure 8:
Representation of heat Map



3.6. Comparative study

In comparative study of research work, shows the result with accuracy, precision, and sensitivity in the following Table 1. It is observed that all algorithms show the same accuracy. The final evaluation exposes the maximum accuracy of 70.11% for the support vector machine, 69.75% for KNN, and 70.11% for the RF Tree maximum.

Table 1.
Comparison of the performance of various algorithms

Algorithms	Accuracy (%)	Precision	Sensitivity
Random Forest tree	70.11	73.12	71.23
KNN	70.11	74.56	74.33
SVM	69.75	73.34	71.76

4. Conclusion

The development of accurate injury prediction models using machine learning (ML) algorithms is crucial for improving patient outcomes and reducing healthcare costs.

1. Data quality and quantity
2. Feature engineering
3. Algorithm selection

Despite analyzing a large prospective data set with extensive anthropometric, clinical, genetic, neuromuscular, and biomechanical measurements, using a variety of ML methods, the predictive ability was too low for ACL injury risk assessment in clinical practice. Therefore, further studies are needed to investigate what type of data and ML approaches should be used for more accurate injury prediction. Multiple classifiers are applied for current research; it is observed that all classifiers predict the same cost of accuracy on similar datasets.

REFERENCE

Armstrong-Heimsoth, A., Johnson, M. L., McCulley, A., Basinger, M., Maki, K., & Davison, D. (2017). Good Googling: a consumer health literacy program empowering parents to find quality health information online. *Journal of Consumer Health on the Internet*, *21*(2), 111-124. http://dx.doi.org/10.1080/15398285.2017.1308191

Bickmore, T. W., Trinh, H., Olafsson, S., O'Leary, T. K., Asadi, R., Rickles, N. M., & Cruz, R. (2018). Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. *Journal of medical Internet research*, *20*(9), e11510. https://doi.org/10.2196/11510

Bini, S. A., Schilling, P. L., Patel, S. P., Kalore, N. V., Ast, M. P., Maratt, J. D., ... & Steele, G. D. (2020). Digital orthopaedics: a glimpse into the future in the midst of a pandemic. *The Journal of arthroplasty*, *35*(7), S68-S73. https://doi.org/10.1016/j.arth.2020.04.048

Buvik, A., Bugge, E., Knutsen, G., Småbrekke, A., & Wilsgaard, T. (2019). Patient reported outcomes with remote orthopaedic consultations by telemedicine: a randomised controlled trial. *Journal of telemedicine and telecare*, *25*(8), 451-459. https://doi.org/10.1177/1357633x18783921

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Zalaudek, I. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of oncology*, *29*(8), 1836-1842. https://doi.org/10.1093/annonc/mdy166

Higgins, J. P. (2016). Smartphone applications for patients' health and fitness. *The American journal of medicine*, *129*(1), 11-19. https://doi.org/10.1016/j.amjmed.2015.05.038

Jauhiainen, S., Kauppi, J. P., Krosshaug, T., Bahr, R., Bartsch, J., & Äyrämö, S. (2022). Predicting ACL injury using machine learning on data from an extensive screening test battery of 880 female elite athletes. *The American Journal of Sports Medicine*, *50*(11), 2917-2924. https://doi.org/10.1177/03635465221112095

Keerthan, J. S., Nagasai, Y., & Shaik, S. (2019). Machine learning algorithms for oil price prediction. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(8), 958-963.

Olczak, J., Fahlberg, N., Maki, A., Razavian, A. S., Jilert, A., Stark, A., ... & Gordon, M. (2017). Artificial intelligence for analyzing orthopedic trauma radiographs: deep learning algorithms—are they on par with humans for diagnosing fractures?. *Acta orthopaedica*, *88*(6), 581-586. https://doi.org/10.1080/17453674.2017.1344459

O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., ... & Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The international journal of medical robotics and computer assisted surgery*, *15*(1), e1968. https://doi.org/10.1002/rcs.1968

Rigamonti, L., Albrecht, U. V., Lutter, C., Tempel, M., Wolfarth, B., & Back, D. A. (2020). Potentials of digitalization in sports medicine: a narrative review. *Current sports medicine reports*, *19*(4), 157-163. https://doi.org/10.1249/jsr.0000000000000704

Sandal, L. F., Stochkendahl, M. J., Svendsen, M. J., Wood, K., Øverås, C. K., Nordstoga, A. L., ... & Søgaard, K. (2019). An app-delivered self-management program for people with low back pain: protocol for the selfBACK randomized controlled trial. *JMIR Research Protocols*, *8*(12), e14720. https://doi.org/10.2196/14720

Shravya, C., Pravalika, K., & Subhani, S. (2019). Prediction of breast cancer using supervised machine learning techniques. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(6), 1106-1110.

Subhani Shaik, D.R. (2017). classification of emg signals based on curvelet transform and random forest tree method. *Journal of Theoretical and Applied Information Technology*, *95*(24).

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?. *PloS one*, *12*(4), e0174944 https://doi.org/10.1371/journal.pone.0174944