OR-ensemble designed for interpretable predictions of rotator cuff injuries

Beau Fiechter, Danny Vink, Donna van Grunsven, Emma Korf, Jasmijn Heupers, Tim Bekema

Abstract

When diagnosing a rotator cuff injury, time and money can be saved by using a machine learning model in the process of diagnosing. The objective of this paper was to create classifier that distinguishes patients from non-patients and whether these individuals have a rotator cuff tear. It was key to build a model that is transparent to a physician. One-sided classifiers were combined using a Boolean OR-operator resulting in an OR-ensemble. A physician can retrace which feature(s) defined the classification. State-of-the-art ensemble methods, that have previously been adopted by studies where a machine learning algorithm is used in the medical sector, were compared to the OR-ensemble. Conclusively, this study demonstrates that the OR-ensemble results in higher scores in addition to having more transparent results.

Keywords: Ensemble method, Flock of Birds, Rotator cuff injury

1. Introduction

Rotator cuff injuries are common among people over sixty years old or people who tend to overload their shoulder during exercise [1]. Medical diagnosis usually involves the use of expensive and time-consuming imaging techniques such as X-rays, ultrasounds, and MRI. This project, OrthoEyes, could be a novel solution by having machine learning algorithms classify the degree of a patient's shoulder injury. This alternative approach could provide both financial and procedural benefits to the medical diagnosis methods stated above.

This study aims to improve the earlier established OR-ensemble by Vuurens, Andrioli, and de Vlugt. This OR-ensemble consists of single one-sided classifiers which each act as standalone deciding factors [2]. The results of the classifiers are combined using a Boolean OR-operator. Therefore, one deviating classifier in the OR-ensemble can decide the outcome. It is important that the OR-ensemble has a high rate of interpretability and the features selected of substantial clinical quality. Even though machine learning models have contributed to contemporary technology, it is difficult to properly apply to the medical field due to its lack of medical accountability. The additions to the OR-ensemble found in this study aim to tackle the issues stated above. This concept is applied on two experiments. One to distinguish patients from non-patients and the other to distinguish between individuals with or without rotator cuff tears. In conclusion, the purpose of this research is to assess in which ways the OR-ensemble can be improved.

The remainder of this paper is structured as follows: Section 2 describes the background and preparation of the dataset. Section 3 provides an analysis of interpretability and the feature selection. Section 4 will discuss the OR-ensemble in general. In successive order, Section 5 will examine the results of this study and provide a comparison to other models. Section 6 is dedicated to discussion points and future research on kinematic analysis of the shoulder and finally, Section 7 describes the conclusion.

2. Data

The dataset consists of positional and rotational data of different exercises performed by four patient groups. The data has been anonymized since it is a medical dataset. The majority of exercises lack a uniform protocol between each patient group. In order to cut out noise during movements, data cleaning had to be applied to all patient groups.

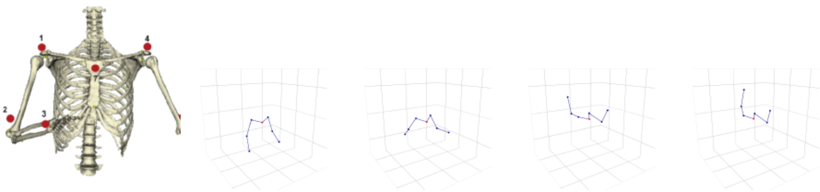


Figure 1: Flock of Birds-System

2.1 Sensor recordings

The dataset of this study was collected by the LUMC using a Flock of Birds (FoB) system. FoB is an electromagnetic motion capture system, which is a useful tool for shoulder kinematic studies [3]. A total of seven FoB-sensors were placed on each participant [figure 1] to simulate the bony landmarks of the shoulder region. The location of these sensors were the wrists, elbows, shoulders, and sternum.

Participants were asked to perform certain exercises that could be examined from a medical viewpoint: Abduction (AB), Anteflexion (AF), Retroflexion (RF) and Exo- and Endo-rotation Low (EL). During the AB-exercise, participants were asked to raise their arms in an arch to the highest point and then slowly return to the original position. During the AF-exercise participants were instructed to extend their arms forwards and lift them up to the highest point. After raising their arms to 180 degrees, the arms would return to their starting position. The RF-exercise can be performed by pointing the arms backwards as far as possible and then retracting them. Lastly, the EL-exercise was being performed by having each participant extend out their elbows with their upper arms raised. By continuing from this start position, participants would extend their upper arms backwards while simultaneously rotating their shoulders.

Met opmerkingen [K(1]: extend your elbows? dat doe je toch juist niet? je houdt de hoek van je elleboog gelijk, terwijl je je polsen zover mogelijk naar voor en achter beweegt

The positional data of each sensor has been recorded during these movements by the FoB-system. these datapoints are in an XYZ-cartesian space. The XY-plane represents the transverse view, the YZ-plane represents the frontal view and the XZ-plane the sagittal view.

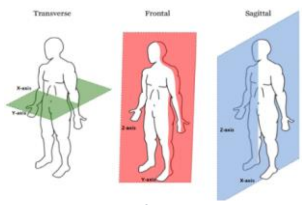


Figure 2: Planes of the human body

2.2 Labels
The dataset consists of four categories in which participants from Category 1 (n=30) have no injury. The participants in Category 2 (n=39), Category 3 (n=37), and Category 4 (n=28) have increased severities of shoulder injuries. Since there are no labels for the specifications of the injuries, it is unknown whether the right or left shoulder has been injured. In addition to this, there are no labels that indicate what kind of rotator cuff injury the participant has.

2.3 Cleaning
Two participants have been excluded, after the data was inspected, because they did not perform all four exercises (AB, AF, RF and EL). Three other participants have been excluded because the sensor, which was supposed to be placed on their sternum, was instead placed on their back between both scapulae.

The data has been cleaned after discovering noise before and after an exercise. Additionally, some participants performed an exercise twice in one recording, while others performed an exercise only once. The execution of an exercise has been sliced and the noise has been deleted. These frames mostly contained either no movement at all or random movements, which were not a part of the exercise. If a participant performed an exercise twice in one recording, the second execution has been stored in a new recording.

Unfortunately, the range of the coordinates in each category was not uniform. For example, the X-coordinates from Category 1 ranged between 700 and 1300, while the X-coordinates from Category 4 ranged between 30 and 50. This was solved by calibrating all the categories. For each participant, the calibration consisted of two parts: translation and scaling. Firstly, the positional data from all sensors was transformed, so that for every frame, the coordinate for the sensor on the sternum (sensor 3) was placed on the coordinate (0, 0, 0). Secondly, scaling was applied to ensure that the positional data from all categories was in the same range. The average length of both upper arms was calculated by taking the length of the 3D-vector between the sensor on the shoulder and the sensor on the elbow. Then the average length of the two upper arms was calculated and scaled down to a length of 1. Based on this scaling, the rest of the positional data was scaled as well.

2.4 Test set
For the final validation of the model, twenty percent of the participants from each category have been randomly selected and isolated.

3. Analysis

The features which are used as input for the model need to be interpretable. These features will be analysed and justified in the upcoming chapter. Additionally, the calculations need to be performed on the data to be used as input.

3.1 Interpretability
From a medical standpoint it is of high importance that the results are interpretable, considering a physician will not make conclusions based on results without supporting evidence. Therefore, the selected features which indicate the decision should be retraceable for each participant. Otherwise, extra diagnosing would be needed by the physician.

3.2 Feature selection
Features need to be selected on their data-scientific significance, as they need to have a certain contribution to the system. However, these features need to be medically justifiable, therefore a patient must be classified based on a data perspective that reflects medical classification methodology.

3.2.1 Medical justification

When researching possible features to add to the OR-ensemble, it is key to choose features that are medically relevant. In this section the significance of these features is described.

Shoulder angles and compensation behaviour

An individual without a shoulder injury will have no trouble performing the abduction and anteflexion movement. This means that the angle of the shoulder will approach 180°. If a participant is unable to approach this angle, it increases the likelihood of the person having a rotator cuff injury. Idem ditto for the maximum height of the wrist during these movements [4]. Besides not being able to reach the optimal angle of the shoulder and height of the wrist, patients with a rotator cuff injury are likely to compensate their lack of height during the movement by extensively raising their injured shoulder [10].

Velocity and acceleration of the movement

Rotator cuff injuries are associated with loss of shoulder stability [5]. Having a decent shoulder stability implies one is able to execute the movement fluently. Similarly, the speed in which the movement is executed could be an indicator of possible injuries of the rotator cuff.

Symmetry of the movement

Loss of strength in the shoulder is, in many cases, a consequence of a rotator cuff injury [6]. To measure this loss of strength, the injured arm was compared to the not-injured arm [7]. This is done by comparing the difference in height, speed, and acceleration of the wrists and elbows for both sides. A greater difference between the left and right arm during an exercise, indicates a higher probability of the participant having a rotator cuff injury.

Deviation in an exercise

Given that injured people will avoid painful movements as much as possible [8], it is feasible to recognise a rotator cuff injury by looking for deviation in an exercise. Conclusively, this means that participants who experience pain 'move around' a painful part of the exercise to avoid feeling this pain.

3.2.2 Calculations

Each of the named indicators, along with some others, can be combined with one or multiple meaningful mathematical operations forming features. For each of these features, a logistic regression model is made. After ensuring the credibility of the chosen feature, the model is added into an ensemble.

Shoulder angle

The maximum shoulder angle seems insightful when diagnosing people with rotator cuff injuries. In the AB and AF movements the Y-axis is irrelevant to observe the main motion. Therefore, the angle of the shoulder is calculated using the locations of the elbow and shoulder in the XZ-plane, for both the left and right side of the body. This is done by creating a 90-degree triangle using the shoulder and elbow point and adding a third point at the height of the elbow and the width of the shoulder. Due to the fact that the locations of the sensors are known, the lengths between the points are calculated. Then, taking the arccosine of the two lengths results in the angle of the shoulder. For every frame of the movement the angle is calculated and using the 'max' operation the maximum shoulder angle will result. The same approach is used to calculate the shoulder angle for the AB movement. Only during this movement, the subject moves in a different plane, as explained in chapter 2.1. In this case the YZ-plane shows the majority of the action and therefore the X-axis can be discarded. The rest of the calculations remain the same as explained in the previous paragraph.

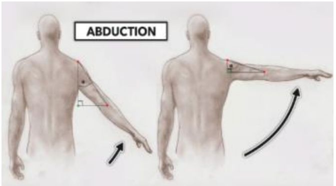


Figure 3: Calculation shoulder angle

Symmetry in position
Continuing, the following features are based on the difference in position between the left and right side of the body. As well as the previous features, the meaningful axes differ per exercise [A]. The distance from the sternum to the wrist, elbow and shoulder is retrieved and the variation between the right and left side is calculated by taking the mirrored difference. These factors are calculated for each frame in the movement. The mathematical operation that is enforced on the feature bases is 'max'.

Furthermore, the positional data of the elbows and wrists along the Z-axis during movements AB, AF and RF give insight into how well the movement was performed. Taking the maximum value of these heights is expected to be a reliable feature to distinguish patients from nonpatients. Regarding the EL movement, only the maximum heights of the wrists along the X-axis are useful to consider for classification. The elbows are moved throughout the movement EL, which is a flaw in the dataset. This has to be taken to account while examining the wrist positioning during the movement. Due to this, the difference in position of the wrist is not taken to account for movement EL.

Velocity and acceleration
The smoothness of a motion is measured by determining the standard deviation of the velocity and acceleration during that movement. A great difference in acceleration within one exercise, implies there is little to no stability in the shoulder. Therefore, large fluctuations in acceleration is an indicator for having rotator cuff tears.

Angular velocity and acceleration
The angular velocities and accelerations of a movement can likewise give a good indication of the shoulder quality of the participant. These are calculated by retrieving the velocity of the right and left elbow, in the desired plane for every movement. By dividing the velocity by the length of the arm, the angular velocity remains. Same is done for the acceleration, only the differential of the angular velocity is taken after the previous calculations. The mathematical operation 'std' is used on the angular velocity, because this gives insight into the degree of variation in the speed whilst not measuring protocol, since participants were instructed to do the movement at different speeds. As for the angular acceleration both the 'std' and 'mean' can be used to build a feature, since the protocol will not be measured by this operation.

4. Model

The OR-ensemble brought forward by this study is based on the combination of multiple one-sided logistic regression models, using a Boolean OR-operator. This results in the option to combine only the features that possess a precision of 1.0. In this chapter the hyperparameter *f* (factor) is introduced and the essence is elaborated on using a feature distribution example. The process of establishing the factor is discussed briefly.

4.1 One-Sided logistic regression

The trained models for each feature assign identifiable patients to their designated categories. Leave one out is applied to optimize the use of the small data set. The model searches for participants who are identifiable. Only certain values that are distinctly lower or higher, which depends on the polarity of the feature, will be selected as those patients are easily identifiable.

To achieve the precision score of 1.0, the model introduces a hyperparameter *f*. This parameter creates a threshold applied for all the different features. The hyperparameter *f* is used to change the value of the threshold. The following formula is used to determine the threshold.

$$t = (1 - f) \cdot \overline{x_0} + f \cdot x_{0,low}$$

The green dots represent the patients from category 2, 3 and 4. The control group, category 1 is shown as the red dots (y=0). Figure 4 shows that some values of the patient (y=1) group match the values of the control group. This results in patients being wrongly identified to a category. If the value of *f* is increased, the threshold boundary moves as seen in figure 5.

The participants labelled as y=1 who resemble the values of y=0 a lot are disregarded, in the placement of participants among the four categories. This means the training data only consists of distinct participants, which makes them simple to differentiate from the y=0 participants.

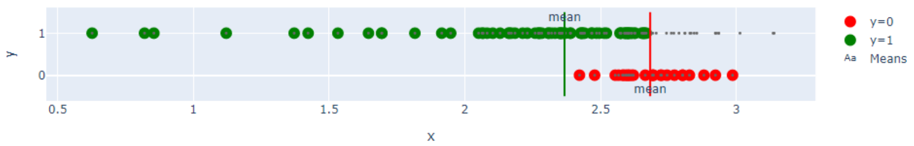


Figure 4:The hyperparameter f is set to 0. Consequently, the threshold lays at the mean of y = 0. Overlap of two groups causes wrong identification in the range between the mean values.

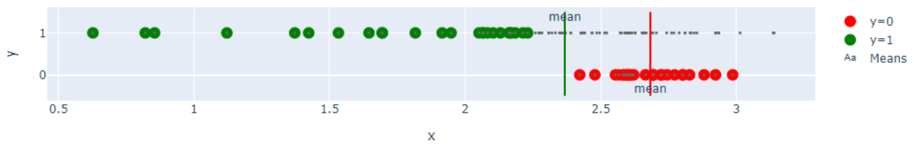


Figure 5: The hyperparameter f is set to 1.7. There is no overlap of green and red dots, which means few false identifications will be made.

4.2 OR-ensemble
The one-sided classifiers, one for each feature, can be combined by state-of the art ensemble algorithms, e.g. AdaBoost and Random Forest. However, these algorithms may be hard to interpret. As was explained in section 3.1, the interpretability of the algorithm is of great importance to the medical domain.

When using the hyperparameter *f*, each one-sided classifier predicts a subset of patients with high precision. The one-sided classifiers with a precision of exactly 1.0 are combined through a Boolean OR-operator. If at least one of the classifiers predicts a participant to be a patient, the ensemble predicts the participant to be a patient as well. The features on a patient are being classified can always be retracted from the model. Therefore, the OR-ensemble created by this method is easy to interpret.
For all one-sided classifiers within the OR-ensemble the hyperparameter *f* is set to the same value. This value is manually picked. Firstly, the results for different values between 1.0 and 2.0 are examined. Secondly, the value with the optimal results is picked.

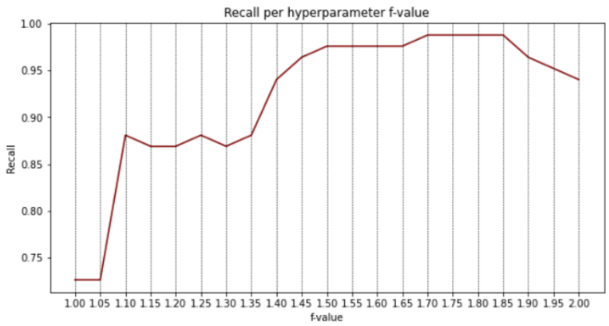


Figure 6: hyperparameter tuning for classifying patients from non-patients

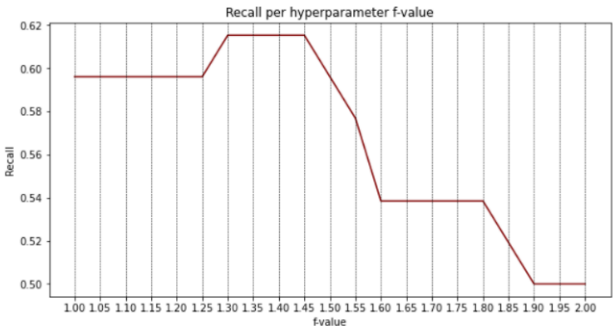


Figure 7: hyperparameter tuning. classifying rotator cuff tears from shoulder pain

5. Results

In the following chapter the results of the two experiments are shown. Both chapters contain the results from the OR-ensemble and other ensemble methods. The first paragraph shows the validating and test results for distinguishing patients from non-patients. The second paragraph shows the validating and test results for distinguishing patients with rotator cuff tears from patients with regular shoulder pain. The final paragraph explains how the different ensemble methods have been executed and tuned.

5.1 Classifying patients from non-patients

These tables show the validation results (table 1) and the test results (table 2) for distinguishing non-patients from patients. Through hyper parameter tuning, shown in figure 6, the factor for the OR-ensemble is set on 1.7. For the validation set, the OR-ensemble gives near perfect results. The test set gives slightly less impressive results.

	OR-ensemble f=1.7	**AdaBoost**	**Random Forest**	**Bagging**	**SVC**
Precision	1.00	0.92	0.95	0.90	0.93
Recall	0.98	0.90	0.95	0.89	0.92
Accuracy	0.99	0.90	0.95	0.89	0.92

Table 1: Validation results for classifying patients from non-patients

	OR-ensemble f=1.7	**AdaBoost**	**Random Forest**	**Bagging**	**SVC**
Precision	0.9	0.74	0.96	0.8	0.87
Recall	0.95	0.76	0.96	0.88	0.88
Accuracy	0.88	0.76	0.96	0.88	0.88

Table 2: Test results for classifying patients from non-patients

5.2 Classifying patients with rotator cuff tears from patients with regular shoulder pain

These tables show the validation results (table 3) and the test results (table 4) for distinguishing patients with and without rotator cuff tears. Through hyper parameter tuning, shown in figure 7, the factor for the OR-ensemble is set on 1.45. For the validation set, the OR-ensemble remains a precision of 1.0 while the recall and accuracy have significantly declined. The test set results in a precision that is not equal to 1.0, whilst this was the goal. Even though the OR-ensemble scores are not outstanding, they are better/equal to the results of the other ensemble methods.

	OR-ensemble f=1.45	**AdaBoost**	**Random Forest**	**Bagging**	**SVC**
Precision	1.00	0.59	0.65	0.60	0.61
Recall	0.61	0.59	0.65	0.59	0.61
Accuracy	0.76	0.59	0.65	0.59	0.61

Table 3: Validation results for classifying patients with rotator cuff tears from patients without tears

	OR-ensemble f=1.45	**AdaBoost**	**Random Forest**	**Bagging**	**SVC**
Precision	0.78	0.40	0.47	0.73	0.73
Recall	0.58	0.63	0.52	0.75	0.73
Accuracy	0.63	0.63	0.49	0.75	0.73

Table 4: Test results for classifying patients with rotator cuff tears from patients without tears

5.3 Comparing ensemble methods

Machine learning has become one of the most prominent new technologies of the decade. However, many machine learning models lack the ability of interpretability and transparency, especially when used within the medical domain [9]. This study examines a small medical dataset and classifies patients based on an OR-ensemble learning method. In order to ensure that the OR-ensemble is the best possible model for the rotator cuff classification problem, the OR-ensemble is being compared to state-of-the-art models that have been successfully used in medical studies [10][11][12][13][14].

The financial upside is a major plus, due to the fact that there will be no false positives predicted while training the model. A non-patient will never have to account for costs of scans, due to the fact that the OR-ensemble does not classify a non-patient as patient by mistake.

5.3.1 Random forest

A Random Forest model uses multiple decision trees to make a final decision. The Random Forest model was applied to the transformed dataset. By automatically optimizing the hyperparameters of the Random Forest model, the score reaches its optimum value. However, the precision score never reached 1.0. This means there will be false positives using the Random Forest. For medical purposes false positives are undesirable, since these will result in unnecessary costs.

5.3.2 AdaBoost
AdaBoost is an ensemble learning method that can be applied to regression problems. The AdaBoost boosting algorithm boosts the model by building a strong learner from the mistakes of several weaker models [15]. As can be seen in [C], the hyperparameters have been tuned to generate the optimal results. When comparing AdaBoost to the OR-ensemble, it can be concluded that the AdaBoost scores notably worse than the OR-ensemble. An important difference between these two ensemble methods, is that the precision in the AdaBoost method will not reach 1.0. This means the model will ultimately consist of false positives. False positives are generally unfavourable, therefore AdaBoost can be seen as a worse alternative compared to the OR-ensemble.

5.3.3 Bagging
Bagging is an ensemble-based algorithm that uses bootstrapping and aggregating techniques. A Bagging Regressor was used to compare the Bagging ensemble against the OR-ensemble. Hyperparameters were tuned by using the same methodology as previous methods as seen in [D]. The amount of decision trees will have no impact on overfitting due to the stochastic nature of Bagging itself. Even though Bagging is best suited for small training data sets [15], the results of Bagging were comparatively lower than the OR-ensemble during both experiments. Similar to other ensemble learning methods, the Bagging ensemble could not uphold a precision of 1.0, therefore being deemed unfit for the medical problem presented in this study.

5.3.4 Linear Support Vector Classification
As for the SVC, the results regarding the optimized hyperparameters were quite reasonable. The results are calculated and plotted for a c-parameter in the range of 10^{-4} to 10^{4}, as shown in [B]. The OR-ensemble scores better when differentiating patients from non-patients. Seen in table 1 and 2, the results of SVC are better concerning the recall and accuracy compared to the OR-ensemble. Yet the OR-ensemble still is the best choice when trying to classify regarding the medical field seeing the precision of 1.0 is reached. This ensures that non-patients are, without a doubt, never classified as a patient.

6. Discussion
The most important finding of this study is that the OR-ensemble is able to distinguish patients from non-patients, in addition to whether they have a rotator cuff injury or not. The precision score of the OR-ensemble will always be 1.0, when testing with the validation set. Therefore, the number of false positives will be zero.

The models that were combined into the OR-ensembles are based on medical tests that are associated with the identification of shoulder injuries and rotator cuff tears. Those logistic regression models represent features based on active movements. In the process of diagnosing a rotator cuff injury, medical specialists look at the passive movements as well [4]. This has not been considered in this research. Important to state is that dataset for this research is relatively small (n=134). Besides the small dataset, the individual logistic regression models have only been trained on the identifiable members of the dataset. Consequently, the model may be trained on insufficient data.

As the data has been collected from different studies, there may be some differences in protocol. To make sure that these differences do not interfere with the quality of the results, it may be beneficial to use the method of the OR-ensemble on new data, which has been generated using one protocol. A last caveat in this research is the lack of background information on the participants. There is no information of the age, gender, and previous injuries of the patients. This information might contribute to indicating whether the subject has a rotator cuff injury or not.

The features were manually picked during the feature selection. Determining the added value for each feature gives insight on the importance of the features. With more information about the importance of a feature, the number of features can be narrowed down.

In this paper, a global factor for separating in- or outliers in the dataset is chosen for all features at once. This might not be a good approach when looking at some of the regression lines for each feature and the equation describing the separation of individuals. Based on the multiplication of the mean values, outliers are filtered out of the dataset. Getting a set on which a logistic regressor can be fitted properly is strongly dependent on the distribution of datapoints for a feature. When means are close to each other, the factor needs to be very high to get any meaningful results for a feature. Vice versa, when the difference in means is large, the factor does not need to be that high at all. Therefore, it might be a better approach to automate the selection of the value for the factor parameter for each feature in some way.

Future research
In addition to the completed research brought forward in this paper, there are some particularly interesting subjects to consider looking into for future research. The final goal of OrthoEyes is to classify a rotator cuff injury of a patient by using a camera in a waiting room of the doctor's office. By providing the doctor with a first indication of the degree of the shoulder injury, a few steps during the research cycle can be skipped.

To accomplish the final goal of this project, it is necessary to dive into the world of cameras and how to position them accordingly. Considering it is relatively easy for cameras to obtain positional data, compared to taking measurements with the FoB system which takes longer to prepare. Doing research on the optimal camera type and the most favourable position of this camera is essential to build forward on the completed work.

Due to the fact that it is hard to obtain rotational data using cameras, it seems unnecessary to continue looking into the rotation matrices and features which are based on this data. In addition to this, the researchers at the LUMC used specific calculations to transfer the placing of the sensors to the bony landmarks of the individual, before starting actual calculations using rotation matrices. The functions used by the LUMC researchers are unavailable for groups working on this project and without these functions it is impossible to use the rotation matrices accurately. Analysing the rotation matrix data can lead to a promising feature but will not lead to a useful contribution to the long-term goal of this project. Seeing the missing information is essential to grasp a full understanding of the matrix data, it seems best to leave the matrices out of further research.

7. Conclusion
The main objective of this study was to create a model that would distinguish patients from non-patients and patients with or without a rotator cuff injury. When creating this model, it was crucial to make it relatively transparent to the physician. Besides high transparency of the model, generating no false positives is key. To accomplish this, the OR-ensemble was created. The OR-ensemble consists of several one-sided logistic regression models, that each represent a chosen feature. Comparing the OR-ensemble to conventional ensemble methods showed that, besides a better interpretability, the OR-ensemble generates better outcomes.

Met opmerkingen [H(2]: dadelijk even controleren of dit echt zo is

Met opmerkingen [GDv(3R2]: Rip random forest

Met opmerkingen [EK4]: Not true toch?

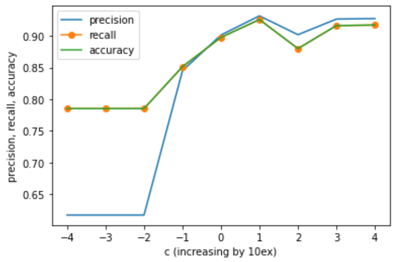
Appendix

[A]

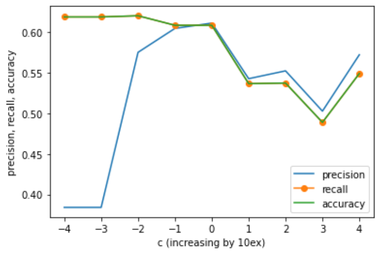
Feature basis	Mathematical operation	Axis / plane	AB	AF	RF	EL
angle_left_shoulder_xz	Max	XZ		✔	✔	

angle_right_shoulder_xz	Max	XZ		✓	✓	
angle_left_shoulder_yz	Max	YZ	✓			
angle_right_shoulder_yz	Max	YZ	✓			
diff_x_wrist	Std	X		✓	✓	✓
diff_x_elbow	Std	X		✓	✓	✓
diff_x_shoulder	Std	X		✓	✓	✓
diff_y_wrist	Std	Y	✓	✓	✓	✓
diff_y_elbow	Std	Y	✓			
diff_z_wrist	Std	Z	✓	✓	✓	
diff_z_elbow	Std	Z	✓	✓	✓	
diff_z_shoulder	Std	Z	✓	✓	✓	
z_elbow	Max	Z	✓	✓	✓	
z_wrist	Max	Z	✓	✓	✓	
x_wrist	Max	X				✓
vel_wrists_x_l	Std	X				✓
vel_wrists_x_r	Std	X				✓
vel_elbows_z_l	Std	Z	✓	✓	✓	
vel_elbows_z_r	Std	Z	✓	✓	✓	
acc_wrists_x_l	Mean, std	X				✓
acc_wrists_x_r	Mean, std	X				✓
acc_elbows_z_l	Mean, std	Z	✓	✓	✓	
acc_elbows_z_r	Mean, std	Z	✓	✓	✓	
angular_vel_xz_elbow_l	Std	XZ		✓	✓	
angular_vel_xz_elbow_r	Std	XZ		✓	✓	
angular_acc_xz_elbow_l	Mean, std	XZ		✓	✓	
angular_acc_xz_elbow_r	Mean, std	XZ		✓	✓	
angular_vel_yz_elbow_l	Std	YZ	✓			
angular_vel_yz_elbow_r	Std	YZ	✓			
angular_acc_yz_elbow_l	Mean, std	YZ	✓			
angular_acc_yz_elbow_r	Mean, std	YZ	✓			

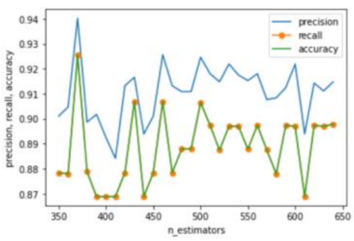
[B] Linear SVC hyper parameter tuning
Category 2, 3 and 4 vs 1
Selected c-factor = 10^1

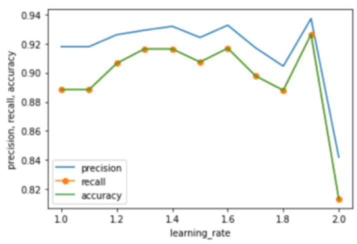


Category 3, 4 vs 2
Selected c-factor = 10^0

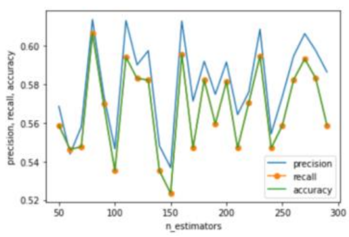


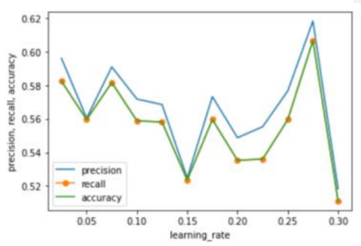
[C] AdaBoost hyper parameter tuning
Category 2, 3, 4 vs 1
n_estimators = 370
learning_rate = 1.9



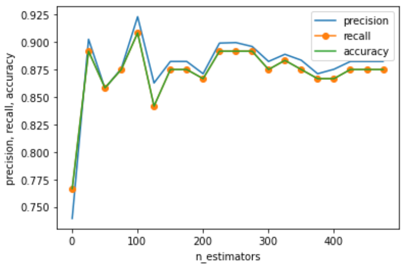


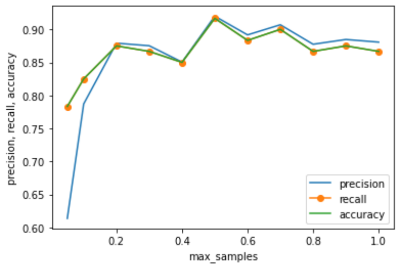
Category 2, 3 vs 4
n_estimators = 80
learning_rate = 0.275



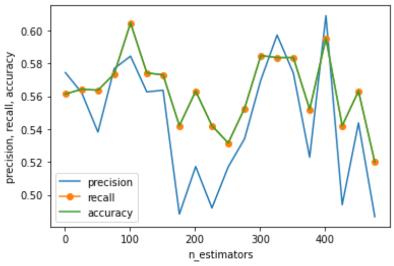


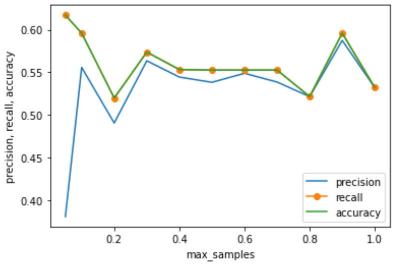
[D] Bagging hyper parameter tuning
Category 2, 3, 4 vs 1
n_estimators = 100
max_samples = 0.5





Category 2, 3 vs 4
n_estimators = 400
max_samples = 0.9





References

[1] van der Zwaal, P., Pekelharing, J. F., Thomassen, B. J. W., van Arkel, S., & van Arkel, E. R. A. (2011, 22 augustus). Rupturen in de rotatorcuff: diagnose en behandeling. https://www.ntvg.nl/artikelen/rupturen-de-rotatorcuff-diagnose-en-behandeling/volledig#toc2
[2] J. Vuurens, T. Andrioli, E. de Vlugt. One-Sided Classification for Interpretable Predictions. 2020.
[3] Meskers, C. G. M., Fraterman, H., van der Helm, F. C. T., Vermeulen, H. M., & Rozing, P. M. (1999). Calibration of the "Flock of Birds" electromagnetic tracking device and its application in shoulder motion studies. Journal of Biomechanics, 32(6), 629–633. https://doi.org/10.1016/s0021-9290(99)00011-1
[4] Jain, N. B., Wilcox, R. B., Katz, J. N., & Higgins, L. D. (2013). Clinical Examination of the Rotator Cuff. PM&R, 5(1), 45–56. https://doi.org/10.1016/j.pmrj.2012.08.019
[5] Morag, Y., Jacobson, J. A., Miller, B., De Maeseneer, M., Girish, G., & Jamadar, D. (2006). MR Imaging of Rotator Cuff Injury: What the Clinician Needs to Know. RadioGraphics, 26(4), 1045–1065. https://doi.org/10.1148/rg.264055087
[6] OrthoInfo. (2017, maart). Rotator Cuff Tears. https://orthoinfo.aaos.org/en/diseases--conditions/rotator-cuff-tears/
[7] Jobe, F. W., & Moynes, D. R. (1982). Delineation of diagnostic criteria and a rehabilitation program for rotator cuff injuries. The American Journal of Sports Medicine, 10(6), 336–339. https://doi.org/10.1177/036354658201000602
[8] Vlaeyen, J. W. S., & Crombez, G. (1999). Fear of movement/(re)injury, avoidance and pain disability in chronic low back pain patients. Manual Therapy, 4(4), 187–195. https://doi.org/10.1054/math.1999.0199
[9] F. K. Dosilovic, M. Brcic, and N. Hlupic. Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*, pages 0210–0215. IEEE, 2018.
[10] Minz, A., & Mahobiya, C. (2017). MR Image Classification Using Adaboost for Brain Tumor Type. 2017 IEEE 7th International Advance Computing Conference (IACC), 701–705. https://doi.org/10.1109/iacc.2017.0146
[11] Yang, F., Wang, H.-, Mi, H., Lin, C.-, & Cai, W.-. (2009). Using random forest for reliable classification and cost-sensitive learning for medical diagnosis. BMC Bioinformatics, 10(Suppl 1), S22. https://doi.org/10.1186/1471-2105-10-s1-s22
[12] Tu, M. C., Shin, D., & Shin, D. K. (2009). Effective Diagnosis of Heart Disease through Bagging Approach. 2009 2nd International Conference on Biomedical Engineering and Informatics, 305–323. https://doi.org/10.1109/bmei.2009.5301650
[13] Vanitha, L. and Venmathi. A. R. (2011). Classification of Medical Images Using Support Vector Machine. 2011 IACSIT Press, Singapore, 63-69.
[14] Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC Medical Informatics and Decision Making, 10(1), 10–16. https://doi.org/10.1186/1472-6947-10-16
[15] Zhang, C., & Ma, Y. (2012). Ensemble Machine Learning. Springer Publishing. https://doi.org/10.1007/978-1-4419-9326-7