IMPROVING CAPABILITIES OF EMG SIGNALS WITH PREDICTION ALGORITHMS[1*]

Emre PARLAK[1], Çağdaş ÖZER[2], Mustafa TAKAOĞLU[3]

[1] *Defense Technologies Institute, Gebze Technical University, Kocaeli, Turkey, emreparlak@gtu.edu.tr, ORCID: 0000-0003-2668-1660*

[2] *Department of Computer Engineering, Istanbul Kültür University, Istanbul, Turkey, c.ozer@iku.edu.tr, ORCID: 0000-0002-0581-7955*

[3] *Department of Computer Engineering, Istanbul Aydın University, Istanbul, Turkey, mustafatakaoglu@aydin.edu.tr, ORCID: 0000-0002-1634-2705*

ABSTRACT

In this study, long-term performance signals were measured; to minimize the changes in the characteristics of the signals due to long term performance of amputated arm patients, it is aimed to improve the read signals by using machine learning algorithms. In our study, the data obtained from the measurements we made through the Armband device of the right arms of seven people were used. While the data were obtained, the hand was turned into a fist, and this movement continued until fatigue occurred in the muscle. Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Tree, Random Forest, and Gradient Boosted Trees algorithms are used to process signals, and 16796 models are created. Data were analyzed based on Accuracy, Classification Error, Area Under Curve, Precision, Recall, F Measure and Specificity. The algorithms that yield the best results were determined in each variable, and the results were shared. This study was orally presented in 10th International Conference on Image Processing, Wavelet and Applications, IWW2019.

Keywords: EMG Signals, Machine Learning Algorithms, Signal Processing, Signal Efficiency, MYO Armband.

1. INTRODUCTION

The use of prosthetic limbs has been applied to millions of people since the 1970s [6]. The prosthesis limbs developed due to limb losses due to various reasons allow people to hold onto life. Today, studies in this field are continuing increasingly. Wiener has proposed the idea of using prosthetic arms using Electromyogram (EMG) signals for real-time motion. These signals were used as a control mechanism for the prosthetic limb systems. Boston Arm(MIT) and the Utah artificial arm are some examples [8].

In this study, we worked with EMG signals. The use of electromyographic (EMG) signals from skeletal muscle advantages of being both convenient and natural [1]. That means EMG signals are directly correlated with the contraction and relaxation of muscle fibers [7]. We're using MYO Armband to detect EMG signals. Our purpose is to maximize signal detection when long term muscle movements happened. When muscles are tired due to long term usage, characteristic of the signal has been changing. When the signal changes, it might affect the working of the mechanism. So with the help of machine learning algorithms, we tried to increase the success of the muscle signal process. This way, with enough data and properly applied machine learning techniques, system will work as intended even if the signal starts to change and the model will predict the correct movement so do the work as its intended. To achieve this, we have used MYO Armband on seven volunteers and get right arm muscle data. After collecting the tired muscle data, first of all different subject's data were compiled within a single dataset. When the dataset is completed, the data were cleaned with some preprocessing steps. These include removing the repeating rows, getting rid of low variance values, high correlated values, applying normalization and principal component analysis. The main reason for using these methods is to correctly predict the required values while minimizing training time, increasing the accuracy of the classification models and so increasing the results of the prediction for the research. When the preprocessing is done, this data is directly used as an input on the eight different machine learning algorithms. Results have been shared on the Results and Discussion part. The importance of our study is to share the results of different machine learning algorithms that worked with our dataset.

2. MATERIALS AND METHODS
In this study, muscle signal measurements were performed with MYO Armband produced by Thalmic Labs. MYO Armband is a wristband with 8 EMG electrodes, three axis accelerometer, three axis gyroscope and three axis magnetic force measurement. This wrist strap can also show the orientation of the arm in 3-dimensional space. The orientation data from the wristband is transmitted to the computer via wireless communication (Bluetooth). After the data is processed with the software prepared in Python programming language, it is sent to the industrial robot in real-time via TCP / IP communication [3].

Dataset collected by MYO Armband are processed and used by machine learning algorithms which are Random Forest, Deep Learning, Gradient Boosted Trees, Decision Tree, Naïve Bayes, Logistic Regression, Generalized Linear Model, and Fast Large Margin. These algorithms are explained in next titles. The reason these algorithms are picked, because the data we are working on requires classification to make the prediction. Although some of the algorithms also works on regression, the main idea is to classifications these data properly and preparing the prediction models based on this accurate classification.

2.1. NAÏVE BAYES
Naive Bayes is a machine learning algorithm used for classifications and has been proven to be effective in areas such as email spam filtering and document categorization [9]. This algorithm is preferred for efficiency, low variance, increased learning, direct estimation of probability, robustness in data corruption, and robustness in incomplete data [10]. The

efficiency of the calculation is important because the performance of the algorithm, especially for big data in modeling and estimation, makes the difference in terms of time, cost and stability, as well as the reliability of the results obtained [2].

2.2. GENERALIZED LINEAR MODEL

The generalized linear model is the algorithm where the usual linear model is generalized [11]. The linear model says that the sum of the model and data is equal to the data. Anova, Ancova and regression analysis have emerged from this model. It is different from usual model in two main aspects, for example, where multiple regression is somewhat a different case: It only contains information about the ranks. Secondly, dependency variable values are estimated from the combination of estimates of the linear variable which is linked to dependency variable by a link function.

2.3. LOGISTIC REGRESSION

Logistic Regression is a statistical management that has one or more arguments and is used to determine a result. The analysis of an existing data set yields two possible results and is used in linear classification problems. Logistic regression contains binary 1 or 0 encoded data.

For a logistic regression analysis in which the machine will find out whether the person is a hyperthyroid or not, the result is negative if the result is 1 and the result is not hyperthyroid. From this example, it can explain the purpose of logistic regression as follows; to find the most suitable model for defining the suitable connection between a group of sovereign variables related to the dependency variable having two-way characteristics.

2.4. FAST LARGE MARGIN

Idea of the fast large margin is to apply a speedy margin learner to a SVM based on a learning plan [4]. Although similar to those presented by conventional SVM or logistic regression applications, this linear classifier can work on a dataset containing millions of samples and attributes. It has been preferred due to its ability to work with large scale data.

2.5. DEEP LEARNING

Deep learning is developed from artificial neural networks and is a widely used algorithm for machine learning. Main idea is to extract automatically the necessary classification for low and high level features. Deep learning models often adopt hierarchical structures to interconnect layers. Having this feature can be more powerful in the feature presentation of deep learning models than shallow machine learning models. While the performance of traditional machine learning methods is often based on users' experiences, deep learning approaches are based on data.

2.6. DECISION TREES

This algorithm has been used for many years as a prediction, classification and rule determination algorithm. The ID3 algorithm has come a long way in finding information. Advantage of the decision tree is being intuitive and easy to understand. The decision tree is not only easy to classify and predict, but it is also very easy to understand the course of

development, classification and forecasting. However, the variable value is a continuous and gradual change in many industrial controls and optimizations, and in most cases it is not necessary to control the exact scope, but to classify for the decision, not to produce the exact predictive value.

2.7. RANDOM FOREST

When communities of decision trees work together, it is known as Random Forest. Each tree in the Random Forest estimates one class, and the class with the most votes becomes the prediction of the model. This algorithm works well where there are classes and lots of data because the better the estimation according to the working principle, the higher the estimation level. Of course, the lower the correlation in the data, the more accurate the predictions will be. Some of these trees may be true and some of them may be wrong, but as the algorithm iterations continue, these trees will move forward and get closer to the result.

2.8. GRADIENT BOOSTED TREE

This algorithm is known for the ability of being able to work with both the Regression and Classification issues [5]. When considered as a tree, it can be said that it stores only scalar values in its leaves. Multiple scalar deciduous trees should be used to address vector regression or multiclass classification problems. The reason for its use is that the data obtained from the trees that have been mis-sampled is used to prevent errors in subsequent trees. So this algorithm learns from its mistakes and tries not to repeat these steps. It is a useful algorithm in terms of this feature and it can produce very good results with sufficient iterations in case of estimation.

3. RESULTS AND DISCUSSION

As explained in the previous chapters of our study, EMG signals measured from seven different individuals with Vocational School Armband were used in Random Forest, Deep Learning, Gradient Boosted Trees, Decision Tree, Naïve Bayes, Logistic Regression, Generalized Linear Model, Fast Large Margin algorithms and 16796 models were created and Receiver Operating Characteristic (ROC) curve is shared in Figure 1.
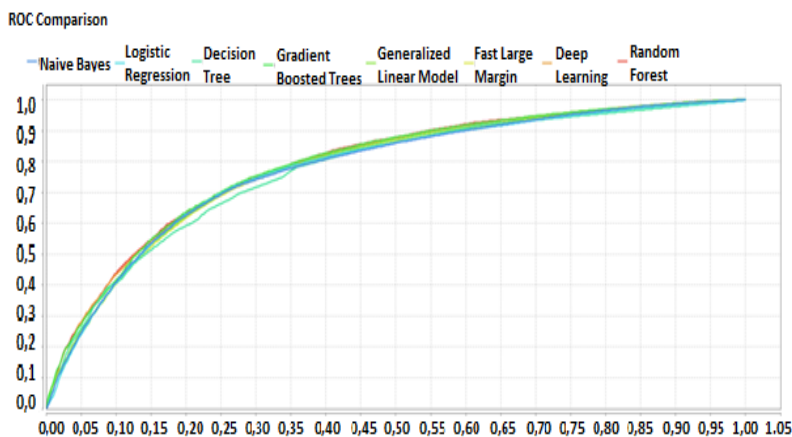

Figure 1: Mechanical model equivalent to the raised water tank

The equations for the Housner (Epstein, 1976) approximations for hydrodynamic pressure are set below [7].

3.1. ACCURACY BASED RESULTS

Accuracy defines the distance between the actual value and the measured value.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

- True Positive (TP): is where the signal is strong and our model correctly predicts as strong signal
- True Negative (TN) is where the signal is weak and our model correctly predicts as weak signal
- False Positive (FP): is where the signal is strong and our model predicts as weak signal
- False Negative (FN): is where the signal is weak and our model predicts as strong signal

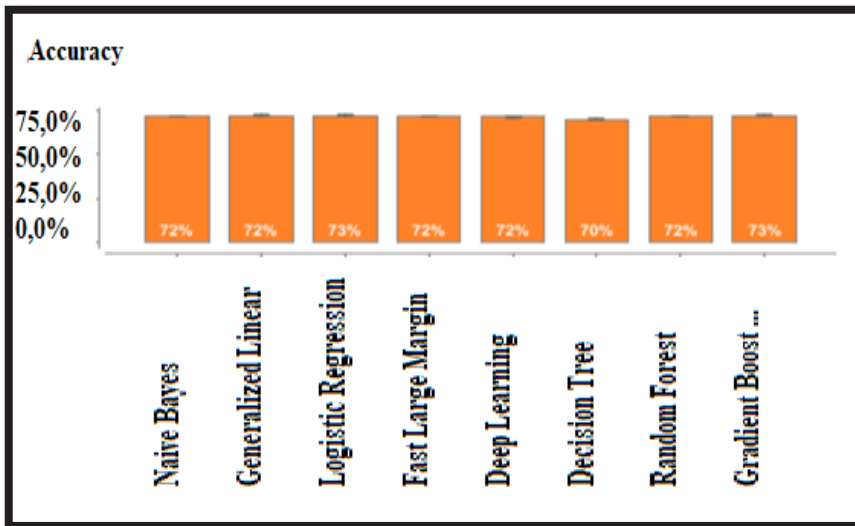After all algorithms worked with our dataset, accuracy based results shared in Figure 2 and Table1.

Figure 1. Accuracy based success rates of all algorithms.

Table 1. Accuracy based results of all algorithms.

Model	Accuracy	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)	Scoring Time (1,000 Rows)
Naïve Bayes	0,7230	0,000974	25382,0	699452,0	1,3	54,1
Generalized Linear Model	0,7239	0,003031	25586,0	696467,0	5,5	73,0
Logistic Regression	0,7255	0,001514	25686,0	514505,0	4,7	56,3
Fast Large Margin	0,7193	0,001400	24998,0	92,6654,0	50,8	56,1
Deep Learning	0,7165	0,001658	24716,0	1115023,0	104,9	106,7
Decision Tree	0,7033	0,002573	23096,0	523277,0	2,0	85,7
Random Forest	0,7212	0,004561	10284,0	1024953,0	11,0	251,4
Gradient Boosted Trees	0,7257	0,005856	10472,0	936112,0	119,5	148,6

3.2. PRECISION BASED RESULTS

Precision defines the distance between the measured values.

$$Precision = \frac{(TP)}{(TP + FP)} \qquad (2)$$

After all algorithms worked with our dataset, precision based results shared in Figure 3 and Table2.
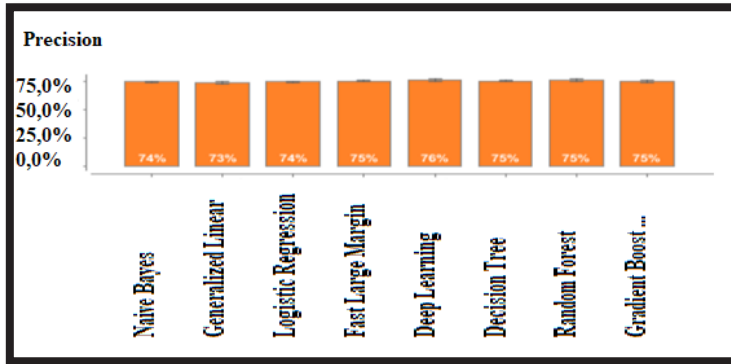

Figure 2. Precision based success rates of all algorithms.

Table 2. Precision based results of all algorithms.

Model	Precision	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)	Scoring Time (1,000 Rows)
Naïve Bayes	0,740990	0,007916	25382,0	699452,0	1,3	54,1
Generalized Linear Model	0,732197	0,008251	25586,0	696467,0	5,5	73,0
Logistic Regression	0,741457	0,004204	25686,0	514505,0	4,7	56,3
Fast Large Margin	0,748666	0,003733	24998,0	926654,0	50,8	56,1
Deep Learning	0,758971	0,008881	24716,0	1115023,0	104,9	106,7
Decision Tree	0,749864	0,005209	23096,0	523277,0	2,0	85,7
Random Forest	0,753676	0,013546	10284,0	1024953,0	11,0	251,4
Gradient Boosted Trees	0,747325	0,014196	10472,0	936112,0	119,5	148,6

3.3. RECALL BASED RESULTS

$$Recall = \frac{(TP)}{(TP + FN)} \tag{3}$$

After all algorithms worked with our dataset, recall based results shared in Figure 4 and Table 3.
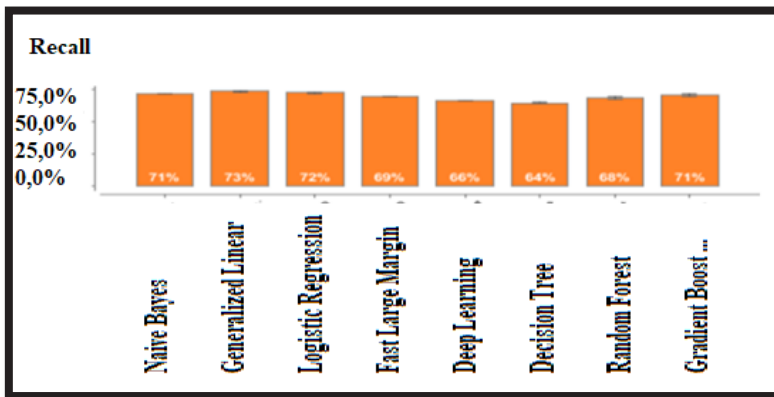

Figure 3. Recall based success rates of all algorithms.

Table 3. Recall based results of all algorithms.

Model	Recall	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)	Scoring Time (1,000 Rows)
Naïve Bayes	0,712751	0,002113	25382,0	699452,0	1,3	54,1
Generalized Linear Model	0,733414	0,004357	25586,0	696467,0	5,5	73,0
Logistic Regression	0,723134	0,002924	25686,0	514505,0	4,7	56,3
Fast Large Margin	0,690873	0,001569	24998,0	926654,0	50,8	56,1
Deep Learning	0,660657	0,003019	24716,0	1115023,0	104,9	106,7
Decision Tree	0,642487	0,003214	23096,0	523277,0	2,0	85,7
Random Forest	0,680806	0,008022	10284,0	1024953,0	11,0	251,4
Gradient Boosted Trees	0,705492	0,008991	10472,0	936112,0	119,5	148,6

3.4. CLASSIFICATION ERROR BASED RESULTS

$$Classification\ Error = \frac{(FP+FN)}{(TP+FP+FN+TN)} \qquad (4)$$

After all algorithms worked with our dataset, classification error-based results shared in Figure 5 and Table 4.
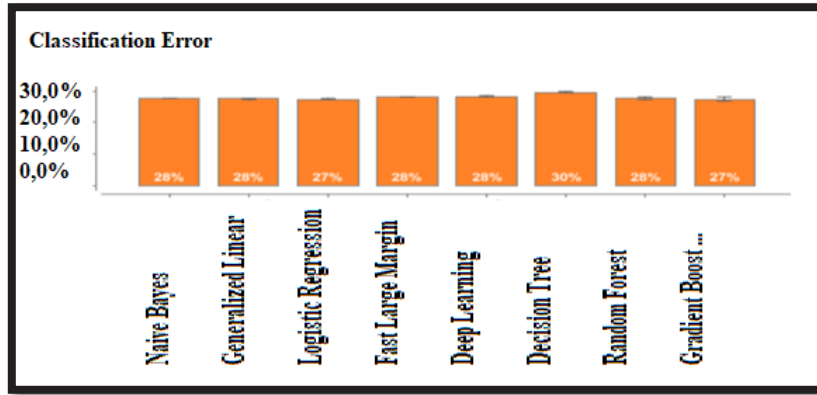

Figure 5. Classification Error based success rates of all algorithms.

Table 4. Classification Error based results of all algorithms.

Model	Classification Error	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)	Scoring Time (1,000 Rows)
Naïve Bayes	0,276901	0,000974	25382,0	699452,0	1,3	54,1
Generalized Linear Model	0,276054	0,003031	25586,0	696467,0	5,5	73,0
Logistic Regression	0,274434	0,001514	25686,0	514505,0	4,7	56,3
Fast Large Margin	0,280686	0,001400	24998,0	926654,0	50,8	56,1
Deep Learning	0,283473	0,001658	24716,0	1115023,0	104,9	106,7
Decision Tree	0,296647	0,002573	23096,0	523277,0	2,0	85,7
Random Forest	0,278770	0,004561	10284,0	1024953,0	11,0	251,4
Gradient Boosted Trees	0,274289	0,005856	10472,0	936112,0	119,5	148,6

3.5. SPECIFICITY BASED RESULTS

$$Specificity = \frac{(TN)}{(TN+FP)} \tag{5}$$

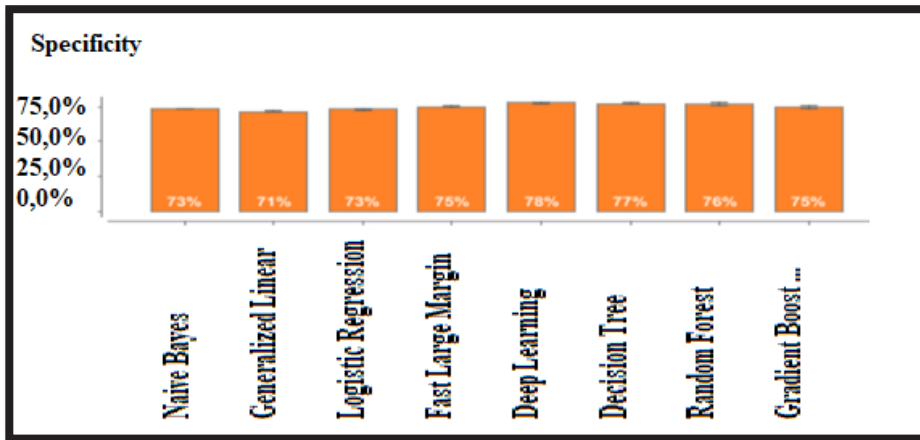After all algorithms worked with our dataset, specificity-based results shared in Figure 6 and Table 5.

Figure 6. Specificity based success rates of all algorithms.

Table 5. Specificity based results of all algorithms.

Model	Specificity	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)	Scoring Time (1,000 Rows)
Naïve Bayes	0,734166	0,003673	25382,0	699452,0	1,3	54,1
Generalized Linear Model	0,713779	0,002292	25586,0	696467,0	5,5	73,0
Logistic Regression	0,728154	0,003146	25686,0	514505,0	4,7	56,3
Fast Large Margin	0,749947	0,003163	24998,0	926654,0	50,8	56,1
Deep Learning	0,776169	0,003125	24716,0	1115023,0	104,9	106,7
Decision Tree	0,768939	0,004299	23096,0	523277,0	2,0	85,7
Random Forest	0,764204	0,011877	10284,0	1024953,0	11,0	251,4
Gradient Boosted Trees	0,747265	0,011986	10472,0	936112,0	119,5	148,6

3.6. F MEASURE BASED RESULTS

$$F\ Measure = \frac{(2*TP)}{(2*TP+FP+FN)} \qquad (6)$$

After all algorithms worked with our dataset, F Measure based results shared in Figure 7 and Table 6.
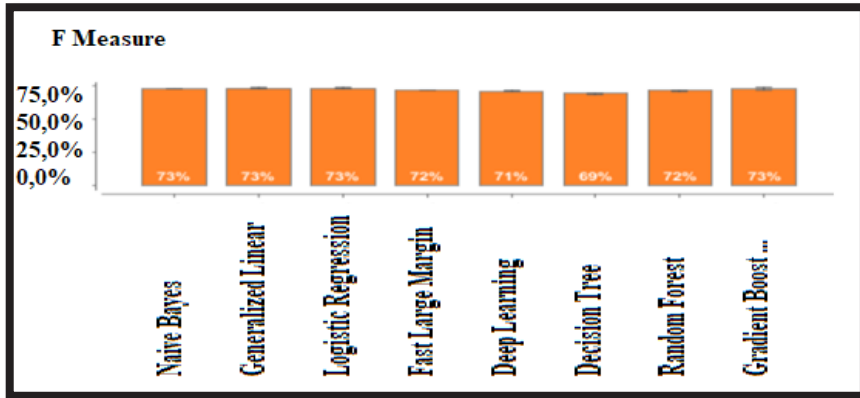

Figure 7. F Measure based success rates of all algorithms.

Table 6. F Measure based results of all algorithms.

Model	F Measure	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)	Scoring Time (1,000 Rows)
Naïve Bayes	0,726577	0,003765	25382,0	699452,0	1,3	54,1
Generalized Linear Model	0,732800	0,006241	25586,0	696467,0	5,5	73,0
Logistic Regression	0,732178	0,003199	25686,0	514505,0	4,7	56,3
Fast Large Margin	0,718608	0,002392	24998,0	926654,0	50,8	56,1
Deep Learning	0,706395	0,005015	24716,0	1115023,0	104,9	106,7
Decision Tree	0,692032	0,003805	23096,0	523277,0	2,0	85,7
Random Forest	0,715304	0,006225	10284,0	1024953,0	11,0	251,4
Gradient Boosted Trees	0,725726	0,007947	10472,0	936112,0	119,5	148,6

3.7. AREA UNDER CURVE (AUC) BASED RESULTS

ROC (Receiver-Operating-Characteristic-Curve) is a graph that calculates the correct positivity on the vertical axis and the false positivity on the horizontal axis, calculated for different threshold values. Area Under Curve is used as a benchmark for superiority of tests. After all algorithms worked with our dataset, Area Under Curve based results shared in Figure 8 and Table 7.
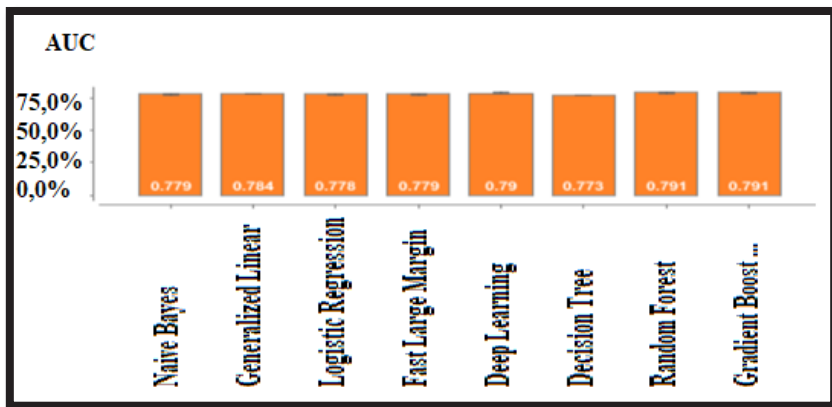

Figure 8. AUC based success rates of all algorithms.

Table 7. AUC based results of all algorithms.

Model	AUC	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)	Scoring Time (1,000 Rows)
Naïve Bayes	0,779166	0,002422	25382,0	699452,0	1,3	54,1
Generalized Linear Model	0,783586	0,002128	25586,0	696467,0	5,5	73,0
Logistic Regression	0,778040	0,001512	25686,0	514505,0	4,7	56,3
Fast Large Margin	0,779310	0,002495	24998,0	926654,0	50,8	56,1
Deep Learning	0,789668	0,002204	24716,0	1115023,0	104,9	106,7
Decision Tree	0,773046	0,001483	23096,0	523277,0	2,0	85,7
Random Forest	0,791139	0,004807	10284,0	1024953,0	11,0	251,4
Gradient Boosted Trees	0,790987	0,004761	10472,0	936112,0	119,5	148,6

4. CONCLUSION

Much artificial intelligence and machine learning methods are being tried to strengthen the signals obtained nowadays. In this study, eight different machine learning algorithms have been tried on the same dataset, and the success rates have been shared. The results obtained in our study are essential in order to predict which machine learning algorithm works with a higher success rate when the data obtained through MYO Armband. The studies to be carried out in this field are of great importance, primarily because they affect the amputee people.

The results obtained in our study can be expressed as follows: Logistic Regression stands out with a success rate of 72.55%. Although the Gradient Boosted Trees algorithm gives almost the same results, it fails when the Training Time and Scoring Time data are examined. When precision-based results are examined, Deep Learning algorithm has come to the forefront with a success rate of 75.89%. When recall based results are examined, the Generalized Linear Model has come to the forefront with a success rate of 73.34%. When sensitivity-based results are examined, the Generalized Linear Model with 73.34% success rate came to the forefront. When the Classification Error based results are examined, Gradient Boosted Trees came to the forefront with a rate of 27.42%. When the results are analyzed based on Specificity, Deep Learning algorithm came to the fore with 77.61%. When F Measure based results are examined, the Generalized Linear Model was successful with a success rate of 73.28%. Finally, when AUC-Based analysis was performed, Random Forest algorithm was used with a success rate of 79.11%.

REFERENCES
[1] Artemiadis, P. K. and Kyriakopoulos, K. J. (2007). EMG-based position and force control of a robot arm: Application to teleoperation and orthosis. IEEE/ASME international conference on advanced intelligent mechatronics, pp. 1-6.

[2] Cichosz, P. (2015). Data mining algorithms: explained using R. United States: John Wiley & Sons.

[3] Erin, K. and Boru, B. (2018). Real time control of industrial robot arm with EMG and gyroscope data. Sakarya University Journal of Science, 22 (2), 509-515.

[4] Fan, R. E., Chang, K. W., Hsieh, C., Wang, X. and Lin, C. (2008). LIBLINEAR: A Library for Large Linear Classification. Journal of Machine Learning Research, 9, 1871-1874.

[5] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232.

[6] Fukuda, O., TSUJI, T., Ohtsuka A. and KANEKO M. (1998). EMG-based Human-Robot Interface for Rehabilitation Aid. Proceedings. IEEE International Conference on Robotics and Automation (Cat. No.98CH36146)*, 4*, 3492-3497.

[7] Li-Zhi, L., Yi-Li, T. and Hsin-Han, C. (2018). EMG-based Control Scheme with SVM Classifier for Assistive Robot Arm. International Automatic Control Conference (CACS), pp.1-5.

[8] Moradi, M., Hashtrudi-Zaad, K., Mountjoy, K. and Morin, E. (2008). An EMG-based force control system for prosthtic arms. Canadian Conference on Electrical and Computer Engineering, pp.1737-1742.

[9] Ramasubramanian, K. and Singh, A. (2017). Machine Learning Using R. *Apress*.

[10] Sammut, C. and Webb, G.I. (2015). Encyclopedia of machine learning and data mining, Springer.

[11] Su, H., Nelder, J. A., Spence, R. and Ismail, M. (1994) Generalized linear models for empirical performance modeling in circuit design. Proceedings of APCCAS'94-Asia Pacific Conference on Circuits and Systems.