Published Ahead of Print on February 6, 2020 as 10.1212/WNL.0000000000009068

ARTICLE CLASS OF EVIDENCE

Accuracy of a machine learning muscle MRI-based tool for the diagnosis of muscular dystrophies

José Verdú-Díaz, Jorge Alonso-Pérez, MD,* Claudia Nuñez-Peralta, MD,* Giorgio Tasca, MD, PhD, John Vissing, MD, PhD, Volker Straub, MD, PhD, Roberto Fernández-Torrón, MD, Jaume Llauger, MD, Isabel Illa, MD, PhD, and Jordi Díaz-Manera, MD, PhD

Correspondence
Dr. Díaz-Manera
jdiazm@santpau.cat

Neurology® 2020;94:1-9. doi:10.1212/WNL.0000000000009068

Abstract

Objective

Genetic diagnosis of muscular dystrophies (MDs) has classically been guided by clinical presentation, muscle biopsy, and muscle MRI data. Muscle MRI suggests diagnosis based on the pattern of muscle fatty replacement. However, patterns overlap between different disorders and knowledge about disease-specific patterns is limited. Our aim was to develop a software-based tool that can recognize muscle MRI patterns and thus aid diagnosis of MDs.

Methods

We collected 976 pelvic and lower limbs T1-weighted muscle MRIs from 10 different MDs. Fatty replacement was quantified using Mercuri score and files containing the numeric data were generated. Random forest supervised machine learning was applied to develop a model useful to identify the correct diagnosis. Two thousand different models were generated and the one with highest accuracy was selected. A new set of 20 MRIs was used to test the accuracy of the model, and the results were compared with diagnoses proposed by 4 specialists in the field.

Results

A total of 976 lower limbs MRIs from 10 different MDs were used. The best model obtained had 95.7% accuracy, with 92.1% sensitivity and 99.4% specificity. When compared with experts on the field, the diagnostic accuracy of the model generated was significantly higher in a new set of 20 MRIs.

Conclusion

Machine learning can help doctors in the diagnosis of muscle dystrophies by analyzing patterns of muscle fatty replacement in muscle MRI. This tool can be helpful in daily clinics and in the interpretation of the results of next-generation sequencing tests.

Classification of evidence

This study provides Class II evidence that a muscle MRI-based artificial intelligence tool accurately diagnoses muscular dystrophies.

RELATED ARTICLE

Machine learning outperforms human experts in MRI pattern analysis of muscular dystrophies
Page XXX

MORE ONLINE

→ **Class of Evidence**
Criteria for rating therapeutic and diagnostic studies
NPub.org/coe

*These authors contributed equally to this work.

From the Neuromuscular Disorders Unit, Neurology Department (J.V.-D., J.A.-P., I.I., J.D.-M.), and Radiology Department (C.N.-P., J.L.), Hospital de la Santa Creu I Sant Pau, Barcelona, Spain; UOC di Neurologia (G.T.), Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy; Copenhagen Neuromuscular Center, Department of Neurology (J.V.), Rigshospitalet, University of Copenhagen, Denmark; John Walton Muscular Dystrophy Research Centre (V.S., J.D.-M.), University of Newcastle, Newcastle Upon Tyne, UK; Hospital Universitario Donostia (R.F.-T.); and Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER) (I.I., J.D.-M.), Madrid, Spain.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Glossary

SHD = facioscapulohumeral muscular dystrophy; **MD** = muscular dystrophy; **NGS** = next-generation sequencing; **NPV** = negative predictive value; **OPMD** = oculopharyngeal muscular dystrophy; **PPV** = positive predictive value.

The classification of muscular dystrophies (MDs) is based on the mutated gene that causes the disease and encompasses 32 disorders.[1–3] Some years ago, clinical data were used to guide Sanger sequencing of individual genes to diagnose patients.[4] Next-generation sequencing (NGS) has increased the diagnostic efficiency substantially in MD, and depending on the cohort, up to 50% of the patients are diagnosed.[5,6] However, NGS has some limitations: the candidate gene sometimes does not fit with the phenotype, potential disease-causing variants in more than one gene can be found, or variants of unknown effect can be identified.[7,8]In all these situations, clinical data and results of complementary tests continue to be of great value to make sense of the results obtained.[9,10] Muscle MRI identifies characteristic patterns of muscle involvement related with specific disorders.[11–23] However, most of these patterns are overlapping, and a high degree of specialization is needed to differentiate one disease from another based on MRI data.

Artificial intelligence is implemented in the decision-making process in medicine in several fields.[24–27] Machine learning uses algorithms to analyze data, learn from them, and then make a determination or predictions.[28] We have applied a machine learning strategy to a large dataset of muscle MRIs performed in patients with genetically confirmed diagnoses of MDs, with the aim to develop an informatic tool that can help in the diagnostic process of these disorders.

Methods

Primary research question

The primary research question was whether machine learning is able to correctly suggest diagnosis of MDs based on muscle MRI of the pelvis, thighs, and leg muscles.

Patients

In this study, we collected 986 muscle MRIs of the pelvic, thigh, and leg muscles of patients with the following genetically confirmed MDs: dystrophinopathies (both Duchenne and Becker MD, n = 46 patients), LGMDR1 (*CAPN3* gene, n = 73), LGMDR2 (*DYSF* gene, n = 181), LGMDR3–6 (sarcoglycanopathies, n = 69), LGMDR9 (*FKRP* gene, n = 40), LGMDR12 (*ANO5* gene, n = 28), MDs caused by mutations in the *LMNA* gene (n = 41), facioscapulohumeral MD (FSHD, n = 269), oculopharyngeal MD (OPMD, n = 171), and Pompe disease (n = 68). Ten MRIs were excluded because they were normal and did not show any fat replacement. Therefore, a total of 976 MRIs were included. Patients included had a wide range of disease severity, including presymptomatic, symptomatic ambulant, and nonambulant patients. Most of the muscle MRIs included in this study also have been used for other studies; however, we added more MRIs from our own database.[11–13,29–34] Patients signed an informed consent and this work was reviewed by the Ethics Committee of the Hospital de la Santa Creu i Sant Pau in Barcelona.

Muscle MRI analysis

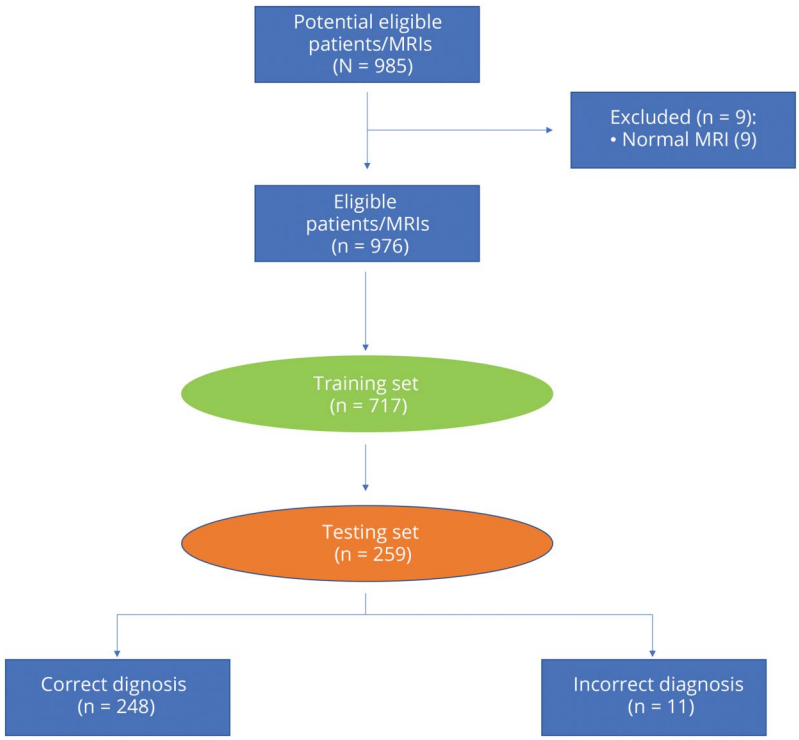
We recollected a total of 986 muscle MRIs (figures 1). Axial T1-weighted images from pelvic, thigh, and leg muscles were analyzed. Fat replacement was quantified using the Mercuri score modified by Fischer et al.[35] as previously reported (data available at datadryad.org/review?doi=doi:10.5061/dryad.7tb88vs). Normal muscle MRIs (n = 10), in which there was no evidence of fat replacement, were removed from the analysis. A final database was generated containing the score for 35 muscles of the right and leg sides of 976 muscle MRIs. This database could potentially contain 68,320 variables for the analysis (70 muscles × 976 patients). However, we were not able to obtain information for 6,475 muscles, because the muscle MRIs were focused on pelvic and thighs or in thighs and legs in some cases. A complete list of the muscles analyzed can be found in the supplemental material.

Random forest

We used random forest machine learning strategy for the analysis of the data to find a model able to distinguish among the different disorders. Random forest is a machine learning tool capable of fitting large datasets and performing both classification and regression tasks.[36] To generate the model, we split the data into 2 groups: one for training the model and the other for validating the model. We analyzed how the accuracy of the models was influenced by the size of the dataset used for training and for validation. This allowed us to obtain better accuracy in the prediction models with low dispersion of the results (figure 2). Based on this function, we used 70% of the MRI to train the model, and 30% of the data to validate it. The software randomly separated the set of MRIs in 2 groups, without any external supervision. We only asked the software to maintain the same proportion of MRIs of a certain diagnosis in the training and testing sets. Forest of 500 trees were grown, and in every split, 8 different variables were sampled. Missing data were imputed using the *rfImpute* function described by Ishwaran et al.[37]

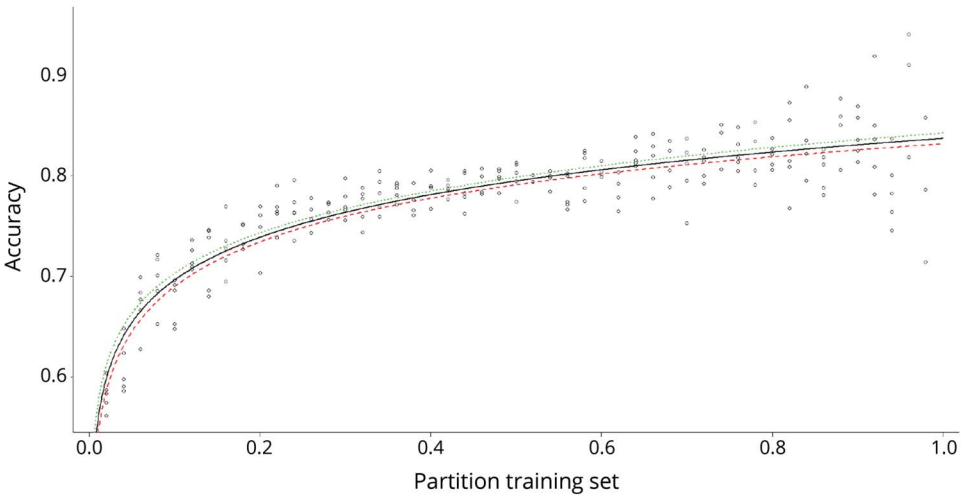
The model provided a list of potential diagnoses for every case, ordering them from the most probable to the least probable. We considered that the model was correct if the first proposed diagnosis was the correct one. For every model generated, we obtained the accuracy, calculated as the

Figure 1 STARD (Standards for Reporting of Diagnostic Accuracy) diagram showing the flow of MRIs throughout the study



percentage of cases predicted successfully. The computer subsequently generated 2,000 models and selected the one with the best accuracy obtained. Once we obtained the best model, we calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for every disease. The global sensitivity, specificity, PPV, and NPV of the model is a mean of all the values obtained in every disease. We studied the intramodel agreement using a test–retest approach, repeating the prediction process with the best mode, and Cohen κ index was obtained. The predictive importance of each muscle in the final model was estimated using Breiman permutation variable importance. The model was generated using R software version 3.5.3 (r-project.org) with the following packages: *randomForest*, *dplyr*, *DT*, *fmsb*, *caret*, *ggplot2*, *lattice*, *stats*, *graphics*, *grDevices*, *utils*, *datasets*, *methods*, and *base*.

Figure 2 Influence of the data partition in the accuracy obtained by the model



Larger training sets produced higher accuracy but were also associated with a higher dispersion of the results obtained. Mean accuracy and SD per every training set size is provided. X axis shows training set size expressed as a percentage of the total size. Y axis shows accuracy results as percentage of correct diagnosis out of the total number of MRIs analyzed.

External validation of the model

We compared the accuracy of the model to predict the diagnosis of muscle dystrophies with experts in the field. To do so, the informatic tool and 4 experts analyzed a new set of 20 MRIs with a confirmed genetic diagnosis of an MD. When the experts analyzed the new set of MRIs, they knew the different diagnostic options to choose but they were blind to the results of the software. These MRIs were not included in the training or the validating set of MRIs used to develop the model. We asked the experts and the software to propose 3 potential diagnoses for every MRI analyzed. If the first option proposed was the correct diagnosis, we assigned 3 points. If the correct diagnosis was the second option proposed, we assigned 2 points, and if it was the third, we assigned 1 point. A final total score was generated for the experts and for the software.

Data availability

All supplementary data are available at doi.org/10.5061/dryad.7tb88vs. Further anonymized data can be made available to qualified investigators upon request to the corresponding author.

Results

Patients and MRIs

We included in the analysis 976 MRIs from the pelvic, thigh, and leg muscles of patients with different MDs. All patients had a confirmed genetic diagnosis of the disease. We analyzed muscle MRIs using Mercuri score modified by Fischer et al.,[35] a total of 70 muscles including both left and right sides (supplemental material, doi.org/10.5061/dryad.7tb88vs). Table e-1 (doi.org/10.5061/dryad.7tb88vs) displays a group of descriptive statistics of the muscles in the model.

The model

We ran 2,000 different random forest models and identified the one with highest diagnostic accuracy rate. The final model selected had a diagnostic accuracy of 95.7% (confidence interval 92.5%–97.8%) with an overall sensitivity of 92.1%, specificity of 99.4%, PPV of 98.06%, and NPV of 99.53%. Internal validation of the model was analyzed using a test–retest strategy obtaining a κ index of 99%. Table 1 describes the mean features of the model for every disease studied.

The testing set of MRIs of the final model contained 259 MRIs of the following diseases: 52 dysferlinopathy, 45 OPMD, 10 *LMNA*, 16 Pompe, 71 FSHD, 17 sarcoglycanopathies, 21 calpainopathy, 9 dystrophinopathy, 7 *ANO5*, and 10 *FKRP*. The predictions made by the testing set are displayed in table 2.

One of the disadvantages of random forest is that it is difficult to understand how the computer is constructing the model. In order to understand which muscles were the most important in the development of the model, we studied the impact of every muscle in what is called node impurity, also known as Gini index. This measure provides information about the importance of every muscle in the final model obtained. Figure 3 shows the mean decrease in node impurity for every muscle studied. The higher the value, the more relevant the muscle in the model. The muscles with a higher impact in the final model generated were the iliacus, the psoas, the tensor fasciae latae, and the gluteus medius. The muscles that contributed less to the final model generated were the popliteus, sartorius, flexor digitorum, and tibialis posterior.

Testing the model: comparison between experts and the software

In order to know if the model generated had higher diagnostic accuracy than human experts, we asked 4 experts in the field to analyze a new set of 20 MRIs from patients with different MDs. We asked the experts and the software to propose 3 potential diagnoses for every MRI analyzed. A final score was obtained depending on the accuracy of the diagnosis as described in the Methods. The software obtained a final score of 55 out of 60 points, while the experts obtained 42, 41, 38, and 31 out of 60 points, respectively (figure 4).

Studying pairs of diseases: resolution of common troubles

Although muscle MRI is useful to distinguish among different muscle disorders, there are diseases that share a similar pattern of muscle involvement, and the distinction between them can be complicated.[7,9] This is the case for LGMDR2 (mutations in *DYSF*) and LGMDR12 (mutations in *ANO5*) or for LGMDR1 (mutations in *CAPN3*) and LGMDR9 (mutations in *FKRP*).[38] We studied the ability of the model generated to distinguish between pairs of diseases,

Table 1 Main features of the model developed by every disease studied

	Dystrophin	*LMNA*	Calpain	Dysferlin	Sarcoglycans	*FKRP*	*ANO5*	Pompe	FSHD	OPMD	Mean value
Accuracy	0.94	0.95	1	0.97	0.91	0.92	0.92	1	0.98	0.97	0.95
Sensitivity	0.88	0.90	1	0.96	0.83	0.8	0.85	1	1	0.97	0.91
Specificity	1	1	1	0.99	1	1	1	1	0.97	0.97	0.99
PPV	1	1	1	0.96	1	1	1	1	0.94	0.89	0.97
NPV	0.99	0.99	1	0.99	0.98	0.99	0.99	1	1	0.99	0.99

Abbreviations: FSHD = facioscapulohumeral muscular dystrophy; NPV = negative predictive value; OPMD = oculopharyngeal muscular dystrophy; PPV = positive predictive value.

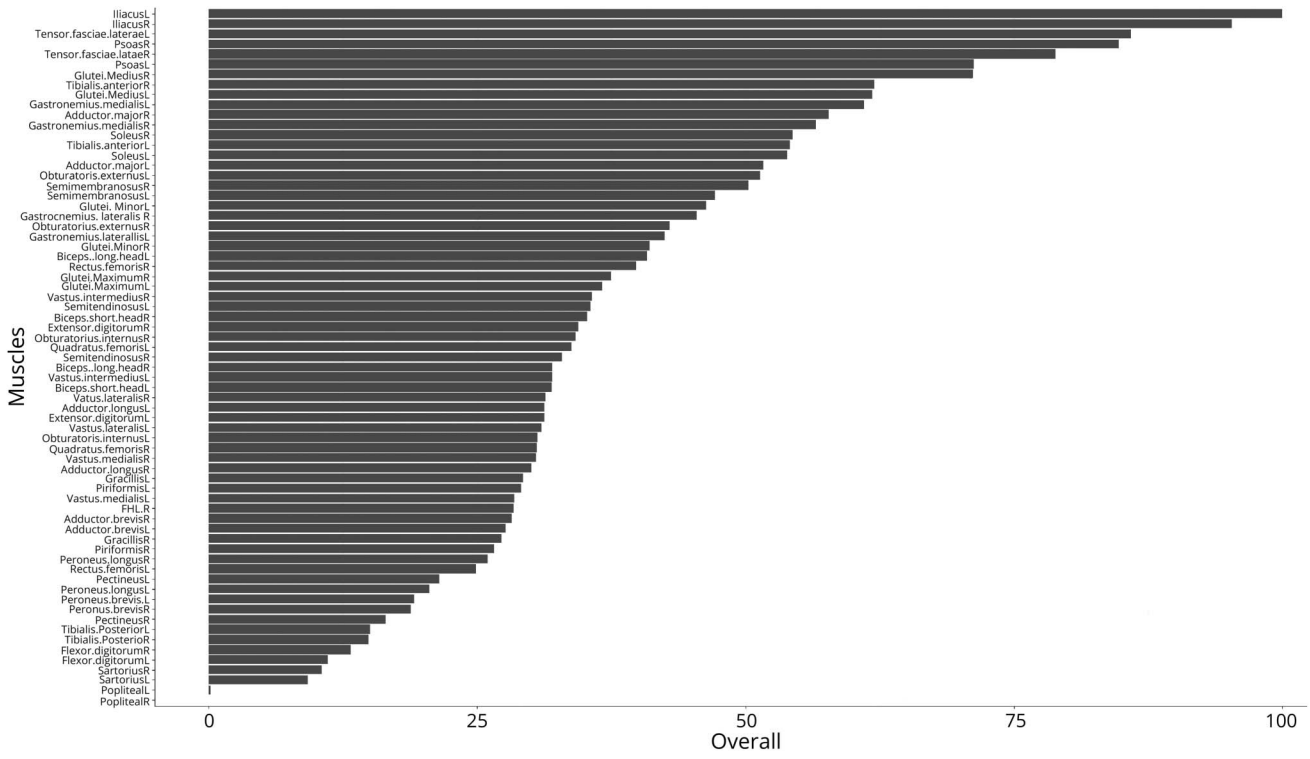
Table 2 Muscle MRIs contained in the testing set of the final model

	Diagnosis									
Prediction	**Dysferlin**	**OPMD**	***LMNA***	**Pompe**	**FSHD**	**Sarcoglycanopathy**	**Calpain**	**Dystrophin**	***ANO5***	***FKRP***
Dysferlin	50	0	0	0	0	1	0	0	1	0
OPMD	0	44	1	0	0	1	0	1	0	2
LMNA	0	0	9	0	0	0	0	0	0	0
Pompe	0	0	0	16	0	0	0	0	0	0
FSHD	2	1	0	0	71	1	0	0	0	0
Sarcoglycanopathy	0	0	0	0	0	15	0	0	0	0
Calpain	0	0	0	0	0	0	21	0	0	0
Dystrophin	0	0	0	0	0	0	0	8	0	0
ANO5	0	0	0	0	0	0	0	0	6	0
FKRP	0	0	0	0	0	0	0	0	0	8

Abbreviations: ANO5 = anoctaminopathy; FKRP = Fukutin-related protein; FSHD = facioscapulohumeral muscle dystrophy; LMNA = laminopathy; OPMD = oculopharyngeal muscle dystrophy.
The table provides the prediction and diagnosis data of the cases contained in the testing set.

including those classically considered problematic, such as LGMDR1 (*CAPN3*) vs LGMDR9 (*FKRP*), LGMDR2 (*DYSF*) vs LGMDR12 (*ANO5*), and LGMDR2 (*DYSF*) vs dystrophinopathies. The model was able to distinguish between all pairs of diseases with an accuracy higher than 90%. Moreover, we decided to study which muscles differentiated one disease from the other analyzing the impact in the Gini index, as described above. Figure 5 displays the

Figure 3 Relative importance of every muscle analyzed



The weight of every muscle in the development of the final model by our software is displayed. Muscles are listed from the most relevant to the least relevant for the development of the model. Bars represent the impact on the Gini index if that specific muscle is removed from the model, ranging from 0 to 100.

Figure 4 Suggested diagnosis for every case analyzed of 20 sets of testing MRIs

Case	Real diagnosis	Suggested diagnosis				
		Software	Expert 1	Expert 2	Expert 3	Expert 4
1	OPMD	OPMD	OPMD	OPMD	FKRP/dysferlin/calpain	OPMD
2	Sarco.	Sarco.	Sarco.	Sarco.	Pompe/sarco.	Sarco.
3	Calpain	Calpain	Calpain	Calpain	Calpain	Calpain
4	FSHD	FSHD	FSHD	FSHD	Dysferlin/FSHD	FSHD
5	Dystrophin	Dystrophin	LMNA/sarco./Pompe	Sarco./calpain/FKRP	Sarco./Pompe	Sarco.
6	Dystrophin	Dystrophin	FKRP/calpain/sarco.	Calpain/dysferlin/FKRP	Dystrophin	Pompe/calpain/ano5
7	ANO5	ANO5	ANO5	ANO5	Dysferlin/ANO5	FSHD/ANO5
8	Dysferlin	FSHD/dysferlin	Dysferlin	Dysferlin	ANO5	Dysferlin
9	Dysferlin	Dysferlin	Dysferlin	LMNA/FSHD/calpain	OPMD	OPMD/dysferlin
10	Pompe	Pompe	Pompe	Pompe	Sarcoglycan/Pompe	Calpain/FKRP/ANO5
11	Sarco.	Sarco.	Sarco.	Sarco.	Pompe/sarco.	Sarco.
12	LMNA	LMNA	FSHD/ANO5	LMNA	LMNA	LMNA
13	Pompe	Pompe	Pompe	Pompe	Pompe	Calpain/ANO5
14	Calpain	ANO5/calpain	FKRP/sarco./calpain	Dystrophin/dysferlin/FKRP	FKRP/calpain	OPMD/LMNA
15	OPMD	OPMD	FKRP	Dystrophin/dysferlin/LMNA	Dystrophin/OPMD	OPMD
16	FSHD	FSHD	FSHD	FSHD	FSHD	FSHD
17	Dystrophin	Dysferlin/LMNA/calpain	FKRP	Dystrophin	Dysferlin/calpain/dystrophin	Dysferlin/ANO5
18	Dysferlin	Dysferlin	Dysferlin	ANO5/dysferlin	ANO5/dysferlin	FSHD/ANO5
19	LMNA	LMNA	Dysferlin/LMNA	LMNA	LMNA	Dysferlin/FKRP
20	Calpain	Calpain	Calpain	Dystrophin/dysferlin/FKRP	Calpain	LMNA/OPMD
Score		**55**	**42**	**41**	**38**	**31**

The table displays the real diagnosis of the patient confirmed by genetic study and the diagnosis suggested by the machine learning software and the 4 experts as described in the Methods. Cells are green if the first diagnosis provided was correct (3 points), blue if the second diagnosis was correct (2 points), yellow if the third diagnosis was correct (1 point), and red if none of the diagnoses provided was correct (0 points). ANO5 = anoctamin 5; FSHD = facioscapulohumeral muscular dystrophy; LMNA = lamin A/C; OPMD = oculopharyngeal muscular dystrophy; Sarco = sarcoglycanopathy.

more important muscles to distinguish among diseases in the comparisons made.

Discussion

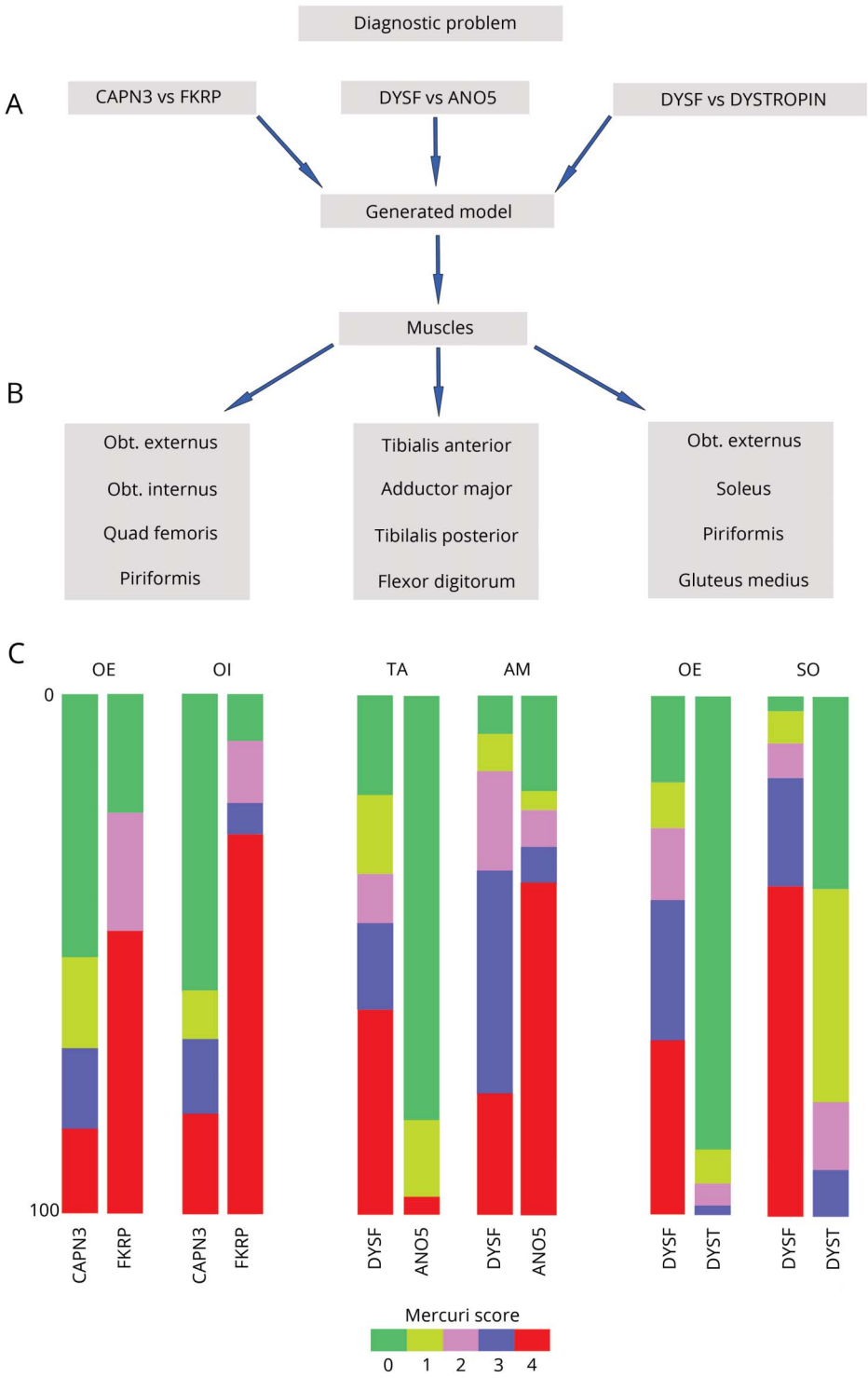
We successfully applied a machine learning strategy, using random forest, to a set of 976 muscle MRIs obtained from patients with already genetically confirmed diagnosis of MD. The model generated has a high accuracy, with high sensitivity, specificity, PPV, and NPV. This model may be of help in the diagnostic process of patients with MDs, facilitating the selection of genes to be analyzed in panels or exomes or in the verification candidate gene variants of unknown significance. Moreover, our work is an example of how the application of artificial intelligence to a large group of data from hundreds of patients can generate models than can be helpful in the diagnostic process of genetic disorders.

The diagnosis of MDs has been classically based on the identification of specific clues from the clinical history, physical examination, or muscle biopsy that have guided the selection of genes to be analyzed using Sanger sequencing.[7,39,40] Muscle MRI can also be helpful in this sense. In recent years, several patterns of muscle involvement have been described that, if identified, can help guide the diagnosis. However, it has been suggested that these MRI-based diagnostic criteria proposed for every disease are not useful in daily clinics, even if MRIs are analyzed by specialists in the field.[41] We demonstrate that a machine-learning strategy can be helpful in identifying a characteristic pattern that can guide genetic testing.

NGS has been a revolution in the field, allowing the analysis of hundreds of genes at the same time, shortening the diagnostic delay and improving the accuracy of the diagnosis.[14,16] However, NGS has several limitations: (1) NGS does not cover all exons of all genes, and therefore the accuracy is not 100%; (2) pathogenic variants can be found in genes that cannot explain the phenotype and there can be some doubts about the diagnosis; (3) variants in more than one gene can be found; and (4) the genetic variants identified can be of unknown significance.[42–44] In all these situations, NGS can suggest a potential diagnosis, but other tests are required. Moreover, NGS is only available in some centers, hence the diagnosis in many other centers continues to be based on the selection of specific genes to be studied using Sanger sequencing. The model that we have developed here may be of help in all these situations. For example, if a diagnosis of calpainopathy is suggested by the model, geneticists can examine the variants found in the *CAPN3* gene in detail. If variants of unknown effect are found in a specific gene—for example, the *DYSF* gene—the model can predict if the MRI is compatible with that specific diagnosis. Moreover, for those centers where NGS is not available, the use of this informatic tool can help guide the selection of genes to be tested using Sanger sequencing.

One of the limitations of machine learning strategies, such as random forest, is that it can be difficult to understand the rationale behind the model.[45,46] We have analyzed which muscles were more important in the generation of the final model by analyzing the impact of every muscle in the Gini index. Doing so, we have obtained a list of muscles relevant for the development of the model. It is noteworthy that many of

Figure 5 Analysis of pairs of diseases using the generated model



The model obtained was able to distinguish with high efficiency between diseases classically considered problematic, as they share a similar pattern of involvement (A). The 4 muscles more relevant to distinguish between every pair are listed in the boxes in (B). We analyzed the distribution of Mercuri score for the 2 most relevant muscles between the diseases compared, from 0% to 100% (C). AM = adductor major; OE = obturator externus; OI = obturator internus; SO = soleus; TA = tibialis anterior.

the muscles that were especially relevant for the model are pelvic muscles, which can be missed if muscle MRI does not include this region. Moreover, previous published studies describing characteristic patterns for specific diseases have for the most part not included this region in the analysis. We recommend scanning the pelvic area whenever possible both for diagnostic and for research studies. Interestingly, the muscles less important for the model are usually not involved in MDs, such as the popliteus, sartorius, tibialis posterior, or flexor digitorum. Moreover, we have also analyzed pairs of diseases and have repeated the same strategy identifying which muscles can provide a clue to distinguish between the diseases compared. This provides the medical community with a set of muscles to assess on MRI in order to reach a likely diagnosis.

Among the limitations of our study, one of the most important is the number of diseases included. We included 10 different MDs that are commonly found in neuromuscular disease units. We included these diseases because we had larger datasets of MRIs available for analysis. However, we have not included other potential diagnoses, such as myofibrillar myopathies or congenital muscle disorders, because we do not have large numbers of MRIs from these patient groups. We do not know what number of MRIs from patients with a specific disease is needed to obtain a model with high accuracy. This can be influenced by many different factors; most importantly, the specificity of the pattern. This is the case for Pompe disease or sarcoglycanopathies. In these diseases, the pattern is highly characteristic and difficult to confound with other MDs owing to the absence of fat replacement of distal muscles of the limbs, which is a common finding in the remaining MDs. Therefore, it is not difficult to distinguish Pompe disease and sarcoglycanopathies from the other MDs.[30,34,47] As a result, the number of patients with sarcoglycanopathy needed for developing a model with high accuracy is probably lower than for other MDs. On the other hand, the patterns described in FSHD are variable, and the number of patients needed to be included likely should be higher. Our aim for the coming years is to improve the informatic tool, including other diseases that can widen the diagnostic possibilities. MDs are rare diseases, and for some of them, there are few patients diagnosed. It seems difficult to foresee a model able to differentiate among all potential genetic myopathies, unless the number of MRIs available from these diseases increases in coming years. Another limitation of this study is the need to quantify muscle fat replacement in an index case in every muscle to run the analysis, which can be time-consuming. There is no software available that automatically quantifies the amount of fat present in every muscle.

Our intention is to give the medical community free access to the tool we have generated by hosting it on a public webpage. The main idea of the project is to allow the software to self-feed from the new MRIs consulted when the final diagnosis is confirmed. Moreover, we want to increase the number of diseases in the protocol, including other relatively common muscle disorders such as myotilinopathy or titinopathy by establishing collaboration among specialized centers. It is important to understand that the tool only predicts diagnosis, which should be confirmed by genetic testing.

We have successfully applied a machine learning strategy to a large dataset obtained from muscle MRIs of 10 different muscle disorders. The model generated was able to distinguish among disorders with higher accuracy than human experts in the field. This tool can be of great help in the diagnostic process of MDs, providing potential genes to be tested or reinforcing the pathogenic role of a mutation found. This study can be considered as a proof of concept that demonstrates that artificial intelligence can be applied to the field of muscle MRI. We need to implement the tool by adding more diseases.

Author contributions

Jose Verdú-Díaz: data acquisition, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and final approval, acquisition of data, statistical analysis. Jorge Alonso-Pérez: data acquisition, analysis or interpretation of data, accepts responsibility for conduct of research and final approval, acquisition of data. Claudia Nuñez-Peralta: data acquisition, accepts responsibility for conduct of research and final approval, acquisition of data. Giorgio Tasca: data acquisition, analysis or interpretation of data, accepts responsibility for conduct of research and final approval. John Vissing: drafting/revising the manuscript, data acquisition, analysis or interpretation of data, accepts responsibility for conduct of research and final approval, acquisition of data. Volker Straub: drafting/revising the manuscript, study concept or design, accepts responsibility for conduct of research and final approval, study supervision, obtaining funding. Roberto Fernández-Torrón: drafting/revising the manuscript, data acquisition, accepts responsibility for conduct of research and final approval, acquisition of data. Jaume Llauger: data acquisition, accepts responsibility for conduct of research and final approval, acquisition of data. Isabel Illa: drafting/revising the manuscript, study concept or design, accepts responsibility for conduct of research and final approval, obtaining funding. Jordi Diaz-Manera: drafting/revising the manuscript, data acquisition, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and final approval, acquisition of data, statistical analysis, study supervision, obtaining funding.

Study funding

This investigation was sponsored by a grant from the Spanish Ministry of Health, Fondos FEDER-ISCIII PI18/01525, to Dr. Jordi Díaz-Manera.

Disclosure

The authors report no disclosures relevant to the manuscript. Go to Neurology.org/N for full disclosures.

Publication history

Received by *Neurology* May 16, 2019. Accepted in final form October 3, 2019.

References

1. Bushby K. Diagnosis and management of the limb girdle muscular dystrophies. Pract Neurol 2009;9:314–323.
2. Straub V, Murphy A, Udd B. 229th ENMC international workshop: limb girdle muscular dystrophies: nomenclature and reformed classification: Naarden, the Netherlands, 17–19 March 2017. Neuromuscul Disord 2018;28:702–710.
3. Khadilkar SV, Patel BA, Lalkaka JA. Making sense of the clinical spectrum of limb girdle muscular dystrophies. Pract Neurol 2018;18:201–210.
4. Harris E, Topf A, Barresi R, et al. Exome sequences versus sequential gene testing in the UK Highly Specialised Service for Limb Girdle Muscular Dystrophy. Orphanet J Rare Dis 2017;12:1–12.
5. Savarese M, Fruscio GD, Magri F, et al. The genetic basis of undiagnosed muscular dystrophies and myopathies: results from 504 patients. Neurology 2018;90:1084.
6. Nallamilli BRR, Chakravorty S, Kesari A, et al. Genetic landscape and novel disease mechanisms from a large LGMD cohort of 4656 patients. Ann Clin Transl Neurol 2018;5:1574–1587.
7. Nigro V, Savarese M. Next-generation sequencing approaches for the diagnosis of skeletal muscle disorders. Curr Opin Neurol 2016;29:621–627.

8. Ghaoui R, Cooper ST, Lek M, et al. Use of whole-exome sequencing for diagnosis of limb-girdle muscular dystrophy: outcomes and lessons learned. JAMA Neurol 2015; 72:1424–1432.
9. Schuelke M, Oien NC, Oldfors A. Myopathology in the times of modern genetics. Neuropathol Appl Neurobiol 2017;43:44–61.
10. Puusepp S, Reinson K, Pajusalu S, et al. Effectiveness of whole exome sequencing in unsolved patients with a clinical suspicion of a mitochondrial disorder in Estonia. Mol Genet Metab Rep 2018;15:80–89.
11. Diaz-Manera J, Llauger J, Gallardo E, Illa I. Muscle MRI in muscular dystrophies. Acta Myol 2015;34:95–108.
12. Bugiardini E, Morrow JM, Shah S, et al. The diagnostic value of MRI pattern recognition in distal myopathies. Front Neurol 2018;9:1–11.
13. Weber MA, Wolf M, Wattjes MP. Imaging patterns of muscle atrophy. Semin Musculoskelet Radiol 2018;22:299–306.
14. Wattjes MP, Kley RA, Fischer D. Neuromuscular imaging in inherited muscle diseases. Eur Radiol 2010;20:2447–2460.
15. Tasca G, Iannaccone E, Monforte M, et al. Muscle MRI in Becker muscular dystrophy. Neuromuscul Disord 2012;22(suppl 2):S100–S106.
16. Diaz-Manera J, Fernandez-Torron R, Llauger J, et al. Muscle MRI in patients with dysferlinopathy: pattern recognition and implications for clinical trials. J Neurol Neurosurg Psychiatry 2018;89:1071–1081.
17. Alonso-Jimenez A, Kroon RHMJM, Alejaldre-Monforte A, et al. Muscle MRI in a large cohort of patients with oculopharyngeal muscular dystrophy. J Neurol Neurosurg Psychiatry 2018;90:576–585.
18. Olive M, Goldfard LG, Shatunov A, Fischer D, Ferrer I. Myotilinopathy: refining the clinical and myopathological phenotype. Brain 2005;128:2315–2326.
19. Olive M, Odgerel Z, Martínez A, Poza JJ, Bragado FG, Zabalza RJ. Clinical and myopathological evaluation of early- and late-onset subtypes of myofibrillar myopathy. Neuromuscul Disord 2011;21:533–542.
20. Díaz J, Woudt L, Suazo L, et al. Broadening the imaging phenotype of dysferlinopathy at different disease stages. Muscle Nerve 2016;54:203–210.
21. Mercuri E, Clemens E, Offiah A, et al. Muscle magnetic resonance imaging involvement in muscular dystrophies with rigidity of the spine. Ann Neurol 2010;67:201–208.
22. Silva AMS, Coimbra-Neto AR, Souza PVS, et al. Clinical and molecular findings in a cohort of ANO5-related myopathy. Ann Clin Transl Neurol 2019;6:1225–1238.
23. Sandell SM, Mahjneh I, Palmio J, Tasca G, Ricci E, Udd BA. "Pathognomonic" muscle imaging findings in DNAJB6 mutated LGMD1D. Eur J Neurol 2013;20:1553–1639.
24. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019; 380:1347–1358.
25. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56.
26. Bloomingdale P, Mager DE. Machine learning models for the prediction of chemotherapy-induced peripheral neuropathy. Pharm Res 2019;36:35.
27. Ferroni P, Zanzotto FM, Riondino S, Scarpato N, Guadagni F, Roselli M. Breast cancer prognosis using a machine learning approach. Cancers 2019;11:328.
28. Obermeyer Z, Lee TH. Lost in thought: the limits of the human mind and the future of medicine. N Engl J Med 2017;377:1209–1211.
29. Sarkozy A, Deschauer M, Carlier R-Y, et al. Muscle MRI findings in limb girdle muscular dystrophy type 2L. Neuromuscul Disord 2012;22(suppl 2):S122–S129.
30. Figueroa-Bonaparte S, Segovia S, Llauger J, et al. Muscle MRI findings in childhood/adult onset Pompe disease correlate with muscle function. PLoS One 2016;11: e0163493.
31. Tasca G, Monforte M, Ottaviani P, et al. Magnetic resonance imaging in a large cohort of facioscapulohumeral muscular dystrophy patients: pattern refinement and implications for clinical trials. Ann Neurol 2016;79:854–864.
32. Willis TA, Hollingsworth KG, Coombs A, et al. Quantitative magnetic resonance imaging in limb-girdle muscular dystrophy 2I: a multinational cross-sectional study. PLoS One 2014;9:e90377.
33. Díaz-Manera J, Alejaldre A, González L, et al. Muscle imaging in muscle dystrophies produced by mutations in the EMD and LMNA genes. Neuromuscul Disord 2016;26: 33–40.
34. Tasca G, Monforte M, Díaz-Manera J, et al. MRI in sarcoglycanopathies: a large international cohort study. J Neurol Neurosurg Psychiatry 2018;89:1–6.
35. Fischer D, Kley RA, Strach K, et al. Distinct muscle imaging patterns in myofibrillar myopathies. Neurology 2008;71:758–765.
36. Gómez-Andrés D, Díaz J, Munell F, et al. Disease duration and disability in dysferlinopathy can be described by muscle imaging using heatmaps and random forests. Muscle Nerve 2019;59:436–444.
37. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. J Am Stat Assoc 2010;105:205–217.
38. Ten Dam L, van der Kooi AL, Rövekamp F, Linssen WH, de Visser M. Comparing clinical data a muscle imaging of DYSF and ANO5 related muscular dystrophies. Neuromuscul Disord 2014;24:1097–1102.
39. Jungbluth H. Myopathology in times of modern imaging. Neuropathol Appl Neurobiol 2017;43:24–43.
40. Beckmann JS, Bushby K. Advances in the molecular genetics of the limb-girdle type of autosomal recessive progressive muscular dystrophy. Curr Opin Neurol 1996;9: 389–393.
41. ten Dam L, Van der Kooi AJ, Van Wattingen M, de Haan RJ, de Visser M. Reliability and accuracy of skeletal muscle imaging in limb-girdle muscular dystrophies. Neurology 2012;79:1716–1723.
42. Silveira-Moriyama L, Paciorkowski AR. Genetic diagnostics for neurologists. Continuum 2018;24:18–36.
43. Lee PH, Lee C, Li X, Wee B, Dwivedi T, Daly M. Principles and methods of in-silico prioritization of non-coding regulatory variants. Hum Genet 2018;137:15–30.
44. Vasli N, Laporte J. Impacts of massively parallel sequencing for genetic diagnosis of neuromuscular disorders. Acta Neuropathol 2013;125:173–185.
45. Le EP, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. Clin Radiol 2019;74:357–366.
46. Grollemund V, Pradat PF, Querin G, et al. Machine learning in amyotrophic lateral sclerosis: achievements, pitfalls, and future directions. Front Neurosci 2019;13:135.
47. Carlier RY, Laforet P, Wary C, et al. Whole-body muscle MRI in 20 patients suffering from late onset Pompe disease: involvement patterns. Neuromuscul Disord 2011;21: 791–799.

Accuracy of a machine learning muscle MRI-based tool for the diagnosis of muscular dystrophies

José Verdú-Díaz, Jorge Alonso-Pérez, Claudia Nuñez-Peralta, et al.

Neurology published online February 6, 2020
DOI 10.1212/WNL.0000000000009068