

Short-term outcome prediction for myasthenia gravis: an explainable machine learning model

Huahua Zhong*, Zhe Ruan*, Chong Yan*, Zhiguo Lv, Xueying Zheng, Li-Ying Goh, Jianying Xi, Jie Song, Lijun Luo, Lan Chu, Song Tan, Chao Zhang, Bitao Bu, Yuwei Da , Ruisheng Duan, Huan Yang , Sushan Luo , Ting Chang  and Chongbo Zhao; The Pan-Yangtze River Delta Alliance for Neuromuscular Disorders (PYDAN)

Abstract

Background: Myasthenia gravis (MG) is an autoimmune disease characterized by muscle weakness and fatigability. The fluctuating nature of the disease course impedes the clinical management.

Objective: The purpose of the study was to establish and validate a machine learning (ML)-based model for predicting the short-term clinical outcome in MG patients with different antibody types.

Methods: We studied 890 MG patients who had regular follow-ups at 11 tertiary centers in China from 1 January 2015 to 31 July 2021 (653 patients for derivation and 237 for validation). The short-term outcome was the modified post-intervention status (PIS) at a 6-month visit. A two-step variable screening was used to determine the factors for model construction and 14 ML algorithms were used for model optimisation.

Results: The derivation cohort included 653 patients from Huashan hospital [age 44.24 (17.22) years, female 57.6%, generalized MG 73.5%], and the validation cohort included 237 patients from 10 independent centers [age 44.24 (17.22) years, female 55.0%, generalized MG 81.2%]. The ML model identified patients who were improved with an area under the receiver operating characteristic curve (AUC) of 0.91 [0.89–0.93], ‘Unchanged’ 0.89 [0.87–0.91], and ‘Worse’ 0.89 [0.85–0.92] in the derivation cohort, whereas identified patients who were improved with an AUC of 0.84 [0.79–0.89], ‘Unchanged’ 0.74 [0.67–0.82], and ‘Worse’ 0.79 [0.70–0.88] in the validation cohort. Both datasets presented a good calibration ability by fitting the expectation slopes. The model is finally explained by 25 simple predictors and transferred to a feasible web tool for an initial assessment.

Conclusion: The explainable, ML-based predictive model can aid in forecasting the short-term outcome for MG with good accuracy in clinical practice.

Keywords: machine learning, myasthenia gravis, prognosis, short-term

Received: 17 May 2022; revised manuscript accepted: 15 January 2023.

Introduction

Myasthenia gravis (MG) is characterized by fatigable weakness in ocular, bulbar, limbs, and respiratory muscles, with an annual incidence of roughly 10–29 cases per million people.¹ It is an autoimmune disease mediated by antibodies targeting the postsynaptic components in the neuromuscular junction.² Of them, 80–85% have

antibodies against the acetylcholine receptor (AChR), while about 5–8% have antibodies against the muscle-specific tyrosine kinase (MuSK) and 7–33% against low-density lipoprotein receptor-related protein 4 (LRP4).³ It can be further divided into different subgroups according to the involved muscle domains, thymoma associations, antibody specificity, and the onset ages.⁴

Ther Adv Neurol Disord

2023, Vol. 16: 1–16

DOI: 10.1177/
17562864231154976

© The Author(s), 2023.
Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:

Sushan Luo
Huashan Rare Disease
Center, Department of
Neurology, Huashan
Hospital, Fudan University,
Shanghai 200040, China.
luosushan@fudan.edu.cn

National Center for
Neurological Disorders,
Shanghai, China

Chongbo Zhao
Huashan Rare Disease
Center, Department of
Neurology, Huashan
Hospital, Fudan University,
Shanghai 200040, China.
zhao_chongbo@fudan.edu.cn

National Center for
Neurological Disorders,
Shanghai, China

Ting Chang
Department of Neurology,
Tangdu Hospital, The Air
Force Medical University,
Xi'an 710000, China.
changting1981@163.com

Huahua Zhong
Chong Yan
Jianying Xi
Jie Song
Huashan Rare Disease
Center, Department of
Neurology, Huashan
Hospital, Fudan University,
Shanghai, China

National Center for
Neurological Disorders,
Shanghai, China

Zhe Ruan
Department of Neurology,
Tangdu Hospital, The Air
Force Medical University,
Xi'an, China

Zhiguo Lv

Department of Neurology,
The Affiliated Hospital
of Changchun University
of Chinese Medicine,
Changchun, China

Xueying Zheng

Department of
Biostatistics, School of
Public Health and Key
Laboratory of Public
Health Safety, Fudan
University, Shanghai,
China

Li-Ying Goh

Shanghai Medical
College, Fudan University,
Shanghai, China

Lijun Luo

Department of Neurology,
Wuhan No.1 Hospital,
Wuhan, China

Lan Chu

Department of Neurology,
The Affiliated Hospital
of Guizhou Medical
University, Guiyang, China

Song Tan

Department of Neurology,
Sichuan Provincial
People's Hospital,
University of Electronic
Science and Technology of
China, Chengdu, China

Chao Zhang

Department of Neurology
and Tianjin Neurological
Institute, Tianjin Medical
University General
Hospital, Tianjin, China

Bitao Bu

Department of Neurology,
Tongji Hospital, Tongji
Medical College, Huazhong
University of Science and
Technology, Wuhan, China

Yuwei Da

Department of Neurology,
Xuanwu Hospital, Capital
Medical University, Beijing,
China

Ruisheng Duan

Department of Neurology,
The First Affiliated
Hospital of Shandong First
Medical University, Jinan,
China

Huan Yang

Department of Neurology,
Xiangya Hospital,
Central South University,
Changsha, China

*Huahua Zhong, Zhe Ruan,
and Chong Yan are co-first
authors.

The hallmark of MG is the highly heterogeneous muscle involvement and fluctuating muscle weakness. Most patients require consistent and even lifelong treatment to reduce relapse or avoid conversion from ocular MG (OMG) to generalized MG (GMG).^{5,6} Although recent advances have been made in targeted immunotherapies,⁷ very few patients have achieved complete stable remission.⁸ A clinically distinct subgroup of patients who are refractory to treatment and continue to experience worsening accounts for 10–30% of MG patients.³ In addition, rapid worsening in respiratory muscles may lead to a myasthenic crisis (MC), which is a life-threatening condition that occurs in 15–20% of patients.⁹ The mortality rate in MC was estimated to be 5–12%,⁹ which leads to a considerable disease burden. The fluctuating nature of muscle weakness and the variable responsiveness to immunotherapies remained the most intractable issues in the clinical management of MG. Furthermore, biologics are emerging as important therapeutic tools that promise to provide better corticosteroid-sparing effects than standard treatments and can even induce remission. These target-specific immunotherapies mainly include anti-complement therapeutics, the anti-FcRn and B cell monoclonals.^{10–12} Early identification of patients prone to worsening may prompt advanced targeted immunotherapies. This defines an urgent clinical need for additional tools to forecast the prognosis and personalize disease management.

Previous retrospective studies have attempted to explore the baseline risk factors for predicting the outcome in MG patients, which included clinical characteristics and relevant biomarkers. For instance, vital capacity, disease duration, and bulbar symptoms were significantly associated with the occurrence of postoperative MC.¹³ A high reduction rate in anti-AChR antibody titers is associated with a favorable outcome at 1-year post-treatment.¹⁴ Peripheral memory B cell percentage has been used to predict relapse in rituximab-treated MG.¹⁵ More recently, adult-onset OMG, abnormal repetitive stimuli (RNS) findings, seropositivity for anti-AChR antibody, and thymoma concurrence were identified as risk factors for generalization.¹⁶ However, these studies mainly focused on specific subtypes, for example, anti-AChR antibody-positive generalized MG; or pre-defined conditions, for example, well-controlled patients treated with steroids, and were not suitable for a consecutive follow-up setting.^{17,18}

In the current study, we aimed to develop and validate a data-driven, machine learning (ML)-based, predictive model to forecast the short-term outcome for MG. The advantages of this model include (1) accommodating three subtypes of MG (AChR, MuSK, negative); (2) being robust and replicable; (3) feasible for health professional evaluations. If successful, the ML-based predictive model would be clinically helpful in informing the aggressiveness of treatment on an individual basis at each visit.

Methods

Patients and criteria

We used ML algorithms to establish a prediction model for short-term outcomes in a derivation cohort and a multicenter-derived validation cohort (Figure 1). The Huashan MG registry (established Jan 1, 2015) is a disease-specific database that comprised the records of 4126 visits in 1560 MG patients, who were mainly referred from the coastal areas in Southeast China. The derivation cohort included 653 patients from the Huashan MG registry from 1 January 2015 through 31 July 2021. The validation cohort comprised 237 MG patients referred by 10 independent tertiary medical centers from northern, western, and southern China from 1 June 2016 through 1 December 2020.

All MG diagnoses were made by a specialist in neurology, and the baseline and follow-up evaluations were made by a neuromuscular specialist. Eligible criteria for the patient enrollment included (1) onset symptoms and signs compatible with OMG or GMG; (2) seropositive for at least one antibody: AChR or MuSK, or LRP4 antibodies; and/or (3) positive repetitive nerve stimuli (RNS); (4) follow-up duration of 6 months or longer from baseline. Exclusion criteria included those participants who had more than 5% missing variables. The 6-month interval was set due to the large sample size and the common implementation in MG studies.^{19–21}

One MG patient was only followed by one center. None of the MG patients in the derivation cohort was enrolled in the multicenter-derived validation cohort. This study protocol and patient data usage were approved by the Ethics Committees of Huashan Hospital and all participating centers.

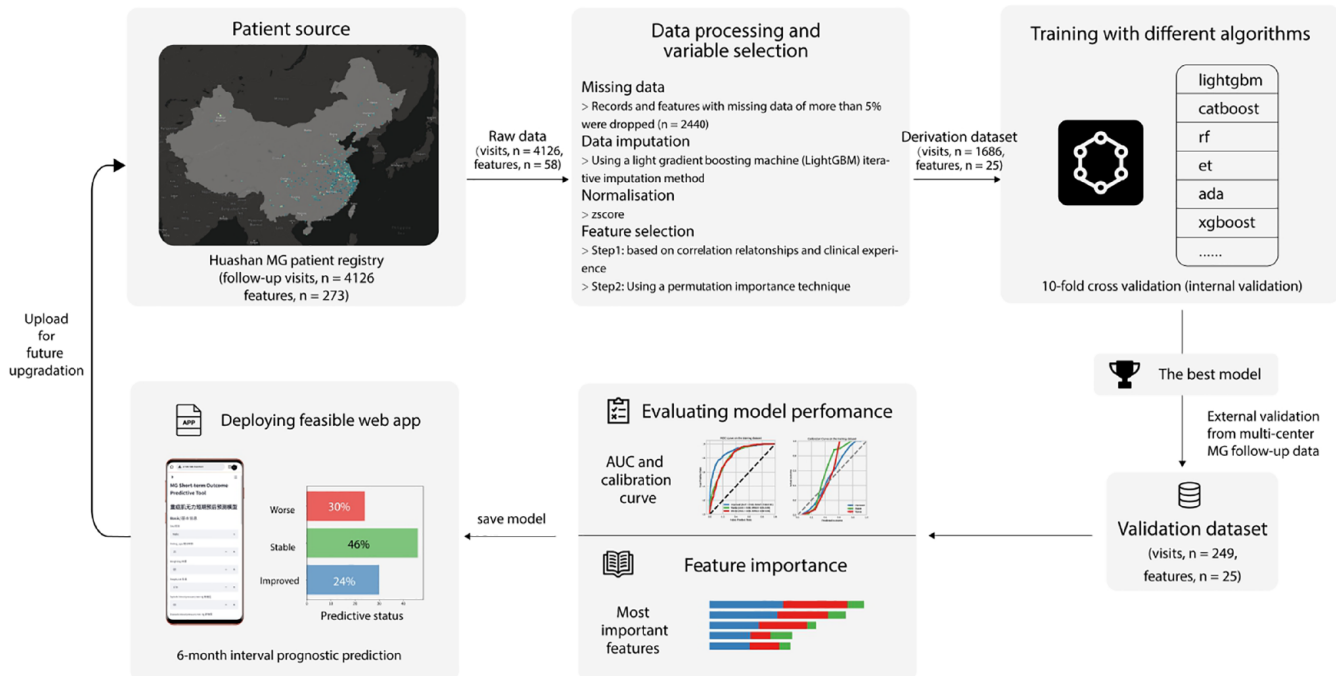


Figure 1. The workflow for the myasthenia gravis (MG) short-term outcome prediction model generation and application. The derivation cohort ($n = 653$) was derived from the Huashan MG patient registry, while the validation cohort ($n = 237$) was from other 10 MG centers in China. The bar plot shows the distribution of patient visiting stages ranging from V1–V7 (V for a visit). 1686 consecutive follow-up records comprised the derivation dataset, and 249 records contained the validation dataset. Commonly used machine learning (ML) algorithms were compared in the training process, including LightGBM (light gradient boosting), catboost, rf (random forest), et (extra trees), ada (adaptive Boosting), and xgboost (extreme gradient boosting), and in this study, the random forest was the best algorithm.

Written informed consent was obtained from each enrolled participant.

A total of 58 baseline features were collected from the derivation cohort, including demographic characteristics ($n = 3$), clinical features ($n = 46$), treatment types ($n = 6$), thymoma status ($n = 1$), and comorbidity ($n = 2$). This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guideline for prognostic studies (Table S1).²² Detailed items from MG scales, including Quantitative Myasthenia Gravis (QMG), MG Activity of Daily Living (ADL), and MG Quality of Life 15-items (QOL-15), were specifically considered instead of the sum scores.²³ Participants were required to discontinue pyridostigmine for at least 10 hours before the clinical assessment.

Primary outcome and outcome measurement

The primary outcome was the modified Post-Intervention Status (PIS) at the 6-month visit,

categorized by the changes in QMG score. We defined the outcome in categorical status as ‘Improved’, ‘Unchanged’, and ‘Worse’, referred to and modified from the Myasthenia Gravis Foundation of America (MGFA) PIS classification.²⁴ The ‘Unchanged’ category was added based on the following reasons: (1) Through analysis of the short-term outcome by minimizing measurement error at each visit, around 25% of patients exhibited a Unchanged status and still did not meet the improvement criteria;^{25,26} (2) during the following time frame of 6 months, the therapeutic efficacy of immunosuppressants have not been fully demonstrated.^{3,27}

Previous studies on minimal clinically significant differences have been set at a 3-point change if baseline QMG > 16 and a 2-point change if baseline QMG ≤ 16 .^{23,25,28} Herein, we defined three categories (1) when the baseline QMG score > 16 , ‘Improved’ means reduced score ≥ 3 , ‘Unchanged’ means increased or reduced score < 3 , and ‘Worse’ means increased score ≥ 3 ; (2) when the baseline QMG score ≤ 16 ,

'Improved' means reduced score ≥ 2 , 'Unchanged' means increased or reduced score < 2 , and 'Worse' means increased score ≥ 2 . In addition, acute MG worsening and the outcome of death was classified as the 'Worse' outcomes.

Data preparation and feature selection

Since MG is a fluctuating disease that can deteriorate within days to months, each visit was considered an independent event for simplifying the model. In the derivation dataset, follow-up records and clinical features with more than 5% missing data were dropped, and the remaining missing data were filled using a light gradient boosting machine (LightGBM) iterative imputation method.²⁹ A two-step feature selection procedure was first processed to filter the most influential features ($n = 58$), including MG scales, primary status (gender, age, height, weight), thymoma states, comorbidities, and treatments. Since some items have similarities in clinical significance in different MG scales (e.g. eyelid droop in ADL and upward gaze ptosis in QMG) or in left and proper comparison (e.g. left- and right-hand grip force in QMG), redundant items were then removed. Height and weight were converted to body mass index (BMI). In Step 1, a correlation matrix was made, and redundant features with correlation efficient $r > 0.6$ or having similar clinical significance were excluded. In Step 2, the selection was algorithm-based, and those features that contributed mainly to the prediction were automatically selected using a permutation importance technique with a 0.6 filtering threshold.

Model development and validation

A total of 14 different ML algorithms were preliminarily tested on the derivation dataset, and seven commonly used model performance metrics including accuracy, the area under the receiver operating characteristic curve (AUC), recall, precision, F1, Kappa, and Matthew's correlation coefficient were calculated to help determine the optimal algorithm. The best-performed algorithm was further fine-tuned to generate a better model. The internal and external validation were processed by a 10-fold cross-validation procedure and an external validation dataset, respectively. The precision of the model was evaluated by calibration plots. Finally, the

interpretation and feature importance of the model with the best performance was sought by a game theory-based SHAP (Shapley Additive Explanations) approach.³⁰

Statistical analysis

Continuous features were reported as mean (SD) and compared by a one-way analysis of variance (ANOVA) test. In contrast, semi-quantitative features (QMG, ADL, and QOL-15 scores) were reported as median (IQR) and compared by Kruskal–Wallis test. Categorical features were reported as percentages (%) and compared by the chi-square test and Fisher's exact test. The model was generated using the PyCaret pipeline (<https://pycaret.org/about>) on Python version 3.8.3 (Python Software Foundation). The specific features were analyzed using tidyverse, tableone, ggstatsplot, and cowplot packages on R version 4.03 (R Foundation for Statistical Computing). The model was finally deployed on a web server using a Streamlit python library (<https://streamlit.io/>).

Results

Clinical cohort comparison

The derivation cohort included 653 MG patients [age 44.24 (17.22) years, female 57.6% (376/653), generalized MG 73.5% (480/653)] and the average visit interval was 6.60 (2.85) months (Table 1). The validation cohort included 237 MG patients [age 44.24 (17.22) years, female 55.0% (130/237), generalized MG 81.2% (192/237)] and the average visiting interval was 7.51 (2.19) months (Table 2). In the derivation cohort, 76.9% (502/653) were AChR antibody-positive, and 4.3% (28/653) were MuSK antibody-positive, and 18.8% (123/653) were seronegative; while in the validation cohort, 75.9% (180/237) were AChR antibody-positive and 4.2% (10/237) were MuSK antibody-positive, and 19.8% (47/237) were seronegative. At the baseline, the thymoma concurrence proportion was higher in the derivation cohort (22.7% versus 14.0%, $p = 0.022$), whereas the proportion for thymectomy was higher in the validation cohort (20.2% versus 10.6%, $p < 0.001$). However, the derivation cohort has similar proportions in other chronic comorbidities (19.5% versus 22.9%, $p = 0.274$) and autoimmune diseases (15.6%

Table 1. Demographic and clinical features of the derivation dataset.

Feature	Improved (n = 629)	Unchanged (n = 841)	Worse (n = 216)	p
Visit interval (months)	6.60 (2.96)	6.51 (1.87)	6.63 (3.52)	0.344
Age (years)	43.11 (16.32)	44.37 (16.44)	45.41 (16.57)	0.145
Gender = male (%)	256 (40.7)	363 (43.2)	78 (36.1)	0.158
Weight (kg)	63.32 (13.55)	64.73 (14.19)	65.21 (13.65)	0.095
Height (cm)	164.86 (7.98)	164.57 (8.40)	164.34 (10.90)	0.698
Waistline (cm)	82.08 (12.18)	83.31 (12.51)	84.42 (11.57)	0.039
Hipline (cm)	94.00 (8.68)	94.25 (10.37)	95.26 (8.57)	0.262
SBP (mmHg)	125.15 (18.14)	125.47 (17.95)	123.81 (18.50)	0.498
DBP (mmHg)	81.71 (10.84)	81.25 (10.93)	80.06 (12.22)	0.178
Heart rate (beats per minute)	83.28 (14.26)	83.00 (13.64)	82.60 (14.30)	0.824
MGFA classification (%)				< 0.001
I	140 (22.3)	391 (46.5)	113 (52.3)	
IIa	189 (30.0)	223 (26.5)	45 (20.8)	
IIb	146 (23.2)	151 (18.0)	36 (16.7)	
IIIa	67 (10.7)	27 (3.2)	13 (6.0)	
IIIb	60 (9.5)	31 (3.7)	8 (3.7)	
IVa	14 (2.2)	2 (0.2)	0 (0.0)	
IVb	9 (1.4)	5 (0.6)	1 (0.5)	
V	4 (0.6)	11 (1.3)	0 (0.0)	
QMG (scores)	10.00 [7.00, 14.00]	4.00 [2.00, 8.00]	4.00 [2.00, 7.00]	< 0.001
ADL (scores)	4.00 [2.00, 7.00]	1.00 [0.00, 4.00]	1.00 [0.00, 3.00]	< 0.001
QOL-15 (scores)	16.00 [7.00, 28.00]	8.00 [2.00, 18.00]	7.00 [1.00, 18.00]	< 0.001
Antibody type (%)				0.524
AChR +	481 (76.4)	641 (67.2)	174 (80.6)	
MuSK +	25 (4.0)	32 (3.8)	10 (4.6)	
Negative (AChR/MuSK/LRP4 -)	123 (19.6)	168 (20.0)	32 (14.8)	
Thymectomy = yes (%)	71 (11.3)	89 (10.6)	15 (6.9)	0.189
Thymoma = yes (%)	158 (25.1)	210 (25.0)	48 (22.2)	0.669
Other autoimmune diseases = yes (%)	118 (18.7)	119 (14.2)	26 (11.9)	0.017

(Continued)

Table 1. (Continued)

Feature	Improved (n = 629)	Unchanged (n = 841)	Worse (n = 216)	p
Comorbidity = yes (%)	130 (20.7)	164 (19.5)	34 (15.6)	0.287
Treatment (pyridostigmine) = yes (%)	515 (81.9)	700 (83.2)	176 (81.5)	0.727
Treatment (corticosteroids) = yes (%)	518 (82.4)	589 (70.0)	154 (71.3)	< 0.001
Treatment (immunosuppressants) = yes (%)	119 (18.9)	215 (25.6)	62 (28.7)	0.002
Treatment (monoclonal antibody drugs) = yes (%)	36 (5.7)	66 (7.8)	20 (9.3)	0.141
Treatment (urgent relief) = yes (%)	16 (2.5)	8 (0.9)	5 (2.3)	0.052

AChR, acetylcholine receptor; ADL, activity of daily living; DBP, diastolic blood pressure; IQR, interquartile range; LRP4, lipoprotein receptor-related protein 4; MGFA, Myasthenia Gravis Foundation of America; MuSK, muscle-specific tyrosine kinase; QMG, quantitative myasthenia gravis; QOL-15, Quality of Life 15 items; SBP, systolic blood pressure; SD, standard deviation.

Continuous features are reported as mean (SD) test, semi-quantitative features are reported as median (IQR), and categorical features are reported as percentage (%). 'Other autoimmune diseases' represents commonly accompanied autoimmune diseases with MG including rheumatoid, systemic lupus erythematosus, leukoderma, and so on. 'Comorbidity' represents a more common comorbidities occurring also in normal population, including hypertension, hyperlipidemia, stroke, tumor, and so on.

versus 17.7%, $p = 0.452$) compared with that in the validation cohort.

The derivation dataset comprised 1686 visits from all categories according to MGFA clinical classification: MGFA I (38.2%, 644/1686), MGFA II (46.9%, 791/1686), MGFA III (12.2%, 206/1686), MGFA IV (1.8%, 30/1686), and MGFA V (0.9%, 15/1686); while for the validation dataset, the proportions were MGFA I (28.9%, 72/249), MGFA II (42.2%, 105/249),

MGFA III (22.1%, 55/249), MGFA IV (5.6%, 14/249), and MGFA V (1.2%, 3/249), respectively.

Clinical predictors with a two-step selecting strategy

A total of 58 baseline variables entered the screening (Table S2). After Step 1 selection, 21 variables were excluded due to being correlated with other features or not being convenient enough. In

Table 2. Demographic and clinical features of the validation dataset.

Feature	Improved (n = 127)	Unchanged (n = 76)	Worse (n = 46)	p
Visit interval (months)	7.37 (1.96)	6.88 (2.55)	7.04 (1.73)	0.211
Age (years)	45.20 (17.18)	46.22 (17.91)	47.00 (18.64)	0.83
Gender = male (%)	61 (48.0)	37 (48.7)	17 (37.0)	0.379
Weight (kg)	67.46 (11.52)	65.81 (11.92)	65.64 (14.11)	0.555
Height (cm)	166.56 (7.25)	164.81 (7.03)	164.30 (7.06)	0.108
Waistline (cm)	83.62 (10.02)	79.22 (7.67)	81.62 (3.70)	0.229
Hipline (cm)	95.27 (8.34)	94.89 (5.51)	90.25 (5.85)	0.219

(Continued)

Table 2. (Continued)

Feature	Improved (<i>n</i> = 127)	Unchanged (<i>n</i> = 76)	Worse (<i>n</i> = 46)	<i>p</i>
SBP (mmHg)	127.96 (17.64)	130.30 (16.53)	130.06 (15.51)	0.835
DBP (mmHg)	77.92 (11.34)	76.40 (8.78)	81.12 (8.69)	0.371
Heart rate (beats per minute)	75.89 (8.64)	76.52 (10.14)	75.65 (8.10)	0.949
MGFA classification (%)				0.201
I	28 (22.0)	29 (38.7)	15 (32.6)	
IIa	24 (18.9)	19 (25.3)	15 (32.6)	
IIb	28 (22.0)	13 (17.3)	6 (13.0)	
IIIa	18 (14.2)	6 (8.0)	4 (8.7)	
IIIb	19 (15.0)	5 (6.7)	3 (6.5)	
Iva	1 (0.8)	1 (1.3)	0 (0.0)	
IVb	7 (5.5)	2 (2.7)	3 (6.5)	
V	2 (1.6)	1 (1.3)	0 (0.0)	
QMG (scores)	11.00 [8.00, 16.00]	6.00 [3.00, 9.00]	4.00 [1.25, 6.00]	< 0.001
ADL (scores)	6.00 [4.00, 8.00]	3.00 [1.00, 5.25]	3.00 [1.25, 6.00]	< 0.001
QOL-15 (scores)	6.00 [0.00, 16.00]	1.00 [0.00, 8.00]	0.00 [0.00, 4.75]	0.001
Antibody type (%)				0.149
AChR +	99 (78.0)	56 (73.7)	36 (78.3)	
MuSK +	4 (3.1)	2 (2.6)	4 (8.7)	
Negative (AChR/MuSK/LRP4 -)	24 (18.9)	18 (23.7)	6 (13.0)	
Thymectomy = yes (%)	21 (17.1)	20 (26.7)	9 (20.0)	0.268
Thymoma = yes (%)	23 (18.7)	13 (17.8)	9 (22.0)	0.858
Other autoimmune diseases = yes (%)	23 (18.1)	11 (14.5)	10 (22.2)	0.552
Comorbidity = yes (%)	33 (26.0)	15 (19.7)	9 (19.6)	0.495
Treatment (pyridostigmine) = yes (%)	122 (96.1)	67 (88.2)	43 (93.5)	0.096
Treatment (corticosteroids) = yes (%)	106 (83.5)	52 (68.4)	43 (93.5)	0.002
Treatment (immunosuppressants) = yes (%)	89 (70.1)	58 (76.3)	26 (56.5)	0.069
Treatment (monoclonal antibody drugs) = yes (%)	34 (26.7)	17 (22.4)	10 (21.7)	0.694
Treatment (urgent relief) = yes (%)	6 (4.7)	3 (3.9)	2 (4.3)	0.966

AChR, acetylcholine receptor; ADL, activity of daily living; DBP, diastolic blood pressure; IQR, interquartile range; LRP4, lipoprotein receptor-related protein 4; MGFA, Myasthenia Gravis Foundation of America; MuSK, muscle-specific tyrosine kinase; QMG, quantitative myasthenia gravis; QOL-15, Quality of Life 15 items; SBP, systolic blood pressure; SD, standard deviation.

contrast to the QMG scale, both ADL and QOL-15 were patient-reported outcomes prone to be swayed by psychological factors.³¹ Hence, their items were less preferred when redundancy appeared with QMG items. Since most QOL-15 items ($n = 12$) were clustered and correlated with each other, and the rest ($n = 3$) also correlated well with the items in other scales, the QOL-15 scale was dropped from our model. In terms of synonyms in predictors, talking, chewing, swallowing, breathing, double vision, and ptosis in the ADL scale were dropped due to the replications in the QMG scale. The MGFA classifications and generalised/ocular types were also filtered out for MG scales due to the overlapped information with other scales. Given that the evaluations on the right body side would be affected by right dominance, we selected the items measured on the left body side, for example, left-arm outstretched, left-hand grip, and left-leg stretch in QMG.³² Those features that do not meet the criteria of simplicity and convenience including waistline, hipline, and vital capacity were also dropped. Next, 37 variables have finally entered Step 2 selection followed by an automatic procedure with a pre-determined filter rate of 0.7.

In the end, 25 clinical features were finally selected out of 58 features after the two-step selection procedure. These features include (1) gender; (2) BMI; (3) age; (4) systolic blood pressure (SBP); (5) diastolic blood pressure (DBP); (6) antibody type (AChR/MuSK/Negative); (7) thymectomy; (8) thymoma; (9) comorbidity; (10) pyridostigmine treatment; (11) corticosteroid treatment; (12) immunosuppressant treatment (Azathioprine, Mycophenolate Mofetil, Cyclosporine, or Tacrolimus); (13) monoclonal antibody treatment (Rituximab or Tocilizumab); (14) QMG: Double vision on lateral gaze; (15) QMG: Ptosis (upward gaze); (16) QMG: Facial muscles; (17) QMG: Swallowing 4 oz. water; (18) QMG: Speech following counting; (19) QMG: Left arm outstretched; (20) QMG: Head lifted (45° supine); (21) QMG: Left leg outstretched; (22) ADL: Chewing; (23) ADL: Breathing; (24) ADL: Impairment of ability to brush teeth; (25) ADL: Impairment of ability to arise.

Random forest classifier was chosen as the best algorithm

Finally, 25 variables have entered the training process. In the preliminarily training test, the

random forest classifier outperformed the other 13 algorithms in 5 of 7 model performance metrics (accuracy, AUC, F1, Kappa, Matthew's correlation coefficient) (Table S2). Then the random forest model was further fine-tuned to achieve a good average AUC score of 0.7976 in the 10-fold cross-validation, which has a better performance than the previous untuned random forest model with an AUC score of 0.7118. Each of three predictive results (Improved, Unchanged, and Worse) showed an AUC score above 0.89 in the derivation dataset, and an AUC score above 0.74 in the validation dataset (Figure 2(a) and (b)). The calibration plots of the two datasets generally fitted the direction and slope of the expectation slope (Figure 2(c) and (d)). Subsequently, the two datasets were combined as a combination dataset to subsequently perform a sensitivity analysis of the model on different MG subtypes. For the application in different MG classifications, the AUC score is above 0.88 in AChR/MuSK/Seronegative MG, and the AUC score is above 0.86 in ocular/generalized MG. In summary, the final model showed good discrimination and calibration abilities in the derivation dataset, moderately good performance in the validation dataset, and moderately good generalization ability in different MG subtypes.

Feature importance determination in the model

To evaluate each feature's influence in the ML model, we use a SHAP method for quantification (Figure 3). The top 4 most important features ($|\text{SHAP value}| > 0.06$) were then selected to be retrogradely analyzed in the derivation dataset with additional information, including QMG left-arm outstretched, corticosteroid treatment, QMG ptosis (upward gaze), and antibody type.

Above two QMG items showed similar composition ratios in the 'Unchanged' and 'Worse' categories, indicating that these two states were difficult to discriminate (Figure 4(a) and (b)). We reviewed the dosage and the regimen of corticosteroids used in both derivation and validation cohorts to further explain the model (Figure 4(c)). Two contradicted trends showed that patients sustained with alternative-day corticosteroid treatment were more prone to worse after 6 months ($p < 0.001$). In contrast, patients treated with increasing dosage are prone to become improved ($p < 0.001$). As for antibody types, there is no significant results in comparisons

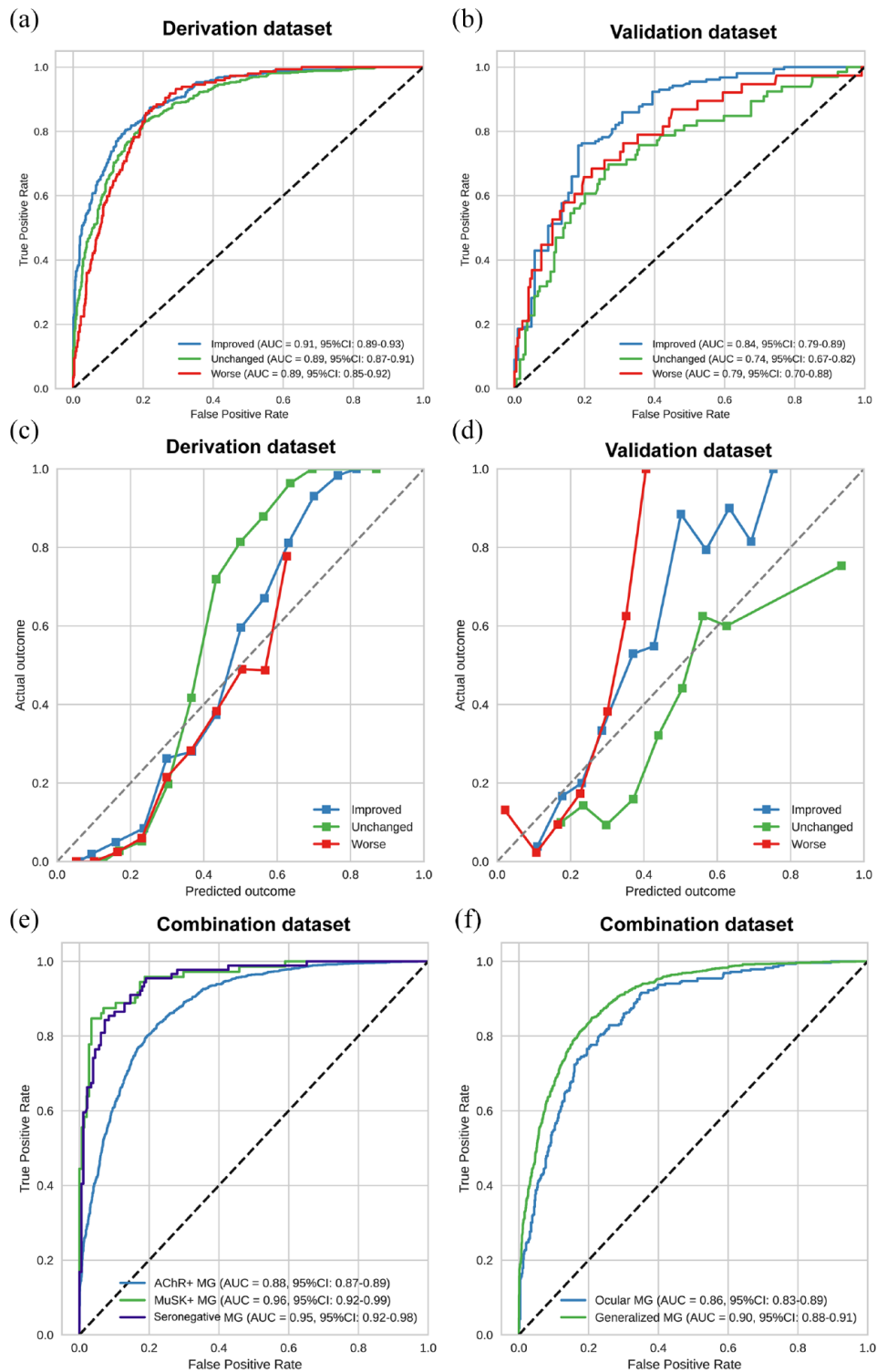


Figure 2. The model performance in terms of discrimination and calibration. (a and b) The area under the curve (AUC) scores of the model in derivation and validation datasets. (c and d) Calibration plots show the calibration ability of the model in derivation and validation datasets. (e and f) The derivation and validation datasets are combined to comprise a combination dataset. The model AUC performance in different MG subtypes in the combination dataset.

Note that seronegative MG represents negative in AChR/MuSK/LRP4 antibodies.

AChR, acetylcholine receptor; LRP4, lipoprotein receptor-related protein 4; MuSK, muscle-specific tyrosine kinase.

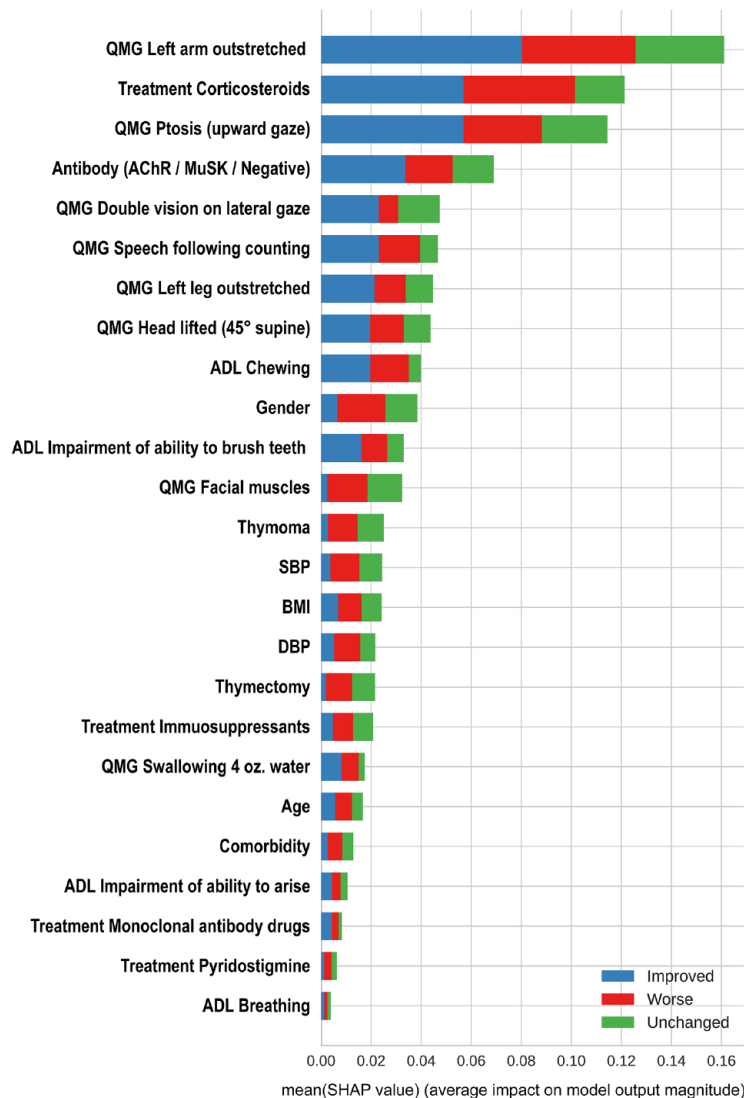


Figure 3. Feature importance of the model.

Summary plot for the 25 clinical features comprise the model calculated by Shapley Additive Explanations (SHAP) method. The features are ordered in descending rank according to their importance to the model. Note that a feature has a different contribution to three results, which is labeled in the plot with different colors. DBP, diastolic blood pressure; SBP, systolic blood pressure.

among three outcomes (Figure 4(d)). Thus, here we generated an explainable ML-based model by screening the best algorithm and quantifying the contribution of each variable.

Explainable machine learning model in clinical settings

We then retrospectively applied the ML model to clinical settings in three MG patients with different conditions to further explain the model

(Figure 5). The expected probabilities of the three categories all started at 0.333 and were pulled by various feature forces in different directions.

Patient 1 is a 49-year-old female with AChR-ab+ GMG for 3 years. She had converted to GMG 1 year after ocular onset. At her first visit (V0, September 2018), she was immunotherapy-naïve with a QMG score of 17 and diagnosed with a thymoma. After treatment with corticosteroids

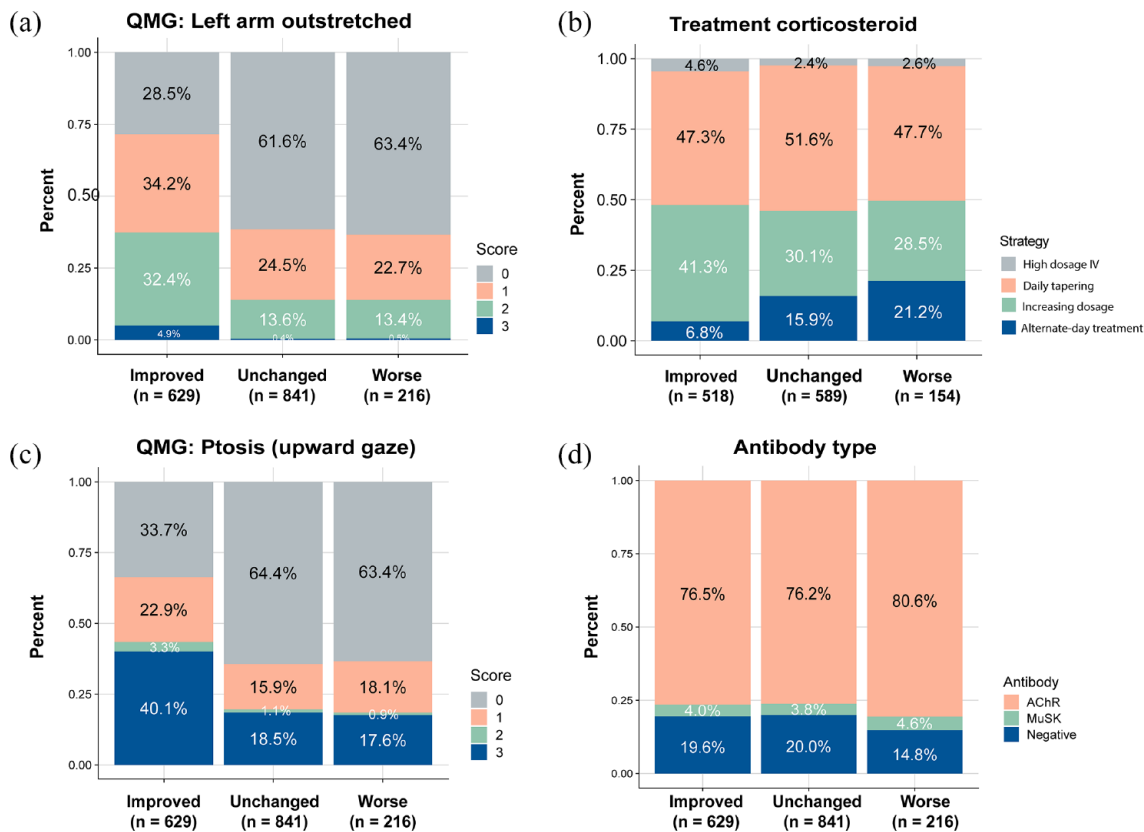


Figure 4. Further analysis of the top 4 most essential features in the model. All the data are sourced from the derivation dataset, and the prognostic categories are in their actual outcomes. (a) The proportional plot of QMG: Left-arm outstretch scores in three outcomes. (b) The constitution in corticosteroid strategies used in those patients (74.8%) treated with corticosteroid in the derivation dataset. (c) The proportional plot of QMG: Ptosis (upward gaze) scores in three outcomes. (d) The proportional plot of antibody types in three outcomes.

and pyridostigmine, the patient significantly improved with a reduced QMG score of 8 at her second visit (V1, March 2019) and the treatment was gradually tapered off, also the patient had a history of thymectomy. At her third visit (V2, September 2019), she had mild weakness in hand gripping with a QMG score of 2. Here we retrospectively used the model to predict the short-term outcome for V3 by inputting the key features. The prognostic prediction for the outcome at the next visit showed a score of 0.11 for 'Improved', 0.41 for 'Unchanged', 0.48 for 'Worse'. The actual outcome at fourth visit (V3, March 2020) was 'Worse' with facial and bulbar muscle weakness and a QMG score of 5 (true positive). From a clinical point of view, this is a middle-aged, overweight woman with a thymoma history. A high BMI and a thymoma history might incline the outcome to 'Worse' rather than 'Unchanged'. This can also be due to corticosteroid tapering.

Patient 2 is a 53-year-old male GMG patient for 12 years. He was double negative for anti-AChR and MuSK antibodies but positive in RNS and neostigmine tests. Oral administration of corticosteroids and azathioprine effectively controlled disease at previous visits (V1–V3, August 2015 to September 2016). At his fifth visit (V4, March 2017), he had slight weakness in the head lifted test with a low QMG score of 1. The prognostic prediction model provided a score of 0.08 for 'Improved', 0.59 for 'Unchanged', 0.26 for 'Worse', with the supporting evidence of a thymectomy history and sustained corticosteroid and azathioprine usage. The actual outcome for this patient is 'Unchanged' at the sixth visit (V5, October 2017), with a small increased QMG score of 3 (true positive).

Patient 3 is a 31-year-old female with GMG for 2 years. The patient has a history of Sjogren's

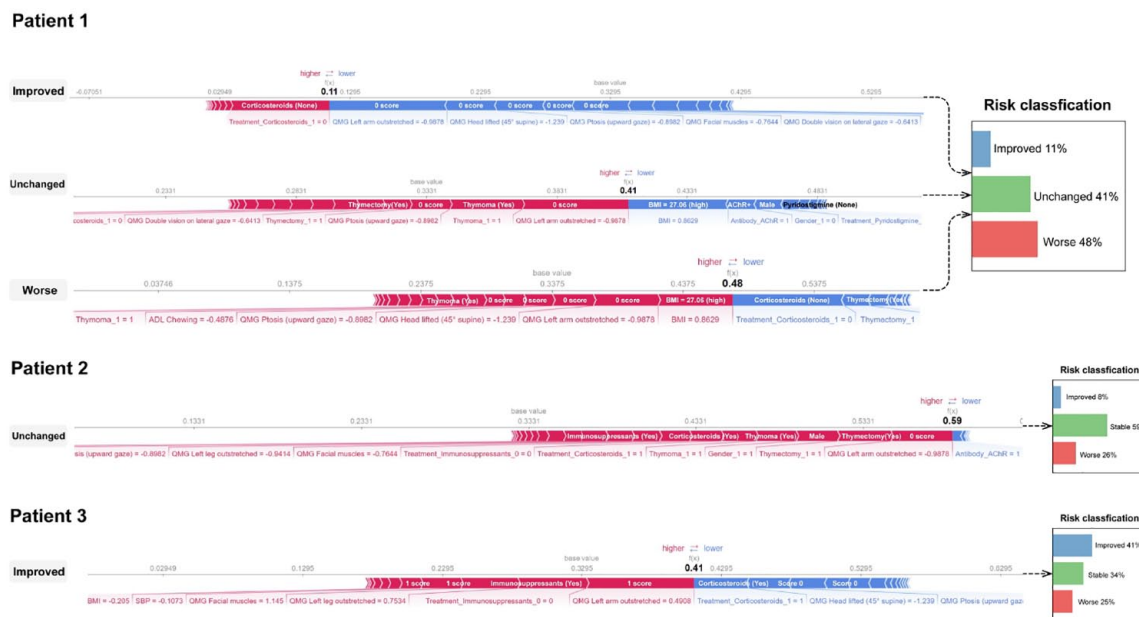


Figure 5. Explainable machine learning model with three patients. Patient 1 is a 'Worse' patient at the next visit, patient 2 an 'Unchanged' case, and patient 3 an 'Improved' case. Note that the model generated three risk scores for one patient (e.g. patient 1); the category with the highest probability score is the result of the model.

syndrome, and the anti-AChR antibody titer was 1.65 nmol/L. After oral corticosteroids and tacrolimus treatment for 6 months, the QMG score significantly decreased from 16 at the first visit (V0, February 2017) to 8 at the second visit (V1, February 2018). From this time point, we used the model to generate a prediction with a score of 0.41 for 'Improved', 0.34 for 'Unchanged', and 0.25 for 'Worse'. The outcome for this patient at her third visit (V2, July 2018) was a significant improvement in ptosis and dyspnea with a decreased QMG score of 3. This case may benefit from the long-lasting effect of tacrolimus.

Discussion

Among MG patients with highly heterogeneous clinical features and responses to the therapies, the prediction for the short-term outcome is more likely to stratify those who had myasthenia worsening and prepared for active therapeutic interventions. ML has been extensively employed in risk stratification and mortality prediction in recent years as an advanced tool in data handling and unilinear relationship mimicking.³³ However, the inherent complexity underlying ML still obfuscates the model interpretation and the clinical relevance, often labeled as a black box.³⁴

Therefore, there is an increasing need for an explainable model with good interpretation and implications for clinical practice. To the best of our knowledge, this is the first attempt to establish a predictive model accommodating all different MG subgroups with explainable scores using the ML technique. This model was validated by a multicenter cohort to verify its external generalization ability, which may provide more convenience for disease surveillance as a web-based tool.

Basic demographic features, subitems of MG scales, antibody types, thymoma status, and different MG-related treatments finally comprised the model. Our model identified 'left arm outstretched' in QMG score as the most influential feature in the prediction, which is consistent with our previous 1-year prognostic model on AChR-ab+ generalized MG.¹⁷ Interestingly, symptoms in specific muscle domains are more stubborn to minimize. According to two large-scale retrospective studies from the United States and Japan,^{35,36} 95% of MG patients eventually experienced an improvement in the first 1–2 years after onset; however, around 35% remained with the slightest weakness in ocular or leg muscles. The treatment-resistant ophthalmoplegia and ptosis were more prevalent in AChR-ab+ /African/Asian MG patients.¹

In contrast, according to the clinical experience, the muscle weakness in the upper limbs is more prone to alleviate than other muscle domains at a 6-month interval, which renders it more sensitive to treatments. Notably, 8 QMG items and 4 ADL items were finally selected for the model construction. QOL-15 scale did not enter the model as a result of correlations between its subitems and other scales revealed in the selection step 1. In addition, factors unrelated to MG symptoms (e.g. the side effect of drugs) may undermine its ability to detect improvement in core MG relevant symptoms.²³ The ADL scores have been previously revealed to correlate better with the oculobulbar domains than generalized muscle domains.³⁷ Hence, items with redundant meanings as QMG items were also not survived at the step 1 selection.

Antibody types which were ranked high in our model as prognosis were heterogeneous in different subgroups. AChR-MG is the most classic subtype, MuSK-MG is characterized by more severe and generalized muscle weakness, and seronegative MG is heterogeneous for including patients with antibodies that have low affinities or have not yet been defined.³⁸ Another ranked high factor was corticosteroid treatment, as it is the most common kind of drugs used to treat MG patients. We identified that the corticosteroid dosage was correlated with the outcomes in the post hoc analysis, namely, patients who were treated with an increasing dosage were more prone to being improved. Interestingly, we also found that the proportion of patients who adopted an alternative-day tapering regimen significantly increased in the ‘Worse’ subgroup, which was consistent with a recent trial supporting that rapid daily tapering of corticosteroids was better than alternative-day regimen in moderate to severe generalized MG.³⁹ As for immunosuppressants and monoclonal antibody drugs, these factors were ranked low in our model. We think this might be explained by different onset times and duration for each drug. Non-steroidal immunosuppressants and monoclonal antibody drugs nowadays used in Chinese patients were mostly tacrolimus and rituximab,^{40,41} and their onset time is relatively longer to achieve ideal drug concentrations and effective duration. Tacrolimus can rapidly improve MG symptoms subjectively within 1 month and objectively at 2–3 months,⁴²

while the time to the peak response after a single cycle of rituximab is 4.5 ± 1 months.⁴³ This heterogeneity in clinical drug selection practice in different centers and the different proportions of immunosuppressants used in two datasets might also explain the relative lower performance in the validation dataset. The reason why thymectomy was not ranked high might be similar, in that benefits of thymectomy are often seen within the first year and will sustain the therapeutic effects through at least 3 years.⁴ While urgent relief treatments were dropped at step 2 selection, due to its scarce data sample size.

There are several limitations in our study: (1) Only the treatment types and no specific drug dosages and exact treatment timeline were enrolled in this model. Future studies with enough sample size in different drugs and detailed regime data may yield more accurate models; (2) the follow-up intervals in both datasets were not technically consistent and the standard deviation is relatively large; (3) the patient source was restricted in consecutively follow-up patients, which may result in a selection bias that those achieved better alleviation were more inclined to be lost in follow-up.

Conclusion

In current study, we have developed a short-term outcome prediction model using ML for MG patients with different antibody types. Our predictive tool may help promote the clinical management of MG patients and build a follow-up surveillance system for professionals.

Declarations

Ethics approval and consent to participate

This study protocol was approved by the Ethics Board of Huashan Hospital, Fudan University (2020-999). Written informed consent was obtained from each enrolled participant for patient data usage in the retrospective analysis.

Consent for publication

Not applicable.

Author contributions

Huahua Zhong: Data curation; Formal analysis; Investigation; Methodology; Writing – original draft.

Zhe Ruan: Data curation; Formal analysis; Investigation; Methodology; Writing – original draft.

Chong Yan: Data curation; Formal analysis; Investigation; Methodology; Writing – original draft.

Zhiguo Lv: Data curation; Investigation; Methodology.

Xueying Zheng: Methodology.

Li-Ying Goh: Investigation; Writing – review & editing.

Jianning Xi: Investigation; Methodology; Writing – review & editing.

Jie Song: Data curation; Investigation.

Lijun Luo: Data curation; Writing – review & editing.

Lan Chu: Data curation; Writing – review & editing.

Song Tan: Data curation; Writing – review & editing.

Chao Zhang: Data curation; Writing – review & editing.

Bitao Bu: Data curation; Writing – review & editing.

Yuwei Da: Data curation; Writing – review & editing.

Ruisheng Duan: Data curation; Writing – review & editing.

Huan Yang: Data curation; Writing – review & editing.

Sushan Luo: Conceptualization; Writing – review & editing.

Ting Chang: Conceptualization; Data curation; Writing – review & editing.

Chongbo Zhao: Conceptualization; Data curation; Funding acquisition; Writing – review & editing.

Acknowledgements

We are most grateful to all MG patients and their families participating in this project and to Dr Wanlong Wu (Renji Hospital South Campus, Shanghai Jiao Tong University School of Medicine) and Dr Zhirui Zhou (Shanghai Cancer Center, Fudan University) for providing

constructive suggestions on model optimization. We also thank Shuang Zhong, Miao Yang, Min Ji, Jinpeng Yang, Dantong Wang, Tianjiao Fang, SHUYU Co, Ltd, and Shanghai Amplicongene Co, Ltd, for their contributions to the Huashan MG patient registry database.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from China's National Natural Science Foundation (No. 81870988, 82071410 and 82001335) and the Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), and ZJLab.

Competing interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Availability of data and materials

The code used for the model generation and the detailed listed features used in this model are available on GitHub (https://github.com/Hirriririr/MG_predictive_tool). The predictive tool implementation of the model is deployed on a web server (<http://47.99.108.154:8501/>). The article data will be shared upon reasonable request to the corresponding author.


Coinvestigators


For The Pan-Yangtze River Delta Alliance for Neuromuscular Disorders

Co-investigators—Zhangyu Zou: Department of Neurology, Fujian Medical University Union Hospital, Fuzhou, China; Qing Ke: Department of Neurology, the first affiliated hospital, Zhejiang University School of Medicine, Hangzhou, China; Jianquan Shi: Department of Neurology, Nanjing First Hospital, Nanjing Medical University, Nanjing, Jiangsu, China; Li Zeng: Department of Neurology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Sichuan, China; Jie Yang: Department of Neurology, Wuhan No.1 hospital, Wuhan, China; Heng Li, Department of Neurology, Central hospital affiliated to Shandong First Medical University, Jinan, China; Song Yang: Department of Neurology, Changzhou First hospital, Changzhou, China; Yali Zhang: Department of Neurology, Inner

Mongolia Medical University, Inner Mongolia, China.

ORCID iDs

Yuwei Da  <https://orcid.org/0000-0002-0318-148X>

Huan Yang  <https://orcid.org/0000-0002-8690-2544>

Sushan Luo  <https://orcid.org/0000-0002-9033-7568>

Ting Chang  <https://orcid.org/0000-0002-7546-8017>

Supplemental material

Supplemental material for this article is available online.

References

- Punga AR, Maddison P, Heckmann JM, *et al.* Epidemiology, diagnostics, and biomarkers of autoimmune neuromuscular junction disorders. *Lancet Neurol* 2022; 21: 176–188.
- Huijbers MG, Marx A, Plomp JJ, *et al.* Advances in the understanding of disease mechanisms of autoimmune neuromuscular junction disorders. *Lancet Neurol* 2022; 21: 163–175.
- Gilhus NE, Tzartos S, Evoli A, *et al.* Myasthenia gravis. *Nat Rev Dis Primers* 2019; 5: 30.
- Wolfe GI, Kaminski HJ, Aban IB, *et al.* Randomized trial of thymectomy in myasthenia gravis. *N Engl J Med* 2016; 375: 511–522.
- Gupta A, Goyal V, Srivastava AK, *et al.* Remission and relapse of myasthenia gravis on long-term azathioprine: an ambispective study. *Muscle Nerve* 2016; 54: 405–412.
- Hendricks TM, Bhatti MT, Hodge DO, *et al.* Incidence, epidemiology, and transformation of ocular myasthenia gravis: a population-based study. *Am J Ophthalmol* 2019; 205: 99–105.
- Verschuuren JJ, Palace J, Murai H, *et al.* Advances and ongoing research in the treatment of autoimmune neuromuscular junction disorders. *Lancet Neurol* 2022; 21: 189–202.
- Andersen JB, Gilhus NE and Sanders DB. Factors affecting outcome in myasthenia gravis. *Muscle & Nerve* 2016; 54: 1041–1049.
- Neumann B, Angstwurm K, Mergenthaler P, *et al.* Myasthenic crisis demanding mechanical ventilation: a multicenter analysis of 250 cases. *Neurology* 2020; 94: e299–e313.
- Dalakas MC. Role of complement, anti-complement therapeutics, and other targeted immunotherapies in myasthenia gravis. *Expert Rev Clin Immunol* 2022; 18: 691–701.
- Dalakas MC, Alexopoulos H and Spaeth PJ. Complement in neurological disorders and emerging complement-targeted therapeutics. *Nat Rev Neurol* 2020; 16: 601–617.
- Dalakas MC. Progress in the therapy of myasthenia gravis: getting closer to effective targeted immunotherapies. *Curr Opin Neurol* 2020; 33: 545–552.
- Kanai T, Uzawa A, Sato Y, *et al.* A clinical predictive score for postoperative myasthenic crisis. *Ann Neurol* 2017; 82: 841–849.
- Kojima Y, Uzawa A, Ozawa Y, *et al.* Rate of change in acetylcholine receptor antibody levels predicts myasthenia gravis outcome. *J Neurol Neurosurg Psychiatry* 2021; 92: 963–968.
- Ruetsch-Chelli C, Bresch S, Seitz-Polski B, *et al.* Memory B cells predict relapse in rituximab-treated myasthenia gravis. *Neurotherapeutics* 2021; 18: 938–948.
- Guo RJ, Gao T, Ruan Z, *et al.* Risk factors for generalization in patients with ocular myasthenia gravis: a multicenter retrospective cohort study. *Neurol Ther* 2022; 11: 73–86.
- Zhao R, Wang Y, Huan X, *et al.* Nomogram for short-term outcome assessment in AChR subtype generalized myasthenia gravis. *J Transl Med* 2021; 19: 1–10.
- Su S, Lei L, Fan Z, *et al.* Clinical predictors of relapse in a cohort of steroid-treated patients with well-controlled myasthenia gravis. *Front Neurol* 2022; 13.
- Brauner S, Eriksson-Dufva A, Hietala MA, *et al.* Comparison between rituximab treatment for new-onset generalized myasthenia gravis and refractory generalized myasthenia gravis. *JAMA Neurol* 2020; 77: 974–981.
- Wanschitz JV, Kaml M, Pfausler B, *et al.* Myasthenic crisis following SARS-CoV-2 infection and delayed virus clearance in a patient treated with rituximab: clinical course and 6-month follow-up. *J Neurol* 2021; 268: 2700–2702.
- Alcantara M, Sarpong E, Barnett C, *et al.* Chronic immunoglobulin maintenance therapy in myasthenia gravis. *Eur J Neurol* 2021; 28: 639–646.

22. Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; 162: 55–63.
23. Thomsen JLS and Andersen H. Outcome measures in clinical trials of patients with myasthenia gravis. *Front Neurol* 2020; 11: 596382.
24. Narayanaswami P, Sanders DB, Wolfe G, *et al.* International consensus guidance for management of myasthenia gravis: 2020 update. *Neurology* 2021; 96: 114–122.
25. Katzberg HD, Barnett C, Merkies IS, *et al.* Minimal clinically important difference in myasthenia gravis: outcomes from a randomized trial. *Muscle Nerve* 2014; 49: 661–665.
26. Howard JF Jr, Nowak RJ, Wolfe GI, *et al.* Clinical effects of the self-administered subcutaneous complement inhibitor zilucoplan in patients with moderate to severe generalized myasthenia gravis: results of a phase 2 randomized, double-blind, placebo-controlled, multicenter clinical trial. *JAMA Neurol* 2020; 77: 582–592.
27. Mantegazza R and Antozzi C. From traditional to targeted immunotherapy in myasthenia gravis: prospects for research. *Front Neurol* 2020; 11: 981.
28. Barnett C, Herbelin L, Dimachkie MM, *et al.* Measuring clinical treatment response in myasthenia gravis. *Neurol Clin* 2018; 36: 339–353.
29. Luo Y. Evaluating the state of the art in missing data imputation for clinical data. *Briefings in Bioinformatics* 2022; 23: bbab489.
30. Lundberg SM, Nair B, Vavilala MS, *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2018; 2: 749–760.
31. Kulaksizoglu IB. Mood and anxiety disorders in patients with myasthenia gravis: aetiology, diagnosis and treatment. *CNS Drugs* 2007; 21: 473–481.
32. Incel NA, Ceceli E, Durukan PB, *et al.* Grip strength: effect of hand dominance. *Singapore Med J* 2002; 43: 234–237.
33. Chang CC, Yeh JH, Chen YM, *et al.* Clinical predictors of prolonged hospital stay in patients with myasthenia gravis: a study using machine learning algorithms. *J Clin Med* 2021; 10: 4393.
34. Collins GS and Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019; 393: 1577–1579.
35. Grob D, Brunner N, Namba T, *et al.* Lifetime course of myasthenia gravis. *Muscle Nerve* 2008; 37: 141–149.
36. Kawaguchi N, Kuwabara S, Nemoto Y, *et al.* Treatment and outcome of myasthenia gravis: retrospective multi-center analysis of 470 Japanese patients, 1999–2000. *J Neurol Sci* 2004; 224: 43–47.
37. de Meel RHP, Raadsheer WF, van Zwet EW, *et al.* Sensitivity of MG-ADL for generalized weakness in myasthenia gravis. *Eur J Neurol* 2019; 26: 947–950.
38. Gilhus NE, Skeie GO, Romi F, *et al.* Myasthenia gravis – autoantibody characteristics and their implications for therapy. *Nat Rev Neurol* 2016; 12: 259–268.
39. Sharshar T, Porcher R, Demeret S, *et al.* Comparison of corticosteroid tapering regimens in myasthenia gravis: a randomized clinical trial. *JAMA Neurol* 2021; 78: 426–433.
40. Lu J, Zhong H, Jing S, *et al.* Low-dose rituximab every 6 months for the treatment of acetylcholine receptor-positive refractory generalized myasthenia gravis. *Muscle Nerve* 2020; 61: 311–315.
41. Zhou L, Liu W, Li W, *et al.* Tacrolimus in the treatment of myasthenia gravis in patients with an inadequate response to glucocorticoid therapy: randomized, double-blind, placebo-controlled study conducted in China. *Ther Adv Neurol Disord* 2017; 10: 315–325.
42. Fan Z, Li Z, Shen F, *et al.* Favorable effects of tacrolimus monotherapy on myasthenia gravis patients. *Front Neurol* 2020; 11: 594152.
43. Anderson D, Phan C, Johnston WS, *et al.* Rituximab in refractory myasthenia gravis: a prospective, open-label study with long-term follow-up. *Ann Clin Transl Neurol* 2016; 3: 552–555.