scientific reports

Check for updates

OPEN

Decoding myasthenia gravis: advanced diagnosis with infrared spectroscopy and machine learning

Feride Severcan[1]✉, Ipek Ozyurt[1], Ayca Dogan[2], Mete Severcan[3], Rafig Gurbanov[4], Fulya Kucukcankurt[5], Birsen Elibol[6,7], Irem Tiftikcioglu[8], Esra Gursoy[9,10], Melike Nur Yangin[11] & Yasar Zorlu[12]

Myasthenia Gravis (MG) is a rare neurological disease. Although there are intensive efforts, the underlying mechanism of MG still has not been fully elucidated, and early diagnosis is still a question mark. Diagnostic paraclinical tests are also time-consuming, burden patients financially, and sometimes all test results can be negative. Therefore, rapid, cost-effective novel methods are essential for the early accurate diagnosis of MG. Here, we aimed to determine MG-induced spectral biomarkers from blood serum using infrared spectroscopy. Furthermore, infrared spectroscopy coupled with multivariate analysis methods e.g., principal component analysis (PCA), support vector machine (SVM), discriminant analysis and Neural Network Classifier were used for rapid MG diagnosis. The detailed spectral characterization studies revealed significant increases in lipid peroxidation; saturated lipid, protein, and DNA concentrations; protein phosphorylation; PO_2^-asym + sym /protein and PO_2^-sym/lipid ratios; as well as structural changes in protein with a significant decrease in lipid dynamics. All these spectral parameters can be used as biomarkers for MG diagnosis and also in MG therapy. Furthermore, MG was diagnosed with 100% accuracy, sensitivity and specificity values by infrared spectroscopy coupled with multivariate analysis methods. In conclusion, FTIR spectroscopy coupled with machine learning technology is advancing towards clinical translation as a rapid, low-cost, sensitive novel approach for MG diagnosis.

Myasthenia Gravis (MG) is a rare disease with an incidence of 15–20 per 100,000 people with variability in different countries[1]. It is an autoimmune disorder that results in impaired postsynaptic neuromuscular transmission. This could be related to toxicity, immunology, or genetics, among other factors[2]. Autoimmune MG is related to autoantibodies, which develop against postsynaptic membrane proteins, resulting in reduced electrical impulse conduction at the neuromuscular junction, which causes clinical muscle weakness[3]. Although MG typically affects different muscles, the sites and severity of muscle involvement vary from patient to patient, and the clinical picture may change over time. Specific clinical symptoms occurring in MG patients—such as double vision, drooping eyelids, and weakness or fatigue in the bulbar, extremity, and cervical muscles—serve as crucial indicators for diagnosis of MG[4]. Facial paralysis, difficulty in speech, exhaustion, and weakness in neck muscles can be accompanied by the main clinical symptoms. Moreover, patients may experience unexplained muscle weakness, heightened fatigue during nighttime, and worsening of symptoms with exertion. If a person has a clinical suspicion of MG, based on the presence of such symptoms, several tests are administered to diagnose and confirm the disease.

[1]Department of Biophysics, Faculty of Medicine, Altinbas University, Istanbul, Türkiye. [2]Department of Physiology, Faculty of Medicine, Altinbas University, Istanbul, Türkiye. [3]Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Türkiye. [4]Department of Bioengineering, Faculty of Engineering, Bilecik Seyh Edebali University, Bilecik, Türkiye. [5]Department of Medical Biology, Faculty of Medicine, Altinbas University, Istanbul, Türkiye. [6]Department of Medical Biology, Faculty of Medicine, Istanbul Medeniyet University, Istanbul, Türkiye. [7]Department of Medical Biology, Faculty of Medicine, Bezmialem Vakif University, Istanbul, Türkiye. [8]Cigli Training and Research Hospital, Neurology Clinic, Bakircay University, İzmir, Türkiye. [9]Department of Neurology, Faculty of Medicine, Bezmialem Vakif University, Istanbul, Türkiye. [10]Basaksehir Cam and Sakura City Hospital, Neurology Clinics, Istanbul, Türkiye. [11]Biomedical Sciences Graduate Program, Institute of Graduate Studies, Altinbas University, Istanbul, Türkiye. [12]Tepecik Educational and Training Hospital, Neurology Department, University of Health Sciences, Izmir, Türkiye. ✉email: feride@metu.edu.tr; feride.severcan@altinbas.edu.tr

Autoantibodies against proteins in the postsynaptic membrane are the most important marker in diagnosing MG. A positive titer of acetylcholine receptor antibody (anti-AChR) is a feature in 85% of individuals with generalized MG and 50% of people with ocular MG[5]. Anti-MuSK (muscle-specific kinase antibody) positivity is detected in 1-10% of all patients. Hence, individuals suspected of having MG are subjected to serological tests, specifically the anti-AChR and anti-MuSK tests. If the test results are positive, additional tests may not be required[6]. However, detecting these in all patients is impossible, and specialized laboratories are also needed. In addition to these antibodies, anti-LRP4 positivity is also found in 3–25% of patients who do not have anti-AChR or anti-MuSK positivity[4]. Anti-LRP4 positivity, a test not routinely administered in clinics, has also been detected in patients with amyotrophic lateral sclerosis (ALS) and various neuroimmune disorders, including MuSK-MG. Therefore, it is not regarded as a specific diagnostic marker for MG[7].

Numerous clinical tests are employed to diagnose MG. Despite evaluations like the edrophonium test, ice pack test, serum autoantibody testing, electrophysiological assessments including repetitive nerve stimulation and single fiber electromyography (SFEMG), as well as computed tomography (CT) or magnetic resonance imaging (MRI) of the thymus, MG could still be difficult to differentiate from other clinical disorders[8]. The fluctuations in symptoms and the subtleness of clinical findings make diagnosing MG difficult. While 13% of the patients experience a delay of more than 5 years in diagnosing the disease, 26% are given nonspecific diagnoses[9]. Particularly among the elderly, the symptoms such as dysphagia, muscle fatigue, slurred speech, and ptosis (droopy eyelid) might be considered age-related and overlooked. Among elderly patients, there are instances where the symptoms of MG have been incorrectly diagnosed as stroke, Parkinson's disease, or motor neuron disease. Clinical cases where the diagnosis of MG is missed may also be encountered. The difficulty in diagnosis of the condition arises when patients have seronegative test results and display normal electrophysiological testing. A recent systematic review highlighted the challenges in diagnosing and treating MG[10].

Extensive research has been conducted for the characterization[11,12], diagnosis, pathophysiology, and treatment of MG[8,13] using biochemical, genetic, and immunologic techniques and modeling studies. Most of the characterization studies have focused on antibodies such as nicotinic AchR, MuSK, and neuromuscular junction[11,12]. Some of the studies investigated the structure-function relationship to detect changes in motor endplate, antibody-receptor interactions, and effects on the functioning of AChR and neuromuscular synaptic transmission. These investigations enhance our comprehension of the underlying pathophysiology and aid in the advancement of diagnostic and therapeutic approaches for MG. There are limited spectroscopic studies regarding MG, including ^{31}P MR Spectroscopy[14] and Proton magnetic resonance spectroscopy[15], which are mainly focused on muscle. Another study investigated the protein profile from the serum of MG patients using surface-enhanced laser desorption/ionization (SELDI) time-of-flight mass spectrometry (TOFMS)[16].

In spite of intensive efforts, the underlying mechanism of MG has still not been fully elucidated due to low accuracy, possible false positive results, low sensitivity and specificity, and problems with the availability and reliability of the existing methods[8,17]. The consensus in the MG literature is that there is a need for better more dependable diagnostic techniques[2,17].

Therefore, novel approaches that are rapid, sensitive, reliable, and relatively low-cost are essential for the characterization and diagnosis of MG disease[18]. Fourier transform infrared (FTIR) spectroscopy coupled with chemometrics meets these criteria as reported by our group for the diagnosis of bladder cancer[19], Multiple Sclerosis[20] and the rapid diagnosis of malignant pleural mesothelioma and its discrimination from lung cancer[21,22] and benign exudative effusions[22]. It has been used by others for the diagnosis of different diseases[23–25]. Blood serum is a biological fluid that is commonly used in clinical practice. It contains a lot of biomolecular information about proteins, lipids, and nucleic acids; thus, applying IR spectroscopic studies to the serum is ideal for detecting disease states. This technique, combined with chemometrics, has also been used in many serum studies as a promising analytical tool for disease diagnosis, e.g., cancers[22,23,26] onco-hematological diseases[27], Multiple Sclerosis[20,28], Osteoarthritis[29], infectious diseases such as Covid 19[30], and neurodegenerative diseases[31–33]. The attenuated total reflectance (ATR) unit of FTIR spectroscopy is preferred in disease diagnosis and characterization studies because a very small sample (for example, one drop of biofluids) is enough to obtain a clear spectrum that contains information about different functional groups belonging to lipids, proteins, carbohydrates, and nucleic acids.

This research aims to assess the feasibility of infrared spectroscopy, specifically ATR-FTIR spectroscopy combined with chemometrics, for rapid, sensitive, and cost-effective diagnosis of MG using human blood serum samples. Notably, this study aims to achieve accurate and swift MG diagnosis without relying on antibody information. Furthermore, this research represents the first investigation into the structural biomolecular foundation of MG and the determination of spectral biomarkers for the diagnosis of this disease.

Results

We conducted this study to characterize MG disease in terms of biomolecular concentration, content, and structure to understand the molecular basis of the disease. This approach will lead to determining spectral biomarkers. In addition, ATR-FTIR spectroscopy coupled with multivariate analysis methods was used to achieve an accurate and rapid novel approach for the diagnosis of MG via blood serum.

Individual and clinical characteristics of the patients

A total of 24 MG patients and age and sex-matched 42 healthy individuals in the control group were included in the study. Power analysis results for MG sample size are given in Supplementary Information (S-Text 1) and the exclusion criterion for MG patients is given the Supplementary Information (S-Text 2). The MG patients had not had any treatment—neither symptomatic therapy nor immunosuppressive/immunomodulator medications. The MG group included 9 males and 15 females. Patients ranged in age from 18 to 73 years (avg. 48.16±16.79

years), males (avg. 53.78±17.70 years), and females (avg. 44.80±15.86 years). The difference between the control and MG groups' ages was not statistically significant. (t=1.60, df=64, p>0.05). Demographic and clinical data of the healthy individuals and MG patients are given in Supplementary Tables 1 and 2 (for details, please see also the Method section).

Spectral characterization and biomarker determination

The average infrared spectra of the control (N=42) and MG (N=24) groups in 3800-3020 cm^{-1}, 3020-2800 cm^{-1}, and 1800-800 cm^{-1} regions are shown in Fig. 1A,B , and C, respectively. All spectra for the MG and control groups are given in Supplementary Figure S1. The bands of interest are labelled in the figure, and the band assignment is shown in Table 1. As seen in Figure 1 and Table 1, the serum sample contains several functional groups that belong to different biomolecules of the system, such as saturated lipids, unsaturated lipids, proteins, and nucleic acids. For a visual demonstration of the variation between the groups, the average spectra of the groups were baseline corrected and then normalized with respect to the amide A band (band 1) for the representation of the 3020-2800 cm^{-1} and 1800-800 cm^{-1} regions and with respect to CH_2 antisymmetric stretching (band 5) for 3800-3020 cm^{-1} region. Peak normalization is a method used to remove inconsistencies in the peak heights of spectra recorded under various settings and gives rough information about the spectral changes between the groups[34].

However, interactive baseline corrected raw spectra were used to accurately determine these variations between the MG and the control spectra (Fig. 2). The concentration information of the related functional groups was obtained from the area under the spectral bands according to Beer-Lambert's law. To eliminate potential experimental distortions, band area ratios were employed instead of area values, enabling precise concentration information for biomolecules, as previously documented[22,35]. The band area ratios and their corresponding interpretations employed in this study are provided in Table 2.

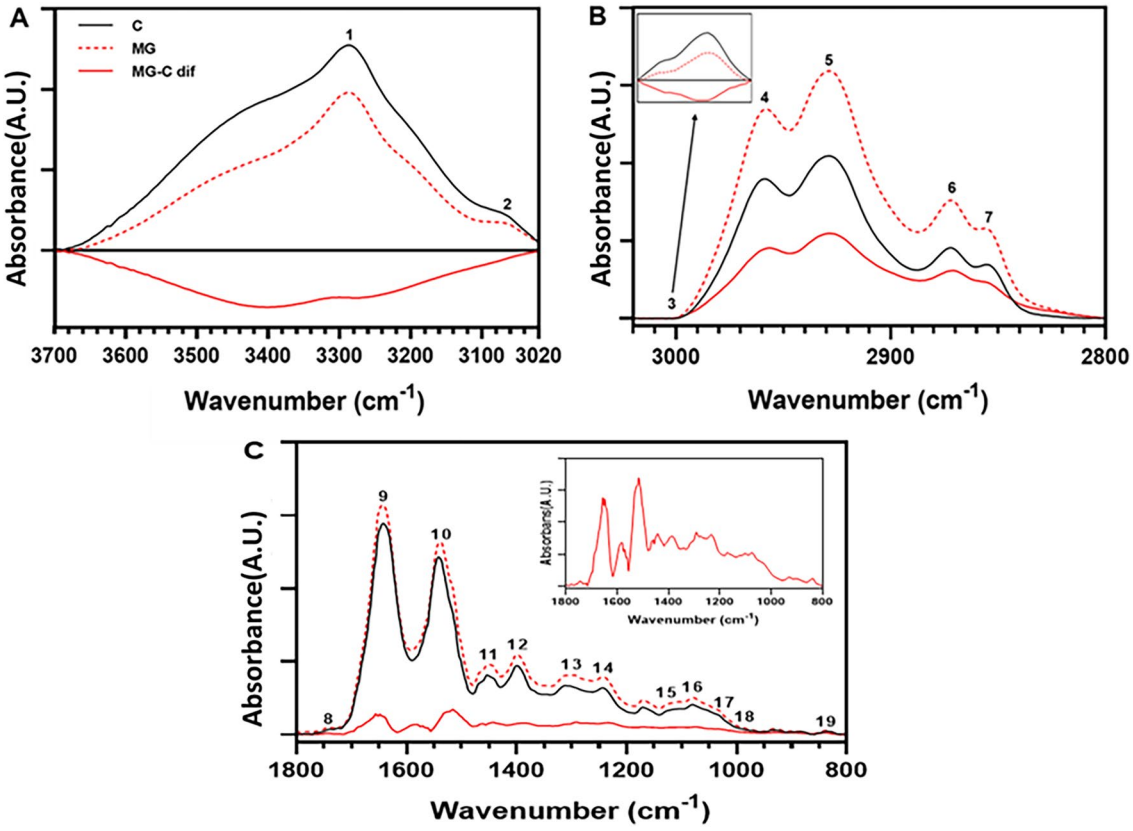

Figure 1. The average serum spectra of the control and MG groups in different regions. The average serum absorbance and difference spectra (solid red line) of the MG (N = 24) and healthy control (N = 42) groups (**A**) in the 3800–3020 cm^{-1} region. The spectra were normalized with respect to the CH_2 antisymmetric band located at 2929 cm^{-1}. (**B**) in the 3020–2800 cm^{-1} regions. The spectra were normalized with respect to the amide A band located at 3286 cm^{-1}. The enlarged olefinic band spectra are given in the sub-panel. (**C**) in the 1800–800 cm^{-1} region. The spectra were normalized with respect to the amide A band located at 3286 cm^{-1}. Enlarged difference spectra are given in the sub-panel. Difference IR spectra were obtained by subtracting the control spectrum from the MG groups' spectra. Band assignments are given in Table 1. Control was represented by C, Myasthenia Gravis by MG, and Difference spectra by MG-C dif. This figure was given for a rough visual representation of the MG-induced spectral changes. The data were generated using OPUS 8.v1 software (https://www.bruker.com/en/products-and-solutions/infrared-and-raman/opus-spectroscopy-software/downloads.html).

Band No	Wavenumber (cm^{-1})	Band assignments
1	3284	Amide A (mainly N–H stretching of proteins with contribution from the O–H stretching of polysaccharides and intermolecular hydrogen bonding of proteins and glycogen)
2	3062	Amide B band: Amide I overtone: proteins
3	3008	Olefinic H–C = C–H stretching vibration: unsaturated lipids
4	2958	CH_3 antisymmetric stretching: equal contribution from lipids and proteins
5	2929	CH_2 antisymmetric stretching: mainly lipids
6	2872	CH_3 symmetric stretching: mainly proteins
7	2854	CH_2 symmetric stretching: mainly lipids
8	1738	C = O stretching: triglycerides and cholesterol esters
9	1642	Amide I (C = O stretching in proteins [80%])
10	1541	Amide II (proteins, N–H bending [60%], and the C-N stretching [40%])
11	1454	CH_2 antisymmetric stretching: mainly lipids, with a minor contribution from proteins
12	1398	COO- symmetric stretching: fatty acids, amino acid side groups
13	1310	Amide III (protein C–N stretching (40%), N–H bending [40%] and C–C stretching [20%])
14	1241	PO_2^- antisymmetric stretching: phospholipids and nucleic acids
15	1127	Ribose: RNA
16	1080	PO_2^- symmetric stretching: nucleic acids and phospholipids
17	1030	C––O stretching: carbohydrates (glucose)
18	989	RNA backbone vibration
19	835	A-form and B-form helix conformation of DNA

Table 1. The band assignment of an FTIR spectrum for a serum sample[35–37].

The detailed numerical summary of the differences between the band area ratios of the control and MG groups and the degree of significance of these variations are given in Table S3. The orientation of the arrows signifies the trends of variations in MG biomolecules relative to the control.

As seen in Table S3, there are significant changes in some of the parameters of the MG group compared to the control group. The parameters that are significant values are depicted in Figure 2 as bar diagrams.

As seen from Table S3 and Fig. 2A.a, a significant decrease ($p<0.0001$) was observed in the amide A band in MG in comparison to the control. This band includes the contribution from the O–H stretching of polysaccharides and intermolecular hydrogen bonding of proteins (≈3284 cm^{-1}) and glycogen (around 3200 cm^{-1}). The variation in the amide A amount was obtained from the derivation of the area of the amide A located at 3280 cm^{-1} to the sum of amide I and II protein band areas and the area of the glycogen band at 1030 cm^{-1} as given in Table 2. To support this, the amide A amount was also obtained directly from the area of the amide A absorbance peak. The result revealed that the amide A amount decreases from 24.68 ± 0.3684 for the control to the 18.56 ± 0.1879 value for MG with a significance value of **** p<0,0001.

The olefinic band located at 3008 cm^{-1} monitors unsaturated lipids (Table 1). A significant decrease ($p<0.0001$) was observed in the unsaturation index (Table 2 and Fig.2A.b), which was obtained from the area ratios of the olefinic band to the saturated lipid band, thus indicating an increase in lipid peroxidation[38,39]. The increase in lipid peroxidation was supported by the decrease in the area of the olefinic band from 2.212±0.0689 for the control to 1.792±0.059 for MG ($p= 0.0065$)[40] This was further supported by a traditional biochemical test. The malondialdehyde(MDA) concentration which is one of the biochemical markers for lipid peroxidation was also increased in the MG group (1.40±0.29 µmol/L) compared to the control group (0.09±0.02 µmol/L) ($t=5.88$, $df=64$, $p<0.0001$). Significant increases for saturated lipids (***p<0,001), proteins (****p<0,0001), and DNA (****p<0,0001) concentrations were obtained for the MG group. The increase in the total protein concentration of the MG patients was also supported by the Bicinchoninic acid (BCA) assay. The protein concentration increased from 96.49±2.63 for the healthy controls to 125.82±5.14 for the MG patients ($t=5.63$, $df=64$, $p<0.0001$).

The bands located at 1241 and 1080 cm^{-1} are assigned as PO_2^- antisymmetric and symmetric stretching, respectively which have contributions from phospholipids and nucleic acids. However, the contribution of nucleic acids (DNA and RNA) to this band in blood serum is small. RNA and DNA-specific infrared absorptions were observed around 1127, 990, and 835 cm^{-1} in control serum samples as reported previously[37,41,42]. The RNA concentration obtained from the area ratios of these RNA bands revealed a significant increase. In addition, the RNA/DNA ratio was calculated by dividing the RNA band at 1127 cm^{-1} into the DNA band at 835 cm^{-1}. A significant decrease (****p<0.0001) was observed in this RNA/DNA ratio (Fig. 2A.k) indicating a potential alteration in nucleic acid metabolism in the disease state. The decrease in the RNA/DNA ratio also implies that DNA is more affected than RNA in the MG patients since the rise in DNA concentration surpasses the increase in RNA concentration. This result also suggests that the decreased RNA/DNA ratio may serve as a valuable biomarker for distinguishing MG patients from healthy individuals, providing insights into the disease's diagnostic potential.

Furthermore, PO_2^- antisym.+ sym. stretching/protein and PO_2^- sym. stretching/total lipid ratios, which were obtained by taking the area ratios of relevant functional groups[22,36], were significantly increased. Protein phosphorylation was calculated from the area ratio of the band located at 1241 cm^{-1} to the band at 2958 cm^{-1} as reported previously[36,43]. A significant increase ($p<0.0001$) was observed in this parameter for the MG group.

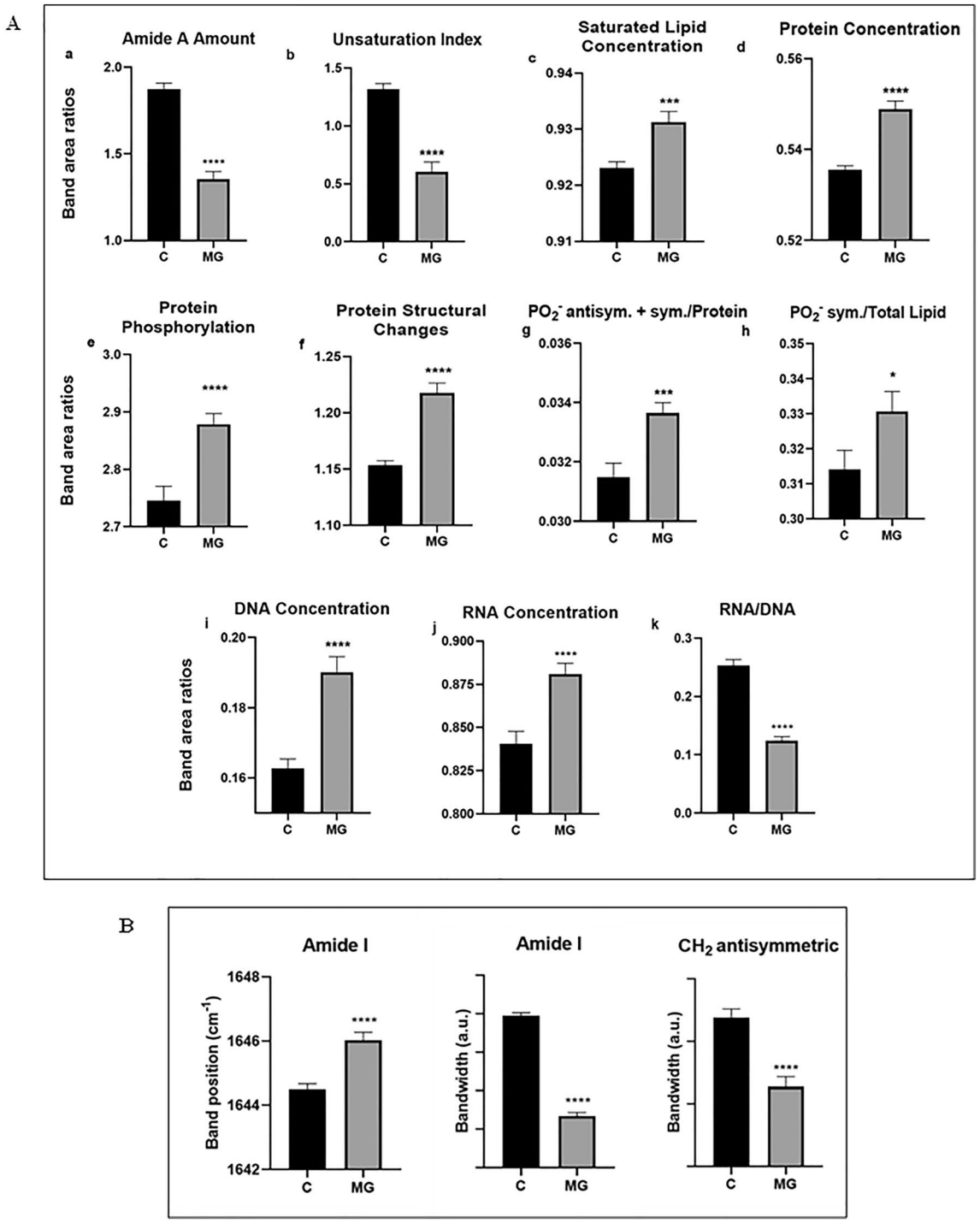

Figure 2. Visual representation of infrared spectral biomarkers for MG diagnosis: (**A**) Comparison of area ratios of MG patient serum biomolecules with the control group, (**a**) Amide A amount, (**b**) Unsaturation index, (**c**) Saturated lipid concentration, (**d**) Protein concentration, (**e**) Protein Phosphorylation, (**f**) Protein structural changes, (**g**) PO_2^- antisym. + sym./Protein, (**h**) PO_2^- sym./Total lipid, (**i**) DNA concentration, (**j**) RNA concentration, and (**k**) RNA/DNA. (**B**) Comparison of protein and lipid band position and bandwidth values (**a**) Amide I band position, (**b**) Amide I bandwidth, (**c**) CH_2 antisym. stretching bandwidth. An unpaired t-test was applied to the samples. (*: indicates that the patient group was compared with the control; *p < 0,05, **p < 0,01, ***p < 0,001), **** p < 0,0001) (C: Control group, MG: Myasthenia Gravis patient group). The figure shows, in addition to MG-induced contextual and concentration changes of serum biomolecules, a decrease in the unsaturation index, i.e., an increase in the lipid peroxidation, variations in the amide I band position and bandwidth, i.e., protein structural changes, and a decrease in the bandwidth of CH_2 antisymmetric stretching lipid band, i.e., a decrease in lipid dynamics. The graphs were generated using GraphPad Prism version 7.0 for Windows, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com.

Bant Area (A) Ratios	Definition
$A_{3286}/A_{1639+1538+1030}$	Amide A amount
A_{3008}/A_{2929}	Unsaturation index
$A_{2929}/A_{2929+2854}$	Saturated lipid concentration
A_{2929}/A_{2958}	Aliphatic chain length
$A_{1639}/A_{1639+1538}$	Protein concentration
$A_{2929+2854}/A_{1639+1538}$	Lipid/protein ratio
$A_{1738}/A_{2928+2854+1454+1738}$	Triglyceride-Cholesterol-Ester Concentration
$A_{1030}/A_{1639+1538}$	Glucose/protein
$A_{1241+1080}/A_{1639+1538}$	PO_2^- antisym. + sym. stretching/protein ratio
A_{1241}/A_{2958}	Protein phosphorylation
$A_{1080}/A_{2958+2929+2854+1454}$	PO_2^- sym. stretching/total lipid ratio
$A_{835}/A_{1241+1080}$	DNA concentration
$A_{989}/A_{989+1127}$	RNA concentration
A_{1127}/A_{835}	RNA/DNA
A_{1639}/A_{1538}	Protein structural changes

Table 2. Band area rations and their definitions.

The effect of MG on structural parameters was also investigated. As seen from Table 2 and Fig. 2A, the area of the amide I to amide II ratio, which monitors protein structural changes, increased significantly ($p<0.0001$) in the MG group compared to the control group[38]. This result was supported by the significant ($p<0.0001$) wavenumber (band location) shift and change in the bandwidth of the amide I band (Table 2 and Figure 2C.a). The bandwidth of CH_2 stretching lipid bands gives information about lipid dynamics[21]. The significant decrease ($p<0.0001$) observed in the bandwidth of the CH_2 antisymmetric stretching band indicates a significant decrease in lipid dynamics i.e., fluidity[21]. We also calculated the albumin-globulin ratio in serum from the signal intensity values of the vector normalized second derivative amide I spectral bands located at 1657 cm^{-1} and 1637 cm^{-1} as suggested by Lasch et al[31] and Diem[44]. These bands belong to alpha –helical (albumin) and beta sheets (globulin) respectively, monitor mainly secondary structural elements of these proteins. This ratio was suggested as a biomarker for diagnosis of some diseases such as bacterial infections, cancer, and some neurological diseases[31,45]. However, in our study, no significant change was detected in this ratio (p-value = 0.45) for MG disease.

All the significant spectral changes presented in Figure 2 can be used as biomarkers for MG diagnosis.

Multivariate analysis results

To develop a diagnosis method for MG using the FTIR spectra of blood serum samples, classifier models were obtained from the Principal Component Analysis (PCA) score plots at 4 different regions of vector-normalized original spectra and the second derivative of vector-normalized spectra. Better classification results were obtained from the PCA score plots of the second derivative spectra. Therefore, here, PCA score plots and the classification outcomes derived from the second derivative spectra are presented. Score plots and classification results obtained from the original spectra are given in Supplementary Figure S2 and Table S4.

Figure 3 shows the PCA score plots in the PC1-PC2 plane for 4 different spectrum regions. The 4 regions studied were defined as follows: whole region (union of the intervals 3700-2400 cm^{-1} and 1800-650 cm^{-1}); lipid region (3000-2830 cm^{-1}); protein region (1700-1500 cm^{-1}); fingerprint region (1480-1000 cm^{-1}). In all 4 cases, the control and MG samples are clustered separately. The best separation of the MG and control clusters is obtained for the protein and fingerprint regions (Figs. 3C and D). The cluster separation is also sufficiently high for the whole region (Fig. 3A) in the PC1-PC2 plane. Therefore, the best classifier performances can be obtained in these intervals. Although high classification performances were obtained for the fingerprint and whole spectral regions they are not as high as those of the protein region.

Table 3 shows the performance results of the support vector machine (SVM), discriminant, and neural network type classifiers that were used in 4 different regions of the spectrum. The classifier models were obtained using the PCA score vectors obtained from the second derivative of the vector-normalized spectra as described in the methods section. The protein region yielded the highest classification performance, with all three types of classifiers achieving perfect scores of 100% accuracy, sensitivity, and specificity. Accuracy is defined as the ratio of the number of correctly classified samples to the total number of samples as percentage. Sensitivity is the true positive rate as percentage, while specificity is the true negative rate as percentage. Such high classification performances are expectable because of the good separation of the control and MG clusters. However, although good classification performances are achieved for the fingerprint and the whole regions, the explained variances for validation scores were slightly lower than those of the calibration scores in these regions (not shown here). Therefore, classifiers modeled for the protein regions should be preferred.

Confusion tables and receiver operating characteristic (ROC) curves of the SVM classifier models (first column of Table 3) are depicted in Figure 4, to illustrate the performance of the classifier models. Here, for the sake of brevity, only the results for SVM-type classifier models are given. As seen from the third row in the figure, corresponding to the model for the protein region, all of the 42 healthy control samples and 24 MG diagnosed samples are correctly classified. The corresponding ROC curve is the ideal ROC curve for 100% sensitivity and

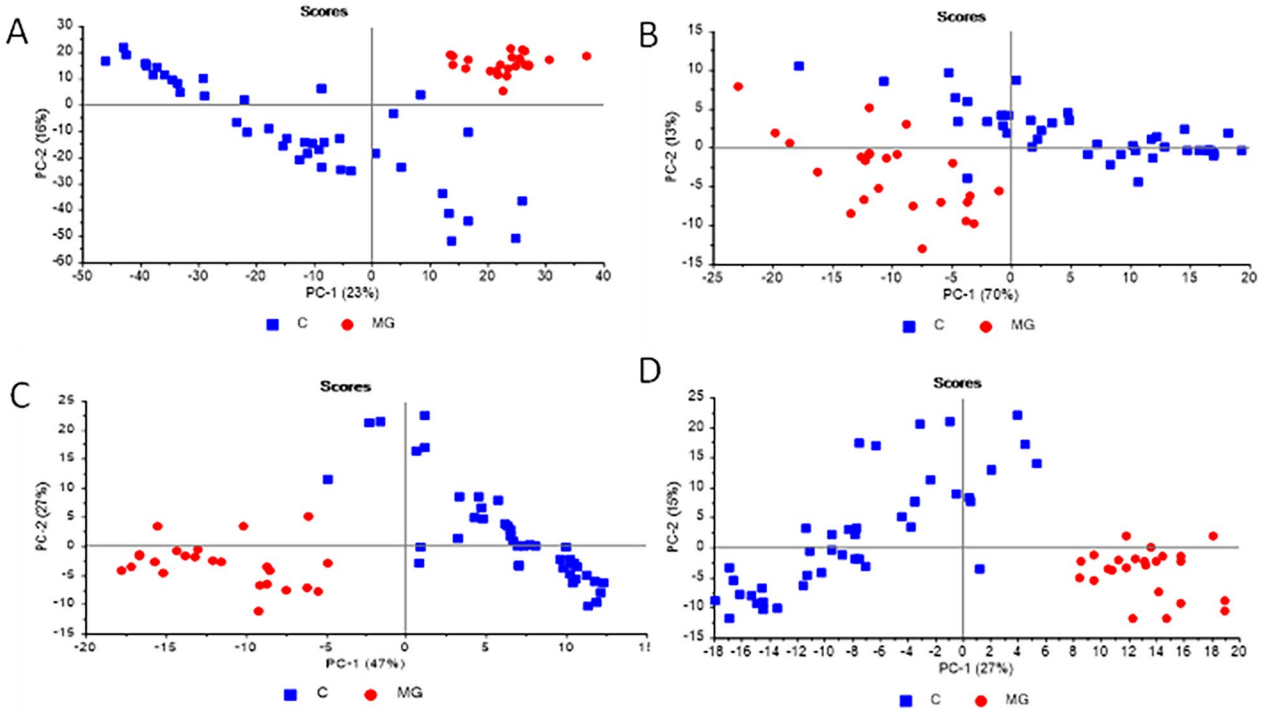

Figure 3. PCA score plots in the PC1-PC2 plane of the Control (**C**), and MG samples obtained from the second derivative of unit vector normalized spectra (**A**) for the whole region, (**B**) for the lipid region, (**C**) for the protein region, and (**D**) for the fingerprint region. PCA scores of the Control and MG samples are separated into two clusters. The figures were generated by "The Unscrambler X" Version 10.3, CAMO Software Inc., Oslo Norway(www.camo.com/unscrambler/).

	SVM	Discriminant	Neural net classifier
Whole region	98.5/95.8/100	100/100/100	97/95.8/97.6
Lipid region	98/100/97.6	97/95.8/97.6	90.9/83.3/95.2
Protein region	100/100/100	100/100/100	100/100/100
Fingerprint region	98.5/100/97.6	100/100/100	98.5/100/97.6

Table 3. Accuracy/Sensitivity/Specificity values (in percentages) of the 3 types of classifier models. Models were obtained for 4 different regions using as inputs the PCA score vectors which were obtained from the vector-normalized second derivative spectra. All of the three types of classifiers perform best in the protein region with 100% accuracy. The table lists the best results of linear and quadratic SVM, linear and quadratic discriminant, and narrow and medium type neural network classifiers. The models were generated via MATLAB Classification Learner App software.

specificity, with an area under curve (AUC) equal to 1. The first, second and fourth rows show the confusion tables and ROC curves of the classifiers for the whole, lipid and fingerprint regions. The SVM classifier for the whole region, wrongly decides 1 MG sample as control and the classifiers for the lipid and fingerprint regions decide 1 control sample as MG. AUC values being close to 1 for these classifiers is an indication of the acceptably high performance of the classifiers.

LDA and SVM type classifier models were also generated using "The Unscrambler X" software. The results and related figures are presented in the supporting material section (Please see Supplementary Text-3, Figs. S3-S5).

Discussion

This study evaluated the potential of infrared study coupled with multivariate analysis techniques as a novel approach for MG diagnosis and characterization from blood serum. Characterization studies using ATR-FTIR spectroscopy on blood serum enabled us to better understand the structural molecular basis of MG and specify spectral biomarkers for the disease diagnosis. The human serum samples were obtained from newly diagnosed MG patients who had not yet initiated any treatment and were compared with samples from a healthy control group. Serum studies that include healthy controls along with patients are important to understand the molecular basis of the diseases. Other advantages are that it is a minimally invasive procedure, and repetitive testing with blood samples for confirmation or during the patient follow-up is easier as reported in Yonar et al.[22] and

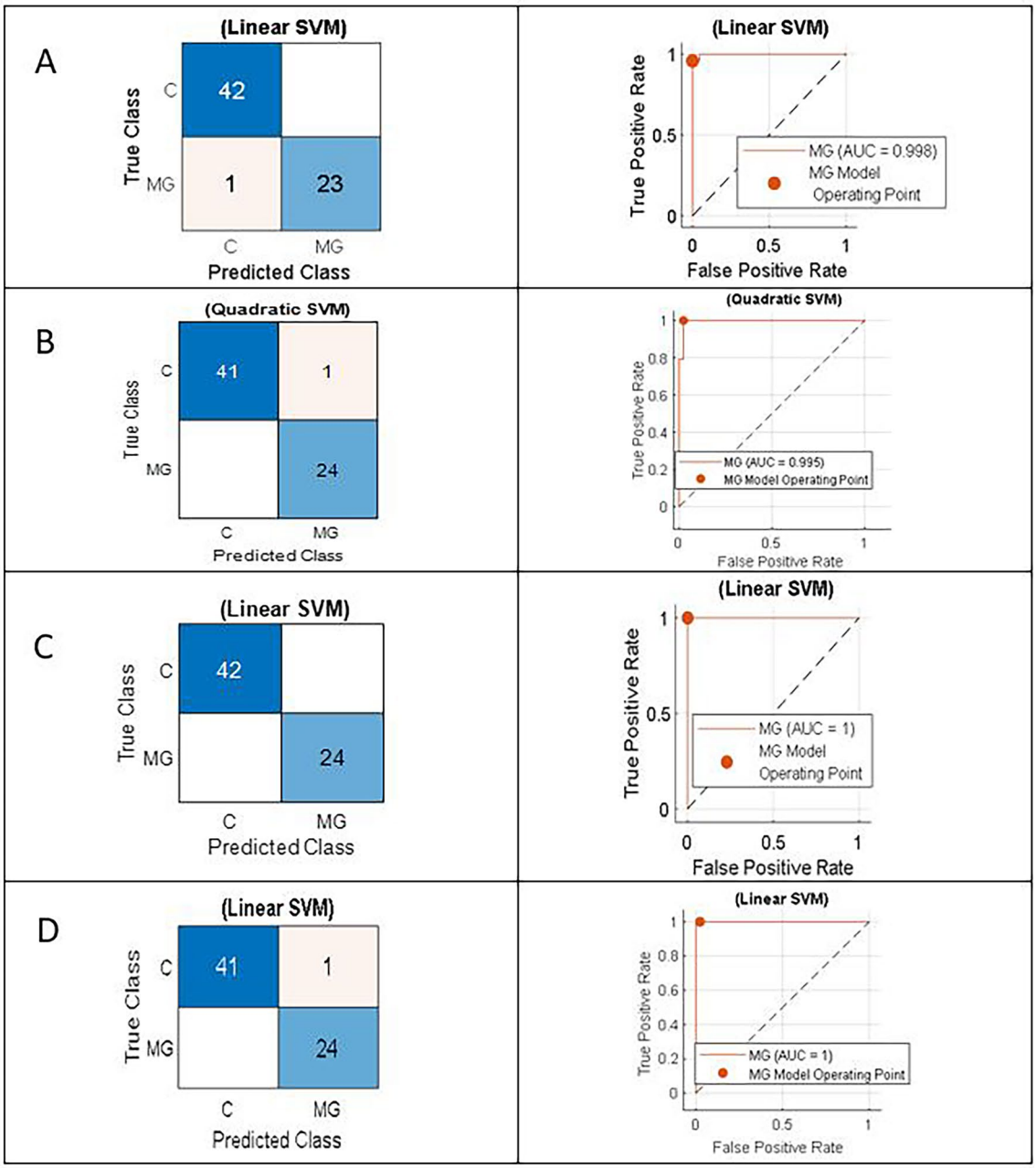

Figure 4. Confusion tables showing the number of observations of the Control and MG samples (First column) and the Receiver Operating Characteristic (ROC) (Second column) for the SVM classifier models obtained for the (**A**) whole, (**B**) lipid, (**C**) protein, and (**D**) fingerprint regions. The SVM classifier correctly predicts all 42 control samples and 24 MG samples by the classifier designed for the protein region. ROC curves show the plots of sensitivity vs. (1-specificity) for the MG mode classifiers. The models were generated via MATLAB R2023a, Classification Learner App software(The MathWorks Inc. https://www.mathworks.com).

in a recent review article[46]. FTIR spectroscopy, which has fingerprint properties, relies on the distinct absorbance bands exhibited by functional groups within molecules. Consequently, structural, and functional changes induced by pathological or environmental factors can be directly detected and monitored using this technique. As a result, FTIR spectroscopy enables the simultaneous acquisition of comprehensive information regarding the composition, concentration, and structure of macromolecules such as lipids, proteins, carbohydrates, and nucleic acids. As previously mentioned, this approach has been widely employed in numerous studies as a promising analytical tool for the diagnosis of diseases based on human serum samples[22–25,31,47–49].

The detailed characterization studies revealed that an increase in saturated lipids, a decrease in unsaturated lipids, and an elevation in lipid peroxidation were observed in MG patients (Figures 2A.c and A.b). The olefinic band located at 3008 cm^{-1} provides information about lipid peroxidation[50]. As seen in Figure 1. B, the olefinic band is one of the weaker bands and is derived from unsaturated lipids. It is known that unsaturated lipids are

more susceptible to lipid peroxidation. Therefore, this band's signal intensity or area is used as a lipid peroxidation index[40]. A decrease in this band's area corresponds to an increase in lipid peroxidation[51]. A significant decrease in the area of this band was observed in the MG group, indicating a reduction in unsaturated lipids and an increase in lipid peroxidation which was also supported by an increase in the MDA concentration of the MG patients' serum samples parallel to the experimental MG studies in the literature[52].

It is known that pathological conditions lead to metabolic alterations which increase protein levels, reactive oxygen species (ROS), oxidative stress, and, subsequently, lipid peroxidation. Enhanced lipid peroxidation is a process wherein oxidants such as free radicals attack lipids containing C=C double bonds, typically characterized by the shortening of the acyl chain length. In this process, long acyl chains are fragmented into shorter acyl chains, forming saturated lipids due to the attack on C=C double bonds[53]. It is believed that the conversion of unsaturated bonds to saturated bonds also contributes to the increase in saturated lipid levels in the patient group[39]. A dramatic increase in protein concentration was observed in the serum of MG patients ($p < 0.0001$) (Fig. 2A.d), using the variations in the area of well-known amide I and amide II protein bands, and also in the serum total protein concentration determined by BCA assay. The increase in protein concentration could be attributed to the overall increase in muscle or glycolytic enzymes, transport proteins, or other proteins such as creatine kinase. In addition, mature neuropeptides, which are released in the extracellular space and travel through the body to reach organs/tissues/cells, may contribute to the increase in protein concentration. Previous studies have also reported an increase in proteins in muscle diseases, suggesting that this increase could serve as a biomarker[54,55]. The relationship between glycogen metabolism and muscle disease has been reported in the literature[56], indicating the dysregulation of enzymes involved in glucose/glycogen metabolism, which affects the levels of glucose/glycogen in the blood. Supporting this, we observed a significant decrease in the amount of amide A, which has a contribution from proteins and glucose/glycogen. Since protein amount increases in the MG group, this decrease observed in the amide A amount is due to a significant decrease in the glucose/glycogen concentration.

Since protein concentration is high in blood serum, some studies about disease diagnosis mainly focused on proteins[31,57] and the others for protein concentration determination[58,59]. However, the weak infrared bands of lipids and nucleic acids (DNA and RNA) give valuable information in disease diagnosis from serum[60–63]. Gahan and Stroun (2010) published a study about the possible role(s) of circulating nucleic acids in plasma and serum (CNAPS) in the diagnosis and prognosis of personalized cancer. It was pointed out that some of the nucleic acids in plasma and serum released from primary tumors can act as messengers between cells and play a role in the initiation of metastases[64]. Yang et al.[61] studied the FTIR spectra of lung cancer patients' serum samples and found that the concentrations of protein, lipid, and nucleic acids were increased in the serum of cancer patients compared to the serum of healthy people. They distinguished the serum of patients with lung cancer from that of healthy people successfully due to changes in nucleic acid[61]. Li et al.[65] determined the ratio "A1243+1081/A1652+1539" to distinguish the serum of lung cancer patients from the serum of patients with benign lung nodules[65]. Cameron and colleagues conducted a study using 641 blood serum samples from brain cancer and control patients. They observed changes in the symmetric PO_2^- stretching vibrations within nucleic acids (1100–1000 cm^{-1})[66]. There are many other studies also analyzed nucleic acid vibrations from serum[67–70]. We studied these bands and observed an increase in PO_2^- antisym.+ sym. stretching/protein ratio which may be attributed to an increase in RNA and/or DNA in addition to an increase in phospholipid levels whose former was discussed above. Indeed, the DNA and RNA levels were found to be significantly increased ($p < 0.0001$) in the MG patient group (Fig. 2A.i and A.j). The increase in RNA concentration can be interpreted as the change in DNA affecting protein expression. DNA is copied into RNA by transcription. In other words, when the expression of a gene increases, the concentration of RNA in the cell also increases. In this case, the amount of protein produced due to translation will also be increased. The increase in DNA, RNA, and protein amounts in our results support each other and show that there are differences at the gene level in the MG patients compared to the controls. In a study, mRNA and IncRNAs were shown to be upregulated in the MG patients compared to the control. Upregulated mRNAs and IncRNAs have been reported to be associated with inflammatory response, lymphocyte proliferation, regulation of B-cell proliferation, and nucleic acid transcription factor activity, which correlates with MG[71].

MG is known as a classic example of a type II hypersensitivity reaction (according to the Gell and Coombs classification). This means that IgG-class autoantibodies target antigens expressed on the cell surface or in the extracellular matrix, leading to organ-specific damage. This damage may be associated with changes in DNA concentration. The observed increase in DNA concentration may originate from injured or dying cells. Damage-associated molecular patterns (DAMPs) can also trigger the production of type I interferons (IFN-I). DAMPs can be endogenous molecules such as DNA or RNA or nuclear components like cytosolic proteins. In this case, it is referred to as sterile inflammation since it does not arise from a pathogenic infection[72]. The increase in DNA concentration was supported by another study from the same group[73], which reported impaired elimination of apoptotic thymocytes that release endogenous nucleic acids due to decreased macrophages in the thymus of AChR-MG. The effect of inflammation should also be considered for MG-induced changes. Evidence shows that dysregulated inflammatory response is associated not only with MG but also with many autoimmune diseases, such as multiple sclerosis, rheumatoid arthritis, and systemic lupus erythematosus[74]. A recent review explained the relationship between MG and inflammation in detail by emphasizing that inflammation plays a significant role in MG, occurring in both generalized and ocular MG[75]. Inflammatory mediators, such as cytokines, are released by immune cells at the neuromuscular junction or thymic tissue. Infections and inflammation in the thymus contribute to MG etiology.

Differences in miRNA expression have been shown in MG[76], and it has been demonstrated that an increase occurs in transcript profiles of genes involved in cell trafficking and apoptosis in the blood[77]. Another study reported that miRNA-155, which is an innate and central regulator of adaptive responses, and miRNA-146a,

which participates in thymic myoid cell-mediated autoantigen stimulation, are overexpressed in MG[78]. The same study suggested that miRNAs' function and epigenetic modifications could be used as biomarkers for the onset and progression of MG. The significant increase observed in nucleic acid bands (1080 cm^{-1}) indicates the presence of oxidative stress in the DNA[79]. Additionally, the results of the current study revealed a significant increase ($p < 0.0001$) in protein phosphorylation, which is a crucial parameter in the regulation of various physiological processes, including signal transduction, transcription, cell survival, and cell death[36,43].

The function of biomolecules is closely related to their structures; therefore, structural changes in proteins affect their normal functions. In addition to spectral quantitative analyses, structural analysis of proteins and lipids was also performed. Since the amide I band primarily arises from C=O stretching (80%), and the amide II band from a combination of C-N stretching (40%) and N-H bending (60%), alterations in the amide I/amide II area ratio serve as an indicator of modifications in protein structure[80]. In the MG patients, a significant increase ($p < 0.0001$) in this parameter was observed. This result was also supported by the significant wavenumber shift of the amide I band position. Elevations in the concentrations of both DNA and RNA reinforce this scenario, as they can impact the process of protein synthesis.

In addition to spectral concentration analyses, lipid order and dynamics were also examined. Lipid order is a structural parameter that reflects the flexibility of lipid acyl chains[40]. It was observed that MG does not influence the structural parameter of lipid chain organization (i.e., its flexibility) but it significantly reduces lipid dynamics, in other words, lipid diffusion. Therefore, the parameter of lipid dynamics can also be used as a biomarker for MG diagnosis.

Supervised chemometrics and machine learning techniques have gained significant attention and application in diagnosing autoimmune and neurodegenerative disorders[81–84]. Herein, supervised machine-learning tools were applied as a comprehensive approach to generating a fast and reliable diagnostic tool that centers around several fingerprint spectral windows covering the range of spectrochemical variables potentially related to the onset of the disease. The best discrimination was obtained at the protein region with 100% accuracy, sensitivity, and specificity values using SVM, discriminant analysis, and neural net classifier.

It is worth noting that the present study does not address the differential diagnosis between suspected MG cases and those afflicted. Even such disorders have been excluded from the study. However, for a more comprehensive differential diagnosis, involving individuals exhibiting similar symptoms to MG but without the condition would be ideal. Nevertheless, such studies entail various clinical, ethical, and conceptual considerations that warrant attention in future MG research. Indeed, we started with the analysis of 29 MG serum samples. They were separated from the control group but showed 2-sub-clusters, one of which has 5 patients. We noticed that, in this sub-cluster of 5 patients, for two of the patients, thyroid autoantibodies (Anti-TPO and Anti-Tg) were found to be high, and they were diagnosed with autoimmune thyroiditis. For another patient, thyroid autoantibodies were not examined but an appearance compatible with thyroiditis was observed on tomography examination, and the fourth patient had a history of diabetic polyneuropathy. There is no disease information for the other MG patient. This patient may also have another disease but has not been diagnosed yet. This may mean that FTIR spectroscopy detects very small molecular changes in the blood that cannot be observed in routine analysis and by clinicians. As we also discussed previously in one of our manuscripts infrared spectroscopy may give information about the disease-induced changes long before they become visible to the clinicians[19]. These patients were not included in the study because they were in the exclusion criteria (Please see S-Text 2).

These advanced analytical approaches integrate data-driven algorithms to extract meaningful patterns and relationships from complex datasets. These methods can effectively classify and differentiate various disease states using supervised models like support vector machines, discriminant analysis and neural networks[85]. Supervised chemometrics and machine learning allow the identification of specific biomarkers, molecular signatures and disease-related patterns that may not be easily identifiable through traditional diagnostic methods. The benefits of employing supervised chemometrics and machine learning techniques include their capability to manage high-dimensional data, account for variable interactions and provide accurate predictions. Additionally, these methods can be optimized for personalized medicine, aiding in patient stratification, treatment selection and disease prognosis[86].

Conclusion

To our understanding, there have been no prior research study that have utilized spectroscopy and machine learning for the diagnosis of MG and the exploration of disease-related occurrences in biofluids. As discussed above in this text, there is the involvement of many different serum biomarkers potentially associated with MG pathogenesis. Amide A amount, unsaturation index, saturated lipid and protein concentrations, DNA and RNA concentrations, protein phosphorylation parameter, protein structural changes, PO_2^- antisym.+sym. stretching/protein, PO_2^- sym. stretching /total lipid and RNA/DNA ratios, amide I protein band's position and bandwidth, and lipid dynamics parameter can be used as biomarkers for MG diagnosis. Moreover, these potential biomarker molecules and/or metabolites are valuable messengers of early disease-related events. Principal Component Analysis (PCA), Support Vector Machine (SVM), discriminant analysis and Neural Network Classifier were applied to the infrared spectra for rapid MG diagnosis which revealed very successful discrimination of the MG disease group from the healthy control group.

This study also highlights that numerous learning algorithms can effectively operate in medical diagnostics when the appropriate data processing, and most importantly, spectral window, is identified and processed.

Regarding the study's limitations, it should be noted that the supervised learning-based diagnostic tools used in this study achieved comparable accuracies ranging from 90% to 100% in correctly predicting newly diagnosed, untreated MG patients from healthy individuals. While the current sample size is deemed sufficient, according to the power analysis result, forthcoming studies can concentrate on predicting MG using the data

features and designs of the infrared spectroscopy-based supervised diagnostic tool proposed in this study by incorporating larger patient cohorts. This may also lead to determining the antibody status of the MG patients since proof-of-concept study with 15 anti-AChR and anti-MuSK negative and 9 at least one of the anti-AChR and anti-MuSK positive MG groups, not mentioned in this manuscript, indicated some spectral differences between different antibody groups. As follow-up studies, the possibility of the separation of these two different antibody groups from each other can be studied with higher patient numbers. If this is succeeded, very rapid and cheap determination of antibodies of MG patients will be achieved. In addition, with the establishment of a spectral database with larger MG and healthy control serum spectra, it will be possible to make a diagnosis of a suspected MG patient from one drop blood serum sample by using ATR-FTIR spectroscopy coupled with chemometrics.

This rapid, inexpensive, automated, and less invasive diagnostic serum test by infrared spectroscopy coupled with chemometrics would be beneficial by reducing current diagnosis times and increasing the chance of successful treatment. Herein, supervised machine-learning tools were applied as a comprehensive approach to generating a fast and reliable diagnostic tool that centers around several fingerprint spectral windows covering the range of spectrochemical variables potentially related to the onset of the disease. The convergence of FTIR spectroscopy with machine learning technology is progressing towards clinical translation (bench-to-bed site), presenting a rapid, cost-effective, and sensitive novel approach for the diagnosis of MG.

Materials and methods

Patients

The experimental protocol of this study was approved by the Scientific Ethical Committee of the Bezmialem Vakif University, Istanbul, Türkiye (Approval no:15.10.2018-6032). The participants' informed consent was obtained under the requirements of the study ethics boards, and all associated methodologies were conducted in compliance with established guidelines and regulations for human biological material.

The power analysis result indicated a minimum sample size of N=10 for MG patients (Please see Supplementary Note 1 in supporting material) by taking the incidence of MG into account[87]. However, we studied larger patient numbers to obtain a reliable statistical analysis. The studied groups are the MG group (N= 24 patients) and age and sex-matched Control group (N= 42 healthy individuals). MG patients had not received any treatment, neither symptomatic therapy like Pyridostigmine (Mestinon®) nor immunosuppressive/immunomodulator medication. When the patients were suspected of having MG in the clinic, the MG diagnosis and follow-up protocol accepted in routine neurological clinical practice was applied. Accordingly, detailed disease characteristics and medical and family history of all patients participating in the study were questioned, and the medical treatments they had received were recorded. Then, detailed physical and neurological examinations were performed. In this context, clinical tests such as the edrophonium test or the pyridostigmine test, the ice-pack test, and the fatigue test were applied to all these patients. As electrophysiological tests, repetitive nerve stimulation (RNS) and, when necessary, single fiber electromyography (SFEMG) tests were applied. Imaging was performed with thorax CT or thorax MRI according to the suitability of the patients in terms of possible thymic pathology. In patients who had been previously definitively diagnosed with MG, the information about these tests was recorded within the scope of the study by examining the hospital records. Anti-AChR antibody levels were checked in a private laboratory that can perform these analyses from the serum of the patients at the newly diagnosed stage, and if they were negative, anti-MuSK antibody levels were determined.

Demographic data of the healthy individuals and MG patients is presented in Supplementary Table S1 and the clinical characteristics of MG patients who did not start any treatment are given in Supplementary Table S2. Exclusion criterion for MG is given in Supplementary text 2.

Preparation of specimens

Fasting blood samples of all patients were collected into EDTA-free blood collection tubes in the morning. Twenty minutes after blood collection, the blood samples were centrifuged at 3500 rpm for 10 minutes to separate the serum from the cellular components. Serum samples were then placed in cryogenic vials and kept at -80 °C until the spectroscopic and biochemical experiments.

Spectral data collection for ATR-FTIR spectroscopy

Before spectroscopic analysis, frozen serum samples were thawed and vortexed at room temperature. Afterward, they were directly put on the ATR crystal and scanned. The acquisition of ATR-FTIR spectra of all samples were collected by Perkin Elmer-Frontier FTIR spectrometer and Miracle, PIKE Zn/Se crystalline ATR unit using Spectrum One software (PerkinElmer Inc., Norwalk, Connecticut, USA). The surface of ATR crystal was first cleaned with sterile phosphate-buffered saline and then by with isopropanol. Then, it was air-dried for 2 min before recording the background interferogram. The cleanliness test was then conducted to ensure there are no signal peaks. Background measurement, showing the presence of any environmental residue on the crystal surface or in the light path, was performed on a clean surface before each measurement. It allows us to get high-quality spectra using background corrections. 1 µL samples of serum were placed on the crystal plate and dried with a mild nitrogen gas flux for 2 min to remove any excess free water. IR spectra of the samples were scanned in the 4000-650 cm^{-1} wavenumber region and 32 scans were taken for each interferogram with 4 cm^{-1} spectral resolution at 20 °C which is the constant temperature of the room where the infrared spectrometer is located.

To eliminate the contribution of molecules in the air, the background spectrum of air was recorded under identical conditions as the samples prior to each sample spectra acquisition and then subtracted automatically from the spectra.

For each sample, four randomly taken aliquots were scanned, which revealed almost identical spectra. The average of these replicates was used for further data analysis. The average spectra of 42 healthy individuals and 24 MG patients are shown in Supplementary Figures S1A and S1B, respectively.

Spectral data analysis

For characterization studies, the quantitative spectral analyses (including spectral band position, bandwidth, and band area ratio analyses) were calculated from the baseline corrected average raw IR spectra using OPUS 5.5 software (Bruker Optics, GmbH, Germany). The average spectra of 42 healthy control and 24 MG samples were used in visual representations. The difference spectra were obtained using the substruction function of the OPUS 5.5 software, which allows the control spectra to be subtracted from the MG spectra. The whole spectra were given in three regions: 3800-3020 cm^{-1}, 3020-2800 cm^{-1}, and 1800-800 cm^{-1}. The spectra were baseline-corrected with the interactive baseline function of OPUS 5.5 software for each region. Then the spectra vector was normalized with respect to the amide A band located at 3286 cm^{-1} for the representation of 3020-2800 cm^{-1} and 1800-800 cm^{-1} regions and with respect to the CH_2 asymmetric band located at 2929 cm^{-1} for the representation of the 3800-3020 cm^{-1} region.

The same software was used to determine band area values and band positions (frequency/ wavenumber). Integrated areas were calculated using the baseline corrected averaged IR spectra according to integration method B, which is implemented in the same data acquisition software package. The left and right baseline points of the bands were determined from the second derivative of the absorbance bands. The band positions were measured according to the center of mass of the width at 75% of band intensity and confirmed with second derivatives of the absorbance peaks. The bandwidth value of the CH_2 antisymmetric stretching band was measured as the width at 75% of the band intensity of the absorption spectra in terms of cm^{-1}.

Biochemical assays

To perform the validation of the alterations in the levels of protein concentration and lipid peroxidation of the MG patients, bicinchoninic acid (BAC) assay and thiobarbituric acid reactive substance (TBARS) assay were performed, respectively.

To determine the total protein concentration in the samples, a commercial BAC assay kit (Pierce BCA Protein Assay Kit, Thermo Fisher Scientific, Boston, MA, USA) was used as described in the manufacturer's instructions. Briefly, the serum samples stored at -80°C were equilibrated to room temperature. After applying the necessary dilutions, the serum samples were mixed with the working solution on a well plate and incubated at 37 °C for 30 minutes. Finally, the absorbance (OD) of each well was measured at 562 nm by Multiskan GO (Thermo Fisher Scientific, Boston, MA, USA). The protein concentration of each sample was determined according to the standard curve and compared to the control samples.

To determine the malondialdehyde (MDA) concentration for measuring the level of lipid peroxidation, a commercial calorimetric TBARS microplate assay kit (Biorbyt Ltd, Cambridge, UK) was used as described in the manufacturer's instructions. Briefly, the serum samples stored at – 80 °C were equilibrated to room temperature. After mixing the samples with assay buffer, they were centrifuged at 12000 x *g* for 10 minutes at 4 °C. The supernatants were collected in new microcentrifuge tubes and a dye reagent was added. Finally, absorbances (OD) of each well were measured at 535 nm by Multiskan GO (Thermo Fisher Scientific, Boston, MA, USA). The levels of MDA concentration for each sample were determined as µmol/L using a standard curve.

Multivariate analysis methods

To discriminate MG samples from control samples, supervised and unsupervised multivariate analysis methods were employed. Principal component analysis (PCA), as an unsupervised multivariate analysis method, is usually used in preliminary analysis to observe vectors representing each spectrum. Most of the variations in the original multidimensional space are preserved in the first few dimensions of the transformed space. The coordinates of the transformed space are the principal components and the plot obtained is called the scores plot. The first principal component (PC) has the highest variation, and the variations of the following PCs are in decreasing order[88].

In the current study, PCA was performed on "The Unscrambler X," Version 10.3, (CAMO Software Inc., Oslo, Norway). The input to The Unscrambler X was baseline corrected 66 spectra, 42 of which represented the control samples and 24 of which represented the MG samples. The spectra covered the wavenumber range 4000-650 cm^{-1}. PCA was performed over the intervals 3000-2830 cm^{-1} (lipid), 1700-1500 cm^{-1} (protein), 1480-1000 cm^{-1} (fingerprint), and the intervals 3700-2400 cm^{-1} and 1800-650 cm^{-1} juxta-positioned side by side. The last interval mentioned covers the whole region excluding the noisy regions above 3700 cm^{-1} and between 2400 cm^{-1} and 1800 cm^{-1}. It is called "whole" in this manuscript. The same wavenumber intervals were used in the supervised classifier designs. PCA was performed on both the unit vector normalized original spectra and the second derivative spectra obtained from the unit vector normalized spectra. The wavenumber intervals for vector normalization and second derivative operations for each region were wider than the PCA interval as follows: For the lipid region, this interval was the 3700-2400 cm^{-1} region; for the protein and fingerprint region, it was 1800-650 cm^{-1}; and 4000-650 cm^{-1} region for the whole region. For the second derivative, Savitzky-Golay filtering with 3rd-order polynomial and 9-point smoothing (4 points on each side) was employed. PCA was performed using the NIPALS algorithm, with mean-centered and standard deviation normalized data, and with full cross-validation.

In this study, three types of supervised classification methods (namely discriminant analysis (DA), support vector machine (SVM) and neural network (NN) classifier) were employed for the classification of MG and healthy control samples. For this purpose, MATLAB Classification Learner App was used. The classification app, in addition to the classifier model, provides a confusion table, true positive and false negative rates, positive

prediction and false discovery rates and receiver operating characteristic (ROC) curve as outputs. Classifier models were obtained for different types of inputs over the spectral ranges defined above for PCA. The inputs to the classification app were the vector normalized spectra with known classes or the corresponding second derivative spectra. 50-fold cross-validation (CV) was used to assess the model performances. PCA was enabled and the first 7 components of the score vectors were used in obtaining the models. Any new spectrum with an unknown class can be evaluated using the model generated by the app to identify its class.

The principles of the three types of classifiers used in this work are briefly discussed below:

Discriminant analysis

The training samples in this case are assumed to be normally distributed. The optimum boundary between the two groups is determined in such a way that the probability of making an error is minimized. For linear discriminant analysis (LDA) the covariance of the class distributions is assumed to be the same. For quadratic discriminant analysis (QDA) the covariances are assumed to be different. In this case, the boundary between the classes is not linear but quadratic.

Support vector machine (SVM)

This classifier determines a hyperplane that separates two different classes in such a way that the closest points in each class to this hyperplane are at equal distances. Those points are called support vectors, and the optimum hyperplane has the highest distance to the support vectors. If the two classes cannot be separated by a hyperplane, then a hyperplane that separates most of the training samples is generated. If the two classes are not linearly separable, different nonlinear kernels can be used for the optimum separation of the classes[22].

Neural network classifier

Artificial neural networks consist of layers of elementary neurons that imitate mathematically the operation of a biological neuron. In general, there is one input layer, one output layer and one or more hidden layers in between. Each neuron sums the weighted outputs of the previous layer and passes the sum through a nonlinear activation function to its output. Neural networks are commonly used for classification and regression applications[89]. Neural networks are trained by repeatedly introducing the training samples to the input and using an algorithm called the backpropagation algorithm, which adjusts the weights and biases of each neuron. For classification applications, fully connected neural networks with different layer sizes are used. In this work, networks with a single hidden layer of varied sizes were used.

Statistical data analysis

To evaluate the significance of differences between the control and MG groups' spectral changes, two-tailed unpaired Student's t-tests with Welch correction were performed in GraphPad Prism 8.0.2 software (La Jolla, CA, USA). The results are presented as the mean ± SD and the degree of significance was recognized as *p<0,05, **p<0,01, ***p<0,001.

Ethics approval and consent to participation

The experimental protocol of this study was approved by the Istanbul Bezmialem Foundation University Ethics Committee (Approval No: 15.10.2018-6032). Informed consent was obtained from each participant throughout the duration of the study.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 6 February 2024; Accepted: 2 July 2024
Published online: 20 August 2024

References

1. Bubuioc, A. M., Kudebayeva, A., Turuspekova, S., Lisnic, V. & Leone, M. A. The epidemiology of myasthenia gravis. *J. Med. Life* **14**, 7 (2021).
2. Sieb, J. P. Myasthenia gravis: An update for the clinician. *Clin. Exp. Immunol.* **175**, 408–418 (2014).
3. Suresh, A. B. & Asuncion, R. M. D. Myasthenia Gravis. *StatPearls* (2023).
4. Yavuz, Z. *Türkiye Cumhuriyeti Ankara Üniversitesi Tip Fakültesi Myastenia Gravis klinik ve demografik özellikleri.*
5. Lennon, V. A. Serologic profile of and distinction from myasthenic myasthenia gravis the lambert-eaton syndrome. *Neurology* **48**, 23S (1997).
6. Gilhus, N. E. *et al.* Myasthenia gravis. *Nat. Rev. Dis. Primers.* **5**, 1–19 (2019).
7. Frykman, H., Kumar, P. & Oger, J. Immunopathology of autoimmune myasthenia gravis: Implications for improved testing algorithms and treatment strategies. *Front. Neurol.* **11**, 596621 (2020).
8. Rousseff, R. T. Diagnosis of myasthenia gravis. *J. Clin. Med.* **10**, 1736 (2021).
9. Gilhus, N. E. *et al.* Myasthenia gravis—Autoantibody characteristics and their implications for therapy. *Nat. Rev. Neurol.* **12**, 259–268 (2016).
10. Yangin, M. N., Zorlu, Y. & Severcan, F. Diagnosis and treatment methods of autoimmune myasthenia gravis: A systematic review. *Aurum J. Health Sci.* **4**, 104–116 (2022).
11. Fuchs, S., Aricha, R., Reuveni, D. & Souroujon, M. C. Experimental autoimmune myasthenia gravis (EAMG): From immunochemical characterization to therapeutic approaches. *J. Autoimmun.* **54**, 51–59 (2014).
12. Takata, K. *et al.* Characterization of pathogenic monoclonal autoantibodies derived from muscle-specific kinase myasthenia gravis patients. *JCI Insight* https://doi.org/10.1172/jci.insight.127167 (2019).

13. Sánchez-Tejerina, D. *et al.* New targeted agents in myasthenia gravis and future therapeutic strategies. *J. Clin. Med.* **11**, 6394 (2022).
14. Ko, S. F. *et al.* 31P MR spectroscopic assessment of muscle in patients with myasthenia gravis before and after thymectomy: Initial experience. *Radiology* **247**, 162–169 (2008).
15. Lavrnic, D. *et al.* Proton magnetic resonance spectroscopy of the intrinsic tongue muscles in patients with myasthenia gravis with different autoantibodies. *J. Neurol. Sci.* **302**, 25–28 (2011).
16. Cheng, C. *et al.* Serum protein profiles in myasthenia gravis. *Ann. Thorac. Surg.* **88**, 1118–1123 (2009).
17. Gastaldi, M. *et al.* Improving laboratory diagnostics in myasthenia gravis. *Expert Rev. Mol. Diagn.* **21**, 579–590 (2021).
18. Skeie, G. O. *et al.* Guidelines for treatment of autoimmune neuromuscular transmission disorders. *Eur. J. Neurol.* **17**, 893–902 (2010).
19. Gok, S. *et al.* Bladder cancer diagnosis from bladder wash by Fourier transform infrared spectroscopy as a novel test for tumor recurrence. *J. Biophotonics* **9**, 967–975 (2016).
20. Yonar, D., Ocek, L., Tiftikcioglu, B. I., Zorlu, Y. & Severcan, F. Relapsing-remitting multiple sclerosis diagnosis from cerebrospinal fluids via fourier transform infrared spectroscopy coupled with multivariate analysis. *Sci. Rep.* https://doi.org/10.1038/s41598-018-19303-3 (2018).
21. Abbas, S. *et al.* Diagnosis of malignant pleural mesothelioma from pleural fluid by fourier transform-infrared spectroscopy coupled with chemometrics. *J. Biomed. Opt.* **23**, 1 (2018).
22. Yonar, D. *et al.* Rapid diagnosis of malignant pleural mesothelioma and its discrimination from lung cancer and benign exudative effusions using blood serum. *Biochim. Biophys. Acta Mol. Basis Dis.* **1868**, 166473 (2022).
23. Sala, A. *et al.* Biofluid diagnostics by FTIR spectroscopy: A platform technology for cancer detection. *Cancer Lett.* **477**, 122–130 (2020).
24. Heraud, P. *et al.* Infrared spectroscopy coupled to cloud-based data management as a tool to diagnose malaria: A pilot study in a malaria-endemic country. *Malar. J.* https://doi.org/10.1186/s12936-019-2945-1 (2019).
25. Naseer, K., Ali, S., Mubarik, S., Zajif Hussain, S. & Qazi, J. Use of ATR-FTIR for detection of salmonella typhi infection in human blood sera. *Infrared Phys. Technol.* **110**, 103473 (2020).
26. Wang, R. & Wang, Y. Fourier transform infrared spectroscopy in oral cancer diagnosis. *Int. J. Mol. Sci.* **22**, 1–21 (2021).
27. Telnaya, E. A. *et al.* Infrared spectroscopy of blood serum from patients with oncohematological diseases. *Biophysics (Russian Federation)* **65**, 981–986 (2020).
28. Kołodziej, M. *et al.* Spectral signature of multiple sclerosis. Preliminary studies of blood fraction by ATR FTIR technique. *Biochem. Biophys. Res. Commun.* **593**, 40–45 (2022).
29. Aziz, A. A., Selvaratnam, V., Fikri, Y. F. B. A., Sani, M. S. A. & Kamarul, T. Diagnosis of osteoarthritis at an early stage via infrared spectroscopy combined chemometrics in human serum: A pilot study. *Processes* **11**, 404 (2023).
30. Zhang, L. *et al.* Fast screening and primary diagnosis of COVID-19 by ATR-FT-IR. *Anal. Chem.* **93**, 2191–2199 (2021).
31. Lasch, P., Beekes, M., Fabian, H. & Naumann, D. *Antemortem Identification of Transmissible Spongiform Encephalopathy (TSE) from Serum by Mid-infrared Spectroscopy* (Wiley Online Library, 2006). https://doi.org/10.1002/9780470027325.s8925.
32. Miller, L. M. Infrared spectroscopy and imaging for understanding neurodegenerative protein-misfolding diseases. In *Vibrational Spectroscopy in Protein Research: From Purified Proteins to Aggregates and Assemblies* (ed. Miller, L. M.) 121–142 (Elsevier, 2020). https://doi.org/10.1016/B978-0-12-818610-7.00005-0.
33. Kavungal, D. *et al.* Artificial intelligence-coupled plasmonic infrared sensor for detection of structural protein biomarkers in neurodegenerative diseases. *Sci. Adv.* **9**, eadg9644 (2023).
34. Bozkurt, O. *et al.* Structural alterations in rat liver proteins due to streptozotocin-induced diabetes and the recovery effect of selenium: Fourier transform infrared microspectroscopy and neural network study. *J. Biomed. Opt.* **17**, 0760231 (2012).
35. Elibol, B., Severcan, M., Jakubowska-Dogru, E., Dursun, I. & Severcan, F. The structural effects of vitamin A deficiency on biological macromolecules due to ethanol consumption and withdrawal: An FTIR study with chemometrics. *J. Biophotonics* https://doi.org/10.1002/jbio.202100377 (2022).
36. Dogan, A., Gurbanov, R., Severcan, M. & Severcan, F. CoronaVac (Sinovac) COVID-19 vaccine-induced molecular changes in healthy human serum by infrared spectroscopy coupled with chemometrics. *Turk J. Biol.* **45**, 549–558 (2021).
37. Whelan, D. R. *et al.* Monitoring the reversible B to A-like transition of DNA in eukaryotic cells using fourier transform infrared spectroscopy. *Nucl. Acids Res.* **39**, 5439–5448 (2011).
38. Ustaoglu, S. G. *et al.* Biomolecular changes and subsequent time-dependent recovery in hippocampal tissue after experimental mild traumatic brain injury. *Sci. Rep.* **11**, 1–13 (2021).
39. Algburi, A. F., Dursun, I. & Garip Ustaoglu, S. The investigation of the effects of postnatal alcohol exposure on molecular content and antioxidant capacity of mice liver tissue. *Life Sci.* **310**, 121102 (2022).
40. Severcan, F., Gorgulu, G., Gorgulu, S. T. & Guray, T. Rapid monitoring of diabetes-induced lipid peroxidation by Fourier transform infrared spectroscopy: Evidence from rat liver microsomal membranes. *Anal. Biochem.* **339**, 36–40 (2005).
41. Naumann, D. Ft-infrared and FT-raman spectroscopy in biomedical research. *Appl. Spectrosc. Rev.* **36**, 239 (2001).
42. Delrue, C. *et al.* Infrared spectroscopy: A new frontier in hematological disease diagnosis. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms242317007 (2023).
43. Ricciardi, V. *et al.* Study of SH-SY5Y cancer cell response to treatment with polyphenol extracts using FT-IR spectroscopy. *Biosensors (Basel)* **7**, 57 (2017).
44. Diem, M. Comments on recent reports on infrared spectral detection of disease markers in blood components. *J. Biophotonics* https://doi.org/10.1002/jbio.201800064 (2018).
45. Suh, B. *et al.* Low albumin-to-globulin ratio associated with cancer incidence and mortality in generally healthy adults. *Ann. Oncol.* **25**, 2260–2266 (2014).
46. Schillebeeckx, E., van Meerbeeck, J. P. & Lamote, K. Clinical utility of diagnostic biomarkers in malignant pleural mesothelioma: A systematic review and meta-analysis. *Eur. Respir. Rev.* **30**, 210057 (2021).
47. Mordechai, S., Shufan, E., Porat Katz, B. S. & Salman, A. Early diagnosis of Alzheimer's disease using infrared spectroscopy of isolated blood samples followed by multivariate analyses. *Analyst* **142**, 1276–1284 (2017).
48. Wang, X., Shen, X., Sheng, D., Chen, X. & Liu, X. FTIR spectroscopic comparison of serum from lung cancer patients and healthy persons. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **122**, 193–197 (2014).
49. Huber, M. *et al.* Infrared molecular fingerprinting of blood-based liquid biopsies for the detection of cancer. *Elife* https://doi.org/10.7554/eLife.68758 (2021).
50. Halliwell, B., Chirico, S., Crawford, M. A., Bjerve, K. S. & Gey, K. F. Lipid peroxidation: Its mechanism, measurement, and significance. *Am. J. Clin. Nutr.* **57**, 715S (1993).
51. Sills, R. H., Moore, D. J. & Mendelsohn, R. Erythrocyte peroxidation: Quantitation by Fourier transform infrared spectroscopy. *Anal. Biochem.* **218**, 118–123 (1994).
52. Chen, W. *et al.* The decreased expression of thioredoxin-1 in brain of mice with experimental autoimmune myasthenia gravis. *Neuromuscul. Disord.* **24**, 726–735 (2014).
53. Ayala, A., Muñoz, M. F. & Argüelles, S. Lipid peroxidation: production, metabolism, and signaling mechanisms of malondialdehyde and 4-Hydroxy-2-Nonenal. *Oxid. Med. Cell. Longev.* **2014**, 360438 (2014).
54. Hathout, Y. *et al.* Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7153–7158 (2015).

55. Ralbovsky, N. M., Dey, P., Galfano, A., Dey, B. K. & Lednev, I. K. Diagnosis of a model of Duchenne muscular dystrophy in blood serum of mdx mice using Raman hyperspectroscopy. *Sci. Rep.* https://doi.org/10.1038/s41598-020-68598-8 (2020).
56. Stapleton, D. I. *et al.* Dysfunctional muscle and liver glycogen metabolism in mdx dystrophic mice. *PLoS One* **9**, e91514 (2014).
57. Nabers, A. *et al.* Amyloid-β-secondary structure distribution in cerebrospinal fluid and blood measured by an immuno-infrared-sensor: A biomarker candidate for Alzheimer's disease. *Anal. Chem.* **88**, 2755–2762 (2016).
58. Perez-Guaita, D. *et al.* Protein determination in serum and whole blood by attenuated total reflectance infrared spectroscopy. *Anal. Bioanal. Chem.* **404**, 649–656 (2012).
59. Spalding, K. *et al.* Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance—Fourier transform infrared (ATR-FTIR) spectroscopy. *Vib. Spectrosc.* **99**, 50–58 (2018).
60. Mabed, M. *et al.* The Potential utility of fourier transform infrared spectroscopy for the diagnosis and the prognosis of non-hodgkin lymphoma. *Hematol. Transfus. Int. J.* **11**, 1 (2023).
61. Yang, X. *et al.* Diagnosis of lung cancer by ATR-FTIR spectroscopy and chemometrics. *Front. Oncol.* https://doi.org/10.3389/fonc.2021.753791 (2021).
62. Rai, V. *et al.* Serum-based diagnostic prediction of oral submucous fibrosis using FTIR spectrometry. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **189**, 322 (2018).
63. Chen, X. *et al.* Characterization of microRNAs in serum: A novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* **18**, 997 (2008).
64. Gahan, P. B. & Stroun, M. The biology of circulating nucleic acids in plasma and serum (CNAPS). In: Kikuchi, Y., Rykova, E. (eds) Extracellular Nucleic Acids. Nucleic Acids and Molecular Biology, vol 25. Springer, Berlin, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12617-8_10.
65. Li, H. *et al.* Comparison of serum from lung cancer patients and from patients with benign lung nodule using FTIR spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **306**, 123596 (2024).
66. Cameron, J. M. *et al.* Stratifying brain tumour histological sub-types: The application of ATR-FTIR serum spectroscopy in secondary care. *Cancers (Basel)* **12**, 1710 (2020).
67. Sala, A. *et al.* Rapid analysis of disease state in liquid human serum combining infrared spectroscopy and "digital drying". *J. Biophotonics* https://doi.org/10.1002/jbio.202000118 (2020).
68. Sheng, D. *et al.* Distinction of leukemia patients' and healthy persons' serum using FTIR spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **101**, 228 (2013).
69. Guleken, Z. *et al.* FTIR- based serum structure analysis in molecular diagnostics of essential thrombocythemia disease. *J. Photochem. Photobiol. B* **245**, 112734 (2023).
70. Ghimire, H., Venkataramani, M., Bian, Z., Liu, Y. & Perera, A. G. U. ATR-FTIR spectral discrimination between normal and tumorous mouse models of lymphoma and melanoma from serum samples. *Sci. Rep.* https://doi.org/10.1038/s41598-017-17027-4 (2017).
71. Zhang, F., Liu, G., Bu, Y., Ma, X. & Hao, J. Expression profile of long noncoding RNAs and mRNAs in peripheral blood mononuclear cells from myasthenia gravis patients. *J. Neuroimmunol.* **299**, 124–129 (2016).
72. Payet, C. A. *et al.* Myasthenia gravis: An acquired interferonopathy?. *Cells* **11**, 1218 (2022).
73. Payet, C. A. *et al.* Central role of macrophages and nucleic acid release in myasthenia gravis thymus. *Ann. Neurol.* **93**, 643–654 (2023).
74. Duan, L., Rao, X. & Sigdel, K. R. Regulation of Inflammation in autoimmune disease. *J. Immunol. Res.* **2019**, 1–2 (2019).
75. Huda, R. Inflammation and autoimmune myasthenia gravis. *Front. Immunol.* https://doi.org/10.3389/fimmu.2023.1110499 (2023).
76. Sabre, L., Punga, T. & Punga, A. R. Circulating miRNAs as potential biomarkers in myasthenia gravis: Tools for personalized medicine. *Front. Immunol.* https://doi.org/10.3389/fimmu.2020.00213 (2020).
77. Park, K. H., Jung, J., Lee, J. H. & Hong, Y. H. Blood transcriptome profiling in myasthenia gravis patients to assess disease activity: A pilot RNA-seq study. *Exp. Neurobiol.* **25**, 40–47 (2016).
78. Leon-Sarmiento, F. E., Leon-Ariza, J. S., Prada, D., Leon-Arizad, D. S. & Rizzo-Sierra, C. V. Sensory aspects in myasthenia gravis: A translational approach. *J. Neurol. Sci.* **368**, 379–388. https://doi.org/10.1016/j.jns.2016.07.014 (2016).
79. Mihoubi, W., Sahli, E., Gargouri, A. & Amiel, C. FTIR spectroscopy of whole cells for the monitoring of yeast apoptosis mediated by p53 over-expression and its suppression by Nigella sativa extracts. *PLoS One.* **12**, e0180680 (2017).
80. Garip Ustaoğlu, Ş, Kaygusuz, H., Bilgin, M. D. & Severcan, F. Novel approaches for COVID-19 diagnosis and treatment: A non-systematic review. *Turk. J. Biol.* **45**, 358 (2021).
81. Mirzaei, G., Adeli, A. & Adeli, H. Imaging and machine learning techniques for diagnosis of Alzheimer's disease. *Rev. Neurosci.* **27**, 857–870 (2016).
82. Aich, S., Choi, K.-W., Pradhan, P. M., Park, J. & Kim, H.-C. Prediction of neurodegenerative diseases based on gait signals using supervised machine learning techniques. *Adv. Sci. Lett.* **24**, 1974–1978 (2018).
83. Callery, E. L. & Rowbottom, A. W. Vibrational spectroscopy and multivariate analysis techniques in the clinical immunology laboratory: A review of current applications and requirements for diagnostic use. *Appl. Spectrosc. Rev.* **57**, 411–440 (2021).
84. dos Santos, R. F., Paraskevaidi, M. & Lima, K. M. G. A review of biospectroscopy coupled with chemometrics for Alzheimer's disease diagnosis. *Appl. Spectrosc. Rev.* https://doi.org/10.1080/05704928.2023.2165091 (2023).
85. Lussier, F., Thibault, V., Charron, B., Wallace, G. Q. & Masson, J. F. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *TrAC Trends Anal. Chem.* **124**, 115796 (2020).
86. Kumar, Y., Koul, A., Singla, R. & Ijaz, M. F. Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda. *J. Ambient Intell. Humaniz. Comput.* **14**, 8459–8486 (2023).
87. Chow, S.-C., Jones Senior Director, B. & Liu Professor, J. *Editor-in-Chief Biostatistics Series Editors.* (2008).
88. Wang, L. & Mizaikoff, B. Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy. *Anal. Bioanal. Chem.* **391**, 1641–1654 (2008).
89. Hajjo, R., Sabbah, D. A., Bardaweel, S. K. & Tropsha, A. Identification of tumor-specific MRI biomarkers using machine learning (ML). *Diagnostics (Basel)* **11**, 742 (2021).

Acknowledgements

We would like to thank The Scientific and Technological Research Institution of Türkiye (TUBITAK) and Altinbas University for their support in conducting this study. We would like to thank Cengiz Beşoğul and Nihal Simşek Ozek for their contribution to sample size determination with power analysis.

Author contributions

Conceptualization: FS, YZ. Sample collection: BIT, EG, FK. Spectra collection: MNY, RG. IR spectral data analysis: IO, FK, AD. Multivariate data analysis: MS, RG. Biochemical analysis: BE. Interpretation: All authors Visualization: FS, IO, AD, MS, RG. Supervision: FS Writing—original draft: FS. Writing—review & editing: MS, RG, IO.

Funding
Scientific and Technical Research Council of Türkiye with Project no: TUBITAK-1003 SBAG-Project numbers: 218S986, 218S987, 218S988, One of the authors (MNY) had a fellowship from TUBITAK-1003 (SBAG-Project numbers: 218S986) and Altinbas University AYP2021-3 projects.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-66501-3.

Correspondence and requests for materials should be addressed to F.S.