Electronic Journal of Mathematical Analysis and Applications
Vol. 9(2) July 2021, pp. 265-272.
ISSN: 2090-729X(online)
http://math-frac.org/Journals/EJMAA/

APPLICATION OF SUPPORT VECTOR MACHINES TO DIAGNOSIS OF LATE-ONSET GLUTARIC ACIDEMIA TYPE 2

B. C. ASLAN

Abstract. Glutaric Acidemia Type 2 (GA2), or Multiple Acyl-CoA Dehydrogenation Deficiency (MADD), is a genetic metabolic disorder affecting amino acid, fatty acid, and choline mechanisms. While most cases present themselves at birth or at an early age, it is also quite possible to have an onset well into adulthood. For these late-onset patients, the road to diagnosis is often long, painful, and frustrating. In addition, due to late diagnosis they can also suffer from long-lasting effects of their worsening symptoms. The goal of this work is to use support vector machines to detect patterns to aim in the diagnostic process of late-diagnosed GA2 patients.

1. Introduction

Glutaric Acidemia Type 2 (GA2), or Multiple Acyl-CoA Dehydrogenation Deficiency (MADD), is a very rare genetic metabolic disorder affecting amino acid, fatty acid, and choline mechanisms, see [4]. It is an autosomal-recessive disorder, which means that when both parents carry a defective gene and both parents pass a copy of the defective gene, a child is born with GA2. Most forms of GA2 are due to a deficiency of two enzymes: electron transfer flavoprotein (ETF, encoded by ETFA and ETFB) or an electron transfer flavoprotein dehydrogenase (ETFDH). Both of these enzymes play an important role in body's ability to break down fats and proteins and turn them into energy. Therefore in its most severe forms, early detection is very crucial and most infants who show severe symptoms shortly after birth do not survive. While most cases present themselves at birth or at an early age, it is also quite possible to start showing symptoms well into adulthood, see [6]. In addition, while some patients show the symptoms suddenly and severely, some patients present with symptoms that worsen slowly over the years.

One goal of this work is to perform a meta analysis of the clinical data available regarding the late-onset patients with GA2 to determine common elements among the late-diagnosed, late-onset patients by introducing a new graphical approach. We define late-onset patients as patients who are diagnosed beyond the infant period, which we assume to be the first 28 days. We also include the patients who

2010 *Mathematics Subject Classification.* 92C50, 92D10, 65S05.
Key words and phrases. glutaric acidemia type 2, late diagnosis, muscle weakness, vomiting, hypoglycemia, data analysis, support vector machines.
Submitted Oct. 9, 2020.

are diagnosed via newborn screening who did not start showing symptoms until years later with the purpose of analyzing the onset of their symptoms. We define a late diagnosis as a diagnosis received 7 or more years after the onset of symptoms. We use support vector machines, a mathematical pattern classification tool used frequently in data mining, to obtain patterns in data. MATLAB plots are utilized to incorporate more information into a single plot to observe patterns in data.

The paper is organized as follows: In section 2, we describe the methods used to obtain and analyze the data. In section 3, results of the analysis is given. In 4 the discussion of the results are given. Finally, in section 5 conclusions of our findings are presented.

2. Methods

Subjects

Recently, all the findings related to 350 late-onset GA2 patients studied in the literature between 1979 and 2014 were compiled in [5]. A detailed table with crucial information about patients and references was also presented. We observe that diagnosis may take up to 30 years for some late-onset patients. Here we would like to examine the possible reasons behind the gap between the age at onset and age at diagnosis, and find patterns that can be detected earlier to shorten the gap. Therefore we can only study the patients for which we have both age at onset and age at diagnosis data available. Out of 350 patients studied, we have such data available for only 103 patients. Therefore our studies will be done based on these 103 patients.

Scatter Plots

Clinical data related to these 103 patients is turned into a numerical data set and presented in terms of color-coded scatter plots produced by using MATLAB. Plots are done as age at onset versus age at diagnosis. A third variable is plotted in a color-coded fashion. $y = x$ line is drawn in the middle to help observe the patients diagnosed at the time of the onset of their symptoms, or diagnosed via newborn screening before they presented with any symptoms. It also makes is easier to observe the time difference between the two events. Further away a marker is upward from the $y = x$ line, the longer it must have taken for that particular patient to get diagnosed.

Support Vector Machines (SVMs)

SVM is used for pattern classification in data analysis; it is a binary learning machine with sophisticated properties. In the case of linearly separable data, as we have here, the support vector machine constructs a hyperplane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized, see [2]. A set of data is said to be linearly separable when data is separated into two clusters such that their respective convex hulls, smallest convex set containing the data points, are disjoint. Because data here is shown to be linearly separable via convex hulls, see [3], a linear SVM is used to find an optimal separating hyperplane to classify the data. Note that due to the dimension of our problem, the hyperplane here is a line. The notation for describing a support vector machine is commonly $\mathbf{w}$, the normal vector to the separation line satisfies the equation $\mathbf{w}^{\mathsf{T}}\mathbf{x} + b = 0$, where b is a measure of offset of the separation line from

the origin, $\mathbf{x}$ is a support vector for data set in the problem such that

$$\begin{aligned} \mathbf{w}^{\mathsf{T}}\mathbf{x}_i + b &> 0, \text{ with } y_i = 1, \\ \mathbf{w}^{\mathsf{T}}\mathbf{x}_j + b &< 0, \text{ with } y_j = -1, \end{aligned}$$

see [1]. The optimal line is found by solving the optimization problem using Lagrange multipliers subject to

$$\sum_{n=1}^{\infty} \lambda_i y_i,$$

where α_i, $i = 1 \dots n$, are Lagrange multipliers. The result is the vector $\mathbf{w}$, given by the following:

$$\mathbf{w} = \sum_{n=1}^{\infty} \lambda_i y_i \mathbf{x}_i,$$

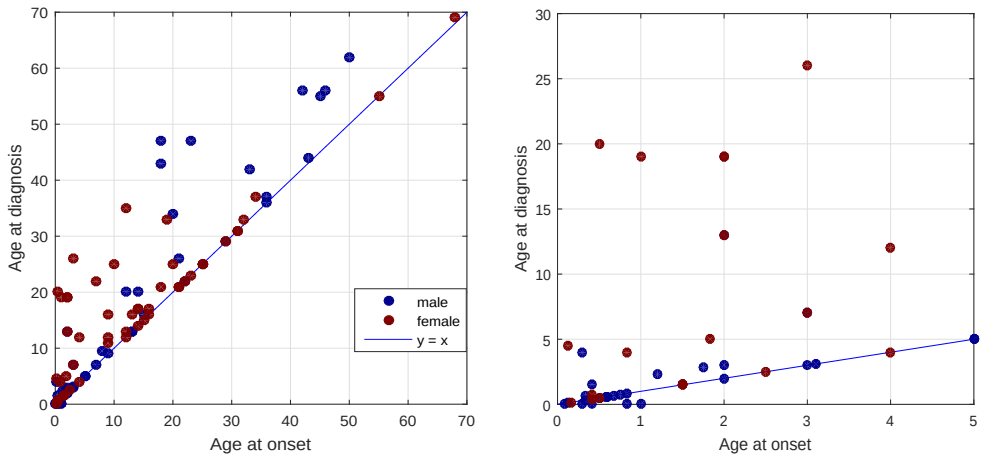
where $\{\mathbf{x}\}_{i=1}^{n}$ are the support vectors.

3. Results

The following analysis is done for 103 patients. All plots are color-coded according to the gender.

Gender Analysis
To establish how the scatter plots represent data, in figure 1(a) we first look at the gender distribution of the patients in the study according to their age when they initially developed symptoms. There are 50 males and 53 females in the study. Zooming in we see in figure 1(b) that in early childhood, boys seem to receive more timely diagnosis than girls. We also see that 5 patients were diagnosed with GA2 via newborn screening, but did not develop symptoms immediately. Since the newborn screening for GA2 is rather new, these numbers might change in the future.

Symptom Analysis
The main goal of this work is to identify patterns among patients who are diagnosed very late. We hope to gain better understanding of the disease presentation in late-onset patients and improve diagnostic process for them. For this purpose, we initially identified 8 commonly observed symptoms in patients in this study. We find that 69 patients had muscle weakness, 28 patients had hypoglycemia, 25 patients had exercise intolerance, 23 patients had vomiting, 17 patients had fatigue, 14 patients had respiratory failure, 9 patients had lipid storage myopathy, 9 patients had heart failure, and 6 patients had seizures. Muscle weakness is the most common symptom at any age of onset. Muscle weakness was also present in 23 of the 25 patients presented with exercise intolerance and in 13 of the 17 patients presented with fatigue. However, only 6 patients experienced all three symptoms altogether. In addition, 17 of the 28 patients who experienced hypoglycemia and 12 of the 23 patients who experienced vomiting also had muscle weakness. Only 18 patients had muscle weakness as their only symptom. Hence, we can say that although muscle weakness is the most common symptom in late-onset GA2 patients, it often accompanies other symptoms. It is most likely those symptoms coupled with the patient's age at onset that changes the course of diagnosis process, making it much longer than usual in some cases.



(a) (b)

FIGURE 1. Gender Data: (a) shows all patients in this study (n=103); (b) shows the patients less than 5 years of age at the age of onset.

We will focus on the three hallmark symptoms of GA2: muscle weakness, vomiting, and hypoglycemia. Muscle weakness occurs in 69 (34 m, 35 f) patients, which is about 67% percent of the patients in the study. We also observe that 39 (15 m, 24 f) patients in the study have been diagnosed 3 or more years after the onset of their symptoms and 29 (12 m, 17 f) of these patients, or 74%, experienced muscle weakness. 20 (9 m, 11 f) patients have been diagnosed 7 or more years after the onset of their symptoms and 16 (7 m, 9 f) of these patients, or 80%, experienced muscle weakness. That is, 80% of the patients diagnosed 7 or more years after the initial onset of their symptoms experienced muscle weakness.We note that there is a pretty even distribution of the number of male and female patients in each of these groups.

Figure 2 shows the patients with muscle weakness who are diagnosed 7 or more years after the onset of their symptoms. We observe that the convex hulls of data sets corresponding to female and male patients are disjoint, and hence the data is separable. By using a linear SVM, we obtain the optimal separating hyperplane, which is the green line given by $y = -x+53$. As explained in the theory of SVMs in [2], patients that lie on the same side of the green line have similar characteristics. For example, data for a female GA2 patient on this plot satisfies the following:

$$\text{age at diagnosis} > 53 - \text{age at onset}.$$

In other words, if a female patients has been trying to get a diagnosis for over 7 years and given her current age and her age at the time of symptom onset satisfies this inequality, then it is worth looking into possibility of GA2. Likewise, data for a male GA2 patient on this plot satisfies the following:

$$\text{age at diagnosis} < 53 - \text{age at onset},$$

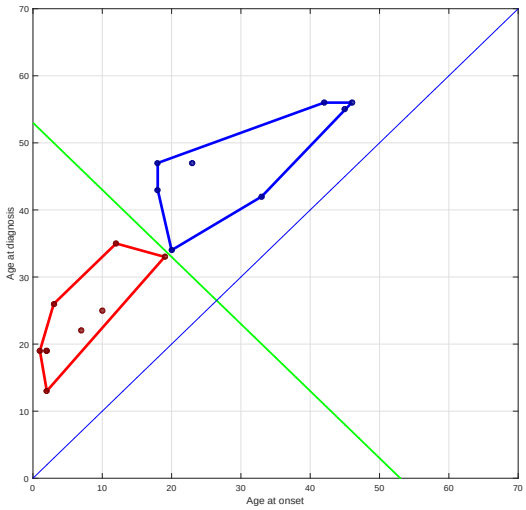


FIGURE 2. Muscle Weakness: Patients having muscle weakness who are diagnosed 7 or more years after their onset (n=16).

and this criteria can be used to determine whether a male patients should be evaluated for a possible GA2.

Next, we focus on patients who experienced vomiting as a symptom. 23 patients in the study experienced vomiting either as a sole symptom or along with other symptoms. We notice in figure 3(a) that patients having onset of their symptoms after their thirties do not complain about vomiting. In addition, we see significantly more females with this particular symptom. In figure 3(b), we observe that 11 patients experienced muscle weakness in addition to vomiting, and 7 of these patients were diagnosed 7 or more years after the onset of their symptoms. That is, about 64% of patients experiencing vomiting and muscle weakness are diagnosed 7 or more years after their onset.

Another common symptom among the late-onset GA2 patients in our study is hypoglycemia. 28 patients in the study experienced hypoglycemia as a symptom. Looking at figure 3(c), we see that not only it occurs in younger patients more commonly, but also patients are diagnosed much quicker. Similar to vomiting, it is also observed more commonly in patients younger than 30. As seen in figure 3(d), 17 of these patients also experienced muscle weakness. In contrast to other symptoms, only 2 patients has been diagnosed over 7 years after the onset of their symptoms, one being postmortem.

4. Discussion

Data related to gender, age at onset, age at diagnosis, and some of the most common symptoms for 103 patients is analyzed by converting the clinical patient data to numerical data, and by using SVMs and MATLAB scatter plots.

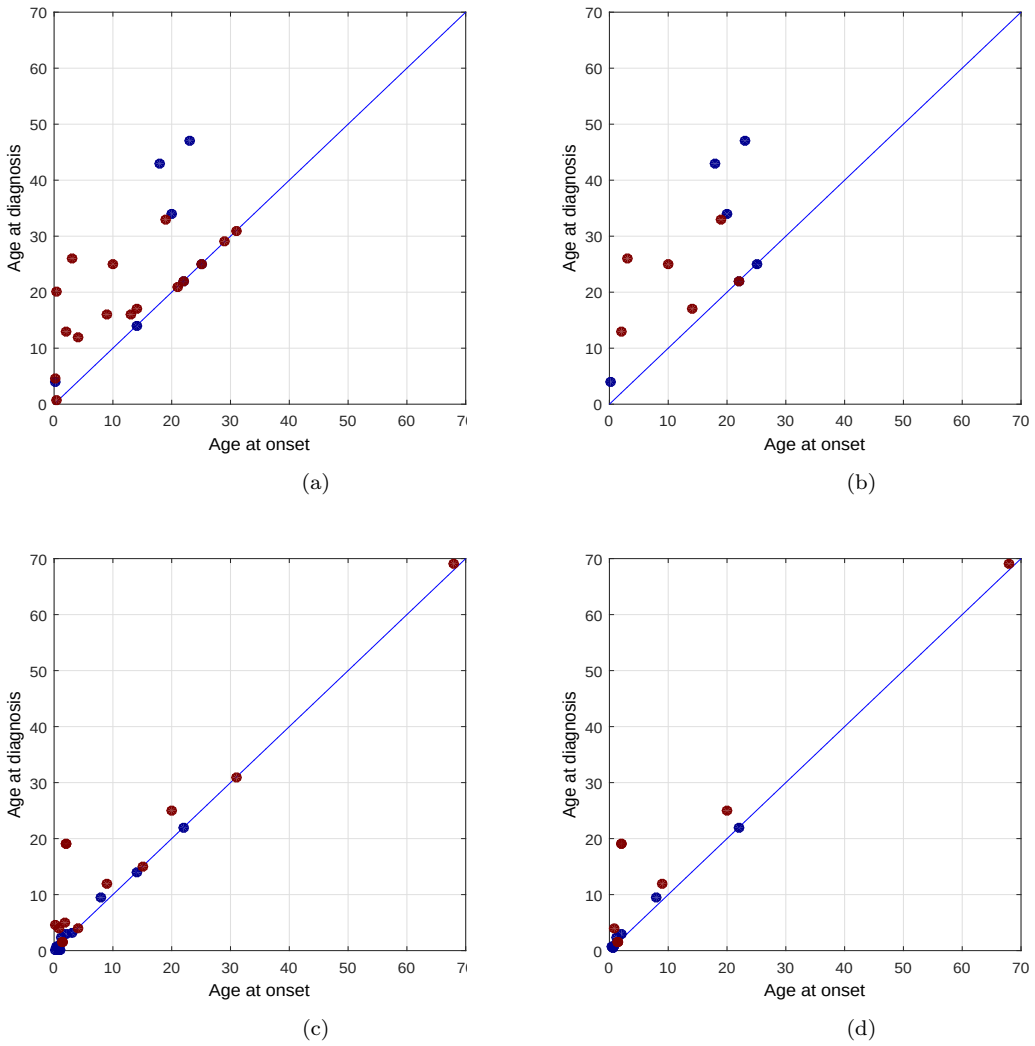


FIGURE 3. Vomiting and Hypoglycemia: (a) shows patients having vomiting as a symptom; (b) shows patients having both vomiting and muscle weakness as a symptom; (c) shows patients having hypoglycemia as a symptom; (d) shows patients having both hypoglycemia and muscle weakness as a symptom.

We observe that patients under age 5 seem to receive diagnosis quicker than older patients. Traditionally, metabolic conditions have been thought of as pediatric diagnoses. The idea of even considering a teenager or an adult to have an inborn error of metabolism is just often not in the mind of the doctor. This is the biggest factor for why the diagnosis is often not made early. In addition, if a woman that has had a "successful" pregnancy, then it is not believed that she has a defective metabolism. As a result, even though a woman presents with GA2 related symptoms, this belief

Symptom	Common explanation
muscle weakness	fatigue
vomiting	gastrointestinal conditions
hypoglycemia	insulin related disorders

TABLE 1. The typical misdiagnoses for the common symptoms of MADD

keeps doctor from even ordering right types of tests to look for a metabolic root of the problem in the first place. When a child presents with hallmark symptoms of GA2, like muscle weakness, hypoglycemia or vomiting, it is taken more seriously and more rigourous blood work including a metabolic workup is done to find the problem. On the other hand, when an older person complains of such symptoms, it is often explained via more trivial, common conditions appropriate to the patient's age and activities at the time. As a result not only those patients suffer longer, their conditions get much worse by the time they receive a diagnosis. This clearly shows the need to raise awareness about metabolic disorders.

More detailed analysis shows that muscle weakness was the most common symptom occurring in 67% of the patients in the study. Most important finding is that 80% of the patients with muscle weakness are diagnosed late, that is they are diagnosed more than 7 years after the onset of their symptoms. Moreover, all the females in this group had their onset before their twenties and all the males had their onset after their late teen years. Although this might be due to their biological differences, it could very well be due to gender bias in reaction to health concerns. Generally, men tend to wait longer before they seek a doctor's advice for a problem they are experiencing. By using SVMs, we determined a criteria based on gender and age at onset to be used in evaluation of patients with long term, over 7 years, muscle weakness. This will be a simple and helpful tool to use in evaluation of patients by physicians, possibly leading to an earlier diagnosis resulting in less damage to patients muscles and increased quality of life.

The analysis also shows that 64% of the patients with vomiting and muscle weakness are diagnosed late. In addition, we observe that vomiting and hypoglycemia, two of the signature symptoms of GA2, occurs in patients having their onset before their thirties. We do not see older patients having these symptoms. While hypoglycemia seems to occur in both females and males at the same frequency, vomiting is reported by significantly more females than males. These patterns can also be used as part of the differential diagnostic process.

In table 1, we present the typical explanations for the commons symptoms of GA2. Instead, we suggest that the following tests be considered if an ongoing muscle problem is present in an older child or adult: H & E Histochemical staining for Light Microscopy; Electron Microscopy; and Enzyme studies for Mitochondrial Enzyme Analysis.

5. Conclusions

By utilizing colored scatter plots as data visualization tool, we can incorporate more information into a single plot of data. This enables us to observe the impact of different parameters to each other, making is easier to see the patterns sometimes. Therefore we suggest the use of such plots when the data size is rather manageable,

especially if the goal is to establish patters within the available data. In addition, using SVMs we can classify the data into groups, which help determine patterns and aid in diagnostic process.

The findings in this paper can be utilized for patients presenting especially with muscle problems. The most important finding of this paper is the optimal hyperplane for patients diagnosed 7 or more years after the onset of muscle weakness. Given the inequalities satisfied for female and male patients and their age at onset, we can determine whether a patient should be evaluated for GA2. Therefore careful consideration should be given to GA2 possibility depending on the patients age, gender, and the length of the course of patients medical problems. We expect that the awareness in the medical community treating adult patients along with the patterns and patient demographics described here should improve and shorten the diagnostic process for adult patients.

References

[1] J. Dean, Big Data, Data Mining, and Machine Learning, Wileyn, 2014.
[2] S. Haykin, Neural Networks and Learning Machines, Pearson, 2008.
[3] I.T. Jolliffe, Principal Component Analysis, Springer, 2002.
[4] S. I. Goodman and F. E. Frerman, Glutaric acidaemia type II (multiple Acyl-Coa dehydrogenation deficiency), J. Inher. Metab. Dis., 7, Suppl 1, 33-37, 1984.
[5] S. Grunert, Clinical and genetical heterogeneity of late-ondet multiple acyl-coenzyme A dehydrogenase deficiency, Orp. J. Rare Dis., 9, 117-125, 2014.
[6] S. Koppel and J. Gottschalk and G. F. Hoffmann and H. R. Waterham and H. Blobel and S. Kolker, Late-onset multiple acyl-CoA dehydrogenase deficiency: a frequently missed diagnosis, Neurology, 67, 8, 1519, 2006.

Beyza C. Aslan
Department of Mathematics and Statistics, University of North Florida, Jacksonville, FL, USA
E-mail address: beyza.aslan@unf.edu