Brief Report

Cytokine Hub Classification of PASC, ME-CFS and other PASC-like Conditions

Bruce K. Patterson[1], Jose Guevara-Coto[2], Edgar B. Francisco[1], Ram Yogendra[3], Purvi Parikh[4], Rodrigo A Mora-Rodríguez[5], Javier Mora[5], Christopher Beaty[1], Gary Kaplan[6], Amiram Katz[7], Joseph A. Bellanti[8]

[1]IncellDx Inc, San Carlos, CA
[2]Department of Computer Science and Informatics (ECCI), Universidad de Costa Rica, San Jose, Costa Rica
[3]Lawrence General Hospital, Lawrence, MA
[4]NYU Langone Health, New York, NY
[5]Lab of Tumor Chemosensitivity, CIET / CICICA, Faculty of Microbiology, Universidad de Costa Rica.
[6]Clinical Associate Professor (Adjunct), Department of Community and Family Medicine, Georgetown University School of Medicine, Washington, DC
[7]Neurology Specialist Affiliated with Norwalk Hospital. Orange, CT
[8]Departments of Pediatrics and Microbiology-Immunology, and the International Center for Interdisciplinary Studies of Immunology, Georgetown University Medical Center, Washington DC.

Corresponding author:
Bruce K. Patterson MD
1541 Industrial Road
San Carlos, CA 94070
Tel: +1.650.777.7630
Fax: +1.650.587.1528
Email: brucep@incelldx.com

Key words:

COVID-19, PASC, long COVID, long hauler, PTLD, ME-CSF

Abbreviations:
PASC, post-acute sequelae of COVID-19; myalgic encephalomyelitis-chronic fatigue syndrome (ME-CFS), post-treatment Lyme disease (PTLD), 42 post-vaccine individuals with PASC-like symptoms (POVIP), classification and regression Trees (CART), IL-interleukin; RANTES-regulation on activation, healthy control T-expressed and secreted; CCR-chemokine receptor; IFN-interferon, TNF-tumor necrosis factor; MIP-macrophage inflammatory protein; GM-CSF-granulocyte-macrophage colony-stimulating factor; VEGF-vascular endothelial growth factor

ABSTRACT

Background: Post-acute sequelae of COVID-19 (PASC) is a growing healthcare and economic concern affecting as many as 10%-30% of those infected with COVID-19. Though the symptoms have been well-documented, they significantly overlap with other common chronic inflammatory conditions which could confound treatment and therapeutic trials.

Methods: A total of 236 patients including 64 with post-acute sequelae of COVID-19 (PASC), 50 with myalgic encephalomyelitis-chronic fatigue syndrome (ME-CFS), 29 with post-treatment Lyme disease (PTLD), and 42 post-vaccine individuals with PASC-like symptoms (POVIP) were enrolled in the study. We performed a 14-plex cytokine/chemokine panel previously described to generate raw data that was normalized and run in a decision tree model using a Classification and Regression Tree (CART) algorithm. The algorithm was used to classify these conditions in distinct groups despite their similar symptoms.

Results: PASC, ME-CSF, POVIP, and Acute COVID-19 disease categories were able to be classified by our cytokine hub based CART algorithm with an average F1 score of 0.61 and high specificity (94%).

Conclusions: Proper classification of these inflammatory conditions with very similar symptoms is critical for proper diagnosis and treatment.

INTRODUCTION

A number of authors have identified overlapping symptom presentations of long Covid with similarities to ME/CSF[1,2], PTLD and other postinfectious chronic medical disorders[3] There are, however, clear etiological differences and pathphysiological differences in these conditions that necessitate precision medicine tailored therapies.

A recent report suggested the use of cytokine hubs to more precisely categorize autoimmune diseases with the stated goal of using the information as therapeutic targets as the expansion of immune based therapy grows[4]. The heterogeneity of immune-mediated inflammatory diseases (IMIDS) described in the prior publication also applies to post-infectious immune-mediated and inflammatory conditions currently in the discussion of post-infectious sequelae of COVID-19 (PASC) and the focus of this report. Unlike the previous publication, PASC-like conditions share many of the same symptoms making diagnosis and, ultimately, treatment more difficult[5-8]. Here, we present a machine learning approach to classifying these symptom related conditions.

METHODS

Patients

After written informed consent was obtained, the immunological lab results were used for the current analysis. 236 patients including 64 with post-acute sequelae of COVID-19 (PASC), 50 with myalgic encephalomyelitis-chronic fatigue syndrome (ME-CFS), 29 with post-treatment Lyme disease (PTLD), and 42 post-vaccine individuals with PASC-like symptoms (POVIP) defined as COVID-negative (nucleoprotein Ab negative, T-cell immunity negative) individuals with PASC-like symptoms 3 months after the last vaccination shot.

Multiplex Cytokine Quantification

Fresh plasma was used quantification of the following analytes: TNF-α, IL-4, IL-13,IL-2, GM-CSF, sCD40L, CCL5 (RANTES), CCL3 (MIP-1α),IL-6, IL-10, IFN-γ, VEGF, IL-8, and CCL4 (MIP-1β) as previously decribed[6].

Data Acquisition and Processing

A dataset comprising six categories [Mild-Moderate acute COVID, Severe acute COVID, PASC, POVIP, ME-CSF and PTLD individuals] was derived from flow cytometry analysis. The data consisted of a total of 16 columns represented by an anonymized sample identifier, and the last column was the class or disease state assigned to the individual. The raw dataset was imported into python 3.9 using the *pandas* library[9]. The identifier column was converted to an index to avoid using in model

construction. The class label was removed from this dataset and assigned into a new variable (*target*).

The dataset comprising the cytokine profiles was normalized using L2 normalization, also known as the Euclidian normalization. The L2 nonrealization approach calculates the distance of a given vector of values, such as the cytokine values for each data point or instance from the origin of the vector space. The implementation of the L2 normalization or Euclidaian norm results in a positive value. The implementation of the L2 norm can be supported by the notion our model is focused on identifying the differences between classes, thus the signal or pattern that characterizes each class and not the effect of the magnitudes on each class[10]. The two datasets (cytokines profiles and targets/labels) were used to train a machine learning classifier using decision trees to classify the individuals according to their disease state.

Multiclass classification of disease states using a decision tree model

The processed feature set and targets (labels) were used to train a decision tree model using the Classification and Regression Trees (CART) algorithm in scikit-learn[11]. The model was fine-tuned using GridSearchCV[12] from scikit-learn. To fine-tune the model, we selected the following hyperparameters: maximum number of features, the minimum number of samples to split a node, the minimum number of samples per node, the maximum depth of the tree. The hyperparameter corresponding to the weight of each class was set to balanced to counter for class imbalances in the dataset. Additionally,

the impurity criterion was defined as the Gini impurity value. The cross validation was set to 10-fold cross validation with 3 repeats.

The resulting decision tree was used to further calculate a confusion matrix in order to visualize the model's predictive power as well as a "leave one out cross-validation" (LOOCV). The confusion matrix was used to generate a classification report with the relevant predictive metrics and the LOOCV was used to calculate the F1-score which uses the harmonic means of precision and recall; combining them into a performance metric ranging from 0 (poor performance) to 1 (perfect performance). The resulting tree was then plotted to visualize the separation of the classes based on the different cytokines, allowing a potential identification of which cytokines and values define each class.

RESULTS and DISCUSSION

The symptoms of post-acute sequelae of COVID-19 (PASC) have been widely reported[5,6] and significantly overlap with myalgic encephalomyelitis-chronic fatigue syndrome (ME-CFS)[7] and with post-treatment Lyme disease (PTLD)[8]. In addition, we included patients, of unknown prevalence, who are post-vaccine individuals with PASC-like symptoms (POVIP) in the present study (Figure 1). These patients experience PASC-like symptoms 3 months minimum post-vaccination in the absence of COVID 19 infection.

By using a decision tree classifier, Classification and Regression Trees (CART), we developed an algorithm to propose a simple but powerful predictive model with high

interpretability, a characteristic of great importance when attempting to understand differences between disease states. Our model had an average weighted F1 score of 0.61, which was variable due to the stochastic nature of both the model and the dataset. As shown in Table 1 and in the confusion matrix (Supplementary Figure 1), the model was robust in its ability to identify four of the six classes of disease states (e.g. Mild to Moderate COVID-19, Severe COVID-19, PASC and ME-CSF). However, some misclassification was demonstrated in the remaining two classes (POVIP and Lyme) that was most likely due to overlapping cytokine hubs. The confusion matrix, however, could indicate that the immune contribution to these two states were similar. Moreover, the addition of clinical data as well as other immunological parameters could potentially separate these two conditions and further increase the model's predictive power.

The highest performance metrics after fine-tuning was provided by the CART decision tree (Fig. 2) when data were plotted using a scikit-learns internal tree plotting function and python's matplotlib (Supplementary Fig. 2). As shown in the plotted tree (Fig 2, Supplementary Fig. 2), we demonstrate that the CART algorithm was capable of constructing splitting and terminal hubs with low Gini impurity values and high F1 scores for those classes shown previously in the confusion matrix (Supplementary Fig. 1). In the POVID and PTLD classes, splitting hubs with higher Gini impurity values were observed. The identification of these hubs supports the possibility that the immune profiles of both Post-Vax and Lyme individuals have similar cytokine patterns.

The cross-sectional nature of data included in this report provided a unique opportunity to explore the effectiveness of measuring cytokine hub signatures beyond populations originally described in the Schett et al. Because our study also included participants

post-vaccination either positive or negative serology, we were able to compare the prevalence of persistent physical symptoms according to these two variables. We were also able to classify patients with symptoms related to COVID-19 infection alone (pure PASC).

CONCLUSIONS

We agree with the conclusions of Schett et al.[1] that targeting of individual cytokines underlying the immunopathogenesis of these conditions may provide a powerful new tool in the treatment of these immunologically-mediated disorders using precision medicine. Further study is necessary to elucidate how pathogen or antigen persistence may contribute to the classification presented here[13,14].

REFERENCES

1. Wong TL, Weitzer DJ. Long COVID and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)-A Systemic Review and Comparison of Clinical Presentation and Symptomatology. Medicina (Kaunas). 2021 Apr 26;57(5):418. doi: 10.3390/medicina57050418. PMID: 33925784; PMCID: PMC8145228.
2. Komaroff AL, Bateman L. Will COVID-19 Lead to Myalgic Encephalomyelitis/Chronic Fatigue Syndrome?. *Front Med (Lausanne)*. 2021;7:606824. Published 2021 Jan 18. doi:10.3389/fmed.2020.606824
3. Auwaerter PG. The Race to Understand Post-COVID-19 Conditions. *Ann Intern Med*. 2021;174(10):1458-1459. doi:10.7326/M21-3072
4. Schett G, McInnes IB, Neurath MF. Reframing Immune-Mediated Inflammatory Diseases through Signature Cytokine Hubs. N Engl J Med. 2021;385(7):628-639.
5. Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. EClinicalMedicine. 2021;38:101019.
6. Patterson BK, Guevara-Coto J, Yogendra R, et. al. Immune-based prediction of COVID-19 severity and chronicity decoded using machine learning. Front Immunol. 2021 Jun 28;12:700782. doi: 10.3389/fimmu.2021.700782.
7. Wong KH, Shapiro ED, Soffer GK. A Review of Post-treatment Lyme Disease Syndrome and Chronic Lyme Disease for the Practicing Immunologist. Clin Rev

Allergy Immunol. 2022 Feb;62(1):264-271. doi: 10.1007/s12016-021-08906-w. Epub 2021 Oct 23. PMID: 34687445.
8. Bateman L, Bested AC, Bonilla HF, Chheda BV, Chu L, Curtin JM, Dempsey TT, Dimmock ME, Dowell TG, Felsenstein D, Kaufman DL, Klimas NG, Komaroff AL, Lapp CW, Levine SM, Montoya JG, Natelson BH, Peterson DL, Podell RN, Rey IR, Ruhoy IS, Vera-Nunez MA, Yellman BP. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Essentials of Diagnosis and Management. Mayo Clin Proc. 2021 Nov;96(11):2861-2878. doi: 10.1016/j.mayocp.2021.07.004. Epub 2021 Aug 25. PMID: 34454716.
9. Mckinney W. *Pandas: A Foundational Python Library for Data Analysis and Statistics*. http://pandas.sf.net
10. Brownlee J. *Basics of Linear Algebra for Machine Learning: Discover the Mathematical Language of Data in Python*.; 2018.
11. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Classif Regres Trees. Published online January 1, 2017:1-358. doi:10.1201/9781315139470
12. Pedregosa F, Michel V, Grisel O, et al. Scikit-Learn: Machine Learning in Python Vol 12.; 2011. Accessed April 17, 2021. http://scikit-learn.sourceforge.net.
13. Patterson BK, Francisco EB, Yogendra R, Long E, Pise A, Rodrigues H, Hall E, Herrera M, Parikh P, Guevara-Coto J, Triche TJ, Scott P, Hekmati S, Maglinte D, Chang X, Mora-Rodríguez RA, Mora J. Persistence of SARS CoV-2 S1 Protein in CD16+ Monocytes in Post-Acute Sequelae of COVID-19 (PASC) up to 15 Months Post-Infection. Front Immunol. 2022 doi: 10.3389/fimmu.2021.746021.
14. Buonsenso D, Piazza M, Boner AL, Bellanti JA. Long COVID: A Proposed Hypothesis-driven Model of Viral Persistence for the Pathophysiology of the Syndrome. Allergy Asthma Proc. 2022 May (in press).

Ethics:
All the patients/participants provided their written informed consent to participate in this study.

Data and materials availability:
All requests for materials and raw data should be addressed to the corresponding author

Competing interests:
B.K.P., E.B.F. are employees of IncellDx, Inc. C.B. is a consultant of IncellDx, Inc

Author contributions:
R.Y., P.P. organized the clinical study and actively recruited patients.
B.K.P, E.B.F, J.G-C. performed experiments and analyzed the data.
J.G-C., R.A.M., J.M., C.B. performed the statistics and bioinformatics
B.K.P., J.A.B., G.K., A.K., J.M., E.B.F, J.G-C., R.A.M. wrote and edited the draft of the manuscript and all authors contributed to revising the manuscript prior to submission.

Funding: None

FIGURE LEGENDS

Table 1: Performance metrics for the classification of the six disease states in this report.

Figure 1. Symptoms in POVIP separated by vaccine type.

Figure 2. Schematic of the fine-tuned decision tree model implemented with CART (Supplementary Figure 2). The plotted decision tree with the highest performance for class separation following fine-tuning with grid search and cross validation. Cytokine origination hubs are round, defining cytokine hubs are in rectangles and disease states are in hexagons. The resulting tree had a maximum depth of 6 splitting hubs, and used the Gini impurity criterion, which measures the probability value of misclassifying a randomly selected event or individual from the dataset if such an element were randomly labeled based on its class distribution in the dataset.

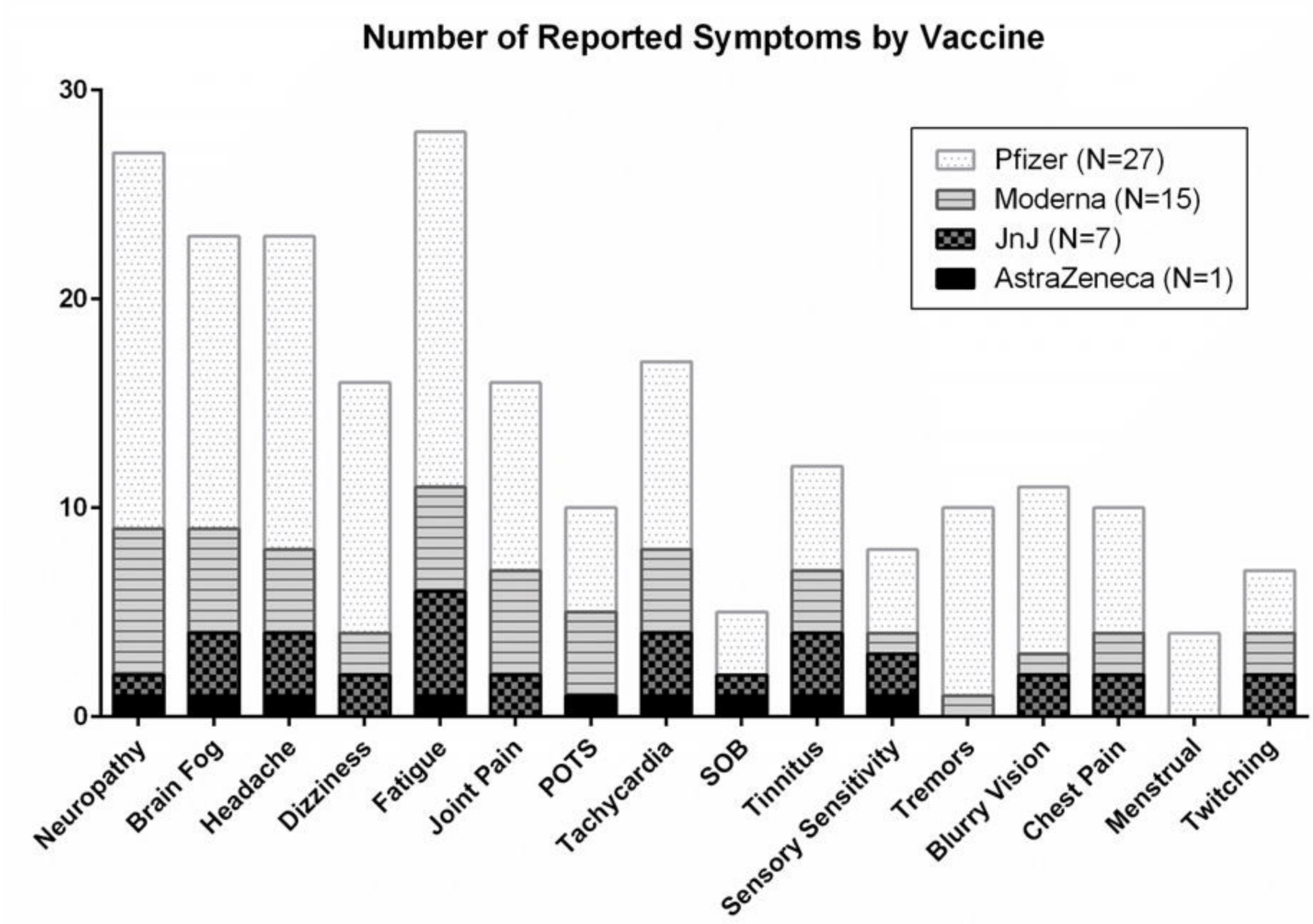
Supplementary Figure 1. Confusion matrix generated by the analysis of the dataset using the fine-tuned CART model. The matrix represents the analysis of the dataset using the highest performing model after tuning and internal cross-validation. The diagonal is the classification of each class as its own (true positives). Left and right of the diagonal for each class are the misclassifications of individuals from that class as belonging to other disease states (false negatives). The colored bar on the right side provides a reference for classification and misclassification, with darker colors (lower values) preferred for the latter.

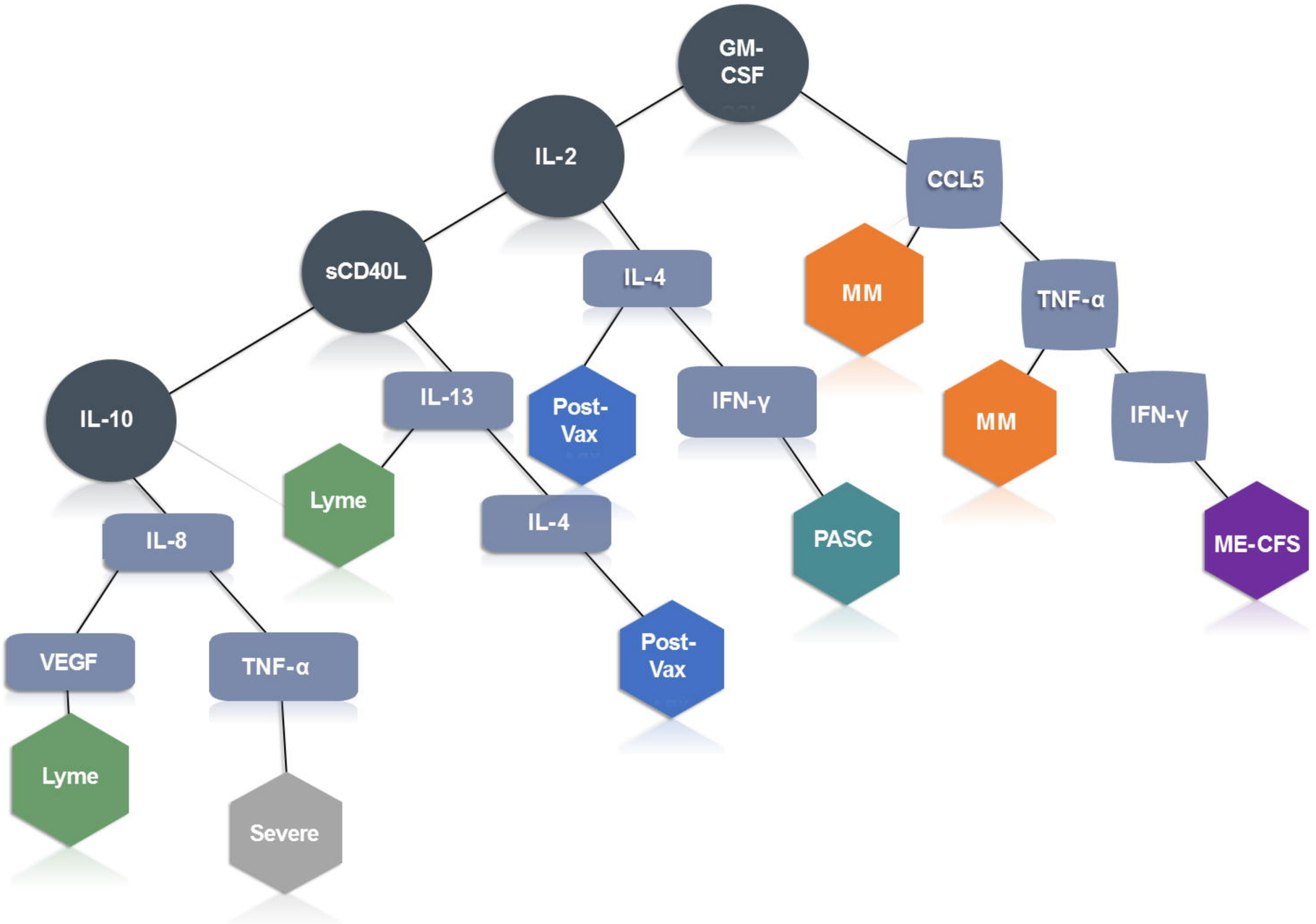
Supplementary Figure 2. Visualization of the fine-tuned decision tree model implemented with CART. The plotted decision tree with the highest performance for class separation following fine-tuning with grid search and cross validation. Included are node impurity scores according to the Gini impurity criterion. Additionally, the tree provides the node splitting threshold for the cytokines (features) used in model construction, and the number of individuals per class for each splitting or terminal node (in square brackets). Each of the six

classes (MM, Severe, PASC, Post-Vax, ME-CSF and Lyme) was assigned a different color upon separation and potential classification.

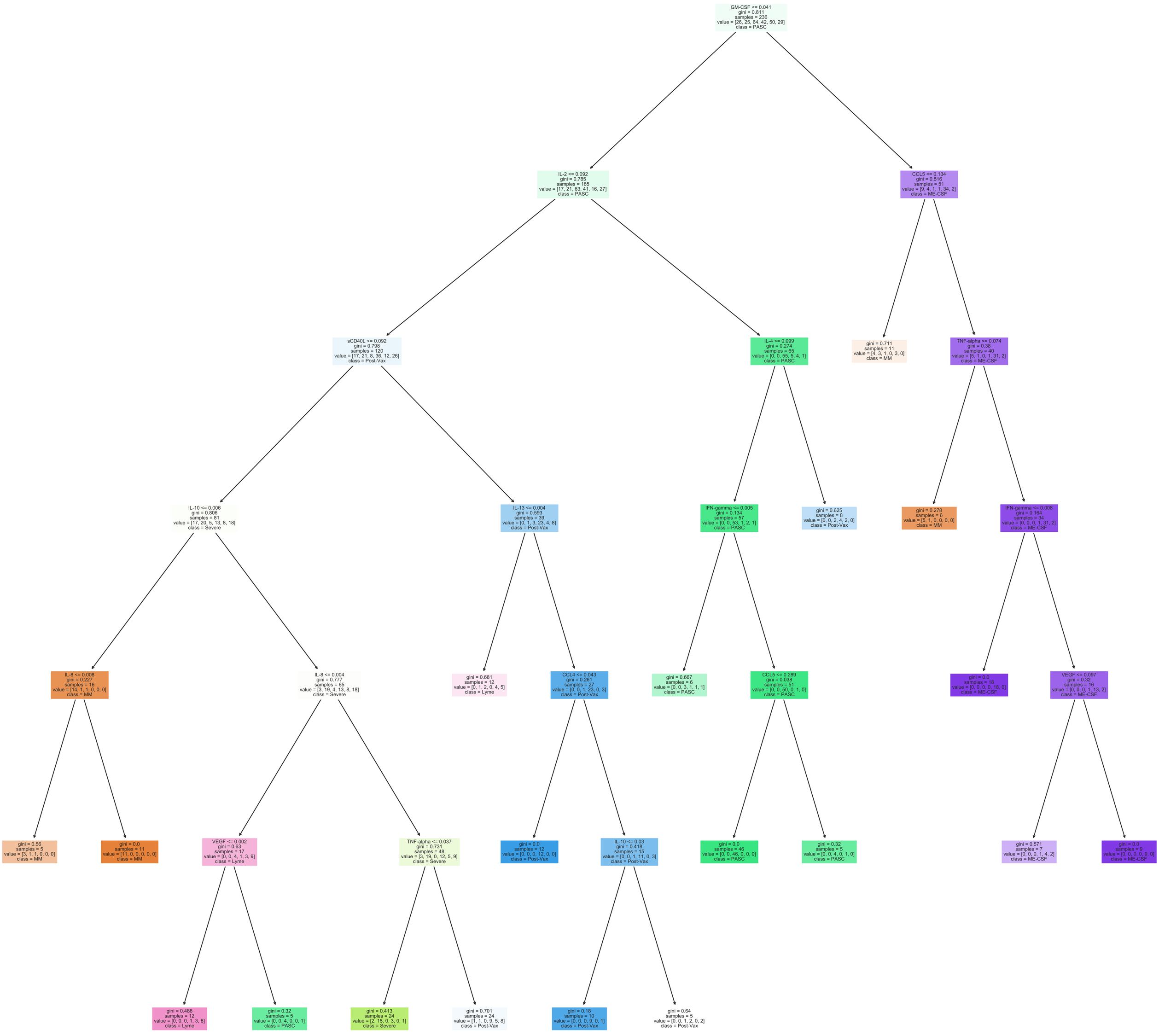
Table 1. Performance metrics for the multi-class decision tree predictor constructed using the CART algorithm. For each class in the dataset the true and false positives and negatives were identified from the confusion matrix. The recall, precision, specificity and the F1 score were calculated for each class and a macro or average value for each metric was also obtained (bottom row).

Class	**Recall**	**Precision**	**Specificity**	**F1 Score**
MM	0.8846	0.7419	0.9619	0.8070
Severe	0.6000	0.7143	0.9716	0.6522
PASC	0.8438	0.8308	0.9360	0.8372
Post-Vax	0.8333	0.6140	0.8866	0.7071
ME-CSF	0.7000	0.6604	0.9032	0.6796
Lyme	0.1724	0.5556	0.9807	0.2632
Average	**0.6724**	**0.6862**	**0.9400**	**0.6577**


Number of Reported Symptoms by Vaccine
30
20
10
0
Pfizer (N=27)
Moderna (N=15)
JnJ (N=7)
AstraZeneca (N=1)
Neuropathy
Brain Fog
Headache
Dizziness
Fatigue
Joint Pain
POTS
Tachycardia
SOB
Tinnitus
Sensory Sensitivity
Tremors
Blurry Vision
Chest Pain
Menstrual
Twitching


GM-CSF
IL-2
CCL5
sCD40L
IL-4
MM
TNF-α
IL-10
IL-13
Post-Vax
IFN-γ
MM
IFN-γ
Lyme
IL-4
PASC
ME-CFS
IL-8
Post-Vax
VEGF
TNF-α
Lyme
Severe


MM
Severe
PASC
Post-Vax
ME-CSF
Lyme
23 2 0 1 0 0
5 18 0 1 0 1
2 0 57 3 0 2
0 3 1 36 1 1
3 0 2 7 31 7
0 1 2 11 2 13
MM
Severe
PASC
Post-Vax
ME-CSF
Lyme
50
40
30
20
10
0


GM-CSF <= 0.041
gini = 0.811
samples = 236
value = [26, 25, 64, 42, 50, 29]
class = PASC
IL-2 <= 0.092
gini = 0.785
samples = 185
value = [17, 21, 63, 41, 16, 27]
class = PASC
CCL5 <= 0.134
gini = 0.516
samples = 51
value = [9, 4, 1, 1, 34, 2]
class = ME-CSF
sCD40L <= 0.092
gini = 0.798
samples = 120
value = [17, 21, 8, 36, 12, 26]
class = Post-Vax
IL-4 <= 0.099
gini = 0.274
samples = 65
value = [0, 0, 55, 5, 4, 1]
class = PASC
gini = 0.711
samples = 11
value = [4, 3, 1, 0, 3, 0]
class = MM
TNF-alpha <= 0.074
gini = 0.38
samples = 40
value = [5, 1, 0, 1, 31, 2]
class = ME-CSF
IL-10 <= 0.006
gini = 0.806
samples = 81
value = [17, 20, 5, 13, 8, 18]
class = Severe
IL-13 <= 0.004
gini = 0.593
samples = 39
value = [0, 1, 3, 23, 4, 8]
class = Post-Vax
IFN-gamma <= 0.005
gini = 0.134
samples = 57
value = [0, 0, 53, 1, 2, 1]
class = PASC
gini = 0.625
samples = 8
value = [0, 0, 2, 4, 2, 0]
class = Post-Vax
gini = 0.278
samples = 6
value = [5, 1, 0, 0, 0, 0]
class = MM
IFN-gamma <= 0.008
gini = 0.164
samples = 34
value = [0, 0, 0, 1, 31, 2]
class = ME-CSF
IL-8 <= 0.008
gini = 0.227
samples = 16
value = [14, 1, 1, 0, 0, 0]
class = MM
IL-8 <= 0.004
gini = 0.777
samples = 65
value = [3, 19, 4, 13, 8, 18]
class = Severe
gini = 0.681
samples = 12
value = [0, 1, 2, 0, 4, 5]
class = Lyme
CCL4 <= 0.043
gini = 0.261
samples = 27
value = [0, 0, 1, 23, 0, 3]
class = Post-Vax
gini = 0.667
samples = 6
value = [0, 0, 3, 1, 1, 1]
class = PASC
CCL5 <= 0.289
gini = 0.038
samples = 51
value = [0, 0, 50, 0, 1, 0]
class = PASC
gini = 0.0
samples = 11
value = [11, 0, 0, 0, 0, 0]
class = MM
VEGF <= 0.002
gini = 0.63
samples = 17
value = [0, 0, 4, 1, 3, 9]
class = Lyme
TNF-alpha <= 0.037
gini = 0.731
samples = 48
value = [3, 19, 0, 12, 5, 9]
class = Severe
gini = 0.0
samples = 12
value = [0, 0, 0, 12, 0, 0]
class = Post-Vax
IL-10 <= 0.03
gini = 0.418
samples = 15
value = [0, 0, 1, 11, 0, 3]
class = Post-Vax
gini = 0.0
samples = 46
value = [0, 0, 46, 0, 0, 0]
class = PASC
gini = 0.32
samples = 5
value = [0, 0, 4, 0, 1, 0]
class = PASC
gini = 0.0
samples = 18
value = [0, 0, 0, 0, 18, 0]
class = ME-CSF
VEGF <= 0.097
gini = 0.32
samples = 16
value = [0, 0, 0, 1, 13, 2]
class = ME-CSF
gini = 0.56
samples = 5
value = [3, 1, 1, 0, 0, 0]
class = MM
gini = 0.0
samples = 11
value = [11, 0, 0, 0, 0, 0]
class = MM
gini = 0.486
samples = 12
value = [0, 0, 0, 1, 3, 8]
class = Lyme
gini = 0.32
samples = 5
value = [0, 0, 4, 0, 0, 1]
class = PASC
gini = 0.413
samples = 24
value = [2, 18, 0, 3, 0, 1]
class = Severe
gini = 0.701
samples = 24
value = [1, 1, 0, 9, 5, 8]
class = Post-Vax
gini = 0.18
samples = 10
value = [0, 0, 0, 9, 0, 1]
class = Post-Vax
gini = 0.64
samples = 5
value = [0, 0, 1, 2, 0, 2]
class = Post-Vax
gini = 0.571
samples = 7
value = [0, 0, 0, 1, 4, 2]
class = ME-CSF
gini = 0.0
samples = 9
value = [0, 0, 0, 0, 9, 0]
class = ME-CSF