

# Who's your data? Primary immune deficiency differential diagnosis prediction via machine learning and data mining of the USIDNET registry

**Jose Alfredo Mendez Barrera**

Autonomous Technological Institute of Mexico

**Samuel Rocha Guzman**

Autonomous Technological Institute of Mexico

**Elisa Hierro Cascajares**

National Institute of Pediatrics

**Elizabeth K Garabedian**

Immunology and Rheumatology at Columbia University

**Ramsay L Fuleihan**

Immunology and Rheumatology at Columbia University

**Kathleen E Sullivan**

National Institute of Pediatrics

**Saul Oswaldo Lugo Reyes** (✉ [dr.lugo.reyes@gmail.com](mailto:dr.lugo.reyes@gmail.com))

National Institute of Pediatrics <https://orcid.org/0000-0002-3730-4150>

---

## Research Article

**Keywords:** Rare diseases, inborn errors of immunity, data mining, extreme gradient boosting, machine learning, diagnosis prediction, classification, primary immune deficiencies, registry

**Posted Date:** January 27th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1211465/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**INTRODUCTION:** There are currently more than 450 primary immune deficiency (PID) diseases, and about 7,000 rare diseases that together afflict around 1 in every 17 humans. Computational aids based on data mining and machine learning might facilitate the diagnostic task by extracting rules from large datasets and making predictions when faced with new problem cases.

**OBJECTIVE:** In a proof-of-concept data mining study, we aimed to predict PID diagnoses with a supervised machine learning algorithm based on classification tree boosting.

**METHODS:** Through a data query at the USIDNET registry we obtained a database of 2,396 patients with common diagnoses of PID, including their clinical and laboratory features. We kept 12 diagnoses and 286 features that were included in the model. We used the XGBoost package with parallel tree boosting for the supervised classification model, and SHAP for variable importance interpretation, on Python v3.7. The patient database was split into training and testing subsets, and after boosting through gradient descent, the predictive model provides measures of diagnostic prediction accuracy and individual feature importance. To correct for imbalanced classification, after a baseline performance test, we used the Class Weighting Hyperparameter, or `scale_pos_weight`.

**RESULTS:** The twelve PID diagnoses were CVID (1,098 patients), DiGeorge syndrome, Chronic granulomatous disease, Congenital agammaglobulinemia, ID not otherwise classified, Specific antibody deficiency, Complement deficiency, Hyper-IgM, Leukocyte adhesion deficiency, ectodermal dysplasia with immune deficiency, Severe combined immune deficiency, and Wiskott-Aldrich syndrome. For CVID, the model found an accuracy on the train sample of 0.80, with an area under the ROC curve (AUC) of 0.80, and a Gini coefficient of 0.60. In the test subset, accuracy was 0.76, AUC 0.75, and Gini 0.51. The positive feature value to predict CVID was highest for upper respiratory infections, asthma, autoimmunity and hypogammaglobulinemia. Features with the highest negative predictive value were high IgE, growth delay, abscess, lymphopenia, and congenital heart disease. For the rest of the diagnoses, accuracy stayed between 0.75 and 0.99, AUC 0.46-0.87, Gini 0.07-0.75, and LogLoss 0.09-8.55. See tables and figures.

**DISCUSSION:** Clinicians should remember to consider the negative predictive features together with the positives. We are calling this a proof-of-concept study to continue with our explorations. A good performance is encouraging, and the feature importance might aid feature selection for future endeavors. In the meantime, we can learn from the rules derived by the model and build a user-friendly decision tree to generate differential diagnoses.

## Introduction:

Inborn errors of immunity, also known as Primary immune deficiencies (PID), are a heterogeneous group of over 450 congenital rare diseases with increased susceptibility to infection, inflammation, autoimmunity, allergy and/or cancer (1). Their estimated global prevalence is 1 in 10,000 humans. The

United States Immunodeficiency Network (USIDNET) is a research consortium that maintains a registry of data from over 4,000 patients with PID ([usidnet.org](http://usidnet.org)).

Patients with rare diseases usually endure an odyssey of several years before reaching a correct diagnosis (2). Diagnostic errors are an important cause of deaths, complaints, complications, and waste (3). The human gene connectome at Rockefeller University has predicted a ceiling of more than 3,100 genes involved in the immune system (4). In the following decade, gene discovery by next-generation sequencing could bring the number of PID diseases to the thousands, one-upping the complexity of the diagnostic task as well.

Computers might help clinicians diagnose and reduce diagnostic errors (5), especially in the field of primary immune deficiencies and other rare diseases, which are usually congenital, with an early and syndromic presentation (6). Artificial intelligence (A.I.) solutions are already implemented in financial services, global positioning/navigation, and biological sciences, among other human endeavors.

Data mining and machine learning are two branches of A.I. that make use of existing large datasets to find patterns and extract rules with algorithms or automated methods that make sense of data to then make predictions about new data sets or problem cases.

Extreme Gradient Boosting (XGBoost) is a robust supervised machine learning algorithm (<https://xgboost.readthedocs.io/en/latest/>). Classification tree boosting consists of the generation of multiple sequential “weak” prediction models, each of which is taken to generate a stronger, more stable model by tweaking the results of the previous model. This optimization algorithm is called a Gradient descent. The average gain of each feature in all trees in which it appears is used to rank the features in importance (7). Hyperparameter optimization, or tuning, is the measure and choosing of the most valuable parameters, by giving them different weights or constraints, thus improving learning rates and minimizing function losses to optimize the model.

For the interpretation of the importance of variables, SHAP (Shapley Additive exPlanations) can be helpful (<https://github.com/slundberg/shap>). This method assigns to each feature an importance value for a unique prediction and facilitates the global and local interpretability of any machine learning model, thus making for a more user-friendly interpretation (8).

## **Objective:**

In this proof-of-concept data mining study, we aimed to predict PID diagnoses using a machine learning algorithm.

## **Methods:**

Through a data query at the USIDNET registry ([usidnet.org](http://usidnet.org)), we obtained a database of 2,396 patients with common diagnoses of primary immune deficiencies (PID), including their clinical and laboratory

features. The database was curated and re-codified to reduce dimensionality and convert continuous variants into categorical (*e.g.*, high/low/normal IgG); we kept 12 diagnoses (rows) and 286 features (columns) that were included in the model.

We used the Extreme Gradient Boost (XGBoost) package with parallel tree boosting for the supervised classification model, and SHAP (Shapley Additive exPlanations) for variable importance interpretation, on Python version 3.7 (Jupyter) and Google Colaboratory notebooks.

Each diagnosis is differentiated or predicted against all others (*i.e.*, CVID versus 11 other diagnoses). The patient database is split into training and testing subsets, and after boosting through gradient descent, the predictive model provides measures of prediction probability: accuracy, area under the ROC curve (AUC), gini index, and Logloss. To correct for imbalanced classification, after a baseline performance test, we used the Class Weighting Hyperparameter, or `scale_pos_weight` for all diagnoses except the most abundant: CVID and DiGeorge Syndrome.

This study design is exempt from Institutional Review Board approval at the National Institute of Pediatrics. We maintain and respect the confidentiality of patients and their families. All authors subscribe and uphold the Helsinki Declaration involving experimentation on human subjects.

## Results:

The twelve PID diagnoses were CVID (1,098 patients), DiGeorge syndrome (DGS, 406), Chronic granulomatous disease (154), Congenital agammaglobulinemia (135), Not otherwise classified (CORE, 132), Specific antibody deficiency (SAD, 117), Complement deficiency (12), Hyper-IgM (HIGM, 46), Leukocyte adhesion deficiency (6), Ectodermal dysplasia with immune deficiency (NEMO, 25), Severe combined immune deficiency (SCID, 202), and Wiskott-Aldrich syndrome (WAS, 63 cases). See figure 1.

For the most frequent diagnosis, Common variable immune deficiency (CVID), the model found an accuracy on the train sample of 0.80, with an area under the ROC curve (AUC) of 0.80, and a Gini coefficient of 0.60. In the test subset, accuracy was 0.76, AUC 0.75, and Gini 0.51. The positive feature value to predict CVID was highest for upper respiratory infections, asthma, autoimmunity and hypogammaglobulinemia. Features with the highest negative predictive value were high IgE, growth delay, abscess, lymphopenia, and congenital heart disease. See figure 2.

For DiGeorge Syndrome (DGS, n=406), the accuracy on the test sample was 0.90, with an AUC of 0.87, and gini index of 0.75. Feature importance was highest for congenital heart disease, facial dysmorphism, palate malformation, and hydroelectrolytic disorders. Top negative features include Low IgA, upper respiratory infections, abscess, diarrhea, and low IgG. Figure 3.

Specific antibody deficiency (SPAD, n=117) had an accuracy on test of 0.84, AUC 0.74, and gini of 0.49. Important positive features were Upper respiratory and ear infections, allergic rhinitis, and Streptococcus;

top negative features: Low IgG, Low IgM, cardiovascular disease, teeth, diarrhea, eczema, gastrointestinal manifestations, and High IgE.

For Agammaglobulinemia (AGAMMA, n=135), accuracy was 0.89, AUC 0.76, gini 0.519; Logloss was 3.84. Top positive features were Low IgM and IgG, Ear, Eye, and Bone infections, and arthritis; top negatives were Candida sp., cardiovascular disease, leukopenia, endocrine-metabolic disorders, and lymph nodes.

For Chronic granulomatous disease (CGD, n=154), accuracy was 0.92, AUC 0.83, gini 0.67, and Logloss 2.69. Top positive predictive features were Abscess, high IgA, Aspergillus sp, lymph nodes, and inflammatory bowel disease. Top negative predictive features: Upper respiratory infections, Low IgM, Asthma, Low IgM, and cytopenias.

After target-variable (y) scaling to correct for imbalanced classification, Complement deficiencies (COMPDEF, n=12) had a prediction accuracy of 0.92, AUC of 0.46, gini of -0.07, and a Logloss of 2.64. Positive predictive features were Skin infection, Upper respiratory, and Mouth infection. Negative: Growth delay, cardiovascular disease, low IgA, Cytopenias, endocrine-metabolic disorders, bronchiectasis, and abscess. See tables and figures for accuracy, numbers, AUC, Gini and LogLoss for the rest of diagnoses.

## Discussion:

We found a good performance of XGBoost to predict any of twelve PID diagnosis in a dataset of over 2,300 patients from the USIDNET registry, with accuracy and AUC between 0.70 and 0.80 for most diseases, and Gini indexes above 0.50.

The strengths of our approach include a large total number of diagnosed patients, a robust machine learning algorithm, and a random train/test split step for cross-validation. The main limitations are a small list of diagnoses, a lack of non-PID diagnoses, and the lack of a prospective validation study.

In the past few years, we have toyed with different approaches to machine learning classification, including linear discriminant analysis, decision tree and random forests (9). In the field of PID diseases, one group in Finland (10–12) and another in Houston at Baylor (13) have attempted machine learning-assisted classification and prediction models, with promising results.

In 2020 at Wuhan, Gao and Ding compared machine learning strategies to predict breast cancer (BC) and cardiovascular disease (CVD) in datasets from the Irving repository (14). They found a 94.74% accuracy for the XGBoost model in BC patients, with the fine needle image of breast lump as the most important predictive value, and a 73.5% accuracy for the CVD dataset XGBoost model, with systolic blood pressure as the most important feature.

There are not many implications for clinical practice yet. The astute clinician should remember to consider the negative predictive features, or counterfactuals, together with the positives. We are calling

this a proof-of-concept study to continue with our explorations. A good performance is encouraging, and the feature importance might aid feature selection for future endeavors.

Next, we want to develop a supervised decision tree or expert system that will make use of both our knowledge base and the most important features from this study, in a user-friendly questionnaire with consecutive classifying questions to predict the genetic diagnosis of new problem cases with suspected PID.

A computational aid to assist the clinical diagnosis of rare diseases might facilitate the work of physicians. Encouraging preliminary results from this and other studies suggest such an aid is feasible to develop. Today still, A.I. stands for both Artificial Intelligence and Almost Implemented.

## Abbreviations

AI, artificial intelligence; AUC, Area under the ROC curve; CGD, chronic granulomatous disease; CVID, common-variable immune deficiency; DGS, DiGeorge syndrome; PID, primary immune deficiencies; ROC, Receiver Operating Characteristic; SCID, severe combined immune deficiency; SHAP, Shapley additive explanations; USIDNET, United States Immunodeficiency Network; XGBoost, extreme gradient boosting.

## Declarations

### ACKNOWLEDGEMENTS:

The U.S. Immunodeficiency Network (USIDNET), a program of the Immune Deficiency Foundation (IDF), is supported by a cooperative agreement, U24AI86837, from the National Institute of Allergy and Infectious Diseases (NIAID). The USIDNET Consortium is composed of over a hundred clinicians who have contributed individually with one to hundreds of patient registrations and their features, available at the registry.

**Funding:** No funding was received for this manuscript.

**CONFLICT OF INTERESTS:** All authors declare no competing interests to disclose.

**Availability of data and material:** All the data supporting this study is available upon request.

**Code availability:** Not applicable. The algorithms, language, workflow, and platforms used are open-source software. Should the reader be interested in a verbatim sequence of steps, we can provide it upon request.

**Consent to participate:** Not applicable, the study does not involve human subjects.

**Authors contributions:** JAMB converted values, deputed the database and reduced dimensionality, applied the algorithm with SRG, analyzed results, critically appraised the rough draft, and approved the

final version of the manuscript. SRG coordinated the application of the algorithm, proposed a solution for imbalanced classification, analyzed results, critically appraised the rough draft, and approved the final version of the manuscript. EHC curated the database, helped depurate and reduce features, built the multidisciplinary team, and approved the final version of the manuscript. EKG contributed with the capture of data for over ten percent of patients in the database, critically appraised the rough draft, and approved the final version of the manuscript. RLF also contributed with the capture of data for over ten percent of patients in the database and approved the final version of the manuscript. KES contributed with over ten percent of patients in the database, established contact with USIDNET, facilitated the international collaboration, critically appraised the rough draft, and approved the final version of the manuscript. The USIDNET Consortium contributed data from thousands of patients with PID, available at the registry. SOLR conceived the study and the publication, wrote the manuscript and coordinated the joint effort.

**Consent for publication:** from all authors and the USIDNET Consortium.

**IRB Ethics Approval:** Exempt study. We protect the confidentiality of patients and families.

## References

1. Tangye SG, Al-Herz W, Bousfiha A, Chatila T, Cunningham-Rundles C, Etzioni A, et al. Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol*. 2020 Jan 1;40(1):24–64.
2. Tan TY, Dillon OJ, Stark Z, Schofield D, Alam K, Shrestha R, et al. Diagnostic impact and cost-effectiveness of whole-exome sequencing for ambulant children with suspected monogenic conditions. *JAMA Pediatr*. 2017 Sep 1;171(9):855–62.
3. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ*. 2016 May;3:353:i2139.
4. Itan Y, Casanova J-L. Novel Primary Immunodeficiency Candidate Genes Predicted by the Human Gene Connectome. *Front Immunol*. 2015;6(April):1–8.
5. Segal M. How doctors think, and how software can help avoid cognitive errors in diagnosis. *Acta Paediatr [Internet]*. 2007/09/14. 2007;96(12):1720–2. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17850393>.
6. Berman JJ. Rare diseases and orphan drugs. First Edit. San Diego: Elsevier; 2014. 354 p.
7. <https://medium.com/@jboscomendoza/tutorial-xgboost-en-python-53e48fc58f73>.
8. SHAP for explainable machine learning [Internet]. [cited 2021 Oct 29]. Available from: <https://meichenlu.com/2018-11-10-SHAP-explainable-machine-learning/>.
9. Murata C, Ramirez A, Ramirez G, Cruz A, Morales J, Lugo-Reyes S. Análisis discriminante para predecir el diagnóstico clínico de inmunodeficiencias primarias: reporte preliminar. *Rev Alerg México*. 2015;(62):125–33.

10. Samarghitean C, Iltanen K, Juhola M, Vihinen M, Rugg G. Machine learning methods for primary immunodeficiency diagnosis. In: 17th biennial meeting of the European Society for Immunodeficiencies, Barcelona. 2016.
11. Samarghitean C. PIDexpert-decision support system for primary immunodeficiencies. [online] [Internet]. University of Tampere. 2008. Available from: <http://bioinf.uta.fi/PIDexpert>.
12. Samarghitean C, Ortutay C, Vihinen M. Systematic classification of primary immunodeficiencies based on clinical, pathological, and laboratory parameters. *J Immunol* [Internet]. 2009 Dec 1 [cited 2013 Nov 8];183(11):7569–75. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19917694>.
13. Rider NL, Cahill G, Motazed T, Wei L, Kurian A, Noroski LM, et al. PI prob: A risk prediction and clinical guidance system for evaluating patients with recurrent infections. *PLoS One* [Internet]. 2021;16(2 February):1–15. Available from: <http://dx.doi.org/10.1371/journal.pone.0237285>.
14. Gao L, Ding Y. Disease prediction via Bayesian hyperparameter optimization and ensemble learning. *BMC Res Notes* [Internet]. 2020 Apr 10 [cited 2021 Oct 29];13(1). Available from: [/pmc/articles/PMC7146897/](http://pmc/articles/PMC7146897/).

## Tables

**Table 1.** Performance measures for the 12 diagnoses predictions, after random train/test split. In all but CVID and DGS, the Class Weight Hyperparameter was used to correct for imbalanced classification.

PIDD	n	Accuracy	AUC	Gini	LogLoss
CVID	1098	0.75	0.75	0.49	8.55
DGS	406	0.897	0.87	0.75	3.55
SCID	202	0.85	0.73	0.47	5.14
CGD	154	0.92	0.83	0.66	2.68
AGAMMA	135	0.88	0.76	0.52	3.84
CORE (ID)	132	0.88	0.72	0.44	4.18
SPAD	117	0.84	0.74	0.49	5.57
WAS	63	0.76	0.74	0.48	8.31
HIGM	46	0.93	0.65	0.30	2.59
NEMO-ID	25	0.97	0.62	0.23	1.01
COMPDEF	12	0.92	0.46	-0.07	2.64
LAD	6	0.997	0.66	0.33	0.09

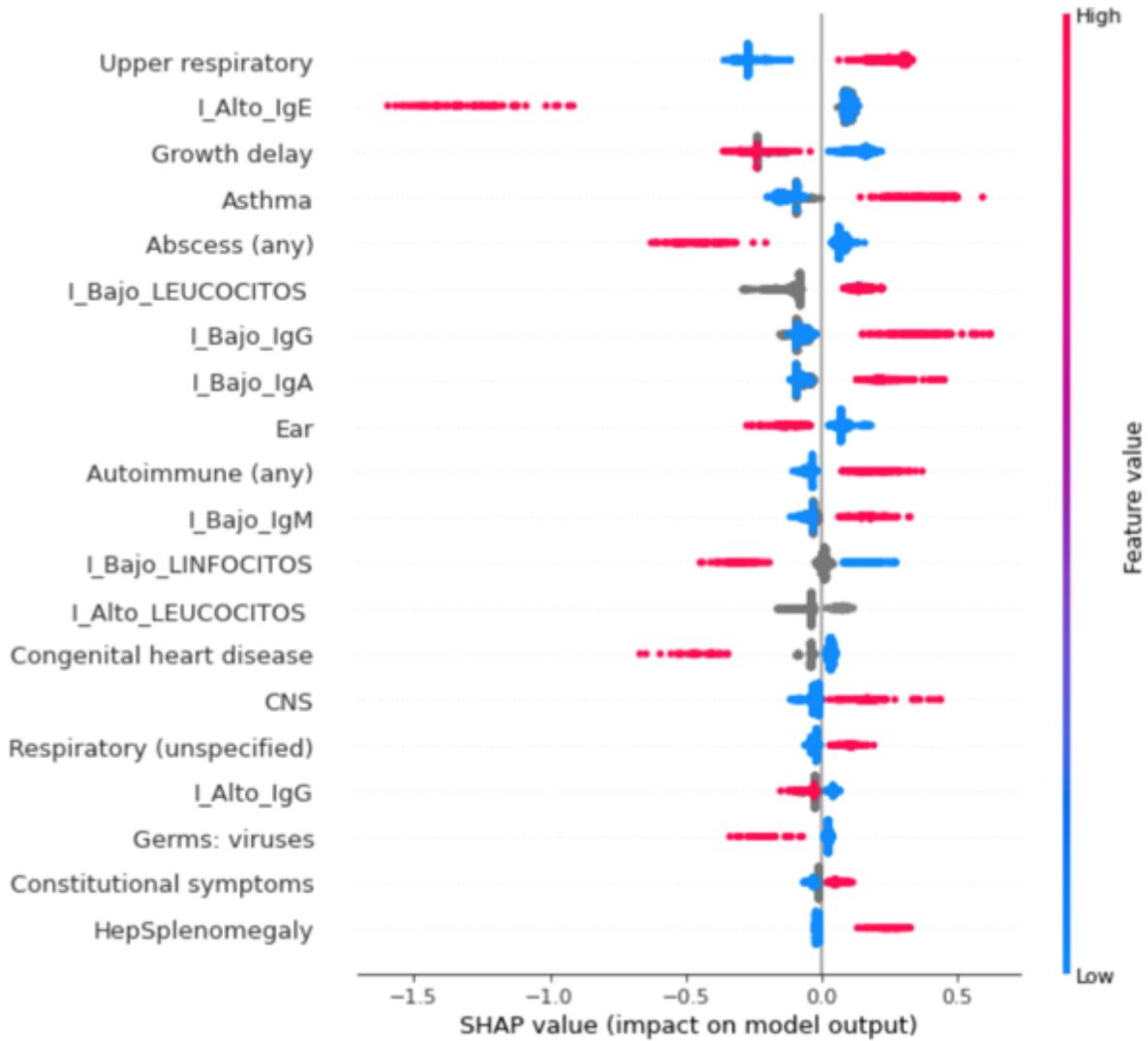
## Figures

	<b>Category</b>	<b>pre_target</b>	<b>target</b>	<b>id_px</b>
<b>0</b>	AB Deficiency	0	0	117
<b>1</b>	AGAMMA	1	0	135
<b>2</b>	CGD	2	0	154
<b>3</b>	COMPDEF	3	0	12
<b>4</b>	CORE	4	0	132
<b>5</b>	CVID	5	1	1098
<b>6</b>	DGS	6	0	406
<b>7</b>	HIGM	7	0	46
<b>8</b>	LAD	8	0	6
<b>9</b>	NEMO	9	0	25
<b>10</b>	SCID	10	0	202
<b>11</b>	WAS	11	0	63
<b>0</b>	TOTAL	[0-11]	[0-1]	2396

**Figure 1**

List of 12 PIDD diagnoses and number of patients from the USIDNET registry. Each target diagnosis (1) is predicted against all others (0).

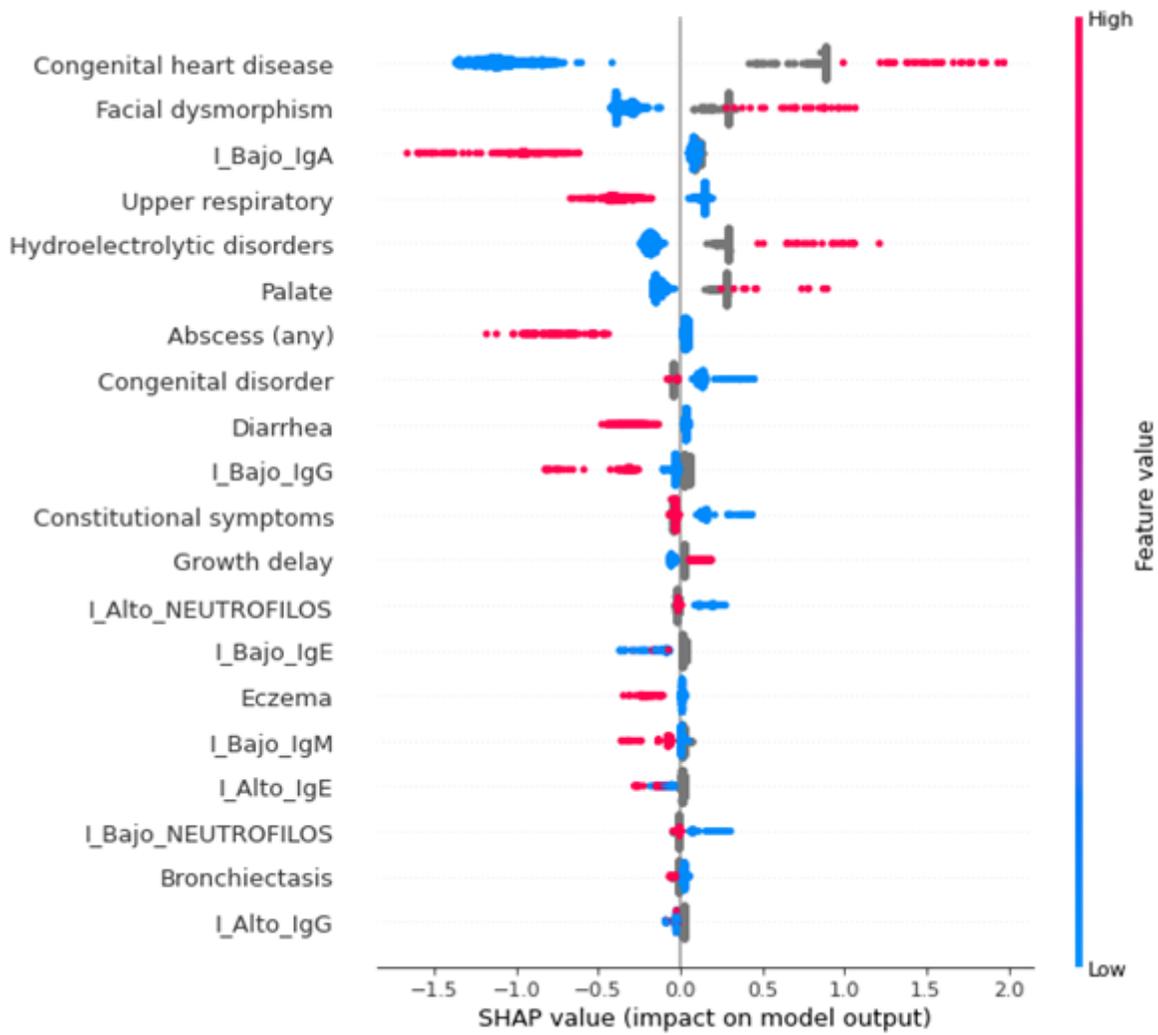
AB deficiency: Specific polysaccharide antibody deficiency; AGAMMA: Congenital agammaglobulinemia; CGD: Chronic granulomatous disease, COMPDEF: complement deficiencies; CORE: immune deficiency, not otherwise specified; CVID: common variable immune deficiency; DGS, DiGeorge syndrome; HIGM, Hyper-IgM syndrome; LAD, Leukocyte adhesion deficiency; NEMO, X-linked ectodermal dysplasia with immunodeficiency; SCID, Severe-combined immune deficiency; WAS, Wiskott-Aldrich syndrome. Table from Google Colab.



**Figure 2**

Twenty most important features to predict the diagnosis CVID and their SHAP values. Higher and red signify greater importance; right/left indicate presence or absence, respectively.

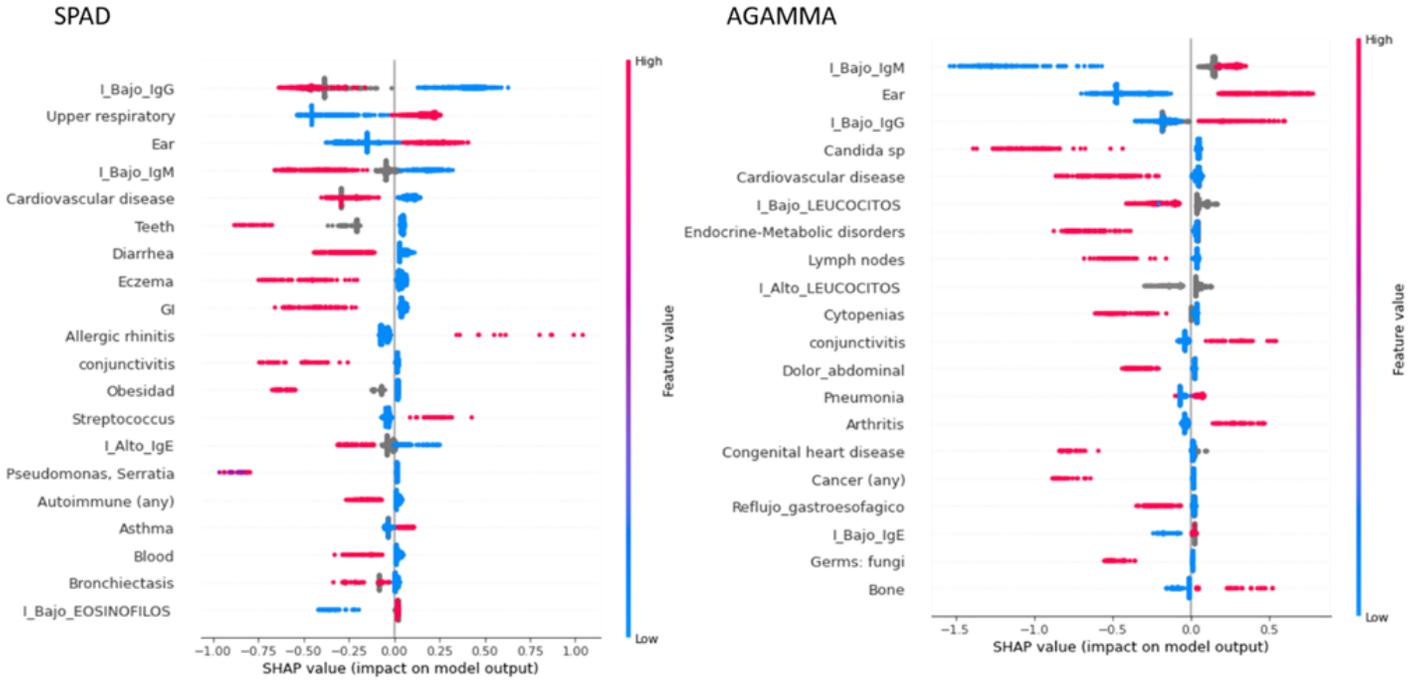
Alto: high, Bajo: low.



**Figure 3**

Top20 features for DiGeorge syndrome diagnosis prediction.

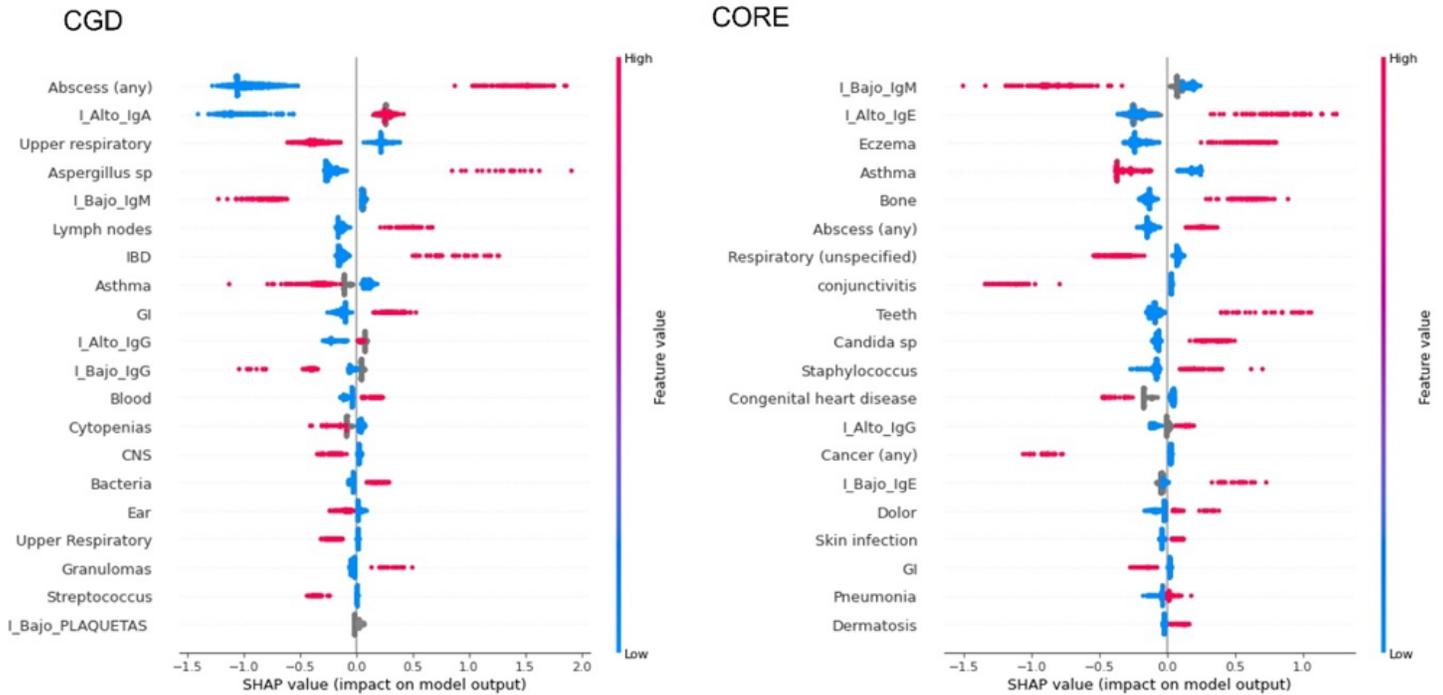
Alto: high, Bajo: low.



**Figure 4**

Top20 features to predict the diagnoses of SPAD (specific polysaccharide antibody deficiency), left, and congenital agammaglobulinemia (AGAMMA), right.

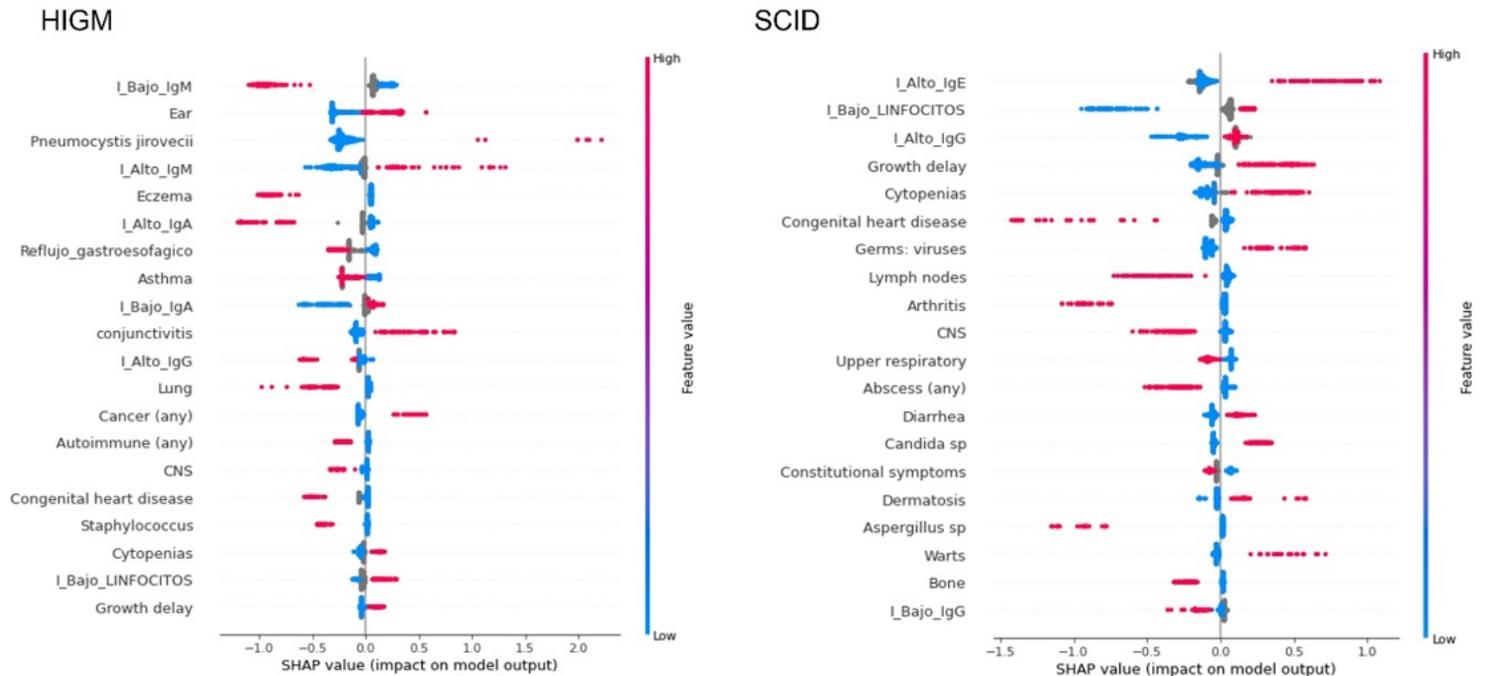
Alto: high, Bajo: low.



**Figure 5**

Top20 features for CGD (Chronic granulomatous disease), left, and CORE (immune deficiency, not otherwise specified), right. Notice how the features in “CORE” suggest Job syndrome.

Alto: high, Bajo: low.

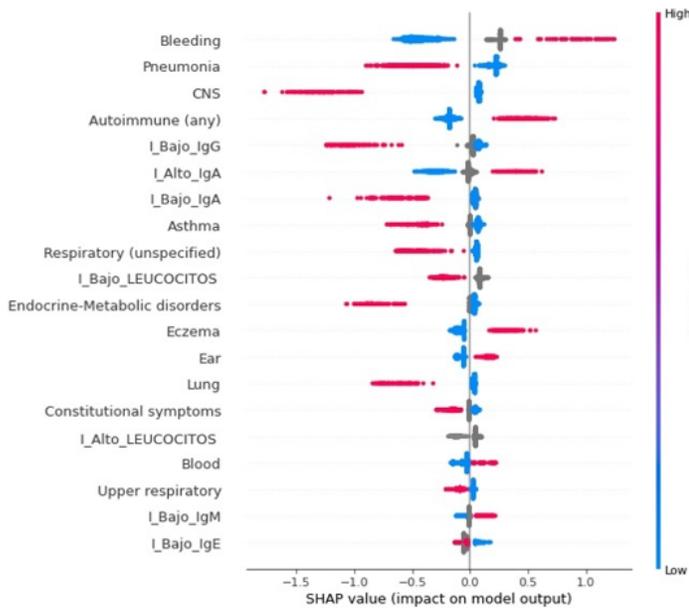


**Figure 6**

Top20 features for HIGM (Hyper-IgM syndrome) and SCID (Severe-combined immune deficiency).

Alto: high, Bajo: low. Reflujo gastroesofágico: gastroesophageal reflux.

## WAS



## NEMO

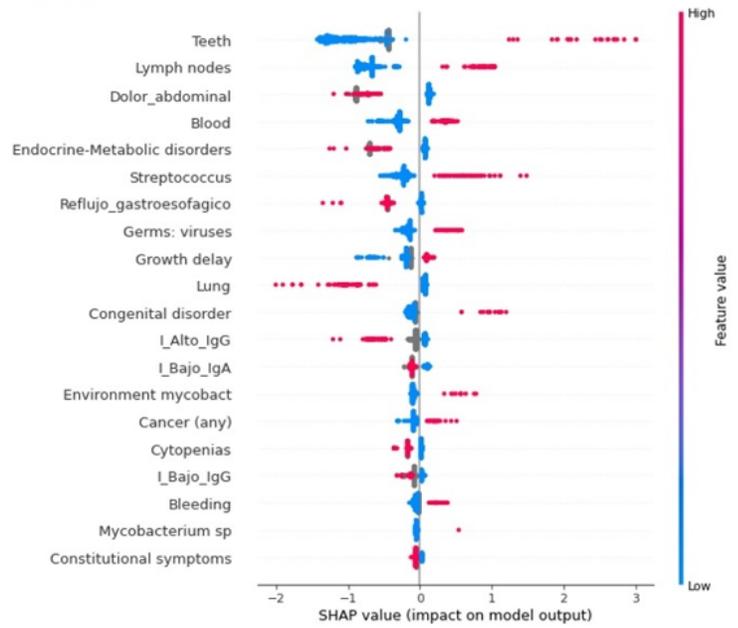
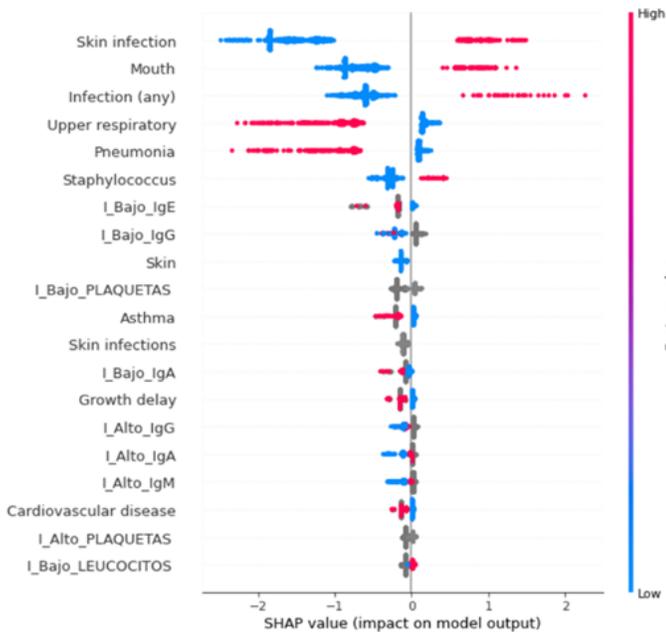


Figure 7

Top20 features for WAS (Wiskott-Aldrich syndrome), left, and NEMO (X-linked anhidrotic ectodermal dysplasia with immunodeficiency), right.

Alto: high, Bajo: low. Leucocitos, leukocytes.

## LAD



## COMPDEF

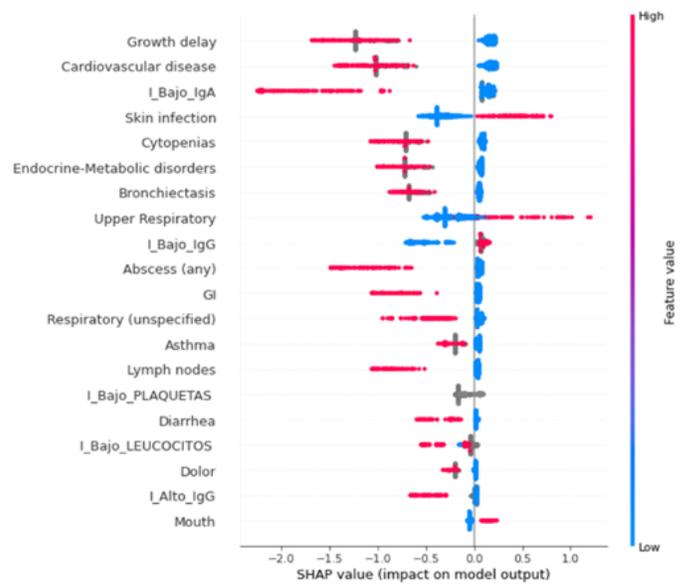
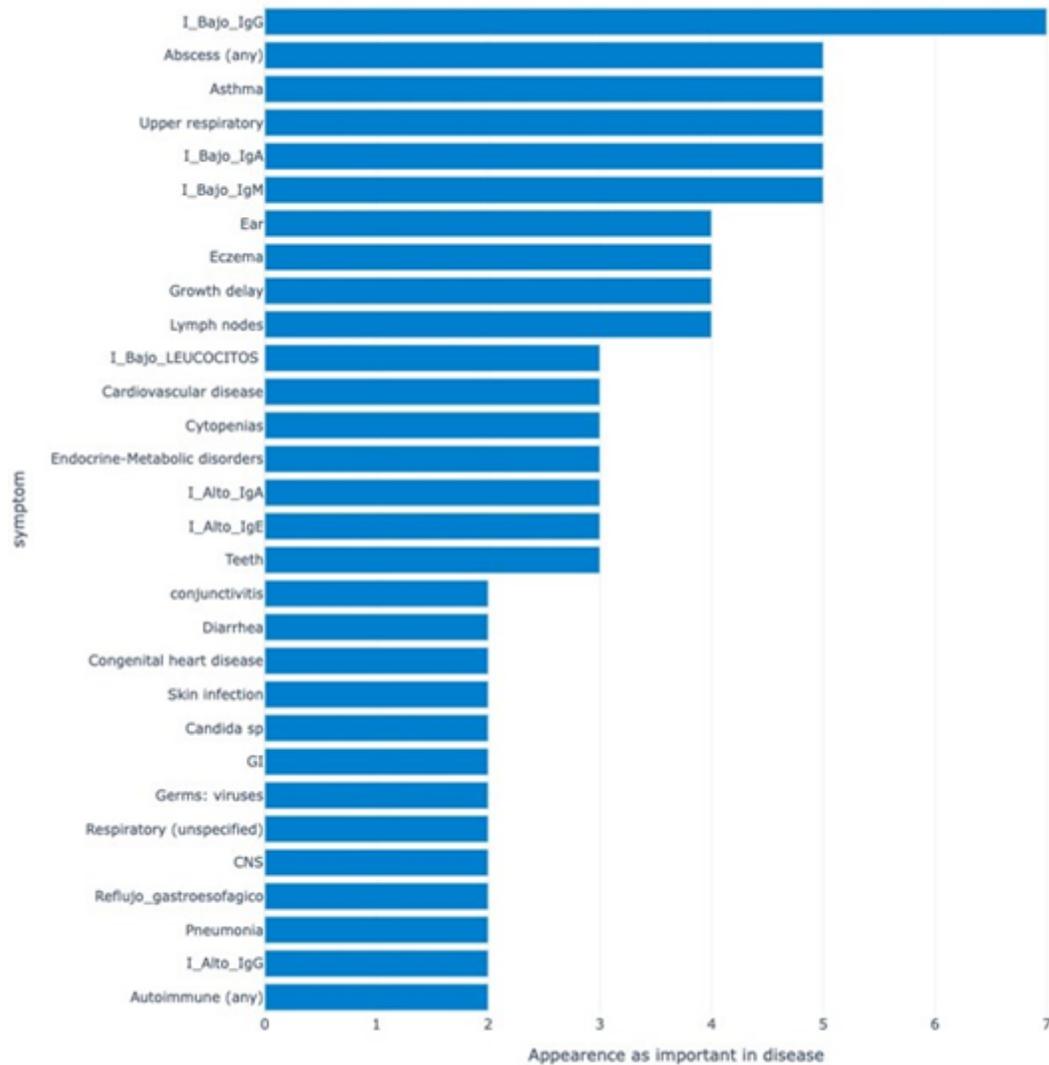


Figure 8

Top20 features for LAD (Leukocyte adhesion deficiency) with n=6, left, and Complement deficiencies (COMPDEF) with n=12, right. Note few important features for LAD, and a predominance of negative or absent features for complement deficiencies.

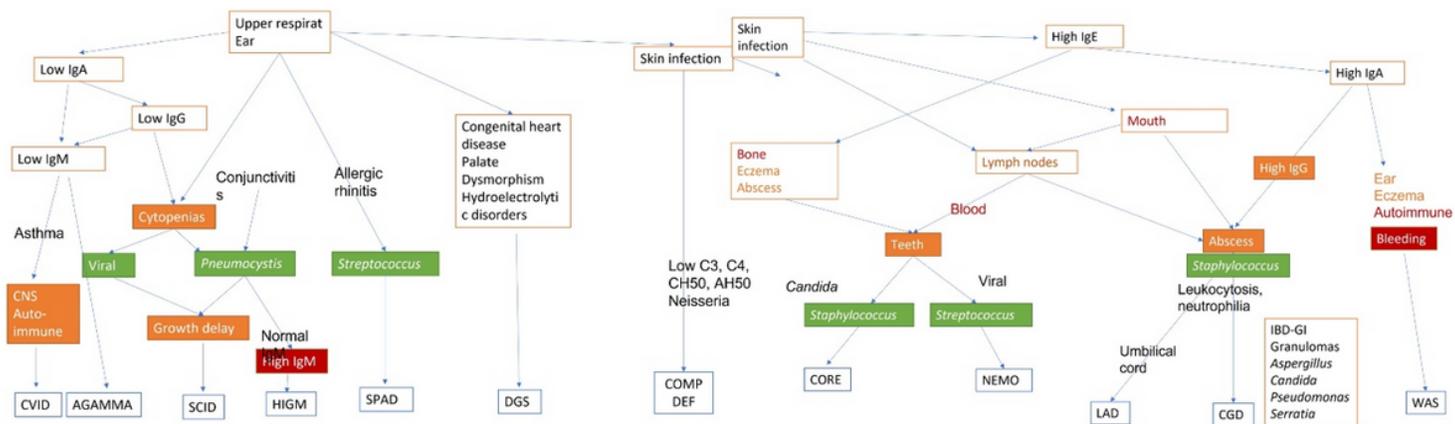
Alto: high, Bajo: low. Leucocitos, leukocytes; plaquetas, platelets.



**Figure 9**

Top20 of most repeated features (absence or presence) across all diagnoses. Low IgG was included as important in 7 predictive models.

Bajo: low, Alto: high, Leucocitos: leukocytes, Reflujo gastroesofágico: gastroesophageal reflux.



**Figure 10**

A flowchart of important questions to ask in the differential diagnosis of 12 classic primary immune deficiencies, derived from the SHAP values provided by our models.

(Supplementary) list of all 238 attributes included in the model after dataset curation and dimensionality reduction.

[Allergic rhinitis', 'Congenital heart disease', 'Cardiovascular disease', 'Autoimmune (any)', 'Heart infection', 'Constitutional symptoms', 'Bleeding', 'Palate', 'Mouth', 'Candida', 'Endocrine-Metabolic disorders', 'GI', 'IBD', 'Kidney', 'Solid tumor', 'Urinary', 'Neuropsychomotor developmental delay', 'Cancer (any)', 'Anemia', 'Arthritis', 'Joint (septic arthritis)', 'Germs: viruses', 'CNS', 'Ear', 'Respiratory (unspecified)', 'Upper Respiratory', 'Lymph nodes', 'Flu', 'Lung', 'Angioedema', 'Eczema', 'Warts', 'Urticaria/Anaphylaxis', 'Molluscum', 'Dermatosis', 'Skin infection', 'Edema', 'Infectious cardiovascular disease', 'Serositis', 'Syndromes', 'Candida sp', 'Teeth', 'Hydroelectrolytic disorders', 'Cholecystitis', 'Jaundice', 'HepSplenomegaly', 'Food allergy', 'Diarrhea', 'Solid Tumor', 'Genital', 'Herpes', 'Proteinuria', 'Growth delay', 'Facial dysmorphism', 'Elevated blood cells', 'Hypogammaglobulinemia', 'Cytopenias', 'Endocrine Autoimmune', 'Bone', 'Bone', 'Leukemia/Lymphoma', 'Eye', 'Pneumonia', 'Congenital disorder', 'Infection (any)', 'Germs: fungi', 'Isolates: VZV', 'Skin conditions (alopecia, vitiligo, psoriasis)', 'Nails', 'Low IgG', 'Low IgM', 'Leukopenia', 'Cardiovascular disease, congenital', 'Bronchiectasis', 'Systemic vasculitis', 'Cirugia\_cardiovascular', 'Isolates: EBV', 'Increased inflammatory markers', 'Abscess (any)', 'Leukoplaquia', 'Sitio\_boca', 'Herpes simplex', 'Sitio\_oidos', 'Mumps', 'Dientes anormales', 'Sitio\_senos paranasales', 'Autoinmunidad\_autoanticuerpos', 'Autoantibodies', 'Absence of lymphoid tissue', 'Skin', 'Alopecia', 'Amenorrea', 'Inborn errors of metabolism', 'Liver or viscerae', 'Autoinmunidad\_hipotiroidismo', 'Sensibilidad\_frio', 'Pubertad\_retrasada', 'Otros\_metabolicos', 'Enterovirus', 'Perinatales', 'Adenomegalia', 'Dolor\_abdominal', 'Sitio\_Herida\_qx', 'Enzimas\_hepaticas\_elevadas', 'Appendicitis', 'Obesidad', 'Sitio\_genitales', 'Ulceras', 'Cirugia\_esplenectomia', 'Bacteria', 'Reflujo\_gastroesofagico', 'Sangrado', 'Cirugia\_abdominal', 'Clostrydium', 'Enfermedad\_celiaca', 'Hepatitis', 'Cryptosporidia', 'Disfagia', 'Parasites',

'Poliposis', 'Colitis', 'Immune deficiency', 'Sitio\_Urinaria', 'Vomito', 'Eosinofilia', 'Isolates: CMV', 'Diarrea crónica', 'Cirugia\_colecistectomia', 'Giardia lamblia', 'HPylori', 'Hospitalizacion', 'Hipoalbuminemia', 'Sitio\_higado', 'Malformacion\_gastrointestinal', 'Hernia', 'Liver abscess', 'Falla\_hepatica/cirrosis', 'Hemangioma', 'Diarrea', 'Desnutricion', 'Granulomas', 'Pseudomonas, Serratia', 'Solid tumors', 'Salmonella sp', 'Cholangitis', 'Spleen', 'Abortos', 'Malformacion\_genitourinaria', 'HSV', 'Human Papillomavirus', 'Sindrome\_hemolitico\_uremico', 'Fistula', 'Malf\_cardiovascular', 'Malf\_genitourinaria', 'Autismo', 'Retraso\_crecimiento', 'Alimentacion\_problematICA', 'Intolerancia\_alimentaria', 'Perdida\_peso', 'Hipoacusia\_sordera', 'Blood', 'Transplant', 'Bone marrow failure', 'Elevated antibody', 'Hemophagocyt', 'Igs\_todaselevadas', 'Antibody deficiency', 'Hemofagocitosis', 'Lymph nodes\xa0', 'Monocitosis', 'Pancitopenia', 'Allergy (any)', 'Malf\_gastrointestinal', 'bone', 'Dedos\_anormales', 'Fracturas', 'Dolor\_toracico', 'Artritis/artralgias', 'Displasia\_cadera', 'SNC', 'Dolor', 'Mialgias', 'Sitio\_hueso', 'Escoliosis', 'Mycobacteria', 'Alergia\_respiratoria', 'Aspergillus sp', 'Sitio\_articulaciones', 'Malf\_esqueletica', 'Osteopenia', 'Mycobacterium sp', 'HPV', 'Linfoma', 'Coccidioides', 'Diabetes', 'Asthma', 'Fever', 'Ataxia', 'Hipertonía', 'Hipotonia', 'Oído', 'Cardiovascular disease', 'Déficit neurológico', 'Convulsiones', 'HZV', 'Telangiectasias', 'Alimentación difícil', 'Retraso desarrollo', 'Mycobacterium Tb', 'Acropaquia', 'Granulomas\xa0', 'Reflujo gastroesofágico', 'Environment mycobact', 'Staphylococcus', 'GramNeg', 'Pneumocystis jirovecii', 'Congenital disease', 'Congenital cardiovascular disease', 'Hipertensión pulmonar', 'Cirugía pulmonar', 'Other infections', 'Artritis/artralgia', 'Severe viral', 'Hair', 'molluscum contagiosum', 'Systemic Lupus', 'Delayed separation of umbilical cord', 'Delayed wound healing', 'Drug allergy', 'Necrosis', 'Skin infections', 'Eritema palmar', 'Fotosensibilidad', 'Piel redundante', 'Ulcera', 'Tejidos blandos', 'Drug Allergy', 'Piel engrosada', 'Acanthamoeba', 'Bowel (intestines)', 'Burkholderia', 'Periodic fever', 'Catheter', 'Hpylori', 'Chronic lung disease', 'CMV/EBV', 'conjunctivitis', 'Mediastinum', 'Muscle', 'Fungemia', 'Measles', 'Streptococcus', 'Escherichia, Klebsiella', 'Other isolates', 'Mononucleosis', 'Mucosae', 'PIC line infection', 'Joints', 'Electrolyte imbalance', 'Upper respiratory', 'Urogenital/perirrectal', 'Inflammation', 'urinary', 'Gérmenes\_atípicos', 'VHH8', 'Aislamiento\_otros\_virus', 'Gérmenes\_piógenos', 'I\_Bajo\_EOSINOFILOS ', 'I\_Bajo\_IgA', 'I\_Bajo\_IgE', 'I\_Bajo\_IgG', 'I\_Bajo\_IgM', 'I\_Bajo\_LEUCOCITOS ', 'I\_Bajo\_LINFOCITOS', 'I\_Bajo\_MONOCITOS ', 'I\_Bajo\_NEUTROFILOS', 'I\_Bajo\_PLAQUETAS ', 'I\_Alto\_EOSINOFILOS ', 'I\_Alto\_IgA', 'I\_Alto\_IgE', 'I\_Alto\_IgG', 'I\_Alto\_IgM', 'I\_Alto\_LEUCOCITOS ', 'I\_Alto\_LINFOCITOS', 'I\_Alto\_MONOCITOS ', 'I\_Alto\_NEUTROFILOS', 'I\_Alto\_PLAQUETAS ']