



A Knowledge-Based System for Evaluation of CIDP Diagnostic Criteria in a Database with 26,000 Nerve Conduction Studies: Computer-Aided Creation of Heuristic Rules and Discovery of New Diagnostic Criteria

Francesco Bonadonna* and James M. Killian

Department of Neurology, Baylor College of Medicine, Houston, TX, USA

OPEN ACCESS

Edited by:

Floriana Grasso,
University of Liverpool, UK

Reviewed by:

Marco Brandizi,
European Bioinformatics Institute
(EMBL-EBI), UK
Jie Zheng,
University of Pennsylvania, USA

*Correspondence:

Francesco Bonadonna
francesb@bcm.edu

Specialty section:

This article was submitted to Digital Health, a section of the journal *Frontiers in ICT*

Received: 08 July 2015

Accepted: 06 November 2015

Published: 16 December 2015

Citation:

Bonadonna F and Killian JM (2015) A Knowledge-Based System for Evaluation of CIDP Diagnostic Criteria in a Database with 26,000 Nerve Conduction Studies: Computer-Aided Creation of Heuristic Rules and Discovery of New Diagnostic Criteria. *Front. ICT* 2:26. doi: 10.3389/fict.2015.00026

This paper describes the implementation and the evaluation of 19 chronic inflammatory demyelinating polyneuropathy (CIDP) diagnostic criteria in a hospital information system (HIS) that contains more than 26,000 motor nerve conduction studies (MNCS), integrated with a knowledge-based system (KBS) that contains and applies neurological knowledge, including the CIDP criteria. The comparison, conducted on 3,750 manual reviewed cases, gave very different results in terms of sensibility and specificity with no one single criterion satisfying for both. Based on personal experience, variations of these CIDP criteria were tried, but the results were not improved. Then, a radically different approach was developed, programming the HIS-KBS to “discover” better criteria in order of their ability to comply with the manual review of cases. The result was a strong reduction of false positives with minimal loss of sensibility. By “reverse engineering” of the computer-generated criteria, it was possible to obtain some new interesting neurologic suggestions, such as the role of H reflex. In conclusion, four points appear of general interest: (1) a large HIS-KBS is fundamental for developing and testing diagnostic criteria and medical procedures, particularly when they are complex; (2) “computer-aided discovery” may create rules that allows the KBS to replicate the human expertise; (3) “reverse engineering” on computer-generated rules may suggest new physiopathological considerations; and (4) this methodology has general application to many other fields of medicine.

Keywords: chronic inflammatory demyelinating polyneuropathy, clinical decision support system, computer-aided discovery, diagnostic criteria, expert system, hospital information system, knowledge-based system, knowledge transfer

INTRODUCTION

A comparative review of complex diagnostic criteria is an ideal task for a hospital information system (HIS) because, like any computerized system, the HIS is able to apply long algorithms in fractions of seconds without errors. Chronic inflammatory demyelinating polyneuropathy (CIDP) is a typical example of a disease for which the diagnostic algorithms are very complex and, furthermore, there

is not a single diagnostic algorithm agreed upon by authors and organizations. This situation is perfectly expressed by the title of the van Dijk paper: *“Too many solutions for one problem”* (Van Dijk, 2006). Many authors (Barohn et al., 1989; Haq et al., 2000; Van Den Bergh et al., 2000; Saperstein et al., 2001; Molenaar et al., 2002; Nicolas et al., 2002; Thaisetthawatkul et al., 2002; Magda et al., 2003; Sander and Latov, 2003; Brainin et al., 2004; Van Den Bergh and Pieret, 2004; Rajabally et al., 2005, 2009; Van Asseldonk et al., 2005; Wilson et al., 2005; Tackenberg et al., 2007; De Sousa et al., 2009; Koski et al., 2009; Viala et al., 2010; Bromberg, 2011) have compared the different criteria and/or proposed a new one. An example is a paper by Bromberg (2011) that compares 16 CIDP diagnostic criteria. We tried to reproduce this work with three important differences: (1) reliability and speed due to the computerized procedure, (2) number of cases because the HIS we use contains more than 26,000 nerve studies, and (3) application on peripheral neuropathy cases that have not been preselected for CIDP.

Note that the 26,000 nerve studies are the activity of a neurologist (Dr. Killian) in a neuromuscular electromyography (EMG) lab performed on cases referred from 1991 to 2013. They are not a representative sample of the population but are a reasonable example of the case mix that was seen by a neurologist.

Our first task was to see how the criteria perform on unselected cases and without the “neurological common sense” that is always present when a neurologist applies a diagnostic criteria to a real case. The second task was to develop, transfer to the knowledge-based system (KBS), and test new criteria in order to reach better results. As it will be reported, the first task was fulfilled but the second one was much more difficult than expected and the results were less than expected. This led to the development of new tools that have been useful for CIDP diagnosis but are also of general methodological interest for any computerized system that deals with complex diagnostic problems.

MATERIALS AND METHODS

The Neurological HIS–KBS

The system used for this paper, called “Euristic,” is described in the methods section because it is the instrument we have used to reach the results described in the following sections.

Euristic has two components: an HIS, with special functions for neurology, and a KBS. As an HIS, it contains general medical folders, specialized neurological medical folders and special extensions for the neurologic subspecialty called “electro-neurodiagnostics” that deals with EMG and motor nerve conduction studies (MNCS).

The other part of Euristic is a KBS, also known as expert system (ES), that is a computer program that “emulates” the behavior of a human expert (Buchanan and Duda, 1982; Ardizzone et al., 1988; Bonadonna, 1990; Lhotska et al., 2001; Sasikumar et al., 2007; Patel et al., 2013). In order to do this, many approaches have been taken in literature. One of the most common uses “concepts” and “rules.” Concepts correspond to a “descriptive knowledge,” and in this case, are nerves, muscles, diseases, symptoms, and so on. They are organized in a hierarchy where concepts are linked by a logical relationship, also known as “IS A.” This organization or concepts

or in general “entities” is also called “ontology.” For example, in **Figure 1**, the concept “FWave lat(ency) del(ay) $\geq 20\%$ and Damp(litude) red(uction) less 20%” IS A “Criteria for Van Den Bergh and EFNS” that IS A “Criteria for CIDP in single nerve.” When dealing with a specific patient study, some concepts may become “true” or “false” depending on the specific patient’s data. Rules represent the “procedural knowledge” and correspond to the way in which a neurologist “teaches” how to verify if a patho-physiologic concept is true or false in a given patient’s study. HIS and KBS are strictly integrated: for example, in our system, the concepts that the KBS finds true for each study are saved in the HIS like any other patient data (EMG, MNCS, etc.).

The dimensions of the KBS (concepts and rules) and of the HIS (patient studies) are both very critical. As it will be clear at the end of this paper, the most interesting results are likely to appear with larger numbers of both KBS and HIS. Our system has been in daily use since 1991 and contains: 25,350 EMG studies, 26,181 MNCS, and a total of 133,020 motor nerves tested. The KBS contains 3,954 “concepts” and 3,382 “rules.” The HIS contains 4,404,161 verified concepts, with a maximum of 822 and an average of 140 per study.

Knowledge Transfer

Teaching the KBS (the so called “Knowledge Transfer” – KT) is a complicated manual task that may become faster with experience. Sometimes it is easy when the neurological concepts and procedures are clearly defined in the medical textbooks or by the experience. For example, a known single concept such as “Delayed Distal Latency (Dlat del)” (see **Figure 1**) becomes true if the parameter DL in milliseconds for that particular nerve is greater than the maximum normal value by at least a certain percentage. In other cases, the diagnostic criteria are complex, as Bromberg writes for CIDP: *“Although some sets (of criteria) are appropriate for formal clinical trials, their complexity makes them difficult to apply in the clinic or electromyography laboratory”* (Bromberg, 2011).

Computer programs have no problems with complexity: once the program learns all the rules needed to verify a concept, it will apply them without errors in a few milliseconds, while for humans, the application of complex criteria will always remain long and error prone.

Two important problems remain: the “implicit clinical sense” and the “real performance.”

- The *implicit clinical sense* is not a problem with the above definition of “Delayed DL”: once the two numerical parameters “normal DL” and “percentage of delay over the normal” are defined, the result is easily reached, by the KBS or a neurologist. It is a problem instead with CIDP diagnostic criteria, for which the KT process has been more difficult because criteria are written for neurologists and not for computer programmers, so definitions are sometimes “incomplete” from the programmer. This happens because the author knows that other neurologists will understand and complete the criteria using their intuition and expertise, i.e., their clinical sense. This problem may be solved by asking the author, or another neurologist, to “explain” every step of the reasoning so that it becomes as clear as in the above “delayed DL” definition. As will be evident later, this implies a reflection on which is the real diagnostic reasoning

Criteria for CIDP in single nerve. Patients 11,041; Studies 11,863; Nerves 17,173

Criteria for Van Den Bergh and EFNS. Patients 6,902; Studies 7,319; Nerves 11,303

- DLat del $\geq 50\%$ (EFNS A). Patients 1,110; Studies 1,188; Nerves 1,766
 - Rule for DLat del $\geq 50\%$ (single nerve, EFNS A)
 - The nerve is motor and D Lat is at least 50% over the upper limit of normal (limited to: Median, Ulnar BE, Ulnar at first DI BE, Tibial, Common Peroneal)
- CV red $\geq 30\%$ (EFNS B). Patients 974; Studies 1,072; Nerves 1,733
 - Rule for CV red $\geq 30\%$ (single nerve, EFNS B)
 - The nerve is motor and CV is at least 30% under the lower limit of normal (limited to: Median, Ulnar BE, Ulnar at first DI BE, Tibial, Common Peroneal)
- FWave lat del $\geq 20\%$ and DAmp red less 20% (EFNS C1). Patients 547; Studies 570; Nerves 712**
 - Rule for FWave lat del $\geq 20\%$ and DAmp red less 20% (single nerve, EFNS C1)
 - The nerve is motor and F-wave lat is at least 20% over the upper limit of normal (limited to: Median, Ulnar BE, Ulnar at first DI BE, Tibial, Common Peroneal)**
 - The nerve is motor and D Amp is up to 20% under the lower limit of normal (limited to: Median, Ulnar BE, Ulnar at first DI BE, Tibial, Common Peroneal)

Hide details Move up Move down Search rules and concepts

Function: **Single nerve conclusion** Last change: **Mar 24 2009**

Motor or Sensory Motor

Whose parameter
F-wave latency

Limited to the following nerves

- Median
- Ulnar BE
- Ulnar at first DI BE
- Tibial
- Common Peroneal

Has one of the values

Or is, with respect to normal

from 20 % to % over

from % to % under

Or is in the numeric range

FIGURE 1 | In this figure, from the “Euristic” HIS–KBS, the concept in the red rectangle is the translation of one of the concepts used in the “EFNS subcriterion c.” The corresponding rule is in green and includes two conditions. The first of the two conditions is enhanced in blue and corresponds to the expanded view in the lower part of the figure. The condition uses a “Function” called “Single nerve conclusion” and tells to the KBS to apply this function to each “Motor” nerve, “limited to the following nerves”: “Median, Ulnar BE, Ulnar at first DI, Tibial, and Common Peroneal,” and to look for the “parameter” “F-wave latency” that must be 20% or more above the normal. If this condition is verified in a patient study, and also the second condition is verified, then the concept in the red rectangle will be considered true. Note that the system tells in how many patients, studies, and nerves, this particular concept has been found true in our archive. Note also the “ontology” of concepts represented graphically by the tree structure: each concept is linked to the one upper in the tree by the “IS A” relationship.

that a skilled neurologist uses in his clinical routine when dealing with complex cases that require a different reasoning than represented in published flow charts.

- The “real performance,” when the KBS is applied to a large set of real patient data, is the final test. If the KBS applies the diagnostic criteria without any mistake but the results are not as expected, then it could be that the original criteria from literature do not “contain” all the expertise that is used by the neurologist. To demonstrate this by use of the KBS is an important result by itself and could be a starting point for further investigation on the KT between physicians and on diagnostic criteria in general.

In conclusion, it can be said that the KT process clarifies the following two problems and offers solutions:

- The first solution is the KT process itself: because the KBS itself “forces” the rules to be complete and unambiguous. So, if they can be written in the KBS and work, it means that they are “complete” and “unambiguous”; otherwise they just not work.
- The second solution is the possibility to test the real performance of the diagnostic criteria on large sets of clinical data and to demonstrate if the criteria perform as they are intended.

“Teaching” 19 CIDP Criteria to the KBS

This section describes how the CIDP diagnostic criteria have been transferred from paper to the KBS.

First Step: Start with Diagnostic Criteria from Literature

The EFNS criterion for definite CIDP is one of the 19 CIDP criteria existing in medical literature. To reach the diagnosis of definite CIDP, “at least one” of seven subcriteria must be true. In particular, subcriterion c is: “Prolongation of F-wave latency $\geq 30\%$ above ULN in two nerves ($\geq 50\%$ if amplitude of distal negative peak CMAP $< 80\%$ of LLN values.)” See Joint Task Force of the EFNS and the PNS (2010), p. 3.

Second Step: “Translate” the Criteria in a Language That the Computer Can Understand

Figure 1 shows how the criteria look when “transferred” in our KBS. The lines in blue are concepts, the lines in green are the names of the rules, and each rule is composed of one or more “conditions” that appear in gray. The enhanced line is the first of the two conditions of “rule for F-Wave lat(ency) del(ay) $\geq 20\%$ and D(istal)Am(plitude) reduction less 20% (EFNS C1).” The condition uses a function called “Single nerve conclusion” and tells the KBS

to apply this function to each “Motor” nerve, limited to Median, Ulnar BE, Ulnar at first DI, Tibial, and Common Peroneal, and to look for the parameter “F-wave latency” that must be 20% or more above the normal (above the normal in this case means “delayed”). The KBS summarizes all that with the statement appearing in blue: “*The nerve is motor and F-wave lat is at least 20% over the upper limit of normal (limited to Median, Ulnar BE, Ulnar at first DI, Tibial, Common Peroneal).*” Note that a lot of work has been done to allow the KBS to describe itself in simple English.

The second “condition” of the rule says: “*The nerve is motor and D Amp is up to 20% under the lower limit of normal (limited to Median, Ulnar BE, Ulnar at first DI, Tibial, Common Peroneal).*” If, for a given nerve, both conditions are true, then the conclusion in blue: “*FWave latency delayed \geq 20% and Damp reduction less 20% (EFNS C1)*” becomes true for that specific nerve.

The essential structure of the KBS is this: a concept, a rule attached to it, one or more conditions per rule and a number of functions that interface the conditions with the patient data. In this way, the KBS in a specific clinical case is able to conclude if each concept is true or false.

Structure of the KBS

Note that after the concept “*FWave latency delayed \geq 20% and Damp reduction less 20% (EFNS C1)*” (in blue in **Figure 1**), there is an indication of how many patients, studies, and nerves have been found in the database with that specific conclusion being true (547 patients and 712 nerves). This, as said before in the introduction, is one of the many examples of the strong integration between HIS and KBS: if the concept has been found true, then it is saved in the database as any other clinical data (numerical or qualitative) and the system can operate on all of them as a whole. It becomes one of the millions of verified concepts that are contained in the Euristic HIS (see The Neurological HIS–KBS).

Note that the conclusion in **Figure 1** is relative to a single nerve. **Figure 2** shows all the concepts at single nerve level that have been created in the KBS in order to transfer the CIDP criteria. In fact, all of them belong to a group called “*Criteria for CIDP in single nerve,*” as can be seen in **Figure 1**. In **Figure 2**, rules and their conditions have been hidden. Note that when single nerve conclusions (we use this term with the same meaning of concept) have been used in more than one criterion, this is indicated in their names, e.g., “*EFNS F equal to Sap1B.*” This means that condition EFNS F is used in EFNS criteria (Hughes et al., 2006) but is identical to condition 1B of Saperstein criteria (and in fact the numbers of patients, studies, and nerves are identical for the two). Clearly, this is redundant, but some redundancy is useful for better “human” reading and changes nothing for the computer.

To confirm the concept in the red box in **Figure 1**, another step is needed: the system must find at least two nerves with that conclusion. This is done by the rule shown in the red box in **Figure 3**.

It is interesting to see that rules in **Figure 3** operate on the results of other rules. The KBS automatically decides which rules have to be applied first in a given case. They are called “*layer zero rules*” because their conditions do not use results from other rules but only raw data. Then, the KBS executes rules in layer 1 that could use raw data but also results from rules of layer 1.

Then, the same for layer 2 and so on. Just to have an idea of the complexity of a KBS, by now the Euristic KBS contains 20 layers. Note that the structure and the functions of the KBS, also called the inference engine, have been completely programmed in Pascal language with the Delphi™ development environment, like also the HIS. The underlying database used, for both the KBS and the HIS, is an open source SQL database management system, called Firebird™.

Another important feature of the KBS is the inheritance in a tree. For example, “*CIDP EFNS Definite*” (**Figure 3**, third line) is the “parent” of the 10 conclusions that appear a level below in the tree. The KBS allows stating that, if a given condition is found true, then also its parent conclusion is set to true. That is exactly what happens in this case for the nine conditions A, B, C1, C2, D, E1, E2, F, and G: the parent conclusion “*CIDP EFNS Definite*” is true if one or more of its first nine “child conclusions” is found true. Condition 10 is a G with a different threshold that stays there for research and statistical purposes but has not the property to set its parent to true.

Incidentally, it is interesting to note that the conceptual hierarchy (the ISA relationship) is not only a descriptive feature, very useful for maintaining a logical order in thousands of concepts in the KBS, but also a tool that allows to write conditions that look much more similar to the natural language.

In the HIS database, the concept “*CIDP EFNS Definite*” has been verified in 1,291 studies and 1,200 patients (**Figure 4**). Obviously, the difference between studies and patients depends on the fact that a number of patients have been studied more than one time.

A consideration about the KT: this kind of work is very close to the process used to explain neurologic or medical reasoning to a “non-neurologist” or “non-medical” person, for example, a computer programmer (Reggia et al., 1985; Sasikumar et al., 2007; Darlington, 2011). In other words, the KBS may be seen as a “non-neurologist, non-physician” that has to reach a correct diagnosis by working only on the data in the electronic patient folder, using a set of rules to reach the diagnosis without any other knowledge or clinical sense, which is exactly what happened with the programmers of Euristic.

The Internal Conclusions

Looking to **Figure 3**, it is possible to observe that instead of creating nine child conclusions, nine rules for verifying them, and a parent conclusion that is the real diagnosis, should have been simpler to have nine rules that point directly to the “parent.” The observation is correct and the final result is exactly the same.

The more complex solution, and in particular the creation of the nine *internal conclusions*, has the following reasons:

1. The original criteria (**Figure 2**) have all these intermediate steps and in this way, the structure in the KBS is more similar to the original one in the paper, and, for this reason, is easier to read for a neurologist or a student.
2. The internal conclusions are saved in the database, so it is possible to study them with database queries and statistical tools in the future. With a single conclusion, this would not be possible.

- [-] Criteria for CIDP in single nerve. Patients 11,067; Studies 11,894; Nerves 17,215
 - [-] Criteria for Van Den Bergh and EFNS. Patients 6,915; Studies 7,335; Nerves 11,325
 - [+] DLat del $\geq 50\%$ (EFNS A). Patients 1,112; Studies 1,190; Nerves 1,769
 - [+] CV red $\geq 30\%$ (EFNS B). Patients 976; Studies 1,074; Nerves 1,735
 - [+] FWave lat del $\geq 20\%$ and DAmpr red less 20% (EFNS C1). Patients 548; Studies 571; Nerves 713
 - [+] FWave Lat del $\geq 50\%$ and DAmpr red $\geq 20\%$ (EFNS C2 = Sap4B). Patients 37; Studies 38; Nerves 45
 - [+] Absent FWave and DAmpr red no more 80% (EFNS D). Patients 1,259; Studies 1,285; Nerves 1,650
 - [+] Absent FWave and DAmpr red no more 60% (EFNS D Baylor). Patients 1,049; Studies 1,067; Nerves 1,305
 - [+] Ampr drop $\geq 50\%$ and DAmpr red no more than 80% (EFNS E). Patients 2,032; Studies 2,103; Nerves 2,578
 - [+] MWave incr $\geq 30\%$ (EFNS F equal to Sap1B). Patients 23; Studies 23; Nerves 28
 - [+] Distal CMAP duration > 12 and DAmpr red no more 20% (modified EFNS G). Patients 655; Studies 662; Nerves 1,037
 - [+] Distal CMAP duration > 9 (EFNS G. Original for internal use). Patients 4,478; Studies 4,648; Nerves 13,379
 - [+] Ampr Drop $\geq 30\%$ and DAmpr reduced no more 80% (EFNS probable). Patients 4,246; Studies 4,426; Nerves 5,485
 - [+] FWave lat del $\geq 15\%$ and DAmpr red less 20% (EFNS C1) Variation for test. Patients 950; Studies 986; Nerves 1,258
 - [+] FWave lat del $\geq 25\%$ and DAmpr red less 20% (EFNS C1) Variation for test. Patients 950; Studies 986; Nerves 1,258
 - [-] Criteria for Saperstein, AAN and INCAT. Patients 9,704; Studies 10,368; Nerves 16,190
 - [+] Ampr Drop $\geq 50\%$ (Sap1A). Patients 2,188; Studies 2,268; Nerves 2,810
 - [+] M wave increase $\geq 30\%$ (Sap1B equal to ENFS F). Patients 23; Studies 23; Nerves 28
 - [+] CV red $\geq 20\%$ and DAmpr red no more than 20% (Sap2A). Patients 1,033; Studies 1,121; Nerves 1,662
 - [+] CV red $\geq 30\%$ and DAmpr red more than 20% (Sap2B). Patients 637; Studies 684; Nerves 910
 - [+] DLat del $\geq 25\%$ and DAmpr red no more than 20% (Sap3A). Patients 1,657; Studies 1,744; Nerves 2,282
 - [+] DLat del $\geq 50\%$ and DAmpr red more than 20% (Sap3B). Patients 676; Studies 711; Nerves 946
 - [+] FWave Lat del $\geq 25\%$ and DAmpr red no more than 20% (Sap4A). Patients 344; Studies 356; Nerves 421
 - [+] FWave del $\geq 50\%$ and DAmpr red $> 20\%$ (Sap4B equal to ENFS C2). Patients 37; Studies 38; Nerves 45
 - [+] MWave incr $> 15\%$ (AAN1B). Patients 32; Studies 33; Nerves 49
 - [+] Ampr Drop $\geq 20\%$ (INCAT). Patients 8,237; Studies 8,690; Nerves 11,583
 - [-] Criteria for Nicolas. Patients 24; Studies 24; Nerves 30
 - [+] MWave incr $\geq 15\%$ (Nicolas TD, No Tibial). Patients 24; Studies 24; Nerves 30
 - [-] Criteria for Albers-Kelly. Patients 5,131; Studies 5,436; Nerves 7,754
 - [+] CV red $\geq 25\%$ (Albers-Kelly). Patients 1,492; Studies 1,620; Nerves 2,634
 - [+] Ampr Drop $\geq 30\%$ (Bromberg and Albers-Kelly). Patients 4,343; Studies 4,530; Nerves 5,637
 - [+] FWave Lat del $\geq 30\%$ (Bromberg and Albers-Kelly). Patients 275; Studies 280; Nerves 362
 - [-] Criteria for Bromberg. Patients 2,266; Studies 2,413; Nerves 3,578
 - [+] DLat del $\geq 30\%$ (Bromberg). Patients 2,266; Studies 2,413; Nerves 3,578
 - [-] Criteria for Magda. Patients 8,918; Studies 9,532; Nerves 15,541
 - [+] Category A - demyelinating-range abnormalities (Magda). Patients 7,039; Studies 7,480; Nerves 11,669
 - [+] Category B - possible partial CB (Magda). Patients 4,343; Studies 4,530; Nerves 5,637
 - [+] Category C - other abnormalities (Magda). Patients 7,887; Studies 8,422; Nerves 13,078
 - [-] Criteria for Tackenberg. Patients 8,122; Studies 8,741; Nerves 15,457
 - [+] CV red $> 10\%$ and DAmpr red no more than 50% (single nerve, Tack1A). Patients 3,639; Studies 3,945; Nerves 6,643
 - [+] CV red $> 20\%$ and DAmpr red $> 50\%$ (single nerve, Tack1B). Patients 1,109; Studies 1,154; Nerves 1,515
 - [+] DLat del $> 15\%$ and normal DAmpr (single nerve, Tack2A). Patients 1,673; Studies 1,741; Nerves 2,167
 - [+] DLat del $> 25\%$ and reduced DAmpr (single nerve, Tack2B). Patients 2,133; Studies 2,255; Nerves 3,225
 - [+] FWave del $> 25\%$ or absent (single nerve, Tack4). Patients 457; Studies 471; Nerves 583
 - [+] Ampr Drop $> 30\%$ (Tack3, AAN1A) No Tibial. Patients 4,343; Studies 4,530; Nerves 5,637
 - [-] Criteria in specific nerves. Patients 341; Studies 351; Nerves 362
 - [+] Ampr drop $\geq 50\%$ and DAmpr red no more than 80% (EFNS E) in Ulnar AE, BE. Patients 341; Studies 351; Nerves 362

FIGURE 2 | Concepts created for CIDP at single nerve level. For each one, the KBS shows the number of patients, studies, and nerves in which the concept was found true. Rules and their conditions are not displayed to limit the size of the figure.

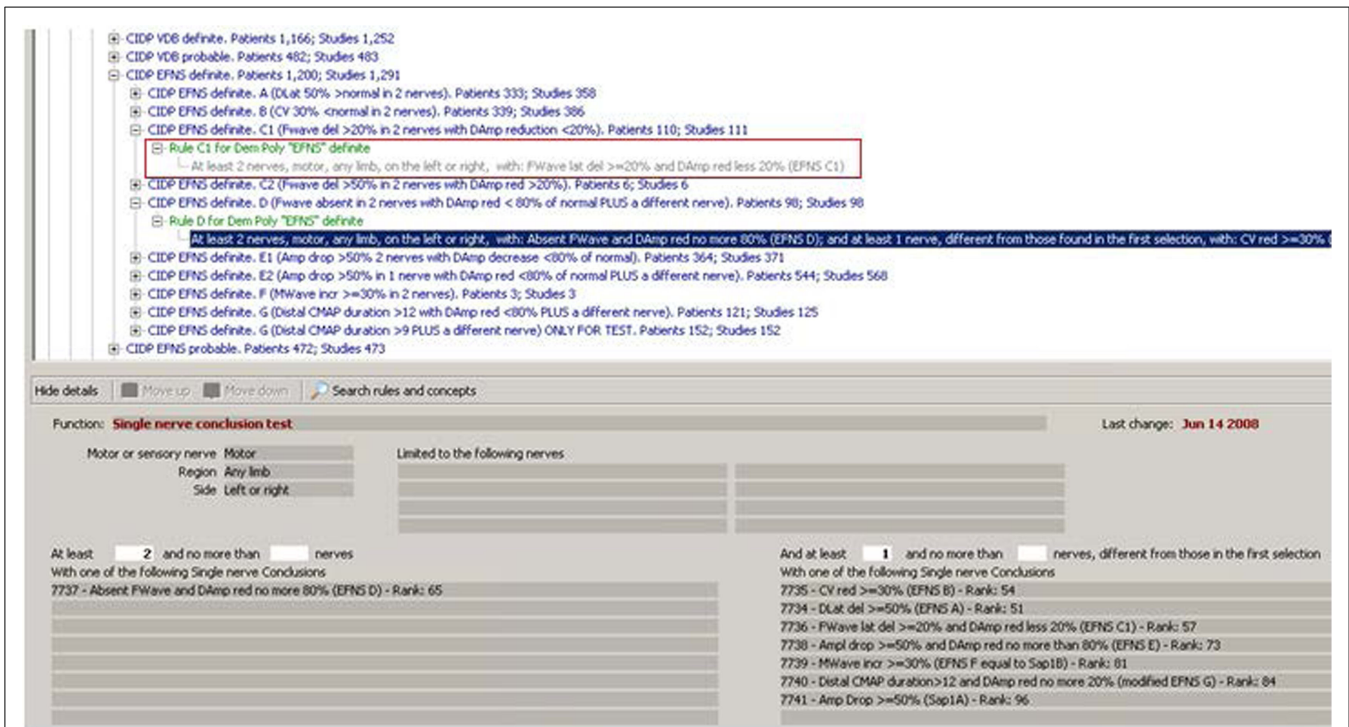


FIGURE 3 | Rules operating on multiple nerves. The rule in the red rectangle is the one that completes the procedure necessary to confirm the concept in Figure 1 ("Rule C1 for Demyelinating Polyneuropathy 'EFNS' Definite"). As can be read in its description, it looks for at least two nerves with the conclusion of Figure 1. The other expanded rule shows a function that operates in a more complex way: it not only searches for at least two nerves with a given result (lower left in the figure), but also verifies that at least one another nerve, different from those with the first conclusion, have one of the conclusions listed (lower right in the figure).

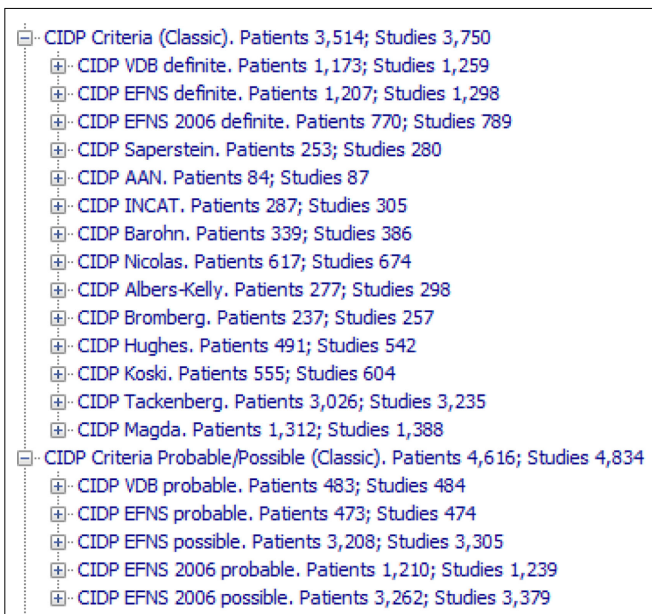


FIGURE 4 | The 19 CIDP criteria (14 for definite CIDP and five for probable or possible CIDP) that have been transferred to the KBS and tested on the archive. They are called "Classic" to distinguish from the ones that we have developed. Note that the criteria are applied to any study in the archive, without the filters that rule out studies without enough data.

3. Dividing a reasoning process or an algorithm into small logical steps is something that makes it possible in the future to reassemble the reasoning in different ways.
4. Practical experience (aka heuristics): doing this has been useful in the past for other sections of the KBS, so this "more steps approach" was adopted also without knowing the way it could be used in the future.
5. No contraindications: as the KBS will become more complex, the number of conclusions associated with each case will be larger but the performance of the KBS will be practically the same (to be exact, it will require more time in the order of milliseconds per case and more space in the archive). Moreover, intermediate conclusions will not appear in the final report and that is why they are called "internal."

Figure 4 shows all the 19 CIDP criteria (14 for definite CIDP and five for probable or possible CIDP) that have been transferred to the KBS and tested on the archive. Figure S1 in Supplementary Material shows the 14 definite conclusions, with all the "intermediate conclusions" one level below.

Note that the "probable" or "possible" criteria in Figure 4 are not just "relaxed" or "a little less than true" criteria, they are exactly like the others only written with different combination of conditions. This is very different from the "fuzzy systems" (Torres and Nieto, 2006; Lee and Wang, 2011) in which any concept may have a continuous degree from false to true.

While they may be very useful in different fields of medicine, they are not suited for a diagnostic reasoning in which each concept, including the “possible” and “probable,” has its strict definition.

Justification

Before going to the results, it is important to see how the KBS “justifies” a conclusion for a given study. **Figure 5** shows a small part of the justification of a study in which the KBS explains why three conclusions have been found true for that patient. In particular, the first one is visible in **Figure 3**. The KBS repeats the definition of the rule (“at least 2 nerves, motor, any limb . . .”) and then says that it is “TRUE” because the nerve “Common Peroneal” on the left and the same nerve on the right have “EFNS D” and because “Ulnar BE” (the different nerve required by original criteria) has “EFNS E.” The second rule is for another CIDP criterion. The third rule, instead, is an example of the many other rules (not for CIDP) that stay in the KBS. In this specific case, there is a rule dealing with EMG results and using the neuroanatomical knowledge inside the system (primary innervations of each nerve) to reach the conclusion of radiculopathy.

Justification may look at first sight long and complicated, but after some time, it becomes easy to read because it always has the same structure, uses different colors and underlines, and because the user may choose to see the justification just for the concept of interest.

In any KBS, the justification is a fundamental function (Reggia et al., 1985; Darlington, 2011): a hypothetical KBS that almost always gave true answers, but without explaining how it reaches them, would be rather unfit as a diagnostic tool and as a teaching tool.

This is the reason why artificial neural networks (ANN) (Pearce et al., 2015) have not been used. Instead, ANN are very useful for

signal analysis like in electrocardiography or in image analysis like in radiology or microscopy.

Similar considerations could be made about case-based reasoning (CBR) (Campillo-Gimenez et al., 2013). Many tools of the KBS are similar to those in CBR, but a justification of a diagnosis based on a similarity with archived cases is not enough because, in order to reach a diagnosis, it is necessary to verify specific steps of a diagnostic reasoning and this, in our experience, is better done with rules.

RESULTS

Manual Review of Cases on 3,750 Studies Selected with the Classical CIDP Criteria

The 19 classic criteria (see **Figure 4**) applied to the whole archive of cases (26,000 nerve conduction studies) bring to 3,750 cases in which one or more that one of the 19 criteria has been found true. Then, the 3,750 selected cases have been manual reviewed with the results summarized in **Figure 6**.

Here are the main considerations:

1. More than 1,400 have been marked as “not enough for CIDP” because not enough upper and lower extremities nerves were tested. Many of them are cases with history and physical examination highly suggestive of carpal tunnel syndrome (CTS) that, for this reason, have been tested in only one limb, but, for many classical CIDP criteria, they were classified as “definite CIDP” while tested in only one arm. In our opinion, if only one limb has been tested, there is no way to confirm a superimposed CTS on CIDP or to exclude CIDP, so the case is marked as “not enough data” not as “CIDP excluded” and does not take part in statistical studies.
2. Two hundred seventeen are confirmed CIDP and 206 are probable CIDP.

Rule D for CIDP “Van Den Bergh” definite: TRUE (Score 100%)

• At least 2 nerves, motor, any limb, on the left or right, with: Absent FWave and DAmpr red no more 80% (EFNS D); and at least 1 nerve, different from those found in the first selection, with: CV red \geq 30% (EFNS B), DLat del \geq 50% (EFNS A), FWave lat del \geq 20% and DAmpr red less 20% (EFNS C1), Ampr drop \geq 50% and DAmpr red no more than 80% (EFNS E), MWave incr \geq 30% (EFNS F equal to Sap1B), Distal CMAP duration $>$ 12 and DAmpr red no more 20% (modified EFNS G), Amp Drop \geq 50% (Sap1A) (Mand): **True** → Common Peroneal: Absent FWave and DAmpr red no more 80% (EFNS D) (L), Common Peroneal: Absent FWave and DAmpr red no more 80% (EFNS D) (R). And, in different nerves: Ulnar BE: Ampr drop \geq 50% and DAmpr red no more than 80% (EFNS E), Amp Drop \geq 50% (Sap1A) (L)
Conclusion: CIDP VDB definite. D (Fwave absent in 2 nerves with DAmpr red $<$ 80% of normal PLUS a different nerve)

Rule for CIDP Albers-Kelly 1 (CV 25% $<$ normal in 1 or more nerves): TRUE (Score 100%)

• At least 1 nerve, motor, any limb, on the left or right, with: CV red \geq 25% (Albers-Kelly) (Mand): **True** → Common Peroneal: CV red \geq 25% (Albers-Kelly) (L), Tibial: CV red \geq 25% (Albers-Kelly) (L), Ulnar BE: CV red \geq 25% (Albers-Kelly) (L), Common Peroneal: CV red \geq 25% (Albers-Kelly) (R)
Conclusion: CIDP Albers-Kelly 1 (CV 25% $<$ normal in 1 or more nerves)

Rule for C8 radiculopathy primary (Left): TRUE (Score 100%)

• At least 2 muscles, primary innervated by C8, with Denervation, mild to complete (NEC): **True** → Dorsal Interosseus 1st: Denervation, mild to complete (L), Abductor Digiti Quinti: Denervation, mild to complete (L)
Conclusion: C8 radiculopathy

FIGURE 5 | Justification – the system explains to the user how a particular conclusion has been reached. The three rules in the figure have only one condition each. The KBS prints the conditions and after (in blue) the word “True” followed by the indication (underlined) of what in that specific patient study has been found to confirm the condition.

- [-] CIDP, AIDP & review (Internal). Patients 3,436; Studies 3,521
 - [+] Not enough data for definite CIDP (by user). Patients 333; Studies 333
 - [+] Not enough data for any CIDP (by user). Patients 1,069; Studies 1,069
 - [+] CIDP definite by user. Patients 217; Studies 217
 - [+] CIDP Definite or Probable by user. Patients 417; Studies 423
 - [+] CIDP Probable by user. Patients 206; Studies 206
 - [+] CIDP excluded by user. Patients 1,708; Studies 1,710
 - [+] CIDP and/or axonal pathology (by user). Patients 48; Studies 48
 - [+] AIDP (by user). Patients 75; Studies 75
 - [+] CIDP by rules and Motor and sensory CIDP (by user). Patients 7; Studies 7
 - [+] CIDP by rules and Motor and sensory polyneuropathy, axonal (by user)
 - [+] CIDP by rules and Polyneuropathy, type undetermined (by user). Patients 33; Studies 33
 - [+] CIDP by rules and Mononeuropathy multiplex (by user). Patients 43; Studies 43

FIGURE 6 | Summary of manual review of cases. The groups marked as “Not enough data” have nerves tested in one limb only and, without nerves from both sides and from upper and lower limbs, it is impossible to confirm or exclude a CIDP. “By user” indicates “by neurologist’s manual review.”

3. An unexpected high number of studies, 1,458, are false positives (FP) at the review and are marked as CIDP excluded.
4. Forty two are CIDP and/or axonal pathology and are cases in which the degree of disease makes it impossible to evaluate the demyelinating or axonal original nature.
5. Seventy four are AIDP (acute inflammatory demyelinating polyneuropathy), an acute pathology neurophysiologically very similar to CIDP.

Note that the 1,400 cases marked as “not enough nerves tested for CIDP” would never have been evaluated for CIDP in clinical practice because they have been sent to the neurologist for a localized problem in the arms and had a neurological history and a physical examination that exclude CIDP. So the “not enough nerves tested” is not a sign of incomplete or wrong examination but just the correct neurological reasoning. But the classic CIDP rules, applied to only the MNCS without the clinical part, cannot discriminate these cases, so it is necessary to find a method to exclude them from further analysis.

RESULTS

Difference Between “Classical” Studies on CIDP and “HIS–KBS” Based Studies

The very large number of “not enough data” cases (1,402) and of “CIDP excluded” cases (1,710) is a consequence of the way the KBS operates in our system. Rules are not applied on selected cases but on all cases, so it is up to the rules to exclude cases that belong to the above two groups.

Translated to daily clinical experience, the “classical” scenario is the one in which a neurologist seeks an expert advice on cases that are already indicative for CIDP but still present some diagnostic problem. A neurologist will never apply CIDP diagnostic criteria to a patient who arrives with sensory symptoms in one or both hands for carpal tunnel or plexopathy. Our scenario, instead, is similar to that of a student that does not trust his clinical intuition, and so tries to consider all the diagnostic criteria to avoid mistakes.

Statistical Analysis Using Only MNCS

Once the rules are applied to all cases in the archive, the analysis starts with the count of true positives (TP), FP, true negatives (TN), and false negatives (FN) defined as follows:

- TP: cases that are “definite CIDP” according to KBS and “Confirmed definite CIDP” by review.
- FP: cases that are “definite CIDP” according to KBS but “CIDP excluded” by review.
- TN: that are NOT “definite CIDP” according to KBS and “CIDP excluded” by review.
- FN: that are NOT “definite CIDP” according to KBS but “Confirmed definite CIDP” by review.

These values are used to calculate the following statistical descriptors:

$$\begin{aligned}
 \text{Sensitivity} &= \text{TP}/(\text{TP} + \text{FN}) \\
 \text{Specificity} &= \text{TN}/(\text{FP} + \text{TN}) \\
 \text{Precision (positive predictive value)} &= \text{TP}/(\text{TP} + \text{FP}) \\
 \text{Negative predictive value (NPV)} &= \text{TN}/(\text{FN} + \text{TN}) \\
 \text{Accuracy} &= (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) \\
 \text{Matthews correlation coefficient (MCC)} &= \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}
 \end{aligned}$$

Accuracy and MCC are the two indicators that take into account true and FP and negatives. All indicators but MCC are expressed in percentage. MCC goes from -1 to $+1$ and could be considered as a correlation coefficient between the observed and the predicted values with $+1$ as a complete correlation, -1 as a complete inverse correlation, and 0 as no correlation. The formulas suggest that MCC should be less sensible than Accuracy to big variation in the relative size of TP, FP, TN, and FN. The following figures will show that in practice.

Figure 7 shows the results sorted by decreasing MCC. The analysis is based only on electrodiagnostic (EDX). No information from history, exam, laboratory, or biopsy have been used.

#	Description	TP	FP	TN	FN	Sensitivity	Specificity	Accuracy	Precision	NPV	Matthews
1	Nicolas	129	184	1,439	81	61.4	88.7	85.5	41.2	94.7	0.424
2	Hughes	114	143	1,480	96	54.3	91.2	87.0	44.4	93.9	0.417
3	Saperstein	82	82	1,541	128	39.0	94.9	88.5	50.0	92.3	0.379
4	Albers-Kelly	80	78	1,545	130	38.1	95.2	88.7	50.6	92.2	0.378
5	VDB definite	163	385	1,238	47	77.6	76.3	76.4	29.7	96.3	0.375
6	Koski	105	152	1,471	105	50.0	90.6	86.0	40.9	93.3	0.373
7	EFNS definite	164	397	1,226	46	78.1	75.5	75.8	29.2	96.4	0.371
8	INCAT	77	81	1,542	133	36.7	95.0	88.3	48.7	92.1	0.359
9	Barohn	92	123	1,500	118	43.8	92.4	86.9	42.8	92.7	0.359
10	Magda	153	363	1,260	57	72.9	77.6	77.1	29.7	95.7	0.358
11	Bromberg	72	71	1,552	138	34.3	95.6	88.6	50.4	91.8	0.355
12	EFNS 2006 definite	96	189	1,434	114	45.7	88.4	83.5	33.7	92.6	0.299
13	Tackenberg	201	823	800	9	95.7	49.3	54.6	19.6	98.9	0.289
14	AAN	35	21	1,602	175	16.7	98.7	89.3	62.5	90.2	0.285

FIGURE 7 | Results of statistical analysis on 14 definite CIDP criterion based only on EDX data. They appear in order of decreasing MCC value. For each column, the higher (best) value is marked in yellow.

The one that performs best in MCC (Nicolas) has neither the best sensitivity (61.4 versus 95.7% of Tackenberg) nor the best specificity (88.7 versus 98.8% of AAN) but has the best “combination” of the two according to MCC. Note that there is one (AAN) that is the best for specificity, accuracy, and precision, but has only 16.7% of sensibility.

In general, the results were less than expected: if the sensitivity is over 40%, then the FP are always more than the TP. The “best” test according to MCC detects only 133 confirmed CIDP cases, on a total of 217, and has a huge 193 FP. The best for accuracy detects only 35 cases on 210. Clearly the classical CIDP criteria, applied to any case without the selection due to clinical information, have no practical use.

This will be corrected by the use of clinical information (Hughes et al., 2001, 2006; Tackenberg et al., 2007; Koski et al., 2009; Joint Task Force of the EFNS and the PNS, 2010; Van den Bergh et al., 2010), but it demonstrates how a diagnostic flow chart could be misleading if applied in a “blind,” “mechanical” way: very often the diagnostic criteria imply some clinical information or, in general, some “clinical sense” that is not explicit in the strict definition of diagnostic criteria.

Statistical Analysis Including Clinical Data

One reason for the disappointing results in Figure 7 is the application of CIDP criteria on all NC studies in our archive, even without clinical signs of CIDP, but with signs of other neurological diseases. The EFNS reports a list of clinical inclusion/exclusion criteria. See Joint Task Force of the EFNS and the PNS (2010), p. 3.

Figure S2 in Supplementary Material shows the effects of including in the rules the clinical information that was available in the medical records. The number of TP is almost the same while the reduction in FP is dramatic (more than 50%). As a result, indicators like accuracy, precision, and MCC are clearly better than before. Nevertheless, the results are far from optimal: the best two criteria for accuracy or MCC have a sensibility lower than 60%, still not enough for a practical use.

At this point, the methodology and the results are similar to the Bromberg paper (Bromberg, 2011). The main difference is the use

of a KBS to apply that methodology in an automatic way on a large number of cases instead of manual calculation on a selected group of cases.

Writing of Other CIDP Criteria

The next logical step was to develop other CIDP criteria based on personal experience and results of CIDP case reviews. Note that this is the way our KBS has been built when clear diagnostic criteria were not available in the literature or when the diagnostic criteria from the literature were not performing well.

Here is a schematic view of the method used to build the KBS:

1. Write concepts and rules according to literature and/or clinical experience.
2. Apply to all cases and analyze them in terms of FP or FN.
3. Modify concepts and rules according to point 2.
4. Repeat the cycle until the results are satisfactory.

Usually, the results were satisfying after a few cycles, but not in the case of the CIDP criteria. Several “generations” of our CIDP criteria were written but, surprisingly, they never performed better than the “classic ones”: typically, when we tried to use more selective criteria, we obtained more specificity but lost sensitivity and vice versa. The persistent disappointing results of many “new CIDP criteria” led to the development of a complete different approach.

Development of New Tools

There are two main assets in the HIS-KBS that can be used:

1. The high number of concepts (and rules) and specially the very high number of “intermediate” concepts.
2. The high number of patient studies.

The first tool (computerized procedure inside the KBS) was developed to better explore the difference between the TP and the FP and is called “*differential association tool*” (Figure 8).

1. It starts with a couple of conclusions, in this case, true positives CIDP (TP) and false positives CIDP (FP).

Item	Associated conclusion	Definite CIDP and definite by review (TP) (205)		Definite CIDP excluded by review (FP) (947)		Diff. %
		Cases	%	Cases	%	
212	Lower thoracic radiculopathy	0	0.00	3	0.31	100.00
213	Major Wallerian and mixed FW C5 radiculopathy	0	0.00	3	0.31	100.00
214	Ulnar BE CV worsening to absent	0	0.00	3	0.31	100.00
215	Ulnar AE CV worsening to moderate	0	0.00	3	0.31	100.00
216	Median II CV recovering to normal	0	0.00	3	0.31	100.00
217	Median II DLat recovering to moderate	0	0.00	3	0.31	100.00
218	Common Peroneal CV recovering to normal	0	0.00	3	0.31	100.00
219	Dorsal-Ulnar nerve, non conductible	0	0.00	3	0.31	100.00
220	Secondary demyelinating polyneuropathy	1	0.48	306	32.31	98.51
221	Inflammatory Demyelinating Polyneuropathies	1	0.48	306	32.31	98.51
222	Polyneuropathy in diabetes (internal, by user)	1	0.48	251	26.50	98.19
223	Demyelinating polyneuropathy secondary to diabetes	1	0.48	180	19.00	97.47
224	Not enough for definite CIDP, upper and lower limbs must be tested	1	0.48	129	13.62	96.48
225	Motor conduction done Upper Limb only	1	0.48	72	7.60	93.68
226	Delayed H Reflex bilateral	1	0.48	67	7.07	93.21
227	Not enough for definite CIDP (by rules)	5	2.43	286	30.20	91.95
228	Not enough for definite CIDP (by rules and not by review)	5	2.43	286	30.20	91.95
229	CIDP with normal motor CV or borderline slowing in lower limbs	2	0.97	112	11.82	91.79
▶ 230	CIDP with CV reduction in 1 limb only	2	0.97	111	11.72	91.72
231	At least 2 Normal/Borderline MCV in the lower limbs and none with higher slowing	2	0.97	93	9.82	90.12
232	Normal common peroneal MCV (bilateral if done)	4	1.95	187	19.74	90.12
233	At least 2 Normal MCV and 1 SCV in the upper limbs	5	2.43	220	23.23	89.54
234	CIDP with normal motor CV or borderline slowing in upper limbs	6	2.92	263	27.77	89.49
235	Normal H Reflex	1	0.48	42	4.43	89.16
236	At least 2 Normal MCV in the lower limbs and none with slowing	2	0.97	84	8.87	89.06
237	Normal H Reflex Soleous	1	0.48	41	4.32	88.89
238	Conduction block neuropathy	1	0.48	41	4.32	88.89
239	At least 2 Normal MCV in the lower limbs	4	1.95	166	17.52	88.87
240	Delayed H Reflex	2	0.97	81	8.55	88.65
241	Delayed H Reflex Soleous	2	0.97	81	8.55	88.65

FIGURE 8 | The “differential association tool.” Given a couple of conclusions to test (the two in brown on the top), the procedure lists all the conclusions that are “associated” with at least one of the two, then, for each “associated conclusion” it counts the number of cases associated with each of the two reference conclusions. The figure shows a small part of the list (there are hundreds of associated conclusions) ordered by the percentage of association with reference conclusion two.

- Then, the tool selects all the studies that have one of the two conclusions.
- Finally, it builds a list of all the other conclusions that are associated to the studies of each group.

Any possible conclusion may, therefore, belong to three groups:

- Associated to conclusion 1 (TP) with various percentage, never to conclusion 2.
- Associated to conclusion 2 (FP) with various percentage, never to conclusion 1.
- Associated to both in various percentages.

Obviously, in a single study, any conclusion may be associated only to TP or to FP because the two are mutually exclusive. For

example, conclusion 230 in **Figure 8** (“CIDP with CV reduction in 1 limb only”) has been found in two studies TP (out of 205 TP) and in 111 studies FP (out of 947 FP).

The associated conclusions are then sorted by decreasing “affinity” with FP and **Figure 8** shows the region of the list in which there are the last ones of group 2 (associated only with TP) and the first ones of group 3 (associated to both TP and FP in various percentage). The possible use of this tool is intuitive:

- If we exclude all the studies that have one or more of the conclusions that are associated only with FP, then the number of FP will be reduced without reducing the TP.
- If we exclude all the studies that have conclusions associated “much more” with FP than with TP, then the number of FP will be reduced “much more” than the TP.

For example, if we exclude all the cases with conclusion 230 (“CIDP with CV reduction in 1 limb only” enhanced in blue in **Figure 8**), then the number of FP will be reduced by 11 while the number of TP only by 2.

Figure 8 is interesting for other reasons too:

1. Conclusions 230, 233, 234, 236, and 239 are typical examples of *internal conclusions*. Internal conclusions could be very simple descriptions, like the five listed above, or also concepts that have a specific meaning in physiopathology, such as the ones in Figure S1 in Supplementary Material. The interesting point is that an internal conclusion like “*At least 2 normal MCV and 1 SCV in the upper limbs*,” that would be of no special interest by itself, suddenly becomes important because the *differential association tool* has shown that it can be used for strongly reducing the FP while maintaining the TP, thus creating better diagnostic criteria.
2. Other conclusions, like 235 (*Normal H Reflex*), are results of other neurological tests that appear in a different light because of the *differential association tool*: if the normal H reflex (but could be any other test result) is associated to 42 FP and only to one TP, not only it is useful to write a better diagnostic rule (like in the above point 1), but there must be a physiopathological reason that deserves to be investigated. In fact, the role of H Reflex in CIDP will be the object of a specific section of this paper.
3. Conclusions from 214 to 218 are not particularly interesting for a diagnostic tool because they are verified in only three studies, but are conceptually interesting because they are conclusions that come from the analysis of sequential studies of the same patient. This is a typical result of a powerful KBS that is able to analyze not only the hundreds of data of a single study, but all the data of all the studies of a patient. This multiplies the complexity of the analysis and prepares the conditions for the “surprises” that the *differential association tool* can produce.

Before using the results of the *differential association tool* for writing a rule that lowers the FP without losing TP, another step is necessary.

If we look to conclusions 235 and 237, it is evident that “Normal H Reflex” includes all the “Normal H reflex Soleus” (the remaining two are normal H reflex FCR). So, it is redundant to exclude 237. It is enough to exclude 235. This simple example can be generalized in the concept of “incremental contribution,” i.e., the contribution of exclusion $N + 1$ to all the N exclusions before. To do this, the “*Incremental contribution tool*” has been created: the target is to reduce FP with no or little reduction of TP. The incremental number of TP is in column C1 R, the incremental number of FP is in column C2 R (Figure S3 in Supplementary Material).

It is interesting to see how the incremental tool changes the perspective: for example, conclusion 230 decreases the FP of only four but reduces the TP of two. This does not change anything about the meaning of considering “*CV reduction in one limb only*” as something that almost exclude CIDP but just says that 107 cases with “*CV reduction in one limb only*” have been already excluded so the incremental contribution is small.

Some exclusions not only become useless because they do not change FP but also appear detrimental. For example, the exclusion

of conclusion 240 leaves FP unchanged at 52 but eliminates one TP (from 188 to 187). Again, this does not change the analysis of the single association but just tells that with those 239 exclusions before, the addition of 240 is useless and detrimental.

At this point, the *Incremental contribution tool* will eliminate all the useless conclusions (those marked with a red square in Figure S3 in Supplementary Material) and is almost ready to create a new rule. It only needs the setting of the cut-off point, i.e., how many rows to include.

The main choices are two:

1. Create a “*Lossless*” rule with only the exclusions that do not reduce the TP (up to row 219 in Figure S4 in Supplementary Material).
2. Create a “*Lossy*” rule that uses also the exclusions that reduce TP but with a much stronger reduction in FP.

The lossless rule can maintain the 205 TP while reducing the FP from 947 to 230. A lossy rule up to row 229 will give 196 TP (losing 9 of them) and only 79 FP.

Figure S4 in Supplementary Material shows the lossless rule. There are 88 exclusions organized automatically by the KBS in 39 conditions by associating the ones that belong to a similar “logical region” of the concepts tree. For example, conclusions belonging to the motor exam occupy one condition, those regarding denervation in EMG another condition and so on. This is useful for the reader, while for the KBS, the order of exclusions changes nothing. The rule was called “specificity filter.”

Statistical Analysis of the New Tool “Specificity Filter”

Figure 9 shows the results of adding the “specificity filter” to the classic CIDP rules. The results should be compared to those with the exclusions suggested by clinical CIDP criteria. The MCC scores are generally slightly better (from 0.604 to 0.686) but the nature of the data is very different. In general, the TP and FP are both higher, with a slight better “combination” of the two according to MCC.

With a more accurate analysis, Figure S2 in Supplementary Material and **Figure 9** (upper) appear very different, much more than the MCC value may suggest. For example, Tackenberg has 199 TP/389 FP in Figure S2 in Supplementary Material but 190 TP/134 FP in **Figure 9** upper: TP are almost the same and FP are reduced by 50%. The best in Figure S2 in Supplementary Material, Nicolas, had 127/52 and in **Figure 9** upper goes to 122/25, again a small reduction in TP but a 50% reduction of FP.

The next step was to “personalize” the filter for each one of the criteria. In fact, the TP and FP were taken from all the manual review (**Figure 6**), so, in theory, it is possible to do the same process by using TP and FP of each one criterion as in Figure S2 in Supplementary Material. Clearly, a specificity rule based on the 133TP/193FP of Nicolas should be different from the rule for Tackenberg that has 207TP/870FP.

These rules were called “*Euristic rules*” because we are not using a clinical theory to build the rules, but only the result of automatic tools (for this reason, they may be called heuristic rules, but we used instead “Euristic” that is the name of our HIS–KBS program). The results are in **Figure 9** (lower).

#	Description	TP	FP	TN	FN	Sensitivity	Specificity	Accuracy	Precision	NPV	Matthews
1	Tackenberg	190	134	1,489	20	90.5	91.7	91.6	58.6	98.7	0.686
2	VDB definite	154	63	1,560	56	73.3	96.1	93.5	71.0	96.5	0.685
3	EFNS definite	155	65	1,558	55	73.8	96.0	93.5	70.5	96.6	0.684
4	Magda	145	57	1,566	65	69.0	96.5	93.3	71.8	96.0	0.667
5	Nicolas	122	25	1,598	88	58.1	98.5	93.8	83.0	94.8	0.663
6	Hughes	108	18	1,605	102	51.4	98.9	93.5	85.7	94.0	0.633
7	Koski	100	11	1,612	110	47.6	99.3	93.4	90.1	93.6	0.627
8	Barohn	90	7	1,616	120	42.9	99.6	93.1	92.8	93.1	0.604
9	Albers-Kelly	77	2	1,621	133	36.7	99.9	92.6	97.5	92.4	0.573
10	Saperstein	78	4	1,619	132	37.1	99.8	92.6	95.1	92.5	0.568
11	Bromberg	69	1	1,622	141	32.9	99.9	92.3	98.6	92.0	0.545
12	INCAT	74	7	1,616	136	35.2	99.6	92.2	91.4	92.2	0.539
13	EFNS 2006 definite	88	40	1,583	122	41.9	97.5	91.2	68.8	92.8	0.493
14	AAN	34	0	1,623	176	16.2	100.0	90.4	100.0	90.2	0.382

#	Description	TP	FP	TN	FN	Sensitivity	Specificity	Accuracy	Precision	NPV	Matthews
1	CIDP Euristic (Tackenberg)	159	63	1,560	51	75.7	96.1	93.8	71.6	96.8	0.701
2	CIDP Euristic (EFNS)	146	44	1,579	64	69.5	97.3	94.1	76.8	96.1	0.698
3	CIDP Euristic (VDB)	151	53	1,570	59	71.9	96.7	93.9	74.0	96.4	0.695
4	CIDP Euristic (Magda)	145	48	1,575	65	69.0	97.0	93.8	75.1	96.0	0.686
5	CIDP Euristic (Nicolas)	127	61	1,562	83	60.5	96.2	92.1	67.5	95.0	0.595
6	CIDP Euristic (EFNS 2006)	91	61	1,562	119	43.3	96.2	90.2	59.9	92.9	0.457

FIGURE 9 | Upper: the 14 definite CIDP criteria with the “specificity filter.” The best Matthews is 0.686 while the best accuracy is 93.8%. **Lower:** the “Euristic” rules. Each rule has a specific set of exclusions generated by the statistical analysis. Only the first three are better than those obtained with specificity filter (**Upper**). Nicolas and EFNS 2006 were better with specificity Filter.

The overall results are very close to those of **Figure 9** (upper). In general, there is a small growth in TP with small or no growth in FP. It was necessary, again, to move to a complete new direction, and the new approach was the CIDP components.

The CIDP Components – A Different Approach for New Tools

The 102 *internal conclusions*, that have been written to reproduce in the KBS the structure of the CIDP criteria, appear in **Figure S1** in Supplementary Material. Applying on the components, the same tools used for the main CIDP criteria were a further logical step ahead. **Figure 10** shows the results. The list of 51 components has been obtained from those in **Figure S1** in Supplementary Material with elimination of duplicates. For example, VDB and EFNS share the same components but component “G” is only used for EFNS.

Each component has been tested with the “specificity filter” as shown in **Figure S4** in Supplementary Material. This time the results were really surprising: a single component, “Albers-Kelly 1” performed alone, in terms of accuracy, precision, and MCC, better than any classical complex CIDP criteria!

The general implications of this result will be discussed at the end of the paper. The most evident at this point are

1. The potential of some of the internal conclusions is crucial.
2. Complex criteria, difficult to remember and apply in consistent way, may be substituted by smaller criteria (the components) with better performance and much easier application!
3. The possibility to test so many components in a KBS–HIS is fundamental: there is no other way to reach these results and create this kind of rules.

Improvement of the Rules Based on “CIDP Components” by Using the “Incremental Contribution Tool”

Figure S5 in Supplementary Material shows the results of the “*incremental contribution tool*.” The difference with the former one (the differential association tool of **Figure 8**) is that, in this case, it works by adding inclusions, not exclusions. The 10 columns marked as “inc” show the cumulative result, while the other values refer to the single component.

For example, the 10 “inc” indicators in row 3 show what happens if we have a rule that selects all the cases with component 1, plus all the cases with component 2, plus all the cases with component 3. In a short form, this could be written as: “select all cases with conclusion 1, or conclusion 2, or conclusion 3.”

Further Improvement of the Rules Based on “CIDP Components” with the “Incremental Contribution Tool” and with the “Second Pass”

The results in **Figure S5** in Supplementary Material show that adding components from a list sorted by MCC does not bring better results than using the first one alone. As with the incremental tool of **Figure S5** in Supplementary Material, there are “useless” rows. For example, row 2 does not change the results of row 1. The reason is that all the 134 TP of Bromberg 1 were already among the 159 TP of Albers-Kelly 1, and the same for the 27 FP. If the component list is sorted in different way (by specificity, accuracy, or precision), the result is not the same: the best MCC value is reached using a few lines instead of only one.

#	Description	TP	FP	TN	FN	Sens	Spec	Accu	Prec	NPV	Matt
1	Albers-Kelly 1	159	45	1,578	51	75.7	97.2	94.8	77.9	96.9	0.739
2	Bromberg 1	134	27	1,596	76	63.8	98.3	94.4	83.2	95.5	0.699
3	Tackenberg with 2 CV	177	106	1,517	33	84.3	93.5	92.4	62.5	97.9	0.685
4	Magda 1+3 (w Cat C)	145	57	1,566	65	69.0	96.5	93.3	71.8	96.0	0.667
5	Saperstein with 2 (2A or 2B)	112	16	1,607	98	53.3	99.0	93.8	87.5	94.3	0.654
6	Tackenberg with 3	151	75	1,548	59	71.9	95.4	92.7	66.8	96.3	0.652
7	Nicolas D	108	13	1,610	102	51.4	99.2	93.7	89.3	94.0	0.649
8	Hughes w CV, DLat or FWave	102	12	1,611	108	48.6	99.3	93.5	89.5	93.7	0.631
9	Magda 1+3	126	43	1,580	84	60.0	97.4	93.1	74.6	95.0	0.631
10	EFNS/VDB B	90	7	1,616	120	42.9	99.6	93.1	92.8	93.1	0.604
11	Tackenberg with 2 DLat	98	23	1,600	112	46.7	98.6	92.6	81.0	93.5	0.580
12	Saperstein with 2 (2A)	84	10	1,613	126	40.0	99.4	92.6	89.4	92.8	0.569
13	Albers-Kelly Bromberg 3	128	83	1,540	82	61.0	94.9	91.0	60.7	94.9	0.557
14	Koski with 5 or more nerves	78	8	1,615	132	37.1	99.5	92.4	90.7	92.4	0.552
15	Albers-Kelly Bromberg 2	78	13	1,610	132	37.1	99.2	92.1	85.7	92.4	0.533
16	INCAT with CB	69	7	1,616	141	32.9	99.6	91.9	90.8	92.0	0.518
17	Saperstein with 2 (3A or 3B)	71	10	1,613	139	33.8	99.4	91.9	87.7	92.1	0.514
18	Nicolas C	60	9	1,614	150	28.6	99.4	91.3	87.0	91.5	0.469
19	EFNS/VDB E2	75	30	1,593	135	35.7	98.2	91.0	71.4	92.2	0.464
20	EFNS/VDB A	55	5	1,618	155	26.2	99.7	91.3	91.7	91.3	0.463
21	Albers-Kelly Bromberg 4	53	6	1,617	157	25.2	99.6	91.1	89.8	91.1	0.449
22	AAN with 2 (1A or 1B)	65	28	1,595	145	31.0	98.3	90.6	69.9	91.7	0.424
23	Tackenberg with 2 CB	64	28	1,595	146	30.5	98.3	90.5	69.6	91.6	0.419
24	Saperstein with 2 (3A)	49	9	1,614	161	23.3	99.4	90.7	84.5	90.9	0.414
25	Nicolas B	41	7	1,616	169	19.5	99.6	90.4	85.4	90.5	0.381
26	Hughes w partial CB or abn TD	41	7	1,616	169	19.5	99.6	90.4	85.4	90.5	0.381
27	EFNS 2006 C1 or C2	34	3	1,620	176	16.2	99.8	90.2	91.9	90.2	0.362
28	Saperstein with 2 (3B)	32	2	1,621	178	15.2	99.9	90.2	94.1	90.1	0.357
29	Saperstein with 2 (2B)	35	6	1,617	175	16.7	99.6	90.1	85.4	90.2	0.351
30	Nicolas A	33	4	1,619	177	15.7	99.8	90.1	89.2	90.1	0.350
31	EFNS/VDB C1 15%	35	7	1,616	175	16.7	99.6	90.1	83.3	90.2	0.346
32	EFNS/VDB C1 25%	35	7	1,616	175	16.7	99.6	90.1	83.3	90.2	0.346
33	Tackenberg with 2 Fwave	27	0	1,623	183	12.9	100.0	90.0	100.0	89.9	0.340
34	EFNS/VDB C1	30	3	1,620	180	14.3	99.8	90.0	90.9	90.0	0.338
35	Saperstein with 2 (1A or 1B)	46	30	1,593	164	21.9	98.2	89.4	60.5	90.7	0.320
36	Saperstein with 2 (4A or 4B)	24	0	1,623	186	11.4	100.0	89.9	100.0	89.7	0.320
37	Saperstein with 2 (1A)	45	30	1,593	165	21.4	98.2	89.4	60.0	90.6	0.315
38	EFNS/VDB E1	42	28	1,595	168	20.0	98.3	89.3	60.0	90.5	0.304
39	Saperstein with 2 (4A)	21	0	1,623	189	10.0	100.0	89.7	100.0	89.6	0.299
40	EFNS/VDB D	23	4	1,619	187	11.0	99.8	89.6	85.2	89.6	0.283
41	Koski with 3-4 nerves	22	3	1,620	188	10.5	99.8	89.6	88.0	89.6	0.283
42	EFNS/VDB G	19	9	1,614	191	9.0	99.4	89.1	67.9	89.4	0.221
43	INCAT without CB	9	0	1,623	201	4.3	100.0	89.0	100.0	89.0	0.195
44	AAN 1B	4	0	1,623	206	1.9	100.0	88.8	100.0	88.7	0.130
45	EFNS/VDB C2	3	0	1,623	207	1.4	100.0	88.7	100.0	88.7	0.113
46	Saperstein with 2 (4B)	3	0	1,623	207	1.4	100.0	88.7	100.0	88.7	0.113
47	AAN 1comb	3	0	1,623	207	1.4	100.0	88.7	100.0	88.7	0.113
48	Saperstein with 2 (1B)	1	0	1,623	209	0.5	100.0	88.6	100.0	88.6	0.065
49	Saperstein with 2 (1comb)	1	0	1,623	209	0.5	100.0	88.6	100.0	88.6	0.065
50	Magda w delayed H	2	4	1,619	208	1.0	99.8	88.4	33.3	88.6	0.039
51	Magda w delayed H (w Cat C)	2	4	1,619	208	1.0	99.8	88.4	33.3	88.6	0.039

FIGURE 10 | Statistics on 51 CIDP components with specificity filter. They appear in order of decreasing MCC. Note that the first one gives a better MCC result (0.739) than any of the complex CIDP criteria, including the one that uses that same component!

The results with this new tool, called “*incremental contribution tool with second pass*,” appear in Figure S6 in Supplementary Material. The idea is simple: exclude the components that, when added, appear useless or detrimental. The procedure does three steps: the first step tests each row, the second clears the useless rows, and the third tests only the ones remaining.

The results were surprising: if components of rows 2, 3, and 4 were not used, the addition of component 5 to component 1 gave a better result than the component 1 alone. And, as evident in Figure S6 in Supplementary Material, the maximum was reached at row 40. The new tool was able to give results that are not possible to suspect by just looking to Figure S5 in Supplementary Material (that is equal but for the second pass). It was the possibility of testing thousands of studies and millions of conclusions over and over again to make this possible.

In conclusion, a new criterion, assembled from only seven components selected from 51 belonging to different original criteria, performs better than any other tested criterion.

Figure S7 in Supplementary Material will show the results of the *incremental contribution tool with second pass* applied to the components sorted by different statistical indicators. For easier reading, the raw results are not displayed.

Figure S7 in Supplementary Material (upper) shows the results on components sorted by Specificity. The maximum is at 0.755 MCC and 16 components are used.

Figure S7 in Supplementary Material (middle) shows the results on components sorted by Accuracy. The maximum is at 0.759 MCC using seven components.

Figure S7 in Supplementary Material (lower) shows the results on components sorted by Sensibility. The maximum is at 0.752 MCC using 14 components.

Finally, **Figure 11** shows the statistical analysis on the rules built according to the results of Figure S7 in Supplementary Material. The criteria are called “*Euristic component rules*” because, as written before, the rules were not created by a neurologist or according to a theory, but using the KBS. As said before, the rules are made

by the components as automatically selected by the tools in Figure S7 in Supplementary Material plus the specificity filter as in Figure S4 in Supplementary Material.

Further Improvement by Adding a Tailored “Specificity Filter” to the “Component” Rules

As a last step, a “specificity filter” was tailored for each one of the rules in **Figure 11**. First, the four rules were tested without the specificity filter. Then, the same procedure as in **Figures 8 and 9** was applied to each one and 4 “specificity filters” were created. **Figure 12** shows the results of the optimized component rules matched to the best results of the classic CIDP criteria.

This concludes the section of this paper about the “discovery” of new CIDP diagnostic criteria by using new tools of the KBS and about their results. The next sections deal with different utilizations of the new tools.

Reverse Engineering

Reverse engineering is the process of discovering the theory that stays behind a certain object or process by studying how the object is made and how it works.

In this case, we have rules generated by computer and not by clinical theories or personal experience. The most important thing is that these rules are not complex arrays of numbers as with neural networks but plain text that is easy to read for any physician, so should be possible to understand why these rules work so well as shown in **Figure 12**.

The first target for reverse engineering is the specificity filter. In **Figure 13**, there are a few of the exclusions that stay in the “Euristic specificity filter.” The reason for some of them is easy to understand: for example “motor radiculopathy” excludes a common source of FP. Instead, two of them have been selected for further analysis because they do not appear directly correlated to the typical causes of FP.

#	Description	TP	FP	TN	FN	Sensitivity	Specificity	Accuracy	Precision	NPV	Matthews
1	CIDP Euristic (Matthews)	179	59	1,564	31	85.2	96.4	95.1	75.2	98.1	0.773
2	CIDP Euristic (Accuracy)	179	59	1,564	31	85.2	96.4	95.1	75.2	98.1	0.773
3	CIDP Euristic (Specificity)	173	52	1,571	37	82.4	96.8	95.1	76.9	97.7	0.768
4	CIDP Euristic (Precision)	178	61	1,562	32	84.8	96.2	94.9	74.5	98.0	0.766

FIGURE 11 | The results with the four Euristic “components” rules.

#	Description	TP	FP	TN	FN	Sensitivity	Specificity	Accuracy	Precision	NPV	Matthews
1	CIDP Euristic (Accuracy) optimized	162	9	1,614	48	77.1	99.4	96.9	94.7	97.1	0.839
2	CIDP Euristic (Matthews) optimized	154	5	1,618	56	73.3	99.7	96.7	96.9	96.7	0.826
3	CIDP Euristic (Specificity) optimized	150	7	1,616	60	71.4	99.6	96.3	95.5	96.4	0.808
4	CIDP Euristic (Precision) optimized	144	3	1,620	66	68.6	99.8	96.2	98.0	96.1	0.802

#	Description	TP	FP	TN	FN	Sensitivity	Specificity	Accuracy	Precision	NPV	Matthews
1	Nicolas	129	184	1,439	81	61.4	88.7	85.5	41.2	94.7	0.424
2	Hughes	114	143	1,480	96	54.3	91.2	87.0	44.4	93.9	0.417

FIGURE 12 | The Euristic component rules with a “specificity filter” tailored for each one. Below are the two best classical CIDP criteria with clinical information. The difference is evident.

☐ Rule for Euristic CIDP specificity filter, Oct 1, 2011.

- ... Confirm Definite CIDP
- ... No evidence of Generalized myopathy by rules but not in dismissal diagnoses
- ... No evidence of 2 or more motor nerves with DLat but no PLat and no CV
- ... No evidence of Upper limb bilat periph sensory deficit (light touch)
- ... No evidence of Suspected motor L5 Radiculopathy on the left or right
- ... No evidence of Mild Upper limb proximal and distal Asymmetric Denervation or Moderate Upper limb distal Asym
- ... No evidence of Multiple moderate denervation, upper and lower limbs
- ... No evidence of Normal EMG of the cervical paraspinous muscles
- ... No evidence of Ulnar motor nerve below elbow, non conductible or Bord. ulnar nerve slowing at the elbow or Si
- ... **No evidence of Dorsal Sural n. lesion, non conductible** or Common Peroneal conduction block with unaffected C
- ... No evidence of Multiple focal mononeuropathy sensory or Multiple focal mononeuropathy mot & sens in legs
- ... No evidence of Mild gener sensory slowing in the legs or Sev. gener sensory slowing in the legs or Generalized
- ... No evidence of Lower limb distal nerve motor Amplitude asymmetry worsening to severe
- ... **No evidence of Normal H Reflex bilateral** or Normal H Reflex, no focal slowings of lower limbs
- ... No evidence of Ulnar BE CV worsening to absent or Median II CV recovering to normal or Median DLat worsenir
- ... No evidence of At least 2 Normal MCV in the lower limbs or At least 2 Normal MCV and 1 SCV in the upper limbs
- ... No evidence of Normal common peroneal MCV (bilateral if done)

FIGURE 13 | The first conditions of the “Euristic specificity filter.” Some lines have been cut on the right for space reason. The two conclusions circled in red could be evaluated in “reverse engineering” because they do not look at typical causes of false positive CIDP.

#	Description	Tot	Absent H	%	Normal H	%	Delayed H	%
1	CIDP, H reflex done and CIDP confirmed	72	69	95.83	1	1.38	2	2.77
2	CIDP, H reflex done and CIDP probable	90	79	87.77	2	2.22	9	10.00
3	CIDP, H reflex done and CIDP excluded	341	218	63.92	42	12.31	81	23.75

#	Description	Tot	Absent H	%	Normal H	%	Delayed H	%
1	CIDP clinical filter, H reflex & CIDP confirmed	72	69	95.83	1	1.38	2	2.77
2	CIDP clinical filter, H reflex & CIDP probable	83	72	86.74	2	2.40	9	10.84
3	CIDP clinical filter, H reflex & CIDP excluded	210	112	53.33	32	15.23	66	31.42

FIGURE 14 | Upper: the distribution of H reflex results in cases with CIDP confirmed, probable, or excluded by review. The absent H is very high in confirmed CIDP (95.83%) but also high in excluded CIDP (63.92%). **Lower:** the same as above but with the “clinical filter” on. Now the number of cases with false positive CIDP is lower, and this is evident in the group with “CIDP excluded.”

Only the second, “Normal H Reflex bilateral,” will be analyzed in this paper.

Reverse Engineering Results – The Role of H Reflex

The “Match tool” and the “specificity rule” show the role of Normal H reflex as an exclusion. **Figure 14** shows the distribution of H reflex results (that may be normal, delayed, or absent) in confirmed, probable, and excluded CIDP. **Figure 14** (upper) uses the EDX as the only criterion; **Figure 14** (lower) includes the clinical information.

The association of Absent H reflex with confirmed CIDP is very strong (over 95%) and this seems to confirm the indication from literature (Lhotska et al., 2001; Amato and Dumitru, 2002; Hadden et al., 2006). But Absent H was also widely represented in excluded CIDP. So, the value of Absent H as an indicator of CIDP is very limited. The neurological explanation is that there are other forms of demyelinating peripheral neuropathies different from

CIDP, for example AIDP, diabetes, hereditary neuropathies, etc. that result in absent H reflex. So, the Absent H reflex is a highly accurate indicator of any form of demyelinating neuropathy but is not specific for CIDP.

The “specificity rule” is not using the “Absent H” result but rather is using the “Normal H” as one of the exclusions. This is the key point and the base for the practical clinical suggestion that comes from the reverse engineering of the “specificity Rule.”

Figure 15 shows the statistical results in a form useful to answer to the question: “What do we expect if the CIDP criteria have been confirmed and the H Reflex test has been done with its three possible results?” Now it is evident that if the result is Normal or Delayed, then CIDP is almost always excluded. It is also evident that Absent H is highly associated with confirmed CIDP but not enough to be used as a single diagnostic test for the reasons shown in **Figure 14**.

This reverse engineering result from a rule created by the KBS could also have an impact on clinical routine for ruling out FP CIDP.

#	Description	Tot	Excluded	%	Confirmed	%	Probable	%
1	CIDP clinical filter and Absent H	253	69	27.27	112	44.26	72	28.45
2	CIDP clinical filter and Normal H	35	1	2.85	32	91.42	2	5.71
3	CIDP clinical filter and Delayed H	77	2	2.59	66	85.71	9	11.68

FIGURE 15 | The distribution seen in an “operative view” as an answer to the question: “what to expect if the CIDP rules apply and H Reflex has been tested?” The three lines show what happens if the H Reflex is “Absent,” “Normal,” or “Delayed.” The answer is clear: if the result of H Reflex test is “Normal” or “Delayed,” then CIDP is almost always excluded. If the result is “Absent,” then the test is not discriminative enough.

DISCUSSION

Advantages

We have seen that complex diagnostic criteria can be “transferred” to a KBS. This brings many advantages:

1. The KT process requires a detailed examination of the criteria in order to find any missing or “implicit” knowledge that is a problem for a computer while a human expert will easily compensate. This process is useful by itself and brings to a better KT not only to the KBS but between individuals like students and physicians.
2. The KBS explains and justifies its results, and this is essential for a medical tool and of great value in a teaching environment and in clinical practice.
3. Complex diagnostic criteria are applied in fractions of a second and without errors, while for a clinician, the process is time consuming and error prone.
4. Physicians tend to integrate complex criteria with clinical sense and personal experience. So, it is difficult to validate and compare different criteria without using a computerized system that, being a “tabula rasa,” needs a complete and unambiguous explanation in order to work.

Knowledge transfer may bring unexpected results:

1. Complex diagnostic criteria may give surprisingly poor results in terms of specificity and sensitivity if tested with an HIS–KBS on a large number of patient studies.
2. With an HIS–KBS, it is possible to create better diagnostic criteria using special processing tools that take advantage of the large number of cases analyzed.
3. The new criteria discovered with the KBS can be smaller and easier to apply than the original ones, while still maintaining the same or better performance.
4. Some of the discovered criteria can be “reverse engineered” to understand why they are made in that way, why they perform well, and if there is a pathophysiological explanation.
5. The described tools can be easily applied to any HIS–KBS to create better criteria and reproduce expert behavior.

There are a few more methodological considerations about the points above. All the procedure described in this paper are essentially a way to explain the performance of an expert (his manual review of cases – see **Figure 6**) by criteria/rules that are generated automatically by the KBS and not dictated by the same expert. This is necessary because an expert is not always able to explain his expertise in a working form. In some cases, as we saw in

this paper, the computer-generated criteria can be more efficient in reproducing the neurologist’s expertise than the criteria that the same neurologist writes in a direct way. Reverse engineering of the computer-generated criteria may reveal mechanisms that could be undiscovered by the same expert.

Another key point is the way the target is reached. As said before, if the computer system perfectly matched the behavior of the expert but is unable to explain how it reaches diagnosis and conclusion, then the utility of the system would be much lower. On the contrary, the method described generates rules that are easy to read, and in some cases, easier than the classic ones because they are shorter while performing better. In addition, because they are rules and not a lot of numerical weights as in ANN, they can justify in natural language how they work.

It is evident that the results strongly depend on the availability of a large HIS integrated with a large KBS. We stress the dimensional aspect because it represents a quantum leap. In fact:

1. A small KBS, developed for the specific task of CIDP diagnosis, will not have that large number of intermediate concepts belonging to many different areas that have been the key for the “component tool.”
2. A small archive of patient studies instead will strongly limit the most part of the tools used for discovery that requires statistical inference from large numbers.
3. Also, a large KBS and a large archive of patient studies could not be enough if the patient studies are limited to just a section of the patient folder, for example, only Nerve Conduction. It would miss a lot of interesting discoveries that come from unexpected relations between any clinical data.

Limits

1. The first limit stays in the source of the diagnostic expertise itself: if the expert makes mistakes, then the method will generate rules that lead to a wrong result. In fact, the method described in this paper does not include the validation of the new discovered criteria. To do this, the new rules must be tested on new cases that have been reviewed by other neurologists. But the aim of this paper is different: the core is the development of a “computer-aided way to generate rules” that reproduce the expert diagnostic behavior when the expert is not able to do it directly (as happened to us) or when the possibilities and the combinations to explore are too complex (for example, the 19 diagnostic criteria with 51 components). The aim has been to reproduce the expertise as it was during the manual review of the 3,750 cases.

2. The second limit comes straight from the “indications” of the method: it is useful only when the diagnostic expertise expressed in books, diagnostic criteria, or personal experience, is not enough to reach a good result. Obviously, when a clear and working solution is available, then trying to automatically generate a new one could be a waste of time. But this is true up to a certain point because, as we saw, the analysis could lead to more efficient “shorter” solutions that may also contain unexpected useful results.
3. The third possible limit stays in the reproducibility of our results. We think that the information in **Figures 1–5** and the relative text is enough for a good programmer to be able to reproduce the core of the KBS. Actually, this is what we have already done because the original KBS from 1989 was written for the Macintosh OS and in the year 2000, it was rewritten from scratch for the Windows OS in a different language, using just the screens and the “behavior” of the old version, not the code.
4. The fourth limit is common with any KBS: it is the complexity. While it is relatively easy to create “small” KBS with a few dozens of rules and concepts, when those numbers grow to hundreds or thousands, there is the risk to lose control of the overall logic. There is not a computer-aided solution even though a few tools may be useful: it is mainly a matter of experience, continuous review, and intuition. But, as we wrote concerning the “internal conclusions,” it is just the complexity that allows the creation of tools like those in this paper. So complexity and size of the KB are at the same time a risk and the root of the most interesting results.
5. A fifth limit is the intense human work that is necessary to do the manual review, to use the tools, and to check the rules. This work could in theory be avoided using tools such as ANN and CBR, but these approaches have other limitations in terms of diagnostic reasoning and justification.

CONCLUSION

If the KBS has a valid and user-friendly “cognitive” interface, its daily clinical use brings no special problems and the learning curve is very fast for residents or other specialists. In our experience, it is about 3 days. There is no particular reaction when the program displays its diagnostic conclusions; it becomes part of the routine workflow like dealing with a teacher that is never tired of explaining. Then, the physician makes his own decisions, confirming or changing those from the KBS.

REFERENCES

- Amato, A. A., and Dumitru, D. (2002). “Chapter 23 acquired neuropathies,” in *Electrodiagnostic Medicine*, 2nd Edn (Elsevier), 937–1041. Available at: http://www.elsevierhealth.com/media/us/samplechapters/9781560534334/Chapter_23_Acquired_Neuropathies.pdf
- Ardizzone, E., Bonadonna, F., Gaglio, S., Marcenò, R., Nicolini, C., Ruggiero, C., et al. (1988). Artificial intelligence techniques for cancer treatment planning. *Med. Inform. (Lond.)* 13, 199–210. doi:10.3109/14639238809010100
- Barohn, R. J., Kissel, J. T., Warmolts, J. R., and Mendell, J. R. (1989). Chronic inflammatory demyelinating polyradiculoneuropathy. Clinical characteristics, course, and recommendations for diagnostic criteria. *Arch. Neurol.* 46, 878–884. doi:10.1001/archneur.1989.00520440064022

The knowledge transfer process is very useful not only for building the KBS but also for teaching and for learning. For example, our KBS is able to print all its knowledge in the form of an “electronic book” that can be read and used without a computer. Moreover, diagnostic criteria written inside the KBS must be “complete” and require no further “clinical sense,” otherwise the system just does not work.

As said before, the dimensions of the HIS–KBS are very important, so, in theory, a huge HIS that includes all the different medical specialties and subspecialties would be the ideal target. In practice, this is very difficult, both for HIS and for KBS. Huge systems become complex and slow to build and maintain, and often unable to support a continuous evolution and creation of tools. The solution could be a “society of expert modules” working in the HIS–KBS, and able to exchange data and “talk” to each other and possibly compete for medical quality and results.

This society of expert modules could always be working in the background even for simple cases because unexpected and rare diseases sometimes occur and usually take a long time before being recognized. Naturally, this implies that the diagnostic criteria in the rules must be able to avoid many FP and the consequent information overload.

We think that is necessary to discover and test new ways for knowledge transfer. Also, if a huge quantity of knowledge is transmitted with books and the modern electronic media, it is always necessary that someone read, understand, and learn it before the knowledge is absorbed. KBS can be a form of active knowledge that can be transmitted and ready to use without being learned, or, even better, making the learning process easier and more effective.

The other big issue is with regard to the “clinical sense.” This could be the expertise developed after many years of daily clinical practice, resulting in special intuition or the memory for thousands of analyzed cases. Such expertise is difficult to transfer, mainly because it is only partially conscious. A KBS, with tools such as those demonstrated in this paper, could capture some of that clinical sense in a way that the expert is not able to express. It is like if the KBS “observes” the way the human expert behaves in his daily clinical work and then creates rules that make the KBS more similar to the human expert of what is possible by manually programming the KBS itself.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fict.2015.00026>

- Bonadonna, F. (1990). HyperShell: an expert system shell in a hypermedia environment – application in medical audiology. *Med. Inform. (Lond.)* 15, 105–114. doi:10.3109/14639239008997662
- Brainin, M., Barnes, M., Baron, J. C., Gilhus, N. E., Hughes, R., Selmaj, K., et al. (2004). Guidance for the preparation of neurological management guidelines by EFNS scientific task forces – revised recommendations 2004. *Eur. J. Neurol.* 11, 577–581. doi:10.1111/j.1468-1331.2004.00867.x
- Bromberg, M. B. (2011). Review of the evolution of electrodiagnostic criteria for chronic inflammatory demyelinating polyradiculoneuropathy. *Muscle Nerve* 43, 780–794. doi:10.1002/mus.22038
- Buchanan, B. G., and Duda, R. O. (1982). *Principles of Rule-Based Expert Systems*. Report No.STAN-C’S82-926. Stanford University. Available at: <http://i.stanford.edu/pub/cstr/reports/cs/tr/82/926/CS-TR-82-926.pdf>

- Campillo-Gimenez, B., Jouini, W., Bayat, S., and Cuggia, M. (2013). Improving case-based reasoning systems by combining K-nearest neighbour algorithm with logistic regression in the prediction of patients' registration on the renal transplant waiting list. *PLoS ONE* 8:e71991. doi:10.1371/journal.pone.0071991
- Darlington, K. W. (2011). *Designing for Explanation in Health Care Applications of Expert Systems*. Available at: <http://sgo.sagepub.com/content/1/1/2158244011408618>
- De Sousa, E. A., Chin, R. L., Sander, H. W., Latov, N., and Brannagan, T. H. III (2009). Demyelinating findings in typical and atypical chronic inflammatory demyelinating polyneuropathy: sensitivity and specificity. *J. Clin. Neuromuscul. Dis.* 10, 163–169. doi:10.1097/CND.0b013e31819a71e1
- Hadden, R. D. M., Nobile-Orazio, E., Sommer, C., Hahn, A., Illa, I., Morra, E., et al. (2006). European federation of neurological societies/peripheral nerve society guideline on management of paraproteinaemic demyelinating neuropathies: report of a joint task force of the European federation of neurological societies and the peripheral nerve society. *Eur. J. Neurol.* 13, 809–818. doi:10.1111/j.1468-1331.2006.01467.x
- Haq, R. U., Fries, T. J., Pendlebury, W. W., Kenny, M. J., Badger, G. J., and Tandan, R. (2000). Chronic inflammatory demyelinating polyradiculoneuropathy. A study of proposed electrodiagnostic and histologic criteria. *Arch. Neurol.* 57, 1745–1750. doi:10.1001/archneur.57.12.1745
- Hughes, R., Bensa, S., Willison, H., Van den Bergh, P., Comi, G., Illa, I., et al. (2001). Randomized controlled trial of intravenous immunoglobulin versus oral prednisolone in chronic inflammatory demyelinating polyradiculoneuropathy. *Ann. Neurol.* 50, 195–201. doi:10.1002/ana.1088
- Hughes, R. A., Bouche, P., Cornblath, D. R., Evers, E., Hadden, R. D., Hahn, A., et al. (2006). European federation of neurological societies/peripheral nerve society guideline on management of chronic inflammatory demyelinating polyradiculoneuropathy: report of a joint task force of the European federation of neurological societies and the peripheral nerve society. *Eur. J. Neurol.* 13, 326–332.
- Joint Task Force of the EFNS and the PNS. (2010). European federation of neurological societies/peripheral nerve society guideline on management of chronic inflammatory demyelinating polyradiculoneuropathy: report of a joint task force of the European federation of neurological societies and the peripheral nerve society – first revision. *J. Peripher. Nerv. Syst.* 15, 1–9.
- Koski, C. L., Baumgarten, M., Magder, L. S., Barohn, R. J., Goldstein, J., Graves, M., et al. (2009). Derivation and validation of diagnostic criteria for chronic inflammatory demyelinating polyneuropathy. *J. Neurol. Sci.* 277, 1–8. doi:10.1016/j.jns.2008.11.015
- Lee, C. S., and Wang, M. H. (2011). A fuzzy expert system for diabetes decision support application. *IEEE Trans. Syst. Man Cybern.* 41, 139/53. doi:10.1109/TSMCB.2010.2048899
- Lhotska, L., Marik, V., and Vlcek, T. (2001). Medical applications of enhanced rule-based expert systems. *Int. J. Med. Inform.* 63, 61–75. doi:10.1016/S1386-5056(01)00172-1
- Magda, P., Latov, N., Brannagan, T. H. III, Weimer, L. H., Chin, R. L., and Sander, H. W. (2003). Comparison of electrodiagnostic abnormalities and criteria in a cohort of patients with chronic inflammatory demyelinating polyneuropathy. *Arch. Neurol.* 60, 1755–1759. doi:10.1001/archneur.60.12.1755
- Molenaar, D. S. M., Vermeulen, M., and de Haan, R. J. (2002). Comparison of electrodiagnostic criteria for demyelination in patients with chronic inflammatory demyelinating polyneuropathy (CIDP). *J. Neurol.* 249, 400–403. doi:10.1007/s004150200029
- Nicolas, G., Maisonobe, T., Le Forestier, N., Léger, J. M., and Bouche, P. (2002). Proposed revised electrophysiological criteria for chronic inflammatory demyelinating polyradiculoneuropathy. *Muscle Nerve* 25, 26–30. doi:10.1002/mus.1214
- Patel, M., Patel, A., and Virparia, P. (2013). Web based fuzzy expert system implementation using jFuzzyLogic and JAX-Web service for diarrhea diagnosis. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* 3, 409–414.
- Pearce, G., Mirtskhulava, L., Bakuria, K., Wong, J., Al-Majeed, S., and Gulua, N. (2015). "Artificial neural network and mobile application in medical diagnosis," in *2015 17th UKSIM-AMSS International Conference on Modelling and Simulation*. Available at: <http://uksim.info/uksim2015/data/8713a061.pdf>
- Rajabally, Y. A., Jacob, S., and Hbahbih, M. (2005). Optimizing the use of electrophysiology in the diagnosis of chronic inflammatory demyelinating polyneuropathy: a study of 20 cases. *J. Peripher. Nerv. Syst.* 10, 282–292. doi:10.1111/j.1085-9489.2005.10306.x
- Rajabally, Y. A., Nicolas, G., Piéret, F., Bouche, P., and Van den Bergh, P. Y. (2009). Validity of diagnostic criteria for chronic inflammatory demyelinating polyneuropathy: a multicentre European study. *J. Neurol. Neurosurg. Psychiatr.* 80, 1364. doi:10.1136/jnnp.2009.179358
- Reggia, J. A., Perricone, B. T., Nau, D. S., and Peng, Y. (1985). Answer justification in diagnostic expert systems – part I: abductive inference and its justification. *IEEE Trans. Biomed. Eng.* BME-32, 263–267. doi:10.1109/TBME.1985.325446
- Sander, H. W., and Latov, N. (2003). Research criteria for defining patients with CIDP. *Neurology* 60(Suppl. 3), S8–S15. doi:10.1212/WNL.60.8_suppl_3.S8
- Saperstein, D. S., Katz, J. S., Amato, A. A., and Barohn, R. J. (2001). Clinical spectrum of chronic acquired demyelinating polyneuropathies. *Muscle Nerve* 24, 311–324. doi:10.1002/1097-4598(200103)24:3<311::AID-MUS1001>3.3.CO;2-1
- Sasikumar, M., Ramani, S., and Raman, S. M. (2007). *A Practical Introduction to Rule Based Expert Systems*. Narosa Publishers. Available at: <http://sigai.cdacmumbai.in/files/ESBook.pdf>
- Tackenberg, B., Lünemann, J. D., Steinbrecher, A., Rothenfusser-Korber, E., Sailer, M., Brück, W., et al. (2007). Classifications and treatment responses in chronic immune-mediated demyelinating polyneuropathy. *Neurology* 68, 1622–1629. doi:10.1212/01.wnl.0000260972.07422.ea
- Thaisetthawatkul, P., Logigian, E. L., and Herrmann, D. N. (2002). Dispersion of the distal compound muscle action potential as a diagnostic criterion for chronic inflammatory demyelinating polyneuropathy. *Neurology* 59, 1526. doi:10.1212/01.WNL.0000034172.47882.20
- Torres, A., and Nieto, J. J. (2006). Fuzzy logics in medicine and bioinformatics. *J. Biomed. Biotechnol.* 2006, 91908. doi:10.1155/JBB/2006/91908
- Van Asseldonk, J. T. H., Van den Berg, L. H., Kalmijn, S., Wokke, J. H., and Franssen, H. (2005). Criteria for demyelination based on the maximum slowing due to axonal degeneration, determined after warming in water at 37°C: diagnostic yield in chronic inflammatory demyelinating polyneuropathy. *Brain* 128, 880–891. doi:10.1093/brain/awh375
- Van den Bergh, P. Y. K., Hadden, R. D. M., Bouche, P., Cornblath, D. R., Hahn, A., Illa, I., et al. (2010). European federation of neurological societies/peripheral nerve society guideline on management of chronic inflammatory demyelinating polyradiculoneuropathy: report of a joint task force of the European federation of neurological societies and the peripheral nerve society – first revision. *Eur. J. Neurol.* 17, 356–363. doi:10.1111/j.1468-1331.2009.02930.x
- Van Den Bergh, P. Y. K., Jacquerye, P. H., and Piéret, F. (2000). Electrodiagnosis of demyelinating neuropathies. *Acta Neurol. Belg.* 100, 188–195.
- Van Den Bergh, P. Y. K., and Pieret, F. (2004). Electrodiagnostic criteria for acute and chronic inflammatory demyelinating polyradiculoneuropathy. *Muscle Nerve* 29, 565–574. doi:10.1002/mus.20022
- Van Dijk, J. G. (2006). Too many solutions for one problem. *Muscle Nerve* 33, 713–714. doi:10.1002/mus.20581
- Viala, K., Maisonobe, T., Stojkovic, T., Koutlidis, R., Ayrignac, X., Musset, L., et al. (2010). A current view of the diagnosis, clinical variants, response to treatment and prognosis of chronic inflammatory demyelinating polyradiculoneuropathy. *J. Peripher. Nerv. Syst.* 15, 50–56. doi:10.1111/j.1529-8027.2010.00251.x
- Wilson, J., Chawla, J., and Fisher, M. (2005). Sensitivity and specificity of electrodiagnostic criteria for CIDP using ROC curves: comparison to patients with diabetic and MGUS associated neuropathies. *J. Neurol. Sci.* 231, 19–28. doi:10.1016/j.jns.2004.12.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Bonadonna and Killian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.