Multiply Sectioned Bayesian Networks For Neuromuscular Diagnosis

Y. Xiang[◇], B. Pant[†], A. Eisen[†], M.P. Beddoes[*], D. Poole[‡]

[◇] Department of Computer Science, University of Regina, Regina, Sask., S4S 0A2 Canada

[†] Neuromuscular Disease Unit, Vancouver General Hospital

[*] Department of Electrical Engineering, University of British Columbia

[‡] Department of Computer Science, University of British Columbia &
the Canadian Institute for Advanced Research

Abstract

A prototype neuromuscular diagnostic system (PAINULIM) that diagnoses painful or impaired upper limbs has been developed based on Bayesian networks. This paper presents nonmathematically the major knowledge representation issues that arose in the development of PAINULIM. Motivated by the computational overhead of large application domains, and the desire to provide a user with an interface that gives a focused display of a subdomain of current interest, we built PAINULIM using the idea of multiply sectioned Bayesian networks. A preliminary evaluation of PAINULIM with 76 patients has demonstrated good clinical performance.

Keywords: Neuromuscular diagnosis, Probabilistic Reasoning, Bayesian Network, Decomposition

Correspondence to: Yang Xiang, Department of Computer Science, University of Regina, Regina, Sask., S4S 0A2 Canada, tel: 306 585 5226, email: yxiang@cs.uregina.ca

1 Introduction

This paper presents results of our research in developing the PAINULIM expert system for neuromuscular diagnosis involving a PAINful or impaired Upper LIMb. Our research involves the development of a general technique of multiply sectioned Bayesian networks and its application in the above mentioned medical domain. The emphasis of this paper is on the application. Readers are referred to [27] for mathematical details.

Bayesian networks [18] combine probability theory with a graphical representation of domain models. Probability theory provides a language which embeds many intuitive inference patterns of reasoning under uncertainty and guarantees the consistency of inference made upon the representation. Graphical domain models convey directly to users the dependence and independence assumptions made in the domains, which facilitates knowledge acquisition and makes the representation more transparent. They also allow quick identification of dependence relations by tracing arcs in the networks and efficient computation in which difficulty associated with general probabilistic reasoning [22] can be avoided when the networks are sparse. Some of the medical systems based on Bayesian networks include QMR-DT [20] in internal medicine, MUNIN [1] in EMG, PATHFINDER [9] and INTELLIPATH [15] in pathology, and QUALICON [25] in nerve conduction studies.

In the area of neuromuscular diagnosis, several (prototype) expert systems have appeared since the early 1980's: LOCALIZE [4] for localization of peripheral nerve lesions; MYOSYS [23] for diagnosing mono- and polyneuropathies; MYOLOG [6] for diagnosing plexus and root lesions; Blinowska and Verroust's system [3] for diagnosing carpal tunnel syndrome; ELECTRODIAGNOSTIC ASSISTANT [11] for diagnosing entrapment neuropathies, plexopathies, and radiculopathies; NEUROP for neuropathy diagnoses [19]; KANDID [5] and MUNIN [1] aiming at diagnosing the complete range of neuromuscular disorders.

Most of the above systems in neuromuscular diagnosis are rule based. Satisfaction with system testing based on constructed cases has been reported, while ELECTRODIAGNOSTIC ASSISTANT reported clinical evaluation with a 78% agreement rate with electromyographers (EMGers), based on 15 cases. As medical diagnosis involves reasoning with uncertain knowledge, and limitations of rule-based systems for reasoning under uncertainty have been identified [7, 8, 18], we have chosen to build PAINULIM based on Bayesian belief networks.

One exception in the above systems to rule-based structure is MUNIN which is based on Bayesian networks for its uncertain reasoning component. The MUNIN project started in the mid 80's in Denmark as part of the European ESPRIT program. MUNIN is planned to be a 'full expert system' for neuromuscular diagnosis. Functionalities to be included are test-planning, test-guidance, test-set-up, signal processing of test results, diagnosis, and treatment recommendation. The intended users of MUNIN range from novice to experienced practitioners. The knowledge base ultimately will include full human neuroanatomy. MUNIN adopts Bayesian networks to represent probabilistic knowledge. Substantial contributions to Bayesian network techniques have been made (e.g., [14, 12]).

The MUNIN system is to be developed in 3 stages. In the first stage, a 'nanohuman' model with 1 muscle and 3 possible diseases has been developed. In the second stage, a 'microhuman' system with 6 muscles and corresponding nerves is to be developed. The last stage will correspond to a model of the full human neuroanatomy.

2 PAINULIM Domain

PAINULIM started in 1990 at the University of British Columbia with cooperation from the Neuromuscular Disease Unit (NDU) of the Vancouver General Hospital (VGH). Rather than attempting to cover the full range of diagnosis as does MUNIN, PAINULIM sets out to cover the more modest goal of performing diagnosis on patients suffering from a painful or impaired upper limb due to diseases of the spinal cord and/or the peripheral

nervous system. About 50% of the patients visiting NDU are covered by the domain of PAINULIM. The 14 most common diseases considered include: amyotrophic lateral sclerosis, Parkinsons disease, anterior horn cell disease, root diseases, plexus lesions, intrinsic cord disease, carpal tunnel syndrome, and medial, ulnar and radial nerve lesions.

PAINULIM requires from its users specific knowledge and experience: (1) minimum competence in clinical medicine especially in neuromuscular diseases; (2) basic knowledge of nerve conduction study techniques; and (3) minimum experience of EMG patterns in common neuromuscular diseases. We have designed PAINULIM with the objective of supporting the following classes of users: (1) students and residents in neurology, physical medicine and neuromuscular diseases; (2) doctors who are practising EMG and nerve conduction studies; (3) experienced EMGers (as a formal peer review/self evaluation); and (4) potentially different labs to adapt uniform procedures and criteria for diagnoses.

PAINULIM uses three major information sources to make diagnostic recommendations: clinical examination, (needle) EMG studies, and nerve conduction studies. MUNIN has explored only clinical examination and EMG studies [1], and NEUROP has explored only clinical examination and nerve conduction studies [19]).

Given the level of user competence mentioned above, PAINULIM relies on users to interpret data acquired during EMG and nerve conduction studies, and takes the interpretation as evidence (features) for diagnostic inference. Based on the evidence, PAINULIM computes the posterior probability distributions for each disease hypothesis and feature variables not yet observed.

PAINULIM represents explicitly the clinically significant disease-feature relations, which is one of the most important parts of the expertise of experienced EMGers. This choice allowed the rapid development of PAINULIM at the clinical level (versus, for example, the microhuman level). A clinical evaluation of PAINULIM will be presented in the latter part of this paper. The performance at clinical level can then provide us with feedback for further extensions and improvements.

Clinical diagnoses are performed in three steps: anatomical, pathological and etiological. PAINULIM currently works mostly at the anatomical level.

In the development of PAINULIM, two of us (B.P. and A.E.) acted as the domain experts. The network structure and probability values were generated from the subjective assessments supplemented by available medical literature.

3 Localization In Large Application Domains

PAINULIM is based on a Bayesian network which contains 83 variables representing 14 diseases and 69 features (evidence), each of which has up to three possible outcomes. The network is multiply connected (multiple paths exists between a pair of nodes) and has 271 arcs and 6795 probability values. After the system has performed satisfactorily at the current representation level, we plan to extend its representation, which will pose even greater demands on computational resources. As commonly applied, the Bayesian network representation does not consider domain structure, but rather lumps all variables of a domain into one homogeneous or flat network. This is appropriate for small domains. In the development of PAINULIM, such representations conflict with the need for

- Efficient computation, and as a result, short response time in system testing; and moderate space requirements. Rather than asking EMGers to come to powerful workstations, we would like to perform knowledge acquisition and system testing in hospital environments where only PCs are commonly available. We would like to continue running on PCs even when we further elaborate and extend the PAINULIM representation.

- A more natural user interface. There are natural stages during diagnosis where EMGers focus their attention on a subdomain. Presenting the overall domain represented by a large Bayesian network to EMGers all the time would distract them with subdomains they are not interested in at the moment. We expand on this point below.

We made the following observation on the PAINULIM domain based on the practice in NDU. An EMGer, examining a patient with a painful or impaired upper limb, may temporarily consider only his findings' implication on a set of disease candidates. He may not start to consider the diagnostic significance of each available laboratory test until he has finished the clinical examination. After each clinical finding (about five findings for an average patient), he dynamically changes the most likely disease candidates, and based on that, he subsequently chooses another question or laboratory test. After the clinical examination, findings highlight certain disease candidates and make others less likely, which may suggest that nerve conduction studies are of no help at all, while EMG tests are diagnostically beneficial (in NDU, about 60% of patients receive only EMG tests, and about 27% of patients receive only nerve conduction studies). Faced with many alternatives in EMG tests, the EMGer would not perform a test unless it is considered diagnostically necessary based on results in previous tests. In the NDU of VGH, for an average patient, about six EMG tests are performed, and about four nerve conduction studies are performed.

From this scenario and related statistics, we see the important phenomenon of *localization*: During the clinical examination, only clinical findings and disease candidates are of current interest to the EMGer. And during EMG tests, only EMG test results and their implications on a subset of the diseases are the focus of the EMGer's attention. Furthermore, for a large percentage (60%+27%=87%) of patients, only one of either EMG or nerve conduction studies is required, but not both. If the EMGer is assisted by a Bayesian-network-based system, a "small" part of the network would be involved repeatedly for about five queries during each diagnostic period (the clinical period or the EMG period for the above scenario); and for 87% of patients certain parts of the network (either the EMG portion or the nerve conduction portion) may not be of interest at all. If we can organize the network in a way corresponding to this localization, we can restrict our computation to the area of interest and reduce the computational cost. We can also present the user with only the subdomain of his current attention.

To summarize, by localization, we mean the following: There are fixed natural subdomains in large domains with respect to human reasoning activities. At any time, a human reasoner often focuses attention on only one of them. He acquires evidence from and forms queries about one of them at a time. He may shift attention from one to another from time to time. When localization exists, a homogeneous network representation and resultant global propagation of evidence are unnecessary, unnatural and inefficient.

There have been attempts of decomposition of homogeneous Bayesian networks to achieve different goals. However, they did not address localization explicitly:

- Relevant early work includes a method by Pearl [17] which allows postponing evidence propagation in a hierarchical network of hypotheses.
- Pruning Bayesian networks with respect to each query instance reduces computational cost [2]. The method does not provide general support for incremental evidence (i.e., all evidence must be entered at one time).
- Heckerman [10] partitions Bayesian networks into small groups of naturally related variables to ease the construction of large networks. But once the construction is finished, the run time representation is still homogeneous.

- Suermondt, Cooper and Heckerman [21] combine cutset conditioning with a clique-tree inference method and convert the original net into a set of clique trees to obtain computational savings. The cutset is chosen mainly based on network topology. This method does not lead to the exploitation of some forms of localization. A decomposition based on general cutsets or d-separation without addressing localization based on domain considerations may end up with unnatural and inefficient partitions. This is because variables preserving localization may be separated into different partitions. A consequence may be frequent switches of computation between the partitions which contain these variables.

4 Exploiting Localization In Bayesian Networks

We have developed a general technique to exploit localization existing in large domains[27]. It allows the representation of localization preserving subdomains by Bayesian subnetworks[1]. The resultant set of subnets is termed a *Multiply Sectioned Bayesian Network (MSBN)*. Each subnet is then transformed into a junction tree[2] to allow efficient inference in each subdomain. Information channels called *linkages* between junction trees are created to allow the propagation of evidence during attention shift. We then have a linked *junction forest* as a secondary representation of the domain. Due to localization, only one junction tree is active at any time. Probabilities obtained are identical to those that would be obtained from the homogeneous network. Although the overall system can be large, computational requirements are governed by the size of only one junction tree.

Section 4.1 illustrates MSBNs with a simple example. Section 5 discusses how MSBNs are related to the representation of PAINULIM. Reasons why some 'obvious' ways to exploit localization do not work, reasons why MSBNs are necessary for exploiting localization, and other major concepts of MSBNs are presented in [28]. A formal mathematical treatment of MSBN is contained in [27].

4.1 Representation of localization in MSBNs

In practice, a MSBN is constructed one subnet at a time. For the convenience in presenting the concepts, in this section we assume that an UnSectioned Bayesian Network (USBN) is given. Our task is to construct the MSBN/junction forest representation such that localization can be exploited. The following overview summarizes the major technical steps in construction of MSBNs. A simple example is used to illustrate the steps.

- The first step is to identify subdomains according to localization.

  Suppose variables in the network Θ (Figure 1) form three subdomains with localization:[3]

$$\begin{aligned} S^1 &= \{A_1, A_2, A_3, H_1, H_2, H_3, H_4\} \\ S^2 &= \{F_1, F_2, H_1, H_2\} \\ S^3 &= \{E_1, E_2, E_3, H_2, H_3, H_4\} \end{aligned}$$

- The second step is to construct interfaces between subnets and section the USBN into a set (a MSBN) of subnets. The set of interfacing variables should be a *d-sepset*[27]. A set of nodes in a Bayesian network is a d-sepset if it graphically separates two subnets and each node in the set has all its parents contained

[1] Bayesian subnetwork is abbreviated as *subnet* throughout the paper.

[2] A *junction tree* of a Bayesian network Θ is a tree where each node is labelled with a subset of nodes in Θ. The intersection of any two nodes in this tree is contained in each node on the unique path between them. For a formal definition of junction trees, see [27].

[3] The network Θ is not associated with any particular application domain.

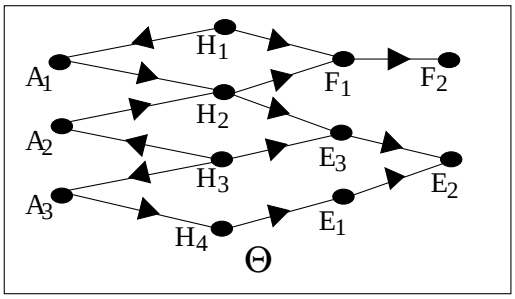

Figure 1: An unsectioned Bayesian network

in one of the two subnets. A node in a d-sepset is called a *d-sepnode*. The idea behind a d-sepset is the d-separation[18] between a *pair* of subnets. When we decide on d-sepset, we must isolate the pair from other subnets. Then the d-sepset makes the pair of subnets conditionally independent. However, when we place the pair back to the original homogeneous network, the d-sepset is not sufficient to guarantee the conditional independence any more. Instead, the union of d-sepsets between a subnet and all its neighbour subnets will be sufficient to guarantee that. The union of d-sepsets between a subnet and all its 'neighbour' subnets d-separates variables in the subnet from variables in the neighbour subnets, and allows all the 'relevant' information to be passed between them [27].

Θ in Figure 1 is sectioned into a MSBN $\{\Theta^1, \Theta^2, \Theta^3\}$ in Figure 2. $I^{12} = \{H_1, H_2\}$ is the d-sepset between Θ^1 and Θ^2; $I^{13} = \{H_2, H_3, H_4\}$ is the d-sepset between Θ^1 and Θ^3; and $I^{23} = \{H_2\}$ is the d-sepset between Θ^2 and Θ^3. If Θ^1 is removed from Θ, then Θ^2 and Θ^3 are conditionally independent given I^{23}. But Θ^2 and Θ^3 are *not* conditionally independent in Θ given I^{23}. Instead the union of I^{23} and I^{12}, i.e., $\{H_1, H_2\}$ render Θ^2 conditionally independent of Θ^1 and Θ^3.

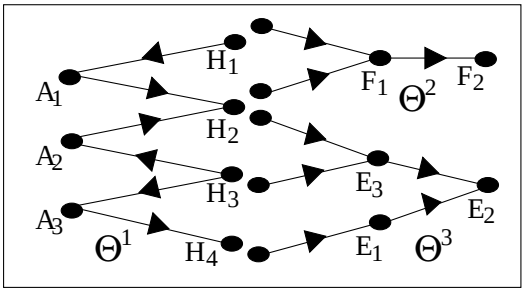

Figure 2: An multiply sectioned Bayesian network from Θ in Figure 1

- Sectioning should have a 'sound' overall structure in order to guarantee correct inference. Intuitively, the structure of a MSBN should ensure that evidence acquired in any subnet be able to propagate to a different subnet through a unique chain of subnets. One rule for a sound sectioning is to have a *covering* subnet which contains all the d-sepnodes between each pair of subnets. A more general rule called 'hypertree' is discussed in [27] which requires only local covering subnets, and a tree structure between the subnets.

  The MSBN in Figure 2 has a covering subnet Θ^1 since $S^1 \supset I^{12} \cup I^{13} \cup I^{23}$. Therefore, the sectioning is sound.

- Once a MSBN has been constructed, we construct probability distributions for each subnet such that a joint distribution equivalent to that of the entire homogeneous network is specified. See [27] for details.
- Transformation of a MSBN to a junction forest requires moralization[4] and triangulation[5] as in other secondary representations [14, 12]. But the modular representation of MSBNs allows the performance of these operations by local computations. This reduces the space requirement to the size of the maximal subnet. Communications between subnets are necessary to ensure correct local computations.

  For example, after moralization and triangulation, the MSBN in Figure 2 is transformed into the set of triangulated graphs $\{\Lambda^1, \Lambda^2, \Lambda^3\}$ in Figure 3.

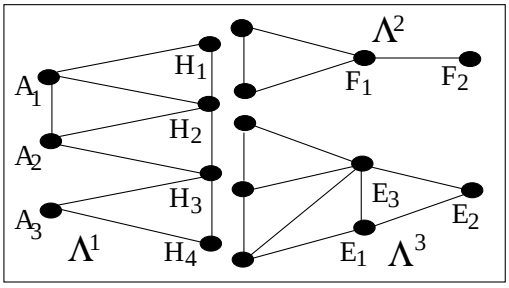

Figure 3: The multiply sectioned Bayesian network in Figure 2 is moralized and triangulated through local computation.

- Each undirected graph (i.e., a moralized and triangulated graph) is converted to a junction tree[6]. Transformation of an undirected graph into a junction tree structure allows efficient computation during diagnostic inference. The transformation is performed in the same way as in [1, 12].

  For example, $\{\Lambda^1, \Lambda^2, \Lambda^3\}$ in Figure 3 is transformed into $\{\Gamma^1, \Gamma^2, \Gamma^3\}$ in Figure 4.
- To finish the structure of the MSBN, we create linkages between the covering subnet and others. A linkage is the intersection of two cliques in different junction trees. The union of all linkages between two adjacent trees equals the d-sepset between them. With linkages created, the set of junction trees becomes a linked junction forest.

  A junction forest consists of a set of linked junction trees, each of which corresponds to a localization preserving subdomain and can be used separately during inference computation. Note that a junction forest allows multiple linkages between a pair of junction trees. These linkages serve as the information channels between junction trees that are used to transfer evidence from one tree to another.

  For example, in Figure 4, linkages between junction trees are indicated with ribbed bands. There is one linkage between clique 1 of Γ^1 and clique 2 of Γ^2, namely, $\{H_1, H_2\}$. The two linkages between Γ^1 and Γ^3 are $\{H_2, H_3\}$ and $\{H_3, H_4\}$.

[4] A Bayesian network is *moralized* if, for each node, links are added between all its parent nodes and directions on the links are dropped. After moralization, a Bayesian network becomes an undirected graph.

[5] An undirected graph is *triangulated* if every cycle of length > 3 has a chord. A *chord* is a link connecting two nonadjacent nodes.

[6] A maximal set of nodes all of which are pairwise linked is called a *clique*. The nodes in the junction tree are cliques of the corresponding undirected graph.

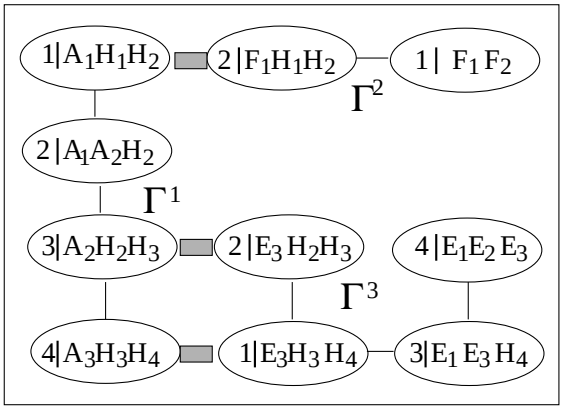

Figure 4: The multiply sectioned Bayesian network in Figure 2 is transformed into a linked junction forest. Linkages between junction trees are indicated with ribbed bands.

- Finally we construct belief tables for each of the cliques and linkages such that a joint distribution for the junction forest equivalent to that of the MSBN is specified. See [27] for details.

4.2 Using the Junction Forest in a Diagnosis

A set of object-oriented operations [27] are defined to perform inference in the linked junction forest. The inference is performed by having one active junction tree. When we want to add evidence or post queries to another junction tree, we shift our attention from the current tree to the target tree.

- The operation **BeliefInitialization** initializes the belief tables of the linked junction forest such that the marginal distribution of each variable can be obtained by local computation at a clique which contains the variable. The initialized forest becomes the permanent representation of the domain and will be reused for each query session.
- The operation **EnterEvidence** is performed to enter evidence into the currently active junction tree and to bring the tree into consistency. Posterior marginal distributions in the active tree can then be obtained by local computation. In the case of incremental evidence, **EnterEvidence** can be applied several times. This is performed in the same way as in [12]. However, in [12], the evidence is propagated to the entire homogeneous network. With the junction forest representation, computational savings are gained by not having to compute all the currently inactive subnets.
- The operation **ShiftAttention** is used to bring a different junction tree of user's attention into active focus. In the case where a covering subnet exists, **ShiftAttention** swaps in and out the covering subnet, and then the target subnet to propagate all the relevant information. In the more general case where the MSBN has an overall 'hypertree' structure, the set of 'intermediate' trees between the previously active tree and the target tree need to be updated. Computational savings are gained by not having to update any of the non-intermediate trees. The posterior marginal distributions in the newly active tree can now be obtained by local computation. They are identical to what would be obtained in a homogeneous network.

  In [12], a global consistency with respect to the entire homogeneous network is maintained before each query is answered. In a junction forest, consistency is maintained with respect to only the subnet of user's current attention. Computational saving are thus gained with no loss of accuracy in answers to queries.

All the above computations consider only one junction tree at a time. No matter how large the overall system, computational requirements are governed by only one subnet/junction tree. This often leads to computational savings in both time and space.

5 Applying MSBNs to the PAINULIM Domain

This section discusses why MSBNs are natural to the PAINULIM domain, and describes the benefits obtained by adopting a MSBN representation.

Using the MSBN representation, we have partitioned the PAINULIM domain into three natural localization preserving subdomains (clinical, EMG and nerve conduction) which are separately represented by three Bayesian subnets: CLINICAL, EMG, and NCV (Figure 5). Abbreviated names are used in Figure 5. For example, the disease node Cts represents carpal tunnel syndrome, Rc81 represents C8T1 root disease, and Mnd represents motor neuron diseases which include anterior horn cell disease and amyotrophic lateral sclerosis. The feature node pn_frm represents pain in forearm, and umcv represents ulnar motor conduction velocity. Figure 6 illustrates the corresponding linked junction forest.

MSBNs require that the interface between each pair of subnets is a d-sepset (Section 4.1). In PAINULIM, the three subnets are interfaced by the disease variables which they share. Topologically, the d-sepset criterion requires that every d-sepnode have all its parents contained in either one of the two subnets. In PAINULIM, disease variables are roots in each subnets without parents. The d-sepset condition is thus trivially true. Conceptually, evidence obtained from a different subdomain might be relevant to the subdomain of current interest, but this relevancy are summarized by the disease variables that the two subdomains share. That is, the disease variables separate the 'interesting' from the 'relevant'. This separating function is achieved through conditional independence which is enforced exactly by the d-sepset criterion.

The incidence of a disease can be influenced by several factors like age, gendar, occupation, etc. The current version of PAINULIM has not considered these factors in its knowledge representation. If evidence is independent from these factors given the diseases, we need to add these variables as parents of disease variables and ensure that the parents for each disease are contained in the same subnet. Then the d-sepset condition is satisfied. In the case these factors also influence the presentation of some evidence, we need to add these variables in the d-sepsets and again the d-sepset condition is satisfied.

MSBNs also require a sound sectioning. The PAINULIM representation satisfies this condition through satisfying the covering subnet rule (Section 4.1). The CLINICAL subnet is the covering subnet which contains all the disease variables. This is natural in PAINULIM domain because at the stage of clinical examination an ENGer would be open to all the possible disease hypotheses. After the clinical examination, he usually narrows the number of hypotheses and chooses either EMG or nerve conduction studies to further differentiate the hypotheses.

The MSBN is implemented in an expert systems shell WEBWEAVR [26], and WEBWEAVR is used for the development of PAINULIM. Adopting the MSBN representation in PAINULIM, we are able to provide users of PAINULIM with a natural interface, and to reduce the computational cost approximately by half relative to using a USBN approach (both space and time) taking into account the repetition of d-sepnodes and the computation required for attention shifts. The reduction of computational cost allows us to use hospital equipment (IBM-compatible PCs) to construct, refine and run PAINULIM interactively with EMGers right in the hospital laboratory. This greatly speeded up the development of PAINULIM. It also enhances the system's usefulness to laboratory personnel.

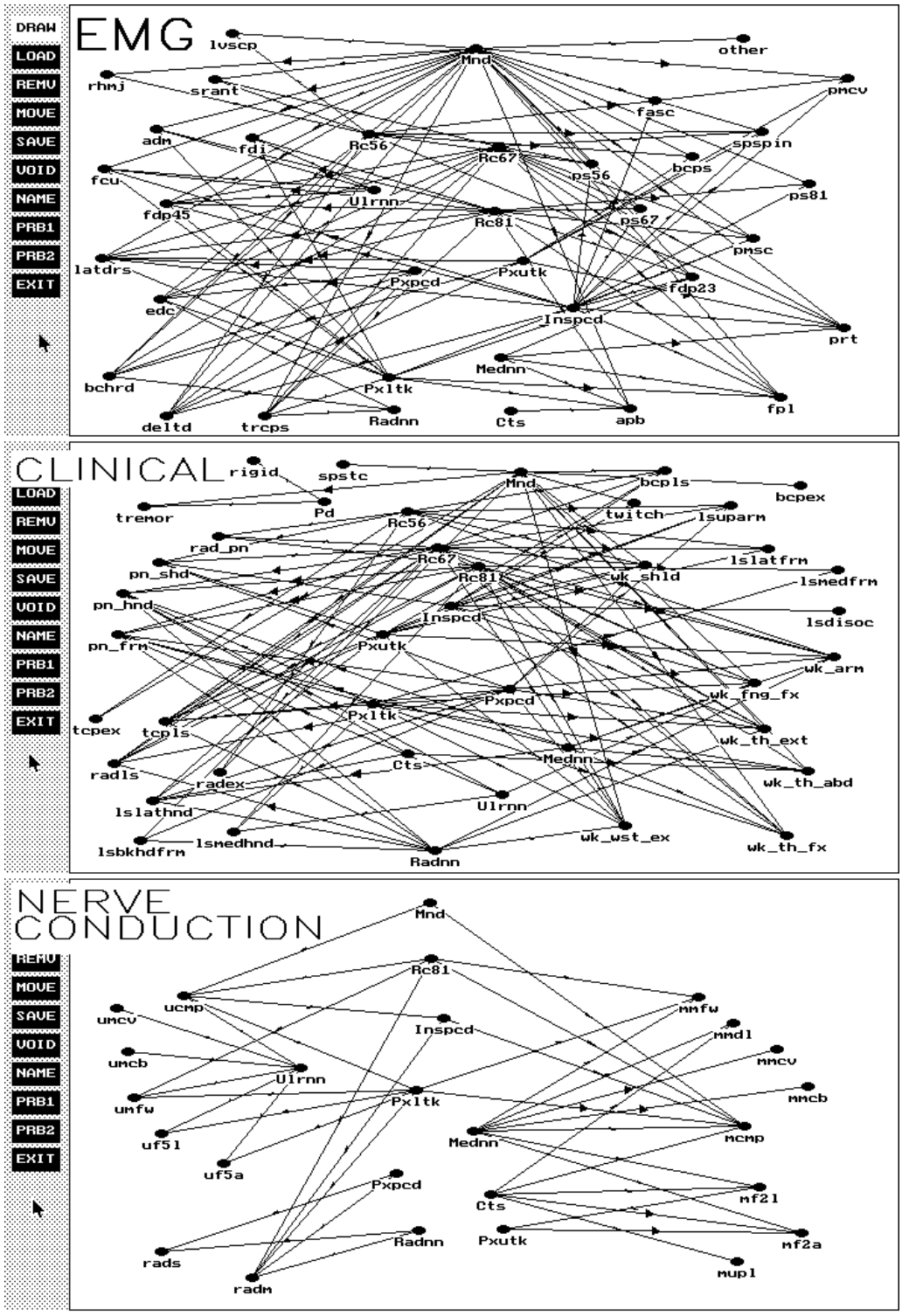

Figure 5: The three subnets of PAINULIM (generated with the WEBWEAVR shell)

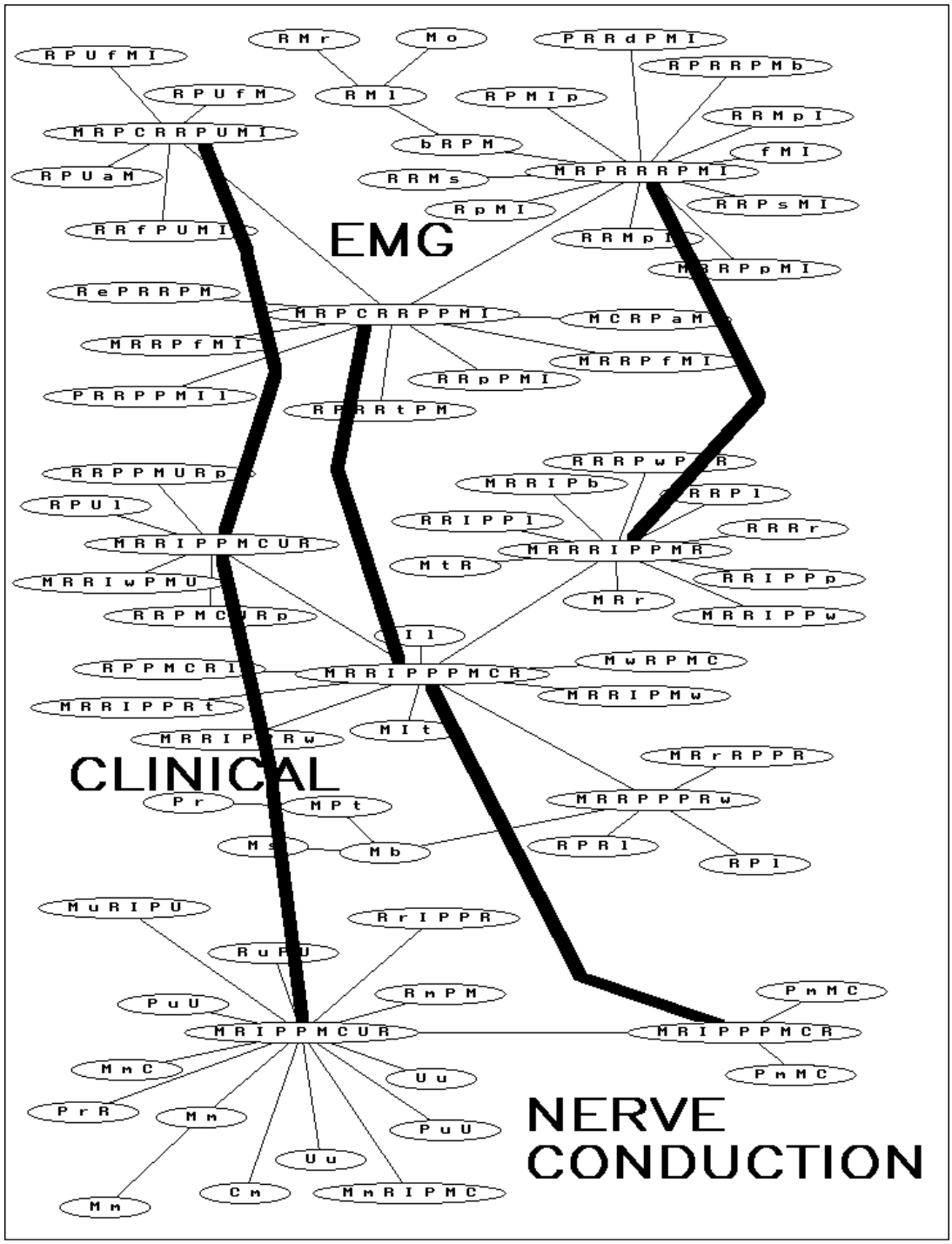

Figure 6: The linked junction forest of PAINULIM. The thick lines represent the linkages between junction trees. The label of each clique is generated by including only the first letter of the name for each node in the clique. For example, the clique label Cm at the bottom is generated from Cts and mupl which are the names of two nodes contained in nerve conduction subnet in Figure 5.

Further extension of PAINULIM may include other conventional tests in its representation. As mentioned in Section 2, the value of many EMG and nerve conduction feature variables are entered by system users based on their subjective interpretation of the recorded signals from EMG machines. Some of these values may be obtained by reasoning with quantities directly obtainable from EMG machines. This requires that the corresponding feature variables be connected to a new subnet at a lower abstraction level. With the MSBN representation, these extensions will not increase the requirement of computational resources since the computational cost of a MSBN is governed by the size of one subnet/junction tree, not by the size of the overall MSBN/junction forest.

Like other exact methods for inference in Bayesian networks, the MSBN requires sparseness of the network topology. With a sparse domain representation, MSBN exploits localization to gain additional computational savings. On the other hand, in a domain where variables are completely connected, there is no localization. We expect that many large domains, which can be represented by sparse Bayesian networks, can benefit from the MSBN representation. We are preparing to test the generality of localization and MSBN representation in other domains.

6 Other Issues in Knowledge Representation

6.1 Multiple Diseases

PAINULIM considers the 14 most common diseases in patients with a painful or impaired upper limb. A patient could suffer from multiple neuromuscular diseases. We have therefore represented each disease by a separate node.

PAINULIM currently represents all disease hypotheses as binary:

'm_or_s': Significant enough (moderate or severe) to prompt further definitive management, e.g., medication, surgery or other tests as X-rays.

'a_or_m': Not as alarming (absent or mild) as the above. A "wait and watch" attitude can be adopted.

A more refined representation is one of our current goals for future extension.

6.2 'Normal' and 'other' diseases

An issue related to the above representation is how to handle unrepresented diseases. As stated earlier, currently PAINULIM represents the 14 most common diseases of the spinal cord and/or peripheral nervous system which present with a painful or impaired upper limb. The diseases in this category that are not represented in PAINULIM include: posterior interosseous nerve lesions, anterior interosseous nerve lesions, axillary nerve lesions, musculocutaneous nerve lesions, suprascapular nerve lesions, and long thoracic nerve lesions. They are not represented in the current version of PAINULIM because their combined incidence is less than 2%.

Other Bayesian network based systems (e.g. PATHFINDER [10] and MUNIN [1]) have used 'normal' and 'other' to represent 'no disease' and 'other diseases' respectively. A 'normal' outcome is used to signal the situation where none of the features is abnormal. The semantics of 'other' is to explain every possible feature pattern uncounted by the diseases explicitly represented.

A medical expert system is typically consulted for a patient who has some complaints. A 'normal' outcome dedicated to the case where everything is normal seems to be unnecessary. The usefulness of a medical expert system comes out of its capability to identify diseases such that a proper management can be made. In the case that 'other' is given as the explanation of complaints, a definitive management cannot be made. Based on these consideration, PAINULIM does not distinguish 'no disease' and 'other diseases'. We represent only the diseases that we want to explicitly represent and exclude 'normal' and 'other' from our explicit representation.

Given a patient with abnormal features, if every disease represented shows high posterior probability for outcome 'a_or_m', then it is interpreted as suffering problems outside the expertise of the current system. The behavior of PAINULIM with unrepresented diseases has been tested in our evaluation which is reported in Section 7.

6.3 Handling Spatial Distribution of Diseases

Neuromuscular diseases spatially distribute along the peripheral nervous system. Each disease involves lesions in certain nerve segment(s). Two aspects of knowledge are mostly involved in an EMGer's daily diagnoses. One is the causal relation of different lesions with corresponding signs and test outcomes. Another is the understanding of neuroanatomy. The anatomical knowledge enters into diagnosis in the following way: a proximal lesion (at a location closer to the centre of body) is likely to produce symptoms which characterize a distal lesion (at a location further away from the centre of body) along the same nerve, but the reverse is not true.

Based on the way in which anatomical knowledge is used in EMGer's practice, we explore the reasoning pattern "explaining away" [18] intrinsic to probability theory and Bayesian network representation. In our network representation, a proximal lesion tends to cause more abnormal signs and test outcomes than a distal lesion. When a patient presents with a feature pattern typical to a distal lesion, the lesion gets enough evidential support and its posterior probability becomes high. While a proximal lesion on the same nerve remains unlikely (explained away) due to insufficient support. On the other hand, if a patient presents with a feature pattern typical to a proximal lesion, the proximal lesion receives strong support and becomes highly likely, the distal lesions on the same nerve cannot explain all the features and will be explained away.

To our knowledge, there is no other system that explicitly exploits explaining away to represent the spatial distribution of diseases. We find it important to explain this to domain experts, as it facilitates knowledge acquisition and system testing. Our experience with PAINULIM makes us believe that many medical domains where diseases are naturally organized into a hierarchy (spatially or anatomically) can be properly represented this way.

7 An Evaluation of PAINULIM

7.1 Case Selection

Seventy six patient cases in NDU were selected. They had been diagnosed by EMGers before being used in the evaluation. Patients ('normal' patients[7]) whose complaints are diagnosed as caused by problems other than neuromuscular diseases and patients of carpal tunnel syndrome (Cts) occupy the majority of the patient population in the NDU of VGH. A case population taken according to natural sequence would not cover properly all the diseases represented in PAINULIM. An evaluation based on such a population would be over-optimistic as we had found that PAINULIM performed very well for 'normal' and Cts cases during system testing. We instead selected the case population such that for each rare disease there was at least one case. We scanned sequentially the cases, which were diagnosed in a few months around the end of 1990 and which presented with a painful and impaired upper limb. After we had selected a few cases with the same disease, we skipped a few cases for this disease. Both typical and difficult cases were included for each disease.

Table 1 lists numbers of cases involved for each disease. If a case involved multiple diseases, that is, either the patient was diagnosed as suffering from multiple diseases or differentiation among several competing disease hypotheses could not be made at the time of diagnosis, then the count for each disease involved was increased

[7]Following the naming in NDU, we will call this class of patients 'normal' which should not be confused with the 'normal' outcome discussed in Section 6.2.

by 1. The total count is 124. Thus the ratio $124/76 = 1.6$ serves as an indication of the number of multiple diseases in the case population. Three cases (labelled 'Unrepresented' in Table 1) not considered by PAINULIM but presenting with a painful or impaired upper limb were also included in the evaluation to test PAINULIM's performance on diseases it does not explicitly model.

disease	count	disease	count
Amyotrophic lateral sclerosis	5	Carpal tunnel syndrome	16
Anterior horn cell disease	1	Medial nerve lesion	5
C56 root disease	8	Ulnar nerve lesion	16
C67 root disease	19	Radial nerve Lesion	6
C81 root disease	6	Intrinsic cord disease	2
Plexus upper trunk lesions	6	Parkinsons disease	3
Plexus lower trunk lesions	8	Normal	12
Plexus post cord lesions	8	Unrepresented	3
subtotal	61		63

Table 1: Number of cases involved for each disease in PAINULIM evaluation

7.2 Performance Rating

In evaluating PATHFINDER, Ng and Abramson [16] used 'classic' cases with known diagnoses. The agreement between the known diagnosis and the disease with top probability was used as an indicator of PATHFINDER's performance. Heckerman [10] asked the expert to compare PATHFINDER's posterior distributions with his own and to give a rating between 0 to 10. He also evaluated PATHFINDER with a decision-theoretic metric using an expert-generated 'gold-standard' distribution based on features observed by another pathologist.

In evaluating PAINULIM, we want to compare it's performance with the diagnoses made by a pool of EMGers who practised in NDU of VGH. This is appropriate since there exists a database in natural language which well documents the symptoms, test results and diagnosis for each previous patient in NDU.[8] In the PAINULIM domain, a human diagnosis can indicate an absence of neuromuscular disease, or the presence of a single disease, or the presence of multiple diseases. Each disease can be judged as being in different degrees of severity. Sometimes the human diagnosis is uncertain because the test studies are not well designed or incomplete. On the other hand, the current version of PAINULIM represents a disease as binary: 'a_or_m' or 'm_or_s' (Section 6.1). Therefore the central issue of evaluation is how to map the posterior distributions (on binary variables) produced by PAINULIM to the EMGer' diagnosis expressed in different severity and confidence.

In the PAINULIM domain, there is no certain way to know the 'real' state of disease(s) given a patient case. One could compare the expert system with EMGers and take the EMGers' judgments as the gold standard. Since no EMGer is perfect, the evaluation conducted this way may have its limitation. Both the system and the EMGer may make the same error, or the system may be correct while the EMGer makes an error. To address the limitation, when there was conflict between the diagnoses made by the original EMGer and PAINULIM, we consulted a second EMGer. The motivation was that the second EMGer, faced with a conflict, would examine the evidence more carefully and hopefully come up with a decision with better quality. When there is disagreement in diagnoses between two EMGers, consultation of a third EMGer is a common practice. Since we were confined by the time and the people who participated in this project, two of us (A.E. and B.P.) acted as the second EMGer. In the following discussion, when we refer to the second EMGer, we will say so explicitly.

[8] PAINULIM was not compared to these cases prior to this evaluation.

For each case, the patient documentation was examined and the values for feature variables were entered to PAINULIM to compute posterior distributions for all diseases represented.

To compare the EMGer's diagnosis with that of PAINULIM, the following procedure was used for each case:

1. Identify the set S of disease variables included in the EMGer's diagnosis. This set is the union of the set of diseases represented in PAINULIM (call it S_1), and a set S_2 of diseases that appeared in the patient file as the EMGer's diagnosis but not represented in PAINULIM. The diseases in S_2 were in fact used to test PAINULIM's performance on diseases it does not explicitly model.

2. Assign the state for each disease in S consistent with the EMGer's diagnosis. If the disease was mentioned in the patient file, the state was determined by the EMGer's statement in natural language that the disease was mild, or moderate, or severe, or uncertain. Otherwise, presumably, the EMGer judged the disease as absent.

3. Seven rules were defined to rate the performance of PAINULIM at each disease in S relative to the EMGer's diagnosis. Only one of the rules was applied to a disease. Rating scales were: superior (SP), excellent (EX), good (GD), fair (FR), poor (PR) and wrong (WR).

   Rule 1 to 4 were general rules. Their applicability to a disease was determined according to only the EMGer's diagnosis (the state of the disease in S). Figure 7 depicts the rating decision ranges for the four rules. The x-axis is labelled with PAINULIM's posterior probability $p(Disease = m\_or\_s|evi.)$ where evi denotes the evidence (patient findings). This is sufficient since the disease variables are binary ($p(D = a\_or\_m|evi.) = 1 - p(D = m\_or\_s|evi)$). The y-axis is labelled with rule applicability conditions.

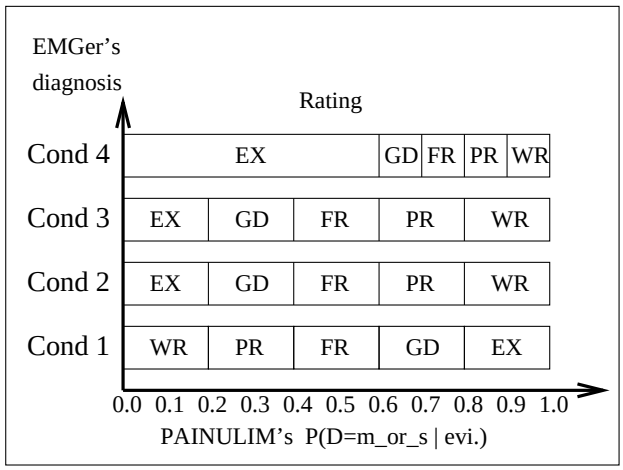

Figure 7: Rule firing conditions for evaluation rule 1 to 4 (see text)

The following conditions determined the first four rules. Rule i is that if **Cond i** holds, then the rating can be obtained by looking at the appropriate value in Figure 7, given PAINULIM's posterior probability.

Cond 1: The disease was judged by the EMGer as moderate or severe.

Cond 2: The disease was judged by the EMGer as absent or mild.

Cond 3: The disease is *not* represented by PAINULIM but was judged by the EMGer as the single disease diagnosis of the case in question.

Cond 4: The disease was judged by the EMGer as uncertain because of evidence which could not be easily explained, or the disease's anatomical site was close to that suspected by the EMGer.

There are three other rules that dealt with situations where the EMGer's diagnosis differed from PAINULIM's diagnosis. These rules had higher priority over Rule 1 to 4. Their applicability to a disease was determined by both the EMGer's diagnosis and PAINULIM's distribution for the disease. In these rules, we distinguish 'the EMGer' who made the diagnosis previously and might have left NDU at the time of evaluation, from the 'second EMGer' who was consulted at the time of evaluation.

Rule 5 If a disease had $p(D = m\_or\_s|evi.) > 0.7$ by PAINULIM, and the disease was different from the original EMGer's diagnosis, and PAINULIM's diagnosis was agreed upon by a second EMGer to whom only the evidence as presented to PAINULIM was available, then the PAINULIM's performance for this disease was rated GD.

Rule 6 If a disease had $p(D = m\_or\_s|evi.) > 0.7$ by PAINULIM, and the disease was different from the original EMGer's diagnosis, and the PAINULIM's diagnosis was agreed upon by a second EMGer to whom the conflict was presented, then the PAINULIM's performance for this disease was rated SP.

Rule 7 If PAINULIM assigned $p(D = m\_or\_s|evi.)$ between 0.7 and 0.3 for a disease D based on evidence ignored by the original EMGer, and a second EMGer considered the evidence for this disease significant and agreed with PAINULIM, then the PAINULIM's performance for this disease was rated SP.

4. For each disease in S, the applicable rule was selected and PAINULIM's performance for this disease was rated by the rule. As noted above, rules 5, 6 and 7, when applicable, overwrite rules 1, 2, 3, and 4.
5. After the rating for each disease was assigned, the lowest rating was given as the rating of PAINULIM's performance for the case in question.

Having presented the evaluation procedure, some explanation of the seven rules is appropriate. For each represented disease, applications of Rule 1, 2, and 4 involved the comparison of the decision range with the distribution for the disease by PAINULIM. For each unrepresented disease, the application of Rule 3 involved the comparison of the decision range with the distribution for each disease represented by PAINULIM. The intuition was that PAINULIM should not confuse the unrepresented disease with any represented disease.

Rules 1, 2 and 3 correspond to the situation where the EMGer had a confident diagnosis. Note that, the corresponding decision ranges in Figure 7 were uniform. The intuition for these rules was that we treated the EMGer's diagnosis as the gold standard if the case was well within his expertise so that he was confident about the diagnosis. For Rule 1, if PAINULIM generated high probability then its behavior was considered satisfactory. The lower the probability, the less satisfactory. For Rule 2 and 3, low probabilities from PAINULIM was desired ($p(D = m\_or\_s|evi) = low$ corresponds to $p(D = a\_or\_m|evi.) = high$). Rule 3 corresponds to the situation where PAINULIM performed on a disease it does not explicitly model (Section 6.2). Since the disease was unrepresented, PAINULIM should behave the same as Rule 2.

Rule 4 corresponds to the situation where the EMGer had an uncertain diagnosis. Note that the decision range is uniform except a wider range of EX. The intuition was that since the EMGer was uncertain about the severity of the disease, PAINULIM should also give a warning but should not be overcommitted with the insufficient evidence.

As mentioned in Section 2, PAINULIM currently works mostly at the anatomical level and cannot process much of the etiological and pathological evidence. Rule 5 handled the situation where the human diagnosis was arrived at by nonanatomical evidence not represented in PAINULIM. Acknowledging the limitation, instead of

failing PAINULIM in these cases, we would like to know if the diagnosis was acceptable at the current level of representation. Such cases, however, amount to only four out of the total 76. When the rule condition was satisfied and unrepresented evidence was present in the patient file, the list of symbolic features as presented to PAINULIM, together with PAINULIM's diagnosis, was presented to the second EMGer. Whether or not he agreed with the diagnosis was asked.

Rules 6 and 7 were designed to identify situations where PAINULIM behaved superior to the original EMGers. When the rule condition was satisfied, the full patient file (including the original EMGer's diagnosis) and the PAINULIM's diagnosis were presented to the second EMGer. Whether he agreed with one or the other was asked. To avoid bias, when the rule succeeded, the second EMGer had to state explicitly the particular pieces of evidence which lead to his disagreement with the original EMGer and to his agreement with PAINULIM. The pieces of evidence were recorded for later inspection. Out of 76 cases, PAINULIM showed superior behavior in twelve cases (five by Rule 6 and seven by Rule 7).

In steps 1 and 2 of our procedure, we extracted the original EMGer's most likely composite hypothesis [18] from the database. The hypothesis is over the set S excluding the diseases judged as uncertain. Our comparison in steps 4 and 5 was performed based on only marginal probabilities, not the joint probability over S. It has been proved [29] that when the marginal probability of an outcome of a variable is high enough, this outcome must be the value of the variable in the most likely composite hypothesis containing the variable. The mathematical detail regarding how high is enough is presented in [29].

7.3 Evaluation Results

The evaluation of 76 patient cases has the following outcomes: SP: 12, EX: 49, GD: 9, FR: 3, PR: 1, WR: 2. The EX or SP rate is 0.80 with 95% confidence interval being $(0.695, 0.885)$ using a standard statistic technique (estimation for confidence intervals in Bernoulli trials [13]). The GD or EX or SP rate is 0.92 with 95% confidence interval being $(0.836, 0.971)$. More detailed results are listed in Table 2.

Cases by EMGer's Diagnosis	SP	EX	GD	FR	PR	WR	subtotal
confident, normal	1	9					10
confident, single mild disease	2	3	1	1			7
confident, single mod. or sev.	2	23	2	2			29
confident, multiple diseases	2	4	2		1	1	10
single unrepresented disease		1	1			1	3
uncertain diagnosis	5	9	3				17
subtotal	12	49	9	3	1	2	76

Table 2: PAINULIM evaluation results

7.4 Discussion of Evaluation Results

For the 56 cases in which original EMGers' diagnoses were confident, PAINULIM performed well when either a single disease was involved or no neuromuscular disease was present (GD or better: 53, FR: 3).

For twelve cases, PAINULIM's behavior was rated SP. In seven of them, PAINULIM indicated a second disease which was not counted properly by the EMGer. Careful examination of the feature presentation by the second EMGer showed that the original diagnosis indeed failed to explain the corresponding features. For example, in one case, the EMGer diagnosed the patient as having mild ulnar nerve lesion (Ulrnn), and PAINULIM generated $p(Ulrnn = m\_or\_s|evi.) = 0.06$ and $p(Cts = m\_or\_s|evi.) = 0.45$ which warned the presence of carpal tunnel

syndrome (Cts). After examining the documentation by the second EMGer, it was found that the original EMGer had not explained the reason for the low amplitude of median sensory potential which was considered due to Cts by PAINULIM.

In the other five cases out of the twelve, PAINULIM provided diagnoses significantly different (different disease or different severity) from the EMGers' diagnoses. The examination by the second EMGer showed that PAINULIM's diagnoses took better account of the evidence.

For 17 cases, the EMGer's diagnosis was either uncertain about the possible diseases, or was confident about the major disease but uncertain about the presence of minor diseases. PAINULIM usually indicated several candidates best matching the available evidence which provided hints for a more complete test study. In five cases out of the 17, PAINULIM's performance was rated SP. This shows that when cases are difficult, PAINULIM might be quite useful to EMGers.

For the two cases diagnosed as a disease in the spinal cord or peripheral nervous system but not represented in PAINULIM, PAINULIM's performance was EX for one and GD for the other. In these two cases, $p(D = a\_or\_m|evi.)$ for every disease D was greater than 0.75. This was interpreted as saying the patient was not in a moderate or severe state for any represented disease, and was suffering from other problems (Section 6.2). The evaluation showed that outside the domain of representation, PAINULIM does not confuse a disease with represented diseases whenever the unrepresented disease has its unique feature pattern.

On the other hand, the evaluation revealed several limitations of PAINULIM. One limitation related to unrepresented diseases. One case in evaluation was diagnosed by the EMGer as central sensory loss which was a disorder in the central nervous system. It was outside the PAINULIM domain but also characterized by a painful or impaired upper limb. When restricted within the PAINULIM domain, the disease presentation was similar to radial nerve lesion (Radnn). PAINULIM could not differentiate and had $p(Radnn = m\_or\_s|evi.) = 0.62$ (the rating is WR). This deficiency calls for the expansion of current PAINULIM domain.

PAINULIM works with limited variables. For example, in the clinical subnet, the variable 'lslathnd', which represents loss of sensation in the lateral hand and/or fingers, does not allow a distinction between the front of the hand (median nerve, carpal tunnel syndrome or C67 root disease) and the back of the hand (radial nerve lesion, plexus posterior cord, C67 root disease). When evidence comes towards one of them but not the other, instantiation of lslathnd will enforce (incorrectly) a group of diseases. Part of our current effort is to further refine our representation of features. This and the following point were the major reasons for the degradation of performance in the case of multiple diseases (one PR and one WR out of 10).

Unsatisfactory performance in some cases was due to the inaccuracy of numerical probabilities. One source of the inaccuracy originated from the experts' limited experience with certain diseases. When a disease is very rare (for example, intrinsic cord disease), the EMGer giving the probabilities may have little experience with the disease and thus provide an inaccurate assessment. A few cases rated as FR or PR or WR were due to inappropriate assessment of conditional probabilities. Further refinement and incorporation of a learning facility [24] are planned for a future version of PAINULIM.

The preliminary evaluation has the following limitations: First, it was performed when PAINULIM had only restricted levels of knowledge representation. As stated earlier, PAINULIM works mostly at the anatomical level and most of the pathological and etiological evidence can not yet be processed. In four out of 76 cases in the evaluation, the ratings were made based on unrealistic situations where the pathological and etiological evidence in the patient cases were removed.

Second, the evaluation used retrospective cases to which the original diagnoses were available. As mentioned

in subsection 7.1, this information was used to obtain an adequate number of cases for each represented disease in the selected cases. An evaluation of PAINULIM based on prospective cases would probabily only use a painful or impaired upper limb as the criterion for case selection. Although we expect an overall better performance of PAINULIM in that situation as explained in subsection 7.1, it remains a hypothesis to be tested.

Third, restricted by the available resource, and influenced by the preliminary nature of the evaluation, the evaluators in this evaluation were also the developers of PAINULIM. This might introduce bias to the results, even though efforts had been made to reduce the bias as much as possible. Finally, it remains to be seen, after further extension, that how PAINULIM's output will influence practising EMGers' diagnostic accuracy.

8 Conclusions

Bayesian networks provide a natural, concise knowledge representation method and a consistent inference formalism for building diagnostic expert systems. Using this formalism, we have been able to represent the diagnostic knowledge at the anatomical level in the domain of neuromuscular diagnosis of patients with a painful or impaired upper limb. This includes the representation of multiple diseases distributed spatially. There have been several expert systems developed in the domain of neuromuscular diagnostic domain in the last decade. But there has been no other such system that has reached the stage of clinical evaluation using a large number of patient cases as has PAINULIM. The performance of PAINULIM supports the notion that a probabilistic approach can be appropriate for some large medical applications.

Substantial advances have been made in probabilistic reasoning using Bayesian networks. However, exact computation in large application domains still incurs large computational overhead. Knowledge acquisition and system testing at medical doctors' work sites can speed up the development of medical expert systems. This is facilitated by systems that can be implemented using inexpensive computational resources commonly available in hospitals. The possibility of deployment of large systems using such resources also increases the usefulness of these systems.

The homogeneous representation of Bayesian networks is a problem we have addressed. We have developed a general solution (MSBN) [27] to the problem and have applied it to PAINULIM by representing and exploiting localization naturally existing in this domain. This allowed reduction of computational cost with no loss of inference accuracy. The MSBN representation also potentially supports the further extension of PAINULIM without the requirement of upgrading computational resources.

We indicate that the applicability of MSBNs in a domain does not require the sharing of all the features of the PAINULIM domain. Although our experience with MSBN is from the domain of neuromuscular diagnosis, we believe that localization is a general feature, and we are preparing to test its generality in other large domains.

Acknowledgements

This work is supported by Operating Grants A3290, OGP0044121, OGP0090307 from NSERC, and CRD3474 from the Centre for Systems Science at Simon Fraser University. Revision of the final draft is partially supported by the University College of the Cariboo. We thank the staff at the Neuromuscular Disease Unit, Vancouver General Hospital for their help and support towards the PAINULIM project, S. Joseph at Simon Fraser University and D. Johnson at the University College of the Cariboo for reviewing early drafts of this paper. Our special thanks are due to G.F. Cooper for his helpful comments and suggestions.

References

[1] S. Andreassen, F.V. Jensen, S.K. Andersen, B. Falck, U. Kjerulff, M. Woldbye, A.R. Sorensen, A. Rosenfalck and F. Jensen, MUNIN - an expert EMG assistant, in: J.E. Desmedt, ed., *Computer-Aided Electromyography and Expert Systems*, (Elsevier, 1989) 255-277.

[2] M. Baker and T.E. Boult, Pruning Bayesian networks for efficient computation, *Proc. 6th Conf. Uncertainty in AI*, Cambridge, Mass. (1990) 257-264.

[3] A. Blinowska and J. Verroust, Building an expert system in electrodiagnosis of neuromuscular diseases, *Electroenceph. Clin. Neurophysiol.* 66 (1987) S10.

[4] M.B. First, B.J. Weimer, S. McLinden and R.A. Miller, LOCALIZE: Computer assisted localization of peripheral nervous system lesions, *Comput. Biomed. Res.* 15 (1982) 525-543.

[5] A. Fuglsang-Frederiksen and S.M. Jeppesen, A rule-based EMG expert system for diagnosing neuromuscular disorders, in: J.E. Desmedt, ed., *Computer-Aided Electromyography and Expert Systems* (Elsevier, 1989) 289-296.

[6] R. Gallardo, M. Gallardo, A. Nodarse, S. Luis, R. Estrada, L. Garcia and O. Padron, Artificial intelligence in EMG diagnosis of cervical roots and brachial plexus lesions, *Electroenceph. Clin. Neurophysiol.* 66 (1987) S37.

[7] D. Heckerman, Probabilistic interpretations for MYCIN's certainty factors, in: L.N. Kanal and J.F. Lemmer, eds., *Uncertainty in Artificial Intelligence* (Elsevier Science, 1986) 167-196.

[8] D. Heckerman and E.J. Horvitz, On the expressiveness of rule-based systems for reasoning with uncertainty, *Proc. AAAI*, Seattle, Washington (1987) 121-126.

[9] D.E. Heckerman, E.J. Horvitz and B.N. Nathwani, Update on the Pathfinder project, *Proc. 13th Symposium on Computer Applications in Medical Care*, Washington, DC (1989) 203-207.

[10] D. Heckerman, *Probabilistic Similarity Networks*, Ph.D. Thesis, Stanford University, Stanford, CA, 1990.

[11] P.W. Jamieson, Computerized interpretation of electromyographic data, *Electroenceph. Clin. Neurophysiol.* 75 (1990) 390-400.

[12] F.V. Jensen, S.L. Lauritzen and K.G. Olesen, Bayesian updating in causal probabilistic networks by local computations, *Computational Statistics Quarterly* 4 (1990) 269-282.

[13] R.J. Larsen and M.L. Marx, *An Introduction to Mathematical Statistics and Its Applications*, (Prentice-Hall, NJ, 1981).

[14] S.L. Lauritzen and D.J. Spiegelhalter, Local computation with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society*, Series B, 50 (1988) 157-244.

[15] B.N. Nathwani, D.E. Heckerman, E.J. Horvitz and T.L. Lincoln, Integrated expert systems and videodisc in surgical pathology: an overview, *Human Pathology* 21 (1) (1990) 11-27.

[16] Keung-Chi Ng and B. Abramson, A sensitivity analysis of PATHFINDER, *Proc. of the Sixth Conference on Uncertainty in Artificial Intelligence* (Cambridge, Mass., 1990) 204-212.

[17] J. Pearl, On evidential reasoning in a hierarchy of hypotheses, *Artificial Intelligence* 28 (1986) 9-15.

[18] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, 1988).

[19] V. Rialle, A. Vila and Y. Besnard, Heterogeneous knowledge representation using a finite automaton and first order logic: A case study in electromyography, *Artificial Intelligence in Medicine* 3 (1991) 65-74.

[20] M. Shwe, B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper, A probabilistic reformulation of the Quick Medical Reference System, Technical Report KSL-90-11, Stanford University, Stanford, CA, 1990.

[21] J. Suermondt, G. Cooper and D. Heckerman, A combination of cutset conditioning with clique-tree propagation in the PATHFINDER system, *Proc. 6th Conf. Uncertainty in AI*, Cambridge, Mass. (1990) 273-279.

[22] P. Szolovits and S.G. Pauker, Categorical and probabilistic reasoning in medical diagnosis, *Artificial Intelligence* 11 (1978) 115-144.

[23] A. Vila, D. Ziebelin and F. Reymond, Experimental EMG expert system as an aid in diagnosis, *Electroenceph. Clin. Neurophysiol.* 61 (1985) S240.

[24] Y. Xiang, M.P. Beddoes and D. Poole, Sequential updating conditional probability in Bayesian networks by posterior probability, *Proc. 8th Biennial Conference of the Canadian Society for Computational Studies of Intelligence*, Ottawa (1990) 21-27.

[25] Y. Xiang, A. Eisen, M. MacNeil and M.P. Beddoes, Quality control in nerve conduction studies with coupled knowledge based system approach, *Muscle and Nerve* 15 (2) (1992) 180-187.

[26] Y. Xiang, B. Pant, A. Eisen, M.P. Beddoes and D. Poole, PAINULIM: a neuromuscular diagnostic aid using multiply sectioned Bayesian networks, *Proceedings of the ISMM International Conference on Mini and Microcomputers in Medicine and Healthcare*, Long Beach, CA (1991) 64-69.

[27] Y. Xiang, D. Poole and M.P. Beddoes, Multiply sectioned Bayesian networks and junction forests for large knowledge based systems, *Computational Intelligence* 9 (2) (1993) 171-220.

[28] Y. Xiang, D. Poole and M.P. Beddoes, Exploring localization in Bayesian networks for large expert systems, *Proc. 8th Conf. Uncertainty in AI*, Stanford, CA (1992) 344-351.

[29] Y. Xiang, Computing the Lower-Bounded Composite Hypothesis by Belief Updating, *Proc. Ninth Canadian Conference on Artificial Intelligence*, Vancouver, BC (1992) 98-105.