STROKEDX: A STROKE DIAGNOSIS PROGRAM

Jeffrey L. Sponsler, MD, MS
Alaska Brain Center, LLC
4551 E Bogard Rd Wasilla AK 99654
jsponsler@akbraincenter.com

ABSTRACT
Decision support for stroke diagnosis is important given the complexity of diagnosis. The project goal was to extend the prototype StrokeDx program with new diagnoses including multiple sclerosis, Brown-Sequard, Weber, Millard-Gubler, and thalamic stroke. Benchmark data files were created to contain symptoms/signs for each new syndrome. Rules were encoded for new diagnoses. StrokeDx employs logic programming to compute a confidence factor for a diagnosis. Each diagnostic rule base was applied to all benchmark datasets. Previous diagnoses included frontal stroke, occipital stroke, Wallenberg syndrome, CADASIL and radial neuropathy. The sensitivity of each diagnostic rule set (for the corresponding benchmark) was 100%. Total diagnosis count is currently 10. The StrokeDx development toolset is extensible and when applied to diagnostic benchmarks is accurate.

KEY WORDS
Expert System, Artificial Intelligence, Neurology, Stroke

1. Introduction

Stroke is the number three cause of mortality and a major source of disability in the US [1]. When an embolus forms and blocks an artery in the brain, ischemia develops causing symptoms that indicate stroke [2]. Computerized stroke diagnosis can support training of medical students and residents and provide aid for medical professionals in the acute care setting. In this report, we describe continued development of an AI program *StrokeDx* that computes stroke diagnoses from patient data using logic programming techniques. Previously, StrokeDx was described [3, 4] and in the current effort has been extended to include new disorders that are either stroke or stroke mimics. Stroke is also called cerebrovascular accident (CVA).

StrokeDx is a hypothesis-driven automated diagnostic system. Further development described herein includes addition of new diagnoses: stroke syndromes, spinal cord syndromes, and multiple sclerosis. To make these changes benchmark files containing signs and symptoms were hand coded. Each benchmark contains over 400 clinical elements (such as “biceps weakness” or “loss of vibration sensation”).

Literature Review

Several AI systems have been constructed for stroke diagnosis. The Anatomic Localizer System [5] employed a decision tree algorithm to compute stroke anatomic locations, contained 30 brain sites, and was similar in performance to human experts. A companion system, Mechanism of Stroke Deducer (MOS) contained knowledge of six stroke types and had 65% accuracy of diagnosis when applied to a patient population [6]. A parent system, MAEISTRO [7] provided a superstructure and user interface for stroke diagnosis and employed Anatomic Localizer and MOS. A program Computerized Medical Decision Making (CMD) using Mount Sinai algorithms produced positive predictive value of 95% for ischemic stroke [8]. A program MICROSTROKE diagnosed stroke and was correct in 72.8% of 250 cases in the Hamburg Stroke Data Bank [9]. An expert system predicts stroke automatically [10]. Neurological expert system development includes diagnosis of epilepsy [11].

Stroke Diagnosis Fundamentals

The current system has accomplished major goals: creation of a set of stroke case benchmarks and a set of hypothesis rule sets for stroke anatomic localization; data transfer from a production electronic medical record into expert system; and data transfer from natural language parsing system into expert system. Other AI work in the Alaska Brain Center lab includes a development environment, NEUROBRIDGE [13] and a digital model of the brachial plexus called PLEXBASE [14]. A natural language front-end called HPARSER [15] is a component of the NEUROBRIDGE.

StrokeDx was designed to emulate the behavior of a neurologist. The clinical analytical process of a neurologist proceeds in this manner: historical information about the symptoms (weakness of the right arm) is obtained. A review of pre-existing conditions (e.g., hypertension) is done. A neurological examination is performed (“strength of right grip is diminished severely”). These data are used to select one (or more) parts of the central nervous system (CNS) or the peripheral nervous system (PNS) where a lesion/disorder might produce these findings. Diagnostic possibilities (usually multiple) are enumerated (stroke, multiple sclerosis, or tumor). Ancillary data (labs, imaging studies) contribute to diagnostic accuracy. MRI diffusion restriction, CT hypodensity, and Babinski are stroke signs [2].

2. Methods and Materials

Development and Software

Goals were enumerated for this research effort: (1) Create benchmark datasets for nervous system disorders (Weber, Miller-Gubler, multiple sclerosis, and Brown-Sequard Syndrome). (2) Create rule trees for these neurological syndromes. (3) Test system components (rules against benchmarks) and report on results. The system uses these engineering tools: Common Lisp [16], Common Lisp Object System (CLOS) [17], and Prolisp [13]. CLOS classes include *examination* and *pxdata*. The examination object is a comprehensive storage object. A pxdata is a single piece of clinical data ("weakness of right biceps"). A Prolisp fact is a pattern (similar to prolog facts). Methods to convert patient pxdata information to Prolisp facts were written.

Benchmark Cases

A *benchmark* case is a set of patient findings that are characteristic of a specific stroke syndrome. The benchmark cases include these diagnoses: frontal stroke, occipital stroke, thalamic stroke, Wallenberg syndrome, Millard-Gubler syndrome, Weber syndrome, Brown-Sequard syndrome, and radial neuropathy. Descriptions of these medical diagnoses can be found in Brazis [2].

The benchmark cases encode patient demographics, stroke risk factors, examination data (e.g., weak right biceps), CT image analysis (hypodensity left frontal), magnetic resonance imaging (MRI) image analysis (diffusion restriction left frontal), computed tomography angiography (CTA) results (occlusion of left middle cerebral artery), and magnetic resonance angiography (MRA) results (occlusion of left middle cerebral artery).

Stroke Benchmarks

Benchmark datasets were created for stroke syndromes. The benchmark for frontal lobe ischemic stroke includes weakness of muscles of the contralateral (CL) arm, weakness of muscles of CL face, CT hypodensity in the ipsilateral (IL) frontal lobe, MRI diffusion restriction IL frontal lobe, CTA occlusion IL middle cerebral artery, MRA occlusion IL middle cerebral artery. The benchmark for occipital stroke includes CL visual field deficit, CT hypodensity in the IL occipital lobe, MRI diffusion restriction IL occipital lobe, CTA occlusion IL posterior cerebral artery, and MRA occlusion IL posterior cerebral artery. The benchmark for Wallenberg Syndrome includes IL loss of facial sensation (pain and temperature), CL loss of pain and temperature sensation of arm and leg, IL CT hypodensity in lateral medulla, IL MRI diffusion restriction in lateral medulla, and CTA and/or MRA occlusion of IL posterior inferior cerebellar artery. The benchmark for thalamic CVA includes CL sensory deficits and MRI showing diffusion restriction in the IL thalamus. The benchmark for Miller-Gubler Syndrome includes MRI diffusion restriction in the anterior lateral pons, IL cranial nerve VI palsy, IL cranial nerve VII palsy, and CL hemiparesis. Benchmark for Weber Syndrome includes MRI diffusion restriction in the midbrain, IL cranial nerve III dysfunction, and CL weakness.

Spine and Peripheral Nerve Disorder Benchmarks

The benchmark for radial neuropathy (associated with weakness of IL wrist extensors, normal brain images, and normal angiogram) was encoded. Arm weakness is common to ischemic stroke (a central nervous system CNS condition) and radial neuropathy (a peripheral nervous system PNS condition. The benchmark for Brown-Sequard Syndrome includes IL weakness, IL loss of touch and vibration, CL loss of pain and temperature sensation.

White Matter Disorder Benchmarks

The benchmark for cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) includes history of dementia, migraines, strokes, family history of dementia, migraines, strokes, subcortical infarcts, and severe diffuse white matter fluid attenuated inversion recovery (FLAIR) hyperintensities [4]. The benchmark for Multiple Sclerosis includes white matter MRI signal changes, cerebrospinal fluid abnormalities, upper motor neuron signs, variable weakness, and variable sensory changes [2].

Electronic Medical Record. We have developed an electronic medical record (NEMR) database management system for a neurology clinic using conventional tools [18]. An interface between NEMR has been developed to transfer patient data to the expert system described in this report. The interface populates the main benchmark data structure. Future work will include tesing StrokeDx against these real cases.

Confidence Factors

For numerical representation of truth this system uses the *confidence factor* (CF). The standard convention for a CF is zero represents false, 0.5 represents unknown, and 1.0 represents true. A mathematical operator, *alpha*, is employed in this system [19]. Applied to confidence factors, alpha combines values synergistically.

3. Running StrokeDx

Benchmark to Prolisp Facts

The first step involves processing the benchmark dataset (a CLOS object) and for each attribute creates a Prolisp fact. A subset of the fact base is listed below. As with Prolog, facts are patterns on which the theorem prover applies rules in a depth-first search. The facts below include propositions about patient symptoms, signs, and test results. For example, the fact (CTA POSTERIOR-CEREBRAL-ARTERY :RIGHT OCCLUDED 1.0) means that the a CTA was done and the right posterior

cerebral artery (PCA) is occluded and this test result has CF 1.

(CTA POSTERIOR-CEREBRAL-ARTERY :RIGHT OCCLUDED 1.0)
(IPSILATERAL :RIGHT :RIGHT)
(BABINSKI :LEFT YES 1.0)
(VISION HOMONYMOUS-HEMIANOPSIA :RIGHT 0.0)
(CONTRALATERAL :RIGHT :LEFT)
(MRA POSTERIOR-CEREBRAL-ARTERY :RIGHT OCCLUDED 1.0)
(MRI-FLAIR-HYPERINTENSITY FRONTAL LEFT WHITE-MATTER 0.0)
(STRENGTH BICEPS RIGHT GRADE-FIVE 1.0)
(MRI-DIFFUSION-POSITIVE FRONTAL LEFT 0.0)

Proof Operator. The *proof* operator tests the diagnosis and binds the *?cf variable* (the computed confidence) and *?hx variable* (not shown in previous examples). Variable ?hx provides a trace of the depth first search through rule space which produced the proved solution. The ?hx data is in lisp format and so can be printed in a clear manner for diagnostic explanation.

Example Prolisp rule: Wallenberg Syndrome

The rule for diagnosis Horner Syndrome is stated here. Three exam findings are required (ptosis, meiosis, and anhidrosis) and the confidence factors are combined using the alpha operator. Ptosis is eyelid droop, meiosis is abnormally small pupil, and anhidrosis is warm/dry face.

```
(pro:define-rule
'(horner-syndrome ?side ?cf)
'((has-ptosis ?side ?p-cf)
(has-meiosis ?side ?m-cf)
(has-anhidrosis face :right ?a-cf)
(alpha3 ?p-cf ?m-cf ?a-cf ?cf)))
```

The top level rule for diagnosis Wallenberg is stated here in an abridged version.

```
(pro:define-rule  '(WALLENBERG ?side ?wallenberg-cf)
 '((contralateral ?side ?cl-side)
(horner-syndrome ?side ?horner-cf)
(has-limb-ataxia ?anatomy ?side ?ataxia-cf)
(loss-of-pain-sensation face ?side ?pain-cf)
(loss-of-temperature-sensation face ?side ?temp-cf)
(loss-of-pain-sensation body ?cl-side ?body-pain-cf)
(loss-of-temperature-sensation body ?cl-side ?body-temp-cf)
(average6 ?horner-cf ?ataxia-cf ?pain-cf ?temp-cf ?body-pain-cf
?body-temp-cf ?wallenberg-cf)))
```

Rule Example: Occipital Stroke

The semantics of this rule: An occipital stroke is diagnosed if there is a contralateral vision deficit and CT shows ipsilateral hypodensity and MRI shows ipsilateral diffusion restriction and angiogram shows ipsilateral PCA occlusion and there are stroke risk factors.

```
(define-rule  '(OCCIPITAL-STROKE ?side ?stroke-cf)
'((contralateral ?side ?cl-side)
(visual-fields homonymous-hemianopsia ?cl-side ?vf-cf)
(ct-hypodensity occipital ?side ?ct-cf
(mri-dwi-positive occipital ?side ?diffusion-cf ?dwi-trace)
(stroke-risk-factors ?risk-cf)
(angiogram-occlusion posterior-cerebral-artery ?side ?angiogram-cf ?angiogram-hx)
(average5 ?vf-cf ?ct-cf ?diffusion-cf ?risk-cf ?angiogram-cf ?stroke-cf)))
```

Diagnosis Brown-Sequard Syndrome

Brown-Sequard syndrome [2] is seen when there is damage to ½ of the spinal cord from the midline laterally. Symptoms include IL motor weakness, IL loss of touch and vibration, CL loss of pain and temperature. The T6 thoracic spinal cord level was selected for this initial benchmark/diagnosis software.

```
PROOF: (BROWN-SEQUARD :RIGHT :T6 ?CF ?HX)
?CF = 1.0
?HX = (BROWN-SEQUARD (SIDE :RIGHT) (LEVEL :T6) (CF 1.0)
(DIMINISHED-TOUCH-AND-VIBRATION-AT-LEVEL (CF 1.0) (LEVEL :T6) (SIDE :RIGHT) (VALUE DIMINISHED)
(DIMINISHED-TOUCH-AT-LEVEL (CF 1.0) (LEVEL :T6) (SIDE :RIGHT) (VALUE DIMINISHED))
(DIMINISHED-VIBRATION-AT-LEVEL (CF 1.0) (LEVEL :T6) (SIDE :RIGHT) (VALUE DIMINISHED)))
(DIMINISHED-PAIN-AT-LEVEL (CF 1.0) (LEVEL :T6) (SIDE :LEFT) (VALUE DIMINISHED)
(DIMINISHED-PAIN-AT-LEVEL (CF 1.0) (LEVEL :T6) (SIDE :LEFT) (VALUE DIMINISHED))
(DIMINISHED-TEMPERATURE-AT-LEVEL (CF 1.0) (LEVEL :T6) (SIDE :LEFT) (VALUE DIMINISHED)))
(LEG-STRENGTH (SIDE :LEFT) (CF 1.0) (COMBINER AVERAGE)
(NORMAL-STRENGTH QUADRICEPS (SIDE :LEFT) (GRADE :GRADE-FIVE) (CF 1.0))
(NORMAL-STRENGTH TIBIALIS-ANTERIOR (SIDE :LEFT) (GRADE :GRADE-FIVE) (CF 1.0)))
(LEG-WEAKNESS (SIDE :RIGHT) (CF 1.0) (COMBINER AVERAGE)
(WEAKNESS QUADRICEPS :RIGHT (GRADE ZERO) (CF 1.0))
(WEAKNESS TIBIALIS-ANTERIOR :RIGHT (GRADE ZERO) (CF 1.0)))
(COMBINER AVERAGE))
```

Diagnosis Millard Gubler Right

Millard-Gubler syndrome represents a lesion of the anterior pons [2]. Below is diagnosis of Millard Gubler applied to its corresponding benchmark dataset.

```
PROOF: (MILLARD-GUBLER :RIGHT ?CF ?HX)
?CF = 0.910
?HX = (MILLARD-GUBLER (SIDE :RIGHT) (CF 0.910)
(ARM-WEAKNESS (WEAKNESS BICEPS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS TRICEPS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS INTEROSSEI :LEFT (GRADE 1) (CF 1.0)) (SIDE :LEFT) (CF 1.0) (COMBINER ALPHA))
(LEG-WEAKNESS (SIDE :LEFT) (CF 1.0) (COMBINER AVERAGE) (WEAKNESS QUADRICEPS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS GASTROCNEMIUS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS BICEPS-FEMORIS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS PERONEUS-LONGUS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS TIBIALIS-ANTERIOR :LEFT (GRADE 1) (CF 1.0)))
```

(CRANIAL-NERVE-6-PALSY (SIDE :RIGHT) (ACTIVITY PARALYZED) (CF 1.0))
(CRANIAL-NERVE-7-PALSY (SIDE :RIGHT) (ACTIVITY PARALYZED) (CF 1.0))
(BABINSKI (SIDE :LEFT) (RESPONSE YES) (CF 1.0))
(CT-DARK-DEFAULT (LOBE ANTERIOR-PONS) (SIDE :RIGHT) (CF 0.5))
(MRI-DWI-BRIGHT (LOBE ANTERIOR-PONS) (SIDE :RIGHT) (CF 1.0)) (COMBINER AVERAGE))

Diagnosis Weber Syndrome Right

Weber Syndrome (midbrain stroke) diagnosis applied to associated benchmark is included below.

PROOF: (WEBER :RIGHT ?CF ?HX)
?CF = 0.848
?HX = (WEBER (SIDE :RIGHT) (CF 0.8482142857142857)
(ARM-WEAKNESS (WEAKNESS BICEPS :LEFT (GRADE 1) (CF 1.0)) (WEAKNESS TRICEPS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS INTEROSSEI :LEFT (GRADE 1) (CF 1.0)) (SIDE :LEFT) (CF 1.0) (COMBINER ALPHA))
(LEG-WEAKNESS (SIDE :LEFT) (CF 1.0) (COMBINER AVERAGE) (WEAKNESS QUADRICEPS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS GASTROCNEMIUS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS BICEPS-FEMORIS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS PERONEUS-LONGUS :LEFT (GRADE 1) (CF 1.0))
(WEAKNESS TIBIALIS-ANTERIOR :LEFT (GRADE 1) (CF 1.0)))
(FACIAL-WEAKNESS (SIDE :LEFT) (RESPONSE UNKNOWN) (DATA DEFAULT) (CF 0.5))
(CRANIAL-NERVE-3-PALSY 1.0 (DILATED-PUPIL :RIGHT 1.0)
(HAS-PTOSIS (CF 1.0) (SIDE :RIGHT) (VALUE YES))
(PARALYZED-EOEM SUPERIOR-RECTUS :RIGHT 1.0)
(PARALYZED-EOEM MEDIAL-RECTUS :RIGHT 1.0)
(PARALYZED-EOEM INFERIOR-RECTUS :RIGHT 1.0))
(BABINSKI (SIDE :LEFT) (RESPONSE YES) (CF 1.0))
(CT-DARK-DEFAULT (LOBE MESENCEPHALON) (SIDE :RIGHT) (CF 0.5))
(MRI-DWI-BRIGHT (LOBE MESENCEPHALON) (SIDE :RIGHT) (CF 1.0)) (COMBINER AVERAGE))

Diagnosis Occipital CVA Right

(proof (occipital-stroke :right ?cf ?hx))
?CF = 1.0
?HX = (STROKE (LOBE OCCIPITAL) (SIDE :RIGHT) (CF 1.0) (HOMONYMOUS-HEMIANOPSIA (SIDE :LEFT) (CF 1.0))(CT-HYPODENSITY (LOBE OCCIPITAL) (SIDE :RIGHT) 1.0) (MRI-DIFFUSION-POSITIVE (LOBE OCCIPITAL) (SIDE :RIGHT) 1.0) (ANGIOGRAM-OCCLUSION(ANGIOGRAM-OCCLUSION
(ARTERY POSTERIOR-CEREBRAL-ARTERY) (SIDE :RIGHT) (CF 1.0)(CTA (CT-ANGIOGRAM-OCCLUSION (ARTERY POSTERIOR-CEREBRAL-ARTERY) (SIDE :RIGHT) (CF 1.0)))(MRA (MR-ANGIOGRAM-OCCLUSION (ARTERY POSTERIOR-CEREBRAL-ARTERY) (SIDE :RIGHT) (CF 1.0)))(COMBINER ALPHA)))

Test Results (Benchmark vs Diagnosis)

The testing algorithm: Loop over each benchmark and apply all stroke rules to that benchmark. Collect confidence factors. These results are included in Table 1 below.

For all benchmarks, the correct diagnosis was found (had the CF closest to 1). For a benchmark, the contralateral diagnosis (left frontal vs. right frontal) calculated CF was found to be correctly 0.5 or less. This is because the benchmark would have all findings (positive or negative) required by the specific diagnosis and no default values.

For certain diagnoses (RT Frontal) applied to another brain locus (LT Occipital), required facts may not have been encoded and thus to have defaulted to unknown (CF 0.5). The confidence in that benchmark/diagnosis pair then tends to have combined confidence of 0.5. After analysis revealed the CF 0.5 results, some benchmark files were augmented with the missing facts (normal patient values) and factors were then closer to the correct zero values.

Discussion of Diagnostic Rule Accuracy

In general, for each benchmark StrokeDx computed the highest CF for the correct diagnosis. Discussions of diagnosis/benchmark scoring follow.

Benchmark CADASIL. The diagnosis CF 1.0 is due to positive lab findings (GOM, Notch3 mutation) that are not present in any other diagnosis.

Benchmark Multiple Sclerosis. Dx CADASIL CF 0.65 reflects the extensive white matter lesion burden seen in both MS and CADASIL. Dx MS matches most criteria and CF 0.93 is calculated. Dx Frontral CVA right and left CF values are low due to absence of MRI stroke findings.

Benchmark Right Frontal CVA. CADASIL rules do not find typical lab findings and historic data is not in this benchmark; CF is 0.5. Dx MS finds no typical CSF findings and no white matter MRI findings producing 0.5 default CF. Findings match for Dx Right Frontal with 0.92 CF. The Left Frontal CVA scores low. Dx left Radial Palsy shows CF 0.75 due to weakness on the same side where stroke would produce plegia.

Benchmark Right Occipital CVA. Frontal CVA rules do not match giving low confidence. The right Occipital CVA rules match well with 1.0 CF; the contralateral rules yield low CF. Wallenberg rules give low CF. Dx Radial palsy defaults to unknown CF. Right Thalamic CVA CF 0.69 is due to sub-goals of normal strength that is seen in occipital CVA and thalamic CVA. Dx Millard-Gubler and Weber are brainstem disorders and few findings match giving low CF.

Benchmark Right Wallenberg. Dx Frontal CVA includes symptoms of CL weakness, no vision changes and CF is computed false on the left due to positive Babinski sign and default on the right. Dx Occipital CVA matches no findings in Wallenberg and CF is low. Dx Right Wallenberg matches but Left Wallenberg CF is zero. Dx Radial Palsy rule includes motor deficit in IL arm and Wallenberg is sensory plus Horner Syndrome; CF is low. Dx Thalamic CVA yields low CF as few symptoms are shared. Dx Brown Sequard includes specific rules for sensory changes at spinal level T6 and matches nothing in

the benchmark. Dx Millard-Gubler yields low CF; infarcts are in different brain regions.

Benchmark Right Thalamic CVA. Dx CADASIL is a white matter disorder and CF is low. Dx MS is a white matter disorder and CF is low. Dx Frontal CVA Right has positive CL Babinski and this coincides with the same finding for this benchmark. Dx Occipital CVA has little in common with thalamic findings; CF is low. Dx Wallenberg has low CF; there are sensory changes in both disorders and rule adjustments are planned. Dx Radial palsy is mainly coded as IL motor deficits and does not match sensory thalamic findings. The right Thalamic CVA rules match well and CF of 0.99 is computed. Dx Brown-Sequard CF is low; there are sensory changes in both disorders and rule adjustments are planned. Dx Millard-Gubler rule includes CL weakness and IL nerve 6 and 7 dysfunction; CF is low. Dx Millard-Gubler rule includes CL weakness and IL nerve 3 dysfunction; CF is low.

Benchmark Right Radial Palsy. The diagnosis radial palsy right applied to this benchmark yielded CF 0.68. The sub-goal radial-weakness was only partially matched because it includes triceps. Triceps muscle is innervated by proximal radial nerve. Radial palsy (focal injury at the mid-humerus) does not involve triceps. The rule should be changed to exclude triceps. Diagnosis Millard-Gubler yielded CF 0.3 (left and right) and this reflects that cranial nerve 6 and 7 were not found to be abnormal. The benchmark is incomplete: this diagnosis ought to score CL weakness higher. Weber CF is 0.3 for left and right sides. The benchmark is incomplete: the Weber diagnosis ought to score CL weakness higher.

Benchmark Right Thalamic Stroke. Dx MS scored CF 0.4 due to low CF for the diffuse white matter heuristic. Unilateral Babinski sign was true for both thalamus and MS and pulled the CF away from zero. Dx right frontal CVA was given CF 0.58 due to positive Babinski sign but CL weakness was not present. Dx radial palsy generates 0.5 CF due to lack of MRI data and so default CF is used. Dx thalamic stroke right yields a satisfactory CF due to the benchmark encoding the complete set of key findings. Dx thalamic stroke left has CF 0.3 due to positive signs being on the contralateral side. Dx Brown-Sequard has CF 0.5 due to key findings being defaulted in this benchmark. Low CF for Millard-Gubler and Weber is due mainly to lack of weakness in any muscle.

Benchmark Right Brown-Sequard T6. This benchmark encodes for syndrome of hemi-cord injury where IL corticospinal tract, IL dorsal column tract, and CL spinothalamic tract are damaged. The benchmark was arbitrarily encoded at the T6 cord level. Left frontal CVA scores higher CF 0.34 than right CVA 0.17 due to the positive right Babinski. Occipital CVA shares few rules with Brown-Sequard and scores correctly low. Negative Wallenberg symptoms (Horner syndrome) score low CF for this benchmark. Thalamic stroke right has negative right Babinski sign and this yields low CF of 0.33. Right Brown-Sequard CF 1.0 reflects the match of all factors for this hypothesis. Rt Weber and Rt Millard Gubler both score CF 0.25 because strength is decreased IL in the benchmark and CL in the diagnosis.

Benchmark Right Millard Gubler. CADASIL rules are false or unknown giving CF 0.5. MS rules give default CF except for positive Babinski sign. Right Millard-Gubler and right frontal CVA agree on left sided weakness and so CF 0.55 is greater than left frontal CVA CF 0.2. Occipital rules bear little in common with this syndrome and the CF is false for this match. Millard Gubler does not produce Horner syndrome or facial sensory changes and therefore CF is low for this pair. Millard-Gubler syndrome produces CL weakness and this matches radial palsy with IL arm weakness giving CF 0.75; the right radial palsy diagnosis has low CF 0.28. Millard Gubler CL weakness matches Brown Sequard IL weakness giving partial match. Rules for Millard Gubler right gave highest CF 0.9. Both Millard and Weber share CL motor weakness and so the right/right analysis yields CF 0.66 and left/right analysis CF is 0.29.

Benchmark Right Weber. CADASIL is a deep white matter disorder and this benchmark is missing findings of this kind giving default CF. The benchmark includes no white matter lesions pulling CF lower to 0.43. Rule for upper motor neuron sign uses combiner function average (0, 1) = 0.5; combiner maximum might produce better behavior. Right Weber and right frontal CVA would injure IL corticospinal fibers thus giving CF 0.5; the R Weber and L Frontal share no corticospinal fibers and CF is 0.2. Occipital CVA rules focus on posterior CNS structures and not brainstem; CF is computed to false. Wallenberg stroke rules focus on posterior CNS structures and not brainstem; CF is computed to false. Radial palsy includes IL arm weakness, Weber includes CL weakness and so the right Weber/left radial palsy diagnosis yields CF 0.75. Dx Thalamic CVA CF 0.4 is low due to absence of MRI findings in thalamus. Diagnoses left Brown and right Weber both include weakness on the left side so CF is 0.5; right Brown Sequard shares no finding with Weber and CF is near zero. Right Weber and Right Millard-Gubler share left sided weakness (corticospinal tract) and so CF is 0.66. Dx Right Weber matches with 0.84 CF.

4. Areas for Future Development

The benchmarks are incomplete. More normal findings will shift away from default confidence and increase the certainty level of confidence values when a hypothesis is incorrect (CF of zero). Weber Syndrome does not yet have rules about Parkinsonism (IL) due to destruction of the substantia nigra.

Aphasia Syndromes. There are number of aphasia syndromes that include Broca's aphasia, Wernicke's aphasia, transcortical sensory and transcortical motor aphasia. These have described symptoms and will be coded into the rule-base as a separate diagnostic module.

Magnetic Resonance Spectroscopy (MRS). MRS technology supports computation of concentrations (in brain tissue) of N-acetyl aspartate, lipids, lactate, creatine, and choline [20]. The ratios of these concentrations can support diagnoses including astrocytoma, stroke, or MS. Encoding a rule-base for MRS diagnosis is planned.

A single multiple sclerosis benchmark file is insufficient. Benchmark augmentation may include optic neuritis, T1 black holes, gadolinium enhancement. Benchmark findings of gadolinium contrast enhancing lesion can be created. Sensory changes (corresponding to no dermatome pattern) are often seen in MS. Unsteady spastic gait is another feature of the disorder.

Another lab prototype is called PLEXBASE [14] and is an object-oriented knowledge base describing the human brachial plexus. PLEXBASE is accessed for clinical computations by Lisp access functions and Prolisp diagnostic rules. Adding diagnostic rule trees for lesions of the brachial plexus is planned and will access the PLEXBASE for anatomy knowledge.

StrokeDx has algorithms to support natural language front-end to transfer patient data for diagnostic analysis [15]; future development will include creating natural language test files that correspond to each benchmark.

The creation of benchmark datasets and diagnostic rule sets for other neurological syndromes such as lacunar stroke syndromes, carpal tunnel syndrome, Parkinson's disease, dementia, epilepsy, and spine disease is planned. Based on the experience reported above, development of benchmarks and rule sets will be relatively synchronous and can be rapidly prototyped. Testing the diagnostic engine against real patient data is also planned.

An NEUROBASE interface program is able to transfer data from the Neurology Electronic Medical Record (NEMR) to the StrokeDx examination data structure [8, 13]. Neurology clinic patient data will be downloaded and StrokeDx rules applied to the cases. This step is important to investigate the real-world application of StrokeDx. There are approximately 500 CVA cases in the Alaska Brain Center electronic medical record database and each case is already labeled with a diagnosis. Test software will compared StrokeDx diagnosis with the case diagnosis giving an accuracy score. The large case volume will test the system intensely.

5. Conclusions

The following conclusions are made from this research effort:

1. This unique prototype has demonstrated the utility of creating benchmarks containing positive and negative patient signs. The benchmarks can exercise diagnostic systems supporting improvements in accuracy. A large library of benchmarks is the goal of future work. The benchmark database is a knowledge resource can be used or testing other diagnostic systems.

2. StrokeDx is a powerful diagnostic system with high sensitivity. StrokeDx uses a robust knowledge framework that is easily extended with new rules. Patient data that is incomplete will trigger defaults (CF 0.5) and diagnostic confidence is less robust. Complete patient data improves diagnostic certainty. The encoding of default rules for missing data has both advantages and disadvantages. The default rules support rule set search completion and so missing data does not cause search to fail. The default value of 0.5 (unknown) semantically is reasonable insofar as the number does not support or deny a diagnosis.

3. The integrated environment of object oriented programming with logic programming continues to demonstrate utility is rapidly developing prototypes.

Dx/Benchmark	CAD	MS	R Fr	R Oc	R Wa	R Ra	R Thal	R Br	R MG	R We
CADASIL	**1.0**	0.65	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Multiple Sclerosis	0.5	**0.93**	0.5	0.5	0.5	0.4	0.42	0.4	0.42	0.43
Frontal CVA L	0.4	0.42	0.25	0.42	0.33	0.25	0.25	0.34	0.2	0.2
Frontal CVA R	0.4	0.42	**0.92**	0.58	0.5	0.25	0.58	0.17	0.55	0.5
Occipital CVA L	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.11	0.15	0.15
Occipital CVA R	0.2	0.2	0.2	**1.0**	0.2	0.2	0.2	0.11	0.15	0.15
Wallenberg L	0.0	0.0	0.0	0.0	0.0	0.0	0.17	0.17	0.16	0.16
Wallenberg R	0.0	0.0	0.0	0.0	**1.0**	0.0	0.17	0.17	0.16	0.16
Radial Palsy LT	0.5	0.5	0.75	0.5	0.5	0.3	0.5	0.5	0.75	0.75
Radial Palsy RT	0.5	0.5	0.25	0.5	0.5	**0.68**	0.5	0.5	0.28	0.28
Thalamic CVA LT	0.7	0.69	0.53	0.52	0.45	0.48	0.3	0.5	0.37	0.36
Thalamic CVA RT	0.7	0.69	0.55	0.69	0.67	0.52	**0.99**	0.33	0.41	0.4
Brown-Seq T6 Lt	0.5	0.5	0.75	0.5	0.5	0.5	0.5	0.0	0.5	0.5
Brown-Seq T6 Rt	0.5	0.5	0.25	0.5	0.5	0.5	0.5	**1.0**	0.03	0.025
Millard-Gubler Lt	0.5	0.5	0.37	0.31	0.43	0.3	0.3	0.5	0.16	0.2
Millard-Gubler Rt	0.4	0.44	0.68	0.43	0.56	0.3	0.4	0.25	**0.9**	0.78
Weber Lt	0.4	0.44	0.3	0.31	0.43	0.3	0.3	0.5	0.29	0.26
Weber Rt	0.4	0.44	0.68	0.43	0.56	0.3	0.4	0.25	0.66	**0.84**

Table 1. Benchmark vs. Diagnoses. The columns are benchmarks and the rows are diagnoses. Each table cell is a diagnostic confidence factor that represents the diagnosis applied to the benchmark. For a given benchmark (e.g., L F) the CF of each diagnosis is given. In this table, C AD = CADASIL, Fr = frontal, Oc = occipital, Wa = Wallenberg, Ra = radial palsy, MG = Millard Gubler, We = Weber, Br = Brown Sequard T6. L is left and R is right.

References

[1] U.S. Centers for Disease Control and Prevention and the Heart Disease and Stroke Statistics - 2007 Update, published by the American Heart Association. Available at http://www.strokecenter.org/patients/stats.htm. Accessed September 2008.

[2] P. Brazis, J. Masdeu, J. Biller. *Localization in Clinical Neurology* / Edition 5. Lippincott Williams & Wilkins, Sept 2006.

[3] J Sponsler, MD. StrokeDX:A logic programming system to diagnose stroke. *Proceedings of the IASTED Conference for Telehealth*, Innsbruck Austria, February 2012.

[4] J Sponsler, MD, M Senta, A Kendrick-Adey. Stroke expert system extended to diagnosis CADASIL. *Proceedings of the IASTED Conference for Telehealth*, Innsbruck Austria, February 2012.

[5] J. Sponsler, F. VanScoy, J. Culberson. The System for Neurological Analysis of Patient Symptoms: An Integrated Artificial Intelligence Prototype. *Proceedings of the IASTED International Conference on Modelling, Simulation and Optimization (MSO2004)*, August 17 -19, 2004, in Kauai, Hawaii.

[5] G. Atkinson, D. Hier, W. McCormick, Development, refinement, and evaluation of a knowledge-base for a stroke expert system. *Proc. of the 9th Annual Symp. on Computer Applications to Medical Care*, Baltimore MD, pp. 320-324, Nov 10-13, 1985.

[6] B. Desai, M. Beachy, K. Biss, M. Evens, D. Hier, H. Hill, W. McCormick, Comparing Bayes rule and the nearest neighbor rule for medical diagnosis. *Proc. 1985 Conf. on Intelligent Systems and Machines*, Rochester MI, Apr 22-24, 1985.

[7] D. Hier D, MAIESTRO: The Michael Reese Hospital - Illinois Institute of Technology Expert System for Stroke, *Proc Annu Symp Comput Appl Med Care*. 1990 November 7: 995–996.

[8] R. Wain, S. Tuhrim, L. D'Autrechy, and J.A. Reggia. *Proc Annu Symp Comput Appl Med Care*. 1991: 94–98. The design and automated testing of an expert system for the differential diagnosis of acute stroke.

[9] K. Spitzer, A. Thie, L. Caplan, K. Kunze, The MICROSTROKE expert system for stroke type diagnosis. *Stroke*. 1989 Oct;20(10):1353-6.

[10] Semak AE, Borisov AV, Churakov AV, Lure TV. Prediction of cerebral stroke using an expert system. *Zh Nevrol Psikhiatr Im S S Korsakova*. 2006; Suppl 17:37-41.

[12] J. Sponsler, M. Werz, M. Kahriman, 2004. Video EEG Expert System: Software to compute seizure focus lateralization and localization prior to epilepsy surgery. *Proceedings of the IASTED International Conference on Modelling, Simulation and Optimization (MSO2004)*, August 17 -19, 2004, in Kauai, Hawaii.

[13] Sponsler J. 2013. NEUROBRIDGE: An AI development environment for neurology. *Proceedings of the 24th IASTED Conference on Modeling and Simulation*, Banff, Canada, July 17-19, 2013.

[14] Sponsler J. 2013. PLEXBASE: A digital model of the brachial plexus. *Proceedings of the 24th IASTED Conference on Modeling and Simulation*, Banff, Canada, July 17-19, 2013.

[15] Sponsler J, HPARSER: Extracting formal patient data from free text history and physical reports using natural language processing software. *Proceedings of the American Medical Informatics Association Annual Symposium*, Washington DC, Nov 3-11, 2001.

[16] G. Steele, *Common Lisp the Language, 2nd Edition*, Digital Press, 1990.

[17] S. Keene, 1989. *Object-Oriented Programming in Common Lisp: A Programmer's Guide to CLOS*, Addison Wesley, 1989, ISBN: 0201175894.

[18] J. Sponsler, F. Pan, 2012. An electronic medical record for neurology. *Proceedings of IASTED Conference on Telehealth*, Feb 2012.

[19] M. Lucks, and I. Gladwell, 1993. A Knowledge Representation for Multiple Criteria Assessment Problems, in *Proceedings of CAIA-93, the Ninth IEEE Conference on Artificial Intelligence for Applications*, Orlando, Florida, March 1-5, 1993, pp. 100-105.

[20] Jansen JF, Backes WH, Nicolay K, Kooi ME (2006). 1H MR spectroscopy of the brain: absolute quantification of metabolites. *Radiology* **240** (2): 318–32.