Journal of
Artificial Intelligence and Soft Computing Techniques

Research Article Open Access

Inherent GIGO Problems with Medical AI

Nelson Hendler[1*], Wallace Judd[2], Dennis Picha[3] and Eddie Tantoco[4]

[1]Former Assistant Professor of Psychiatry and Neurosurgery at Johns Hopkins University School of Medicine for 31 years, and Associate Professor of Physiology at University of Maryland School of Dental Surgery, United States, currently CEO of Mensana Clinic Diagnostics, LLC

[2]Former Vice-President of Research and Development for Kelly Services, United States, currently the President of Authentic Testing Corporation

[3]Chief Executive Officer of Veritycare Management, United States

[4]Former Vice President of Marriott Corporation and Starwood Hotels and Resorts Worldwide, United States, currently representative of the Boulos family

***Corresponding author:** Nelson Hendler, Assistant Professor of Psychiatry and Neurosurgery at Johns Hopkins University School of Medicine for 31 years, and Associate Professor of Physiology at University of Maryland School of Dental Surgery, United States, Tel: 443-277-0306, E-mail: Docnelse@aol.com

Received Date: February 05, 2024 **Accepted Date:** March 07, 2024 **Published Date:** March **08,** 2024

Citation: : Nelson Hendler (2024) Inherent GIGO Problems with Medical AI. J Artif Intel Soft Comp Tech 1: 1-10.

Abstract

The medical literature abounds with articles which report a misdiagnosis rate ranging from 35% to 67% for a variety of disorders, including pneumonia, and heart disease, low back and neck pain, and headache. Primary care physicians missed 68 out of 190 diagnoses (35%) according to a 2013 study, with pneumonia and congestive heart failure the most commonly missed. The two leading causes for misdiagnosis were ordering the wrong diagnostic tests (57%), and faulty history taking (56%).

Keywords: GIGO; Inaccurate Data Bases; Expert Systems; Predictive Analytics; On-Line Diagnostic Tests; Misdiagnosis Of Medical Problems

Background

Diagnostic errors lead to permanent damage or death for as many as 160,000 patients each year, according to researchers at Johns Hopkins University [1]. Not only are diagnostic problems more common than other medical mistakes—and more likely to harm patients—but they're also the leading cause of malpractice claims, accounting for 35% of nearly $39 billion in payouts in the U.S. from 1986 to 2010, measured in 2011 dollars, according to Johns Hopkins [1].

Misdiagnoses of some diseases range from 71% to 97% (RSD, electrical injuries, and fibromyalgia) [2-5]. This high rate of misdiagnosis is costly to insurance companies and other payers, as well as to employers of chronic pain patients, where 13% of the workforce lose productive time, estimated to cost industry $61 billion a year [6] . Furthermore, misdiagnosis creates protracted treatment and psychological problems for the patients themselves. Of all the misdiagnosed disorders, the most prevalent problem is chronic pain, which, according to the Academy of Pain Medicine, accounts for 100,000,000 patients in the United States alone [7]. The annual cost of health care for pain ranges from $560 billion to $635 billion (in 2010 dollars) in the United States, which includes the medical costs of pain care and the economic costs, related to disability, lost wages and productivity [7]. Insurance companies and physicians could improve patient care if they had a mechanism which would address the two leading cause of misdiagnosis: faulty history taking and ordering the wrong medical tests [1]. Therefore, a valuable tool for any health care system would be a questionnaire which could provide accurate diagnoses, and, based on the accurate diagnoses, predict the outcome of an expensive medical laboratory tests, to allow physicians to determine, using "evidence based medicine," which tests would be diagnostic and which test would be of no value. This concept is best exemplified by the Ottawa Ankle Rules, and Ottawa Knee Rules, developed in Canadian emergency rooms. They developed a questionnaire, using "predictive analytic techniques," which could predict which patient would or would not have abnormal ankle or knee X-rays. When the use of the Ottawa Ankle and Knee Rules was applied in emergency rooms, for the selection or denial of patients for ankle or knee X-rays, it decreased ankle and knee radiography up to 26 percent, with cost savings of up to $50,000,000 per year [8-11]. This significant savings was just in the city of Ottawa, and just for ankle and knee pain. If these techniques were applied to other cities and other conditions, the extrapolated savings would be billions of dollars a year.

Later research by the group from Ottawa demonstrated that "expert system" evaluations, based on predictive analytic research, were more accurate in predicting the results of cervical spine X-rays, and CT, than unstructured physician judgment [12-14]. "Expert systems" are best defined as computer software which attempts to mimic the reasoning of a human specialist, and improve the decision making process. This commonly employs a "self-improving system," where the results of a decision are evaluated for accuracy and then fed back to the system, to improve the decision-making process. Theoretically, this leads to more accurate decisions.

These same "predictive analytic" techniques allowed physicians from Mensana Clinic and Johns Hopkins Hospital to develop a Pain Validity Test, an "expert system" which could predict with 95% accuracy who would have moderate or severe abnormalities on medical testing, and predict, with 85% to 100% accuracy, who would have no abnormalities, or only mild ones [15-18]. This "expert system" can be used to determine if a patient is faking, malingering or drug seeking, or if the patient has a valid complaint of pain [15-18, 51,52] The Pain Validity Test has always been admitted as evidence in legal cases in nine states. (see Appendix A for a partial list of these cases).

Past research reports indicate that 40% to 67% of chronic pain patients involved in litigation are misdiagnosed [19,20]. When evaluating complex regional pain syndrome, (CRPS), formerly called reflex sympathetic dystrophy (RSD), Hendler found that 71% of the patients, and Dellon found that 80% of the patients, who were told they had only CRPS I, actually had nerve entrapment syndromes [2,3]. These errors in diagnoses are costly to the patient and the insurance industry alike, since they prolong or result in inappropriate treatment.

Another complicating factor in medical evaluations is the specificity and sensitivity of medical testing. If a test is too sensitive, it can lead to "false positive" results being reported, i.e. claiming there is a pathological condition when one does not exist. Conversely, if a test is very specific, so that the only pathology it detects will be just a certain diagnosis, it will lead to "false negative" reports, i.e. overlooking a medical condition which does exist. So there is an inverse relationship between sensitivity and specificity of medical testing. This concept is best summarized by Table 1

Table 1: Specificity versus Sensitivity

Goal: Let's catch a tuna	Goal: Let's catch a tuna
SENSITIVE TEST	**SPECIFIC TEST**
a ***small*** mesh net	a ***big*** mesh net
Will catch a lot of fish	Everything in the net will be a tuna
Will never miss catching a tuna	*The* smaller fish escaped the net
Will also get mackerel, perch, spot and sea trout which will require further sorting	We have missed some smaller tuna which we would have wanted to keep
False positive results	False negative results
Sensitive but not specific	Specific but not sensitive

A practical example of sensitivity and specificity is the MRI, a test used by most physicians to detect anatomical abnormalities, such as disc damage in the spine. The medical literature shows that the MRI has a 28% false positive rate, and a 78% false negative rate for detecting spinal disc damage, compared to a provocative discogram. [21,22], but most clinical practices do not employ provocative discograms to diagnose vertebral disc damage. Moreover, difference in the technique of performing provocative discograms, with both a provocation and anesthetic component, [23], or if a discogram was merely anatomical versus physiological produce different results [24]. Similar problems arise when physicians use merely a CT, which misses pathology 56% of the time compared to a 3D-CT, in unoperated patients, and 76% of the time in patients who have had previous surgery [25].

One of the factors leading to overlooked or missed diagnoses is the history taking techniques of physicians. Traditionally, physicians would take a careful history, derive a diagnosis and differential diagnosis, and use medical testing to confirm or reject the various diagnoses. However, a shift in the medical evaluation paradigm has occurred. Recent research documents that after a physician entered the room, patients were able to speak an average of 12 seconds, before being interrupted by the physician. The time with patients averaged 11 minutes, with the patient speaking for about 4 minutes of the 11 minutes [26]. Computer use during the office visit accounted for more interruptions than beepers. Another study confirmed the truncated time physicians spend with patients. The average face-to-face patient care time measured by direct observation in this recent study was 10.7 minutes. When researcher evaluated the time spent on "visit-specific" work outside the examination room and combined it with face-to-face time, the average time per patient visit was 13.3 minutes. [27]. Therefore, this increasingly prevalent process follows the format of first getting a list of the symptoms, then getting medical testing pertinent to the symptoms to establish or eliminate diagnoses, and finally reaching a diagnosis, i.e. using medical testing to make a diagnosis. Unfortunately, with an inadequate history, the chance of selecting incorrect medical testing increases, leading to an erroneous diagnosis.

An article by Donlin Long, MD, PhD, published at the time he was the chairman of neurosurgery at Johns Hopkins Hospital, summarizes the compellation of these errors. Dr. Long and his colleagues published an article reporting their evaluation of 70 patients, who had their symptoms for longer than 3 months, who were referred to Johns Hopkins Hospital with normal MRI, X-ray and CT studies [28]. In the absence of abnormal medical testing, no clear diagnosis had been established by the referring doctors, other than cervical sprain or strain. By definition, all 70 of these patients had been misdiagnosed, since sprains and strains are defined as a self-limited injury, which lasts no more than one to six weeks (29,30). Cervical disc disease, facet syndrome, and anteriolysthesis were the most commonly overlooked diagnoses. When properly diagnosed and tested, Dr. Long found that 95% of the patients needed interventional testing, such as facet blocks, root blocks, and provocative discograms, to confirm diagnoses. After the diagnostic testing was performed, 63% of the patients were found to be candidates for anterior or posterior cervical fusions, and 93% of those had good or excellent results post-operatively [28].

All of the above medical facts impact upon the use of artificial intelligence for the diagnosing medical problems. While there are many methodologies employed in the creation of an artificial intelligence application, this paper will limit discussions to the development of "expert systems," which use computer scoring of questionnaires to duplicate diagnoses made by physicians.

Some authors feel only limited progress has been made in expert systems [31]. Engelbrecht feels that the quality of knowledge used to create the system, and the availability of patient data are the two main problems confronting any developer of an expert system, and advocates an electronic medical record system to correct one component of the problem [32]. Babic concurs with the value of the longitudinal collection of clinical data, and data mining to develop expert systems [33]. Those expert systems that seem to have the best results are the ones that focus on a narrow and highly specialized area of medicine. One questionnaire, to cover 32 rheumatologic diseases, consists of 60 questions, was tested on 358 patients [34]. The diagnosis correlation rate between questionnaire and clinical diagnosis was 74.4%, and an error rate of 25.6%, with 44% of the errors attributed to "information deficits of the computer using standardized questions." [34]. However, a later version called "RHEUMA" was used prospectively in 51 outpatients, and achieved a 90% correlation with clinical experts [35]. Several groups have approached the diagnosis of jaundice. ICTERUS produced a 70% accuracy rate while 'Jaundice' also had a 70% overall accuracy rate [36, 37]. An expert system for vertigo was reported, and it generated and accuracy rate of 65%, [38]. This later was reported as OtoNeurological Expert (ONE), which generated the exact same results reported in the earlier article [39]. There was a 76% agreement for diagnosis of depression, between an expert system and a clinician [40]. When a Computer Assisted Diagnostic Interview (CADI) was used to diagnosis a broad range of psychiatric disorders, there was an 85.7% agreement level with three clinicians [41]. Former Johns Hopkins Hospital doctors have developed "expert systems" which address patients with the highest level of misdiagnosis. For headaches, where 35%-70% of patients were mistakenly told they have migraine headaches, the former Johns Hopkins Hospital doctors developed an "expert system" questionnaire which scores answers using Bayesian analytic methods, which gives diagnoses with a 94% correlation of diagnoses of former Johns Hopkins Hospital doctors [42]. For patients with chronic pain problems, where the misdiagnosis rate has been reported to be 40%-80%, even as high as 97% [2-5, 28], the former Johns Hopkins Hospital doctors have an "expert system" questionnaire which gives diagnoses with a 96% correlation with diagnosis of Johns Hopkins Hospital doctors [43]. Finally, in a combined research project with University of Rome, the Diagnostic Paradigm for Chronic Pain was able to predict with 100% accuracy what a surgeon would find intra-operatively, based on the pre-operative diagnosis [44].

The expert systems were designed to be self-improving, using results to re-examine the data, and suggest different type of questions, with improved diagnosis and outcomes. An accurate diagnosis is the ultimate predictive analytic tool. It allows a physician to know the origin of the problem, and select the proper treatment to address the cause of the pathology. However, while the value of improving on the accuracy of diagnosis is obvious, from a cost containment perspective [1,6], the real gold standard is outcome results. This can take the form of elimination of unnecessary testing or surgery, reduced mortality for a given diagnosis, reduced number of doctor visits, reduced use of medication, the selection of the proper type of medication, and a host of quantifiable and objective outcome measures. The major thrust of artificial intelligence programs in medicine is the creation of methodology which improves the accuracy of diagnosis, and ultimately improved outcomes [45]. But the major problem with this approach is the collection of accurate data. Almost all medical artificial intelligence systems utilize the concept of "data mining" which is analyzing massive amounts of data, in the hopes that the "law of big numbers" (the guiding concept in actuarial analysis used by insurance companies) [46] will make their results more accurate. Therein lies the problem with the use of artificial intelligence for improving medical care and containing costs. As reported above, the medical diagnoses are erroneous 35%-80% of the time or more, and the medical testing has false negative rates ranging from 56% to 78%. These errors are multiplicative if taken in tandem. No matter how elegant the analysis of the data, if the data are incorrect, then the garbage in will yield nothing more than garbage out producing a classic case of GIGO.

One way to reduce the errors inherent in reviews of electronic medical records is the application of patient-generated health data (PGHD), where the patient themselves reports symptoms and outcomes, rather than rely on physician notes and interpretations [47]. This methodology would at least afford the opportunity of a more comprehensive and accurate history. This is the methodology used by the "expert systems" designed by former Johns Hopkins Hospital doctors [42, 43]. Another process for optimizing accuracy of artificial intelligence results is to utilize reported outcome studies, and retrospectively determine what factors led to the best results. Unfortunately, this methodology is also prone to multiple sources of errors. In a classic meta-analysis of the efficacy of treatments for reflex sympathetic dystrophy (RSD), Payne found physicians' reported improvement of 12% to 97% of their treated cases [48]. The sources of error ranged from accuracy of diagnosis, to methodology used

to treated RSD, to definition of criteria needed to be considered a treatment success. Of course, outcome results of treatment reported by a physician are subject to bias. The more objective the criteria to measure outcome, the easier it is to quantify data. As an example, research reports which measure "pain relief" are using a very subjective criteria, compared to quantification of narcotic use before and after treatment. The ultimate outcome measures would be validated by third parties, not use physician self-reporting.

The outputs of AI are clearly skewed if the data being used are not consistent with reality. Anyone who works with artificial intelligence knows that the quality of the data goes a long way toward determine the quality of the result. That is the problem to be addressed, and therein lies the opportunity. Mitigating the effects of garbage data becomes a moral imperative when an algorithm is being used to make to invest billions of dollars in a medical treatment program. The problem gets even worse when the garbage is derived from malfeasance or bias. To use machine learning effectively, a researcher should embrace the potential for garbage data and anticipate it. That approach means that these data need to be challenged at the time they become data --- before they are used to create invalid, possibly dangerous outcomes. Mikey Shulman is the head of machine learning for Kensho Technologies, acquired by Standard and Poor (S&P Global) last year. He insists that his teams develop a deep understanding of the datasets they use. "A lot of this comes down to the machine-learning practitioner getting to know the data and getting to know the ways in which it's flawed," he says. [49]. Accurate data will allow machine learning to lower costs and get better outcomes. Therefore there is a need to build tools and methodologies which will help validate the data going into an AI system. If done correctly, everyone benefits.

One preliminary step toward devising a viable AI processing system was reported by Hendler and Joshi (50). They describe the Large Analytic Bayesian System (LABS) which uses Bayesian analytics to list, on a symptom by symptom basis, the likelihood of having abnormal medical testing from a list of possible medical tests associated with a symptom, or cluster of symptoms. LABS produces a rank ordered list of medical tests, comparing the symptom(s) with the frequency and severity of abnormal medical tests. Seventy-eight medical charts with evaluations including all pertinent medical tests, and a completed Diagnostic Paradigm with 2008 possible symptoms were reviewed. On a symptom by symptom basis, medical test results in the chart were compiled using the Large Analytic Bayesian System (LABS) and created a rank ordered list of abnormal medical tests from a list of possible 107 medical tests. This resulted in a 2008 by 107 matrix, which was analyzed by the use of a program called the Diagnostic Test Manager. The results clearly demonstrated that physiological testing, such as root blocks, facet blocks and provocative discograms had nearly double the frequency of abnormalities compared to anatomical tests, such as X-ray, CT scans and MRIs for the same symptoms in the same patient. This evidence-based approach will help physicians reduce the use of unnecessary tests, improve patient care, and reduces medical costs [49,50].

In summary, the potential of AI to enhance medical practice is enormous, ranging from increased accuracy of test interpretation, to more accurate diagnosis with the improved patient benefit. This is clearly demonstrated by a number of the studies above, which achieved accuracy well over 90% [15,16,17,18,28,35,42,43,44,51,52] However, these high accuracy studies all shared a common feature. There were granular in nature, evaluating small number of patients (51-251) with great attention to the individual clinical aspects of each case. Perhaps there is a lesson to be learned from reviewing examples from the basic biological literature, where the variables are less complex, and methodology can be better tested [53-58]. The methodology applied from this research can be applied to clinical evaluations, i.e. the utilization of feedback to enhance machine learning, which is essential for a self-improving system. This is the very essence of artificial intelligence.

On the other hand, data mining of large populations, where the accuracy of individual cases is not determined, nor verified, is susceptible to the GIGO phenomena. No collection of data should be assumed to be accurate, without controlling for all the variable outlined above. Only in this fashion will the full benefit of artificial intelligence in medicine be realized.

References

1. Landro, L (2013) The Biggest Mistake Doctors Make, The Wall Street J.

2. Hendler N (2002) Differential Diagnosis of Complex Regional Pain Syndrome, Pan-Arab Journal of Neurosurgery, pp 1-9, October.

3. Dellon LA, Andronian, E Rosson, GD (2009) CRPS of the upper or lower extremity: surgical treatment outcomes, J. Brachial Plex Peripher Nerve Inj, Feb 20: 4: 1.

4. Hendler N (2005) Overlooked Diagnosis in Electric Shock and Lightning Strike Survivors, Journal of Occupational and Enviromental Medicine 47: 796-805.

5. Hendler N, Romano T (2016) Fibromyalgia Over-Diagnosed 97% of The Time: Chronic Pain Due To Thoracic Outlet Syndrome, Acromo-Clavicular Joint Syndrome, Disrupted Disc, Nerve Entrapments, Facet Syndrome and Other Disorders Mistakenly Called Fibromyalgia, J Anesthesia & Pain Medicine 1: 1-7.

6. Walter F Stewart, Judith A Ricci, Elsbeth Chee, David Morganstein, Richard Lipton (2003) Lost Productive Time and Cost due to Common Pain Conditions in the US Workforce. JAMA 290: 2443-54.

7. Institute of Medicine Report from the Committee on Advancing Pain Research, Care, and Education: Relieving Pain in America, A Blueprint for Transforming Prevention, Care, Education and Research. The National Academies Press, 2011.

8. Ravaud P (1997)Use of Ottowa Ankle Rules Reduces Number of Radiology Requests," JAMA 277: 1935-39.

9. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, et al. (1993) Decision rules for the use of radiography in acute ankle injuries. Refinement and prospective validation. JAMA 269: 1127-32.

10. Stiell IG, Greenberg GH, Wells GA, McDowell I, Cwinn AA, et al. (1996) Prospective validation of a decision rule for the use of radiographs in acute knee injuries, JAMA 275: 611-5.

11. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR (1992) A study to develop clinical decision rules for the use of radiography in acute ankle injuries, Ann Emerg Med 21: 384-90.

12. Bandiera G, Stiell IG, Wells GA, Clement C, De Maio V, et al. (2003) Canadian C-Spine and CT Head Study Group. The Canadian C-spine rule performs better than unstructured physician judgment. Ann Emerg Med 42: 395-402.

13. Dickinson G, Stiell IG, Schull M, Brison R, Clement CM, Vandemheen KL, et al. (2004) Retrospective application of the NEXUS low-risk criteria for cervical spine radiography in Canadian emergency departments, Ann Emerg Med. 43: 507-14.

14. Stiell IG, Wells GA, Vandemheen KL, Clement CM, Lesiuk H, et al. (2001) The Canadian C-spine rule for radiography in alert and stable trauma patients. JAMA. Oct 17;2865: 1841-8.

15. Hendler N, Mollett A, Viernstein M, Schroeder D, Rybock J, et al. (1985) A Comparison Between the MMPI and the 'Mensana Clinic Back Pain Test' for Validating the Complaint of Chronic Back Pain in Women." Pain 23: 243-251.

16. Hendler N, Mollett A, Viernstein M, Schroeder D, Rybock J, et al. (1985) A Comparison Between the MMPI and the 'Hendler Back Pain Test' for Validating the Complaint of Chronic Back Pain in Men." The Journal of Neurological & Orthopaedic Medicine & Surgery 6: 333-7.

17. Hendler N, Mollett A, Talo S, Levin S (1988) A Comparison Between the Minnesota Multiphasic Personality Inventory and the 'Mensana Clinic Back Pain Test' for Validating the Complaint of Chronic Back Pain. J Occupational Medicine 30: 98-102.

18. Hendler N, Baker A (2008) An Internet questionnaire to predict the presence or absence of organic pathology in chronic back, neck and limb pain patients, Pan Arab J Neurosurgery 12: 15-24

19. Hendler N, Kozikowski J (1993) Overlooked Physical Diagnoses in Chronic Pain Patients Involved in Litigation, Psychosomatics 34: 494-501.

20. Hendler N, Bergson C, Morrison C (1996) Overlooked

Physical Diagnoses in Chronic Pain Patients Involved in Litigation, Part 2, Psychosomatics 37: 509-517.

21. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, et al. (1994) Magnetic resonance imaging of the lumbar spine in people without back pain, N Engl J Med 331: 69-73.

22. Braithwaite I, White J, Saifuddin A, Renton P, Taylor BA (1998) Vertebral end-plate (Modic) changes on lumbar spine MRI: correlation with pain reproduction at lumbar discography. Eur Spine J 7: 363-8.

23. Alamin T, Kim MJ, Agarwal V (2011) Provocative lumbar discography versus functional anesthetic discography: a comparison of the results of two different diagnostic techniques in 52 patients with chronic low back pain, Spine J 11: 756-65.

24. Stretanski M, Vu L (2021) Fluoroscopy Discography Assessment, Protocols, and Interpretation, StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing.

25. Zinreich SJ, Long DM, Davis R, Quinn CB, McAfee PC, Wang H (1990) Three-dimensional CT imaging in postsurgical "failed back" syndrome. J Comput Assist Tomogr 14: 574-80.

26. Rhoades DR, McFarland KF, Finch WH, Johnson AO (2001) Speaking and interruptions during primary care office visits. Fam Med 33: 528-32.

27. Gottschalk A, Flocke S (2005) Time Spent in Face-to-Face Patient Care and Work Outside the Examination Room, Ann Fam Med 3: 488-493.

28. Long D, Davis R, Speed W, Hendler N (2006) Fusion For Occult Post-Traumatic Cervical Facet Injury." Neurosurg Quart 16: 129-35.

29. Bonica JJ, Teitz D (1987) The Management of Pain. Lea & Febiger; 2nd edition, p. 375-376, April 1990. 30: 87-1592.

31. Metaxiotis KS, Samouilidis JE (2000) Expert systems in medicine: academic exercise or practical tool? J Med Eng Technol 24: 68-72.

32. Engelbrecht R (1997) [Expert systems for medicine--functions and developments]. Zentralbl Gynakol 119: 428-434.

33. Babic A (1999) Knowledge discovery for advanced clinical data management and analysis. Stud Health Technol Inform 68: 409-413.

34. Schewe S, Herzer P, Krüger K (1990) Prospective application of an expert system for the medical history of joint pain. Klin Wochenschr 68: 466-471.

35. Schewe S, Schreiber MA (1993) Stepwise development of a clinical expert system in rheumatology. Clin Investig 71: 139- 144.

36. Molino G, Marzuoli M, Molino F, Battista S, Bar F, et al. (2000) Validation of ICTERUS, a knowledge-based expert system for Jaundice diagnosis. Methods Inf Med 39: 311-318.

37. Cammà C, Garofalo G, Almasio P, Tinè F, Craxì A, et al. (1991) A performance evaluation of the expert system 'Jaundice' in comparison with that of three hepatologists. J Hepatol 13: 279-285.

38. Kentala E, Auramo Y, Juhola M, Pyykkö I (1998) Comparison between diagnoses of human experts and a neurotologic expert system. Ann Otol Rhinol Laryngol 107: 135-140.

39. Kentala EL, Laurikkala JP, Viikki K, Auramo Y, Juhola M, et al. (2001) Experiences of otoneurological expert system for vertigo. Scand Audiol Suppl : 90-91.

40. Cawthorpe D (2001) An evaluation of a computer-based psychiatric assessment: evidence for expanded use. Cyberpsychol Behav 4: 503-510.

41. Miller PR, Dasher R, Collins R, Griffiths P, Brown F (2001) Inpatient Diagnostic Assessments: 1. Accuracy of Structured vs. Unstructured Interviews. Psychiatry Res. 105: 255-264.

42. Landi, A, Speed, W, Hendler, N, (2018) Comparison of Clinical Diagnoses versus Computerized Test (ExpertSystem) Diagnoses from the Headache Diagnostic Paradigm (Expert System), SciFed Journal of Headache and Pain, 1:1,1-8.

43. Hendler, N., Berzoksky, C. and Davis, R.J. Comparison of Clinical Diagnoses Versus Computerized Test Diagnoses Using the Mensana Clinic Diagnostic Paradigm (Expert System)

for Diagnosing Chronic Pain in the Neck, Back and Limbs, Pan Arab Journal of Neurosurgery, pp:8-17, October, 2007

44. Landi, A, Davis, R, Hendler, N and Tailor, A, Diagnoses from an On-Line Expert System for Chronic Pain Confirmed by Intra-Operative Findings, Journal of Anesthesia & Pain Medicine, Vol. 1, No. 1 pp. 1-7, 2016.

45. Mintz, Y, Brodie, R, Introduction to artificial intelligence in medicine, Minim Invasive Ther Allied Technol, Apr;28(2):73-81, 2019

46. Rouillard, A. Gundersen, G, Fernandez, N, Wang, Z, Monteiro, C, McDermott, C, Ma'ayan, A, The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins, Database (Oxford), Jul 3;2016

47. Jim, H, Hoogland, A, Brownstein, N, Barata, A, Dicker, Knoop, H, Gonzalez, B, Perkins, R, Rollison, D, Gilbert, S, Nanda, R, Berglund, A, Mitchell, R, Johnstone,P, Innovations in research and clinical care using patient-generated health data, CA Cancer J Clin, May;70(3):182-199, 2020.

48. Payne R (1986) Neuropathic Pain Syndromes, with Special Reference to causalgia, and reflex sympathetic dystrophy, Clinical J. of Pain 2: 59-73.

49. Avoiding Garbage in Machine Learning, S&P Global 2019.

50. Hendler N, Joshi K (2021) Predicting Abnormal Medical Tests, on a Symptom by Symptom Basis, Using the Large Analytic Bayesian System (LABS): A Bayesian Solution Applied to Diagnostic Test Management, Journal of Biogeneric Science and Research 7: 1-9.

51. Hendler, N, Cashen, A, Hendler, S, Brigham, C, Osborne, P, LeRoy, P., Graybill, T, Catlett, L., Gronblad, M. A Multi-Center Study for Validating The Complaint of Chronic Back, Neck and Limb Pain Using "The Mensana Clinic Pain Validity Test." Forensic Examiner 14: 41-9.

52. Hendler N (2017) An Internet based Questionnaire to Identify Drug Seeking Behavior in a Patient in the ED and Office. J Anesth Crit Care Open Access 8: 00306.

53. Z Xu (2020) Classification, identification, and growth stage estimation of microalgae based on transmission hyperspectral microscopic imaging and machine learning, Optics Express 28: 30686-30700.

54. Z Xu (2020) Light-sheet microscopy for surface topography measurements and quantitative analysis, Sensors 20: 2842.

55. C Jiao (2021) Machine learning classification of origins and varieties of Tetrastigma hemsleyanum using a dual-mode microscopic hyperspectral imager, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 261: 120054.

56. T Wang (2022) Smartphone imaging spectrometer for egg/meat freshness monitoring, Analytical Methods 14: 508.

57. Z Xu (2020) Multi-mode microscopic hyperspectral imager for the sensing of biological samples, Applied Sciences 10: 4876.

58. L Luo (2020) A parameter-free calibration process for a Scheimpflug LIDAR for volumetric profiling, Progress in Electromagnetics Research 169: 117-27.

Appendix A-Partial List of Medical-Legal Cases Where The Pain Validity Test Has Been Accepted as Accurate Methodology

Attorney name	Court	Case Number	City and State
C. Richard Cranwell	Circuit Court of the	Case No. CL	Roanoak, VA
Cranwell and Moore	City of Roanoke	95000860-00	
Thomas Vesper	*Circuit Court of*	*Case No. M-2003*	Atlantic City,
	Atlantic Co.	*45395 AC*	New Jersey
Alan Legum	District Court of	Case No. CV PI	State of Idaho
	Fourth Judicial District	0100197D,Idaho	
Alan Legum	Circuit court for	Case No: C-2001-	Glen Burnie, MD.
	Anne Arundel County	73911 OT	
Charles Sickles	Fairfax County Circuit	Case # 194133	Fairfax, VA.
	Court		
Scott Fruge	Judicial District Court	docket # 470,239	Parish of East
	Baton Rouge, LA	Div. N	Baton Rouge
Stephen Mastaitis	Workers compensation	Claim NYS WCB	New York State
	hearing New York State	#6931-1354:	
	Administrative Law Judge	D/A 8/30/93.	
Mark Cavanaugh	Philadelphia, PA. Court	docket Dec. Term	Philadelphia, Pa.
	of Common Pleas	2001 No. 000532	
Kevin McCarthy	*Circuit Court for Anne*	*Case No.*	Glen Burnie, Md.
	Arundel County	*C2002-83654.*	
Marc Neff	Denver, Co.	Case No. DC	Denver, Colorado
		430-345-0232	
Dusan Bratic	Dauphin Co. PA	62-S-2001	Dauphin Co., PA
Scudder Stevens	Bureau of Workers	Workers Comp	Harrisburg, PA
	Compensation, PA	Claim 1285129-01	
Alan Lancaster	Circuit Court of	Civil Action: No:	Jackson, MS
	Lowndes County, MS	96-020-CV1	
Matthew Williams	Philadelpia, PA	Work Comp	Philadelphia, PA
Martin Banks Pond		Claim 214323-03	
Lehocky & Wilson			
Alan Legum	US District Court	Civil action No.	Columbia, MD
	Columbia, MD	03-1931 (PLF)	