

Artificial Intelligence: Bayesian versus Heuristic Method for Diagnostic Decision Support

Peter L. Elkin¹ Daniel R. Schlegel¹ Michael Anderson² Jordan Komm^{1,2} Gregoire Ficheur¹
Leslie Bisson²

¹Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, New York, United States

²Department of Orthopedics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, New York, United States

Address for correspondence Peter L. Elkin, MD, Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY 10128, United States (e-mail: ontolimatics@gmail.com).

Appl Clin Inform 2018;9:432–439.

Abstract

Keywords

- ▶ clinical decision support
- ▶ process management tools
- ▶ specific types
- ▶ clinical information systems
- ▶ computer-assisted diagnosis
- ▶ new diagnosis
- ▶ disease management
- ▶ clinical informatics
- ▶ professional training
- ▶ education
- ▶ decision support algorithm
- ▶ computer-assisted decision making
- ▶ knowledge modeling and representation
- ▶ knowledge management
- ▶ clinical information

Evoking strength is one of the important contributions of the field of Biomedical Informatics to the discipline of Artificial Intelligence. The University at Buffalo's Orthopedics Department wanted to create an expert system to assist patients with self-diagnosis of knee problems and to thereby facilitate referral to the right orthopedic subspecialist. They had two independent sports medicine physicians review 469 cases. A board-certified orthopedic sports medicine practitioner, L.B., reviewed any disagreements until a gold standard diagnosis was reached. For each case, the patients entered 126 potential answers to 26 questions into a Web interface. These were modeled by an expert sports medicine physician and the answers were reviewed by L.B. For each finding, the clinician specified the sensitivity (term frequency) and both specificity (Sp) and the heuristic evoking strength (ES). Heuristics are methods of reasoning with only partial evidence. An expert system was constructed that reflected the posttest odds of disease-ranked list for each case. We compare the accuracy of using Sp to that of using ES (original model, $p < 0.0008$; term importance * disease importance [DtimesTI] model, $p < 0.0001$: Wilcoxon ranked sum test). For patient referral assignment, Sp in the DtimesTI model was superior to the use of ES. By the fifth diagnosis, the advantage was lost and so there is no difference between the techniques when serving as a reminder system.

received
February 6, 2018
accepted after revision
April 17, 2018

Copyright © 2018 Schattauer

DOI <https://doi.org/10.1055/s-0038-1656547>.
ISSN 1869-0327.

Background and Significance

Expert systems use heuristics which are methods of reasoning with only partial evidence.¹ This requires experts in the field to encode knowledge about how they reason into the system. This has traditionally been done by either specifying weightings such as evoking strength (ES) which is defined as *given the manifestation (finding, test result, etc.) how strongly should you think of the diagnosis*. The other method used frequently is feature selection in machine-learning algorithms.² Bayesian approaches use conditional probabilities often in the form of sensitivity and specificity (Sp) to define and combine probabilities of a diagnosis being true. For many years, leaders in medicine have felt that there was something special about the heuristics used to create a differential diagnosis.³ This article looks to answer the question regarding which method performs better in a data set of patients who present with knee pain and their diagnoses established by the consensus of two sports medicine trained orthopedic surgeons.

Ledley and Lusted, in 1959, predicted that computers would help doctors in the diagnostic process.⁴ The first abdominal pain diagnosis program was designed by Tim de Dombal at the University of Leeds using a pure Bayesian approach. The system classified cases as either appendicitis, diverticulitis, perforated ulcers, cholecystitis small-bowel obstruction, pancreatitis, and nonspecific abdominal pain using a training set of thousands of patient data cases.⁵ At Stanford University, Shortliffe et al developed MYCIN, which provided consultation regarding the empiric antibiotic management of infectious diseases.⁶ MYCIN used the If/Then logic production rules consisting of conditional statements (e.g., If the location of the infection is the lung and the patient is of an age range 55–65 Then the likely organisms causing the infection are Streptococcus pneumoniae).⁷ This methodology is one type of artificial intelligence (AI), which includes machine-learning methods such as random forest, deep learning, and Bayesian nets.¹

Clinical judgment, biostatistics, epidemiology, and evidence-based medicine as described by John Paul, Alvan Feinstein, and David Sackett were work that influenced the field. They made a categorization of the types of diagnosis and of the consecutive use of them. This classification of diagnoses classes, according to the magnitudes of the sensitivity and Sp, is as follows: (1) discovery (or detection = high sensitivity), (2) exclusion (or differential = high sensitivity and Sp), and (3) confirmation (or positive = highest Sp) added to our understanding of scientific rigor.

While at the University of Utah, Warner and colleagues developed the Health Evaluation Through Logical Processing (HELP) system which was integrated into a hospital information system and provided direct clinical decision support.^{3,8} The Arden Syntax was used to specify the rules employed in the HELP system.⁹ These medical logic modules were contained rule sets which can be reused.¹⁰ The Iliad system was designed by Warner and employed a pure Bayesian approach to decision support.

Miller et al designed a diagnostic decision support system named the Quick Medical Reference (QMR).¹¹ QMR was used by a consult service at the University of Pittsburg which con-

tended that a physician with a computerized clinical diagnostic decision support system was more accurate at making diagnoses than the physician alone.¹² QMR is a rule-based system that maps manifestations to diagnoses using heuristics.¹³

DXplain, a diagnostic clinical decision support system, was developed by Barnett et al in the 1980s. He led the Laboratory of Computer Science at Massachusetts General Hospital in Boston.^{14,15} DXplain serves as a diagnostic decision support tool and incorporates into its knowledge base clinical probabilities for approximately 6,000 clinical manifestations (history, pulmonary embolism findings, laboratory data, X-ray data, and elements of the past medical history) as connected to each known diagnosis (~2,300) and uses that information to generate a differential diagnosis^{16,17} associated with the patient's manifestations. DXplain makes use of an interactive Web-based human-computer interface to collect clinical information and makes use of a modified form of Bayesian logic to produce a ranked list of diagnoses that might be associated with the clinical manifestations. DXplain uses this same knowledge base and logic to suggest findings that, if present, would differentiate between the various diagnoses in the differential diagnosis list provided. The system provides other information such as disease descriptions and references for each of the diagnoses in its knowledgebase.¹⁸

DXplain has been used by thousands of practicing physicians and medical students. Sixteen years ago, DXplain was made available over the Internet to hospitals, medical schools, and health care organizations.¹⁹ QMR and DXplain both use the concept of ES in their calculations. Iliad by the University of Utah uses Sp in its calculations.²⁰ DXplain uses the heuristic ES in its modeling. A heuristic is a method of reasoning with partial evidence. This study is the first head-to-head trial of ES versus Sp.

Expert systems have been shown to improve pain management after total knee arthroplasty (TKA) and total hip arthroplasty.²¹ Farion et al compared a Bayesian predication model, clinical score, and physicians in the diagnosis of pediatric asthma in the emergency room.²²

Zhou et al developed a machine-learning algorithm for disease phenotypes in primary care electronic health records and was tested in identifying rheumatoid arthritis.² Qureshi et al reported a hierarchical machine-learning technique for distinguishing types of attention deficit disorder using structural magnetic resonance imaging (MRI) data.²³ Ye et al used support vector machines to predict cancer type in full-text articles.²⁴

ES is defined as how much you should think of a disorder given a finding or manifestation. This embodies a sense of how important it is not to miss this disorder. Sp is defined as how often do you see the disorder given the finding (i.e., true negative rate [TN]/[TN + false positive rate (FP)]).

Methods

The University at Buffalo's Orthopedics Department wanted to create an expert system to assist patients with self-diagnosis of problems in their knees and to thereby facilitate referral to the right orthopedic subspecialist. Then two sports medicine physicians independently reviewed 469 patient cases (see **Table 1**). The chair of the department,

Table 1 Demographics of the patient population

| | |
|-----------------------------------|-------------|
| Age range | 1–84 |
| Age median | 47 |
| Age mean | 44 |
| % Female | 50 |
| BMI range | 16.28–97.64 |
| BMI mean | 28.9 |
| BMI median | 27.4 |
| Percent due to injury | 24.5 |
| Percent due to sports injury | 22.8 |
| Percent with pain in wrists/hands | 25.3 |
| Percent who had knee surgery | 33.4 |

Abbreviation: BMI, body mass index.

L.B., reviewed any disagreements and the two experts had a discussion of the case until a gold standard diagnosis was reached. The data came from a Web site where for each case the patients answered 26 questions with a total number of potential responses of 126. Each possible disorder was modeled into an expert system by an expert sports medicine physician and the answers were reviewed by a second clinician L.B. For each finding associated with a disorder (see [Table 2](#)), the clinician specified the sensitivity (equivalent to the term frequency [TF]) and both the Sp and ES. Where the two sports medicine experts disagreed, they worked together and discussed the issue until they developed a final consensus value.

The TF was defined as *given the diagnosis how often do you see the finding*. It is the same as sensitivity (true positive rate [TP]/[TP + false negative rate (FN)]). This fact is often present in the biomedical literature.

The Sp was defined as *given the finding how often do you have the diagnosis*. This is represented as the TN/(TN + FP) rates.

Table 2 List of disorders considered in this evaluation and their prevalence in the orthopedic practice

| Disease name | Prevalence |
|--|------------|
| ACL tear | 0.11727 |
| Patellar chondromalacia/ patelofemoral syndrome | 0.25159 |
| Patellar arthritis | 0.14712 |
| Meniscus tear | 0.40085 |
| Patellar tendinitis | 0.02132 |
| Patellar instability | 0.02985 |
| Patellar contusion/ saphenous nerve contusion | 0.05756 |
| MCL tear | 0.05330 |
| Popliteal cyst | 0.02985 |
| Osteoarthritis w/wo exacerbation | 0.43496 |

Abbreviations: ACL, anterior cruciate ligament; MCL, medial collateral ligament.

Table 3 Patient data entry regarding their knee pain and the final diagnosis selected by the sports medicine orthopedic surgeons

| Question | Answer |
|--|--|
| Sex | Male |
| Age | 48 |
| BMI | 29.0 |
| Which knee hurts? Where? | Left only; Lower front |
| Is the current issue a sports injury? | Yes; Basketball |
| Does the pain worsen when you perform specific activities? | Yes; When running, walking, and using stairs |
| How long have you had the pain? | Weeks |
| Have you had prior surgery? | No |
| Have you had a previous dislocation? | No |
| Do you have pain in your hands or wrists? | No |
| Do you have swelling in your knee? | No |
| Is this due to a specific injury? | No |
| Have you previously had an injection? | No |
| Final diagnoses | Meniscus tear |

Abbreviation: BMI, body mass index.

The ES was defined as *given the finding how strongly should you consider the diagnosis*. Instead of a formula here, we leave it up to the clinical judgment of the individual clinician to determine this value.

Four hundred and sixty-nine patients with knee pain entered data into a Web site with 26 questions and together 126 potential combinations of questions and answers (see [Table 3](#)). This served as the primary data which was used to develop the differential diagnoses.

The expert system written in Java, performed the following calculation to determine the weight of each diagnosis for each case:

1. For each finding in the case, obtain the weights for TF, Sp, and ES from the expert-derived database described above.
2. Sum the weights based on each of the models used in the experiment (positive likelihood ratio multiplied by the term importance (TI) and the positive likelihood ratio multiplied by the TI times the disease importance (DI) (DI*times*TI).
3. Multiply this sum times the disease prevalence (see [Table 1](#)) to obtain the posttest odds.
4. Use the relative posttest odds to order the differential diagnosis list from highest to lowest score.

An expert system was constructed that reflected the posttest odds of disease-ranked list for each case. This was built by having two sports medicine board-certified orthopedic surgeons (L.B. and M.A.) specify the following attributes. For each disorder, they specified the DI on a scale of 1 to 5 with 5 being

the most important diagnoses not to miss. For each term the same orthopedists provided a TI which signified on a 1 to 5 scale how contributory was the finding or manifestation. An example might be prior history of ipsilateral knee surgery was highly contributory where have you had an injection might be less contributory. Then, for each disorder the same orthopedists specified for each term the TF which is defined as given the disorder how often to you see the manifestation present. The ES which is given the manifestation how strongly should you consider the disease. The same orthopedists also provided a Sp for the manifestation for that disorder. This was done for all disorders in the knowledgebase. The orthopedic experts used the biomedical literature, the clinical guidelines in their field, and their training and experience when developing the knowledgebase. As the primary outcome of this study, we compare the accuracy of using Sp to that of using ES in providing the correct diagnosis. We also compare the accuracy at each rank order position on the weighted differential diagnosis list from first position to 10th. For example: what is the chance that the correct diagnosis will be in the top five diagnoses?

The results were analyzed by using the positive likelihood ratio with Sp and with ES substituted for the Sp. In each case, we generated a ranked list of diagnoses for the case and determined where on the list the gold standard diagnosis fell. We graph them as cumulative results so the second rank was the chance that the gold standard diagnosis was either first or second on the ranked list. This generated two graphs. The graphs were compared using the Wilcoxon signed rank sum test.

We started with an original model, which is analogous to the posttest odds of disease. In DXplain, for example, the formula used $TI * TF$, analogous to sensitivity. We normalize this by the positive likelihood ratio ($Sensitivity/(1-Specificity)$).

We ran an original model:

$$P = prevalence * \sum_{i=1}^{term} termImportance_i * \frac{termFrequency_i}{1 - evokingStrength_i}$$

And we added in to the equation the DltimesTI model:

$$P = prevalence * \sum_{i=1}^{term} diseaseImportance * termImportance_i * \frac{termFrequency_i}{1 - evokingStrength_i}$$

The ES and Sp were substituted for each other in their respective models.

Results

We compared results from multiple models to determine what would provide the greatest predictive power. This included the positive likelihood ratio ($TF/(1-(Sp/Es))$), and the likelihood ratio multiplied by the TI and the DI.

In **Table 4**, we present the main results. Here, the rank is the rank on the differential diagnosis list and the number is the number of cases that, for example, the top diagnosis was the correct diagnosis by the orthopedic experts. The second is the number where the correct diagnosis either first or second on the list and so on. The percentages show the percent of the cases that fell into, for example, either the first or second on the list was the correct diagnosis. The various methods can be compared from these data.

Fig. 1 shows the comparison of the Bayesian (Sp) versus heuristic (ES) for the DltimesTI model approaches graphically. At the x-axis is the rank in the differential where the diagnosis was found (e.g., top diagnosis, top two, top three, etc.). At the y-axis is the number of cases where were found in that group (e.g., top diagnosis, top two, top three, etc.). **Fig. 2** is the same graph for the original (TI-only) model.

Table 5 shows the method for calculating the mean rank using each of the two methods (Bayesian vs. heuristic) using the DltimesTI model. **Table 4** also shows the number of cases where there was correct diagnosis at each rank level (again the top diagnosis, top two, top three, etc.).

Fig. 3 shows the number of cases at each rank (e.g., top diagnosis, top two, top three, etc.) in a tabular rather than graphical form.

We used the Wilcoxon signed rank test with two-sided continuity correction to compare the original ES versus SP and

Table 4 Expert system results: % chance of having a certain rank level

| Rank | ES-Gold standard original | ES-original % | ES-Gold standard DltimesTI | ES-DltimesTI % | Sp-Gold standard original | Sp-Original % | Sp-Gold standard DltimesTI | Sp-DltimesTI % |
|------|---------------------------|---------------|----------------------------|----------------|---------------------------|---------------|----------------------------|----------------|
| 1 | 203 | 43.28358 | 191 | 40.724947 | 203 | 43.28358 | 224 | 47.761194 |
| 2 | 300 | 63.96588 | 277 | 59.061834 | 320 | 68.23028 | 320 | 68.230277 |
| 3 | 355 | 75.69296 | 345 | 73.560768 | 393 | 83.79531 | 404 | 86.140725 |
| 4 | 416 | 88.69936 | 419 | 89.339019 | 424 | 90.40512 | 434 | 92.537313 |
| 5 | 441 | 94.02985 | 441 | 94.029851 | 437 | 93.17697 | 437 | 93.176972 |
| 6 | 448 | 95.52239 | 448 | 95.522388 | 451 | 96.16205 | 448 | 95.522388 |
| 7 | 460 | 98.08102 | 457 | 97.441365 | 456 | 97.22814 | 455 | 97.014925 |
| 8 | 465 | 99.14712 | 462 | 98.507463 | 462 | 98.50746 | 462 | 98.507463 |
| 9 | 469 | 100 | 467 | 99.573561 | 464 | 98.9339 | 464 | 98.933902 |
| 10 | 469 | 100 | 469 | 100 | 469 | 100 | 469 | 100 |

Abbreviations: ES, evoking strength; DltimesTI, term importance * disease importance; Sp, specificity.

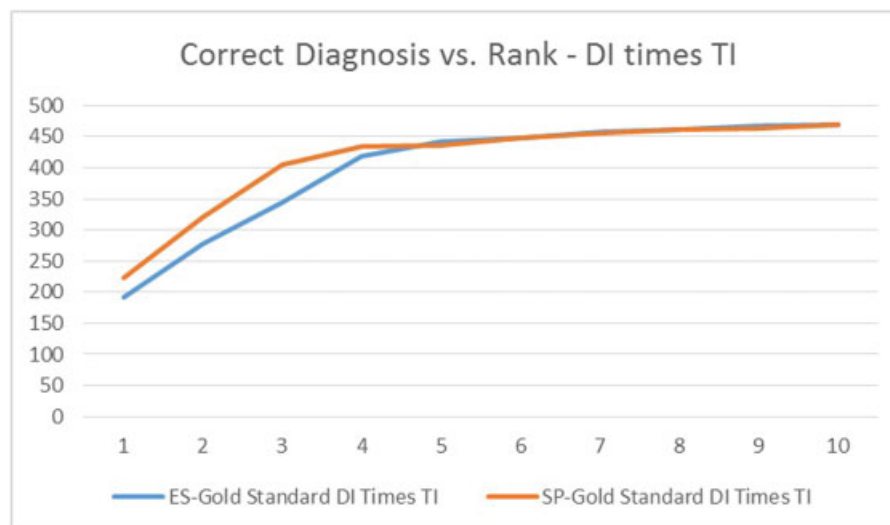


Fig. 1 Difference between the specificity and evoking strength in artificial intelligence models.

the DItimesTI ES versus Sp. In the case comparing the original ES versus Sp, the p -value was < 0.0007 . In the case of DItimesTI ES versus Sp, the p -value was < 0.001 using the two-sided Wilcoxon signed rank test with continuity correction (see [Fig. 4](#)).

The results show that the Sp was statistically significantly better at predicting the knee disorder diagnosis in this clinical trial. The absolute difference in means was small at 0.1 for the original formula and for 0.3 for the DItimesTI model. The DItimesTI model took into account how important the clinician thought that diagnosis was not to miss. This may have accounted for some of the information contained in the composite concept of ES. Overall, the results show the power of using DI in AI models for diagnostic clinical decision support.

Discussion

This study shows that the use of Sp statistically outperformed ES in this expert system. As we added DI to the formula, we picked up information that we believe is part of the uniqueness of the ES heuristic. In doing so, we see an improvement in the two scores. The fact that this effect was greater with the DItimesTI method suggests that DI makes up

some of the performance gains with the ES heuristic. Upon questioning, we found that the Sp values were easier for the clinicians to come up with as compared with ES.

That said, all the expert systems converge to similar accuracy levels after about four or five diagnoses down on the differential diagnosis list. Meaning that the chance that the diagnosis is the top five diagnoses is the same for each approach. The absolute difference is between 4 and 7% in the ability to get the first diagnosis right and increases to 8 to 12.6% difference at a three diagnosis list. This seems both clinically and statistically significant. However, by a five diagnosis list any advantage of Sp over ES is lost. So, as a reminder system the two approaches are equivalent. As a classifier of patients into an exact diagnosis group, Sp has an advantage over ES. These systems were designed to be reminder systems. For that function, either method is equivalent.

The performance of diagnostic clinical decision support systems is integrally related to the ability of experts to model the information. It is possible that the increased comfort with the idea of Sp by clinicians led to the increase in performance seen in this study.

Future research should look at the ability of machine-learning algorithms to predict the correct diagnosis from this type of training data.

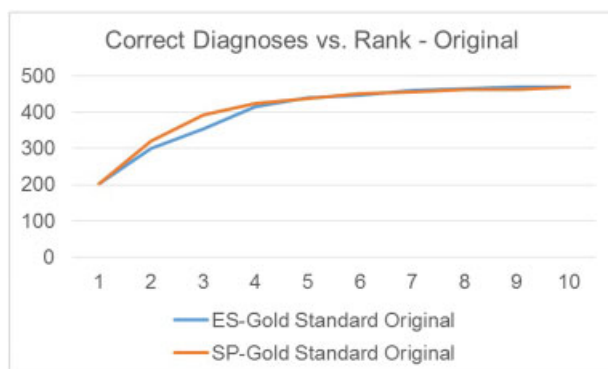


Fig. 2 The original model results.

Conclusion

Bayes theorem still has a lot to teach us about patient diagnosis. In this study, using patient-entered data in an expert system for the determination of the cause of a patient's knee pain from patient-provided information utilizing Sp outperformed the one based on ES for making the correct diagnosis. The significance of this performance difference depends directly on the use case for the expert system. In this case where routing the patient more frequently to the right clinician Sp as used in a positive likelihood ratio was found to provide higher accuracy in making the exact diagnosis and in lists up to three diagnoses long. As

Table 5 Mean rank of the true diagnosis on the computer-generated differential diagnosis list

| DITimesTI_Sp_Rank | | | DITimesTI_ES_Rank | | | |
|--------------------|--------------|------------|--------------------|--------------|------------|--------------|
| | Value | | | Value | | |
| Number | 469 | | Number | 469 | | |
| Mean | 2.215 | | Mean | 2.522 | | |
| Standard deviation | 1.757 | | Standard deviation | 1.799 | | |
| 95% CI | 2.056, 2.375 | | 95% CI | 2.359, 2.686 | | |
| Minimum | 1 | | Minimum | 1 | | |
| Quartile | 1 | | Quartile | 1 | | |
| Median | 2 | | Median | 2 | | |
| Quartile 3 | 3 | | Quartile 3 | 4 | | |
| Maximum | 10 | | Maximum | 10 | | |
| Rank | Number | Percentage | Rank | Number | Percentage | 95% CI |
| 1 | 224 | 48 | 1 | 191 | 41 | 36.27, 45.34 |
| 2 | 96 | 20 | 2 | 86 | 18 | 15, 22.2 |
| 3 | 87 | 19 | 3 | 68 | 14 | 11.5, 18.09 |
| 4 | 27 | 6 | 4 | 74 | 16 | 12.66, 19:47 |
| 5 | 3 | 1 | 5 | 22 | 5 | 3.03, 7:13 |
| 6 | 11 | 2 | 6 | 7 | 1 | 0.66, 3.19 |
| 7 | 7 | 1 | 7 | 9 | 2 | 0.94, 3.74 |
| 8 | 7 | 1 | 8 | 5 | 1 | 0.39, 2.62 |
| 9 | 2 | 0 | 9 | 5 | 1 | 0.39, 2.62 |
| 10 | 3 | 1 | 10 | 2 | 0 | 0.05, 1.53 |
| Total | 469 | 100 | Total | 469 | 100 | |

Abbreviations: CI, confidence interval; ES, evoking strength; DITimesTI, term importance * disease importance; Sp, specificity.

a reminder system, neither approach truly outperformed the other as by the fifth diagnosis in the differential diagnosis list the accuracies were not statistically significantly different.

We believe that knowledge bases such as this orthopedic knee pain database are useful sets of assertional knowledge that can drive medical decision making. This type of knowledge when indexed by ontologies and used in a consistent

fashion has the capacity to improve clinical care through diagnostic and therapeutic clinical decision support.

Future research will seek to determine how machine-learning algorithms such as c-trees or random forest compare with this Bayesian expert knowledge base approach. Other avenues of future research will be to implement this for other disorders and for other specialties. Additionally,

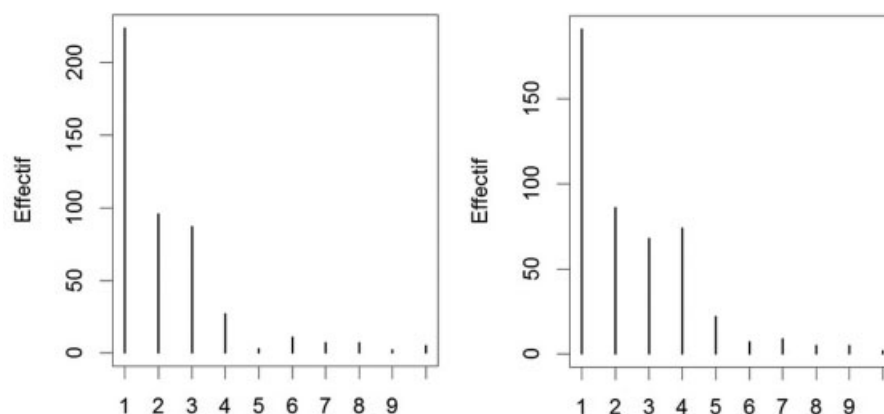


Fig. 3 Graphical view of the distribution of term importance * disease importance (DITimesTI) with specificity (Sp) on the left and evoking strength (ES) on the right by number of correct answers by rank order in the differential diagnosis list. The two-sided Wilcoxon signed rank test with continuity correction showed a $p < 0.001$ when comparing the original formula with ES and the original formula with Sp. We also used the same test to compare the DITimesTI models with ES and with Sp and the $p < 0.001$ as well.

db\$DITimesTI_ES_Rank and db\$DITimesTI_SP_Rank

Paired t-test
two.sided
Mean of the difference
0.307036247334755
p.value
5.69331331894901e-11

Wilcoxon signed rank test with continuity correction
two.sided
Mean of the difference

p.value
1.25443705152435e-10

Fig. 4 Bivariate analysis comparing the original formula specificity (Sp) versus evoking strength (ES) (left) and the term importance * disease importance (DITimesTI) formula Sp versus ES (right).

one might look at electronic health record data as the input source to the expert system rather than patient-provided information.

Systematized care of patients whether health care or self-care requires this level of rigor so that we can effectively extend our health care delivery system toward comprehensive care for all patients each and every day.

Clinical Relevance Statement

The relevance to Clinical Informatics stems from the articles guidance regarding how best to build clinical decision support systems that provide real time point of care clinical decision support to clinicians in support of direct patient care.

Multiple Choice Questions

- In crafting a clinical prediction rule, the posttest odds of disease is:
 - Equal to the pretest odds \times the positive likelihood ratio given a positive test result.
 - Equal to the pretest odds \times the positive likelihood ratio given a negative test result.
 - Equal to the pretest odds adjusted by the negative likelihood ratio given a positive test result.
 - Equal to the pretest odds adjusted by the negative likelihood ratio given a negative test result.

Correct Answer: The correct answer is option a.

- The prevalence of disease is:
 - The number of new cases each year.
 - The number of ICD-10 codes of the disease in your data warehouse.
 - The percent of people in a given time frame that have the disorder.
 - The number of SNOMED CT codes of the disease in your data warehouse.

Correct Answer: The correct answer is option c.

Protection of Human and Animal Subjects

The study obtained IRB approval # 1690612E.

Funding

Funding for this project was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH) under award number UL1TR001412. The content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH. This work has been supported in part by a training grant from the National Library of Medicine LM1012495.

Conflict of Interest

None.

References

- Shortliffe EH, Buchanan BG. Artificial intelligence. *N Engl J Med* 1980;302(26):1482
- Zhou SM, Fernandez-Gutierrez F, Kennedy J, et al; UK Biobank Follow-up and Outcomes Group. Defining disease phenotypes in primary care electronic health records by a machine learning approach: a case study in identifying rheumatoid arthritis. *PLoS One* 2016;11(05):e0154515
- Warner HR, Olmsted CM, Rutherford BD. HELP—a program for medical decision-making. *Comput Biomed Res* 1972;5(01):65–74
- Ledley RS, Lusted LB. The role of computers in medical diagnosis. *Med Dok* 1961;5:70–78
- de Dombal T, Clamp S, Margulies M, Chan M. Computer training for doctors and students. *BMJ* 1994;309(6963):1234–1235
- Shortliffe EH, Axline SG, Buchanan BG, Merigan TC, Cohen SN. An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Comput Biomed Res* 1973;6(06):544–560
- Shortliffe EH, Davis R, Axline SG, Buchanan BG, Green CC, Cohen SN. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system. *Comput Biomed Res* 1975;8(04):303–320
- Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP system. *J Med Syst* 1983;7(02):87–102
- Hripcsak G, Wigertz OB, Kahn MG, Clayton PD, Pryor TA. ASTM E31.15 on health knowledge representation: the Arden Syntax. *Stud Health Technol Inform* 1993;6:105–112
- Pryor TA. The use of medical logic modules at LDS hospital. *Comput Biol Med* 1994;24(05):391–395
- Miller R, Masarie FE, Myers JD. Quick medical reference (QMR) for diagnostic assistance. *MD Comput* 1986;3(05):34–48
- Bankowitz RA, McNeil MA, Challinor SM, Parker RC, Kapoor WN, Miller RA. A computer-assisted medical diagnostic consultation service. Implementation and prospective evaluation of a prototype. *Ann Intern Med* 1989;110(10):824–832
- Giuse DA, Giuse NB, Miller RA. A tool for the computer-assisted creation of QMR medical knowledge base disease profiles. *Proc Annu Symp Comput Appl Med Care* 1991:978–979
- Barnett GO, Hoffer EP, Packer MS, et al. DXplain - important issues in the development of a computer based decision support system. *Proc Annu Symp Comput Appl Med Care* 1990:1013
- Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain. An evolving diagnostic decision-support system. *JAMA* 1987;258(01):67–74
- Packer MS, Hoffer EP, Barnett GO, et al. Evolution of DXplain: a decision support system. *Proc Annu Symp Comput Appl Med Care* 1989:949–951
- Elkin PL, Barnett GO, Famiglietti KT, Kim RJ. Closing the loop on diagnostic decision support systems. *Proc Annu Symp Comput Appl Med Care* 1990:589–593

- 18 Elkin PL, McLatchey J, Packer M, et al. Automated batch searching of MEDLINE for DXplain. *Proc Annu Symp Comput Appl Med Care* 1989;436-440
- 19 Barnett GO, Famiglietti KT, Kim RJ, Hoffer EP, Feldman MJ. DXplain on the Internet. *Proc AMIA Symp* 1998;607-611
- 20 Bouhaddou O, Warner H. An interactive patient information and education system (Medical HouseCall) based on a physician expert system (Iliad). *Medinfo* 1995;8(Pt 2):1181-1185
- 21 Urban MK, Chiu T, Wolfe S, Magid S. Electronic ordering system improves postoperative pain management after total knee or hip arthroplasty. *Appl Clin Inform* 2015;6(03):591-599
- 22 Farion KJ, Wilk S, Michalowski W, O'Sullivan D, Sayyad-Shirabad J. Comparing predictions made by a prediction model, clinical score, and physicians: pediatric asthma exacerbations in the emergency department. *Appl Clin Inform* 2013;4(03):376-391
- 23 Qureshi MN, Min B, Jo HJ, Lee B. Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: structural MRI study. *PLoS One* 2016;11(08):e0160697
- 24 Ye Z, Tafti AP, He KY, Wang K, He MM. SparkText: biomedical text mining on big data framework. *PLoS One* 2016;11(09):e0162721