



Clinical Scenario



Samantha Darren - March 3, 2011

History of Present Illness

Patient has red eye with pain, inflammation, blurred vision, floating spots and sensitivity to light.

Symptoms

Question

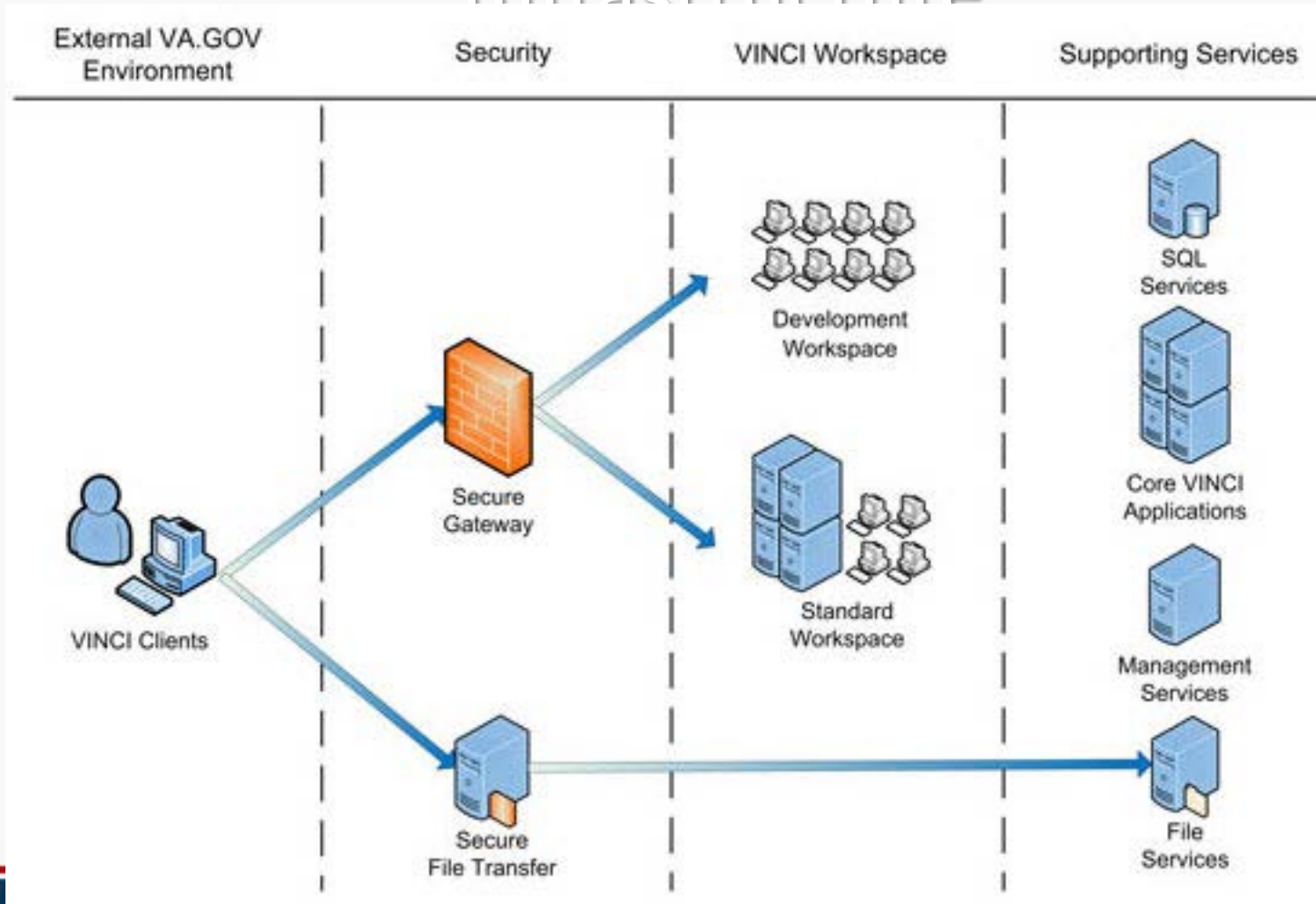
What condition has red eye, pain, inflammation, blurred vision, floating spots and sensitivity to light?





Clinical Challenge

VINCI – VA Informatics and Computing Infrastructure



Clinical Scenario



Samantha Darren - March 3, 2011

History of Present Illness

Patient has red eye with pain, inflammation, blurred vision, floating spots and sensitivity to light.

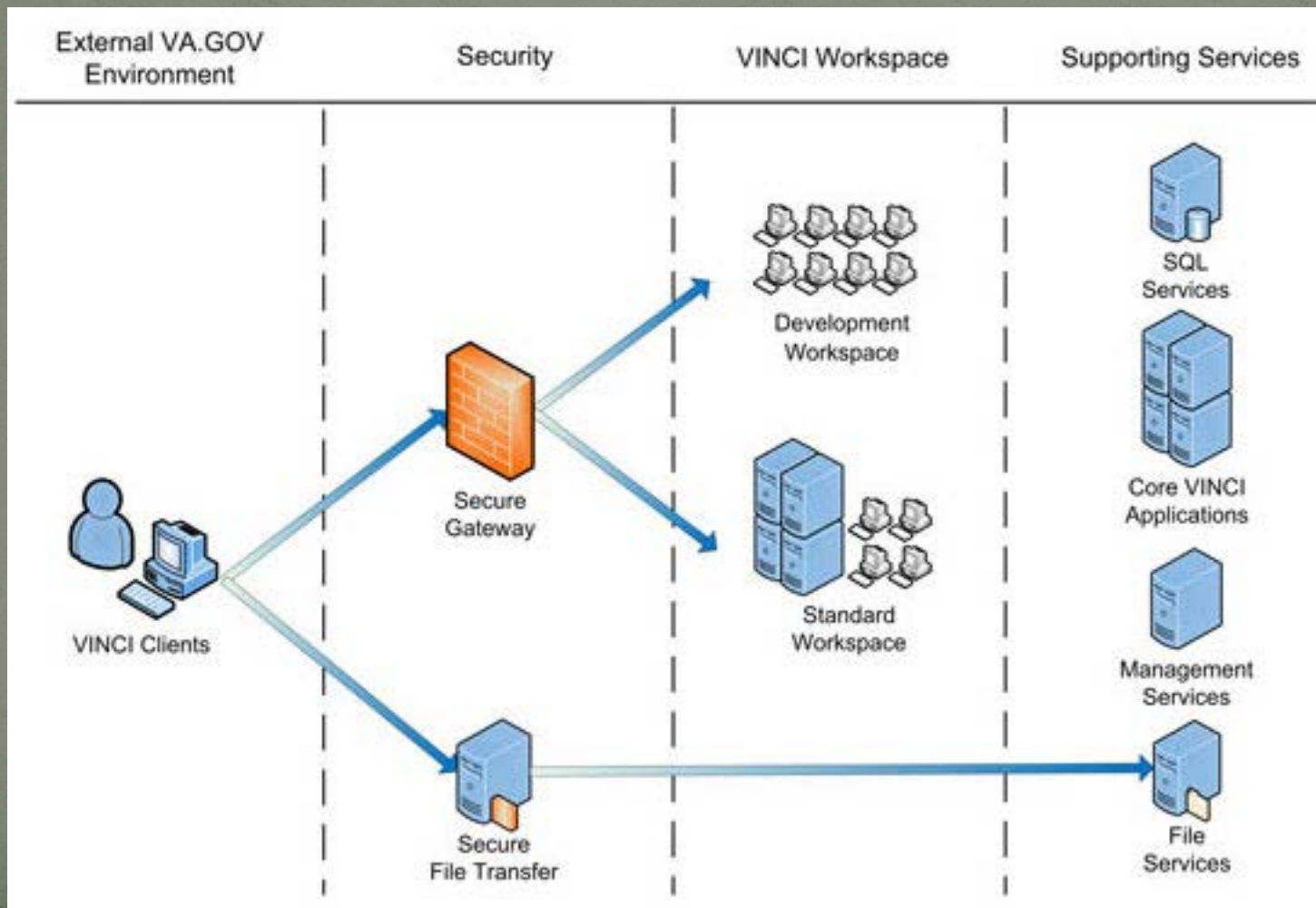
Symptoms

Question

What condition has red eye, pain, inflammation, blurred vision, floating spots and sensitivity to light?

Clinical Challenge

VINCI – VA Informatics and Computing Infrastructure



IBM and the Jeopardy Challenge

For copy of the video, please contact the
speaker:

Eliot Siegel, M. D., Director, Baltimore
Veterans Affairs Medical Center Radiology

Year of Artificial Intelligence in Med



- 2011 will likely be remembered as the year of the re-emergence of artificial intelligence in medicine with Watson and of course, Siri, arguably the best feature of the new iPhone 4S and 5
- 2011 may well be the year that AI finally gets real traction in the medical informatics community and in medicine in general including the lay population
- Biggest contribution of Dr. Watson software in addition to Deep Q/A may be excitement to overcome inertia of the past

IBM and Jeopardy: A New Era?

- The Jeopardy match between the two best human players of all time and the IBM Deep Q/A software, “Watson” captured the spotlight and stimulated the imagination of the entire world
- The subsequent announcement of IBM’s involvement in the creation of “Dr. Watson” has created an incredible interest in the healthcare community about the potential breakthrough technology as well as the potential pitfalls of the use of “artificial intelligence” in medicine.

Why Jeopardy?

For copy of the video, please contact the
speaker:

Eliot Siegel, M. D., Director, Baltimore
Veterans Affairs Medical Center Radiology

Dr. Watson Overview and History



- Initially had opportunity to visit IBM team about a year and a half ago
- Engaged Jeopardy team and discussed the potential for medical applications as next steps after Jeopardy Challenge
- Began initial research with IBM approximately one year ago
- Current grant with IBM for initial exploratory work with physician helping team to understand the medical domain and challenges
- Worked together on deeper understanding of the medical domain using multiple resources

Introduction



- Deep Q/A is unique and exciting because it represents a fundamentally new approach that creates tools to rapidly mine a dynamic and non-predefined database
- Represents a potential fundamental change in opportunities for Artificial Intelligence applications in medicine
- But in some ways Watson is a “special needs” student
- How does one train a system that is so remarkable at Jeopardy! questions and apply to medicine?

- Watson can process 500 gigabytes, the equivalent of a million books, per second
- Hardware cost has been estimated at about \$3 million
- 80 TeraFLOPs , 49th in the Top 50 Supercomputers list
- Content was stored in Watson's RAM for the game because data stored on hard drives too slow to process

Deep Q/A

- Massively parallel, component based pipeline architecture
- Uses extensible set of structured and unstructured content sources
- Uses broad range of pluggable search and scoring components

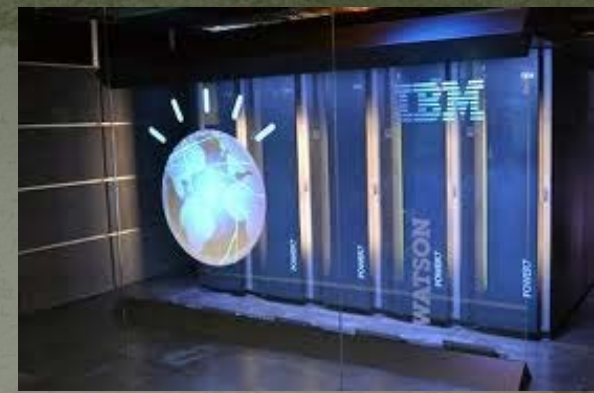
Deep Q/A

- These allow integration of many different analytic techniques
- Input from scorers is weighed and combined using machine learning to generate a set of ranked candidate answers and associated confidence values
- Each answer is linked to its supporting evidence

Deep Q/A

- Does not map question to database of answers
- Represents software architecture for analyzing natural language content in both questions and knowledge sources
- Discovers and evaluates potential answers and gathers and scores evidence for those answers using unstructured sources such as natural language documents and structured sources such as relational and knowledge databases

Hardware

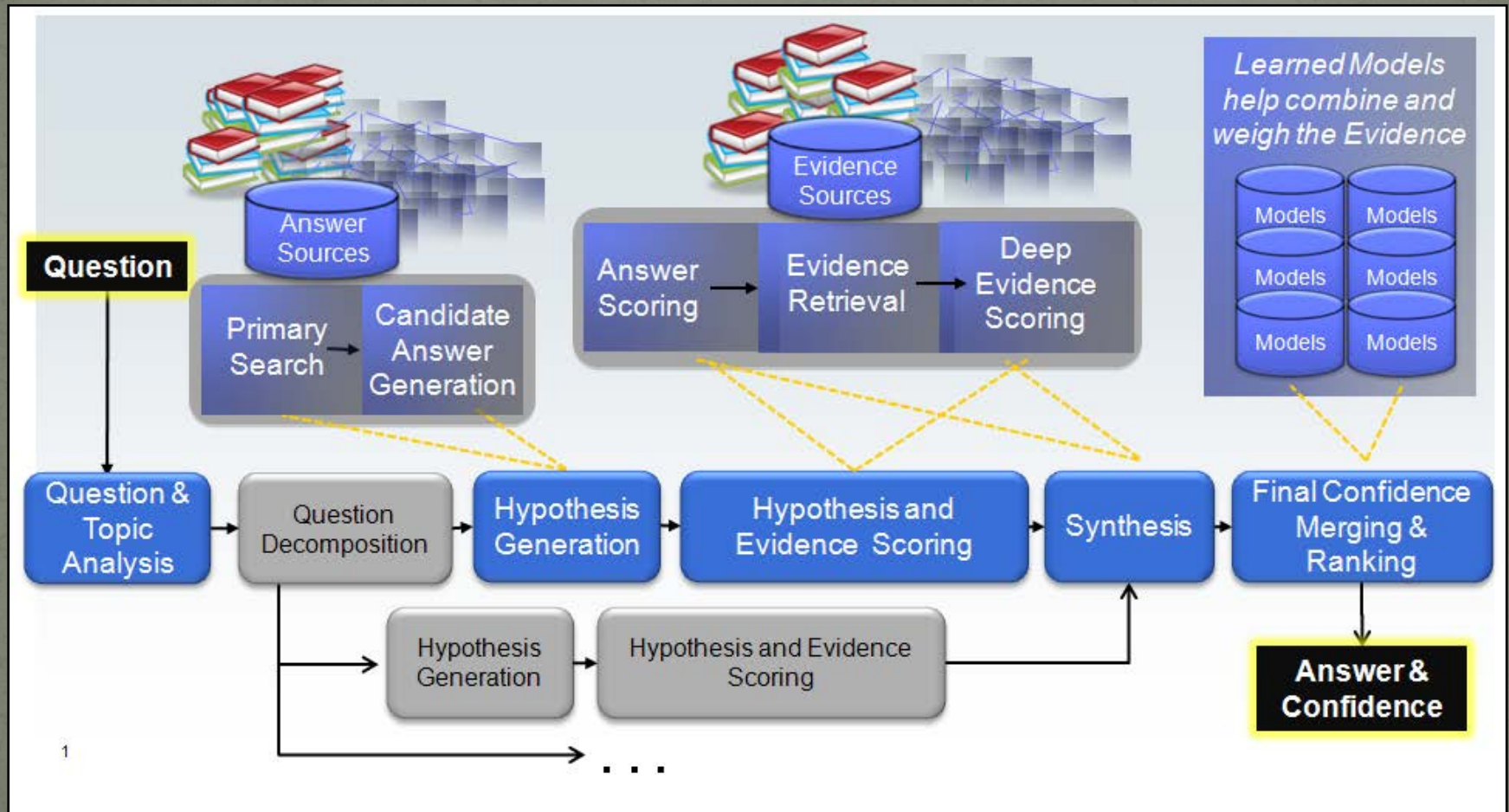


- Cluster of ninety IBM Power 750 servers (plus additional I/O, network and cluster controller nodes in 10 racks) with a total of 2880 POWER7 processor cores and 16 Terabytes of RAM
- Each Power 750 server uses a 3.5 GHz POWER7 eight core processor, with four threads per core
- The POWER7 processor's massively parallel processing capability is an ideal match for Watson's IBM DeepQA software which is **embarrassingly parallel** (that is a workload that is easily split up into multiple parallel tasks)

Software

- Watson's software was written in both Java and C++ and uses Apache Hadoop framework for distributed computing
- Apache **UIMA (Unstructured Information Management Architecture)** framework
- IBM's DeepQA software and SUSE Linux Enterprise Server 11 operating system
- “More than **100 different techniques** are used to analyze natural language, identify sources, find and generate hypotheses, find and score evidence, and merge and rank hypotheses.”

High Level View of DeepQA Architecture



The Science Behind an Answer



Deep QA Process

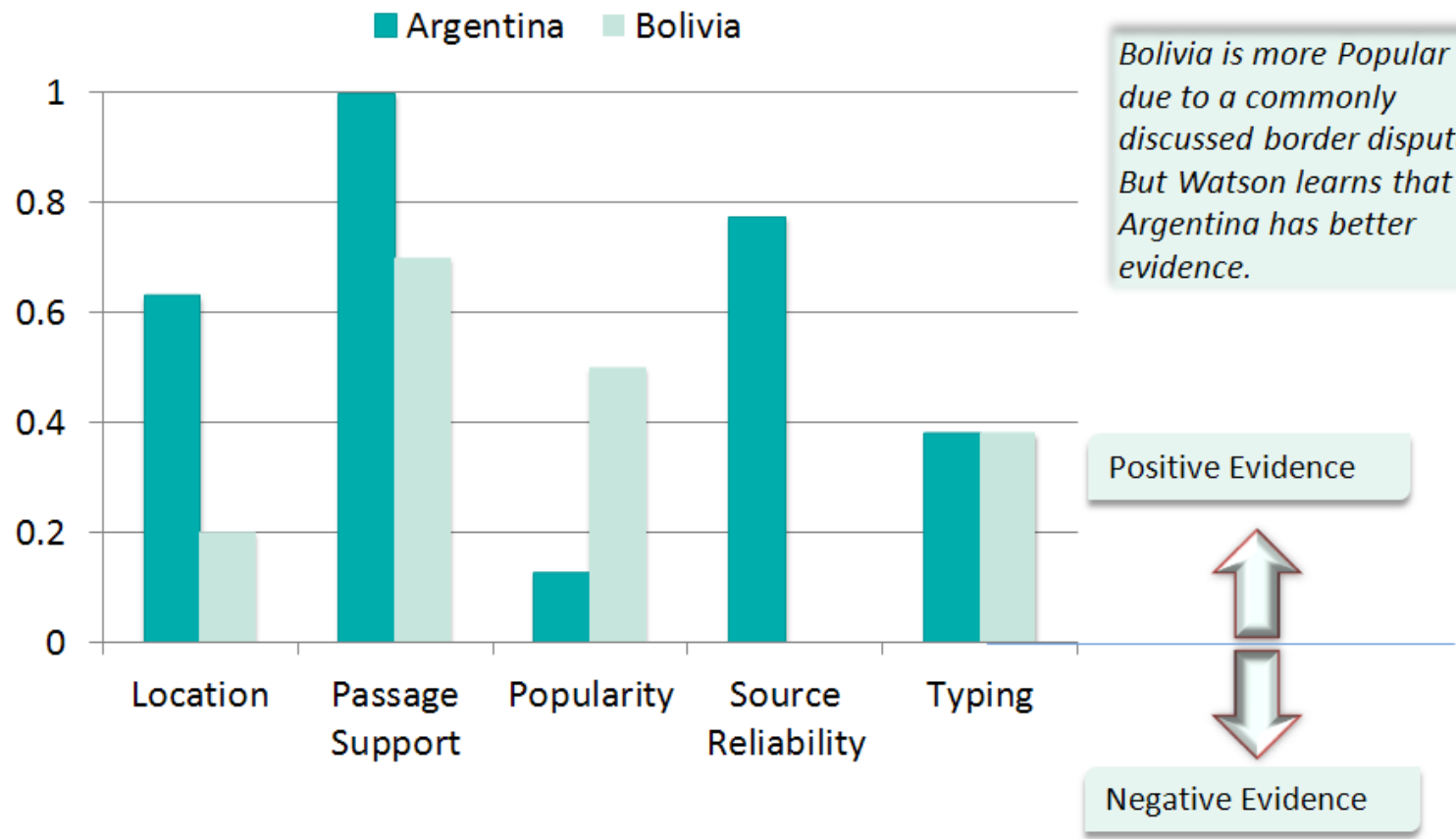
- Analyzes input question and generates many possible candidate answers through broad search of volumes of content
- Hypothesis is formed based on consideration of each candidate answer in context of original question and topic
 - For each of these, DeepQA spawns independent thread attempting to prove it
 - Searches content sources for evidence supporting or refuting each hypothesis
 - Applies hundreds of algorithms for each evidence hypothesis pair that dissects and analyzes along different dimensions of evidence

Types of Dimensions of Evidence

- Type classification
- Time
- Geography
- Popularity
- Passage support
- Source reliability
- Semantic relatedness

Dimensions of Evidence for Jeopardy!

Clue: Chile shares its longest land border with this country.

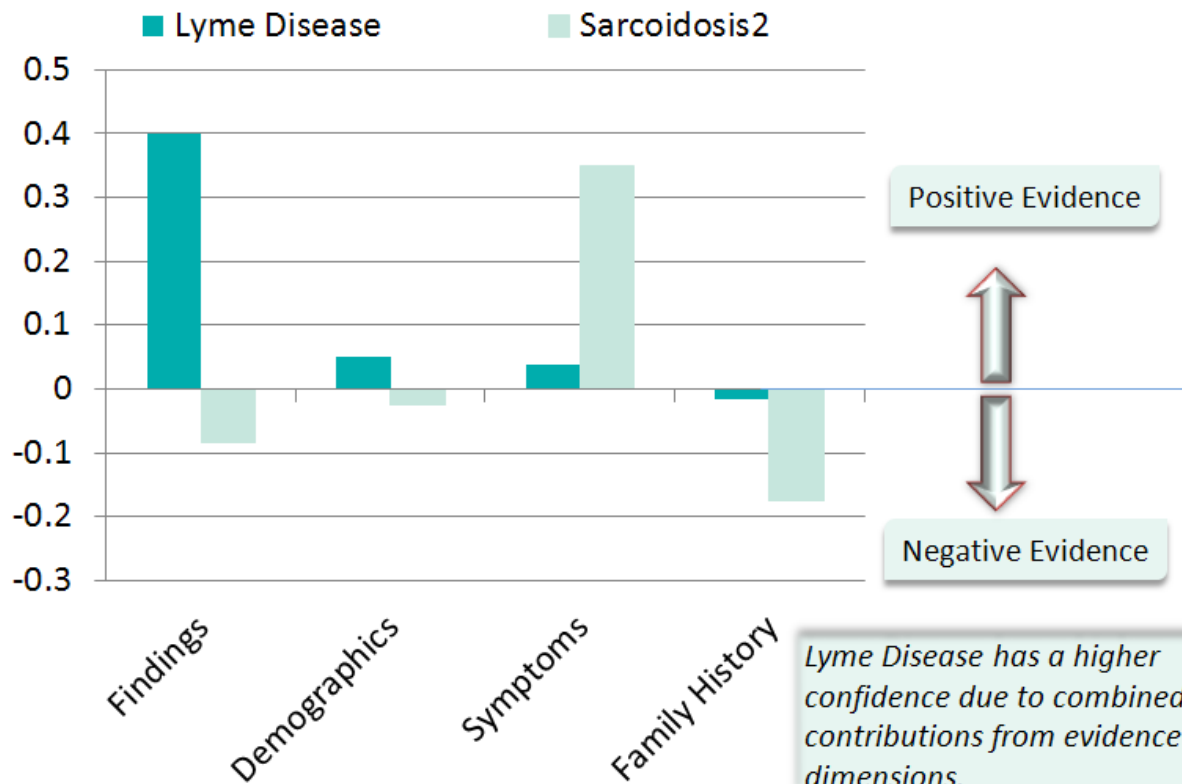


Scoring Features

- These features/scores are then combined based on their learned potential for predicting the right answer resulting in a ranked list of candidate answers, each with a confidence score indicating degree to which the answer is believed to be correct, along with links back to the evidence

Deep QA for Differential Diagnosis

Question: What are diseases, disorders, or causes of uveitis with circular rash, fever, headache, and family history of arthritis in a patient who lives in Connecticut.



Advantages of Dr. Watson Approach

- Represents new architecture for evaluating unstructured content
- Different from traditional expert systems using forward reasoning (data to conclusions) or backward reasoning
- Unlike systems such as Stanford's Mycin that used If-Then statements :
 - **If**
 - The stain of the organism is grampos and the morphology of the organism is coccus and the growth conformation of the organism is chains
 - **Then**
 - There is suggestive evidence that the identity of the organism is streptococcus

Advantages of Watson Approach

- If then approach is costly and difficult to develop and maintain
- Traditional expert systems are brittle because underlying reasoning engine requires perfect match between input data and existing rule forms
- Not all rule forms can be known in advance for all forms that the input data may take

Advantages of Watson Approach

- Watson uses NLP and variety of search techniques to generate likely candidate answers in hypothesis generation (analogous to forward chaining”)
- Uses evidence collection and scoring (analogous to “backward chaining”)
- These make DeepQA more flexible, maintainable, and scalable as well as cost efficient in terms of staying current with vast amounts of new information

Clinical Setting

- Deep QA can develop diagnostic support tool using the context of an input case (information about patient's medical condition)
- Generates ranked list of differential diagnoses with associated confidences
- The dimensions of evidence include
 - Symptoms
 - Findings
 - Patient history
 - Family history
 - Demographics
 - Current medications
 - Many others

Is There A Need for Artificial Intelligence
In Medicine?

Do Physicians Need Assistance?

Motivation for Artificial Intelligence Software in Medicine

- Schiff
 - Diagnostic errors far outnumber other medical errors by 2-4X
- Elstein
 - Diagnostic error rate of about 15% in line with autopsy studies
- Singh and Graber
 - Diagnostic errors are single largest contributor to ambulatory malpractice claims (40% in some studies) and cost about \$300,000 per claim
- Graber
 - Literature review of causes of diagnostic error suggest 65% system related (e.g. communication) and 75% had cognitive related factors

Cognitive Errors

Graber et al Diagnostic Error in Internal Medicine, Arch Intern Med 2005; 165:1493-1499

- Cognitive errors primary due to “faulty synthesis or flawed processing of the available information”
- Predominant cause of cognitive error was premature closure (satisfaction of search in diagnostic imaging)
 - Failure to continue considering reasonable alternatives after an initial diagnosis was reached

Cognitive Errors

- Other contributors to cognitive errors
 - Faulty context generation – lack of awareness of aspects of patient info relevant to diagnosis
 - Misjudging salience of a finding
 - Faulty detection or perception
 - Failed use of heuristics – assuming single rather than multifactorial cause of patient symptoms

Cognitive Errors

- Graber suggested augmenting “a clinician’s inherent metacognitive skills by using expert systems”
- Suggested that clinicians continue to miss diagnostic information and “one likely contributing factor is the overwhelming volume of alerts, reminders, and other diagnostic information in the Electronic Health Record”