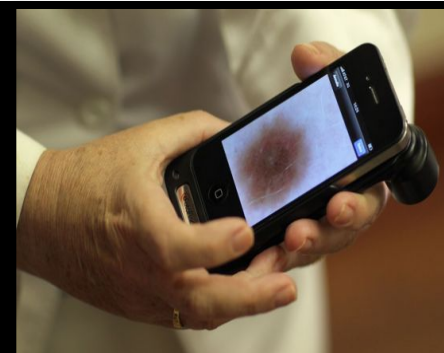
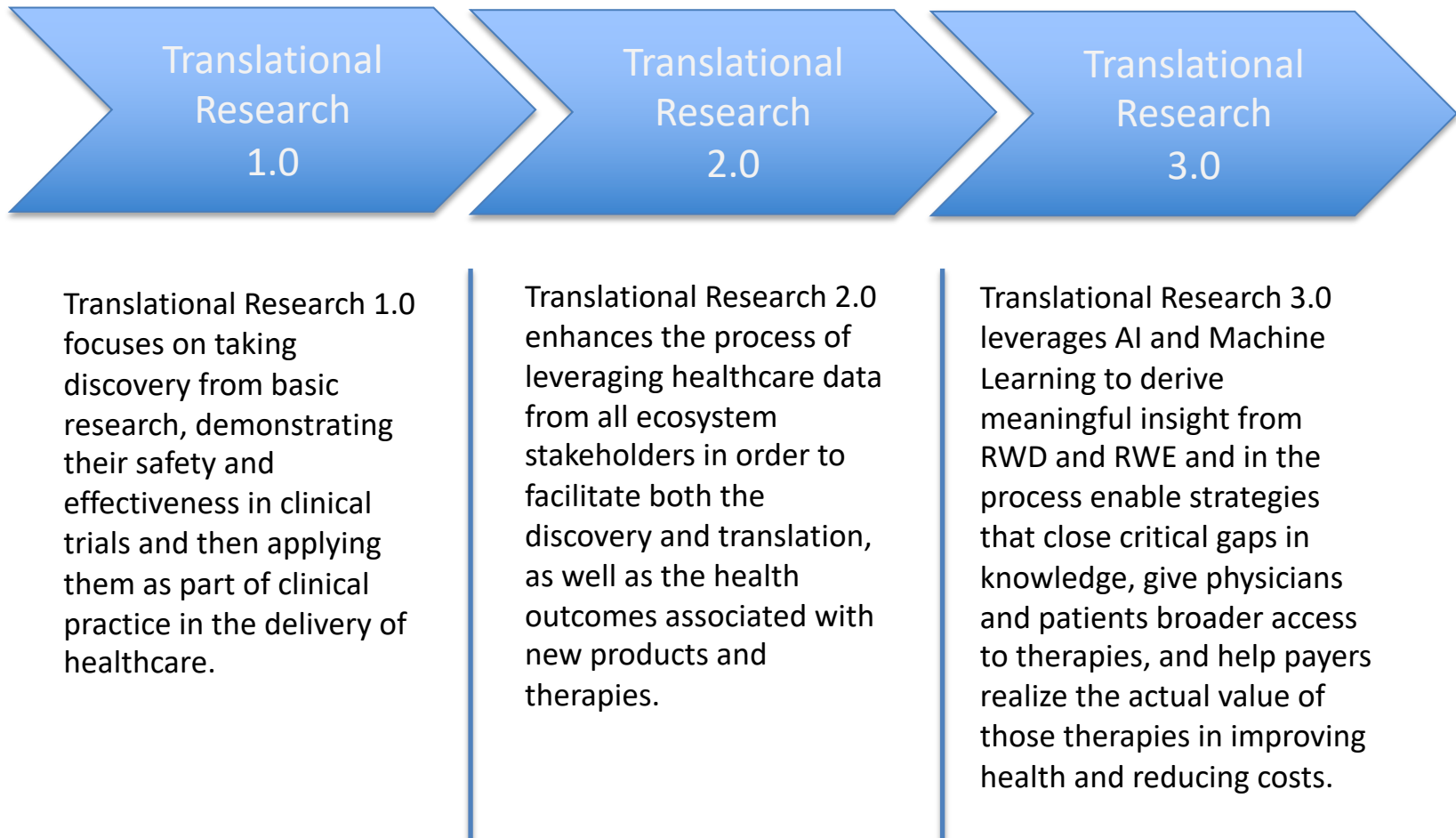


Accelerating Translational Research by Leveraging AI and ML

Dr. Chris Asakiewicz



The Path to Translational Research 3.0

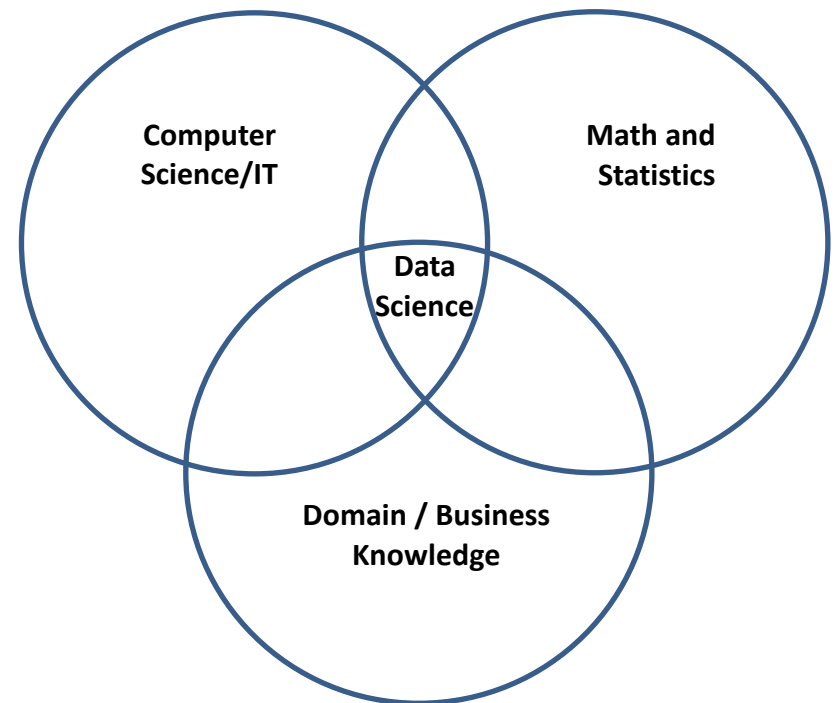


The Path to Translational Research 3.0

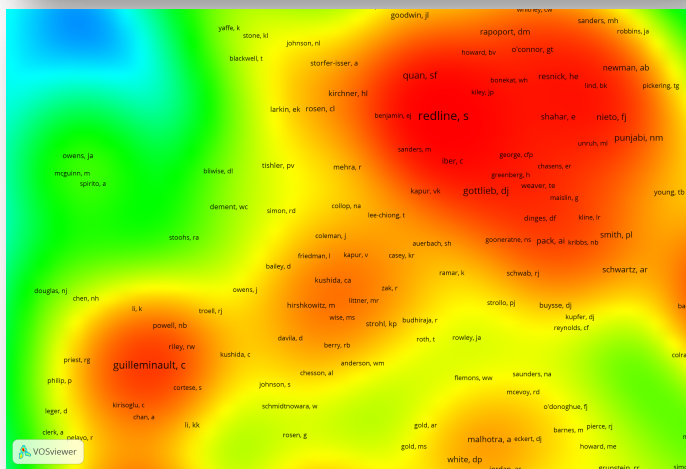
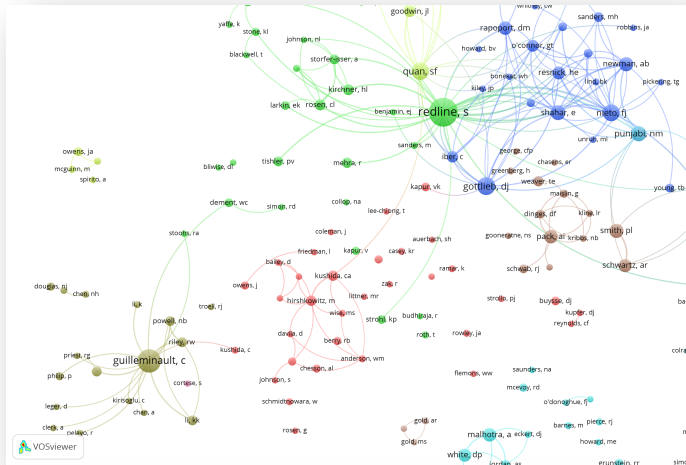
Challenges

- The ability to collect, manage and use large heterogeneous sets of data.
- The need to employ knowledge-anchored methods to discover and test hypotheses.
- The provision of systematic and extensible platforms capable of expediting data integration and analysis workflows.
- The collaborative dissemination of evidence and knowledge.

Opportunities

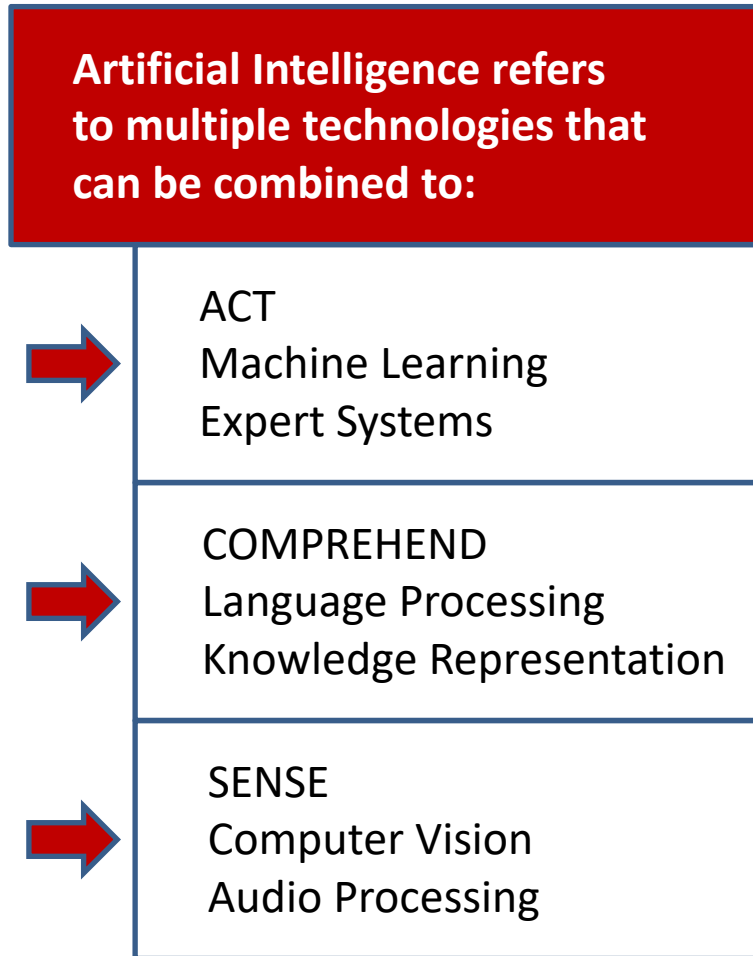


DECISION ANALYTICS AND BIG DATA



Healthcare and Life Sciences

The availability of large quantities of clinical, biological, and healthcare data promises unprecedented advances for the diagnosis and treatment of disease as well as the improvement of patient health; however, the appropriate use of this data mandates changes in the traditional paradigms of data management, analysis, and interpretation.



Concepts

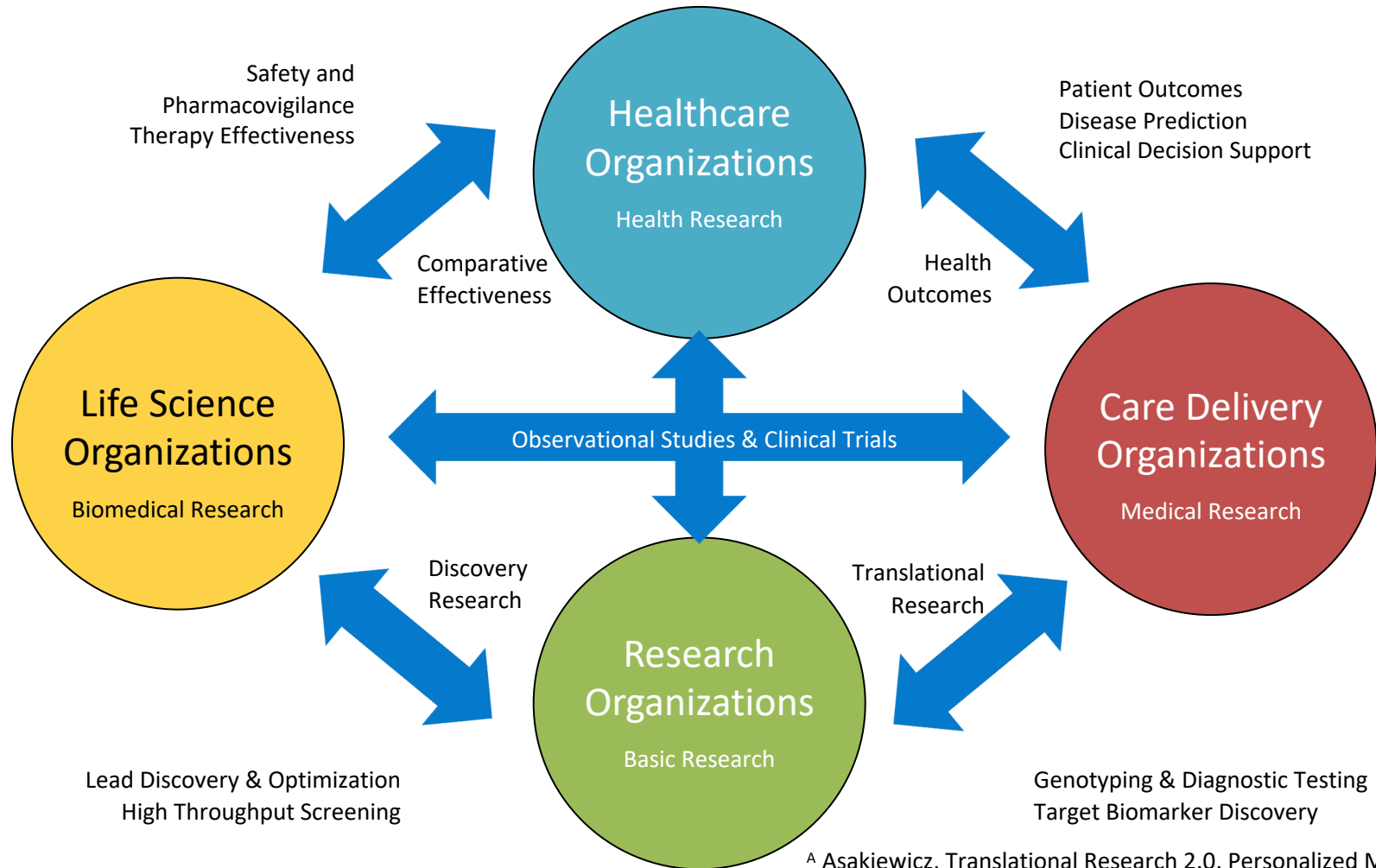
AI refers to multiple technologies that can be combined in different ways to sense, comprehend and act with the ability to learn from experience and adapt over time.

AI is a broad area of computer science that makes machines and computer programs capable of problem solving and learning, like a human brain. AI includes Natural Language Processing (“NLP”) and translation, pattern recognition, visual perception and decision making.

Machine Learning (“ML”), one of the most exciting areas of AI, involves the development of computational approaches to automatically make sense of data – this technology leverages the insight that learning is a dynamic process, made possible through examples and experiences as opposed to pre-defined rules.

Source: Accenture, 2016, “Why is AI the Future of Growth”

MEDICAL AND HEALTHCARE RESEARCH Interconnected Ecosystem ^A



^A Asakiewicz, Translational Research 2.0, Personalized Medicine, 2014

MEDICAL AND HEALTHCARE RESEARCH Interconnected Ecosystem

Health Care - AI is able to make sense of vast amounts of medical history and genetic data and provide insights that can be used to diagnose patients, determine the appropriate course of treatment and monitor that treatment.

Drug Development - With its ability to parse huge amounts of data and find patterns in ways that humans cannot, AI technology can be an invaluable supplement to the clinical data required to bring a drug product to market.

Digital Imaging - MRIs, cat scans, x-rays and all other diagnostic scans produce data in the form of digital images. These images can hold the key to a patient's condition and ultimate survival, but they must be expertly read and analyzed by doctors and used to create the most effective treatment plan.

Diagnosis and Treatment - Beyond simply retrofitting the AI technology that already exists, developments are also underway to fine tune it to meet niche health care demands. Deep-learning algorithms can help doctors diagnose and treat patients more accurately and effectively.

Detection and Prevention - Prevention is the best medicine and, in this vein, wearable medical devices equipped with AI to detect and predict potentially life-threatening events such as blood-glucose spikes and heart attacks are under development.

APPLIED TO THE ANALYSIS OF RWD AND RWE

Definitions of Real-World Data and Real-World Evidence (FDA Framework— Real-World Evidence Program)

Section 505F(b) of the FD&C Act defines RWE as “data regarding the usage, or the potential benefits or risks, of a drug derived from sources other than traditional clinical trials” (21 U.S.C. 355g(b)). In developing its RWE program, FDA believes it is helpful to distinguish between the sources of RWD and the evidence derived from that data. Evaluating RWE in the context of regulatory decision-making depends not only on the evaluation of the methodologies used to generate the evidence but also on the reliability and relevance of the underlying RWD; these constructs may raise different types of considerations. FDA defines RWD and RWE as follows:

- **Real-World Data (RWD)** are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.
- **Real-World Evidence (RWE)** is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.

APPLIED TO THE ANALYSIS OF RWD AND RWE

RWD and RWE include:

- Insurance claims and billing activities
- Pharmacy and specialty pharmacy dispensing data
- Electronic health records (EHR) data
- Aggregated clinical lab-testing data
- Genomic data Patient registries related to specific diseases and products
- Patient-generated data (for instance, from in-home, mobile and wearable devices and sensors)
- Patient-reported outcomes data
- Behavioral data and insights collected from social media
- Purpose-built databases

APPLIED TO THE ANALYSIS OF RWD AND RWE

RWD and RWE-informed efforts include:

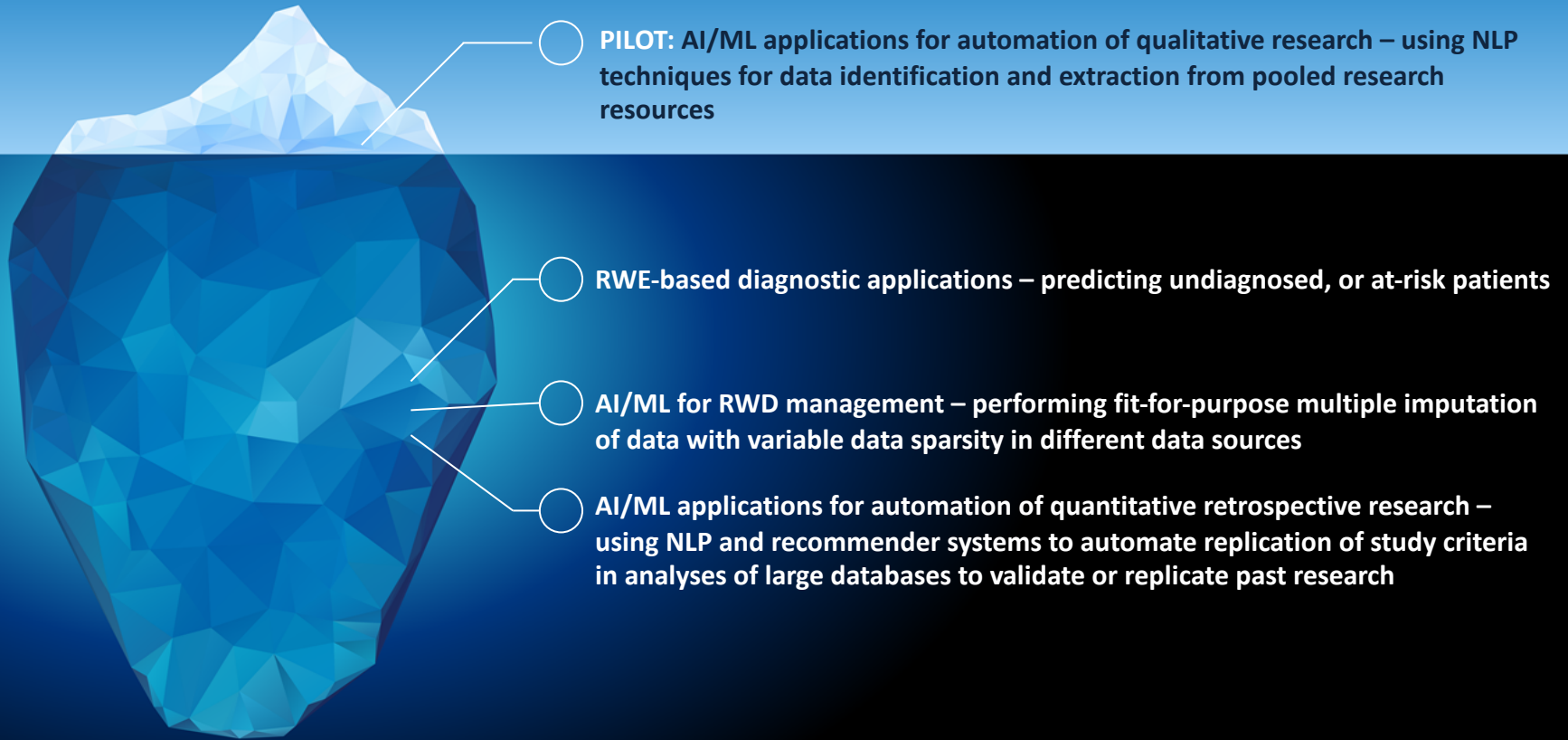
- Identifying new drug targets and better biomarkers
- Finding new uses for existing and late-stage products
- Identifying undiagnosed, under-diagnosed and misdiagnosed illness
- Identifying sites with higher enrollment probability and capability to aid clinical trial recruitment
- Predicting patients who have higher risk for adverse events
- Predicting medication compliance (or lack thereof) in defined patient cohorts (which can inform mitigation efforts)
- Using AI and machine-learning techniques to recognize trends and patterns
- in social media and healthcare sensor data (to inform clinical study development and define study objectives, patient-relevant endpoints and study procedures)

APPLIED TO THE ANALYSIS OF RWD AND RWE

RWD and RWE Issues include:

- Heterogeneous data and lack of interoperability
- Data incompleteness, inconsistency and inaccuracy
- Data ownership and data-use agreement status
- Data privacy regulation / patient-level authorization / ongoing patient consent management for RWD usage
- Lack of data standards
- Lack of standard unique patient identifiers to enable linking across data sources
- Multiple data sources and the need for data aggregation
- Shortage of biomedical and informatics expertise to successfully leverage modeling, simulation and AI and machine learning opportunities

AI AND ML APPLIED TO THE ANALYSIS OF RWD AND RWE



Genesis Research and Stevens School of Business

**Using NLP Techniques for Data Identification and Extraction
from Pooled Research Sources
Genesis Research and Stevens School of Business**

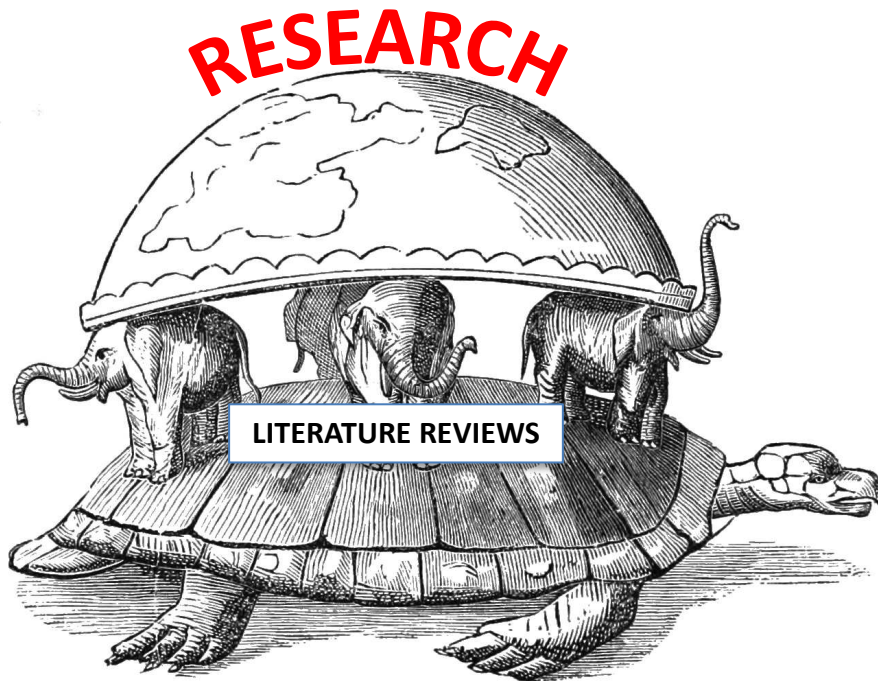
Dr. Chris Asakiewicz, Stevens School of Business

Dr. Rong Liu, Stevens School of Business

Dr. Frank Corvino, Genesis Research

Dr. Marko Zivkovic, Genesis Research

A critical analysis of a segment of a published body of knowledge through summary, classification, and comparison of prior research studies, reviews of literature, and theoretical articles¹



A comprehensive recapitulation on the given scholarship from past to present to:

- Explore existing information
- Ascertain relevant formative works
- Identify key researchers
- Establish current landscape of main ideas, conclusions, and theories
- Clarify research methodologies
- Expose gaps in existing research
- Detect relationships among studies
- Provide context for your own research

¹ University of Wisconsin Writing Center

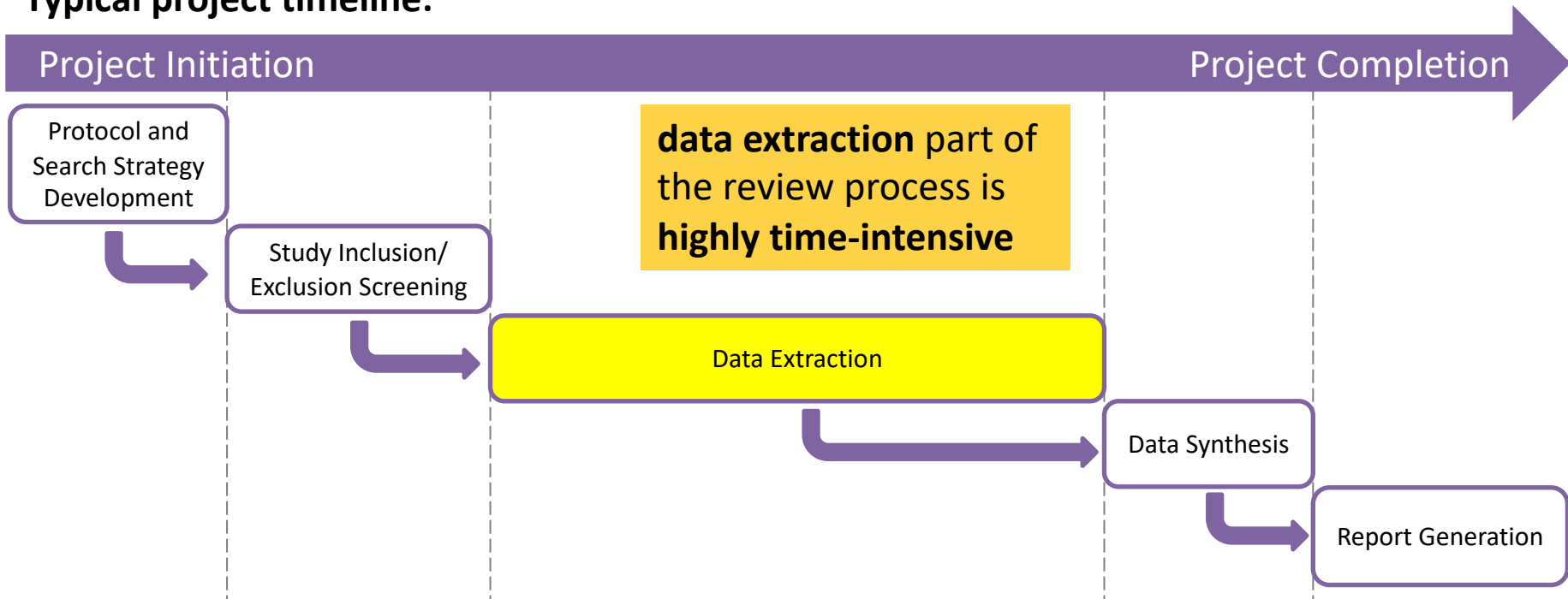
- High quality systematic reviews require two independent researchers to screen studies and extract data¹
- There is a high prevalence error in data extraction for literature reviews and meta-analyses.^{3,4}

**Mean time to
complete a review:**



67.3 weeks²

Typical project timeline:



First-stage screening

Includes screening titles and abstracts.

The largest amount of material to go through but limited in number of sentences per article (abstracts are usually short).

Tasking, but not too demanding. No extraction needed!

Can be automated!!!

DistillerSR



Beverley Shea, Lex M Bouter, Jeremy M Grimshaw, Daniel Francis, Zulma Ortiz, George A Wells, Peter S Tugwell, Maarten Boers (2006). Scope for improvement in the quality of reporting of systematic reviews. From the Cochrane Musculoskeletal Group. *The Journal of rheumatology*. 33(1), 9

Reference Labels:
Add Labels here

OBJECTIVE: To assess the quality of reporting in Cochrane musculoskeletal systematic reviews (excluding back and injury reviews).

METHODS: This study assessed all the Cochrane Musculoskeletal Group's systematic reviews from Issue 4, 2000, of the Cochrane Library Database of Systematic Reviews. Two reviewers independently extracted data and assessed quality. Two assessment tools were used, including an 18 item checklist and flow chart developed by the Quality of Reporting of Meta-analysis (QUOROM) consensus group, and a 10 item scale, the Oxman-Guyatt Overview Quality Assessment Questionnaire (OQAQ). One question on the latter scale (item 10) scores overall quality on a 7 point scale, with high scores indicating superior quality. Data were analyzed using univariate approaches.

RESULTS: The 57 systematic reviews assessed were found to have good overall quality, with scores on individual items revealing only minor flaws. Documenting the flow of included and excluded studies and summarizing the results are 2 areas needing improvement in reporting. According to the Oxman-Guyatt scale the overall scientific quality of the Cochrane musculoskeletal reviews was good (mean 5.02/5% CI 3.71-6.33).

CONCLUSION: Our study found that the reporting quality of Cochrane musculoskeletal systematic reviews was generally good, although there was room for improvement. For example, it might be feasible to

1. Did the study evaluate the quality of systematic reviews?
☐ Yes ☐ No ☒ Can't tell

2. Did the study use a comprehensive search strategy to identify the systematic reviews included for evaluation?
☐ Yes ☐ No ☒ Can't tell

3. Did the study use a measurement tool (eg, AMSTAR) or another set of predefined criteria to assess the quality of the systematic review?
☐ Yes ☒ No ☐ Can't tell

Submit Form and go to: This Form - Next Reference or Skip to Next

DistillerSR's Intuitive 5 Step Process

Upload References > Create Forms > Assign Reviewers & Begin Screening > Monitor > Export Data

The difficult part

Each article must be read in full.

No useful automation... yet!!!

Title and abstracts identified and screened

(n=6,585)

Excluded by screening title and/or abstracts (n=5,035):

- Indication not of interest (n=2,950)
- Animal study (n=104)
- Study design (case studies) (n=1,137)
- Non-English (n=638)
- English articles with no access (n=206)

Full text articles (n=1,550):

- Full text retrieved and assessed for eligibility (n=1,538)
- Awaiting access to full text (English articles) (n=12)

Excluded by full text screening (n=845):

- Indication not of interest (n=252)
- Animal/in-vitro study (n=65)
- Study design (n=431)
- Outcome not of interest (n=97)

Gray literature (n= 21)

Total number of references included for data extraction

(n=726)

Data Extraction Example

 **Article 1**
"Maintenance zinc therapy after initial penicillamine chelation to treat symptomatic hepatic Wilson's disease in resource constrained setting"

 Read the article in detail

Identify + highlight + extract values for aspects of interest



 **Article 2**
"A genetic study of Wilson's disease in the United Kingdom"

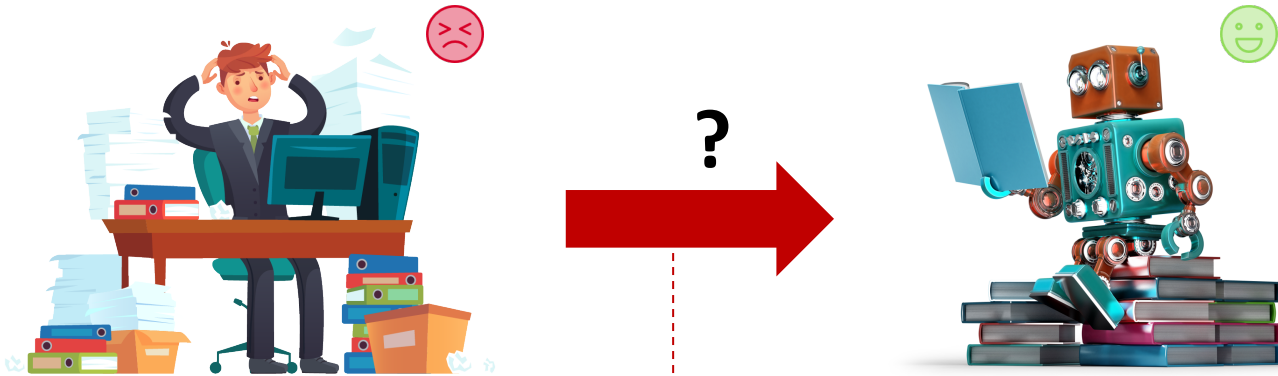
 Read the article in detail

Identify + highlight + extract values for aspects of interest

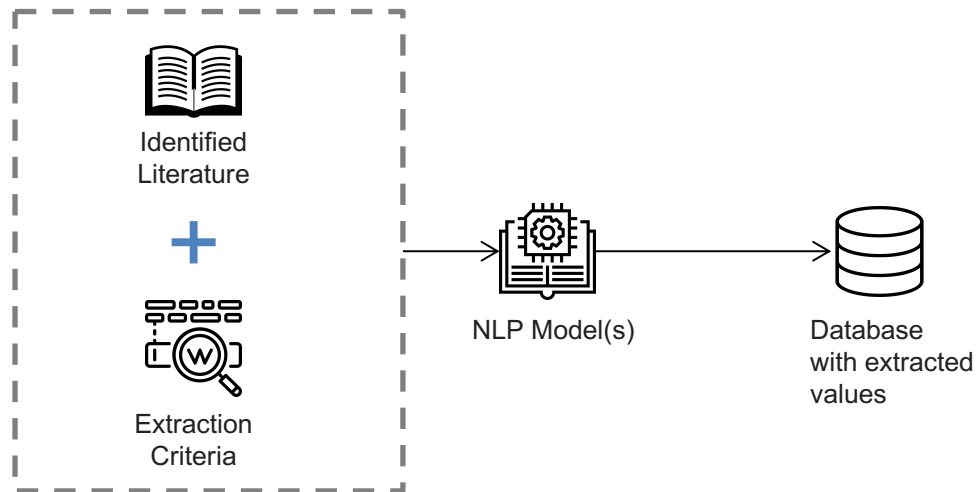


+ 724 other articles

	A	B	C	D	E	F	G	H	
	Author	Year	Title	Data Source Type	Data Source Name	Study Design	Study Period	Study Length	Country
1	Abdelgaffar	2011	Phenotypic and genetic characterization of Patient Charts	NR	Yassin Abdelgaffar Charity Ce	Retrospective Cohort	1992 to 2009	17 years	Egypt
2	Abdelgaffar	2008	WILSON DISEASE (WD) PRESENTING TO THE Patient Charts	NR	Retrospective Cohort	NR	Mean follow-up: 6.4 years (range 1-14)	NR	India
3	Abdelgaffar	2014	Zinc mono-therapy in pre-symptomatic Wilson's disease in central Patient Charts	NR	Children's Hospital of Fudan U	Retrospective Cohort	January 2006 to March 2013	3 years	China
4	Abdelgaffar	2014	Clinical profile of Wilson's disease in central Patient Charts	NR	Chaitram Hospital and Resear	Prospective Cohort	NR	NR	India
5	Aggarwal	2017	Neurological worsening in patients undergo Patient Charts	NR	Wilson's Disease Clinic	Prospective Cohort	2005-2016	11 years	India
6	Aggarwal	2014	Complete neurological recovery in Wilson's disease Patient Charts	NR	NR	Prospective Cohort	2005-2013	NR	NR
7	Aggarwal	2014	A novel objective brain MRI assessment scale Patient Examinations	NR	Validation	NR	NR	NR	NR
8	Aggarwal	2009	A novel global assessment scale for Wilson's NR	NR	Jaslok Hospital and Research	Validation	April 2006-June 2007	NR	India
9	Ala	2015	The psychiatric presentations of Wilson's dis Patient Charts	NR	University of Michigan	Retrospective Cohort	NR	NR	United S
10	Ala	2015	Prospective pilot study of a single daily dose Clinical Trial Data	NR	NC01472874	Clinical trial	November 2011-July 2012	12 months	India
11	Alam	2018	AARAC-ACLF score: BEST PREDICTOR OF OLI NR	NR	NR	Prospective Cohort	NR	90 days	India
12	Alam	2018	Hepatic encephalopathy and bilirubin progn Patient Charts	NR	Institute of Liver and Biliary Sc	Retrospective Cohort	January 2011 to January 2017	6 years	India
13	Alam	2017	Predicting "poor outcome" in decompensate Patient Charts	NR	Retrospective Cohort	NR	2011-2016	NR	India
14	Alam	2017	Low albumin and high alanine transaminase Patient Charts	NR	Institute of Liver and Biliary Sc	Retrospective Cohort	NR	3 years	India
15	Alam	2014	Prediction of poor outcome on medical ther Patient Charts	NR	Institute of Liver and Biliary Sc	Cross Sectional Study	NR	NR	India
16	Alam	2013	Retinal Neurodegeneration in Wilson's Disease Patient Examinations	NR	Heinrich-Heine University	Prospective Case-Control	NR	Mean follow-up time: 9.8 (SD 1.0)	Germany
17	Al-Hilouy	2012	LONG TERM OUTCOMES OF LIVER TRANSPL Patient Charts	NR	University Hospitals Birmingham	Retrospective Cohort	1987-2011	NR	United K
18	Al-Hilouy	2012	Presentation, diagnosis and outcome of pre Patient Charts	NR	King Fahad National Guard Hos	Retrospective Cohort	January 1994 - July 2008	mean (SD) follow up period: 6	Saudi Ar
19	Antczak-Kowalska	2018	Gastroscopy findings in Wilson's disease pati Patient Examinations	NR	2nd department of Neurology, Prospective Cohort	NR	NR	NR	Poland
20	Antczak-Kowalska	2018	Myocardial integrated ultrasono Patient Examinations	NR	NR	Prospective Case-Control	NR	NR	Turkey
	Author	Year	Title	Does this publication have	Incidence	Prevalence	Mortality	Other	Notes
21	Abdelgaffar	2011	Phenotypic and genetic charac	Yes	NR	NR	NR	6 patients (5 males and 1 fema	NR
22	Abdelgaffar	2008	WILSON DISEASE (WD) PRESE	No	NR	NR	NR	NR	NR
23	Abdelgaffar	2014	Zinc mono-therapy in pre-sym	No	NR	NR	NR	NR	NR
24	Abdelgaffar	2014	Clinical profile of Wilson's dis	No	NR	NR	NR	NR	NR
25	Aggarwal	2017	Neurological worsening in pati	No	NR	NR	NR	NR	NR
26	Aggarwal	2014	Complete neurological recovery	Yes	NR	NR	NR	There was no mortality over t	NR
27	Aggarwal	2014	A novel objective brain MRI as	No	NR	NR	NR	NR	NR
28	Aggarwal	2009	A novel global assessment scal	No	NR	NR	NR	NR	NR
29	Ala	2015	The psychiatric presentations (	No	NR	NR	NR	NR	NR
30	Ala	2018	Prospective pilot study of a sin	No	NR	NR	NR	NR	NR
31	Alam	2018	aarac-aclf score: best predicto	No	NR	NR	NR	NR	NR
32	Alam	2017	Hepatic encephalopathy and b	No	NR	NR	NR	NR	NR
33	Alam	2017	Predicting "poor outcome" in	No	NR	NR	NR	NR	NR
34	Alam	2014	Low albumin and high alanine N	No	NR	NR	NR	NR	NR
	Author	Year	Title	Does this publication have	Criteria	Biomarkers	Differential Diagnoses	Common Misdiagnoses	Timing of Diag
35	Abdelgaffar	2011	Phenotypic and genetic charac	No	NR	NR	NR	NR	NR
36	Abdelgaffar	2008	WILSON DISEASE (WD) PRESE	No	NR	NR	NR	NR	NR
37	Abdelgaffar	2014	Zinc mono-therapy in pre-sym	No	NR	NR	NR	NR	NR
38	Abdelgaffar	2014	Clinical profile of Wilson's dis	No	NR	NR	NR	NR	NR
39	Aggarwal	2017	Neurological worsening in pati	No	NR	NR	NR	NR	NR
40	Aggarwal	2014	Complete neurological recovery	No	NR	NR	NR	NR	NR
41	Aggarwal	2014	A novel objective brain MRI as	Yes	NR	NR	NR	NR	NR
42	Aggarwal	2009	A novel global assessment scal	No	NR	NR	NR	NR	NR
43	Ala	2015	The psychiatric presentations (	No	NR	NR	NR	NR	NR
44	Ala	2018	Prospective pilot study of a sin	No	NR	NR	NR	NR	NR
45	Alam	2018	aarac-aclf score: best predicto	No	NR	NR	NR	NR	NR
46	Alam	2017	Hepatic encephalopathy and b	No	NR	NR	NR	NR	NR
47	Alam	2017	Predicting "poor outcome" in	No	NR	NR	NR	NR	NR
48	Alam	2014	Low albumin and high alanine N	No	NR	NR	NR	NR	NR
49	Al-Hilouy	2012	LONG TERM OUTCOMES OF LI	No	NR	NR	NR	NR	NR
50	Al-Hilouy	2012	Presentation, diagnosis and ou	No	NR	NR	NR	NR	NR
51	Antczak-Kowalska	2018	Gastroscopy findings in Wilson	No	NR	NR	NR	NR	NR
52	Antczak-Kowalska	2018	Myocardial integrated ultrason	No	NR	NR	NR	NR	NR
53	Bandmann	2012	A genetic study of Wilson's di	No	NR	NR	NR	NR	NR
54	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
55	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
56	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
57	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
58	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
59	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
60	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
61	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
62	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
63	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
64	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
65	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
66	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
67	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
68	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
69	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
70	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
71	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
72	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
73	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
74	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
75	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
76	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
77	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
78	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
79	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
80	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
81	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
82	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
83	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
84	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
85	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
86	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
87	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
88	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
89	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
90	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
91	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
92	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
93	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
94	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
95	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
96	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
97	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
98	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
99	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR
100	Bandmann	2016	Neurological features and mar	No	NR	NR	NR	NR	NR

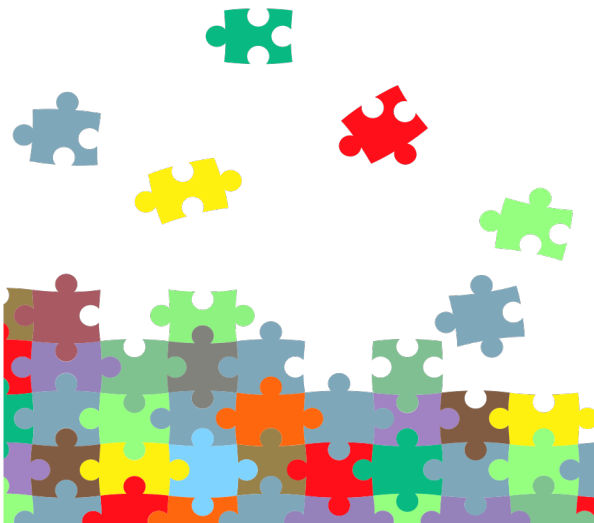


Basic Idea



To date there are **no comprehensive, suitable and intuitive solutions** for automated data extraction of healthcare-related information from pooled research resources.

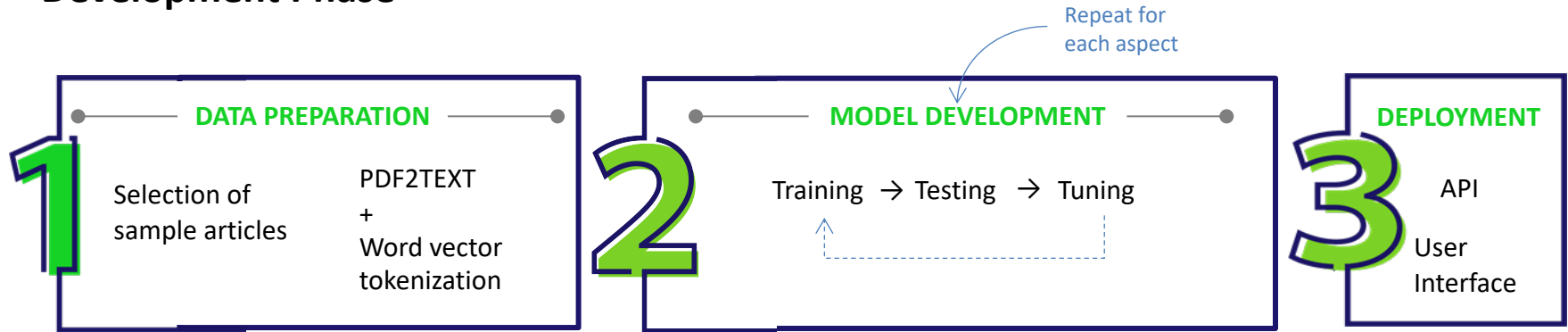
- Kiritchenko et al. developed ExaCT, tool that assists users with locating and extracting key trial characteristics such as eligibility criteria, sample size, drug dosage, and primary outcomes from full-text journal articles.¹
- Wei et al. developed deep learning approaches for medication and adverse drug event extraction from clinical text (2019).²
- Basu et al. proposed a novel framework to expedite systematic reviews by automatically building information extraction training corpora (2019).³



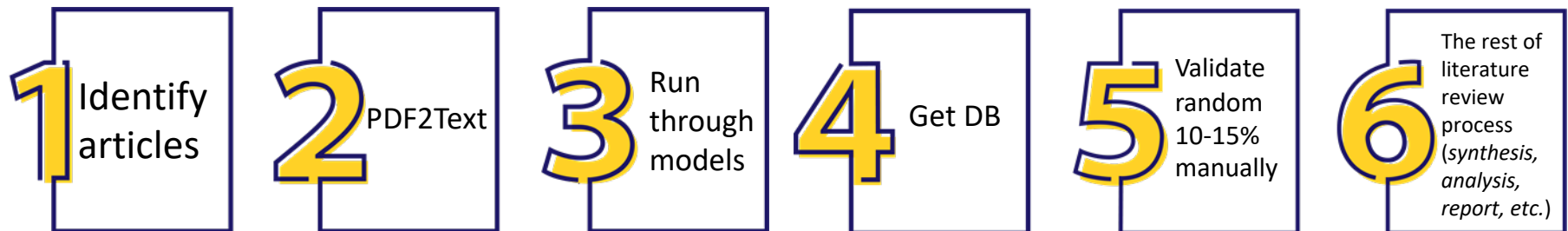
There is other promising research on the horizon, but each addresses a different aspect of the data extraction process.

- Researchers in Denmark compared results of text mining full-text articles versus only their abstracts. Results indicated mining full-text articles consistently outperformed using abstracts only. While the study demonstrated the usefulness of mining different corpora, the Named Recognition Model depended heavily on the dictionaries and stop word lists and was very sensitive to ambiguous words.⁴
- The non-profit eLife Sciences is exploring whether computer vision can be used to provide a high-accuracy method to convert PDF to XML in order to overcome the limitations of unstructured text in PDFs.⁵
- NotesAid is a tool developed to assist patients in understanding their medical health records and physician notes through the use of NLP to link medical terms to lay definitions.⁶

Development Phase



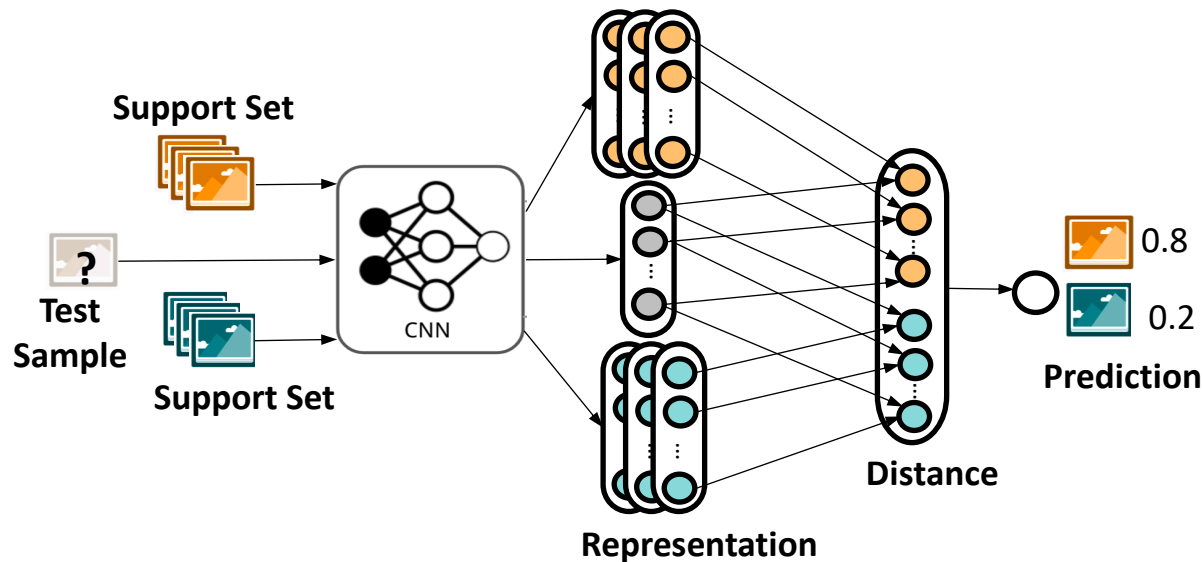
Use Phase



No training needed!

■ Background: Few Shot Learning

- Deep learning models usually require sufficient high-quality data to achieve good performance
- However, it is expensive, and sometimes unrealistic, to obtain sufficient data
- Few-shot learning discovers patterns in such challenging scenarios where data is limited



N-way, K-shot Learning Model

- **n-way**: # of classes
- **k-shot**: # of samples in each support set, k usually is small (e.g. < 20)

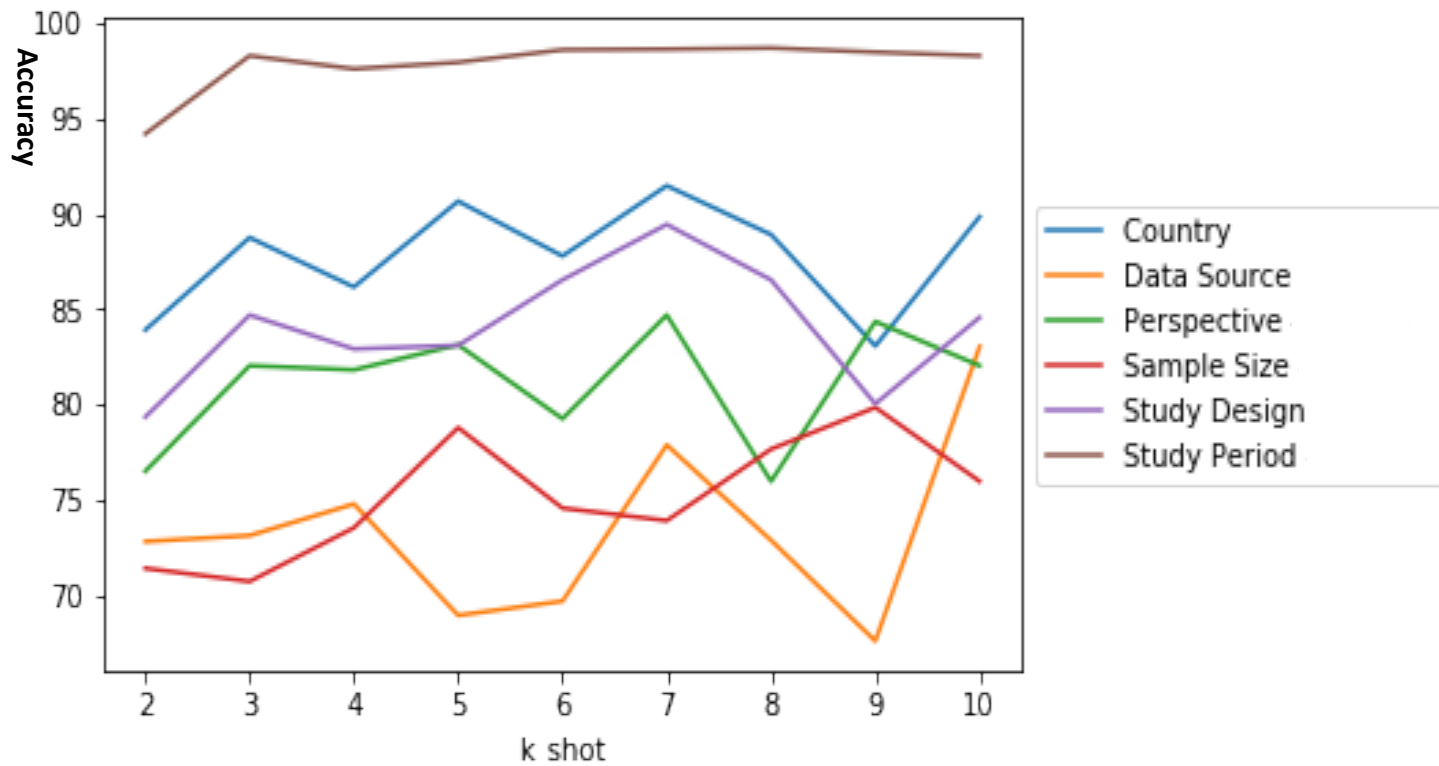
- **Objective:** Identify sentences containing aspects of interest (i.e. **classes**) in research literature

Perspective	Study Design	Data Source	...
Country	Study Period	Sample Size	...

- **Preprocessing:** each sentence in a research article is tokenized into words, and words are mapped to pre-trained Glove word vectors in order to capture meaning

Model Training and Validation:

- ~40 labeled samples for each class, 2/3 (~27) for training, the rest for testing
- The best k varies by model
- Study Period has high accuracy, but Data Source and Sample Size are relatively difficult to identify (e.g., The study observation period was from 2000 to 2016 → Study period)



- Performance is reasonably good considering small sample size
- **Capturing semantics** instead of just keyword matching
 - e.g. Cross sectional **retrospective** study was done → Perspective
 - e.g. ... analysis of an **ongoing** real-life study in CNB → Perspective
- **Multiple-aspect** detection: from a single sentence, the model can extract information relevant to multiple aspects
 - e.g. This is **retrospective** analysis of patients ... Paulbrousse Hospital Villejuif **France** → Perspective, Country
- However, there are a few challenges:
 - **Overfitting** due to small sample set
 - A sentence with **unseen patterns** can be easily misidentified, due to equal distance to both support sets
 - e.g. Poster presentations s630 journal of hepatology 2018 vol. → mislabeled as Study Design, Data Source
 - A sentence alone may not have the **right context** to make prediction
 - e.g. Thirty four patients (32%) were transplanted at mean age of years. → labeled as Sample size; however, we need to identify the total WD sample size

- Model testing and improvement
 - Continuously testing our model with a wide variety of articles
 - Improve the model to be more robust
 - Regularization to overcome overfitting
 - How to deal with unseen patterns
 - Incorporate broader context of each sentence into learning
- Generate summary reports based on identified aspects
 - Extract specific aspect values from identified sentences: e.g. sample size, study period
 - Provide summary statistics
- Transfer learning: apply the model to different literature domains
- Implement the entire machine learning pipeline

Enhanced Data Analysis and Decision Making in Medical and Healthcare Research

Early Disease Detection - Prevention is the best medicine and in this vein, wearable medical devices equipped with AI to detect and predict potentially life-threatening events such as blood-glucose spikes and heart attacks are under development.

Diagnosis and Treatment - Beyond simply retrofitting the AI technology that already exists, developments are also underway to fine tune it to meet niche health care demands. Deep-learning algorithms can help doctors diagnose and treat patients more accurately and effectively.

Imaging and Sensing - MRIs, cat scans, x-rays and all other diagnostic scans produce data in the form of digital images. These images can hold the key to a patient's condition and ultimate survival, but they must be expertly read and analyzed by doctors and used to create the most effective treatment plan.

Smart Health Care - AI is able to make sense of vast amounts of medical history and genetic data and provide insights that can be used to diagnose patients, determine the appropriate course of treatment and monitor that treatment.

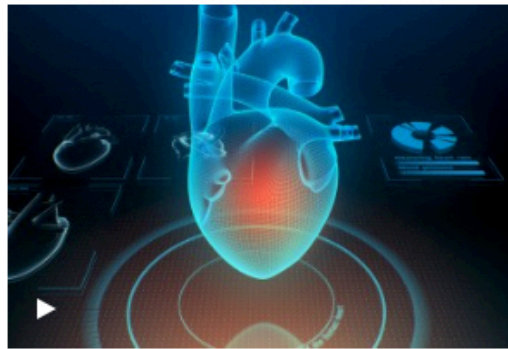
Enhanced Data Analysis and Decision Making in Medical and Healthcare Research

Early Disease Detection



Professors K.P. Subbalakshmi and R. Chandramouli have built an AI-fueled application that has demonstrated it can detect early-warning signs of Alzheimer's disease, dementia and aphasia by analyzing text and voice based communications.

Imaging and Sensing



Stevens professor Negar Tavassolian develops new ways of imaging the surface of the skin and sensing our heartbeats, then processes those signals with new algorithms curated to intelligently detect skin cancer tumors and irregular heartbeats.

Diagnosis and Treatment



Professor Samantha Kleinberg uses machine learning to analyze sensor data from stroke patients in medical centers, research that will better inform stroke diagnosis and treatment.

STEVENS INSTITUTE FOR ARTIFICIAL INTELLIGENCE

<https://www.stevens.edu/research-entrepreneurship/research-centers-labs/stevens-institute-artificial-intelligence>

Enhanced Data Analysis and Decision Making in Medical and Healthcare Research

Real Time Health Monitoring System

Ankit Kumar

Instructor: Dr. Christopher Asakiewicz



Problem Statement

To create a real-time health monitoring system which is user specific using sensors from a smart watch and/or an Arduino device. This application should be able to monitor health features like heart rate, steps count and body temperature in real time and should be able to warn the user or any emergency services of any undesired or serious condition.

Architectural Approach



The architecture flow will consist of following steps:

1. Data will be generated and stored in a file
2. Data will be stream lined using apache Kafka
3. A real-time data visualization will be setup using any of the visualization tool

Tools used

1. JSON file parsing for initial data analysis
2. Apache Kafka for real time streaming
3. Arduino programming for pulling temperature data in real time
4. Python for data cleaning
5. Tableau for visualization

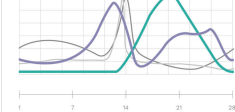
Variables

1. **Heart rate** - Through Smart watch sensor
2. **Steps Count** - Through Smart watch
3. **Body Temperature** - Through Arduino

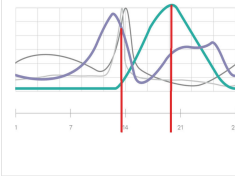
Trigger Cases

1. **Fever** - High temp, low heart rate and steps
2. **Long term unconsciousness** - low heart rate, body temperature and steps count
3. **Heart attack** - high heart rate in a very short time interval

Results



The last part of the project is to be able to visualize results in real time and to be able to plot streams of data and show a trigger result if any abnormal use case occurs.



Business Impact

Our health data analysis application can pull data in real-time from device sensors. Using this system authorities, friends and relatives can easily monitor health of a loved one and can respond immediately in case of an emergency. From a business perspective this can attract customers suffering from a medical condition.

<http://www.stevens.edu/howe/academics/graduate/business-intelligence-analytics>

World's Best Fitness Assistant

Cognitive Computing

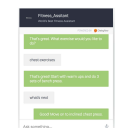
Anand Rai, Jaya Prasad Jayakumar, Saketh Patibandla

Instructor: Christopher Asakiewicz



Text and Speech Enabled Assistant

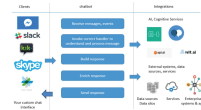
+ Message @Fitness Assistant



Introduction

- We are fitness freaks but never found a good application that guides us while working out.
- There are a few that exist, with the drawback of being static Chatbots which answer predefined questions.
- Our Fitness Assistant, which is AI Driven and uses Natural Language Processing to understand questions and gives the best answer.
- This is a prototype and we are going to build this application for all exercises and food habits.

Chatbot Framework



Natural Language Processing

- Natural Language Processing is the key to an efficient chatbot.
- Every question is stored for future analysis and the knowledge base gives the answer to the question *even if the question is not hardcoded*.



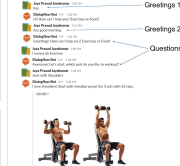
- The NLP searches for the *entity* in the question which is similar to noun in a sentence and narrows down to the answer.

Anatomy of an NLP

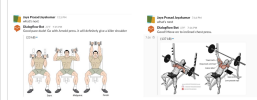


Fitness Assistant

Welcome



Intelligence of our BOT
Same Questions but Different answers based on context



Future Scopes

- This application will be improved by adding in all varieties of exercises and also with multiple sources of knowledge base.
- This app will be *One Stop Solution* for all fitness freaks for exercises and nutrients, we plan to build an in app module to track human behavior and our bot will suggest them what to consume for betterment of their life.

<http://www.stevens.edu/howe/academics/graduate/business-intelligence-analytics>

DengueAI: Predicting Disease Spread

Vicky Rana, Pradeepkumar Prabakaran

Instructor: Christopher Asakiewicz



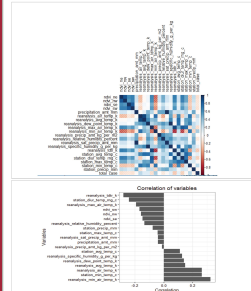
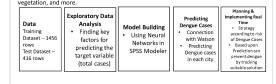
Motivation behind the project

Dengue fever is a mosquito-borne disease that occurs in tropical and sub-tropical parts of the world. In mild cases, symptoms are similar to the flu: fever, rash, and muscle and joint pain. In severe cases, dengue fever can cause severe bleeding, low blood pressure, and even death. Because it is carried by mosquitoes, the transmission dynamics of dengue are related to climate variables such as temperature and precipitation. Although the relationship to climate is complex, a growing number of scientists argue that climate change is likely to produce distributional shifts that will have significant public health implications worldwide.

In recent years dengue fever has been spreading. Historically, the disease has been most prevalent in Southeast Asia and the Pacific Islands. These days many of the nearly half billion cases per year are occurring in Latin America. So, using environmental data collected by various U.S. Federal Government agencies we will be predicting the number of dengue fever cases reported each week in different cities (San Juan and Bogota).

Approach

Accurate dengue predictions would help public health workers and the people around the world take steps to reduce the impact of these epidemics. So, our primary goal is to predict the number of dengue cases each week in each location based on environmental variables describing changes in temperature, precipitation, vegetation, and more.

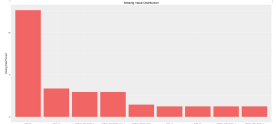
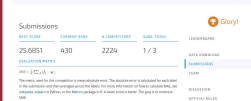


Predictive Model (SPSS Modeler)

1. We used SPSS Modeler for model creation
2. We created model using below mentioned algorithms.
- 3) Neural Networks
- 4) Random Forest
- 5) Linear SVM

We got the best result from Neural Network with mean absolute error as around 25 %.

Dengue Predicting Disease Spread



Prediction [IBM Watson]

Using IBM Watson we can see the trend that will be occurring for the Dengue possibilities in each of the city and can take appropriate action accordingly.

Conclusion

Our primary goal is to predict the number of dengue cases each week in each location based on environmental variables describing changes in temperature, precipitation, vegetation, and more. Using the model that we have built, we can use it for predicting the possibilities of dengue in a region based on some of the variables. This model can be used by governments, administrations, hospitals, social campaigns etc. in order to reduce the impact of these epidemics and can take early precautionary actions.

<http://www.stevens.edu/howe/academics/graduate/business-intelligence-analytics>

Smart Healthcare



Christopher J. Asakiewicz, PhD.

**Professor and Program Director Business Intelligence and Analytics
Director of the Stevens Alliance for Innovation and Leadership (S.A.I.L.)**

Stevens Institute of Technology

Castle Point on Hudson

Hoboken, New Jersey 07030

Email: Christopher.Asakiewicz@stevens.edu

Web: www.drchrisasakiewicz.com

Phone: (201) 216-8012

