平成29 年度 医薬品評価委員会　臨床評価部会 総会
H30年　2月21日　東京証券会館

国立循環器病研究センターの事例を中心とした医療AI の応用に関して

国立循環器病研究センター
予防医学疫学情報部
西村　邦宏

人工知能を応用した研究

- 医療費の20%を占める脳卒中･循環器分野において、人工知能の活用は未だ不十分

【国内事例】東京大学医科学研究所がん最適治療

- 東京大学医科学研究所 宮野研究室
- 今後の日本でのクリニカル・シークエンス、
  がんの個別化医療の社会実装を見据えた研究。
- 人工知能が病名を突き止め患者の命を救う　(東京大学医科学研究所)

がん最適治療 人工知能で

日本人向けに研究

患者別に選択肢、最短10分

日本ＩＢＭ･東大

日本経済新聞

【国内事例】京都大学iPS細胞研究所がWatson採用

IBM Watson、京大iPS細胞研究所が採用

2016/11/09 16:02

2012年のノーベル生理学・医学賞受賞者、山中伸弥氏（京都大学 教授）が所長を務める京都大学 iPS細胞研究所は、米IBM社のコグニティブコンピューティングシステム「IBM Watson」の採用を決めた。導入するのは「Watson for Genomics（Watson Genomic Analytics）」と呼ぶゲノム医療基盤で、日本の医療研究機関としては、2015年7月に導入した東京大学 医科学研究所に続く2施設目となる（関連記事）。

Watson for Genomicsは、がん研究で蓄積されたビッグデータを使い、クラウド上に構築されたゲノム医療基盤。がん細胞の遺伝子変異の解析や、がんの疾病メカニズムの分析、適切な薬剤処方のための分析などを担う。

日本経済新聞

BPニュースセレクト

IBM Watson、京大iPS細胞研究所が採用

2016/11/09 23:00

2012年のノーベル生理学・医学賞受賞者、山中伸弥氏（京都大学 教授）が所長を務める京都大学 iPS細胞研究所は、米IBMのコグニティブコンピューティングシステム「IBM Watson」の採用を決めた。

導入するのは「Watson for Genomics（Watson Genomic Analytics）」と呼ぶゲノム医療基盤で、日本の医療研究機関としては、20…

毎日新聞　7月14日

毎日新聞

写真　動画　有料会員向け　プレミア

トップ　社会　政治　経済　国際　サイエンス　スポーツ　オピニオン　カルチャー　ライフ

総合　医療　科学・技術　環境

[PR]

人工知能

関連ニュースは

糖尿病分析　電子カルテで病状把握へ

毎日新聞 2016年7月14日 07時00分（最終更新 7月14日 07時00分）

愛知県　医療　話題　連載　サイエンス　すべて表示する

人工知能（ＡＩ）で生活習慣病患者の電子カルテを分析し、医療や保険に応用しようと、日本ＩＢＭと藤田保健衛生大学（愛知県豊明市）、第一生命保険が共同研究を始めることが１３日、分かった。カルテに記載された症状や病歴を分析して糖尿病患者の傾向を引き出し、より効率的な治療方法を見つけたり、新たな保険商品の開発につなげたりする狙い。ＡＩを活用した大規模な電子カルテの分析は国内初という。

本日の議題

- 自己紹介
- AI 機械学習 Deep learning
- いわゆるビックデータ（循環器の全国登録の解析）の解析(数理シュミュレーション、機械学習）
- 個人のリスク予測への応用例(画像診断のDeep learning, アンサンブル学習によるリスク予測）
- 自然言語処理によるデータベース作成（ワトソン）

経歴

■1996年……京都大学医学部卒業

■1996-1998年…総合診療部臨床研修

■1998-2000年…聖路加国際病院内科医員

■2000-2001年…神戸大学医学部大学院

■2001-2005年…ハバード大学公衆衛生大学院

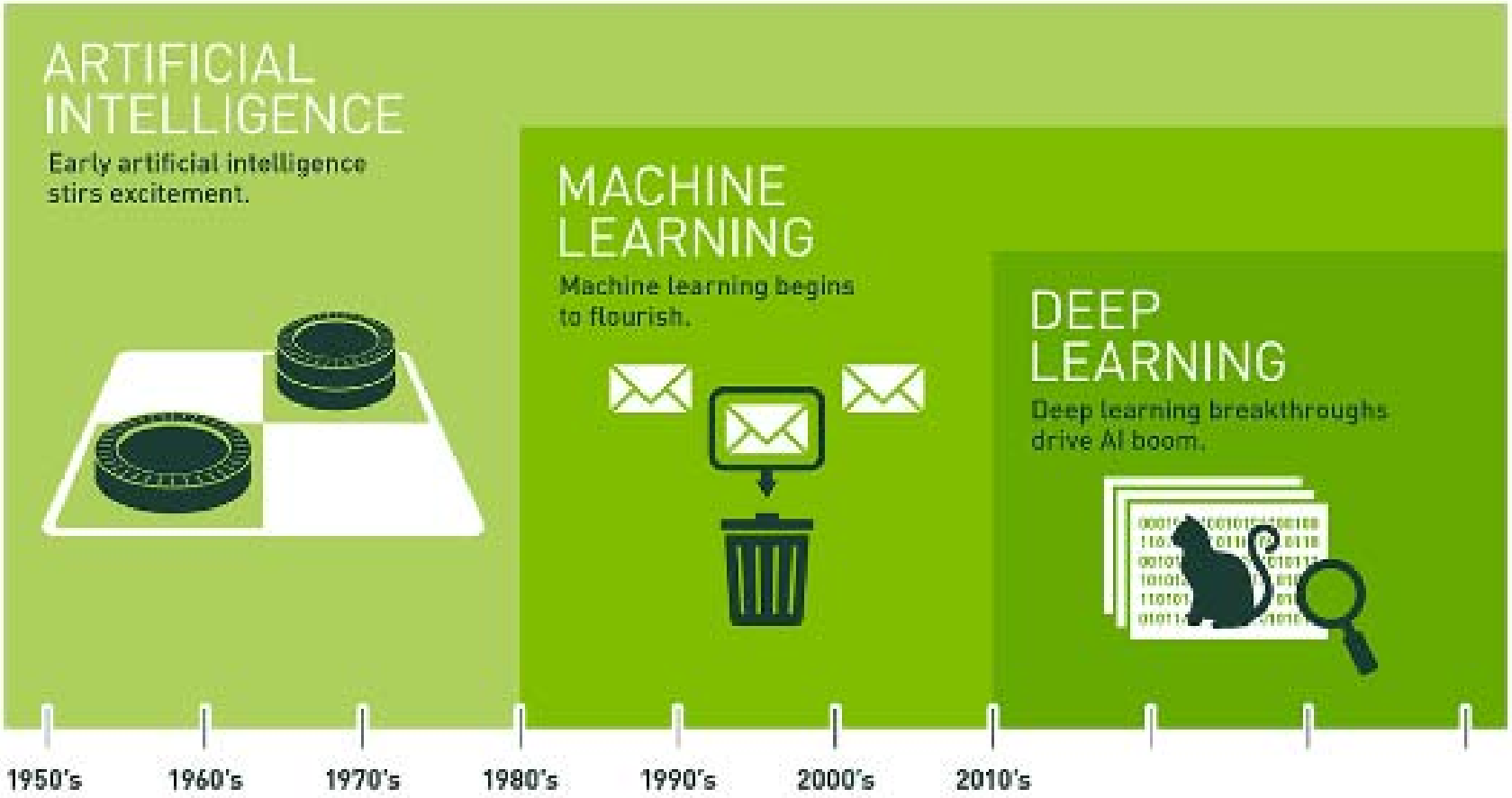
Harvard School of Public health

■2005年-2010…神戸大学大学院医学系研究科
立証検査医学講座助教,准教授

■2010年……国立循環器病研究センター
EBMリスク解析室長

■2018年……同予防医学疫学情報部部長

AI　機械学習　Deep Learning



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

全国登録によるビックデータ解析
JROAD　J-ASPECT研究

循環器疾患診療実態調査
The Japanese Registry Of All cardiac and vascular Diseases (JROAD)

- 2004年より日本循環器学会が主導で行われている全国調査。
- JROADは、① 施設概要（循環器医療の供給度）、② 検査や治療の実施状況（循環器医療の必要度）から構成されている。
- 2013年度よりJROADのデータセンターはNCVCにおかれ、共同研究として運用している。
- 過去５年間 2013~2017年度調査ではいずれも循環器専門医研修施設・研修関連施設の登録率は100%を達成（資格更新のための条件）。

2017年施設分類	施設数	入力完了施設	回答率
研修施設	1,017	1,017	100%
研修関連施設	336	336	100%
合計	1,353	1,353	100%

http://www.j-circ.or.jp/jittai_chosa/jittai_chosa2017web.pdfより作成

JROAD-DPC

JROAD-DPCのQI分布とAHA annual reportの比較

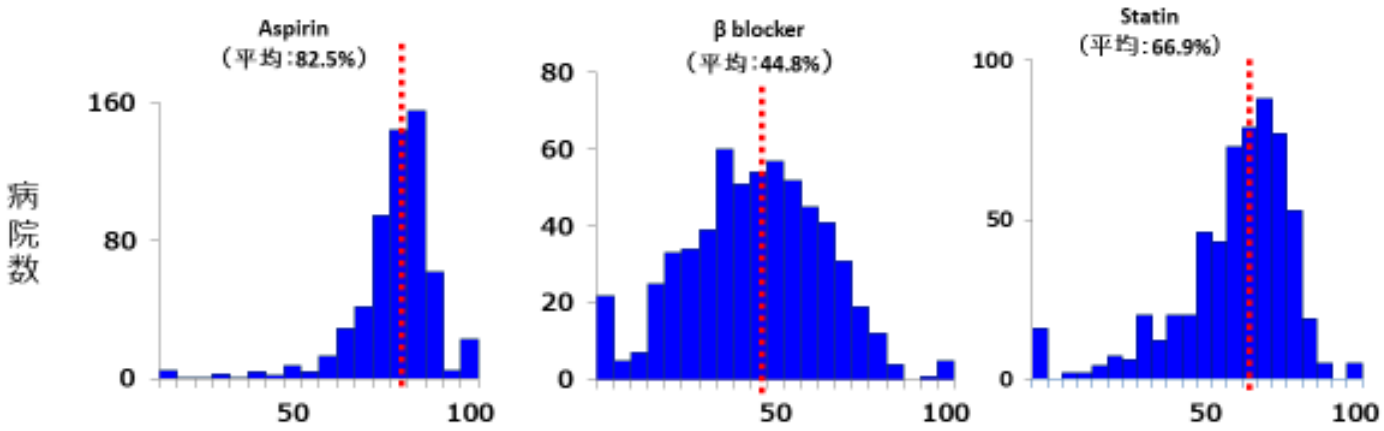


Table 21-3. Time Trends in GWTG-ACS Quality-of-Care Measures, 2006–2011

Quality-of-Care Measure	2006	2007	2008	2009	2010*	2011*
Aspirin within 24 h of admission	94.7	92.8	91.2	90.9	97	97.6
Aspirin at discharge	94.4	95.8	94.9	95.5	98	98.3
β-Blockers at discharge	92.8	94.6	94.5	94.9	96	96.7
Lipid-lowering medication at discharge	84.5	85.6	81.6	86.8	92†	98.4†
Lipid therapy at discharge if LDL cholesterol >100 mg/dL	89.1	90.7	91.9	92.5	NM	NM
ARB/ACEI at discharge for patients with LVEF <40%	87.3	91.1	91.9	91.9	86	87.8
Adult smoking cessation advice/counseling	94.3	97.4	98.4	98.4	98	98.4
Cardiac rehabilitation referral for AMI patients	71.1	63.6	52.0	49.1	75	76.5

Values are percentages.
In the ACTION registry (Acute Coronary Treatment and Intervention Outcomes Registry), the unadjusted in-hospital mortality rate for 2011 was 4.6% (95% confidence interval, 4.4%–4.7%); excludes transfer-out patients). The American Heart Association's (AHA's) GWTG-CAD has now merged into the ACTION registry.
GWTG-ACS indicates Get With The Guidelines–Acute Coronary Syndrome; LDL, low-density lipoprotein; NM, not measured; ARB/ACEI, angiotensin receptor blocker/angiotensin-converting enzyme inhibitor; LVEF, left ventricular ejection fraction; and AMI, acute myocardial infarction.
*Measures from 2006–2009 are from the AHA's GWTG-CAD (Coronary Artery Disease) registry. 2010/2011 measures are from the AHA's ACTION registry (Acute Coronary Treatment and Intervention Outcomes Registry.
†Represents statin use.

現在24の公募研究が実施中、AMED
厚生科研班（小川班、小室班、安田班、飯原班など多数）が利用

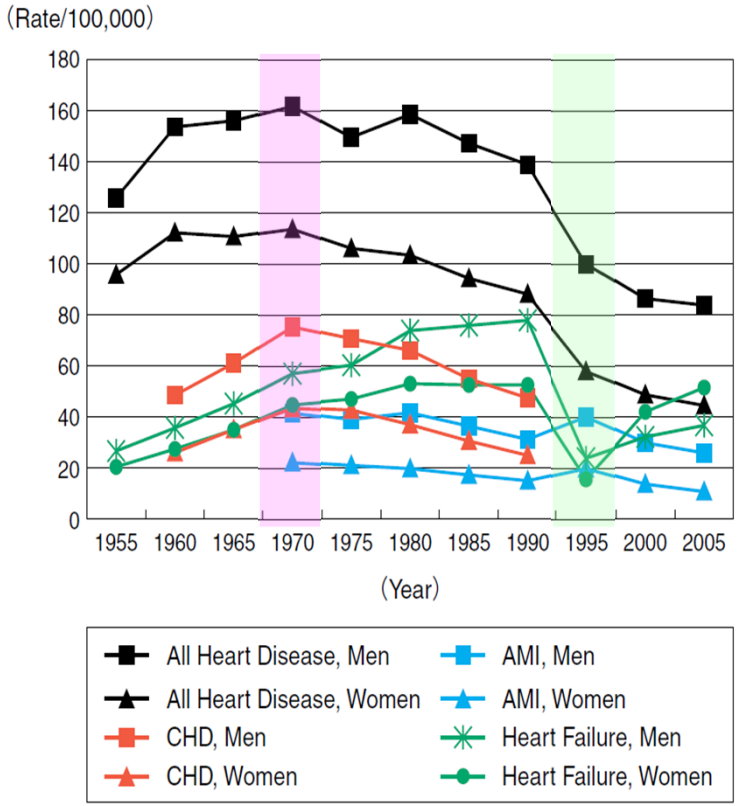
Yasuda, Nishimura et al, Cir J 201

JROAD-DPC
急性心筋梗塞 入院期間中の内服薬の種類

本邦ではAMI入院中(n=38,463)に平均9.3種類の処方が行われている

AMI_EFレセプト電算コード_処方種類	
平均	9.39
標準誤差	0.019
median	9
最頻値　(モード)	8
標準偏差	3.78
分散	14.32
範囲	34
Min	1
Max	35
標本数	38463
最大値(1)	35
最小値(1)	1
信頼区間(95.0%)	0.037

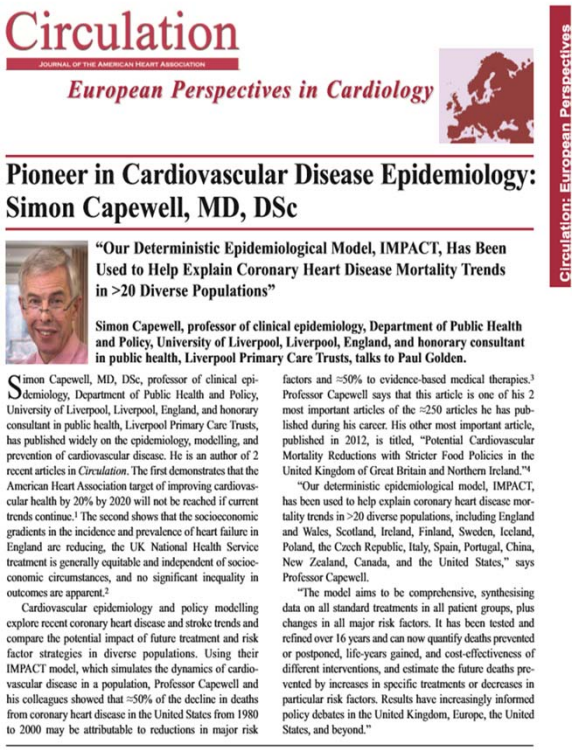
循環器の死亡は減少しているが理由は？



- Age-adjusted all heart disease and coronary heart disease mortality peaked around 1970.
- The sudden change in mortality from the all heart disease and heart failure before and after 1995 was due to the change in the diagnostic approach for heart failure.
- Mortality from AMI maintained a low rate with a slightly downward trend.

J Atheroscler Thromb 2010 7 14: 278-86.

ビックデータによる　循環器死亡全体による数理シュミュレーション
国際共同研究　Impact　Model



Circulation

JOURNAL OF THE AMERICAN HEART ASSOCIATION

European Perspectives in Cardiology

Circulation: European Perspectives

Pioneer in Cardiovascular Disease Epidemiology: Simon Capewell, MD, DSc

"Our Deterministic Epidemiological Model, IMPACT, Has Been Used to Help Explain Coronary Heart Disease Mortality Trends in >20 Diverse Populations"

Simon Capewell, professor of clinical epidemiology, Department of Public Health and Policy, University of Liverpool, Liverpool, England, and honorary consultant in public health, Liverpool Primary Care Trusts, talks to Paul Golden.

Simon Capewell, MD, DSc, professor of clinical epidemiology, Department of Public Health and Policy, University of Liverpool, Liverpool, England, and honorary consultant in public health, Liverpool Primary Care Trusts, has published widely on the epidemiology, modelling, and prevention of cardiovascular disease. He is an author of 2 recent articles in *Circulation*. The first demonstrates that the American Heart Association target of improving cardiovascular health by 20% by 2020 will not be reached if current trends continue.[1] The second shows that the socioeconomic gradients in the incidence and prevalence of heart failure in England are reducing, the UK National Health Service treatment is generally equitable and independent of socioeconomic circumstances, and no significant inequality in outcomes are apparent.[2]

Cardiovascular epidemiology and policy modelling explore recent coronary heart disease and stroke trends and compare the potential impact of future treatment and risk factor strategies in diverse populations. Using their IMPACT model, which simulates the dynamics of cardiovascular disease in a population, Professor Capewell and his colleagues showed that ≈50% of the decline in deaths from coronary heart disease in the United States from 1980 to 2000 may be attributable to reductions in major risk factors and ≈50% to evidence-based medical therapies.[3] Professor Capewell says that this article is one of his 2 most important articles of the ≈250 articles he has published during his career. His other most important article, published in 2012, is titled, "Potential Cardiovascular Mortality Reductions with Stricter Food Policies in the United Kingdom of Great Britain and Northern Ireland."[4]

"Our deterministic epidemiological model, IMPACT, has been used to help explain coronary heart disease mortality trends in >20 diverse populations, including England and Wales, Scotland, Ireland, Finland, Sweden, Iceland, Poland, the Czech Republic, Italy, Spain, Portugal, China, New Zealand, Canada, and the United States," says Professor Capewell.

"The model aims to be comprehensive, synthesising data on all standard treatments in all patient groups, plus changes in all major risk factors. It has been tested and refined over 16 years and can now quantify deaths prevented or postponed, life-years gained, and cost-effectiveness of different interventions, and estimate the future deaths prevented by increases in specific treatments or decreases in particular risk factors. Results have increasingly informed policy debates in the United Kingdom, Europe, the United States, and beyond."

Design:

- **IMPACT model**
  - a extensively validated comprehensive CHD mortality model
    - ✓ England & Wales (Circulation)
    - ✓ Scotland/Ireland (BMJ)
    - ✓ Sweden(Eur heart J)
    - ✓ USA (NEJM)
    - ✓ Canada (JAMA)
    - ✓ China(Circulation)
    - ✓ And so on
- Target population: aged 35-84

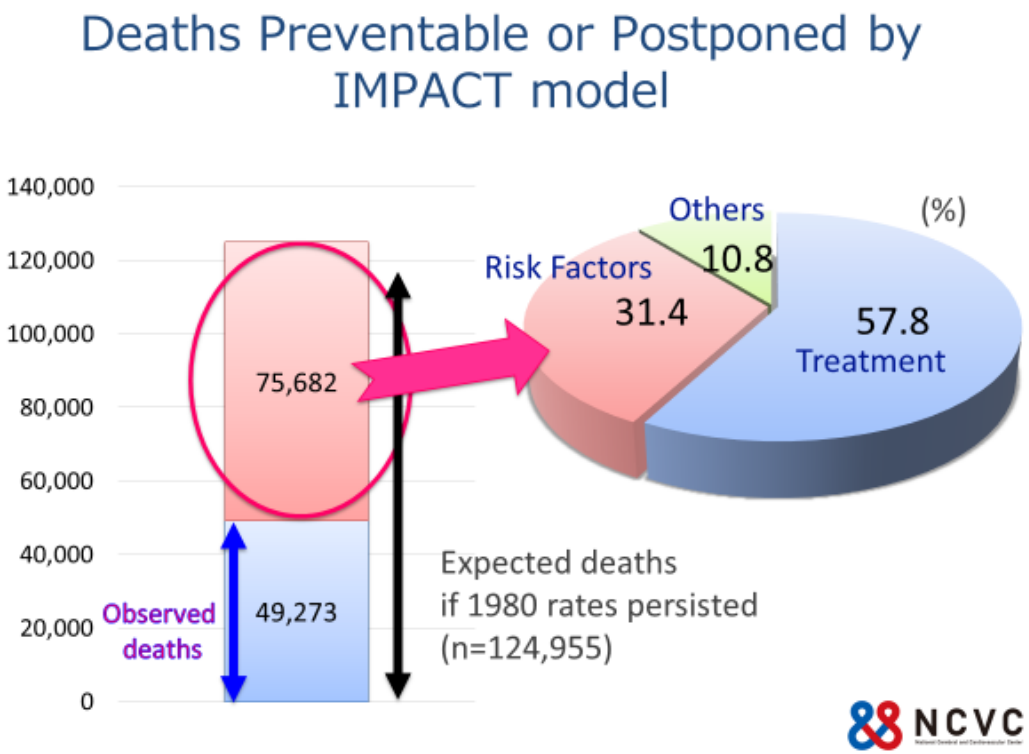
IMPACTモデル構成要素

Explaining the changes in CHD mortality in Japan between 1980 and 2012

急性期治療	二次予防	リスクファクター
• 心肺蘇生	• アスピリン	• 収縮期血圧
• 血栓溶解療法	• B遮断薬	• 総コレステロール
• アスピリン	• ACE阻害薬	• 喫煙
• IIB/IIIA阻害薬	• スタチン	• 糖尿病
• B遮断薬	• ワーファリン	• 体格指数（BMI）
• ACE阻害薬	• 心臓リハビリ	
• PCI		
• CABG		
• 心臓リハビリ		

Circulation under review

研究成果の概要—IMPACT Model



日本においては欧米にくらべ治療の普及（ACE/ARBなど）が大きい。減塩による高血圧減少効果が大きいが、コレステロール、糖尿病、肥満により将来は増加に転じることを示唆する結果

Circulation under review

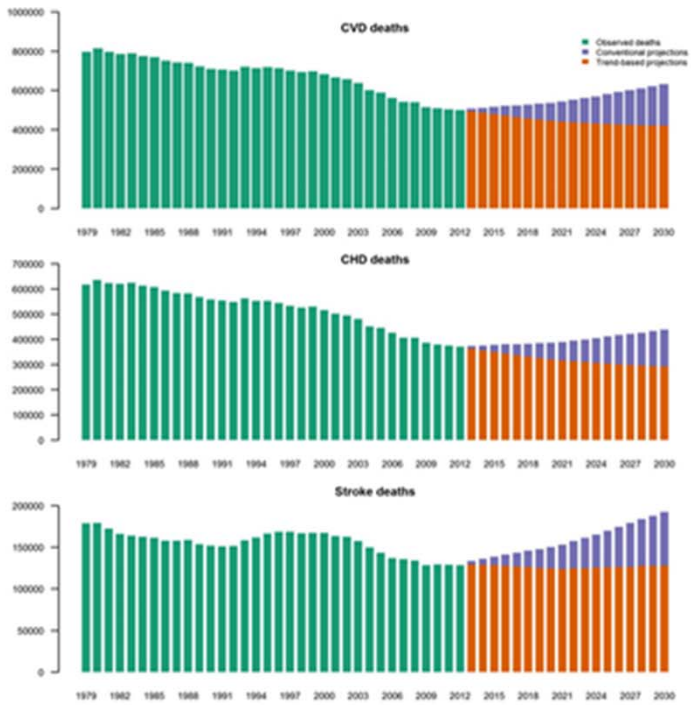
Summary of population risk factor changes

RISK FACTORS	Risk factor levels 1980	Risk factor levels 2012	Risk factor change	DPP	% DPP
Systolic BP	140.9	132.5	-8.4 mmHg	18,128	24.0%
Smoking	34.9%	21.6%	-13.3%	8,298	11.0%
Physical inactivity	89.8%	72.1%	-17.8%	5,641	7.5%
Cholesterol	4.93	5.22	0.29 mmol	-1,600	-2.1%
BMI	22.6	23.3	0.7	-1985	-2.6%
Diabetes	4.4%	7.3%	2.9%	-4722	-6.2%
Total				**23,760**	**31.4%**

コレステロール、肥満、糖尿病が死亡増加につながる

Precise estimate for Japanese future and Policy impacts are needed- Bayesian model based estimate and modeling study

Total number of cardiovascular disease (CVD) deaths, 1979 to 2030; observed deaths, 1979 to 2012; and conventional projections and trend-based projections, 2013 to 2030.



Japanese CHD might be increased by increased DM, Obesity , and Cholesterol

We have good statistical data, but methodology for precise evaluation is needed

日本はこれから増加に転じる可能性

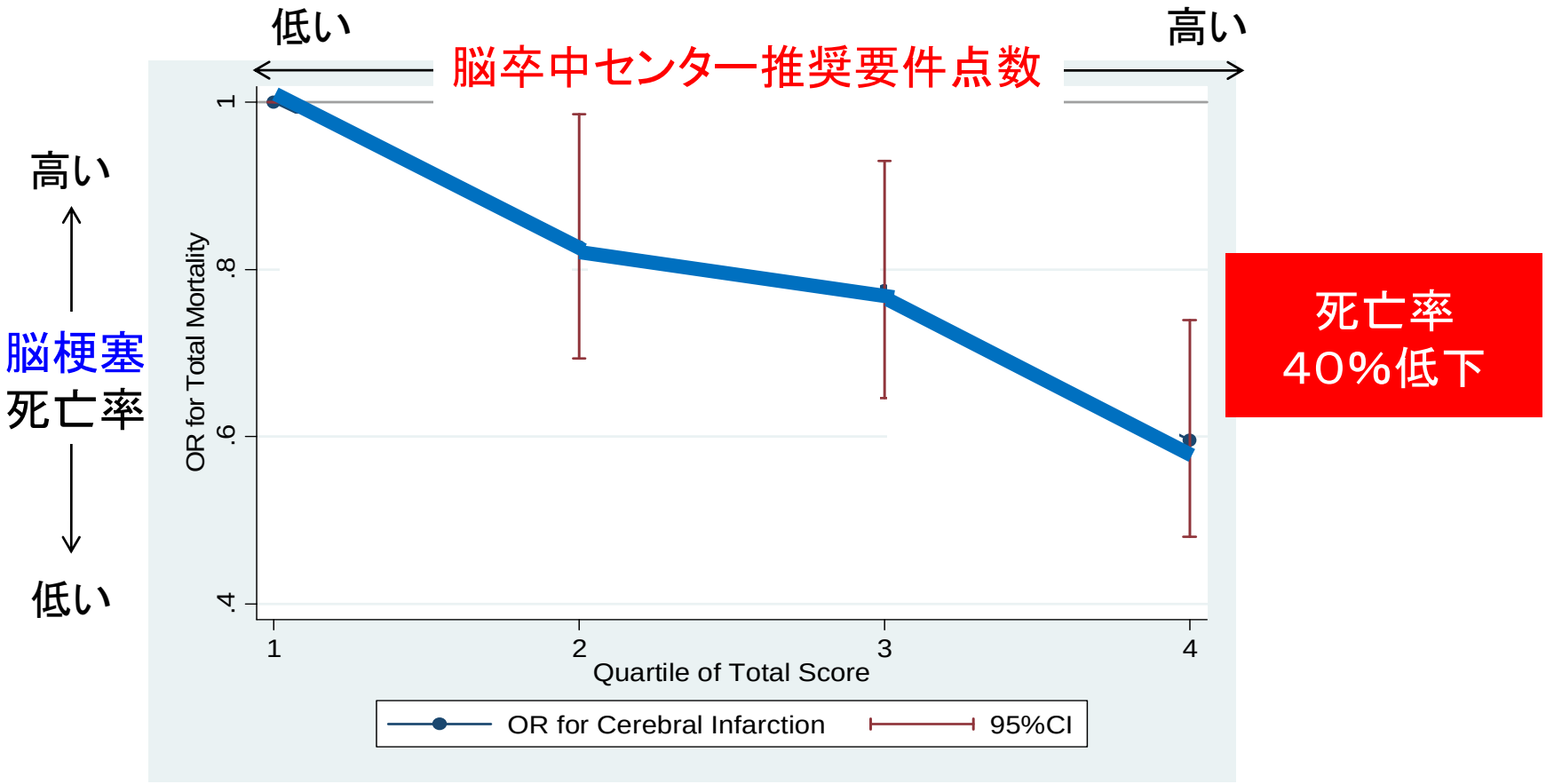
Circulation. 2016;133:967-978

包括的脳卒中センターの整備は、脳卒中死亡を激減させうる！

脳卒中医療の質と死亡率（ J-ASPECT Study）

（ 本邦32，671例の解析）

PLoS One. 2014 May 14;9(5):e96819.



脳卒中　循環器疾患は季節変動、気温、気圧などにより医療資源需要が大きく変化する

Effect of Seasonal and Monthly Variation in Weather and Air Pollution Factors on Stroke Incidence in Seoul, Korea

Myung-Hoon Han, MD; Hyeong-Joong Yi, MD, PhD; Young-Soo Kim, MD, PhD; Young-Seo Kim, MD, PhD

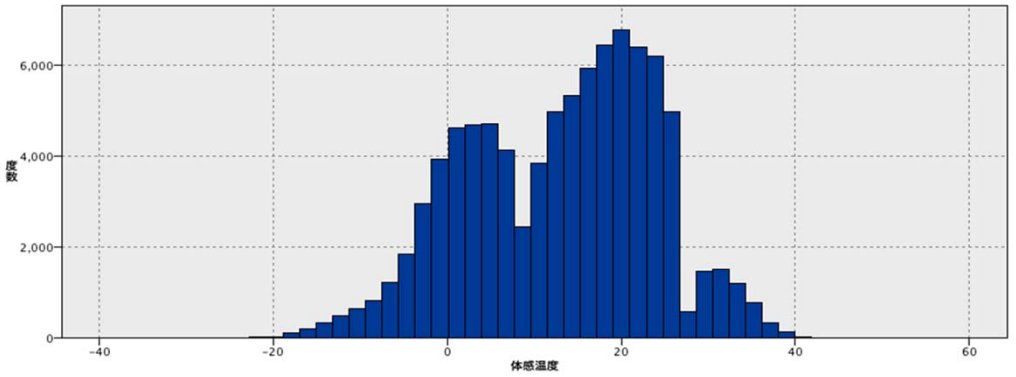
Background and Purpose—The purpose of the present study was to determine whether seasonal and monthly variations in stroke incidence exist and whether they are related to meteorologic and air pollution parameters under similar weather and environmental conditions in selected areas of Seongdong district, Seoul, South Korea.

Methods—From January 1, 2004, to December 31, 2013, 3001 consecutive stroke events were registered in residents of selected areas of Seongdong district, Seoul, South Korea. The authors calculated the stroke attack rate per 100 000 people per month and the relative risk of stroke incidence associated with meteorologic and air pollution parameters. We also analyzed odds ratios with a 95% confidence interval for seasonal and monthly stroke incidence.

Results—The incidence of stroke in September was significantly higher (odds ratio, 1.233; 95% confidence interval, 1.042–1.468) compared with January. The seasonal ischemic stroke incidence in summer (odds ratio, 1.183; 95% confidence interval, 1.056–1.345) was significantly higher than in winter, whereas the seasonal incidence of intracerebral hemorrhage relative to winter was not significant. The mean temperature was positively correlated with ischemic stroke (relative risk, 1.006; P=0.003), and nitrogen dioxide (relative risk, 1.262; P=0.001) showed a strong positive correlation with intracerebral hemorrhage incidence among the older age group.

Conclusions—We demonstrated distinct patterns of seasonal and monthly variation in the incidence of stroke and its subtypes through consideration of the meteorologic and air pollution parameters. We therefore expect that these findings may enhance our understanding of the relationships between stroke and weather and pollutants. (*Stroke*. 2015;46:927-935. DOI: 10.1161/STROKEAHA.114.007950.)

地域内発症頻度の予測因子ー救急搬送情報予備的解析
-脳卒中発症頻度と体感温度との関係-



救急搬送の時点の気候データ約１０万件から体感温度（気温、風速、湿度）により発症頻度が予測可能
-脳卒中、循環器の発症天気予報が可能
-地域消防での啓発活動などにも応用可能

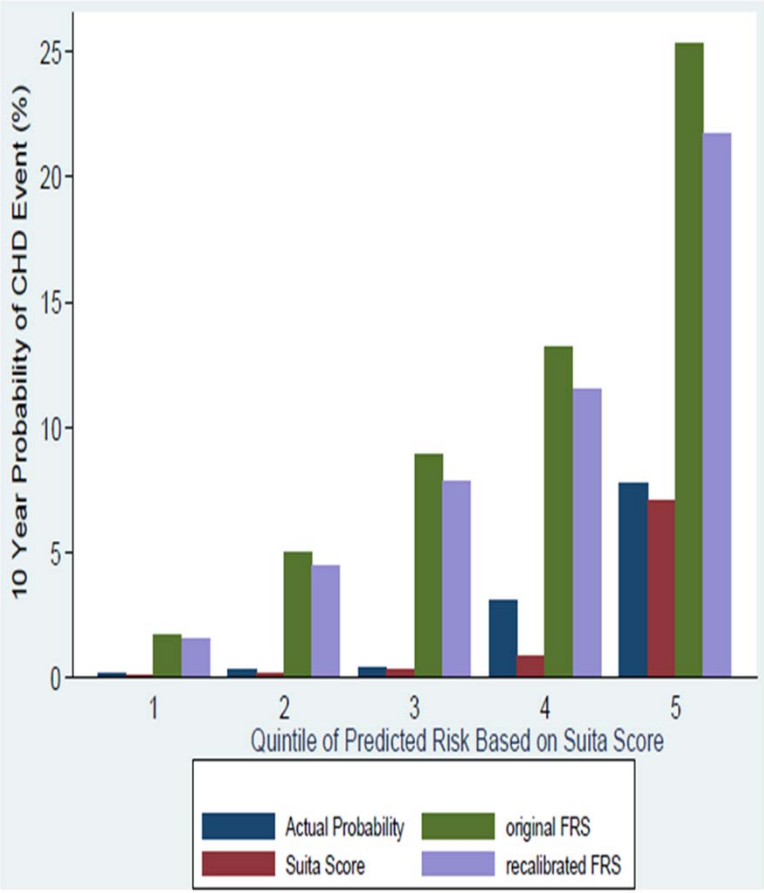
平たく言えば“循環器天気予報”を開発中

個人のリスク予測の精緻化

吹田研究による循環器リスクの定量化

Risk Factor	
Variable	Score
Age	
35-44	30
45-54	38
55-64	45
65-69	51
>=70	53
Female	−7
Current Smoker	5
DM	6
Blood pressure	
Optimal Blood pressure	−7
Stage 1 hypertension	4
Stage2 hypertension	6
LDL(mg/dl)	
<100	0
100-139	5
140-159	7
160-179	10
>=180	11
HDL(mg/dl)	
<40	0
40-59	−5
>=60	−6
CKD	
eGFR>60	0
Stage 3	3
Stage 4 or 5	14
Total Score	A

Predicted Probability of CHD in 10 years	
Total Score	Probability (%)
35<=	<1
36-40	1
41-45	2
46-50	3
51-55	5
56-60	9
61-65	14
66-70	22
71<	>28



J Atheroscler Thromb. 2014;21(8):784-98

日本人の冠動脈疾患リスクを算出するスコアとして動脈硬化学会のガイドラインで使用
同年の動脈硬化学会JAT賞を受賞　約7割5分は予測できるがあとのひとは外れる

研究成果の概要ー企業との共同研究（AI関連）
SRL　みらかホールディングスとの研究

従来のACS関連の予後予測

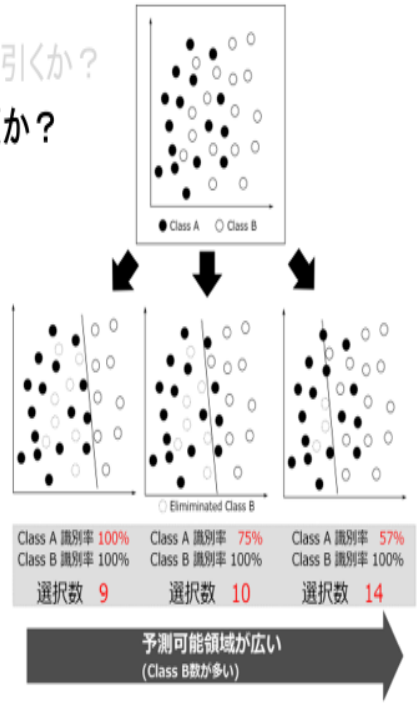
Table VI. Summary of Clinical Risk Scores for ACS, Modified After Morrow [15].

	PURSUIT		TIMI	GRACE	FRISC	
Population	UA/NSTEMI		UA/NSTEMI	All ACS	UA/NSTEMI	
Outcome	Death	Death/MI			Death	Death/MI
Key elements	5		7	8	7	
Age	X		X	X	X	
Gender	X				X	
Prior MI/CAD			X		X	
DM, CRF's			X		X	
Symptoms/History	X		X			
Use of aspirin			X			
Weight						
HR				X		
SBP				X		
CHF/Killip class	X			X		
ECG	X		X	X	X	
CKMB/cTn			X	X	X	
Serum Cr				X		
Serum Interl-6/CRP				X	X	
Cardiac Arrest						
Possible max score	18		7	372	7	
c-statistic	0.84	0.67	0.65	0.83	0.77	0.70
Computer needed				Yes		

- MIを含むACS患者の正確な予後を予測するのは従来のFramingham typeのリスク予測方式では限界（的中率70%-80%程度）

識別線のトレードオフ：予測力vs.識別精度

- どのような形の線を引くか？
- どの場所に線を引くか？

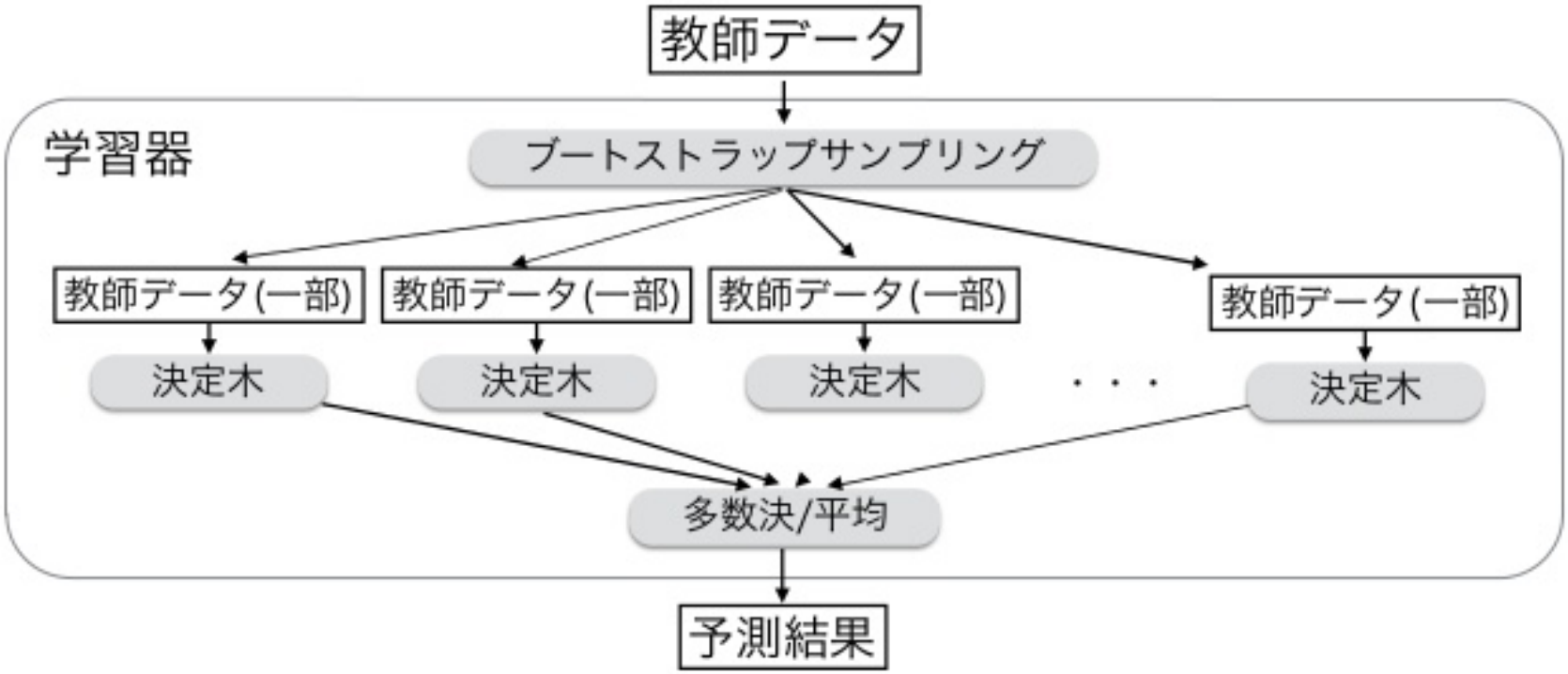


予測可能領域は広くなるが、
誤識別（エラー）の数も大きくなる

古典的な線形統計解析モデルでは予測精度に限界がある

Random Forestとは？

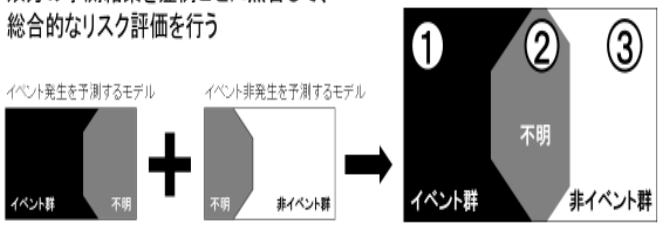
複数の**決定木**で**集団学習**させる事で、精度を高めた集団学習モデル。
2001年 Leo Breimanによって提案された。



研究成果の概要ー企業との共同研究（AI関連）
みらかホールディングスとの共同研究

目標＝双方向の予測モデルの結合

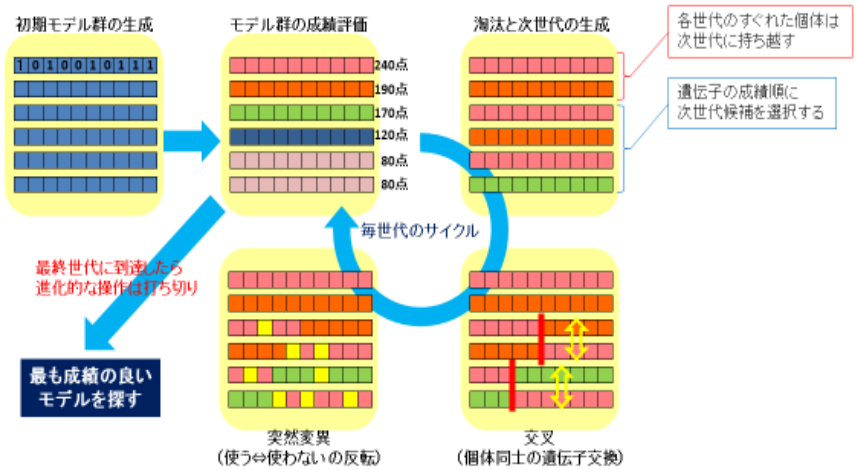
双方の予測結果を症例ごとに照合して、
総合的なリスク評価を行う



組	【イベント発生】 モデルの予測	【イベント非発生】 モデルの予測	総合判断	備考
①	YES	不明	イベント発生 リスク【高】	
②	不明	不明	（予測不明／判断保留） リスク【中?】	検査値が 重なる領域
③	不明	YES	リスク【低】	除外診断
④	YES	YES	予測エラー	なるべく少なく

遺伝的アルゴリズムによる最適化

① ランダムなデータと変数の組み合わせで作った初期モデル群から
② 進化的な操作によって徐々に集団全体の成績を上げてゆき
③ 最終世代のモデルの中から最良モデル（群）を得る

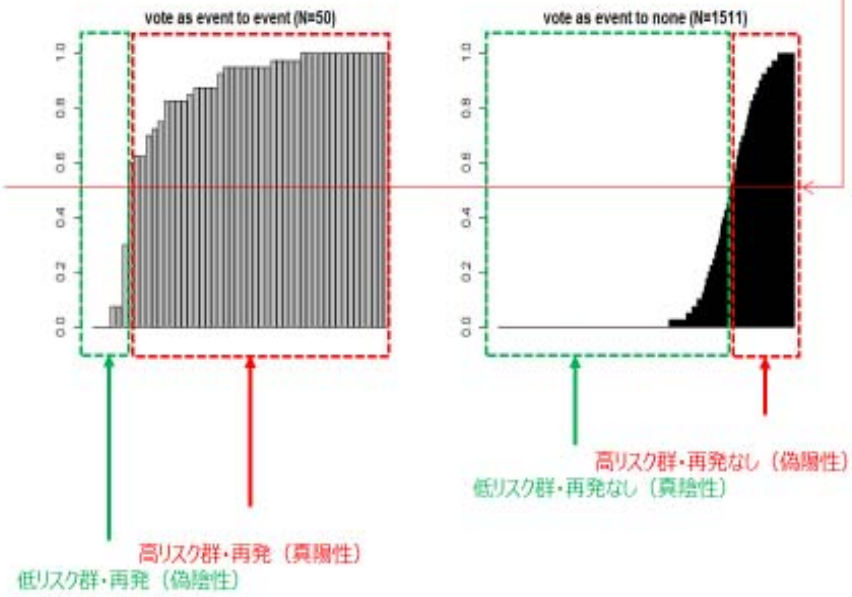


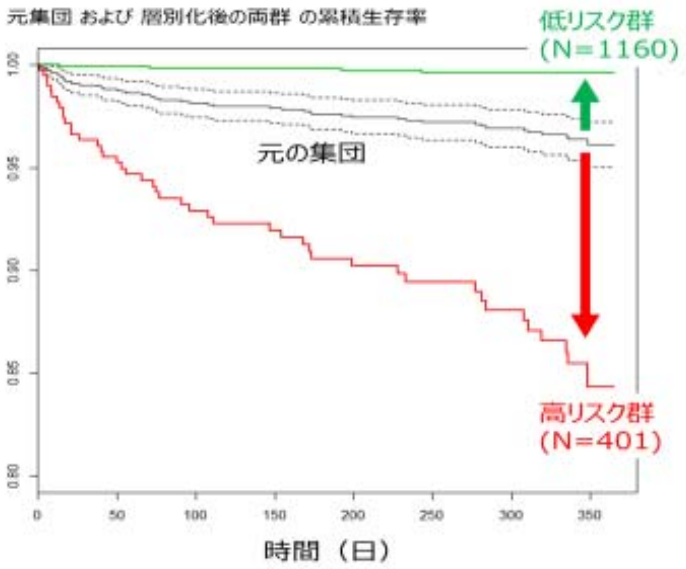
1つのSVM、Random Forestでは8割　モデル同士戦わせてさらに多数決
（特許出願中）
人工知能による予測モデル最適化の促進

研究成果の概要ー企業との共同研究（AI関連）
特許出願中

患者層別化

◆ 投票値を得点化して層別化
- 投票率≦0.5 ： 低リスク群
- 投票率＞0.5 ： 高リスク群



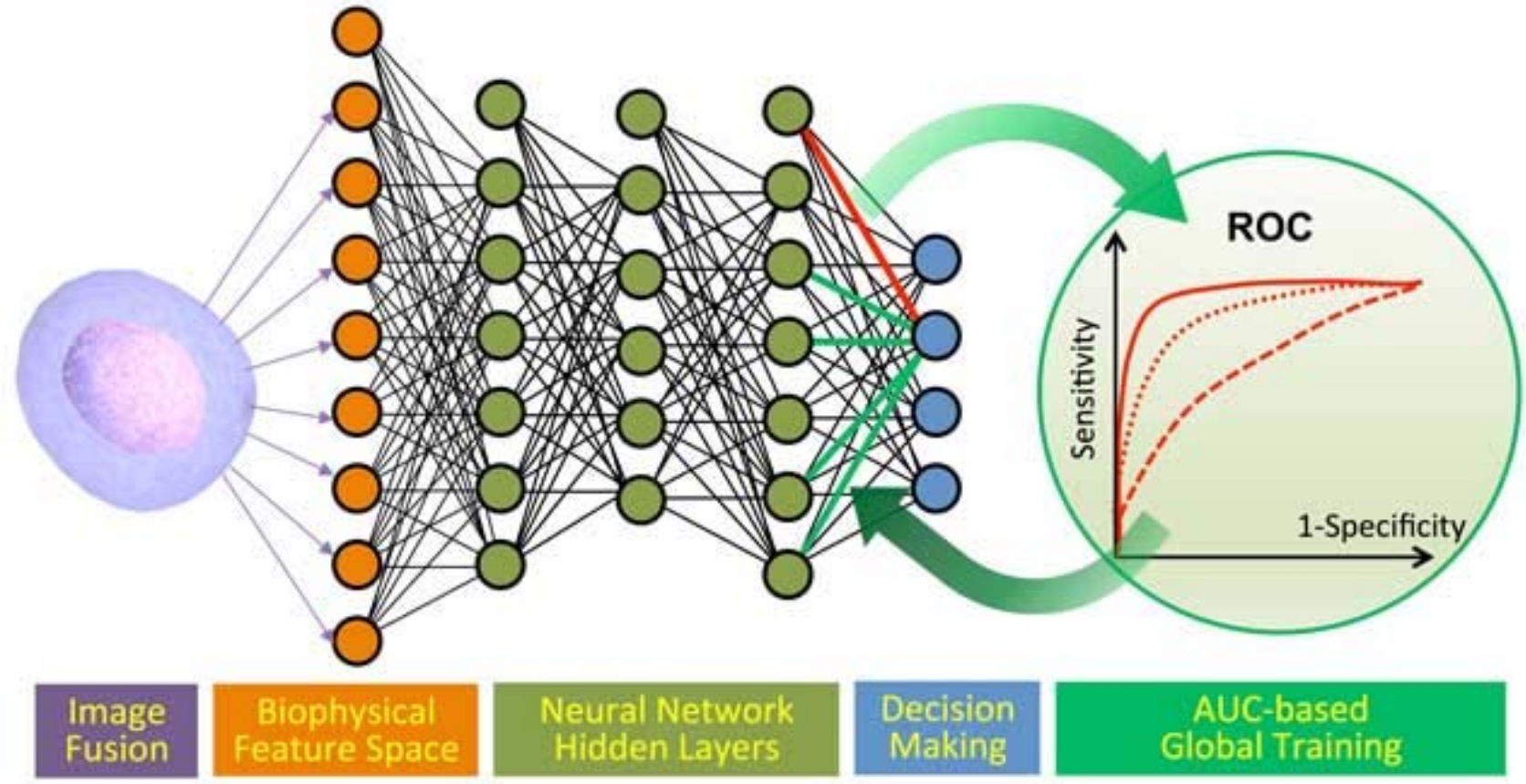


高リスク群　AUC　0.880（validation set）
低リスク群　AUC　0.879（validation set）

Deep learning

- ■DNN（Deep Neural Network，ディープニューラルネットワーク）
- ニューラルネットワーク（NN）というパターン認識をするように設計された、人間や動物の脳神経回路をモデルとしたアルゴリズムを多層構造化したもの。
- ■CNN（Convolutional Neural Network，畳み込みニューラルネットワーク）
- 局所的な情報の抽象化及び位置普遍性をもたせた順伝播型ニューラルネットワークを利用したアルゴリズム。DNNを2次元データに対応させたもので、画像に対して高いパターン認識能力を示します。
- 主な用途：画像認識
- ■RNN（Recurrent Neural Network，再帰型ニューラルネットワーク）
- 音声、動画データのような可変長のデータを扱えるようにするために中間層に再帰的な構造をもたせた双方向に信号が伝播するニューラルネットワークを利用したアルゴリズム。
- 主な用途：音声認識、動画認識、自然言語処理

Deep learning (イメージ図)



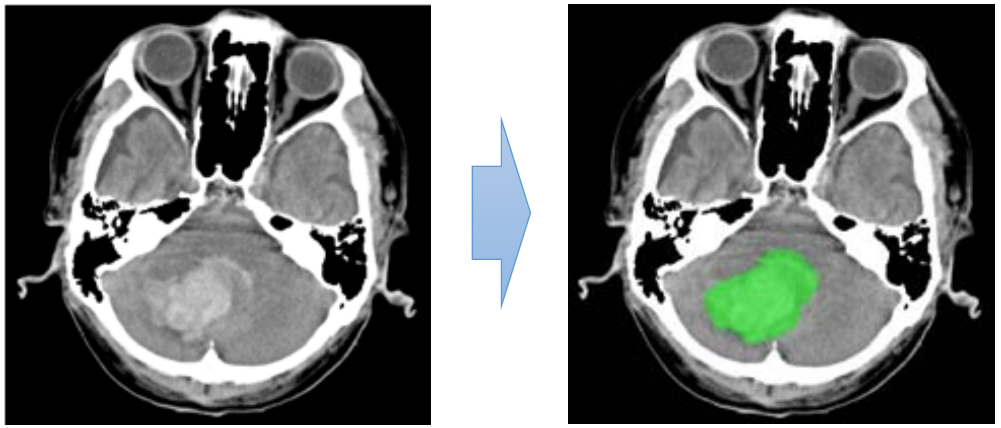
企業との共同研究（AI関連）

脳内出血種増大診断システムー九州大学、九州工業大学、富士フィルムとの共同研究

現状

■ 脳実質内出血における血腫体積を自動計測するアルゴリズムを開発

- 深層学習のモデル改良により、予定性能を概ね達成
  感度: 100%, 特異度: 95%, 臨床画像診断一致率: 74%（評価用データ：75症例）
- 2月より未知データによる検証を開始予定（PMDA事前面談すみ来年度承認申請予定）



今後の課題

■ 脳室内出血併発症例への対応（脳室内出血を体積算出対象外とする）

脳出血高リスク群を抽出する革新的な臨床診断意思決定システムの開発に関する研究
来年度診断補助機器として致死的脳内血種増大のリスク

自然言語処理によるデータベース作成

自然言語処理（ natural language processing：NLP）
人間がふだん日常的に使っている自然言語をコンピュータに理解させる技術

・JROAD、JROAD-DPCを土台とした包括的循環器病全国登録システムの確立
・J-ASPECT、日本脳卒中データバンクを土台とする包括的脳卒中全国登録システムの確立

目　標

1. 悉皆性の向上	2. データの拡充	3. データ活用の促進

事業内容

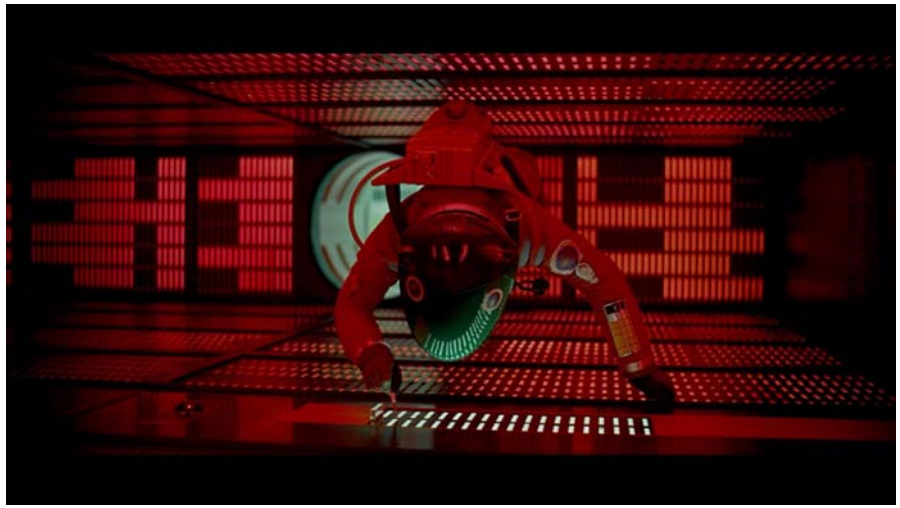
1. 脳卒中と循環器病の診療に対する医療の質評価指標の確立
2. 重要疾患、我が国独自の疾患の予後追跡調査システムの構築
3. 診療情報標準化とICTおよびAI技術によるデータ自動収集システムの構築
4. 医療計画、診療ガイドラインに資する脳卒中と循環器病の統合登録システムの構築
5. 民間企業を含めた予防治療開発研究への活用基盤の構築

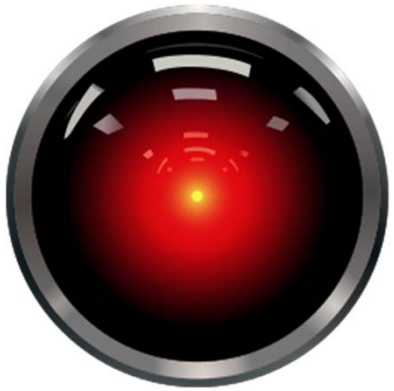
期待される成果

① 過剰医療の抑制による医療費の適正化
② 医療資源の適正配分に向けたヘルステクノロジーアセスメントの促進
③ 臨床試験治験、症例登録、市販後調査などへの活用
④ ＱＩ指標遵守率の向上による医療の質改善および循環器死亡減少の達成
⑤ 地域医療計画への情報活用
⑥ 高齢者医療の全体像把握
⑦ 登録データに基づくシミュレーション将来予測と医療政策評価

より詳細な臨床情報が研究、臨床試験への活用には必要、しかし多忙な医療関係者の手入力は困難

→欠損値、誤入力 の増加、登録数低下
→手入力による全国登録では件数が伸びにくいAIで自動入力できないだろうか？



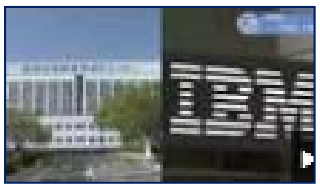


HAL ９０００
に入力してもらいたい

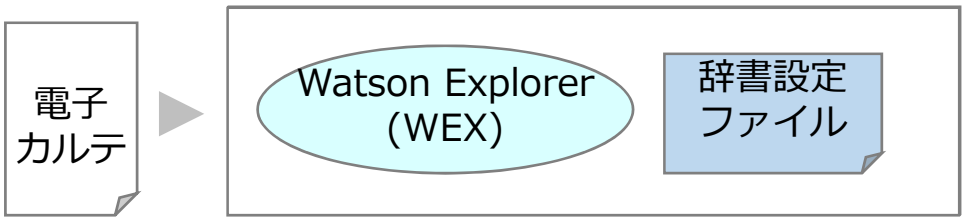
IBMワトソンプロジェクト

“心筋梗塞や脳卒中などの患者のカルテを人工知能に大量に分析させることで、将来的に、心臓や脳の血管の病気になるかどうかを予測するプロジェクトを、大阪・吹田市の国立循環器病研究センターなどが始めることになりました“

出所：2017年4月24日
NHKニュース



従来からの構造化データと、カルテから抽出した症候データの統合分析



循環器疾患の発症リスクの予測や重症化防止のため、人工知能（ＡＩ）を用いるプロジェクトを開始

2017年4月24日
国立循環器病研究センター

国立循環器病研究センター（大阪府吹田市、理事長：小川久雄、略称：国循）は、循環器疾患の発症リスクの予測や重篤化を防止する健康改善、治療モデルの構築に向け、膨大な様々な形式のデータをデータベース化し、人工知能（ＡＩ）を用いて解析、予測するプロジェクトを開始いたしました。
データベース化に関しては日本アイ・ビー・エム株式会社（東京都中央区、代表取締役社長：エリー・キーナン、以下「日本IBM」とします）の協力を得て実施いたします。

背景

高齢化が世界に例をみない速度で進む我が国において、脳卒中、心筋梗塞を中心とする心脳血管疾患は、日本人の死因の25.5%を占め、国民医療費の約20%を循環器病が占めています（がんは13%）。更に急性冠症候群（急性心筋梗塞・不安定狭心症・冠動突然死）はいまだ死亡率が高く、個人の生活の質を損なうだけでなく社会負担を著しく増加させることが知られており、平成20年の厚生労働省調査では、心筋梗塞・狭心症の患者数は80万人強と報告されています。更に高齢者医療費の最大の原因である脳卒中の患者数は今後も更に増加し、2020年には300万人に達すると予想されています。
このような状況のもとで、心脳血管疾患発症や重篤化のリスクを正確に予測し、適切なタイミングで適切な健康指導や治療を実施できるようにするためには、その症状や治療内容を正確に把握する必要がありますが、症状などは電子カルテなどに自由記述されているケースが多く、従来の技術ではデータベース化が困難でした。

研究手法と成果

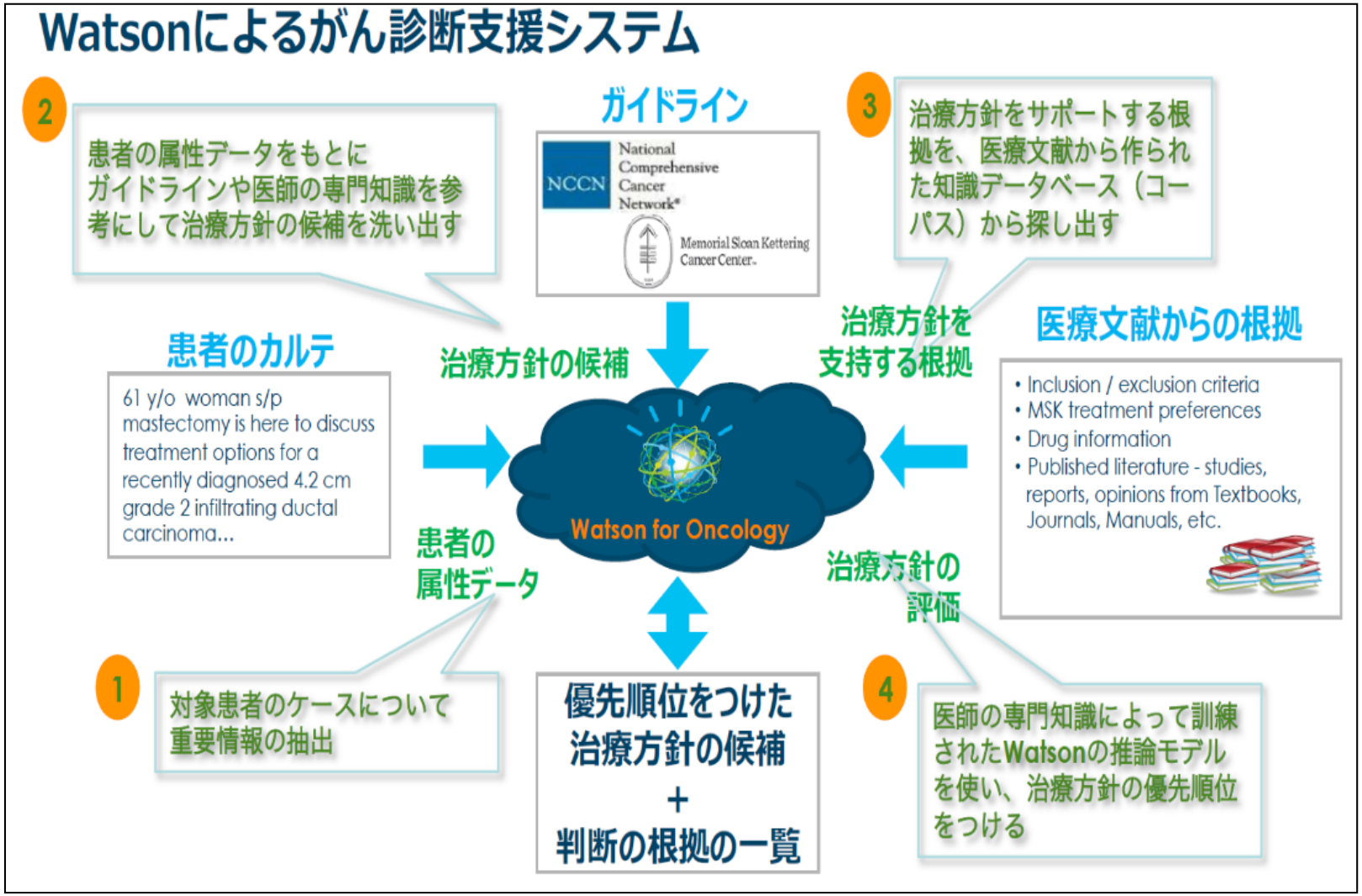
日本IBMとの取り組みにより、IBM Watson Explorerを用いておよそ1500人分の患者の電子カルテの記載から、胸痛、夜間発作性呼吸困難などの症状を自動的に抽出するロジックを確立しました。従来からあるGRACEスコアなどの心臓病の予後予測スコアに加えることで、心脳血管疾患の二次発症の予測能の向上が可能であることを、本年3月の日本循環器学会（金沢）において一部報告を行っております。

今後の展望と課題

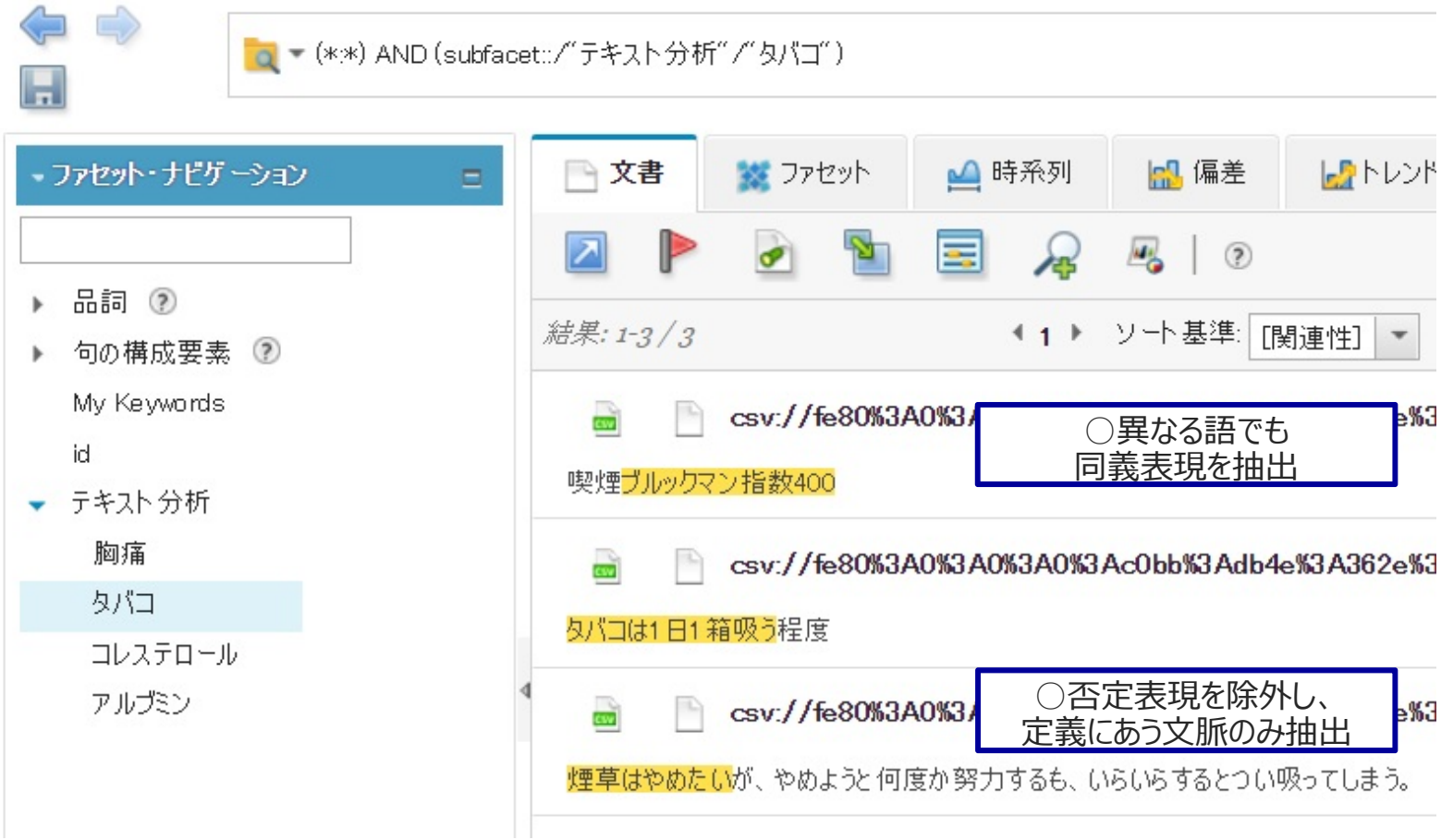
心脳血管疾患発症や重篤化のリスクを正確に予測し、適切なタイミングで適切な健康指導や治療を実施できるようにすることは、国民の健康寿命の延伸、医療費抑制の観点から重要であり、国循は国民の健康増進のためイノベーション創出に向けた取組みを進めていきます。
なお、本研究は、日本医療研究開発機構研究費（循環器疾患・糖尿病等生活習慣病対策実用化研究事業）「循環器疾患の発症予測・重症化予測に基づいた診療体系に関する研究」（研究代表者：安田聡（国立循環器病研究センター副院長））にて行われました。

※IBM Watson は、世界の多くの国で登録されたInternational Business Machines Corporation の商標です。

Watson 研究・医療応用



IBM Watson Explorer



心臓エコー・カテーテル検査値抽出

- WEX辞書チューニングを行い、退院サマリーの検査所見から心臓エコー検査値23因子、心臓カテーテル検査値22因子を抽出

心臓エコー

No	因子名
1	IVSd
2	LVPWd
3	LVDs
4	AoDd
5	LADs
6	LVDd
7	RVDd
8	LAV(m_Simpson)
9	DcT
10	MR
11	AR
12	TR
13	TR_flow_peak_PG
14	IVC呼吸性変動
15	TAPSE
16	EF(m_Simpson)
17	IVCφ_短径
18	IVCφ_長径
19	E
20	A
21	E／A
22	E／e’（average）
23	effusion

心臓カテーテル

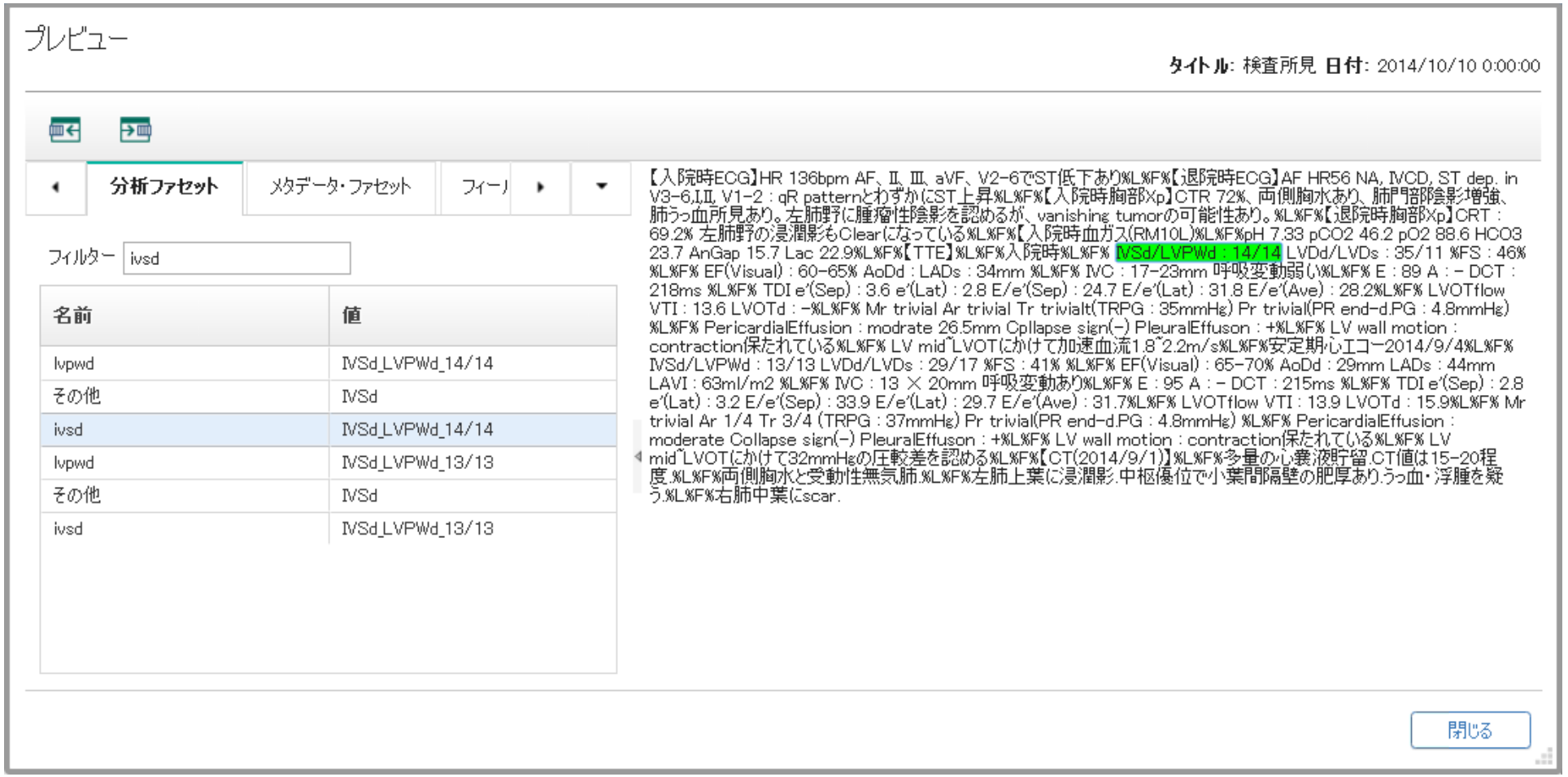
No	因子名	備考
1	LAD_6	
2	LAD_7	
3	LAD_8	
4	LCA_5	
5	LCx 11	
6	LCx 12	
7	LCx 12_1	
8	LCx 12_2	
9	LCx 13	
10	RCA-1	
11	RCA-3	
12	RCA-2	
13	LCx 14-1	
14	LCx 14-2	
15	LCX 14-3	
16	RCA-4PD	
17	RCA-4AV	
18	RCA-4AV-2	対象文章がなく、未実施
19	LAD_D1_9	
20	LAD_D1_91	
21	LAD_D1_92	
22	LAD_D2_10	
23	LAD_D2_101	
24	LAD_D2_102	対象文章がなく、未実施

No	因子名	備考
25	EDV	
26	EDVI	
27	ESV	
28	ESVI	
29	SV	
30	SVI	
31	LVEF	
32	RA(mean)	
33	PAmain(sys)	対象文章がなく、未実施
34	PAmain(dia)	対象文章がなく、未実施
35	Pamain (sys/dia/mean)	
36	PCW(mean)	
37	CO(Thermo)	
38	CO(Fick)	
39	CI(Thermo)	
40	CI(Fick)	
41	HR(Thermo)	対象文章がなく、未実施
42	HR(Fick)	対象文章がなく、未実施
43	HR	

心臓エコー・カテーテル検査値抽出

- 心臓エコー検査値の抽出結果例として、IVSDの抽出結果の一部を示します。

心臓エコー IVSD　左室拡張終末期径



ピナクルレジストリ日付因子の抽出

- WEX辞書チューニングを行い、退院サマリーよりピナクルレジストリ日付項目に関する、23因子を抽出しました。
- 現在のWEXの仕様上、抽出結果に検査値など日付以外の数詞が含まれる場合がございます。
  そこで、WEX以外のプログラムを用い、抽出結果のフィルタリングを行い、抽出精度を向上させています。

No	因子名
1	冠動脈バイパス術日
2	頭蓋内出血診断日
3	不安定狭心症診断日
4	安定狭心症診断日
5	心房細動診断日
6	心房粗動診断日
7	虚血性脳卒中診断日
8	出血性脳卒中診断日
9	CRT-D植込み日
10	CRT植込み日
11	ICD植込み日
12	心不全診断日

No	因子名
13	心筋梗塞診断日
14	一過性脳虚血性発作診断日
15	冠動脈疾患診断日
16	末梢動脈疾患診断日
17	冠動脈形成術－ベアメタルステント実施日
18	冠動脈形成術－薬剤溶出性ステント実施日
19	冠動脈形成術－ステントなし実施日
20	弁形成術あるいは弁置換術日
21	高血圧診断日
22	糖尿病診断日
23	冠動脈形成術－ステント種類不明実施日

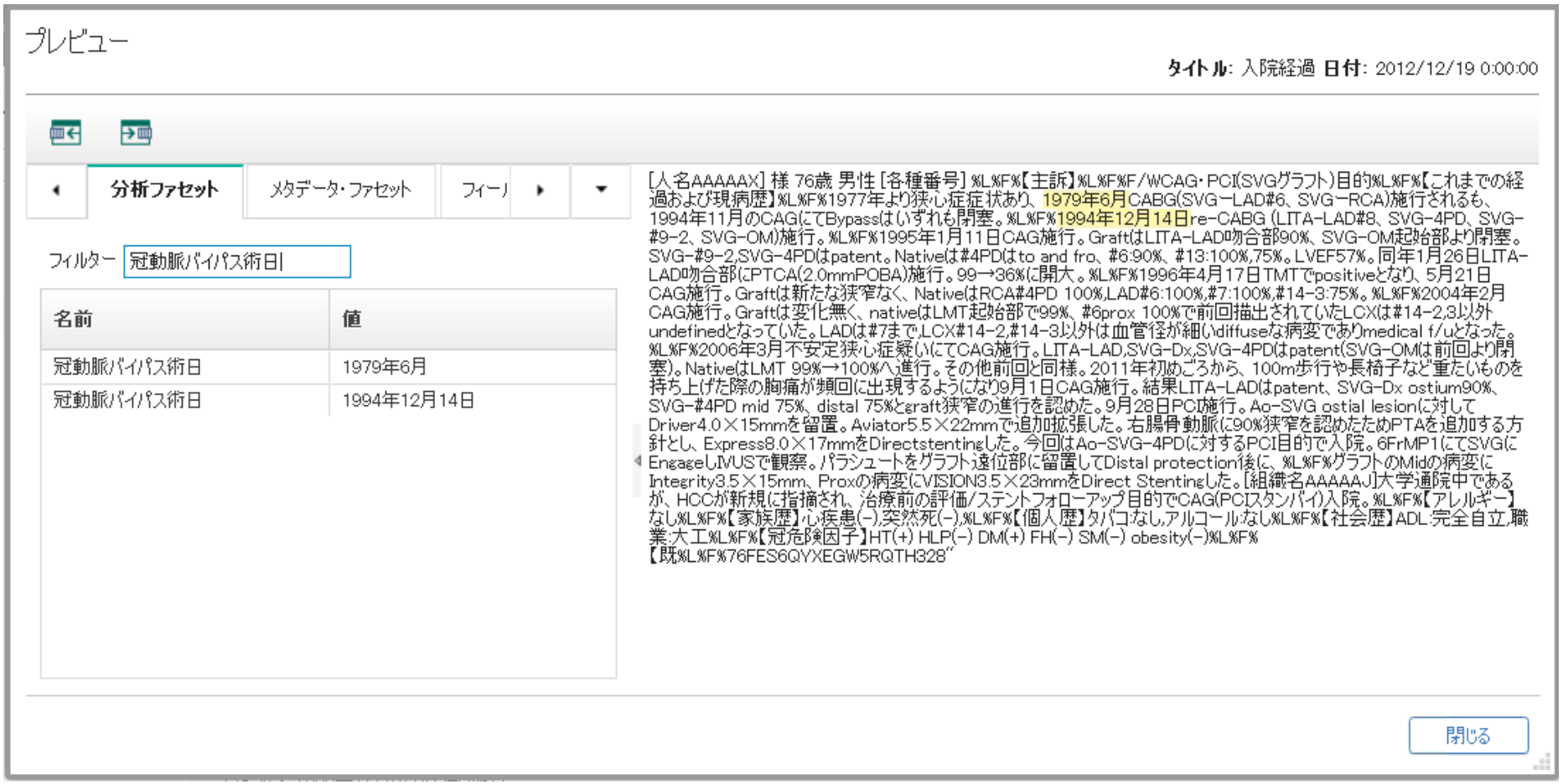
抽出結果のフィルタリング条件

1. 下記の日付を意味する文字列を含む
   – 年・月・日・H・平成・昭和・大正
2. 下記の日付形式の表記を含む
   – スラッシュでの記載：yyyy/MM/dd または MM/dd
   – ハイフンでの記載：yyyy-MM-dd または MM-dd
   – ピリオドでの記載：yyyy.MM.dd または MM.dd

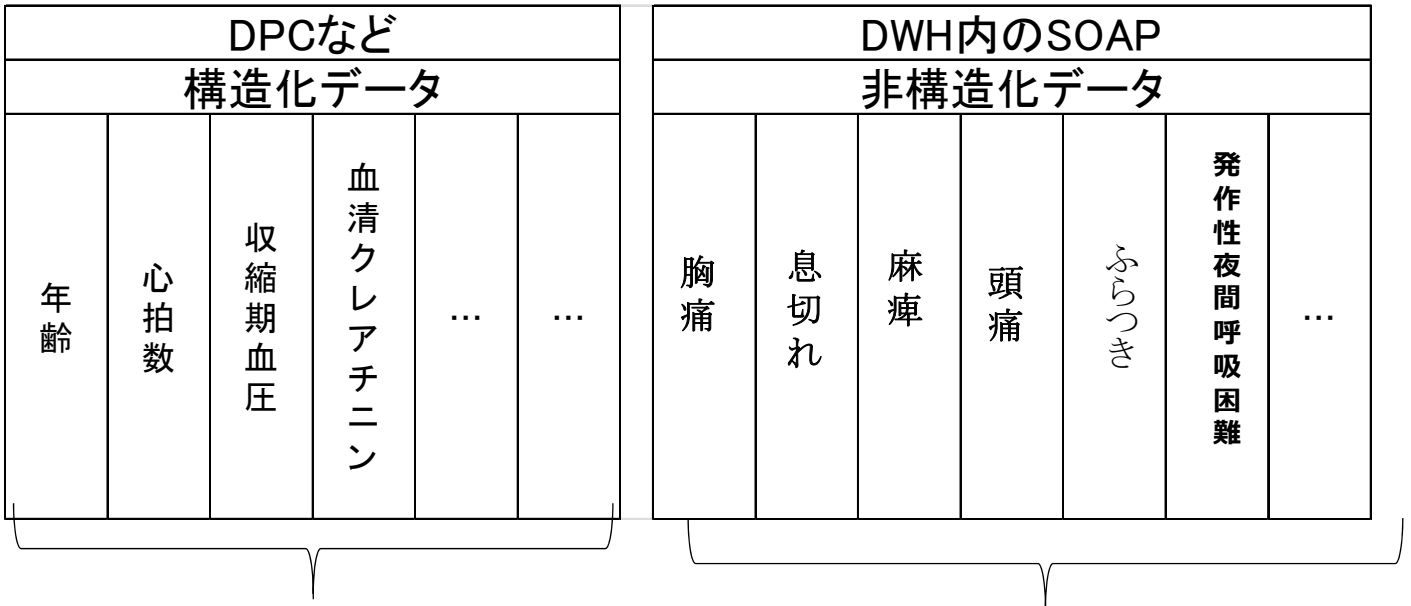
ピナクルレジストリ日付因子の抽出

- ピナクルレジストリ日付因子の抽出結果例として、冠動脈バイパス術日の抽出結果の一部を示します。

冠動脈バイパス術日



人工知能（AI）の活用によるData抽出



DPC, SS-MIXなどから
の従来型データ抽出

人工知能の自然言語処理
辞書の作成（AIに指示）
⇓
SOAP文章理解
（臨床家が正解、不正解をAIに教育）
⇓
SOAPからのデータ抽出・確認

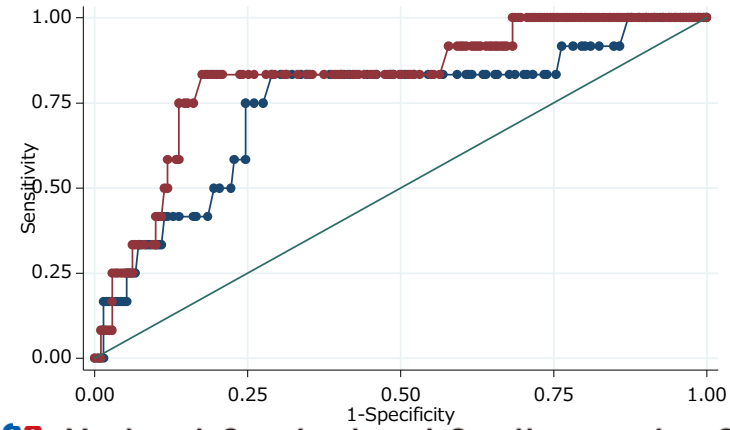
除外条件、略語の文脈を踏まえた処理（例 CP → 胸痛 なのかCost phrenic angle なのか),日本語特有の否定語、述語の位置が離れるなどの複雑な条件の処理を行う

AIによる症候抽出と予後予測の改善

- 症状の有無（大項目8項目、小項目25項目）について予測因子としての有用性を検討した（年度末時点では12項目）
- 専門医2名、統計学者2名が辞書チューニングを行い、ＡＩによる教師あり学習をサポートした。
- MIDAS研究の約1500人を対象として、入院中に自由記載されたカルテのSOAP記事を使用した（看護記録含む）
- 約60万行のカルテ記事を読み込み、症候の出現頻度を患者ごとに集積した
- 文章単位での解析が可能なため症候の出現時期、時間が同定可能であった
- 辞書チューニングの過程で抽出された構文からは、看護師、医学部生、研修医程度の精度の症候抽出は可能であった
- 重複を除いた１万４千件確認時点では問題になる記述はない

- データ取得対象者中ＡＭＩ患者(Ｎ＝225例)
- アウトカムは心臓死+ＭＩ再発+stroke
- Grace本来は心筋梗塞ー院内死亡、一年後死亡予測

変数	HR(95%CI)	P-value
GRACE スコア	1.02(1.01,1.03)	<0.001
いびき（他覚的）	1.44(0.96,2.16)	0.078
胸痛（放散痛）	0.34(0.10,1.16)	0.083
呼吸困難感	1.02(1.00,1.03)	0.019
浮腫	1.04(1.00,1.07)	0.042
無尿	1.15(1.00,1.32)	0.050
睡眠時無呼吸	1.04(1.01,1.07)	0.020

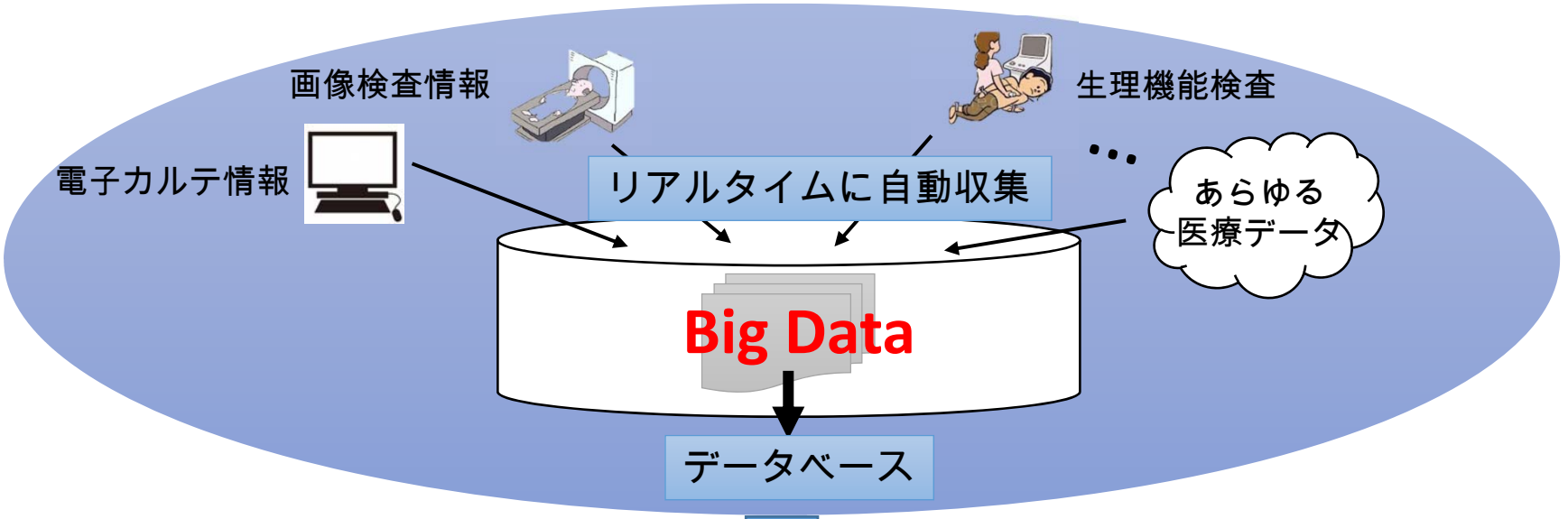


電子カルテ情報の活用（国循院内レジストリの構築）

「医療情報（リソース）の断片化」は臨床研究における情報科学の効率的な使用を妨げ、医療ビッグデータ研究の実施とそこからの社会還元へのプロセスの大きな課題となっている。

国循院内レジストリの３つの特徴

1. 断片化されたあらゆる診療情報を自動抽出しリアルタイムにDB構築
2. 米国心臓病学会主導のレジストリ項目（PINNACLE）に準拠
3. 診療科横断的DBを実現。院内外の臨床研究とリンク

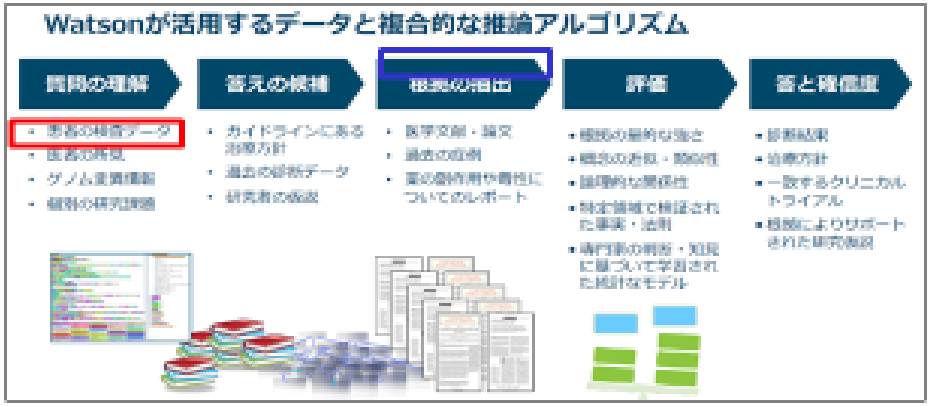


- **電子カルテから完全に自動抽出可能な項目はACC/AHA NCDR　PINNACLE registry 約400項目中　５割強程度　残りは手入力が必要**
  （AMED　安田班報告書　国循　安田先生のご厚意による）

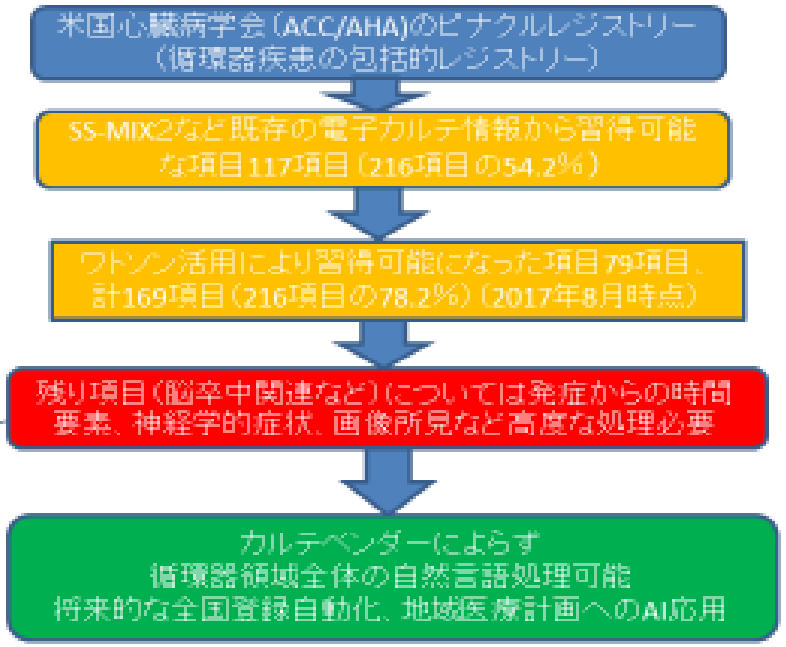
電子カルテ中のチェック項目の自然言語処理による自動登録システムを活用したAI supported Guideline

- カルテテキストデータ分析の実現により人工知能活用に向けた重要な第一歩を踏み出した（循環器領域では国内では皆無）
- 次なる一手は、「ゲノム、論文等の活用」「画像＋テキスト、検査値＋テキスト、音声＋テキスト分析等、異質データの統合分析」により、予後予測や効率的・創造的な創薬に役立てること

→ 現在日本においては、健康食品等での民間資本でのAI参入はある（ダイエット等）
→ 循環器・脳卒中（特に介護、寝たきり）は国民医療費でがん全体に匹敵する
→ 介護、寝たきり状態のケアは公的医療費で賄われており、影響が長期的であることから、民間資本の短期的投資スキームにはそぐわない
→ 公的医療費削減のためには、公正中立なナショナルセンターによるAI開発が必要



出所：日本IBM講演資料



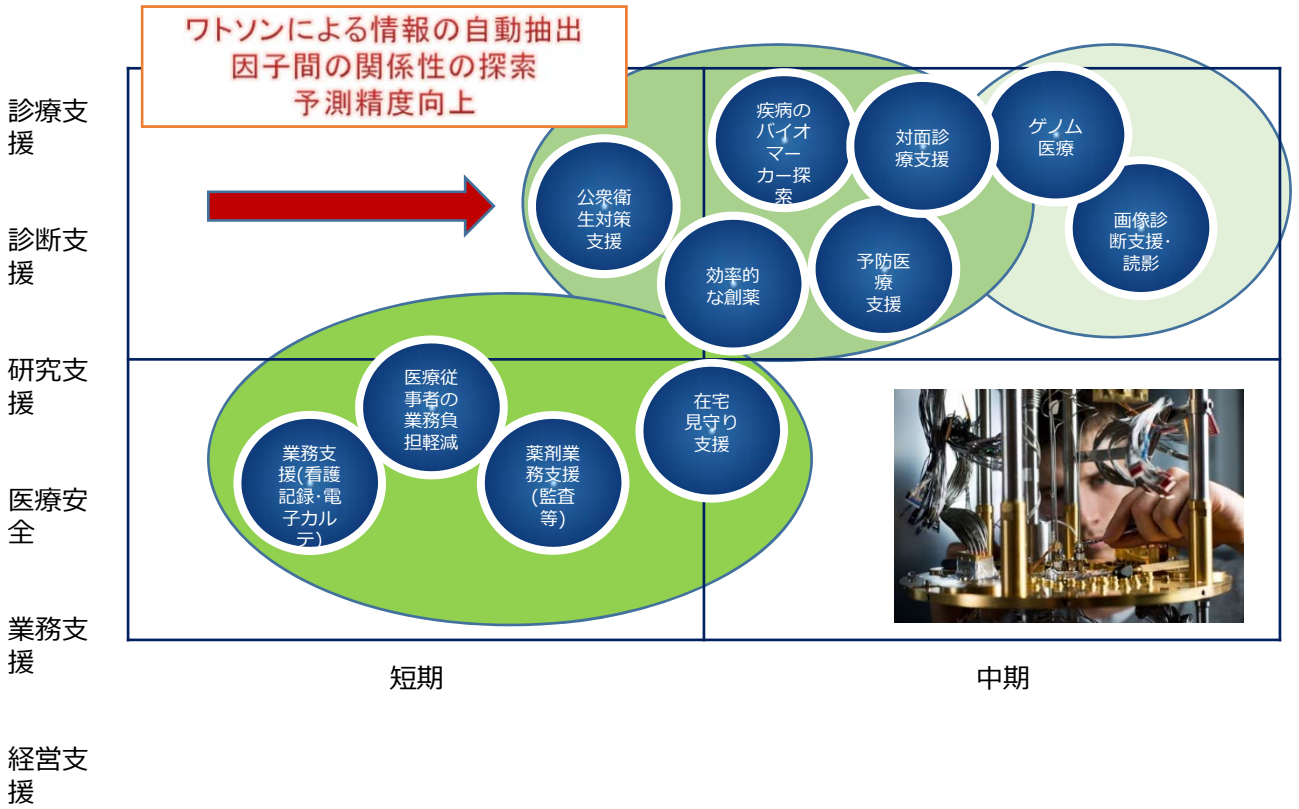
1.国循　循環器疾患（脳卒中、大動脈疾患等含む）に関して、電子カルテ情報を機動的に収集可能（症例数としては各疾患で大学病院数施設/年間）
2.脳、心臓は相互に合併、続発が多いが、国循は診療科間の垣根が低く相互連携が容易
3.すでに診療科医師と疫学者、統計学者、医療情報部の間で高度な自然言語処理の辞書に関しての経験を積んでおり、国内最速で循環器領域の情報自動抽出システムの構築が可能

SS-MIX+ワトソン　循環器では１年で必要項目の85%まで抽出可能となった　追加で半年から１年で自動化達成の見込み

現在　心エコー、心臓カテーテル検査に関する所見約60項目が年度内に完成予定

III.　将来像－人工知能を応用した研究

国循：AIによるコグニティブコンピューテイング活用が想定される領域



現在共同研究中のみらかホールデイングス（SRL社）、日本IBM社、SAS社などに加え
フィリップスなどOICへの入居を検討中の企業とも幅広く協力を予定

ご清聴ありがとうございました