

Uso de Técnicas de Inteligência Artificial e de Modelagem Conceitual Semântica aplicadas à análise de Big Data na área da Saúde

Grupo de Pesquisa em Tecnologias da Informação e Comunicação
Aplicadas à Saúde

- Modelagem Conceitual Baseada em Ontologias
- Big Data em Saúde
- Por que Deep Learning?
- Aplicações em Saúde

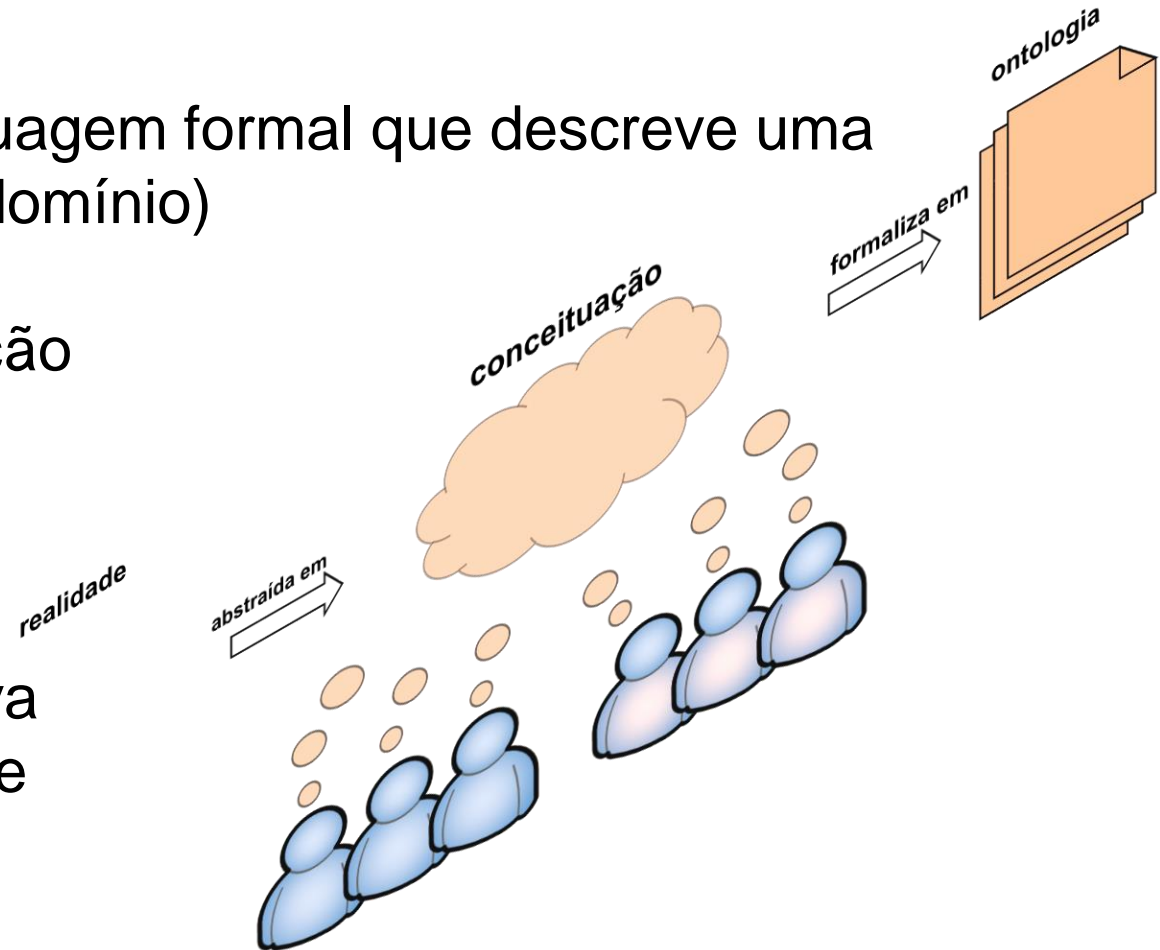
ONTOLOGIAS: O QUE SÃO E PARA QUE SERVEM

Uma especificação formal e explícita de uma conceituação compartilhada

Artefato representado em uma linguagem formal que descreve uma determinada porção da realidade (domínio)

Teoria sobre uma determinada porção da realidade deve ser aceita e compartilhada por uma comunidade de interessados

A conceituação sobre o domínio leva em conta estrutura, funcionalidade e comportamento



Teorias ontológicas formais desenvolvidas e aplicadas na solução de problemas nos campos da Ciência da Computação, Ciência da Informação e, em especial, **Modelagem Conceitual**

Ontologias em Ciência da Computação: Inteligência Artificial, Engenharia de Software, Bancos de Dados, Web Semântica

Um artefato, constituído de conceitos e relações, suas definições e um conjunto de axiomas formalizados, usados para criar novas relações e para restringir as suas interpretações

Uma ontologia é uma descrição **parcial** e **explícita** de uma **conceituação compartilhada**

Uma conceituação, por sua vez, corresponde ao conjunto de conceitos e relações usados para **interligar abstrações de entidades de um dado universo de discurso**

Uma ontologia é uma especificação conceitual que descreve o conhecimento sobre um universo de discurso. Ela define um **vocabulário específico** usado para descrever uma certa realidade e um conjunto de **decisões explícitas para estabelecer precisamente o significado pretendido desse vocabulário**

ONTOLOGIA PARA SOLUÇÃO EM SAÚDE

Modelo conceitual dos domínios relacionados com o SUS

Domínios estudados por unidade operacional e integrados continuamente

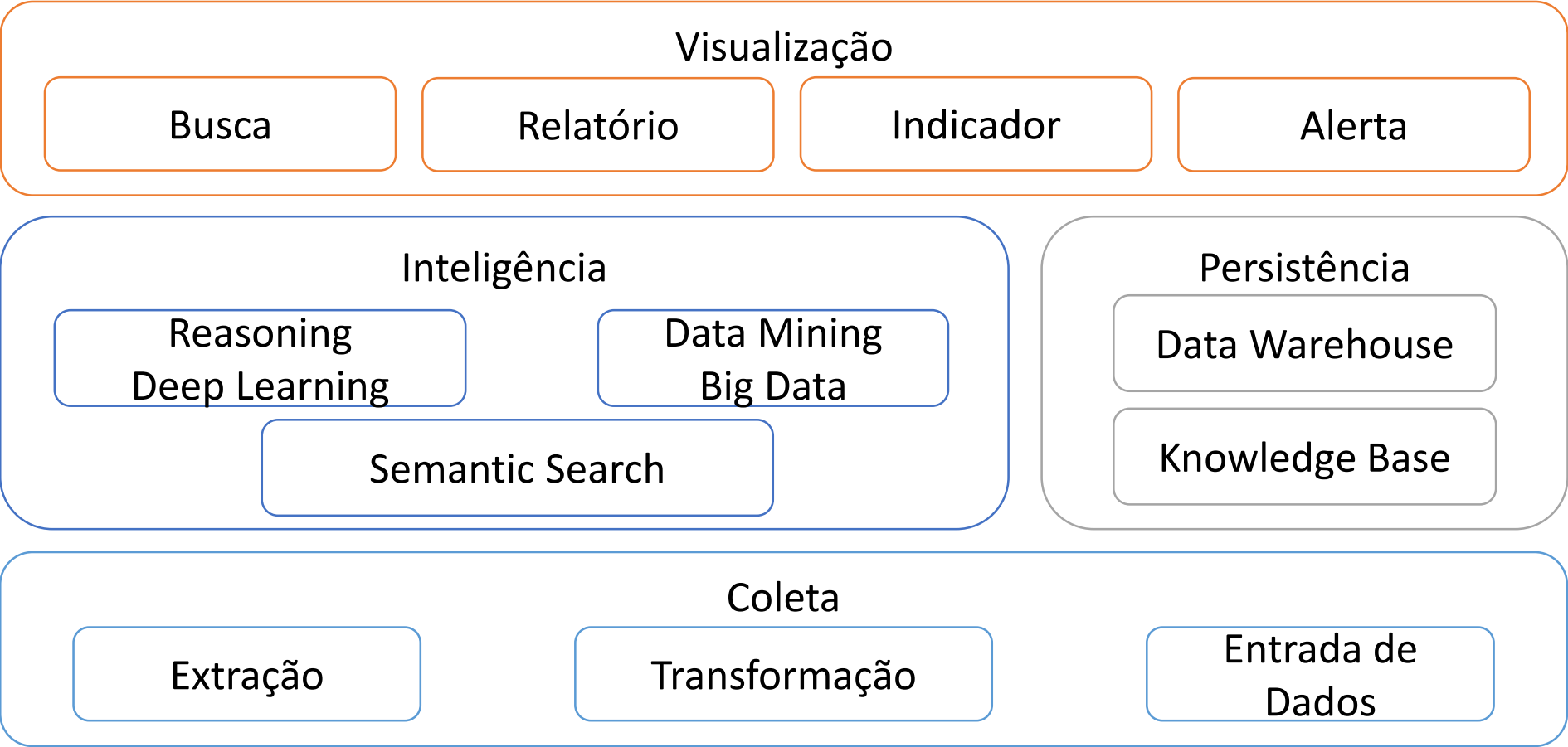
Uma análise geral leva à construção de uma ontologia global

Deve ser construída com base nos objetivos do SUS tanto nos aspectos estratégicos, como operacionais

Modelagem Conceitual

- Ontologias na Gestão Inteligente em Sistema de Saúde
 - Ontologias para Integração
 - Ontologias para Inferências
 - Ontologias para Busca

Exemplos de Componentes da Gestão Inteligente em Sistema de Saúde



Ontologias para Integração

- Problema:
 - Informação de baixa qualidade para tomada de decisão no contexto de saúde pública
 - Exemplo Prático:

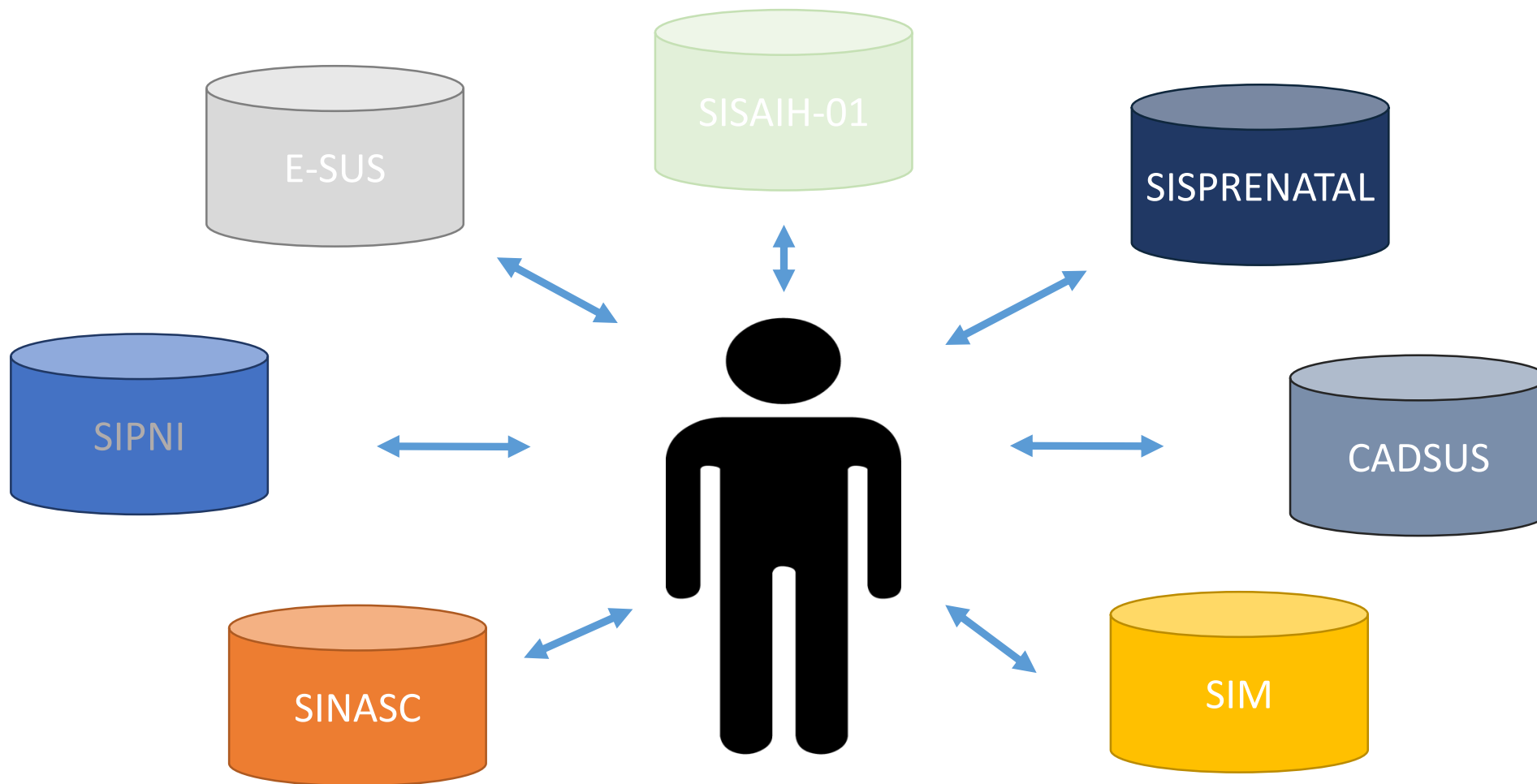
Para identificar o motivo do aumento da ocorrência de partos prematuros, é importante analisar dados clínicos, sociais e do sistema de saúde, os quais estão espalhados em diferentes sistemas

- Dados clínicos: histórico de gestações anteriores, doenças
- Dados sociais: condição de moradia, renda familiar
- Dados do sistema: cobertura vacinal, oferta de profissionais

Ontologias para Integração

- Cenário tecnológico:
 - Informações sobre pessoas estão espalhadas por diversos sistemas públicos de saúde que **não trocam informações**
 - Sistemas com diferentes arquiteturas e tecnologias:
 - Bases locais, integradas ou na nuvem
 - Sistemas desenvolvidos por diferentes profissionais
 - Sistemas projetados para operar sozinhos
 - Ausência de código que identifique uma pessoa em todos sistemas – CNS não utilizado na maioria dos casos

Ontologias para Integração



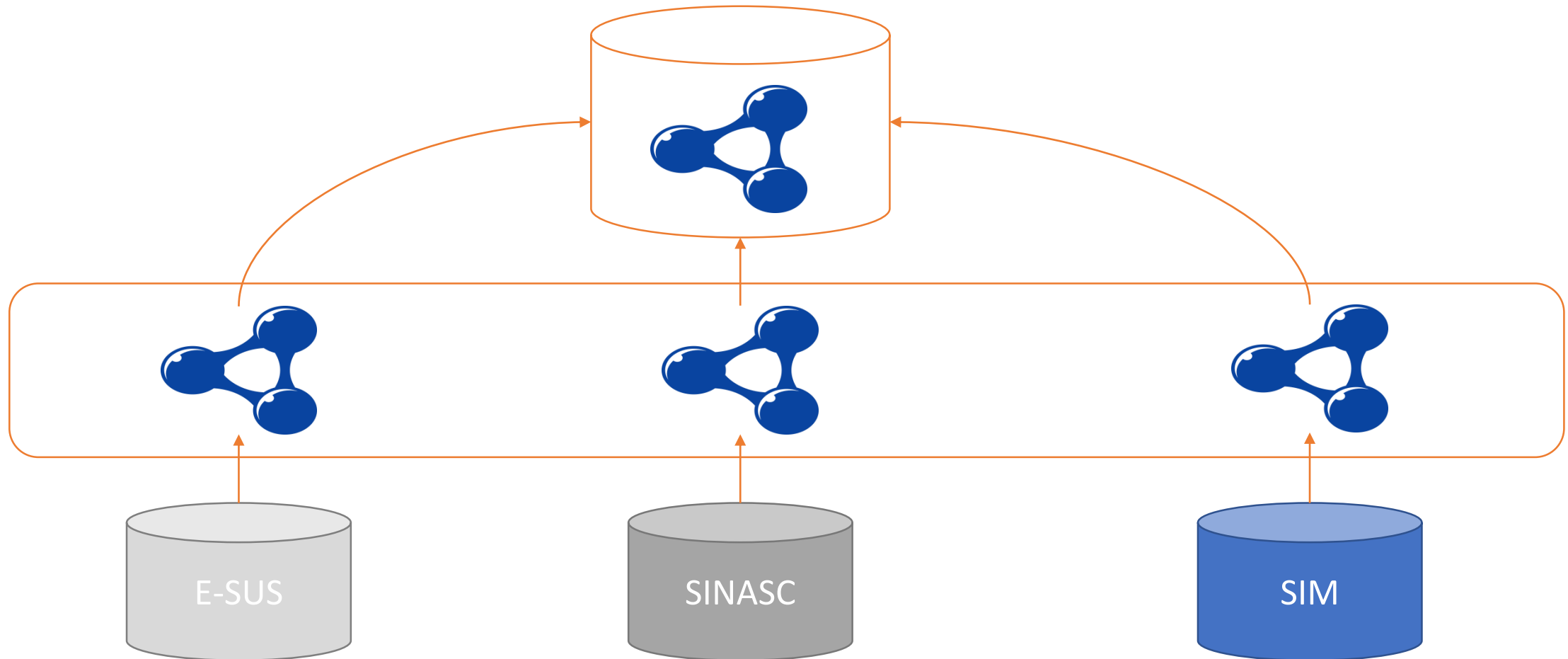
Ontologias para Integração

- Alguns Sistemas Utilizados:
 - **SIM**: Sistema de Informação de Mortalidade
 - **SINASC**: Sistema de Informação de Nascidos Vivos
 - **E-SUS AB**: Sistema de apoio aos serviços de atenção básica à saúde
 - **SI-PNI**: Sistema de Informação do Programa Nacional de Imunização
 - **SISPRENATAL**: Sistema de Informação de Pré-Natal
 - **CADSUS**: Cadastro Único do Cidação
 - **SISAIH-01**: Sistema para internações hospitalares

Ontologias para Integração

- O papel da Ontologia na integração dos sistemas
 - Para o problema **semântico**:
 - Modelo de referência independente de tecnologia (OntoUML)
 - Define a semântica dos dados de forma precisa e inequívoca
 - Para o problema **sintático**:
 - Modelo de implementação (RDF/OWL)
 - Provê um vocabulário comum para troca de informações
 - Permite a identificação de relações entre indivíduos
 - Permite a criação de *mashups*

Ontologias para Integração



Ontologias para Inferências

- Problema:
 - Conhecendo os **fatores de risco** da população acompanhada pela rede cegonha, é possível agir de forma **preventiva** e evitar óbitos materno-infantis
 - Ex: O primeiro filho (risco 1) de uma mãe com baixa escolaridade (risco 2), com histórico de tabagismo (risco 3) que mora em uma região não coberta pelo programa de saúde da família (risco 4) e possui um cobertura vacinal contra sarampo baixa (risco 5)

Ontologias para Inferências

- Solução:
 - Fornecer para os gestores de saúde pública ferramentas de controle dos riscos para **gestantes, puérperas, e crianças até 2 anos.**
 - **Dashboards** para informações analíticas
 - **Relatório** para informações mais detalhadas

Estado

Município

Período

Filtrar

Risco de Óbito Materno

50
GESTANTES

Risco Alto

Risco de Óbito Infantil

100
CRIANÇAS

Risco Medio

Risco de Parto Prematuro

100
GESTANTES

Risco Baixo

Risco de Óbito Fetal

52
BEBÊS

Risco Alto

Estado

Município

Período

Filtrar

Risco de Óbito Materno

Risco Clínico

Risco	Quantidade
Hipertensão	3
Descolamento de Placenta	2
...	

20
GESTANTES

Risco Alto

Risco Social

Risco	Quantidade
Baixa Escolaridade	3
Gravidez na Adolescência	2
...	

23
GESTANTES

Risco Médio

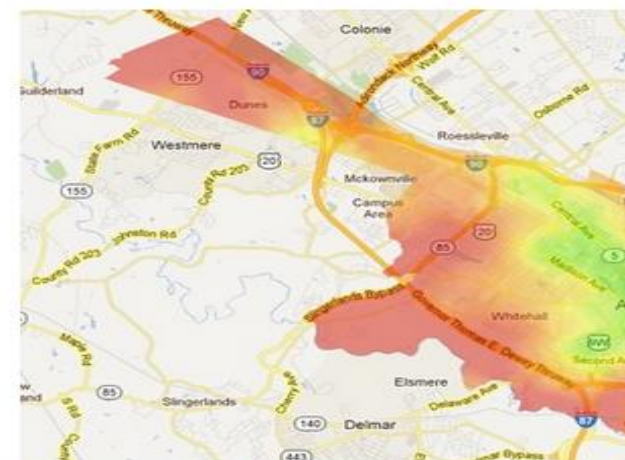
Risco do Sistema de Saúde

Risco	Quantidade
Ausência de Medicamento: Ácido Fólico	20
Área sem equipe de Saúde da Família	4
...	

3
GESTANTES

Risco Baixo

Mapa de Risco



Risco Alto

Risco Médio

Risco Baixo

BIG DATA EM SAÚDE

Big Data em Saúde

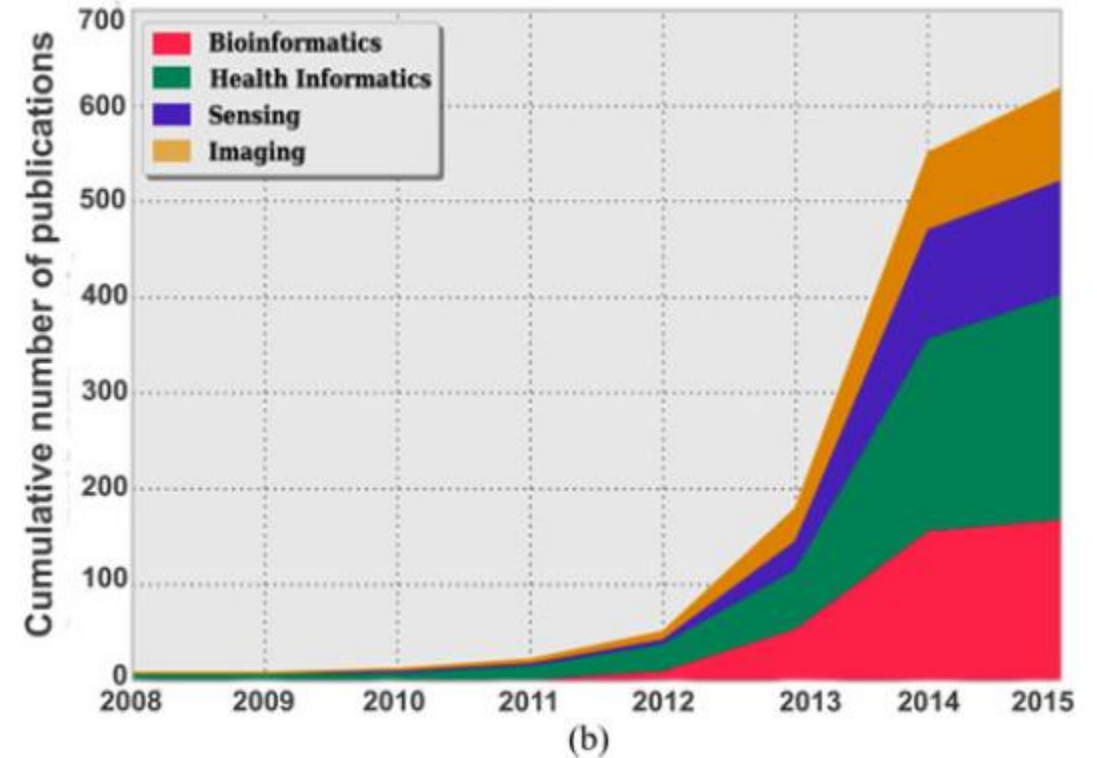
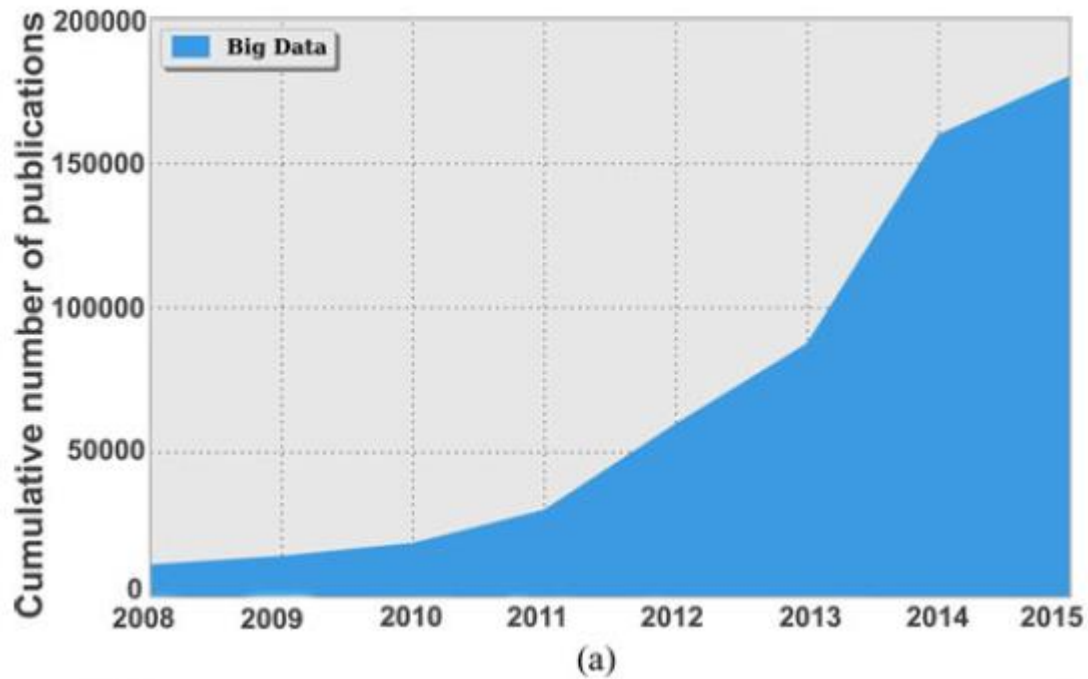


Fig. 1. (a) Cumulative number of publications referring to “big data” indexed by Google Scholar. (b) Cumulative number of publications per health research area referring to “big data,” as indexed in IEEE Xplore, ACM Digital library, PubMed (National Library of Medicine, Bethesda, MD), Web of Science, and Scopus.

Big Data em Saúde



Fig. 2. Six V's of big data (value, volume, velocity, variety, veracity, and variability), which also apply to health data.

Big Data em Saúde – Um exemplo de aplicação

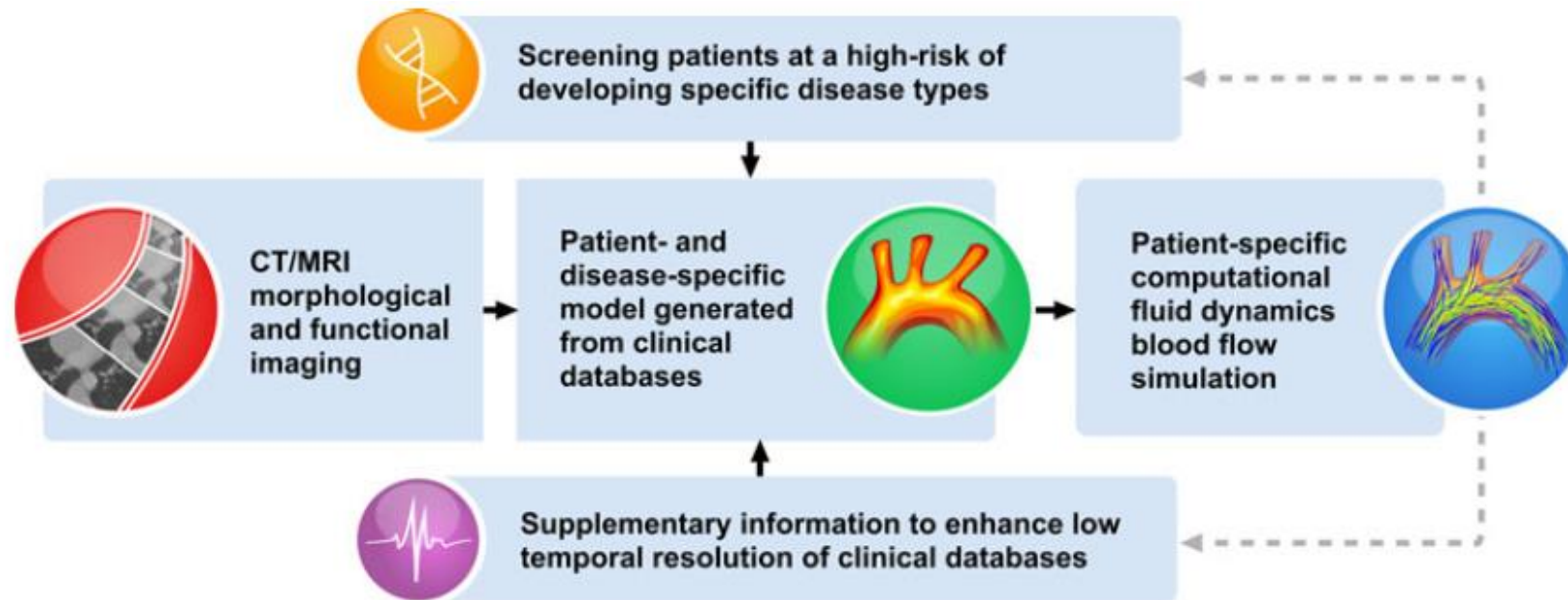


Fig. 3. Integration of imaging, modeling, and real-time sensing for the management of disease progression and planning of intervention procedures. This example of thoracic aortic dissection illustrates how risk stratification and subject-specific haemodynamic modeling substantiated with long-term continuous monitoring are used to guide the clinical decision process.

Big Data em Saúde

- Existe uma grande variedade de tipos de dados de saúde:
 - Informática para saúde e informática médica
 - EHR (Prontuário Eletrônico)
 - Saúde social
 - Saúde pública e fatores ambientais
 - Bioinformática translacional
 - Farmagenômica
 - Genômica translacional
 - Omics e bancos de dados de larga escala

Big Data em Saúde

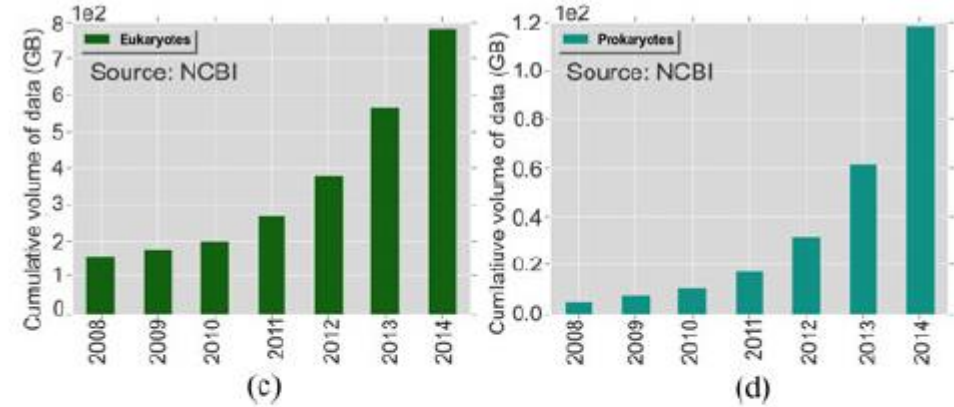
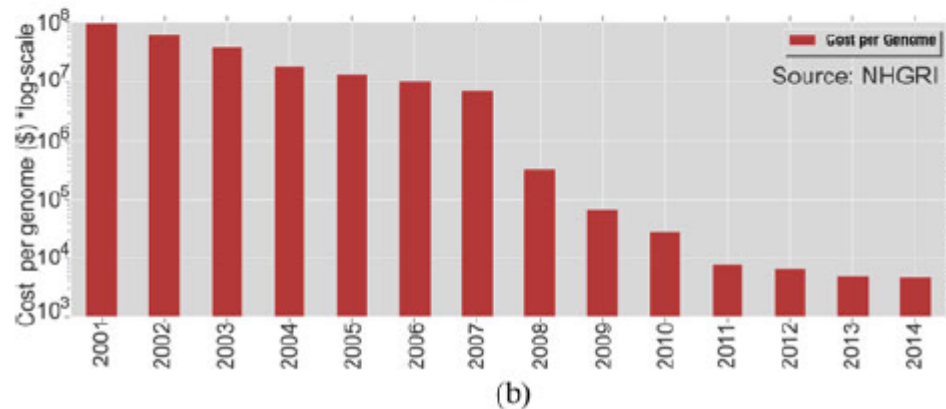
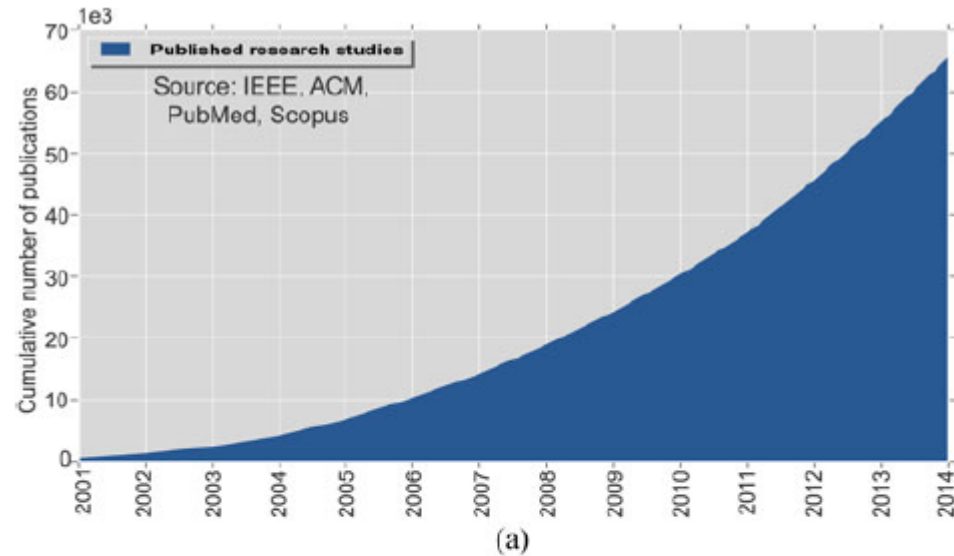


Fig. 4. a) Number of research studies sequencing DNA or genomes (source: PubMed, Web of Science, Scopus, IEEE, ACM). b) Sequencing cost per humanized genome (source: National Human Genome Research Institute, NHGRI). Total volume of genomic data per year reported by completed studies for c) eukaryotes and d) prokaryotes in 1e2 GB (source: National Center for Biotechnology Information).

Big Data em Saúde

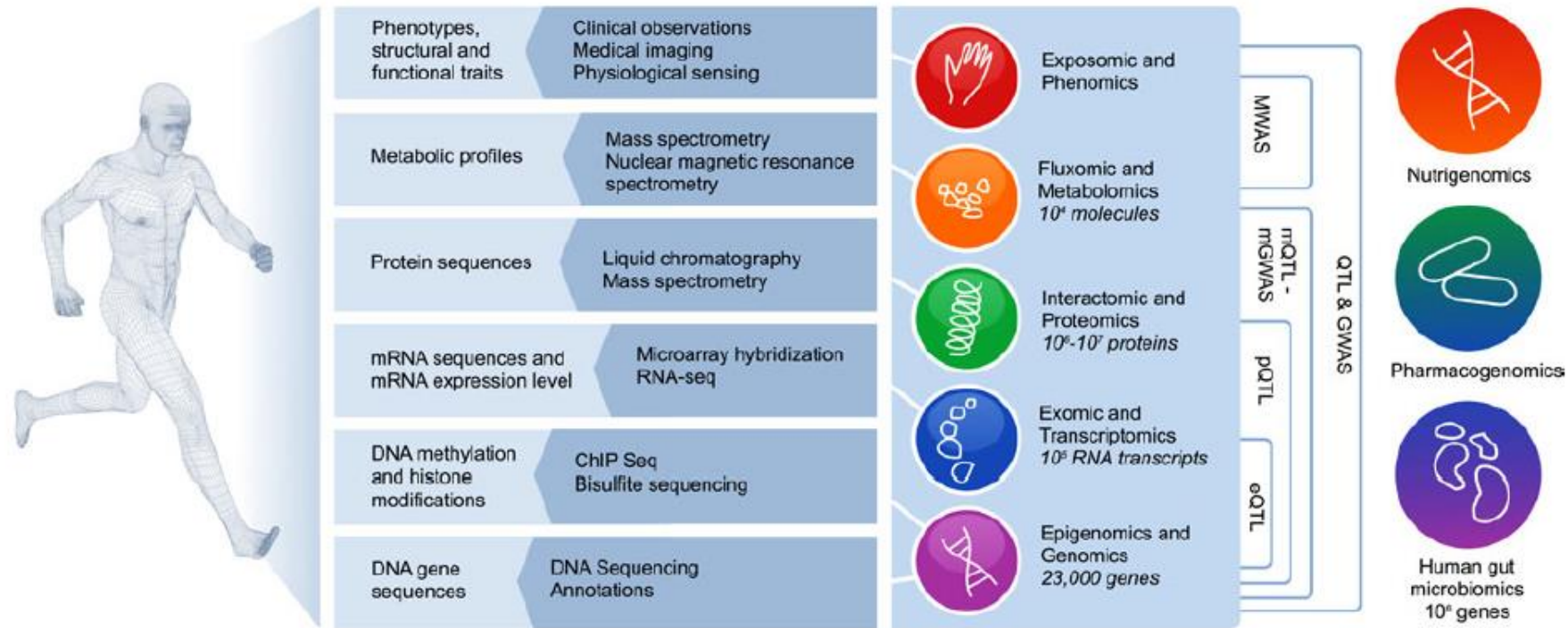


Fig. 5. Outline of the “OMICS” approach for studying disease mechanisms. OMICS aims at collectively characterizing and quantifying groups of biological molecules that translate into the structure, function, and dynamics of an organism. The OMICS profile of each individual, including the genome, transcriptome, proteome, and metabolome, should be eventually linked up with phenotypes obtained from clinical observations, medical images, and physiological signals. Different acquisition technologies are required to collect data at each biological level. Interaction within each level and across different levels as well as with the environment, including nutrition, food, drugs, traditional Chinese medicine, and gut microbiome presents grand challenges in future bioinformatics research.

Big Data em Saúde

- Existe uma grande variedade de tipos de dados de saúde:
 - Redes de sensores
 - Wearables e sensores de ambiente
 - Dados de sensores para estratificação de informações de pacientes
 - Saúde móvel
 - Processamento de imagem
 - Diagnósticos baseados na morfologia
 - Compreensão do Cérebro

Big Data em Saúde

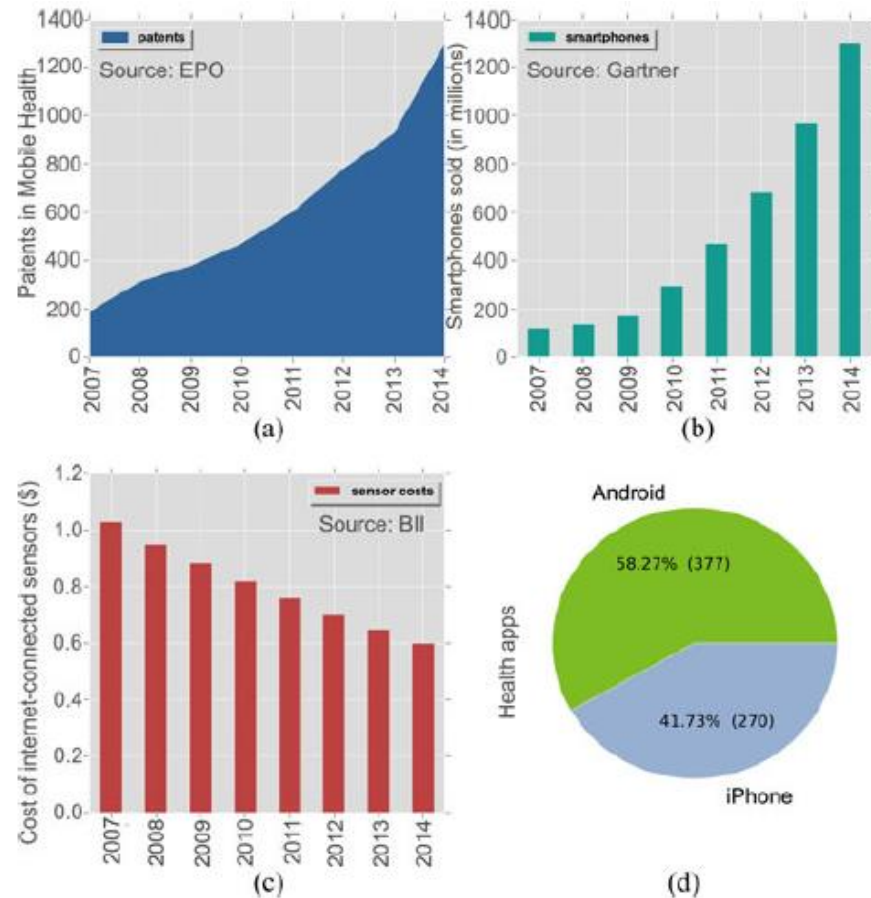
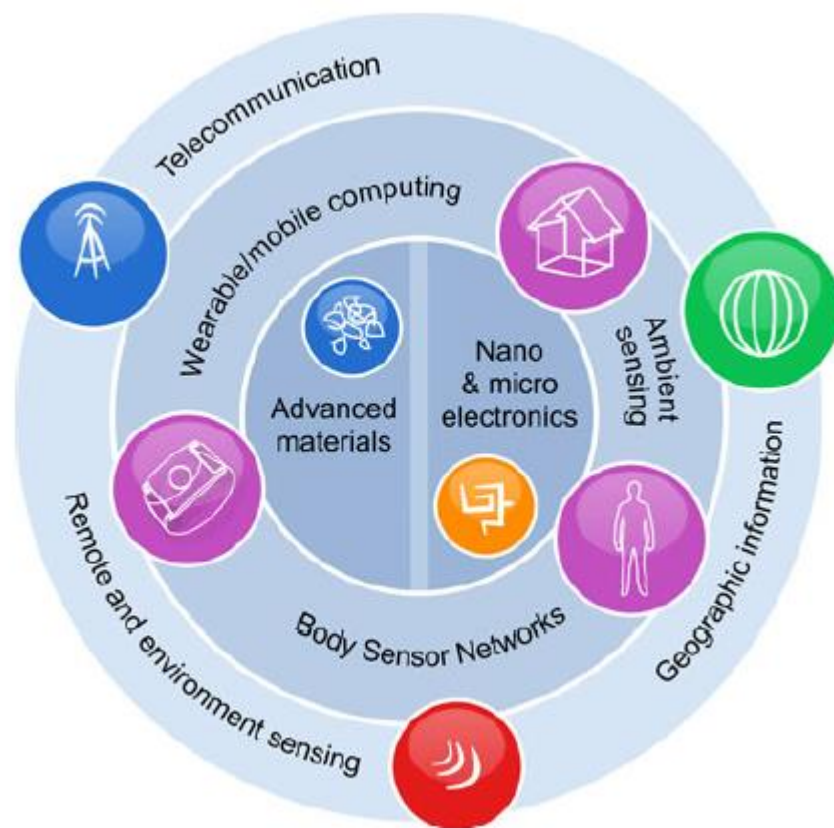


Fig. 6. a) Evolution of the number of patents published in the area of mobile health (source: European Patent Office); b) evolution of the number of smartphones sold per year in million units (source: Gartner); c) evolution of the cost of Internet-enabled sensors in dollars (source: Business Intelligence International); d) number of mobile health apps published in Google play and iTunes as of May 2015.

Big Data em Saúde

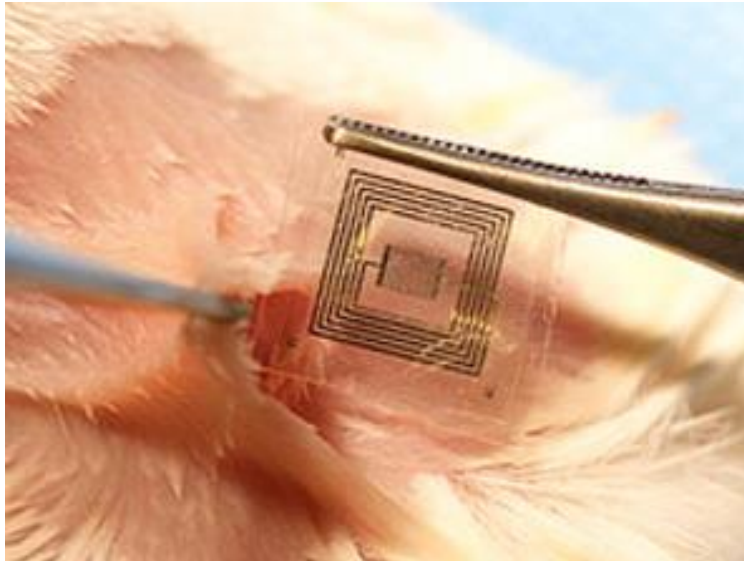


Por que Big Data AGORA em Saúde?

- As Organizações de Saúde estão em uma “encruzilhada” de mudança de rumo – elas devem acolher AGORA novas fontes clínicas de dados ... para melhorar o atendimento de saúde (p. ex., genômica, “wearable”);
- O aumento dos custos em espiral tem levado o setor de saúde a buscar novas formas de aumentar a eficiência operacional e reduzir os custos ... big data é uma alternativa!;
- O aumento dos custos é a maior motivação para as Organizações de Saúde adotarem a “análise de dados” (big data) ... 59% delas afirmam isso nos EUA;
- Em termos de Tecnologia da Informação (TI), a indústria de Saúde (“healthcare”) está atrasada vinte anos em relação a outras indústrias (ver [Healthcare Informatics and MedTech Innovation](#), PwC, 25.apr.2014) – big data ajuda na eliminação desse “gap” tecnológico;
- Aumento exponencial do volume de dados de saúde e, como consequência, a busca de processamentos eficientes para manipular esses dados.



Exemplo ... Uma nova fronteira na tecnologia “wearable” = medicamentos



- Um novo implante eletrônico pode combater a infecção bacteriana no interior do corpo e, em seguida, sem causar danos dissolver.
- O dispositivo simples, passivo é feito inteiramente de seda e magnésio, que são ambos, materiais biocompatíveis benignos.
- O dispositivo, que pode ser carregado com antibióticos, pode ser implantado no interior do corpo durante a cirurgia e, em seguida, accionado a partir do lado de fora sem fios para “matar” bactérias usando calor ou pelo uso de drogas de libertação.
- Uma vez que o seu trabalho é feito, ele se dissolve completamente.

[Silk Implants Fight Bacterial Infection Then Vanish](#), IEEE Spectrum, 28.jan.2015

Big Data em Saúde: Análise Preditiva

ANÁLISE PREDITIVA (“*PREDICTIVE ANALYTICS*”) é o processo de aprendizado (usando algoritmos de “*machine learning*”) a partir dos dados históricos com o objetivo de fazer previsões sobre o futuro (ou algo desconhecido).

PARA A SAÚDE, a Análise Preditiva proporcionará que as melhores decisões sejam tomadas, permitindo o atendimento personalizado para cada indivíduo.

Big Data em Saúde: Análise Preditiva

Passo	Tarefas
1. Agregação de Dados (<i>Training data aggregation</i>)	<ul style="list-style-type: none">• Obter o banco de dados que será utilizado para modelagem;• Elaborar a formatação dos dados quando necessário;• Limpar os dados (verificar registros inválidos, como pacientes com 300 anos);• Disponibilizar essas informações em um formato diretamente "interpretável" pela ferramenta a ser utilizada.
2. Busca de Relacionamentos (<i>Search Relationship</i>)	<ul style="list-style-type: none">• Estudar quais as informações serão úteis para fazer a predição da variável de interesse (p. ex., ... seria buscar por meios estatísticos quais as informações do PEP podem ser úteis para prever se aquele paciente tem uma chance elevada de retornar ao PS num prazo de 90 dias);• Estudar essas relações (é quando se começa a escrever os algoritmos de Inteligência Artificial que trabalharão com a parte de matemática pesada da predição).
3. Coleta de dados do caso (<i>Case Data Collection</i>)	<ul style="list-style-type: none">• Selecionar dentro do banco de dados de pacientes aqueles que tenham uma certa combinação de "preditores" (identificados no passo 2) para investigação mais detalhada.
4. Caracterização Individual do Caso (<i>Individual case characterization</i>)	<ul style="list-style-type: none">• Utilizar os modelos/algoritmos do passo 2 para caracterizar os pacientes do passo 3 e, assim, descrever a resposta de interesse.
5. Recomendação /Contextualização (<i>Recommendation / contextualization</i>)	<ul style="list-style-type: none">• Efetuar uma decisão baseada no passo 4.
6. Captura de Performance (<i>Performance capture</i>)	<ul style="list-style-type: none">• Determinar quais pacientes atendem ao critério de busca;• Usar a informação encontrada, para corrigir o modelo.

Big Data em Saúde: Oportunidades

Analisar os dados clínicos e operacionais pode “abrir muitas portas” nas Organizações de saúde (p. ex., radiografia de pacientes de alto custo)



Saúde Pública

- Melhorar a saúde enquanto reduzimos os custos
- Compartilhar os dados de diferentes instituições
- Gerenciar os custos e os recursos em tempo real
- Melhor gerenciamento do gasto do dinheiro



Provedores

- Medir, rastrear e melhorar a performance
- Melhorar os resultados de saúde e a segurança do paciente
- Melhorar a precisão das estimativas de custos



Seguradoras de Saúde

- Responder a questões sobre crescimento futuro, rentabilidade e sustentabilidade
- Comandar decisões
- Detectar e prevenir fraudes

Big Data em Saúde: Resultados Alcançados

As Organizações de Saúde que já implementaram a Análise Preditiva estão colhendo benefícios ... Interessante destacar que elas têm conseguido **mais benefícios nos dados clínicos do que naqueles operacionais!** (*CDW Analytics in Healthcare 2015*)



→ Destaques nos Benefícios Clínicos

- Melhoria no cuidado do paciente **82%**
- Redução nas taxas de reinternação **63%**
- Melhoria nos resultados globais de saúde **62%**



→ Destaques nos Benefícios Operacionais

- Melhorias nas capacidades financeiras **54%**
- Melhoria na performance operacional do hospital **50%**
- Melhoria na gestão de tomada de decisão **49%**

61% das Organizações afirmam que seus índices de resultados gerencias de 30 dias têm melhorado com a Análise Preditiva e que 81% esperam essa melhoria depois de fazer a implementação completa dos Algoritmos de Predição!

Big Data e Machine Learning

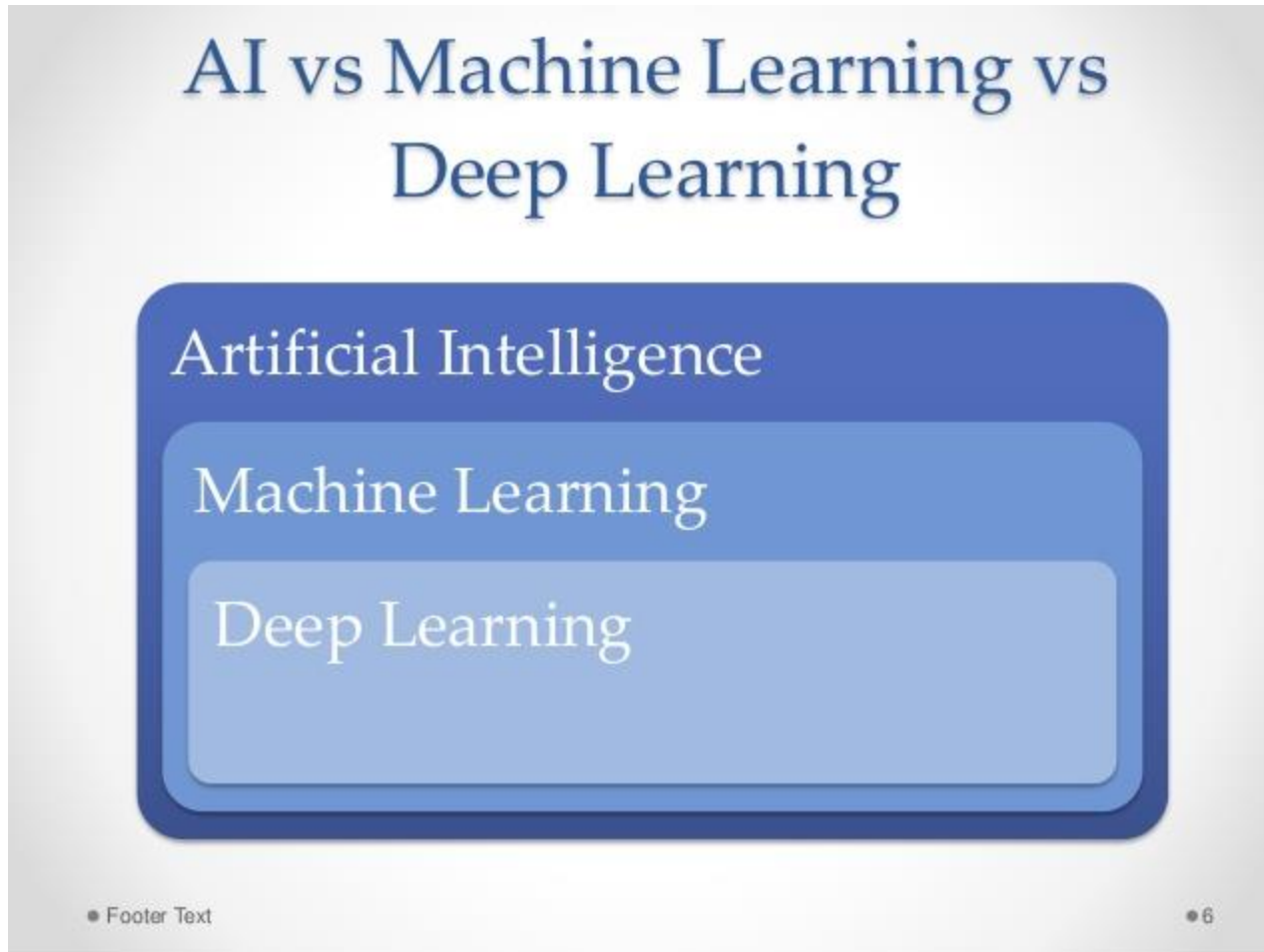
- O volume e complexidade de dados envolvidos em Big Data são tão grandes que as tradicionais técnicas de análise de dados mostram-se muito limitadas para sua análise
- A manipulação desse grande volume de dados em formato digital tem grandes oportunidades na área da saúde
- Contudo, essa manipulação revela-se um grande desafio:
 - Grande volume e variedade de dados
 - As técnicas tradicionais de análise de dados têm se mostrado limitadas quanto à capacidade e tempo de resposta para tal tarefa

DEEP LEARNING EM SAÚDE

Inteligência Artificial, Machine Learning, Deep Learning

- As técnicas de inteligência artificial têm sido apontadas como importantes aliadas para manipulação de grandes e complexos volumes de dados.
- Dentre essas técnicas, as de machine learning (aprendizado de máquina) têm tido preferência, com destaque para Deep Learning mais recentemente.

Inteligência Artificial, Machine Learning, Deep Learning



- **ARTIFICIAL INTELLIGENCE (AI)**

AI é o termo mais amplo, aplicável a uma técnica que permite que os computadores possam imitar a inteligência humana, usando a lógica, regras **if-then**, árvores de decisão e aprendizagem de máquina (incluindo a aprendizagem de profundidade).

- **MACHINE LEARNING (ML)**

O subconjunto de AI, que inclui técnicas estatísticas mais sofisticadas que permitem que as máquinas possam melhorar as tarefas com experiência. A categoria inclui aprendizagem profunda.

- **DEEP LEARNING**

O subconjunto de ML composto por algoritmos que permitem ao software treinar-se para executar tarefas, como fala e reconhecimento de imagem, resultando em redes neurais multicamadas para grandes quantidades de dados.

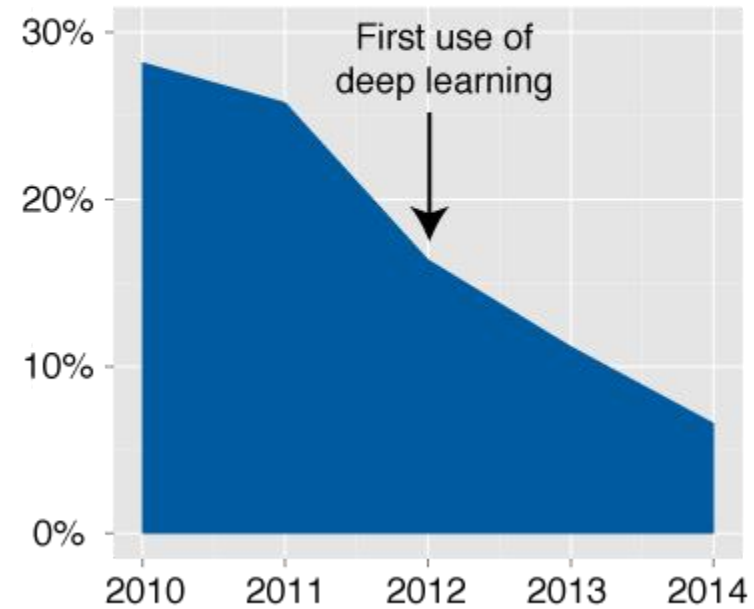
<http://fortune.com/ai-artificial-intelligence-deep-machine-learning/>

Por que Deep Learning?

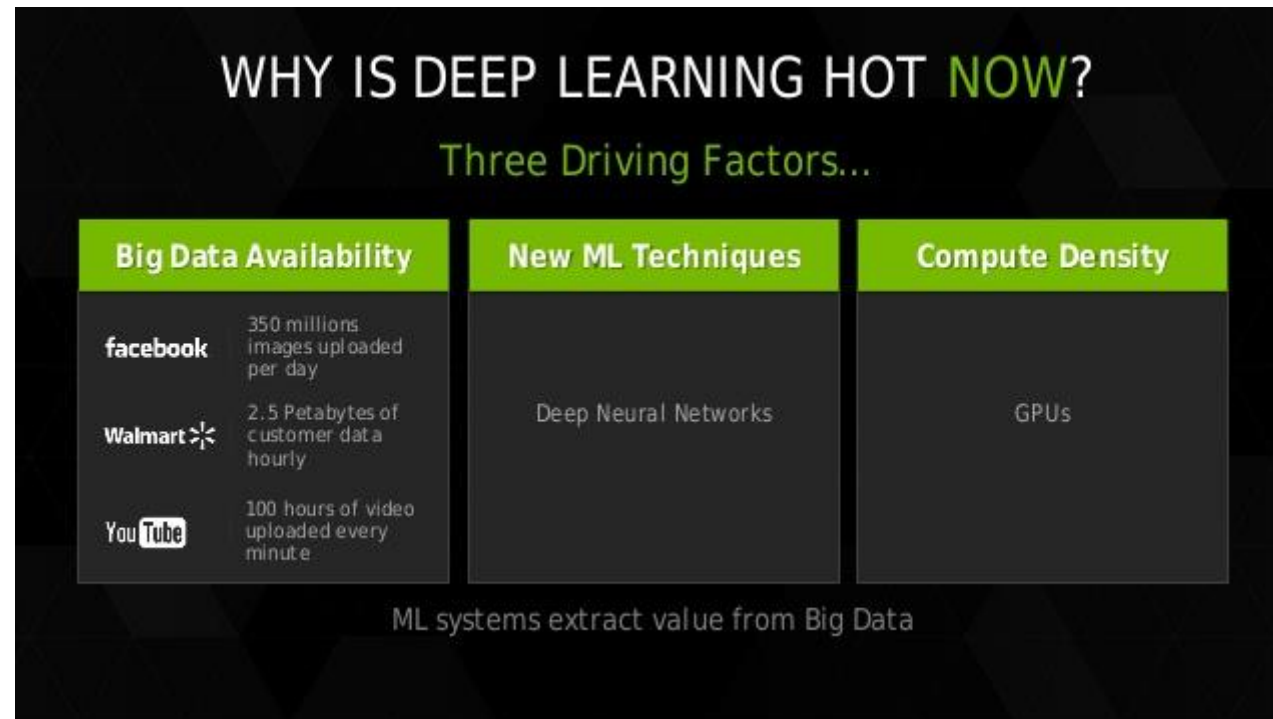
ImageNet examples



Object classification error rate



Por que Deep Learning?



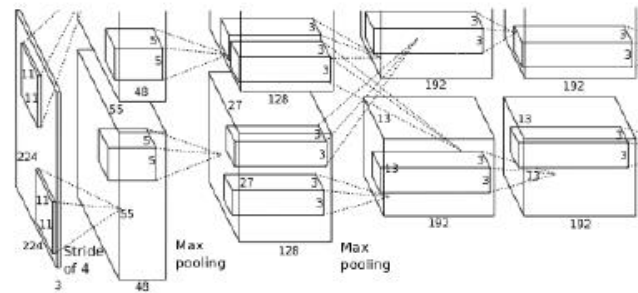
Por que Deep Learning?

The Deep Learning "Computer Vision Recipe"



Big Data: ImageNet

+



Deep Convolutional Neural Network

+



Backprop on GPU

=

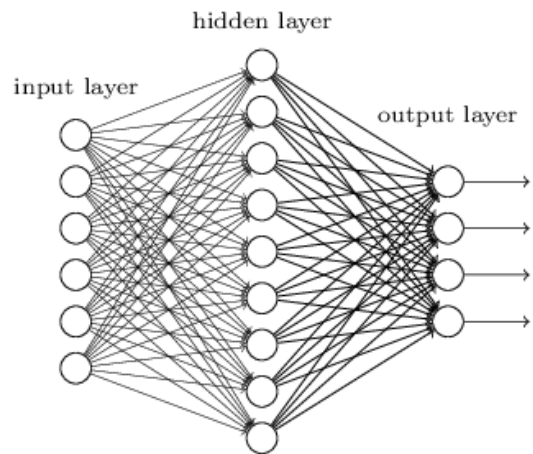


Learned Weights

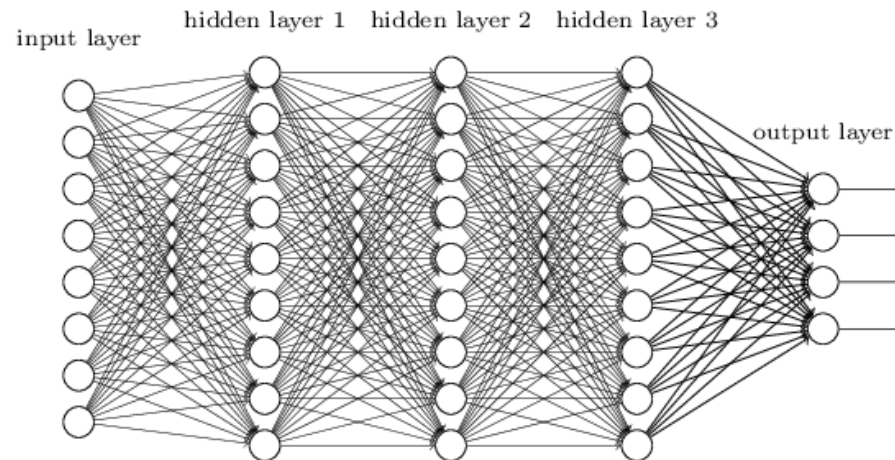
Por que Deep Learning?

- Dentre as diferenças que destacam as técnicas de Deep Learning estão:
 - Múltiplas camadas
 - Consegue trabalhar várias dimensões

"Non-deep" feedforward neural network



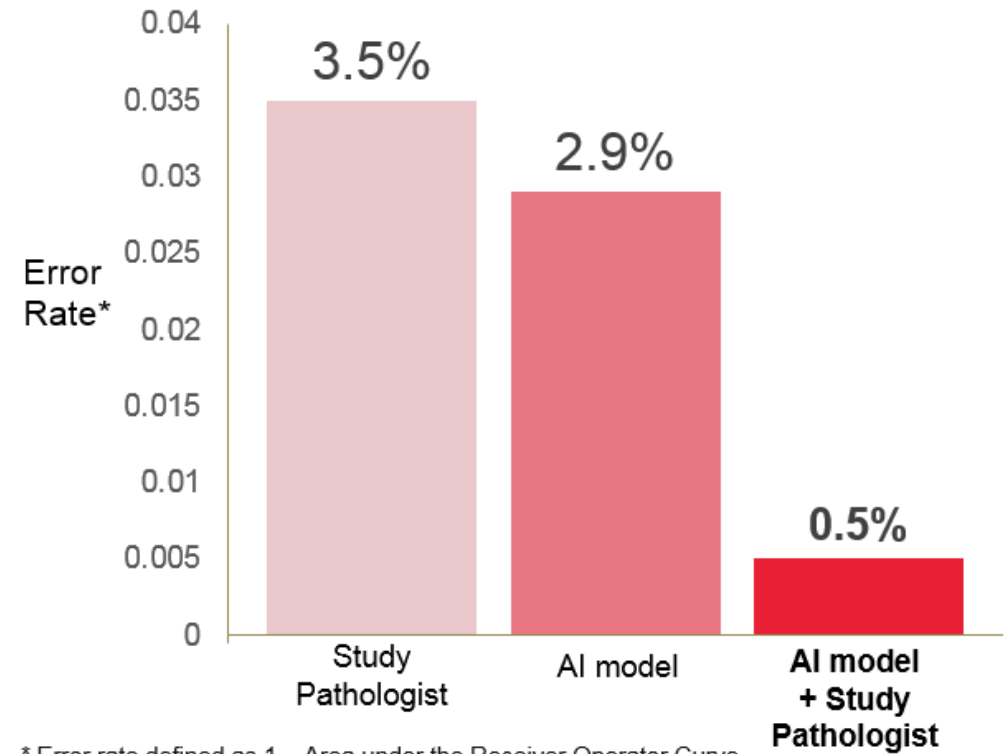
Deep neural network



Aplicações na Área da Saúde

- Análise de Imagens Médicas
- Em um trabalho desenvolvido pela Harvard Medical School e publicado no International Symposium on Biomedical Imaging de 2016 obtiveram os seguintes resultados em mamografias

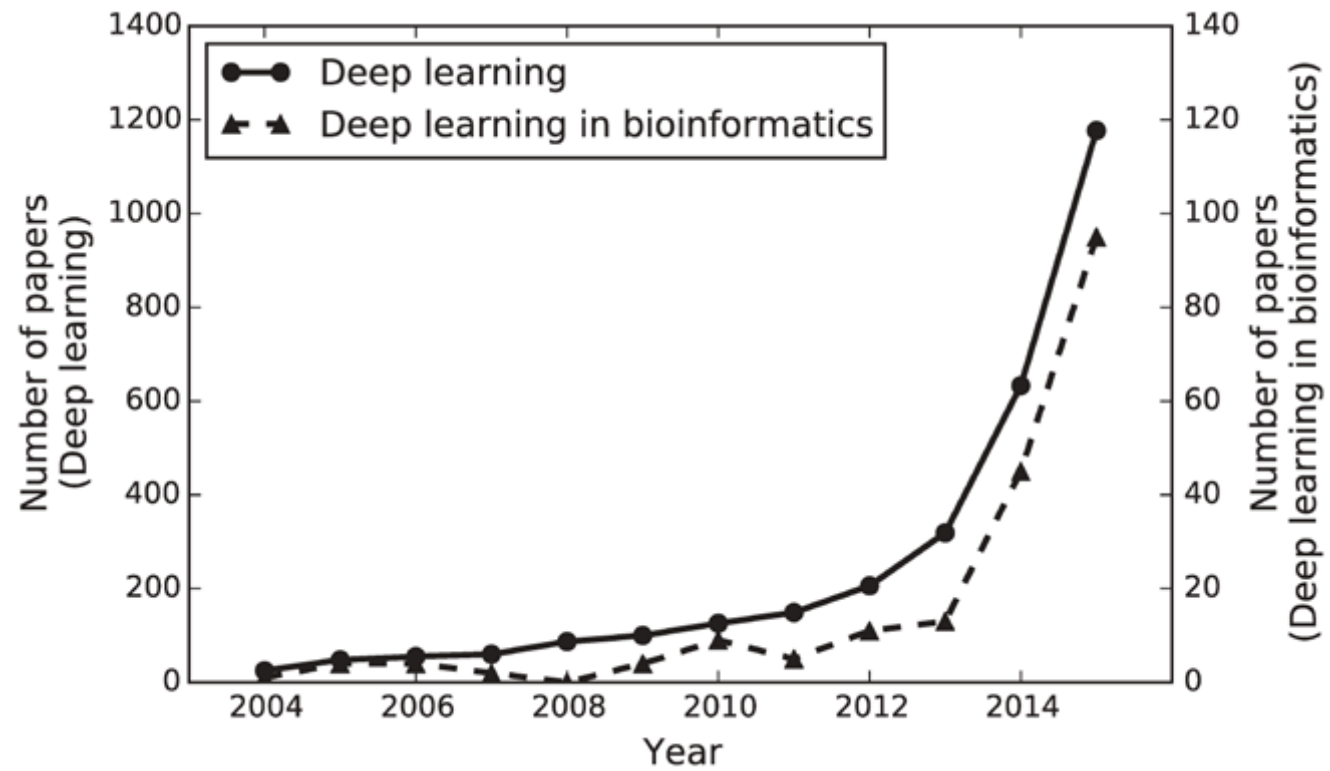
(AI + Pathologist) > Pathologist



* Error rate defined as 1 – Area under the Receiver Operator Curve

** A study pathologist, blinded to the ground truth diagnoses, independently scored all evaluation slides.

Aplicações na Área de Saúde



Aplicações na Área da saúde

- Análise de Imagens
 - O artigo “EARLY DIAGNOSIS OF ALZHEIMER’S DISEASE WITH DEEP LEARNING” propõe o uso de deep learning para identificação de pessoas com pré-disposição a desenvolver Alzheimer

NC: Normal Control

MCI: Mild Cognitive Impairment

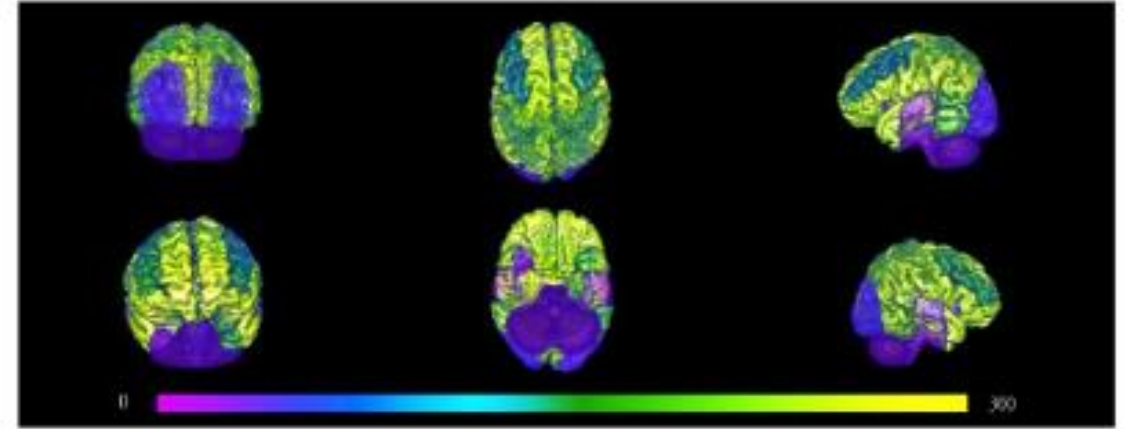


Table 1: The mean values of the binary classification performance (%) with pre-computed features from MRI and PET images

Methods	AD vs. NC			MCI vs. NC		
	ACC	SEN	SPE	ACC	SEN	SPE
SK-SVM	84.40	84.64	84.31	76.81	56.14	86.24
MK-SVM	86.42	84.98	87.83	77.25	55.48	87.10
Proposed	87.76	88.57	87.22	76.92	74.29	78.13

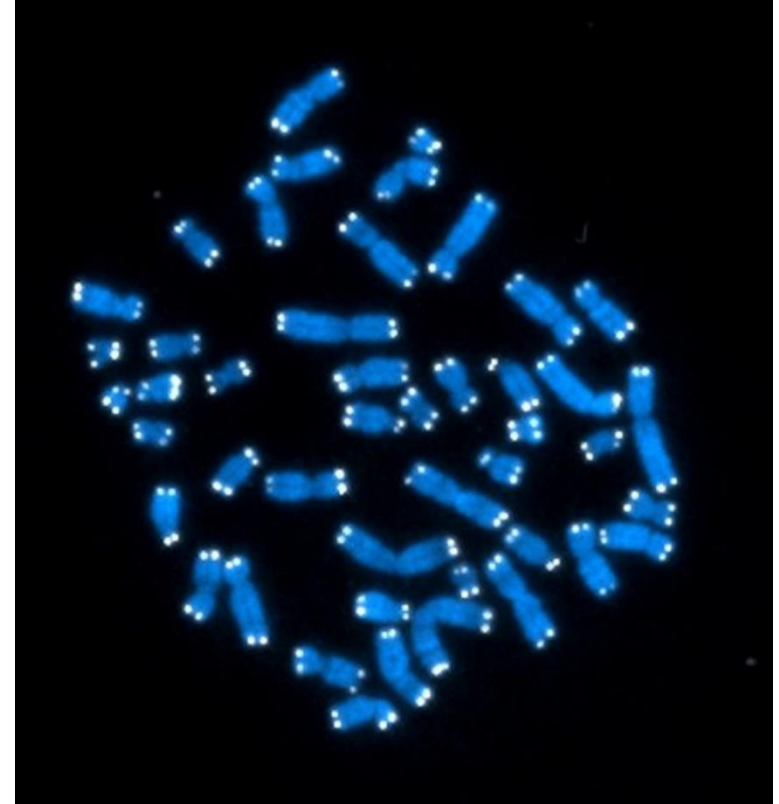
Aplicações na Área da Saúde

- Análise Preditiva em EHR - Eletronic Health Records (Prontuário Eletrônico)
- O artigo “Deep Patient: Un Unsupervised Representation to Predict the Future Patients From Eletronic Health Records” (*Nature Scientific Reports* **6**,(2016)) obtiveram as seguintes precisões de predições

Time Interval = 1 year (76,214 patients)			
Patient Representation	AUC-ROC	Classification Threshold = 0.6	
		Accuracy	F-Score
RawFeat	0.659	0.805	0.084
PCA	0.696	0.879	0.104
GMM	0.632	0.891	0.072
K-Means	0.672	0.887	0.093
ICA	0.695	0.882	0.101
DeepPatient	0.773[*]	0.929[*]	0.181[*]

Aplicações na Área da Saúde

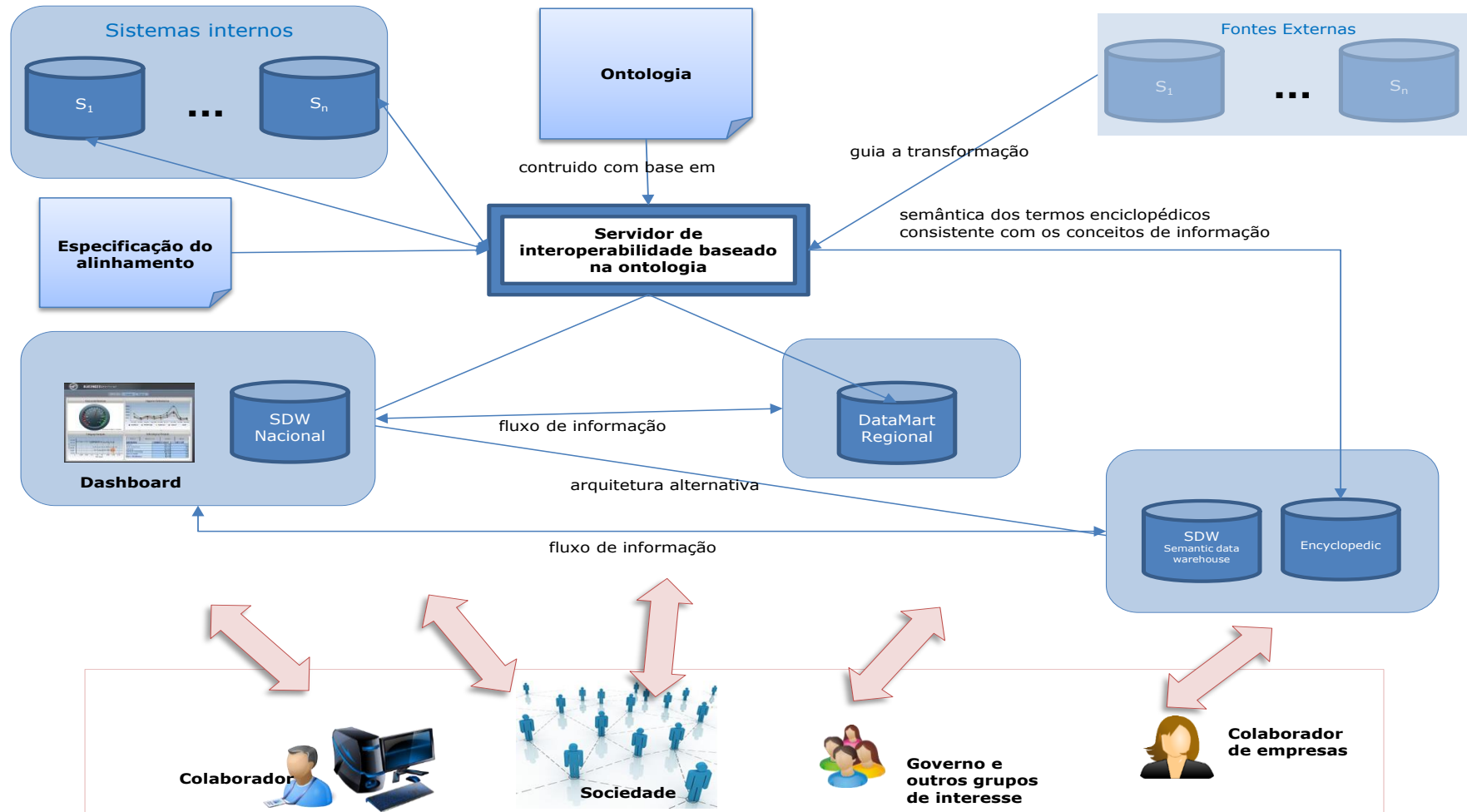
- Genômica
 - O artigo Machine Learning in Genomic Medicine aponta Deep Learning como um dos propulsores principais para as pesquisas da área como:
 - CRISP (Chesapeake Regional Information System for our Patients)
 - Estudo de Riscos de Doenças
 - Mutações genéticas
 - A Deep Genomics é uma companhia criada há pouco tempo, por professores e acadêmicos da Universidade de Toronto, com propósito de desenvolver e aplicar algoritmos de deep learning em genômica

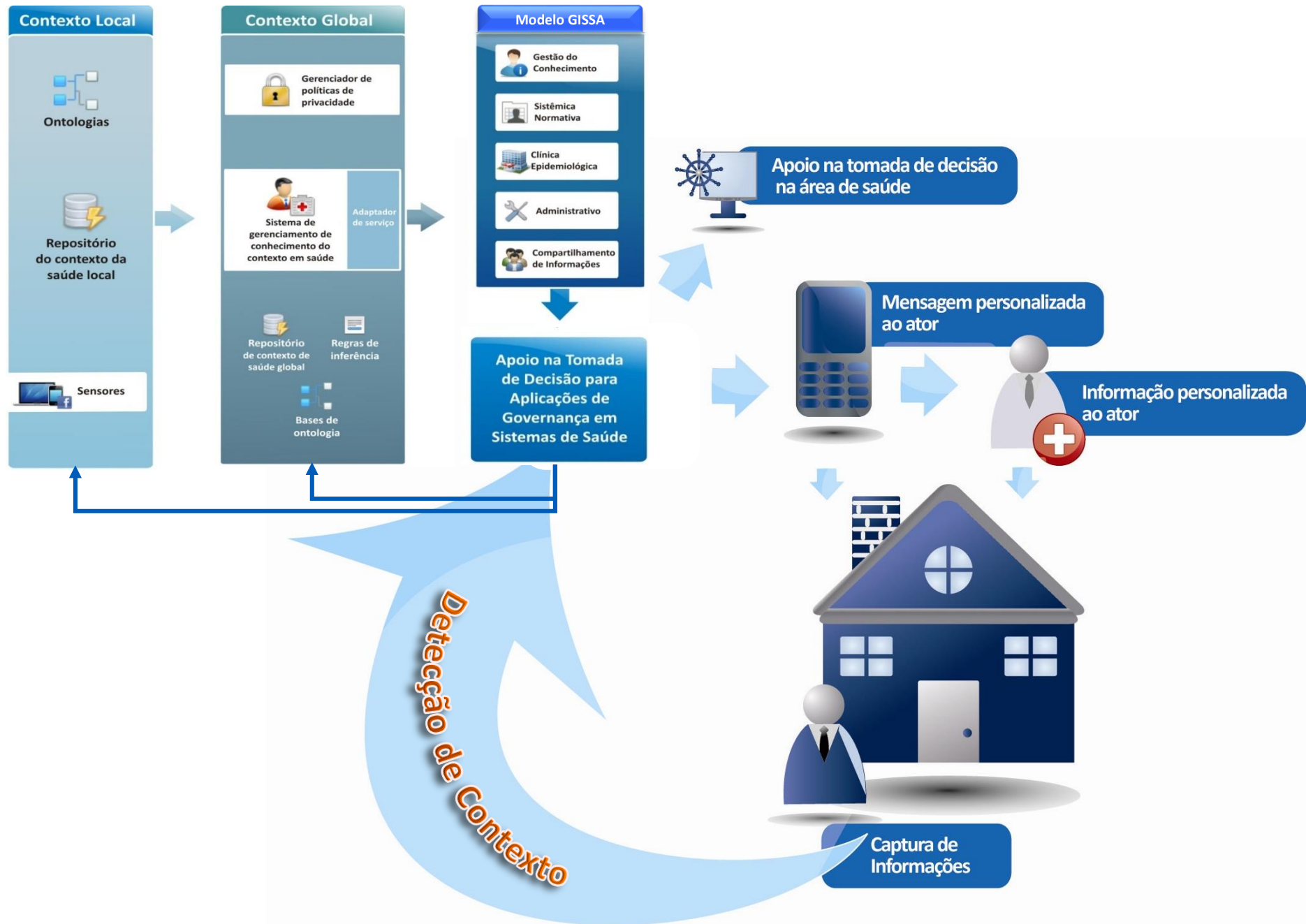


Um sumário das aplicações de deep learning na área biomédica

	Omics		Biomedical imaging		Biomedical signal processing	
	Research topics	Reference	Research topics	Reference	Research topics	Reference
Deep neural networks	Protein structure	[84-87]	Anomaly classification	[122-124]	Brain decoding	[158-163]
	Gene expression regulation	[93-98]	Segmentation	[133]	Anomaly classification	[171-175]
	Protein classification	[108]	Recognition	[142, 143]		
	Anomaly classification	[111]	Brain decoding	[149, 150]		
Convolutional neural networks	Gene expression regulation	[99-104]	Anomaly classification	[125-132]	Brain decoding	[164-167]
			Segmentation	[134-140]	Anomaly classification	[176]
			Recognition	[144-147]		
Recurrent neural networks	Protein structure	[88-90]			Brain decoding	[168]
	Gene expression regulation	[105-107]			Anomaly classification	[177, 178]
	Protein classification	[109, 110]				
Emergent architectures	Protein structure	[91, 92]	Segmentation	[141]	Brain decoding	[169, 170]

Semantic Datawarehouse







Grupo de Pesquisa em Tecnologias da Informação Aplicadas à Saúde

- Prof. Dr. Anilton S Garcia – LabTel/UFES – anilton.garcia@ufes.br
- Prof. Dr. Hélder R de O Rocha – DCEL/CEUNES/UFES - helder.rocha@ufes.br
- Prof. Dr. Jorge A Samadello – DEL/CT/UFES
- Prof. Dr. Leonardo J Silvestre – DCEL/CEUNES/UFES - leonardo.silvestre@ufes.br
- Prof. Dra. Luciana Lee – DCEL/CEUNES/UFES - luciana.lee@ufes.br
- Prof. Dr. Patrick M Ciarelli – DEL/CT/UFES
- Prof. Dra. Silvia Rissino – DCEL/CEUNES/UFES - silvia.rissino@ufes.br
- Prof. Dr. Wilian H Hisatugu – DCEL/CEUNES/UFES - wilian.hisatugu@ufes.br

Muito Obrigado ... !!!

Anilton Salles Garcia

E-Mail: anilton.garcia@ufes.br

Eduardo Prado

E-mail: eprado.sc@gmail.com

Wilian Hiroshi Hisatugu

E-mail: wilian.hisatugu@ufes.br