


IADB


Intégration et Analyse de Données Biomédicales
Biomedical Data Integration and Analysis

BCL, CHUN, I3S, Inria CRISAM, MSI




UNIVERSITÉ CÔTE D'AZUR

1




2



Intégration et Analyse de Données Biomédicales

- Objectifs
 - Analyser des données historiques provenant des institutions médicales régionales pour en extraire des connaissances
 - Développer un système d'aide au diagnostic et au pronostique médical
- Méthodes
 - Intégration de très grands ensembles de données
 - Création de modèles par apprentissage automatique
- Défis
 - Traitement de données sensibles, hétérogènes et morcelées
 - Développement de modèles pertinents pour un usage médical
 - Puissance de calcul limitée

3



Biomedical Data Integration and Analysis

- Objectives of the project
 - Analyze historical data from regional medical institutions to extract knowledge
 - Develop a system to assist in medical diagnosis and prognosis
- Methods of operation
 - Integration of very large data sets
 - Creation of models by automatic learning
- Challenges
 - Processing of sensitive, heterogeneous and fragmented data
 - Development of relevant models for medical use
 - Limited computing power

4

Répondre à des questions telles que...

- Identifier la morbidité primaire ou la principale catégorie clinique
- Estimer la probabilité de co-morbidité
- Diagnostiquer des affections cycliques

- Prédire
 - La probabilité de nouveau séjour
 - La durée de la prochaine hospitalisation
 - Les prochaines procédures cliniques
 - ...

- Certaines questions peuvent porter sur des populations, d'autres sur des individus précis

5

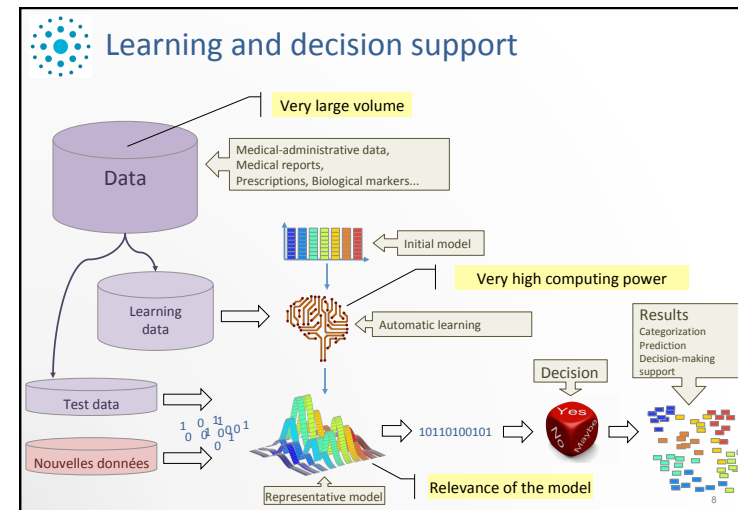
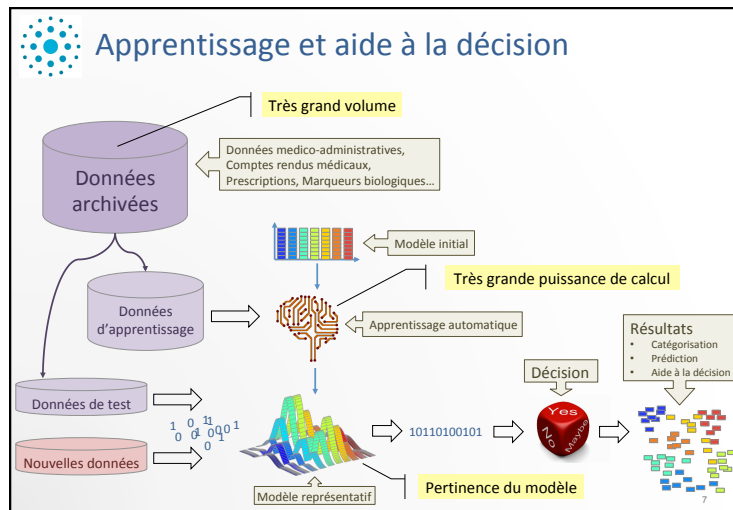
Answer questions such as....

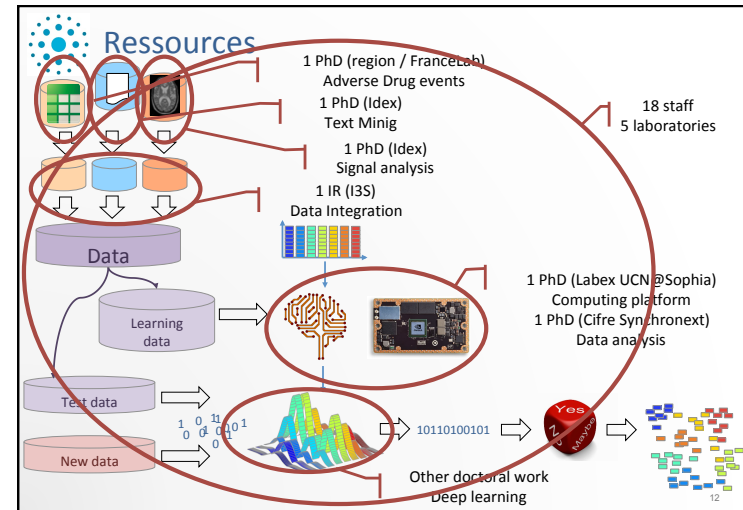
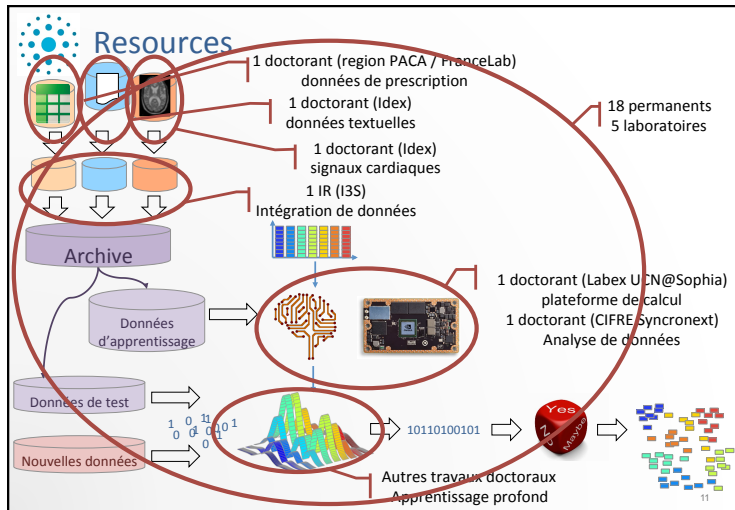
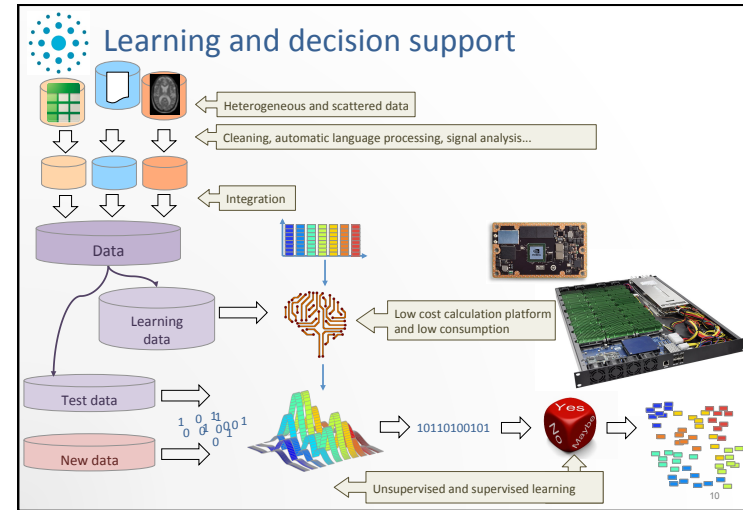
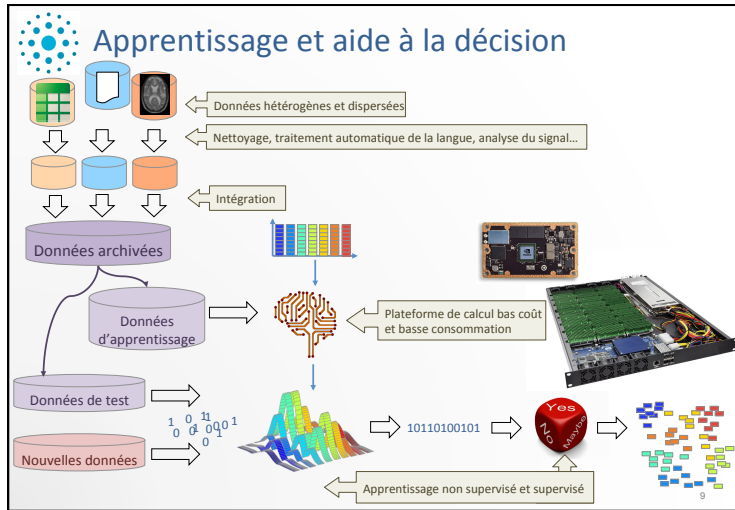
- Identify the primary morbidity or main clinical category
- Estimate the probability of co-morbidity
- Diagnose cyclical conditions


- Predict
 - The probability of a new stay
 - The duration of the next hospitalization
 - The next clinical procedures
 - ...

- Some questions may focus on specific populations, others on specific individuals

6








Partenariat

- I3S SPARKS
 - Intégration de données hétérogènes + apprentissage machine
- Inria-I3S Wimmics
 - Modélisation et exploitation des connaissances
- I3S SIS
 - Analyse d'images, traitement du signal
- BCL
 - Analyse du langage, interprétation de données textuelles
- CHUN
 - Cohortes de données, études médicales
- MSI UCA
 - Plateforme, support, lien avec l'industrie


13



Partnership

- CNRS I3S SPARKS + Inria-I3S Wimmics
 - Heterogeneous data integration + machine learning
- CNRS I3S SIS
 - Image analysis, signal processing
- CNRS BCL
 - Language analysis, interpretation of textual data
- CHUN
 - Data cohorts, medical studies
- MSI UCA
 - Platform, support, link with industry


14



Accès aux données

- Sources de données
 - Locales
 - DIIM du CHUN: données médico-administratives (PMSI) et rapports d'hospitalisation
 - PRIMEGE PACA: plateforme régionale d'information en médecine générale
 - Bases de données d'images provenant d'études cliniques
 - Data from competitions
- Les données sont
 - **Sensibles**: souvent dégradées par le processus d'anonymisation
 - **Hétérogènes**: nécessitent transformation et interprétation
 - **Evolutives**: les formats changent d'année en année
 - **Distribués**: nécessitent intégration
- **Problème principal**
 - Pas de clé d'intégration // résultat du processus d'anonymisation
= Pas de cohortes = pas d'apprentissage = pas de résultats médical
- Medical Data Centre de la MSI
 - Plateforme d'accès au données
 - Travail amont de gestion des droits et de collecte indispensable

15



Access to data

- Data sources
 - Local
 - CHUN DIIM: Medico-administrative data (PMSI) and hospitalization reports
 - PRIMEGE PACA: regional information platform in general medicine
 - Image databases from clinical studies
 - Data from competitions
- The data are
 - Sensitive: often degraded by the anonymization process
 - Heterogeneous: require transformation and interpretation
 - Scalable: formats change from year to year
 - Distributed: require integration
- **Main problem**
 - No integration key // result of the anonymization process
= No cohorts = no learning = no medical outcomes
- Medical Data Centre of the MSI
 - Data access platform
 - Upstream work on rights management and collection is essential

16

Caractère disruptif

- L'apprentissage profond a impacté de nombreux domaines
 - Reconnaissance d'images, reconnaissance de la parole, systèmes autonomes...
 - Potentiel de rupture sur l'exploitation des données de santé
 - Très forte croissance dans le domaine médical depuis 2015
- Compétition internationale exacerbée
 - Investissement massif aux Etats Unis notamment:
 - Pour la numérisation systématique des données de santé (certified EHR)
 - Dans les plateformes de calcul intensif pour l'apprentissage
 - Course au développement de modèles pour la santé
 - Secteur privé très actif
- Contexte national
 - Législation contraignante et en cours d'évolution
 - PMSI collectées systématiquement mais de nombreuses données sont perdues
 - Moins de données = résultats moins pertinents
- Grand éventail de compétences à UCA
 - Renforcer les liens entre les producteurs de données et les concepteurs d'outils d'analyse
 - Nécessité de travailler le chainage, l'intégration et la consolidation des données
 - On butte sur des aspects législatifs très important

17

Disruptive nature

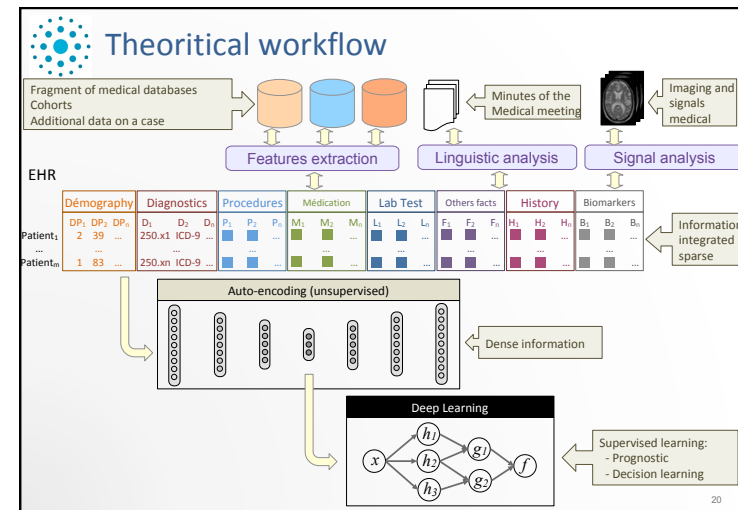
- Deep learning has impacted many areas
 - Image recognition, speech recognition, autonomous systems...
 - Potential for disruption in the exploitation of health data
 - Very strong growth in the medical sector since 2015
- Exacerbated international competition
 - Massive investment in the United States in particular:
 - For the systematic digitisation of health data (certified EHR)
 - In intensive computing platforms for learning
 - A race to develop models for health
 - Very active private sector
- National context
 - Binding and evolving legislation
 - PMSI collected systematically but a lot of data is lost
 - Less data = less relevant results
- Wide range of skills at UCA
 - Strengthen links between data producers and designers of analytical tools
 - Need to work on data chaining, integration and consolidation
 - We come up against very important legislative aspects

18

It is time to move on to some more specific presentations on the work in progress.

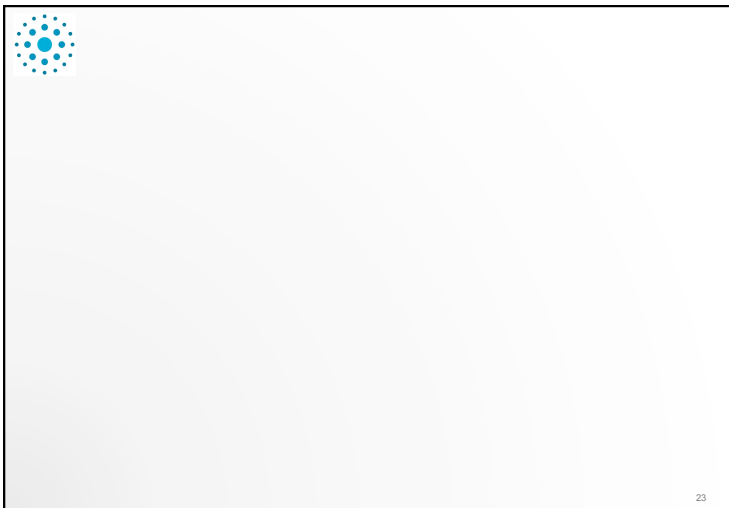
But first, let's look at how they should interact.


19





Conclusion



-  EHR... Medical data
- Generic definition
 - Systematized collection of patient and population electronically-stored health information in a digital format
 - Shared across different health care settings
 - Include a range of data, demographics, medical history, medication and allergies, immunization status, laboratory test results, radiology images, vital signs, and billing information.
 - National context – France
 - 'billing' oriented
 - demographics + Information necessary for the reimbursement of medical procedures
 - No medical results
 - Necessity to build more integrated data
 - on going work in some CHU

