AI SUMMIT

COLUMBUS, OH • OCTOBER 25–27, 2022

AI/ML Point of Care Use Case

Leveraging AI/ML to Aid Clinicians in the Diagnostic Process for Rare Diseases

AI – Artificial Intelligence
ML – Machine Learning

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

AFDO RAPS

MAT-US-2207976 v1.0 – P
Expiration date 10/26/22

Sanofi Medical Affairs for scientific and medical discussions only. Do not photograph, copy or distribute.

Why is AI needed in rare diseases?

- Gaucher is a debilitating, progressive, rare disease that is under-recognized
- Electronic Health Record (EHR) data provides an opportunity to screen for undiagnosed patients, allowing for earlier identification and appropriate management via:
    - Retrospective searches
    - Prospective prompts at point of care
- US healthcare system is disparate and siloed
- Newer companies work to integrate these systems by linking electronic health care records in collaboration with health care systems, laboratories and insurance companies
- Ability to create massive, anonymized, aggregated healthcare datasets

AI/ML Development

Goal:

1. Determine whether an advanced analytics approach could be used to create diagnostic algorithm(s) for rare diseases
2. Compare algorithm performance to a standard clinical diagnostic algorithm

Plan:

1. Use licensed access to very large anonymized dataset to create algorithm, perform validation
2. Follow with study to evaluate in a real-world setting – diagnostic testing

Interactions
Inpatient, ER, cardiologist, radiology…

Diagnoses
Splenomegaly, fatigue, anxiety…

Clinic notes
Pain, paresthesia, angina…

Data Set

Procedures
Pacemaker, biopsy, transplant

Labs
Hgb, platelets, BMI, eGFR…

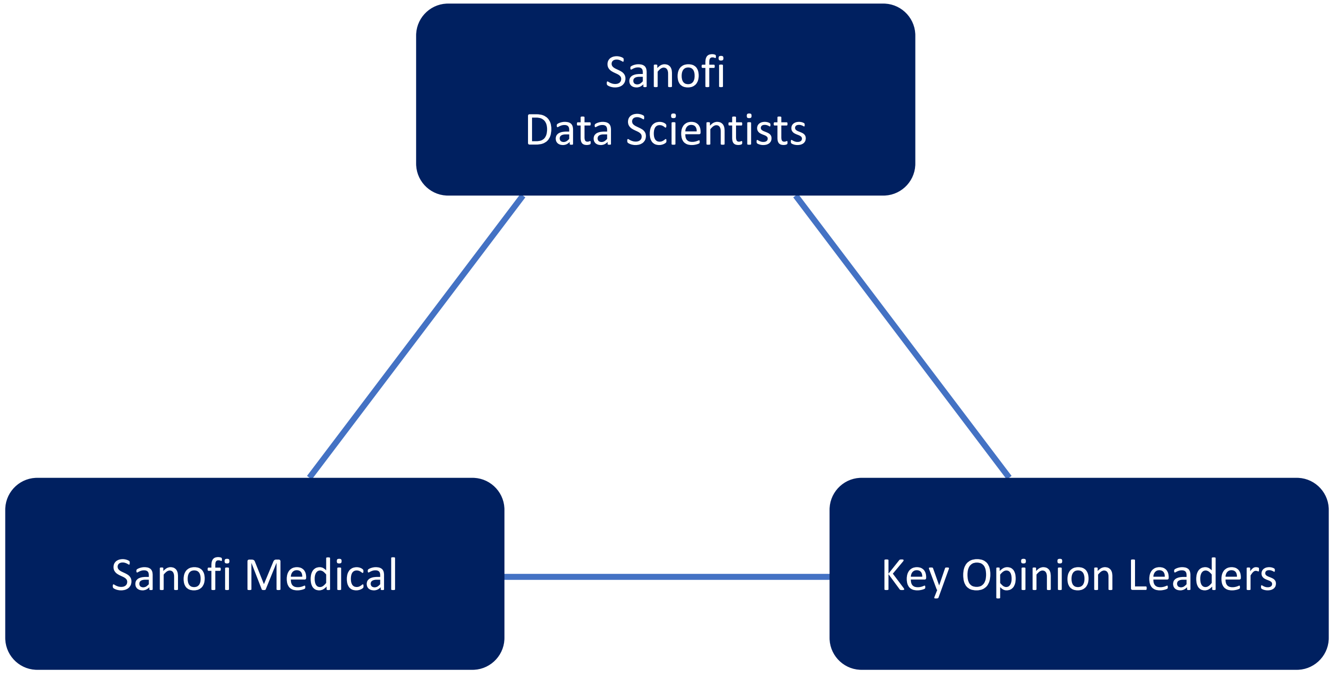
Treatments
Medication Administration, Rx written, Patient Reported Rx

- **2007-19**
- **>50 US IDNs**
- **>700 hospitals/7000 clinics**
- **≈103MM patients**

Algorithm Development Phase

1. Data Structuring
    - Provide features of Gaucher disease from the literature
    - Define "Gaucher disease" for the algorithm, apply to the dataset patients
    - Separate identified Gaucher patients into training cohort and test cohort
2. Train the Algorithm with the **Training** Cohort (1:500)
    - Allow the algorithm to further learn from the Gaucher pts in the training database, looking for p
    - Apply various analytical methods to determine optimal approach
3. Evaluate Model Performance on the **Test** Cohort (1:10,000)
    - Numbers of true patients identified (true positives)
    - Description of highly suspected patients (controls that could be undx Gaucher)
    - Types of patients identified (younger vs older, earlier vs later disease impact)

Engage Clinicians for Development



Training and Test Phases

1. Data Structuring
   - Provide features of Gaucher disease from the literature
   - Define "Gaucher disease" for the algorithm, apply to the dataset patients
   - Separate identified Gaucher patients into training cohort and test cohort
2. Train the Algorithm with the **Training** Cohort (1:500)
   - Allow the algorithm to further learn from the Gaucher pts in the training database, looking for patterns
   - Apply various analytical methods to determine optimal approach
3. Evaluate Model Performance on the **Test** Cohort (1:10,000)
   - Numbers of true patients identified (true positives)
   - Description of highly suspected patients (controls that could be undx Gaucher)
   - Types of patients identified (younger vs older, earlier vs later disease impact)

LITERATURE FEATURES	▪ Symptoms from literature ▪ Enriched with SDS, labs, procedures
DATA DRIVEN FEATURES	▪ Features that the model picked up as differentiators of GD vs. controls
DEMOGRAPHICS	▪ Region, race, gender, age
HEALTHCARE INTERACTIONS	▪ Healthcare provider specialty ▪ Healthcare encounter

▲ **Age Model** ← → ▲ **Prevalence Model**

Redundant Approach to Identify Features

- Thrombocytopenia example:
  - Diagnostic billing codes (ICD): e.g., D69.6 thrombocytopenia, unspecified
  - Procedure billing codes (CPT): e.g., 86965 under Transfusion Medicine Procedures
  - Laboratory values: platelet count <100,000 UI/mL
  - SDS terms: “thrombocytopenia” – mentioned in clinical notes
- Treatment examples:
  - J-codes: J1785 – used to bill for the infusion, drug-specific, permanent code
  - C-codes: C9294 – used to bill for the infusion, drug-specific, a temporary code
  - Drug codes: 58468198301 – drug-specific

Training Results

Age model

Prevalence model

Random Gaucher patients
N=100

TESTING DATASET

Patients
N=1,000,000 (ratio 1:1000)
Gaucher patients
N=100

Age model — **17x fewer patients**

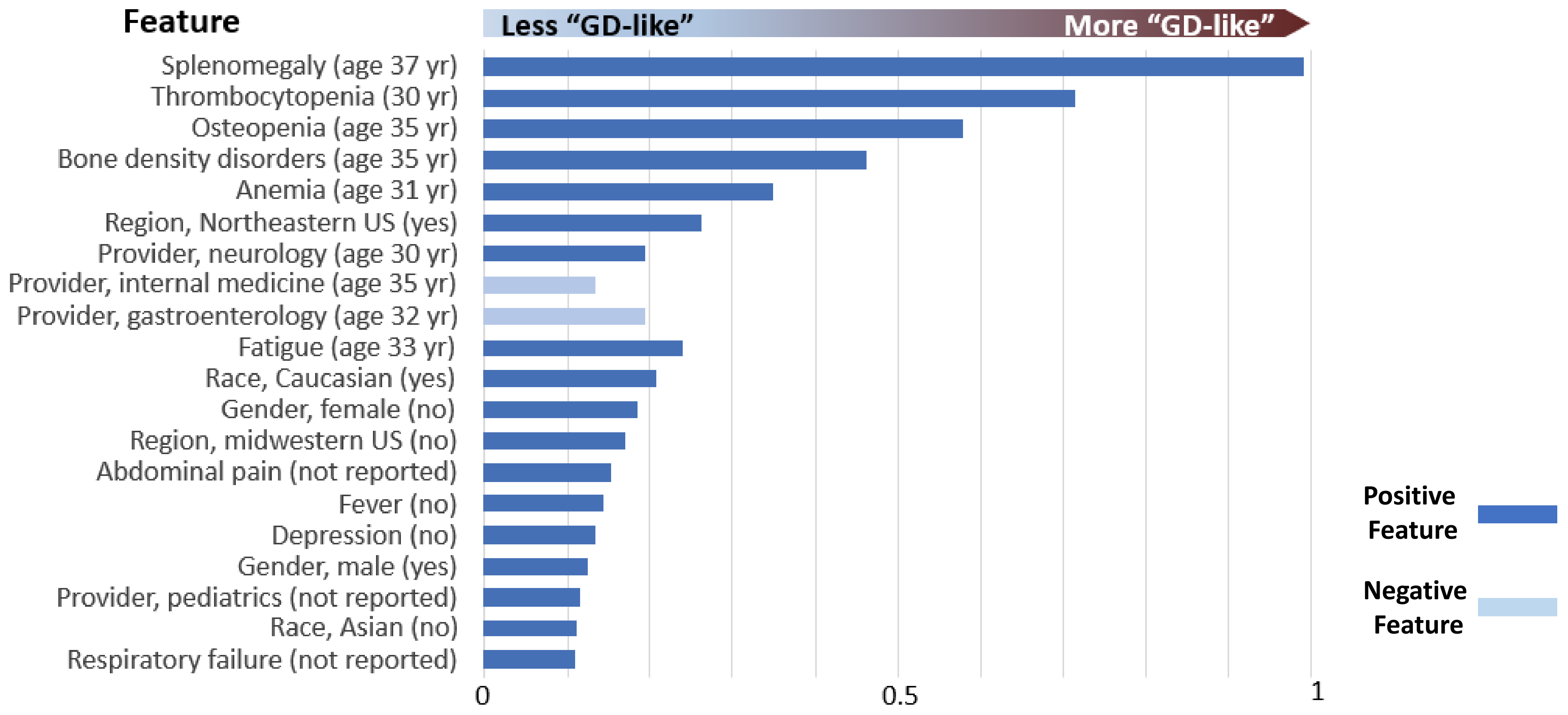
Prevalence model — **7x fewer patients**

vs.

Clinical Diagnostic Algorithm

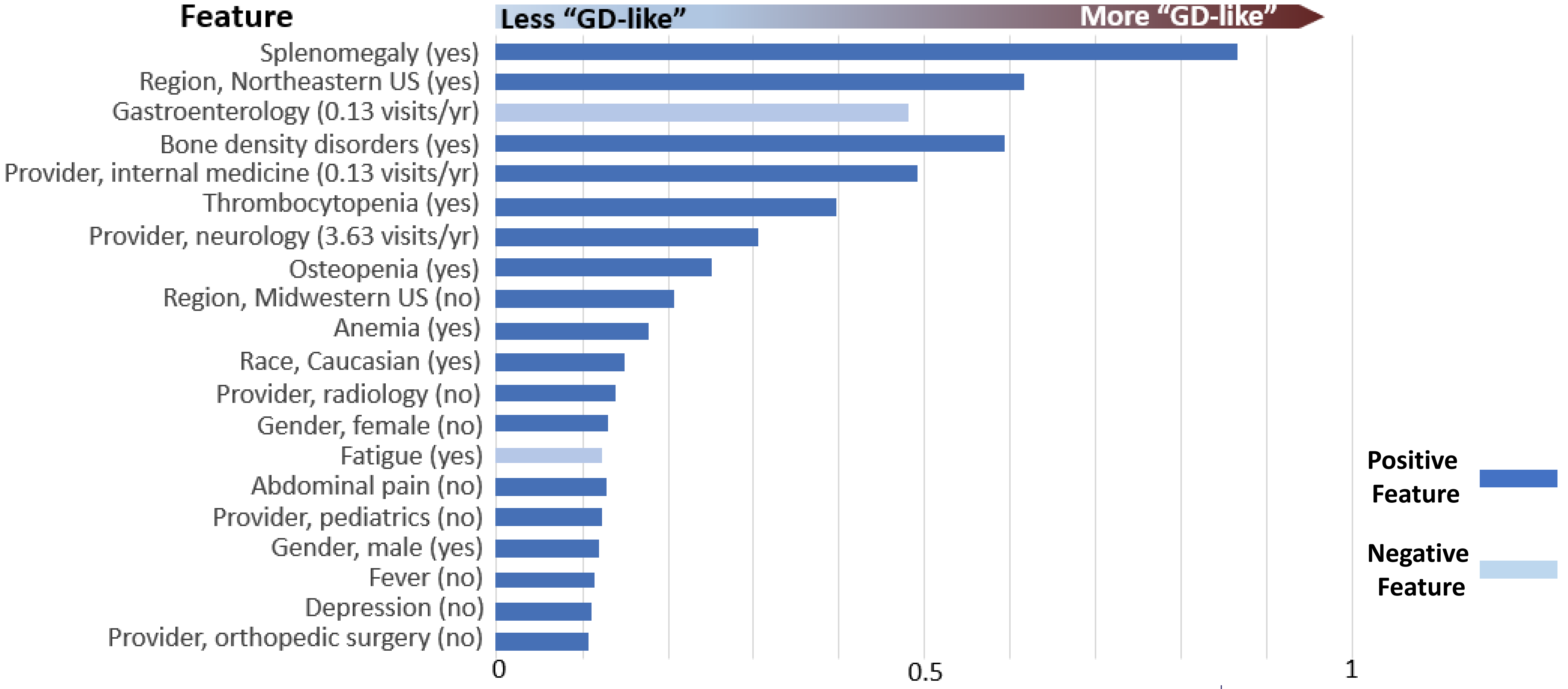
AI SUMMIT
COLUMBUS, OH • OCTOBER 25–27, 2022

Top Predictive Features – Age Model for Patient 1



AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

Top Predictive Features – Prevalence Model for Patient 1



Criteria for Exclusion
Clinical Decision Support Software 520(o)(1)(e) FD&C Act

1. Not intended to acquire, process, or analyze a medical image or a signal from an in vitro diagnostic device or a pattern or signal from a signal acquisition system
2. Intended for the purpose of displaying, analyzing, or printing medical information about a patient or other medical information
3. Intended for the purpose of supporting or providing recommendations to a health care professional about prevention, diagnosis, or treatment of a disease or condition
4. Intended for the purpose of enabling such health care professional to independently review the basis for such recommendations that such software presents so that it is not the intent that such health care professional rely primarily on any of such recommendations to make a clinical diagnosis or treatment decision regarding an individual patient

FDA Approval – Software as a Medical Device (SaMD)

- Engaged FDA via the Digital Health Center of Excellence
  - Deemed that algorithm doesn't meet criterion 3
  - Office of Health Technology 7 (OHT 7: in Vitro Diagnostics and Radiological Health)

Clinical Assessment & Evaluation for FDA Approval

- 3-4 sites
  - Select sites that together represent diverse racial and ethnic populations
- Identifying the top 50 ranked patients per site by the algorithm
- Implementation in 2023

Deployment

- Clinical infrastructure
  - Lysosomal storage disease clinician
  - Supporting specialists
- Bioinformatics team
  - Data size and quality
  - Health system priority

AI SUMMIT
COLUMBUS, OH • OCTOBER 25–27, 2022

Cross Functional Team

- Innovative approach to identifying patients
- Internal stakeholder involvement helps understand:
  - Execution
  - Timelines
  - Cost
  - Organizational risk
- External KOL stakeholder involvement provides:
  - Real-world patient management expertise
  - Insights from prior experience with big data projects
  - Impressions of this approach
  - Operational aspect to consider

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

Insights - Implementation of a Clinical Decision Support Algorithm

- Engage Clinicians, especially those on the front lines
- Clinical Evaluation
  - Study expands from clinical to bioinformatics
  - CROs less experienced at executing bioinformatics phase
- Data
  - Development of predictive models
  - Deployment
- Cross-functional teams – align across the business unit
  - Exploratory vs good development practices
  - Innovative way to reduce diagnostic delay

Acknowledgements

Real World Evidence and Data Science Teams - Sanofi

Amanda Wilson

Alexandra Dumitriu

Quinten

Alexandra Chiorean

Martin Montmerle

Margot Blanchon

Marie Génin

Simon Gosset

Mélissa Rollot

US and Global Medical

Lisa Sniderman King

Neha Shah

Mario Aguiar

Veronica Munoz

Judy Hull

Davorka Sekulik

External Consultants

Dr. Pramod Mistry

Dr. Neal Weinreb

Dr. François Modave

Strategic guidance and support:

Alaa Hamed, Jennifer Ibrahim, and Cliona Molony

Thank you.

Patrick Pavlick

Diagnostic and Informatics Director, Sanofi

Patrick.pavlick@sanofi.com

908.842.4390