XAI w medycynie

Katarzyna Kobylińska

Plan prezentacji:

1. XAI w modelach przesiewowych raka płuca:

- How to increase precision in defining lung cancer screening cohort – regression equation vs. explainable machine learning

2. XAI w medycynie - przegląd artykułów:

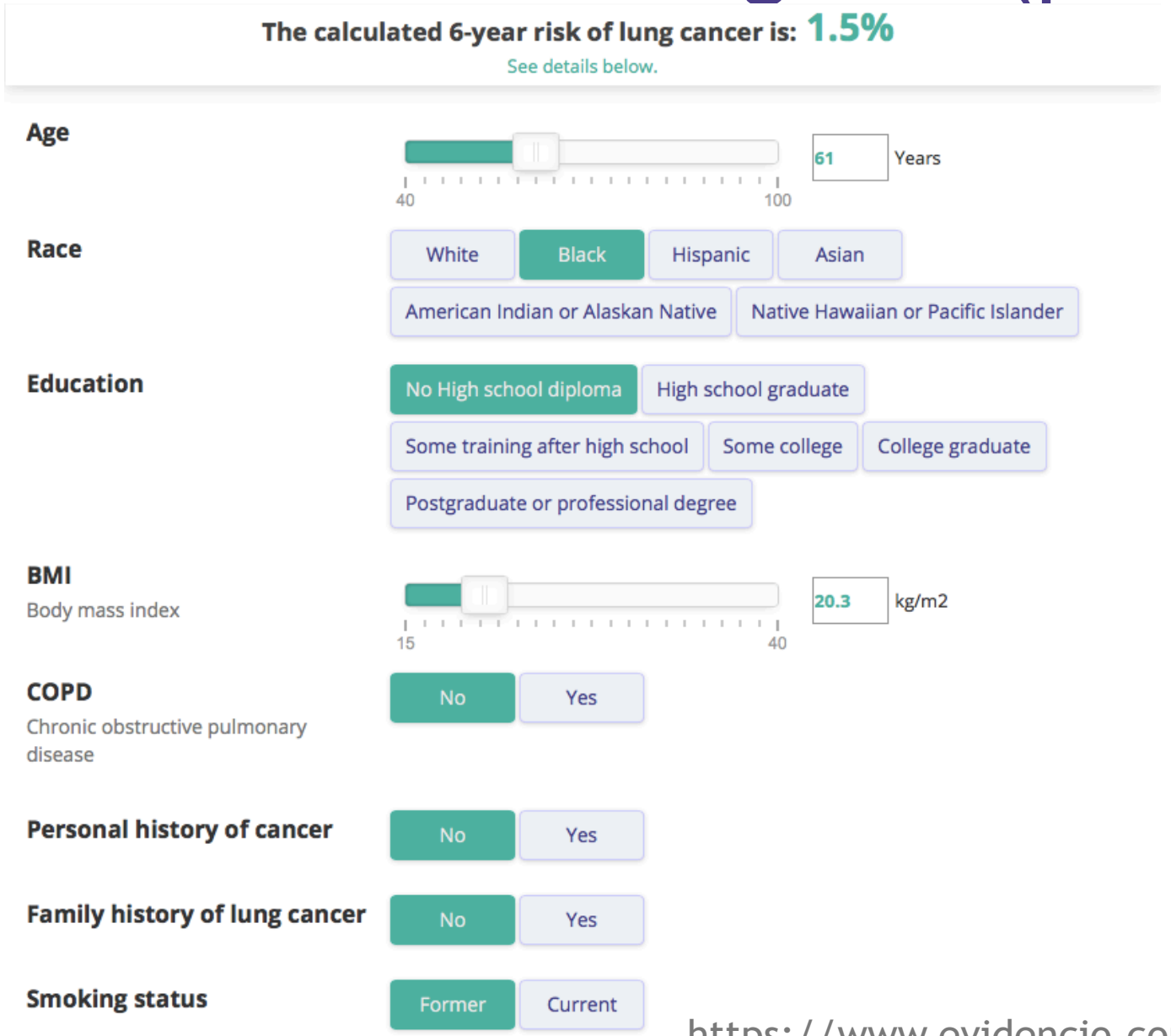
- Drug discovery with explainable artificial intelligence
- Explainable Machine Learning applied to Single-Nucleotide Polymorphisms for Systemic Lupus Erythematosus Prediction
- Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare

XAI for lung cancer screening models

Translational Lung Cancer Research

K. Kobylińska, T. Orłowski, M. Adamek, P.Biecek

Modele screeningowe (przesiewowe)



- Zmniejszają umieralność na raka płuca
- Mają na celu selekcję osób najbardziej zagrożonych ryzykiem raka płuca
- Modele popularne w USA

https://www.evidencio.com/models/show/992

Model Bacha:

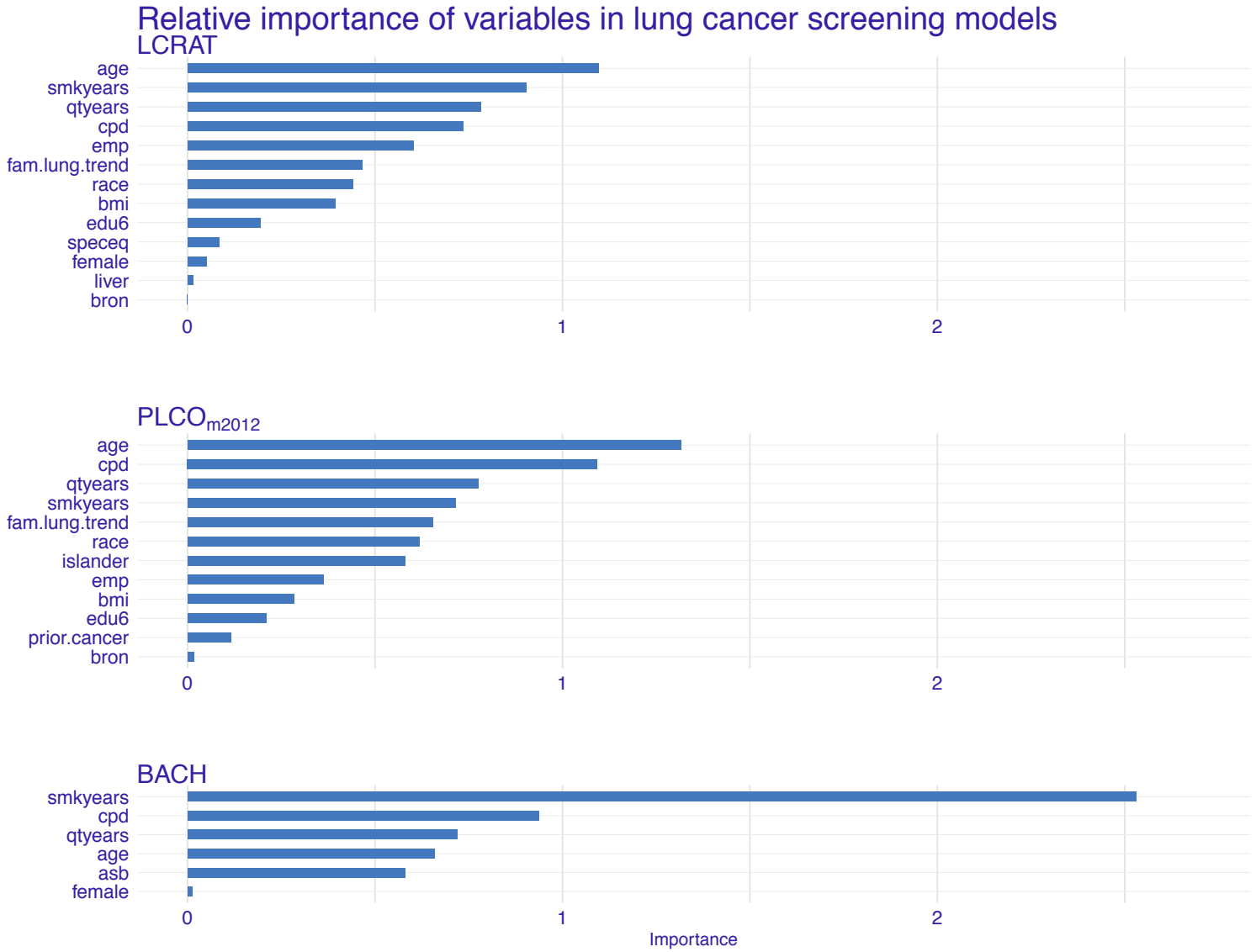
$$risk = \sum_{i=0}^{n} \left(1 - S_0^{exp(\beta X_i)}\right)\left(S_1^{exp(\beta X_i)}\right)\prod_{j<i}\left(S_0^{exp\left(\beta X_j\right)}\right)\left(S_1^{exp\left(\beta X_j\right)}\right)$$

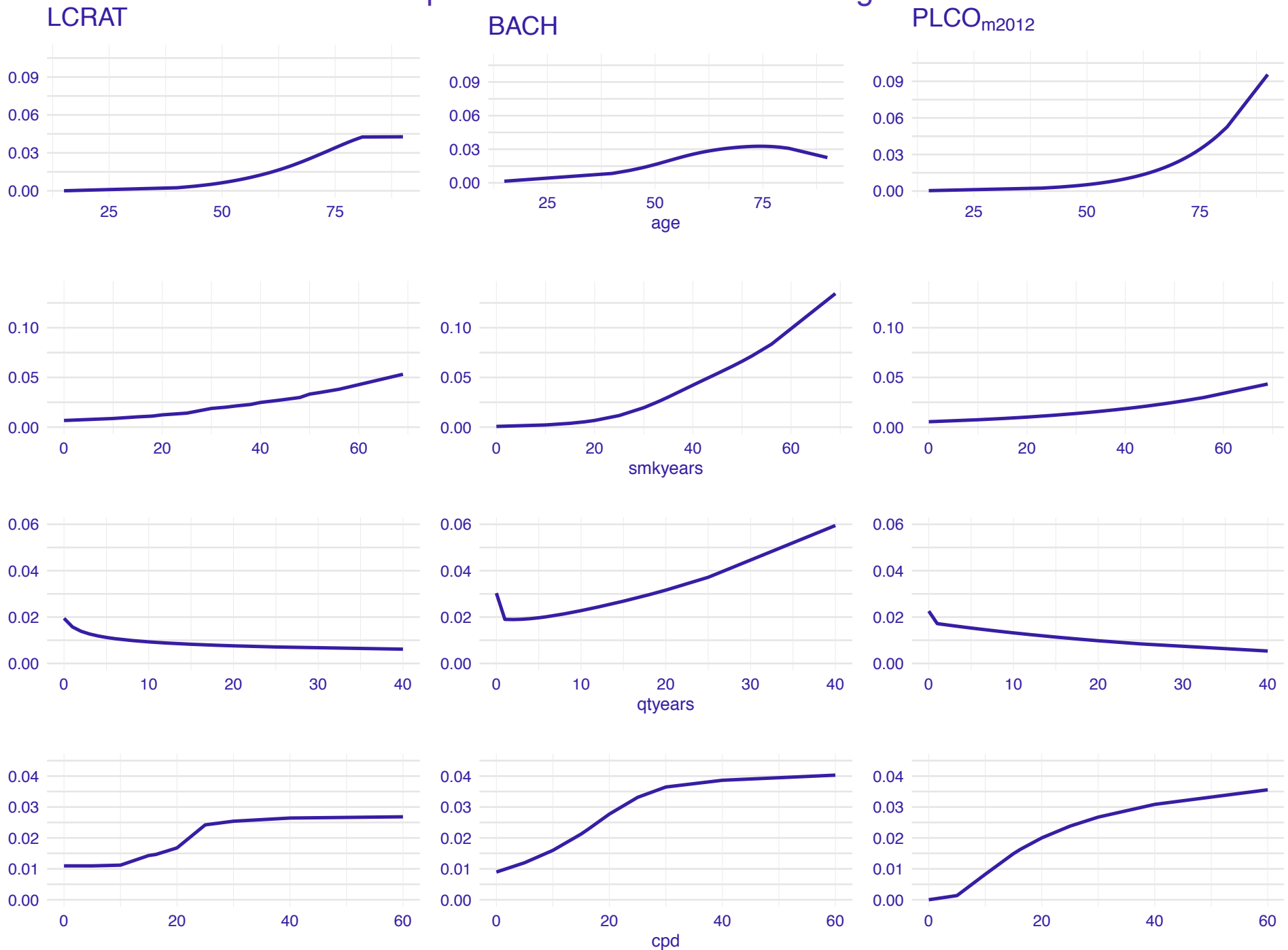
Variable	Expression	Coefficient
intercept		−9.7960571
age		0.070322812
age_2	$(age - 53.459001)^3 * I(age > 53))$	−0.00009382122
age_3	$(age - 61.954825)^3 * I(age > 61)$	0.00018282661
age_4	$(age - 70.910335)^3 * I(age > 70)$	−0.000089005389
female		−0.05827261
qtyears		−0.085684793
$qtyears_2$	$(qtyears)^3$	0.0065499693
$qtyears_3$	$(qtyears - 0.50513347)^3 * I(qtyears > 0)$	−0.0068305845
$qtyears_4$	$(qtyears - 12.295688)^3 * I(qtyears > 12)$	0.00028061519
smkyears		0.11425297
smkyears	$(smkyears - 27.6577)^3 * I(smkyears > 27)$	−0.000080091477
$smkyears_3$	$(smkyears - 40)^3 * I(smkyears > 40)$	0.00017069483
$smkyears_4$	$(smkyears-50.910335)^3*I(smkyears > 50)$	−0.000090603358
cpd		0.060818386
cpd_2	$(cpd - 15)^3 * I(cpd > 15)$	−0.00014652216
cpd_3	$(cpd - 20.185718)^3 * I(cpd > 20)$	0.00018486938
cpd_4	$(cpd - 40)^3 * I(cpd > 40)$	−0.000038347226
asbestos		0.2153936

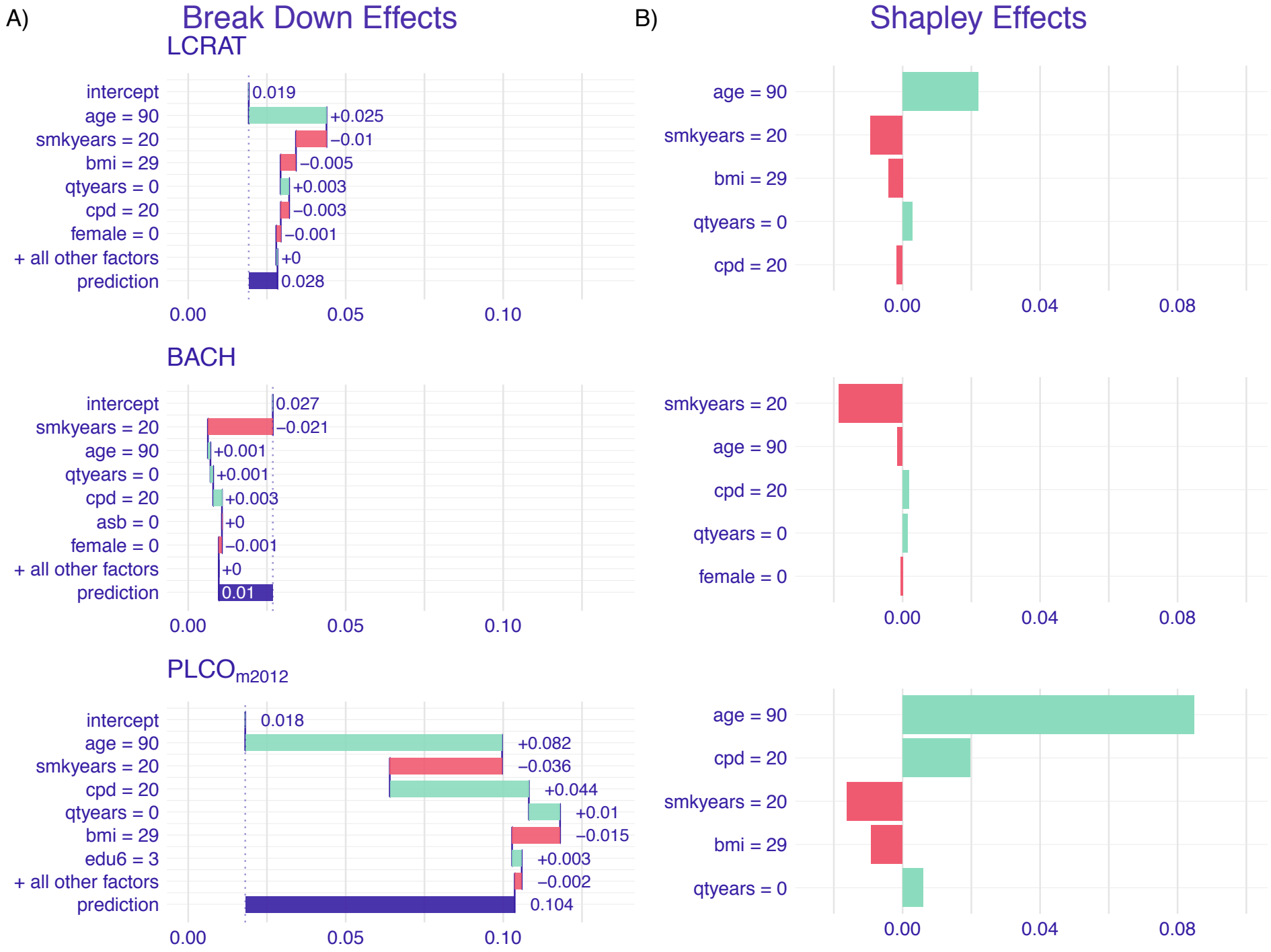
Dane

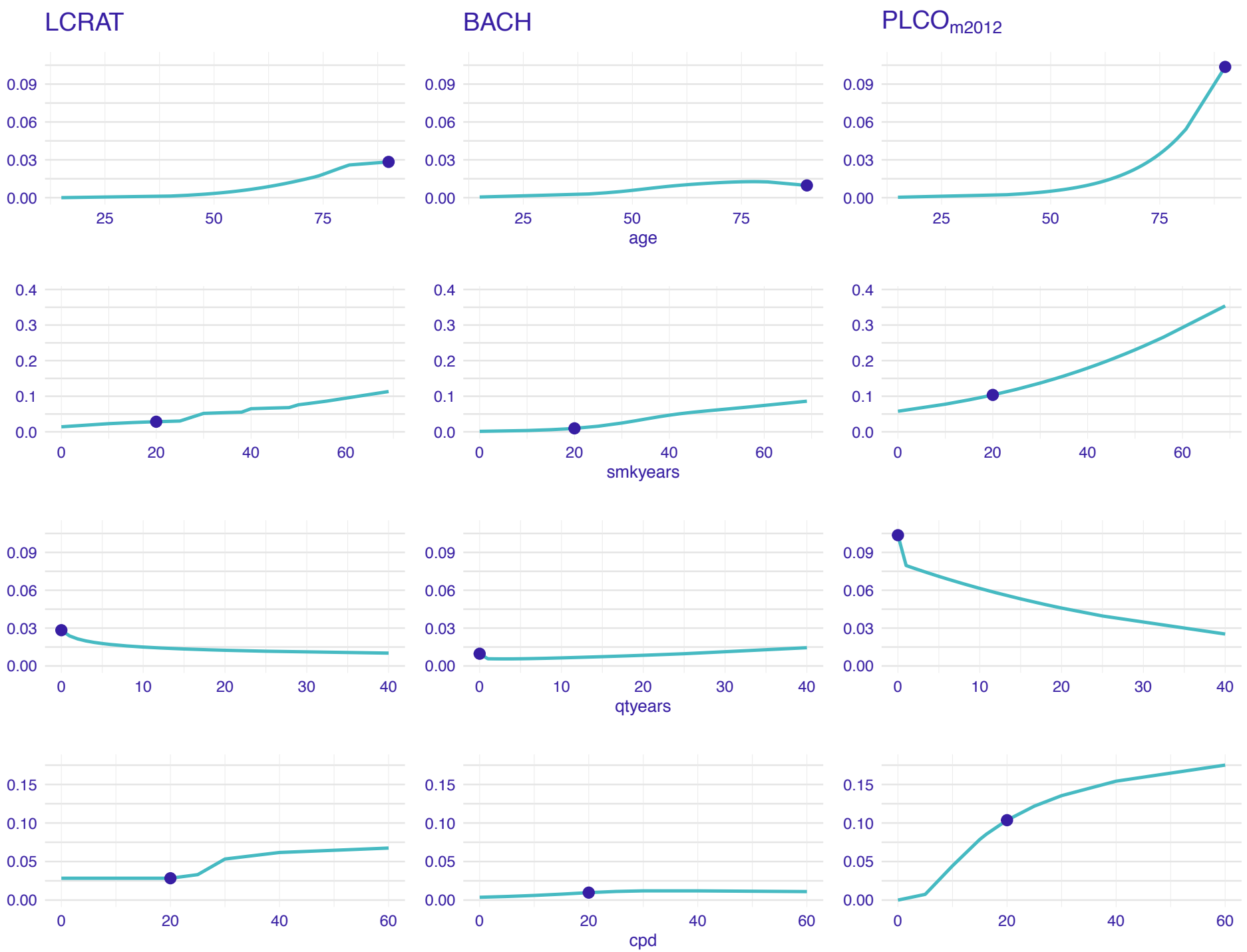
- baza chorych na operacyjnego raka płuca w Polsce w latach 2002- 2016
- pochodzi z Instytutu Gruźlicy i Chorób Płuc
- ~ 34,000 pacjentów
- dane zawierają:
    - historię palenia
    - patologiczne cechy guza
    - wyniki badań
    - symptomy choroby
    - choroby współistniejące

Porównanie 3 modeli screeningowych za pomocą XAI:




Partial Dependence Profiles for screening models
LCRAT
BACH
PLCOm2012
0.09
0.06
0.03
0.00
25
50
75
age
0.10
0.05
0
20
40
60
smkyears
0.04
0.02
10
30
qtyears
0.01
cpd


A)
Break Down Effects
LCRAT
intercept
age = 90
smkyears = 20
bmi = 29
qtyears = 0
cpd = 20
female = 0
+ all other factors
prediction
0.019
+0.025
−0.01
−0.005
+0.003
−0.003
−0.001
+0
0.028
0.00
0.05
0.10
BACH
intercept
smkyears = 20
age = 90
qtyears = 0
cpd = 20
asb = 0
female = 0
+ all other factors
prediction
0.027
−0.021
+0.001
+0.001
+0.003
+0
−0.001
+0
0.01
0.00
0.05
0.10
PLCO$_{m2012}$
intercept
age = 90
smkyears = 20
cpd = 20
qtyears = 0
bmi = 29
edu6 = 3
+ all other factors
prediction
0.018
+0.082
−0.036
+0.044
+0.01
−0.015
+0.003
−0.002
0.104
0.00
0.05
0.10
B)
Shapley Effects
age = 90
smkyears = 20
bmi = 29
qtyears = 0
cpd = 20
0.00
0.04
0.08
smkyears = 20
age = 90
cpd = 20
qtyears = 0
female = 0
0.00
0.04
0.08
age = 90
cpd = 20
smkyears = 20
bmi = 29
qtyears = 0
0.00
0.04
0.08


LCRAT
BACH
PLCO$_{m2012}$
0.09
0.06
0.03
0.00
25
50
75
age
0.4
0.3
0.2
0.1
0.0
0
20
40
60
smkyears
0.09
0.06
0.03
0.00
0
10
20
30
40
qtyears
0.15
0.10
0.05
0.00
0
20
40
60
cpd

Wnioski:

- Model Bacha działa nieintuicyjnie dla zmiennej określającej czas, który minął od rzucenia palenia
- Wprowadzenie XAI dla środowiska pulmonologów, onkologów - w przyszłości zastosowanie XAI do modeli screeningowych opartych na ML
- Lokalna analiza modeli bardzo istotna w kontekście spersonalizowanej medycyny i określenia ryzyka dla konkretnego pacjenta

XAI w medycynie - przegląd

Drug discovery with explainable artificial intelligence

Nature Machine Intelligence

José Jiménez-Luna , Francesca Grisoni and Gisbert Schneider

Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland.

Table 1 | Computational approaches towards explainable AI in drug discovery and related disciplines, categorized according to the respective methodological concept

Family	Aim	Methods	Reported applications in drug discovery
Feature attribution	Determine local feature importance towards a prediction	• Gradient based • Surrogate models • Perturbation based	Ligand pharmacophore identification[55,71,79,80], structural alerts for adverse effect[67], protein–ligand interaction profiling[72]
Instance based	Compute a subset of features that need to be present or absent to guarantee or change a prediction	• Anchors • Counterfactual instances • Contrastive explanations	Not reported
Graph convolution based	Interpret models within the message-passing framework	• Subgraph approaches • Attention based	Retrosynthesis elucidation[101], toxicophore and pharmacophore identification[41], ADMET[a][102,103] reactivity prediction[104]
Self-explaining	Develop models that are explainable by design	• Prototype based • Self-explaining neural networks • Concept learning • Natural language explanations	Not reported
Uncertainty estimation	Quantify the reliability of a prediction	• Ensemble based • Probabilistic • Other approaches	Reaction prediction[147], active learning[148], molecular activity prediction[168]

For each family of approaches, a brief description of its aim is provided, along with specific methods and reported applications in drug discovery. 'Not reported' refers to families of methods that, to the best of our knowledge, have not been yet applied in drug discovery. Potential applications of these are discussed in the main text. [a]ADMET: absorption, distribution, metabolism, excretion and toxicity.

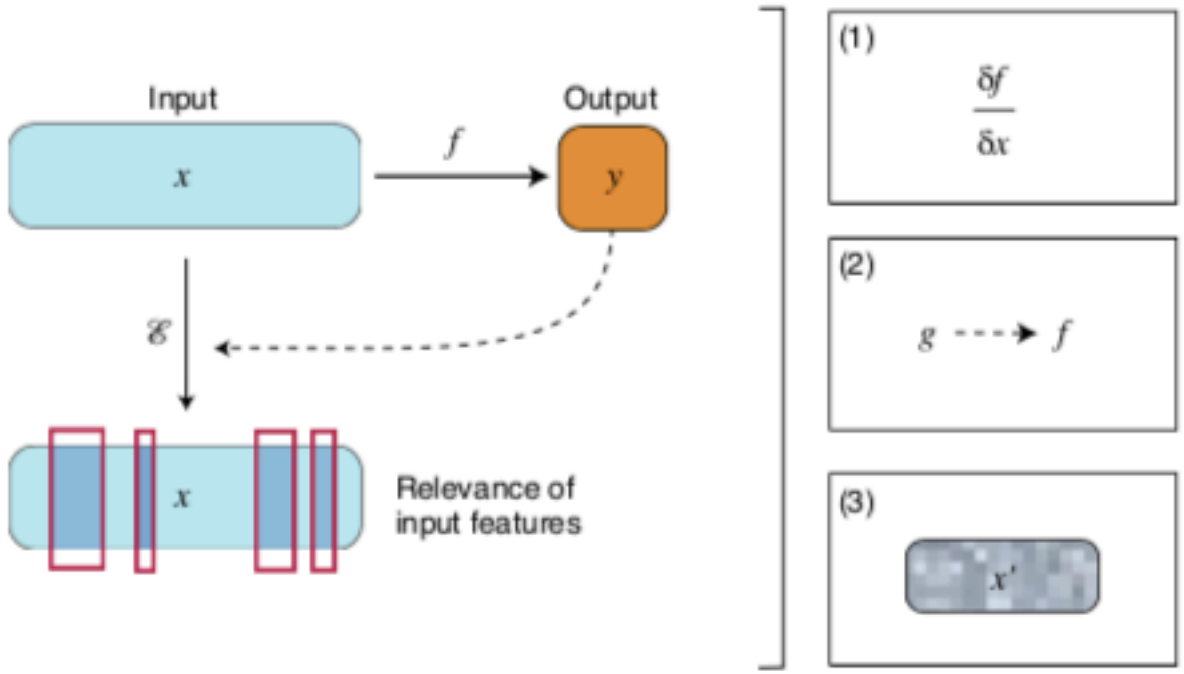
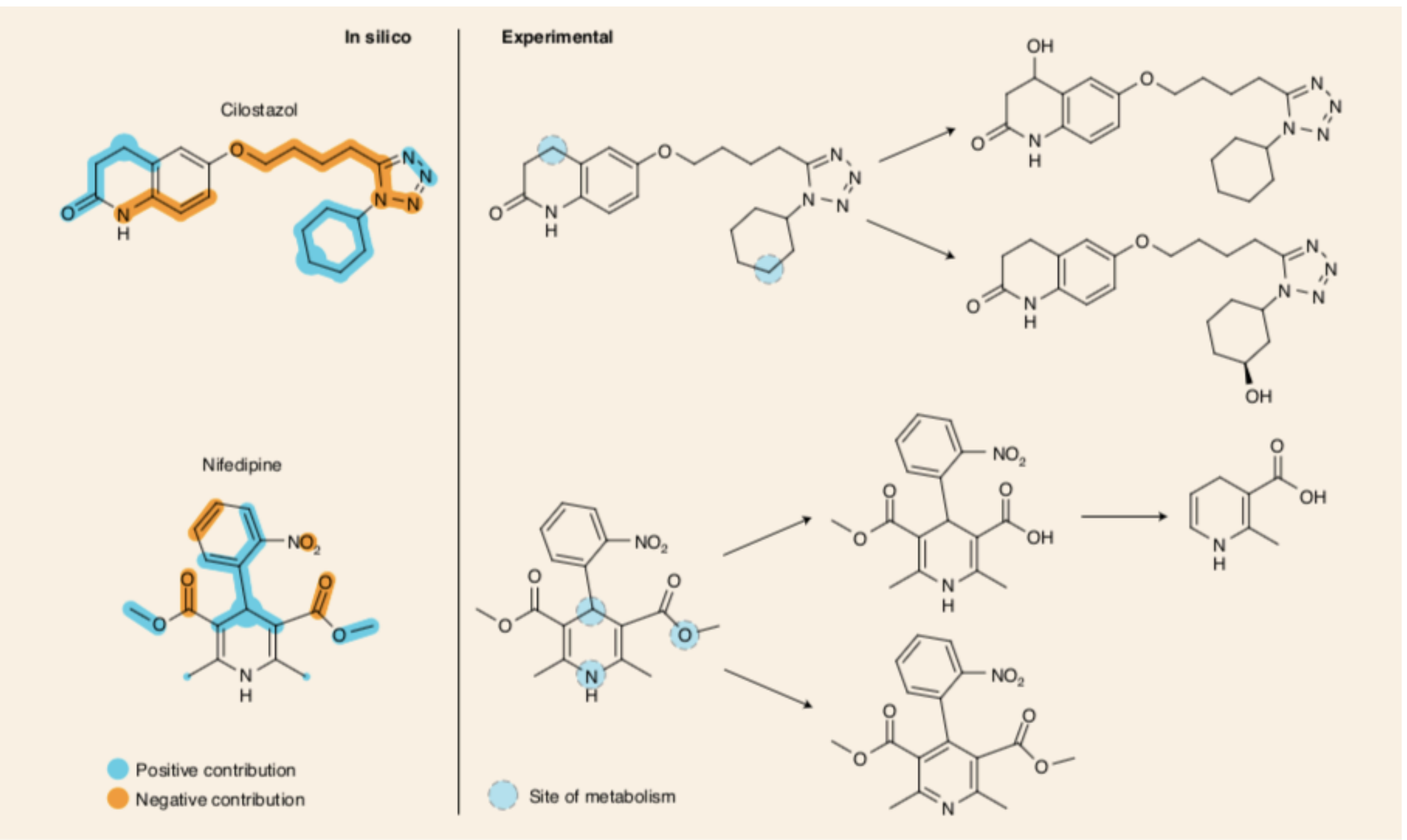


Fig. 1 | Feature attribution methods. Given a neural network model *f*, which computes the prediction $y = f(x)$ for input sample *x*, a feature attribution method $\mathcal{E}$ outputs the relevance of every input feature of *x* for the prediction. There are three basic approaches to determine feature relevance: (1) gradient-based methods, computing the gradient of the network *f* with respect to the input *x*, (2) surrogate methods, which approximate *f* with a human-interpretable model *g*, and (3) perturbation-based methods, which modify the original input to measure the respective changes in the output.


In silico
Experimental
Cilostazol
Nifedipine
Positive contribution
Negative contribution
Site of metabolism

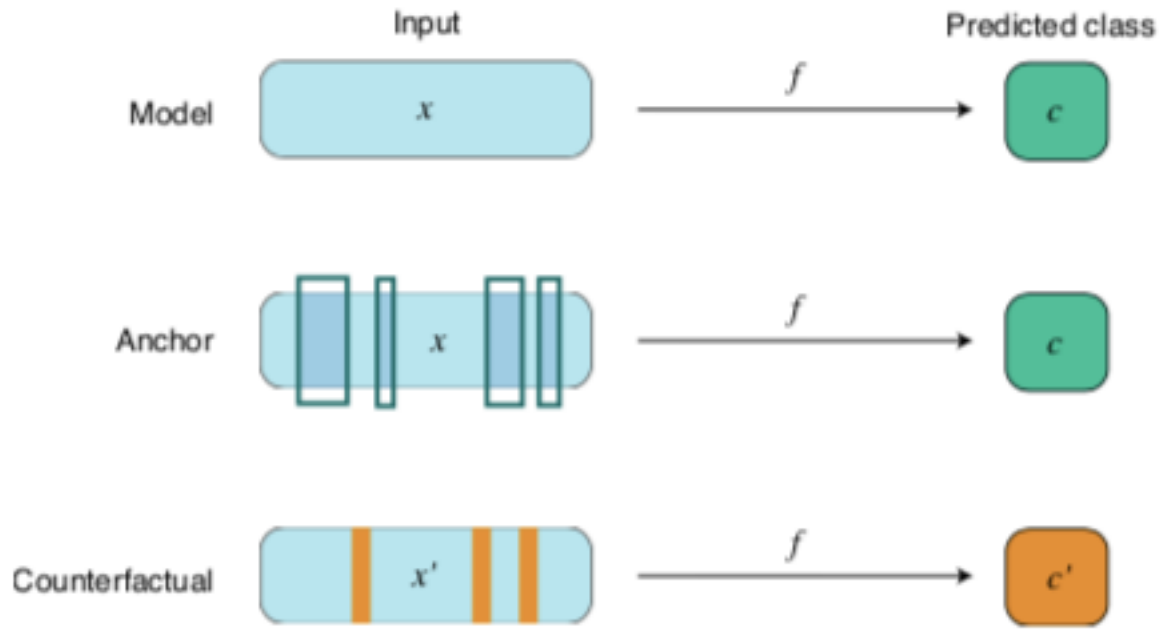


Fig. 2 | Instance-based model interpretation. Given a model *f*, input instance *x* and the respective predicted class *c*, so-called anchor algorithms identify a minimal subset of features of *x* that are sufficient to preserve the predicted class assignment *c*. Counterfactual search generates a new instance *x'* that lies close in feature space to *x* but is classified differently by the model, as belonging to class *c'*.

Wnioski:

- **Dogłębna znajomość dziedziny problemowej jest kluczowa** dla określenia, które decyzje modelowe wymagają dalszych wyjaśnień, które typy odpowiedzi są znaczące dla użytkownika, a które są trywialne lub oczekiwane.
- W przypadku podejmowania decyzji przez ludzi wyjaśnienia wygenerowane za pomocą XAI muszą być nietrywialne, niesztuczne i dostatecznie pouczające dla odpowiedniej społeczności naukowej.

MI

XAI w medycynie - przegląd

Explainable Machine Learning applied to Single-Nucleotide Polymorphisms for Systemic Lupus Erythematosus Prediction

Conference: 11th International Conference on Information, Intelligence, Systems and Applications (IISA 2020)At: Athens, Greece

Marc Jermaine Pontiveros, Cherica Tee, Geoffrey Aserios Solano, Michael Tee

University of the Philippines Manil, Philippine General Hospital

Modelowanie ryzyka zachorowania na toczeń

SLE - autoimmunologiczna choroba atakująca wiele narządów, wykryta zbyt późno często prowadzi do śmierci.

Przyczyna nie jest znana, podejrzewa się że wpływ na zachorowanie mają czynniki genetyczne

- Ryzyko zachorowania na toczeń uwzględniając polimorfizmy nukleotydów.

CROSS-VALIDATION METRICS OF THE BEST PERFORMING MODELS (MEAN (STD))

Model	Acc	AUC	AUCPR	Pre	Rec	Spe
GLM_best	75.63 (6.67)	76.44 (7.21)	75.35 (6.97)	71.07 (8.13)	87.58 (7.44)	60.54 (20.17)
RF_best	77.25 (5.32)	77.14 (4.70)	73.77 (7.98)	74.07 (7.16)	85.79 (9.57)	65.87 (19.02)
GBM_best	76.17 (4.94)	76.88 (3.85)	74.88 (2.90)	71.74 (5.04)	88.46 (8.02)	60.39 (18.58)

Fig. 3. Arena dashboard. Users can select model of interest and drag plots available into the working space. Users can choose different observations or different features to study and pages can be added to create another blank working space. Users can navigate across pages containing plots

Fig. 4. Separate and merged feature importance plots for the three models. Features are ranked differently and some identified features are exclusive to a model

Wnioski:

- Zidentyfikowanie istotnych cech: np. rs12734338
- Modele w inny sposób traktują te same cechy
- Analiza lokalna pomocna w ramach spersonalizowanej medycyny
- Dashboard, który umożliwia dalszą eksplorację

XAI w medycynie - przegląd

Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare

Nature

Davide Cirillo, Silvina Catuara-Solarz, Czuee Morey, Emre Guney, Laia Subirats, Simona Mellino, Annalisa Gigante, Alfonso Valencia, María José Rementeria, Antonella Santuccione Chadha & Nikolaos Mavridis

Barcelona Supercomputing Center

Cele pracy:

- Podkreślenie głównych dostępnych typów danych biomedycznych oraz **roli sztucznej inteligencji w zrozumieniu płci i różnic między płcią w medycynie**
- Istniejące i potencjalne uprzedzenia oraz ich wkład w tworzenie spersonalizowanych interwencji terapeutycznych. Badanie kwestii płci związanej z generowaniem i gromadzeniem danych eksperymentalnych, klinicznych
- Przegląd wielu technologii w celu analizy i wdrażania danych (analiza dużych zbiorów danych, nlp, robotyka)

Desirable Bias:

- Płeć stanowi istotne źródło zmienności w wielu badaniach medycznych, wpływające na czynniki ryzyka, wiek zachorowania, objawy, rokowanie, biomarkery i skuteczność leczenia.
- Różnice płci odnotowano w chorobach tj.: cukrzyca, zaburzenia sercowo-naczyniowe, choroby neurologiczne, nowotwory
- Uwzględnienie płci i różnic między płciami w celu postawienia precyzyjnej diagnozy

Spektrum Autyzmu:
Brak obecnie uwzględnienia wykazanych zależności różnic płci w symptomatologii związanej z zaburzeniami komunikacji i interakcji społecznej, zachowaniami ekspresyjnymi, rozmową wzajemną, gestami niewerbalnymi w celach diagnostycznych

Zaburzenia sercowo- naczyniowe:
Chociaż udokumentowano, że mężczyźni i kobiety różnie reagują na wiele leków sercowo-naczyniowych, tj: statyny, czy beta-blokery, przyjęte metody leczenia nie uwzględniają różnic płciowych

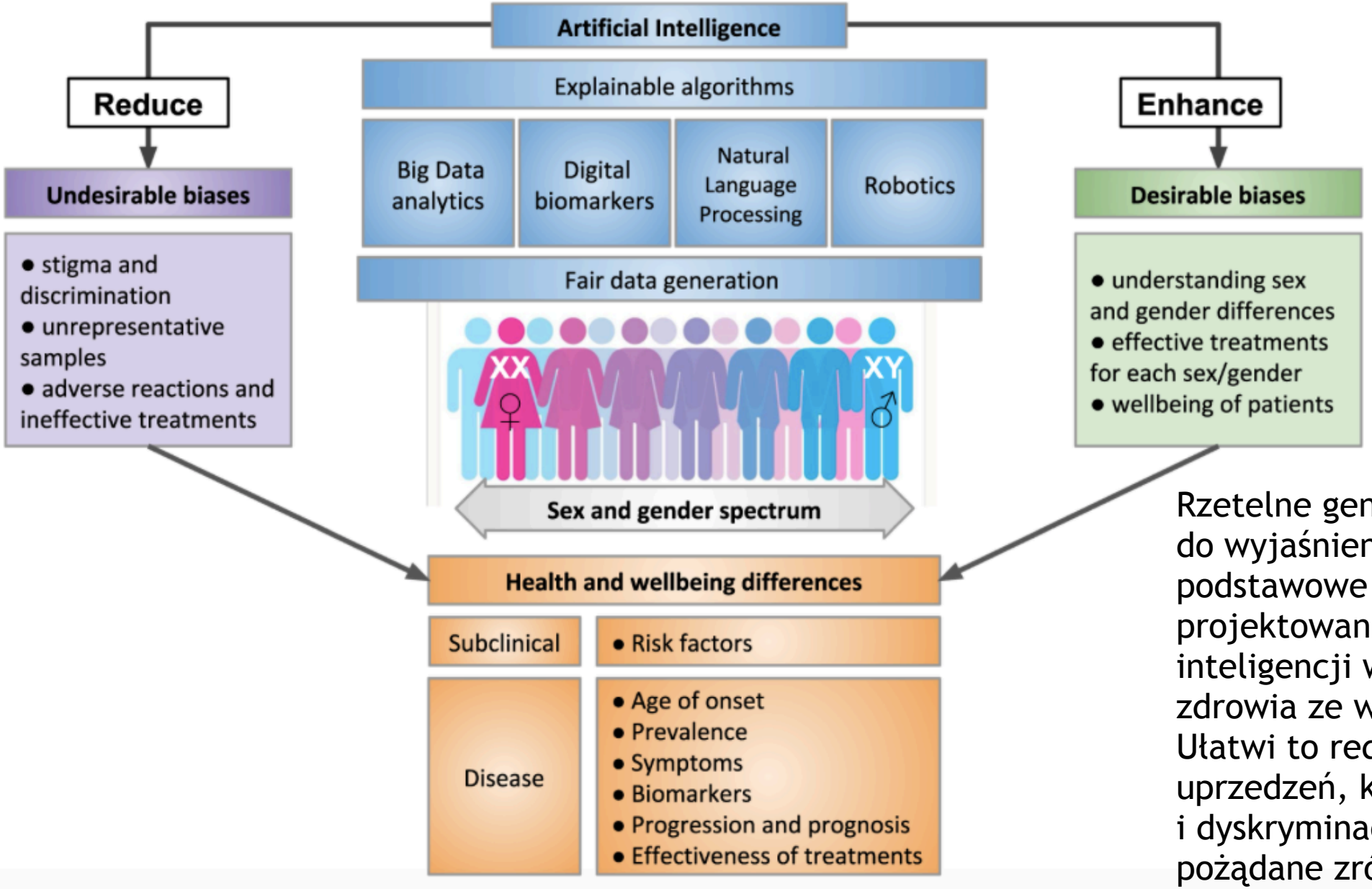
Undesirable Bias:

- Przejaw niezamierzonej bądź niepotrzebnej dyskryminacji
- Dzieje się tak, gdy różnicuje się stan zdrowia ze względu na płeć pomimo braku wyczerpujących dowodów na ich poparcie lub są oparte na wypaczonych dowodach.

  Depresja:
  Na przykład badania epidemiologiczne wskazują, że częściej występuje depresja wśród kobiet, jednak może to wynikać z wypaczonego rozpoznania ze względu na skale depresji mierzącą objawy, które częściej występują u kobiet

- Istnieje wiele źródeł niepożądanych błędów, które mogą zostać przypadkowo wprowadzone w algorytmach sztucznej inteligencji. Najczęstszą z nich jest brak reprezentatywnej próby populacji w uczącym zbiorze danych. W niektórych przypadkach w całej populacji może występować uprzedzenie wynikające z przyczyn społecznych, historycznych lub instytucjonalnych. W innych przypadkach sam algorytm, a nie zbiór danych uczących, może wprowadzać odchylenie, zaciemniając nieodłączną dyskryminację lub wywołując nieuzasadnioną lub nieistotną selektywność.

MI



Rzetelne generowanie danych i możliwe do wyjaśnienia algorytmy to podstawowe wymagania dotyczące projektowania i stosowania sztucznej inteligencji w celu optymalizacji zdrowia ze względu na płeć.
Ułatwi to redukcję niepożądanych uprzedzeń, które propagują nierówność i dyskryminację, oraz będzie promować pożądane zróżnicowanie, które pomoże w rozwoju medycyny precyzyjnej.

Wyjaśnienia modeli AI:

- coraz ważniejsze by móc wyciągać wnioski kliniczne, które mają wpływa na życie pacjentów
- pomogłyby uzasadnić prognozy kliniczne gdy są one zróżnicowane dla pacjentów o różnej płci
- pozwoliłyby na znalezienie potencjalnych błędnych wniosków wynikających z modelowania z błędnymi danymi
- pomogłyby w zidentyfikowaniu niepożądanych uprzedzeń (undesirable biases)
- pomogłyby w odkryciu różnic między płciami w celu spersonalizowanej profilaktyki i terapii.

Przykłady XAI w zastosowaniach medycznych:

Choroby siatkówki:

jest w badanie, w którym algorytm uczenia maszynowego wydał zalecenia dotyczące dziesiątek chorób siatkówki, zwracając uwagę na określone struktury w tomografii optycznej, które mogą prowadzić do niejednoznacznej interpretacji.

Choroby sercowo- naczyniowe

Innym przykładem jest model głębokiego uczenia do przewidywania czynników ryzyka sercowo-naczyniowego na podstawie obrazów siatkówki, wskazujący, które cechy anatomiczne, takie jak tarcza nerwu wzrokowego lub naczynia krwionośne, zostały wykorzystane do wygenerowania prognoz.