Lingnan University 嶺南大學 香港 Hong Kong

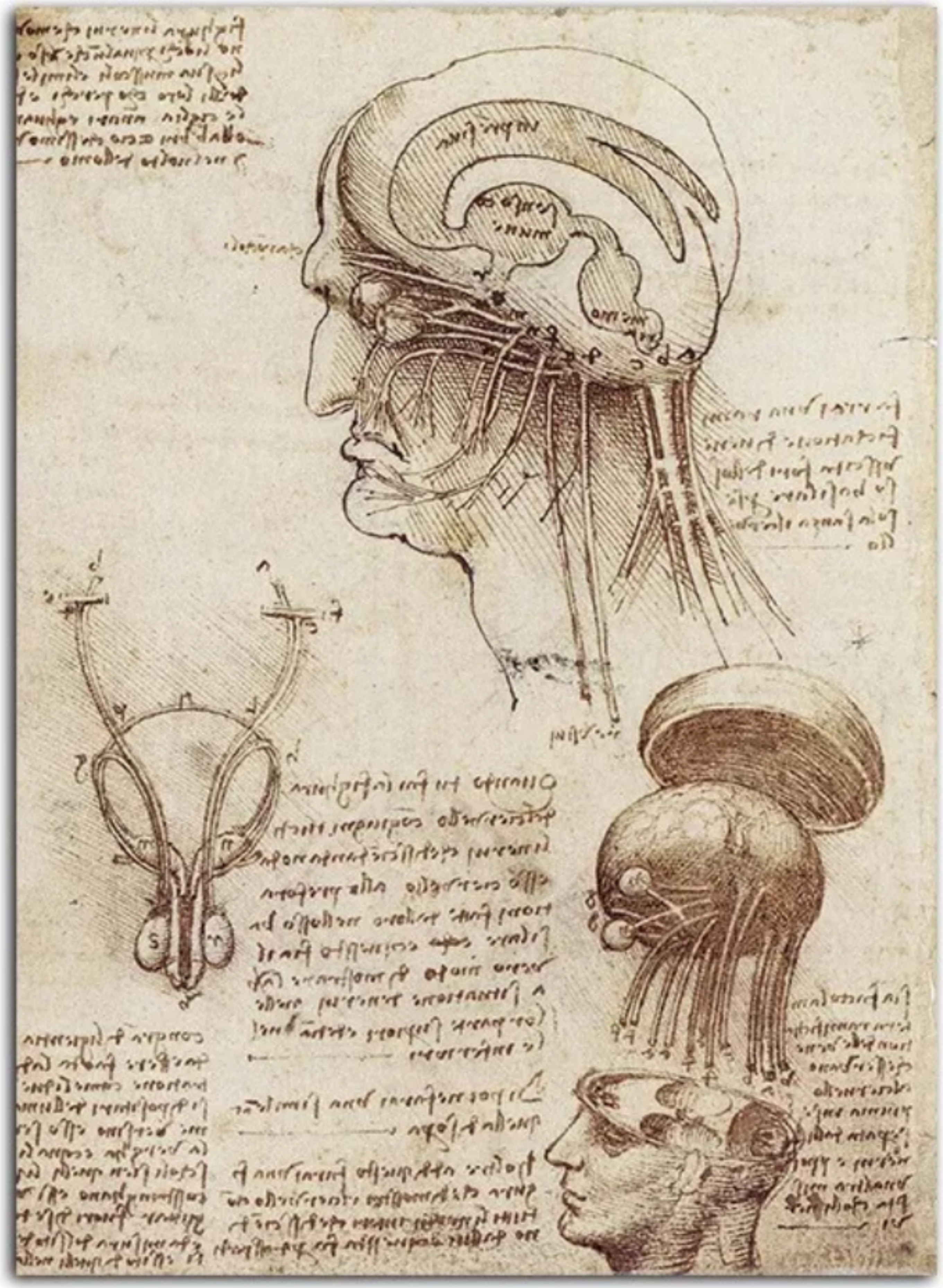
Leveraging Machine Learning for Cutting-Edge Innovations in Social Science Research

Korea University

Department of Sociology

Eun Kyong SHIN

申恩卿



Eun Kyong Shin

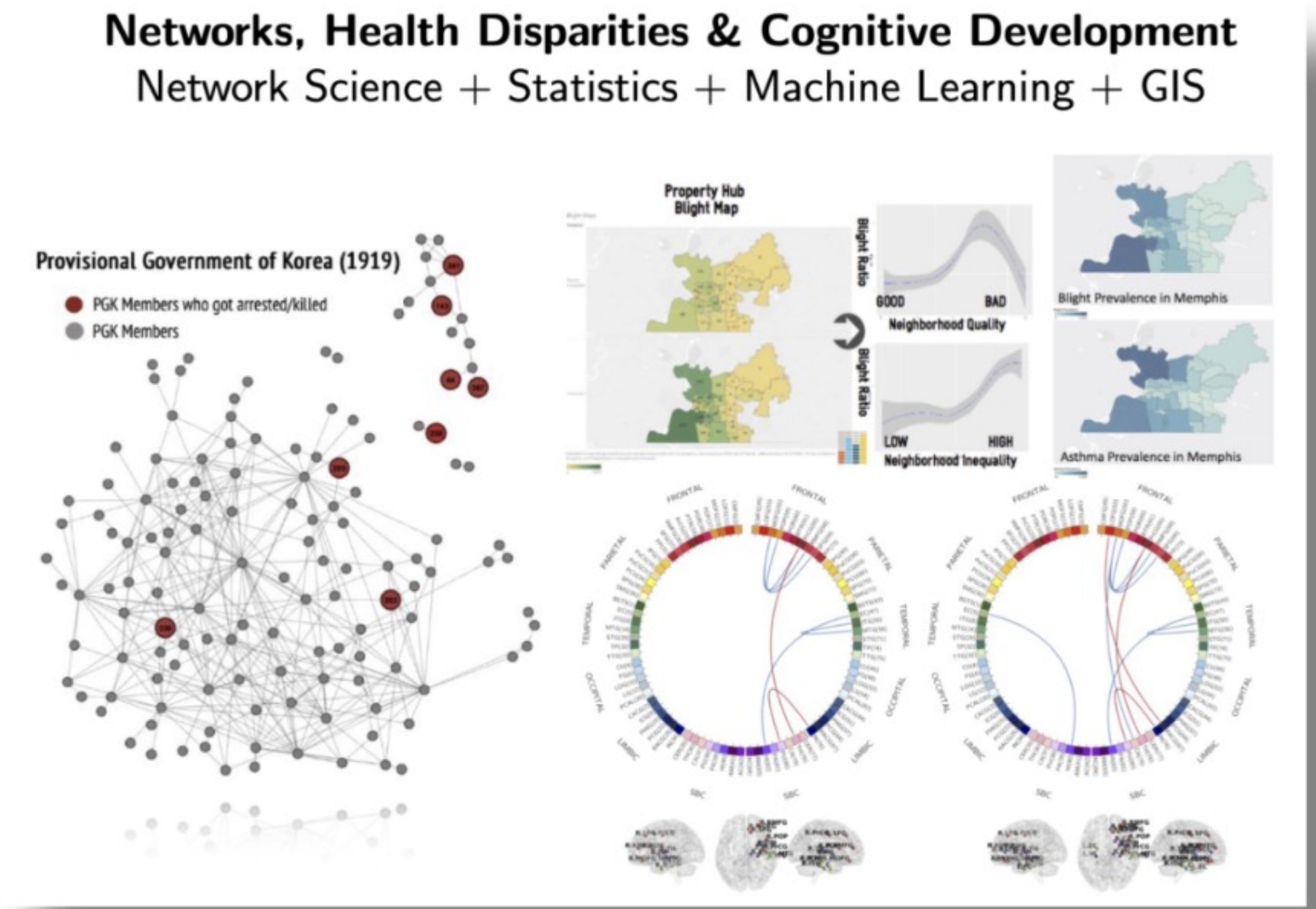
Korea Univerity

Columbia University

Colubia Law School

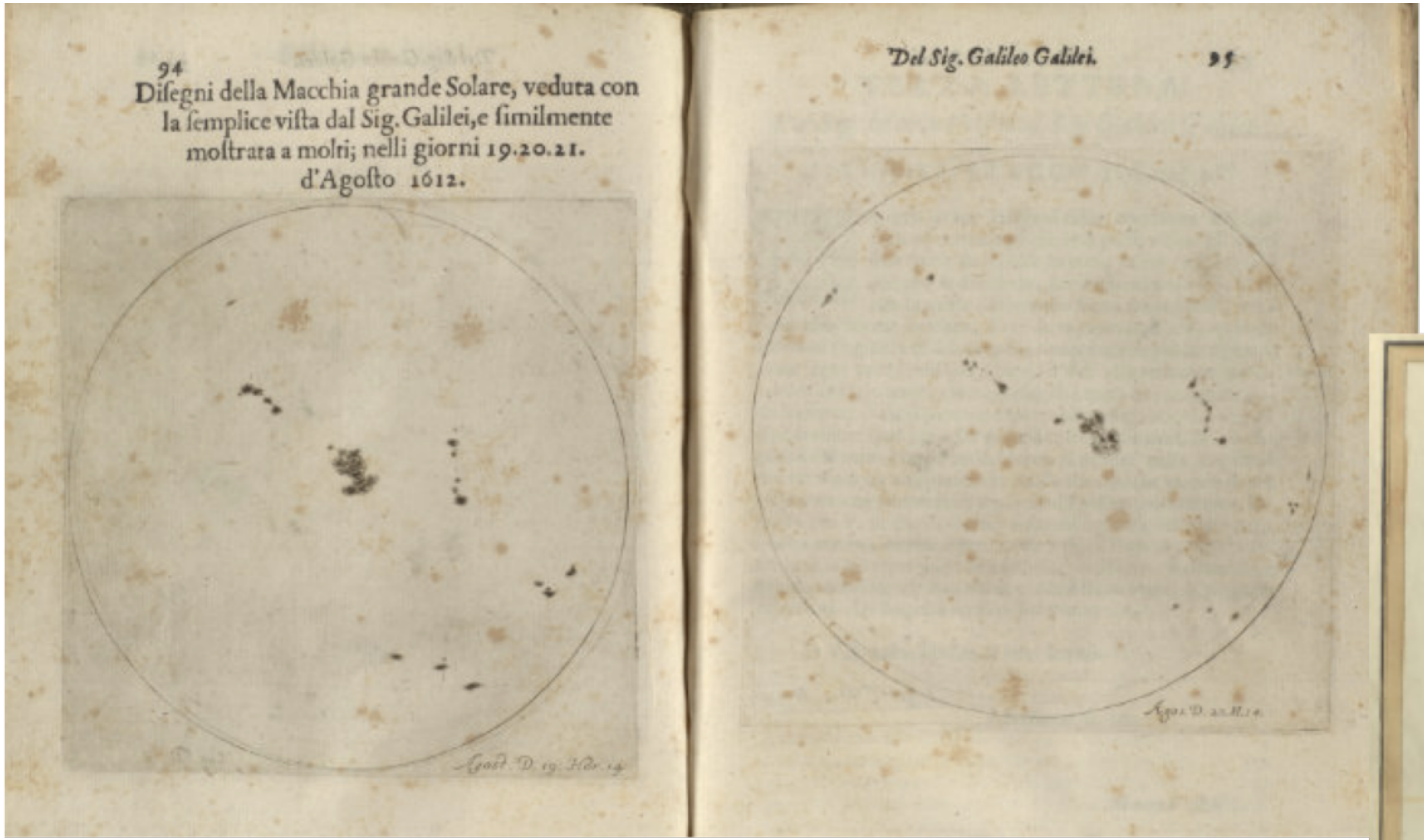
U of T Medical School

LAB SOCIOMARKERS

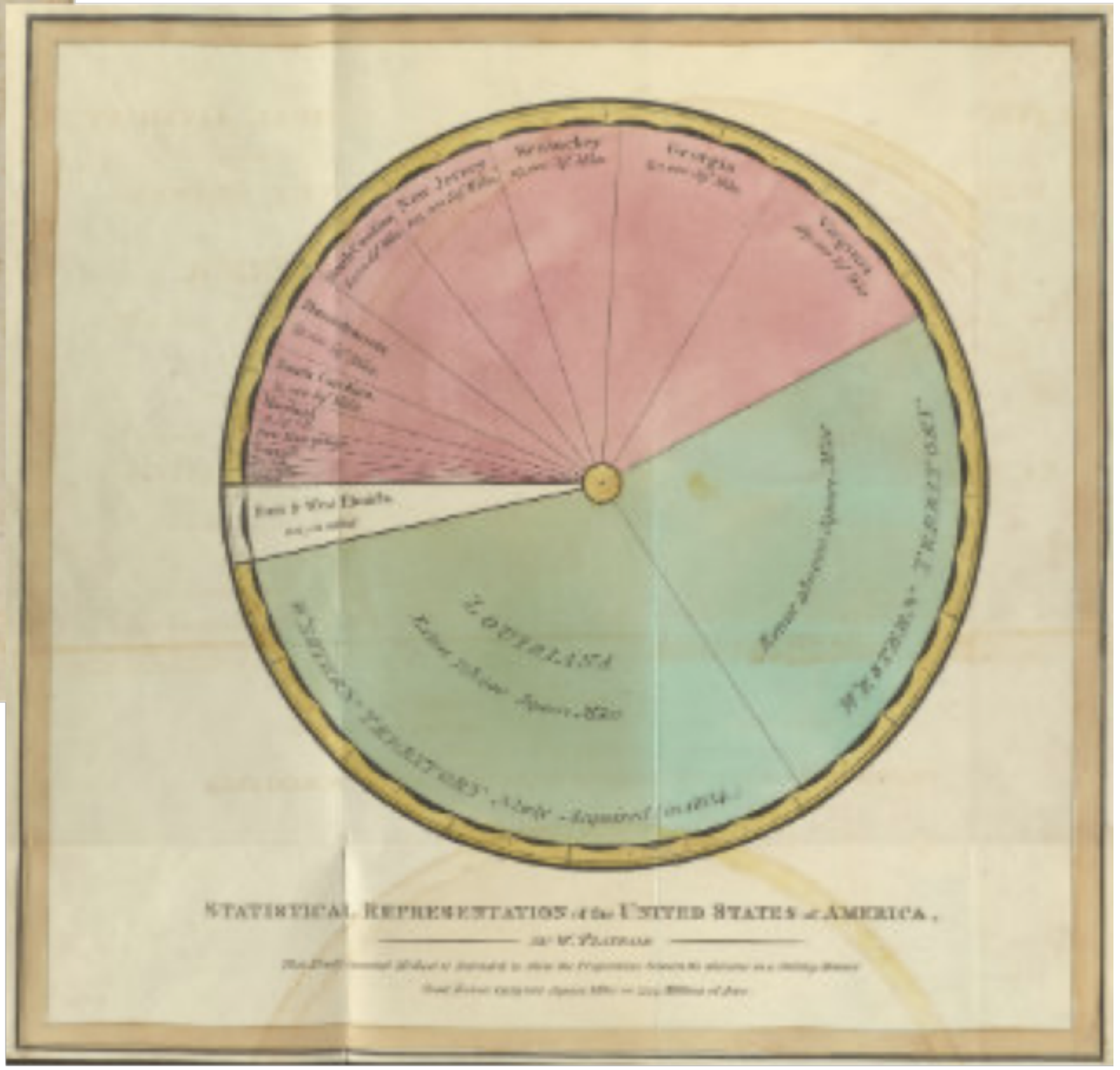


What is DATA?

Why do we need them?



Illustrations of the location, size and shape of sunspots that Galileo observed by projecting the sun onto paper through a telescope.



https://www2.lehigh.edu/news/a-history-of-data-visualization

1925 미국 인구조사

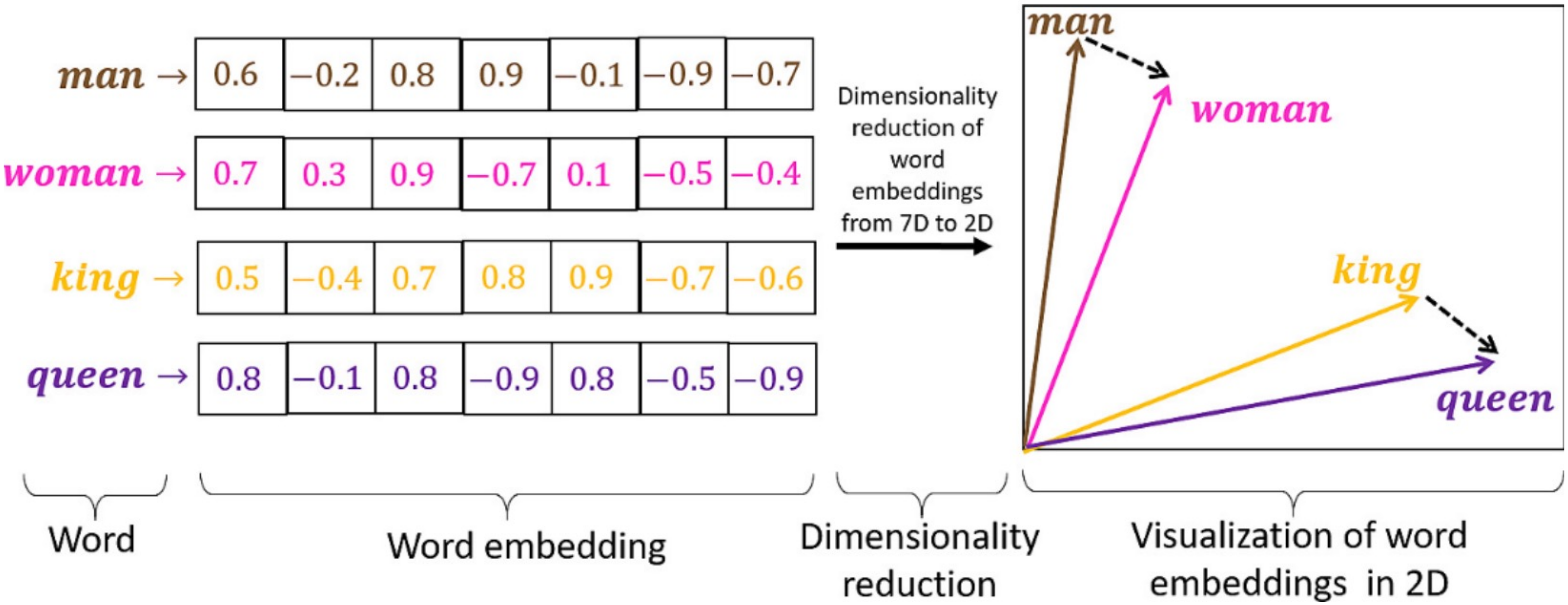
1960년 한국 농촌 국세조사

https://theme.archives.go.kr/next/koreaOfRecord/census.do

What is big data?

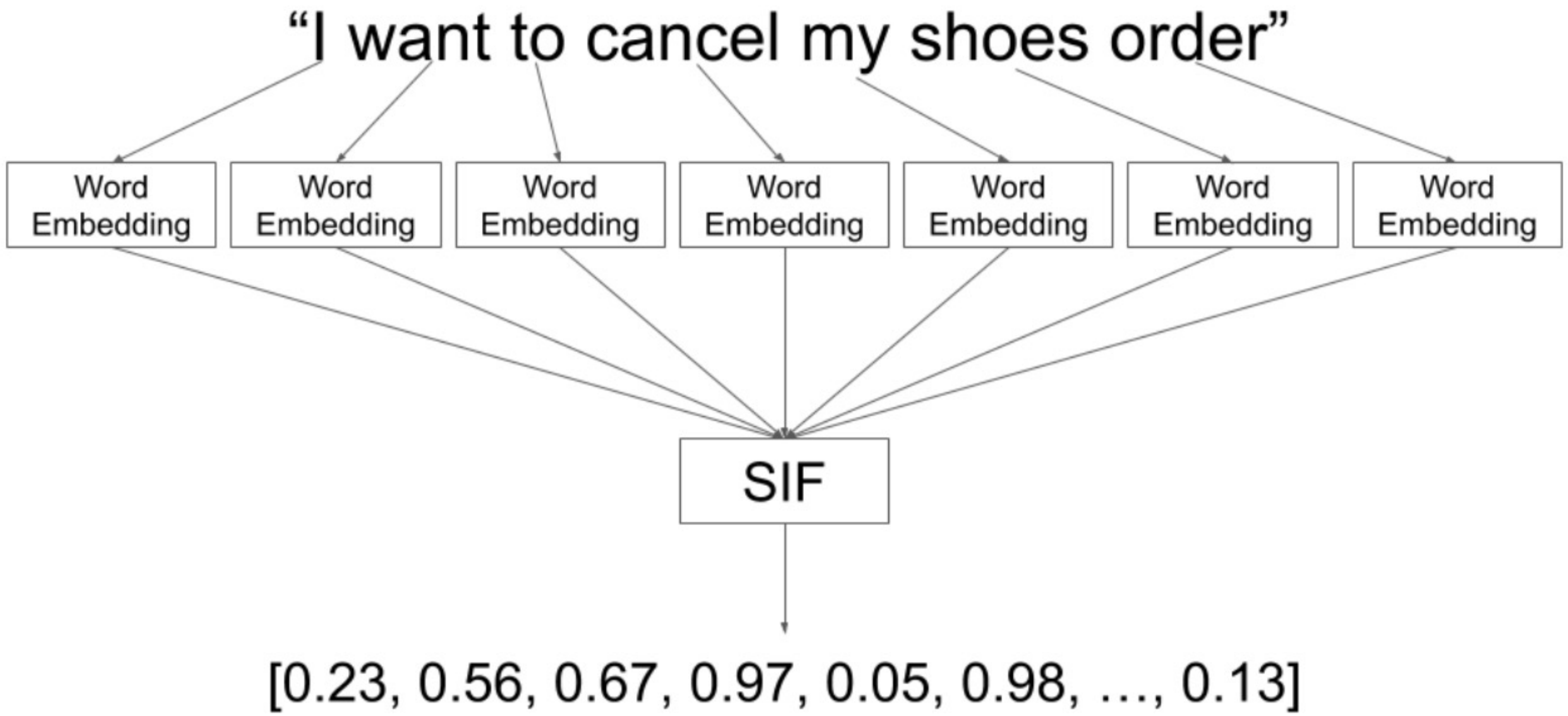
chatGPT-4 IQ?

Word Embedding



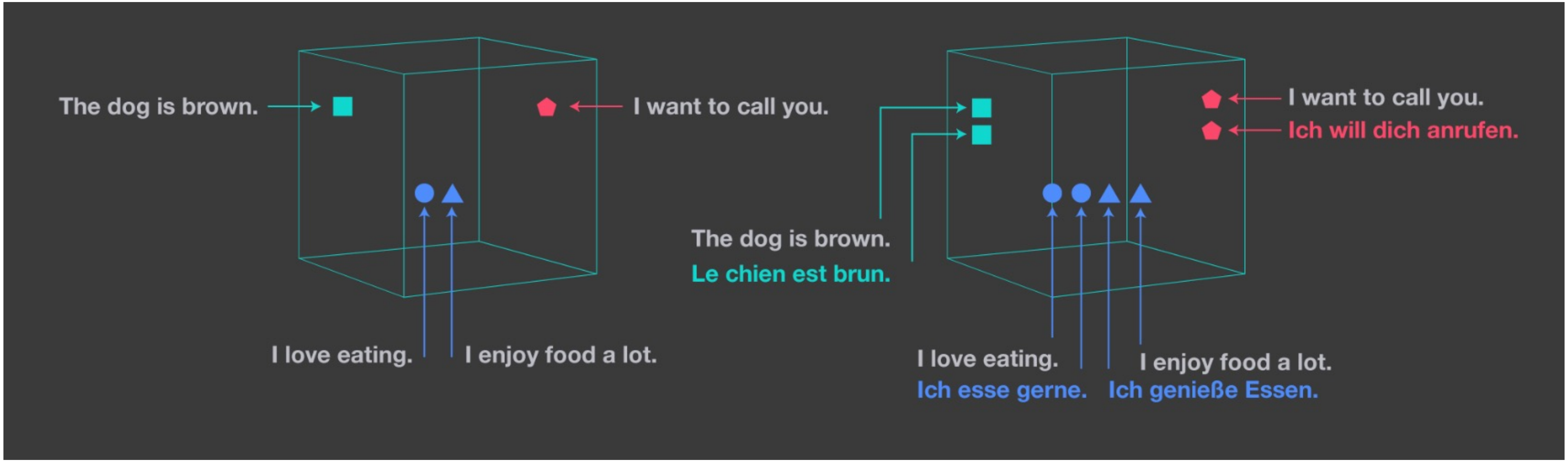
https://airbyte.com/data-engineering-resources/sentence-word-embeddings

Sentence Embedding

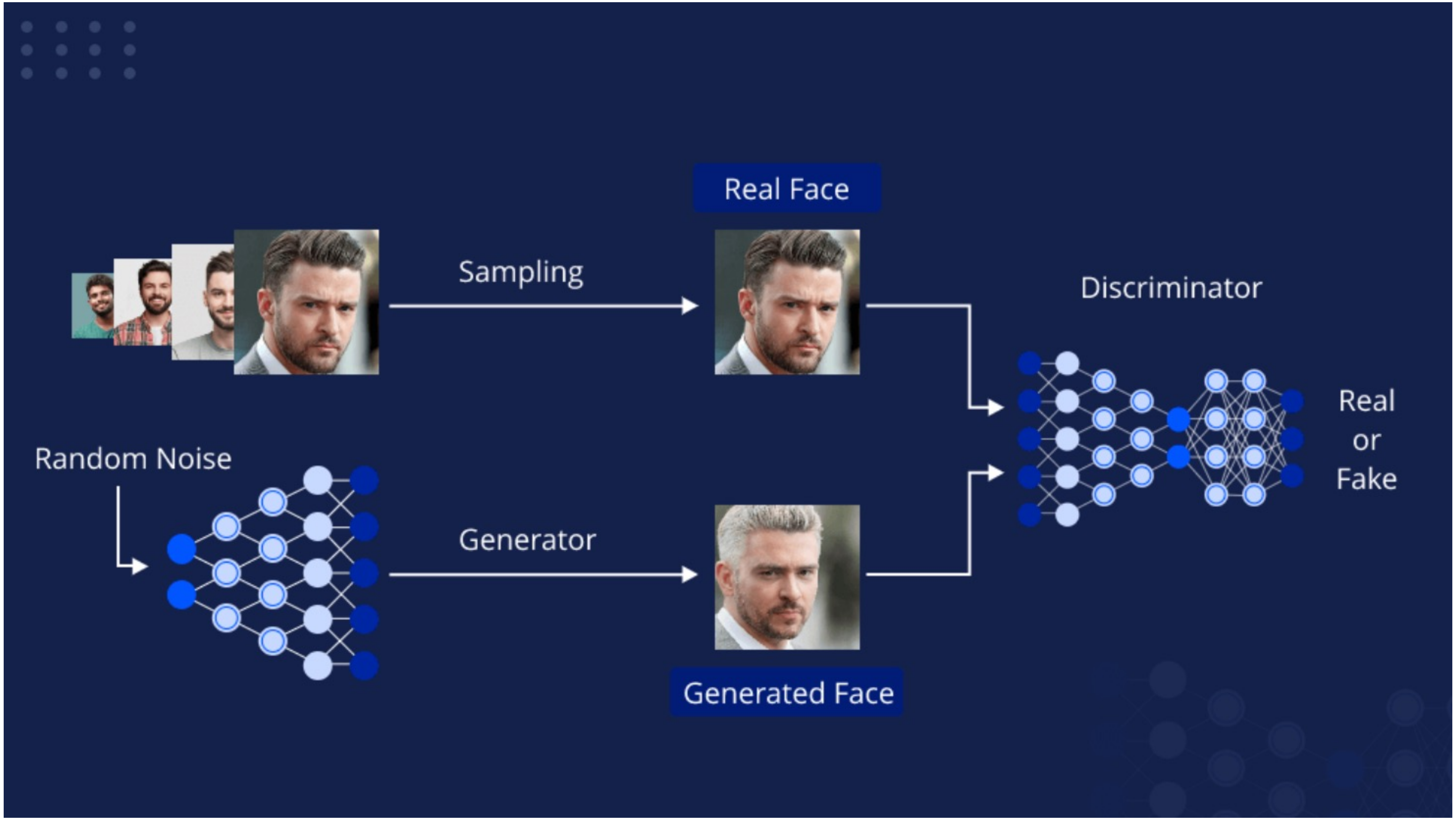


https://airbyte.com/data-engineering-resources/sentence-word-embeddings

Universal sentence embeddings



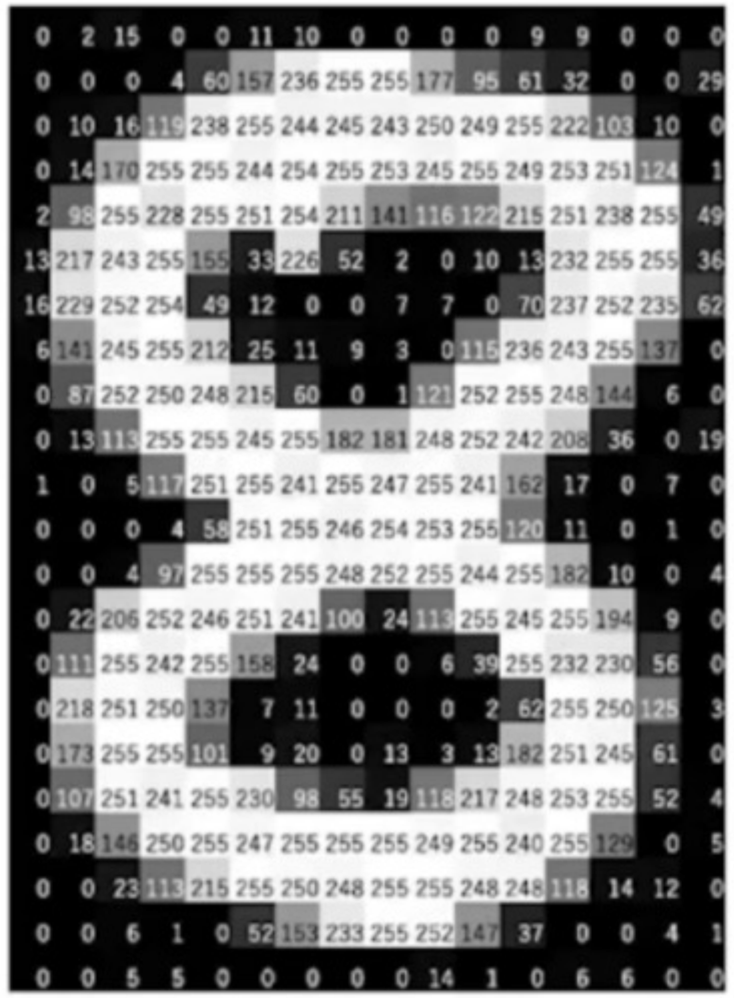
https://engineering.fb.com/2019/01/22/ai-research/laser-multilingual-sentence-embeddings/



https://www.leewayhertz.com/a-guide-on-generative-ai-models-for-image-synthesis/

Iamge Data

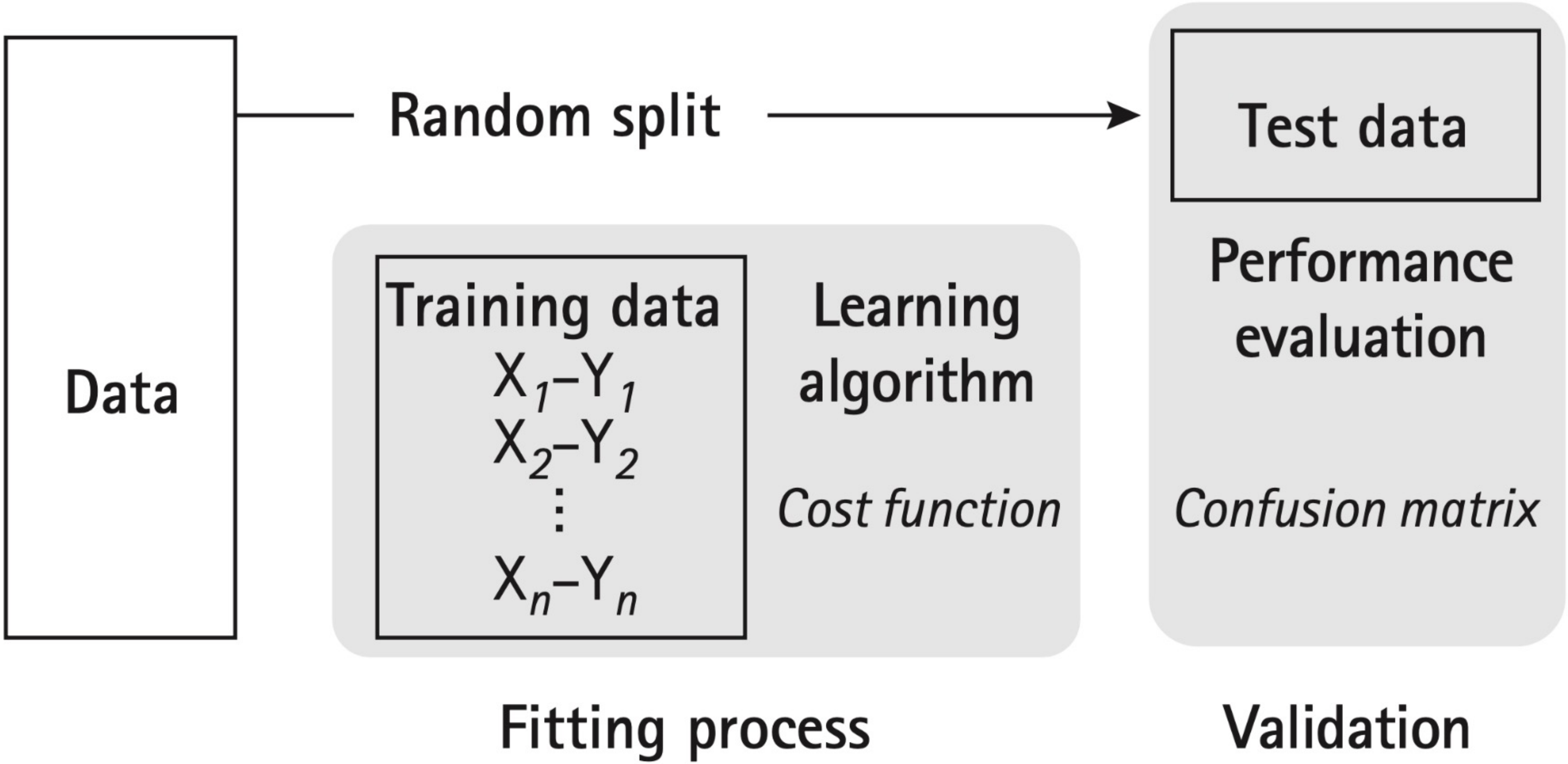
24 x 16



grayscale or b&w image
we have pixel values
(intensity)
ranging from 0 to 255.

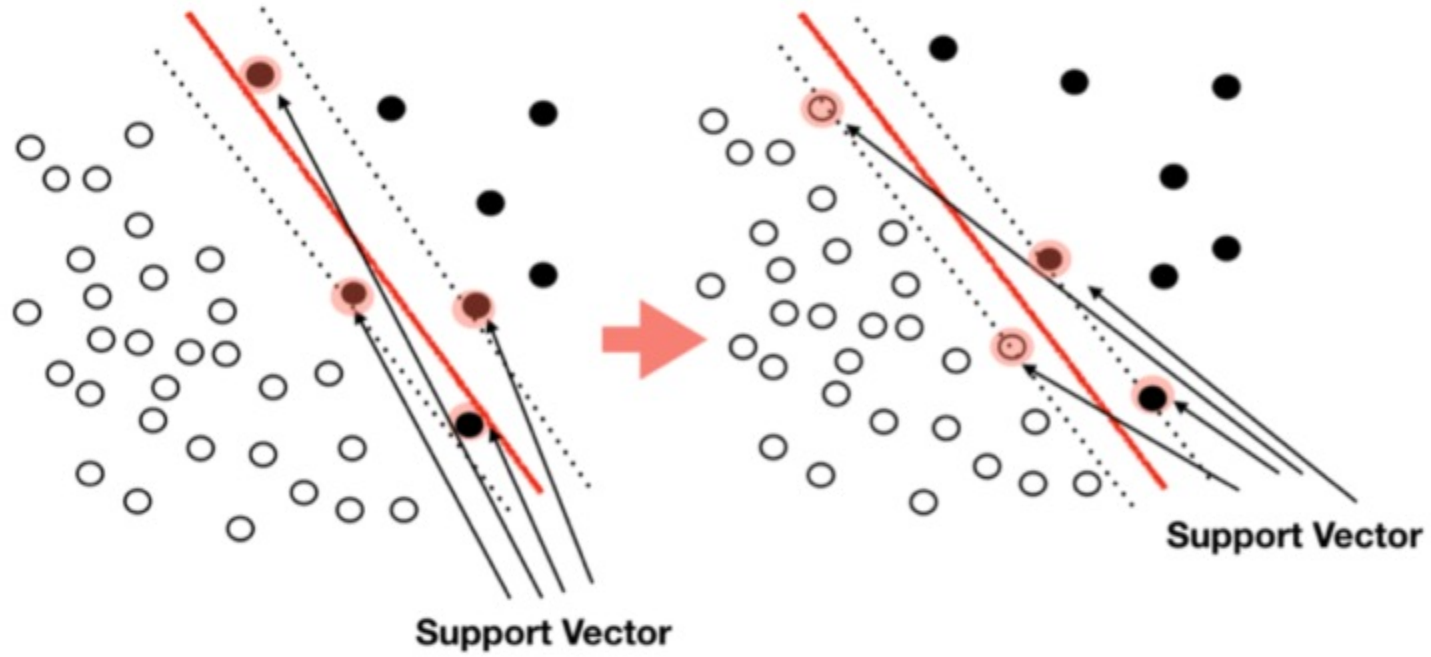
Colour Image

What is Machine Learning?


Data
Random split
Test data
Training data
X1–Y1
X2–Y2
⋮
Xn–Yn
Learning algorithm
Cost function
Performance evaluation
Confusion matrix
Fitting process
Validation

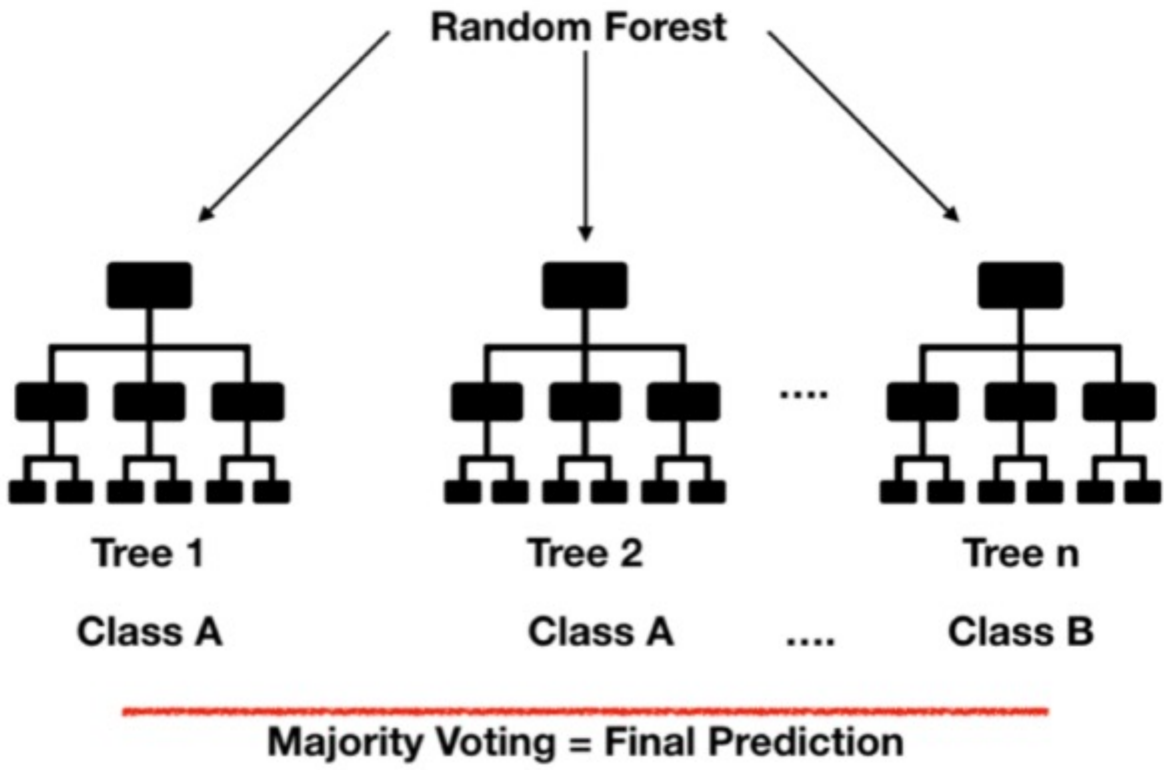
Regression Model

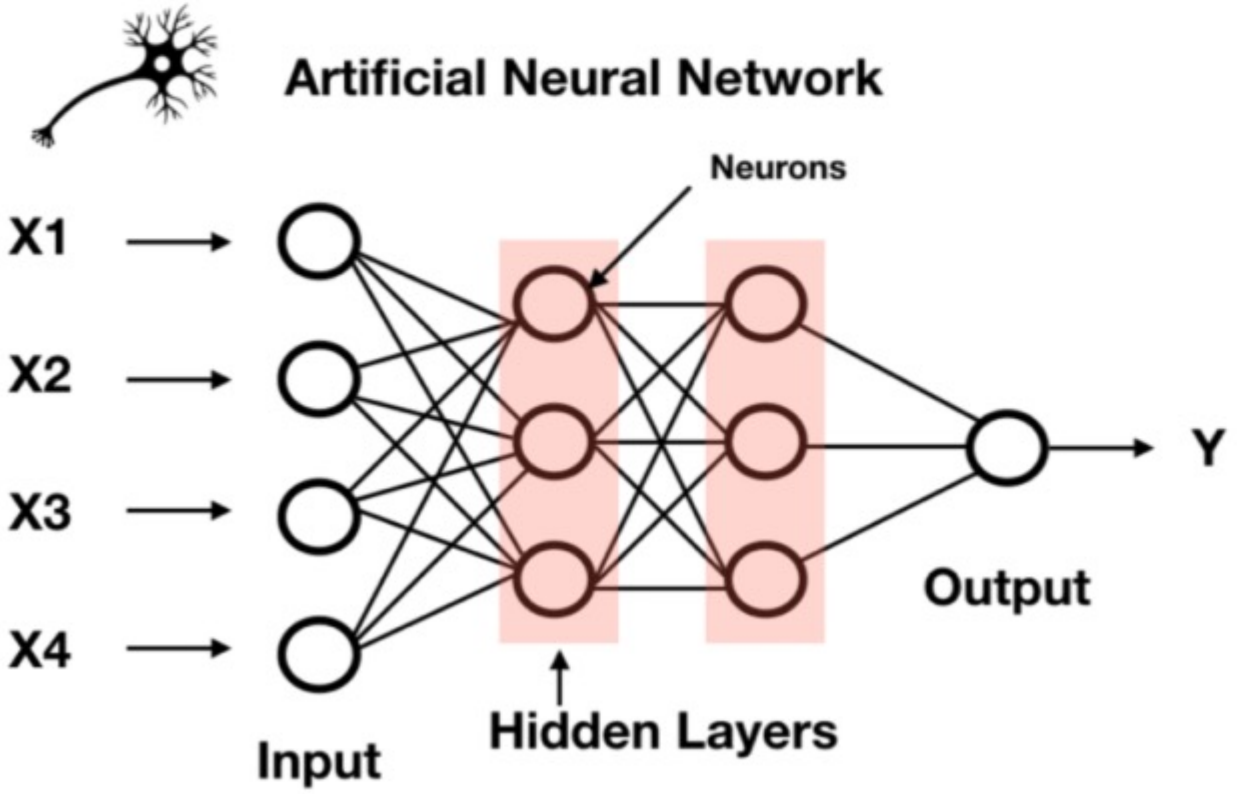
Support Vector Machine



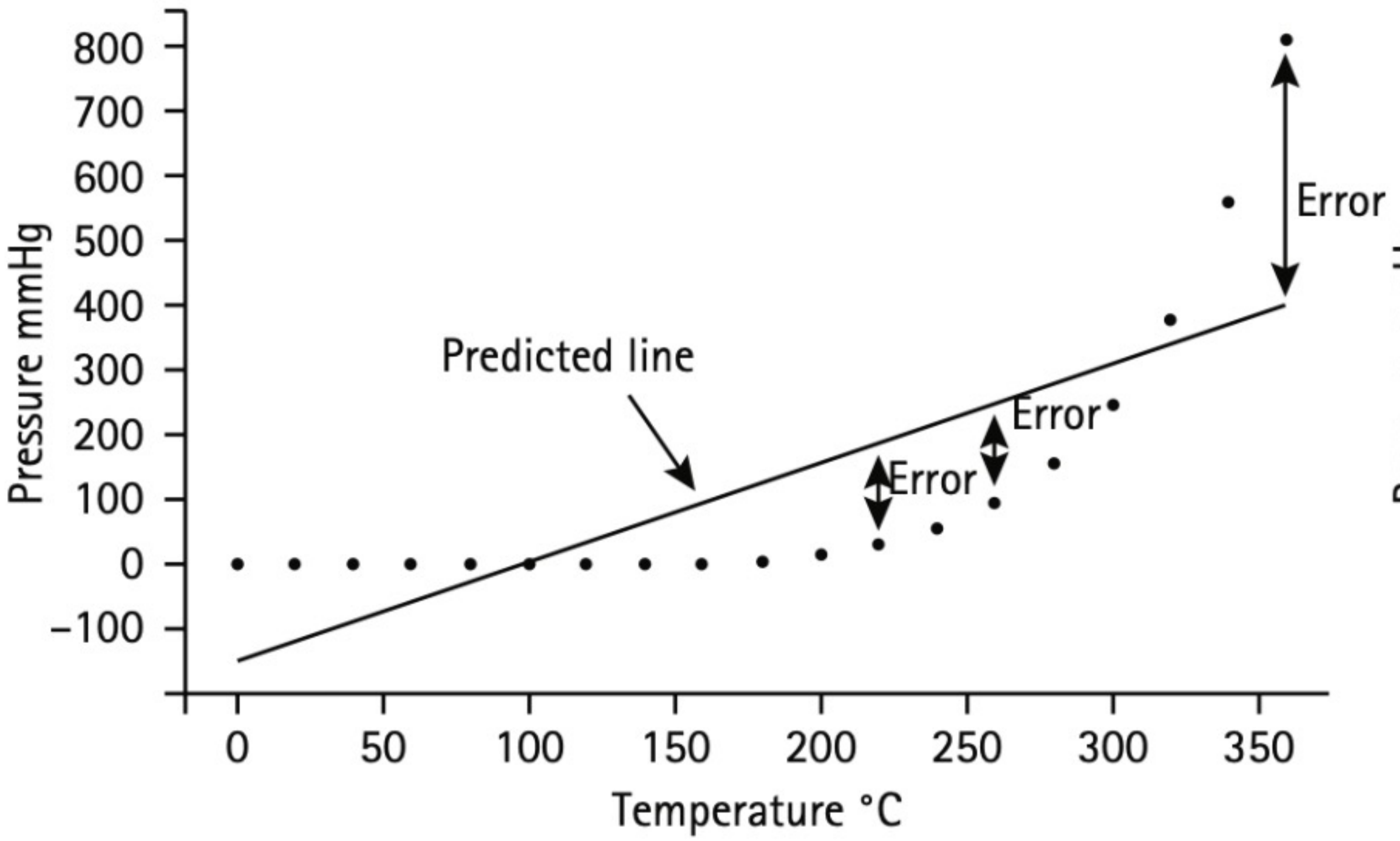
Random Forest Model

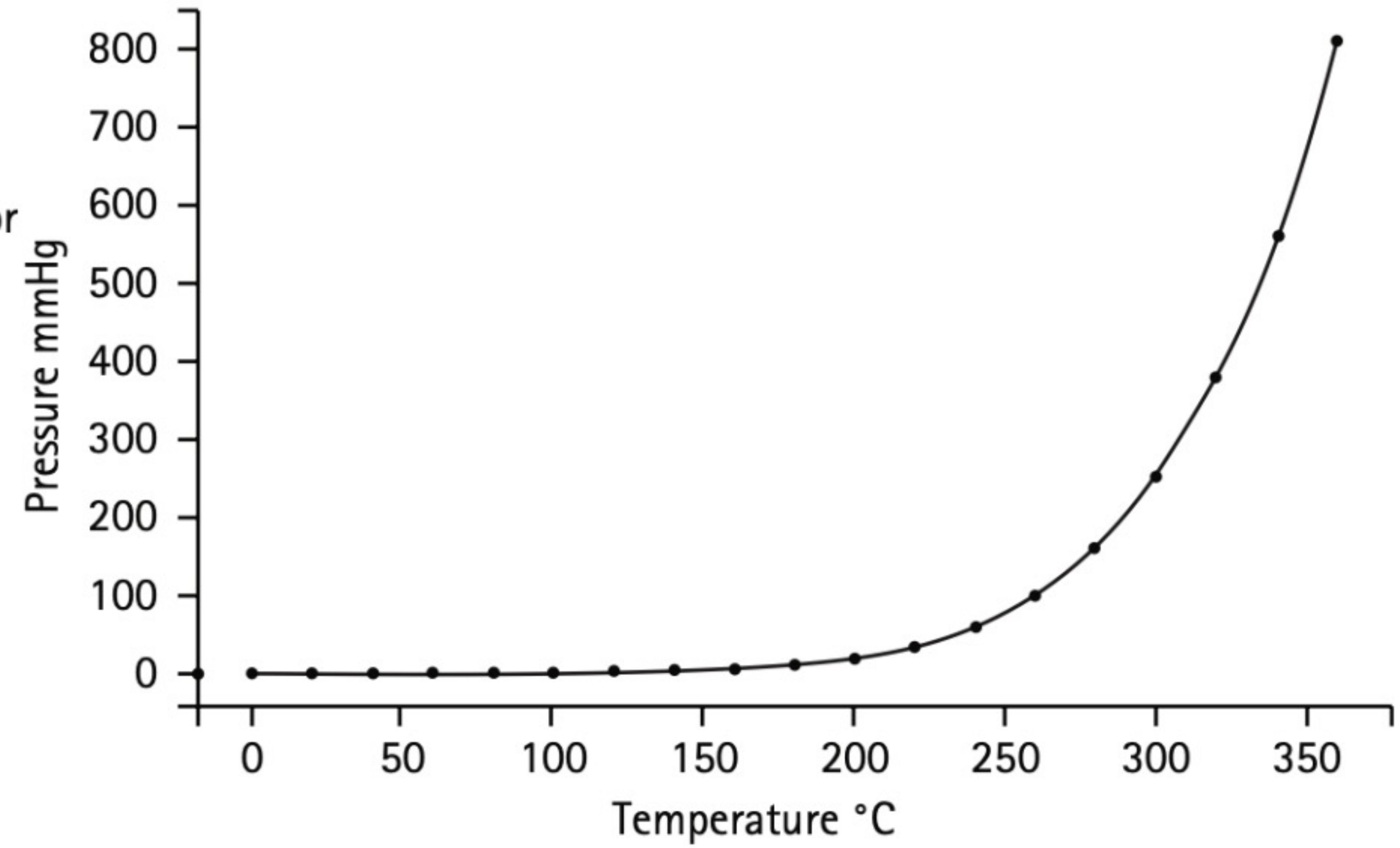
Instance



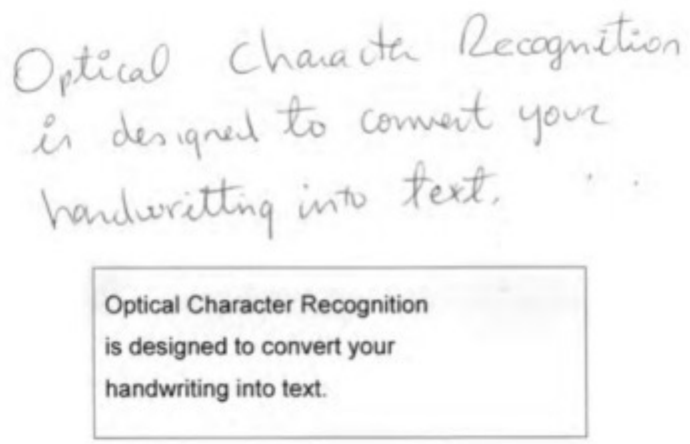


Fitting Process





- IBM's Deep Blue, which beat chess grand master Garry Kasparov in 1997 (Rule-based, a machine that was capable of imagining an average of 200,000,000 positions per second)
- Google Deep Mind's AlphaGo, which beat Lee Sedol at Go in 2016 (ML based, by training itself on a large data set of expert moves)



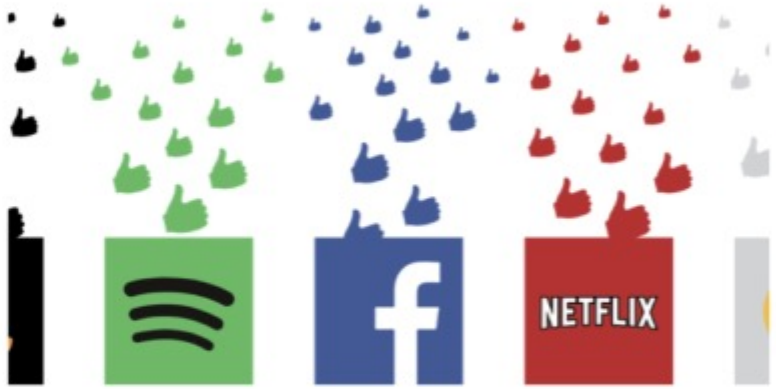
Optical Character Recognition

Information Filtering

Facial Recognition

Automatic Speech Recognition

Customer Clustering



Recommendation System

Automated Driving System

Machine Learning Application

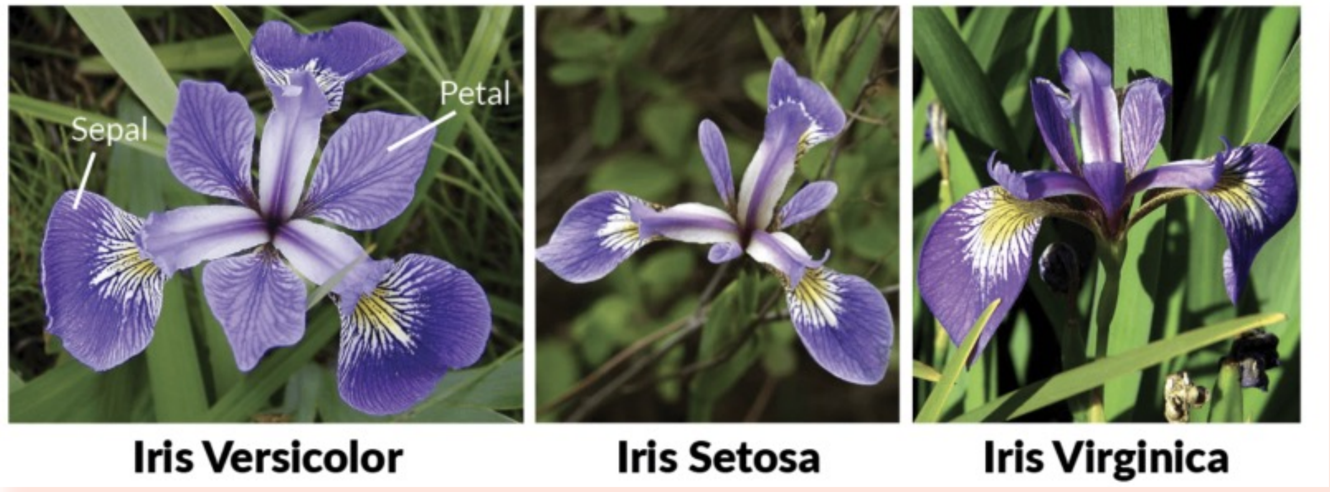
Health Intelligence

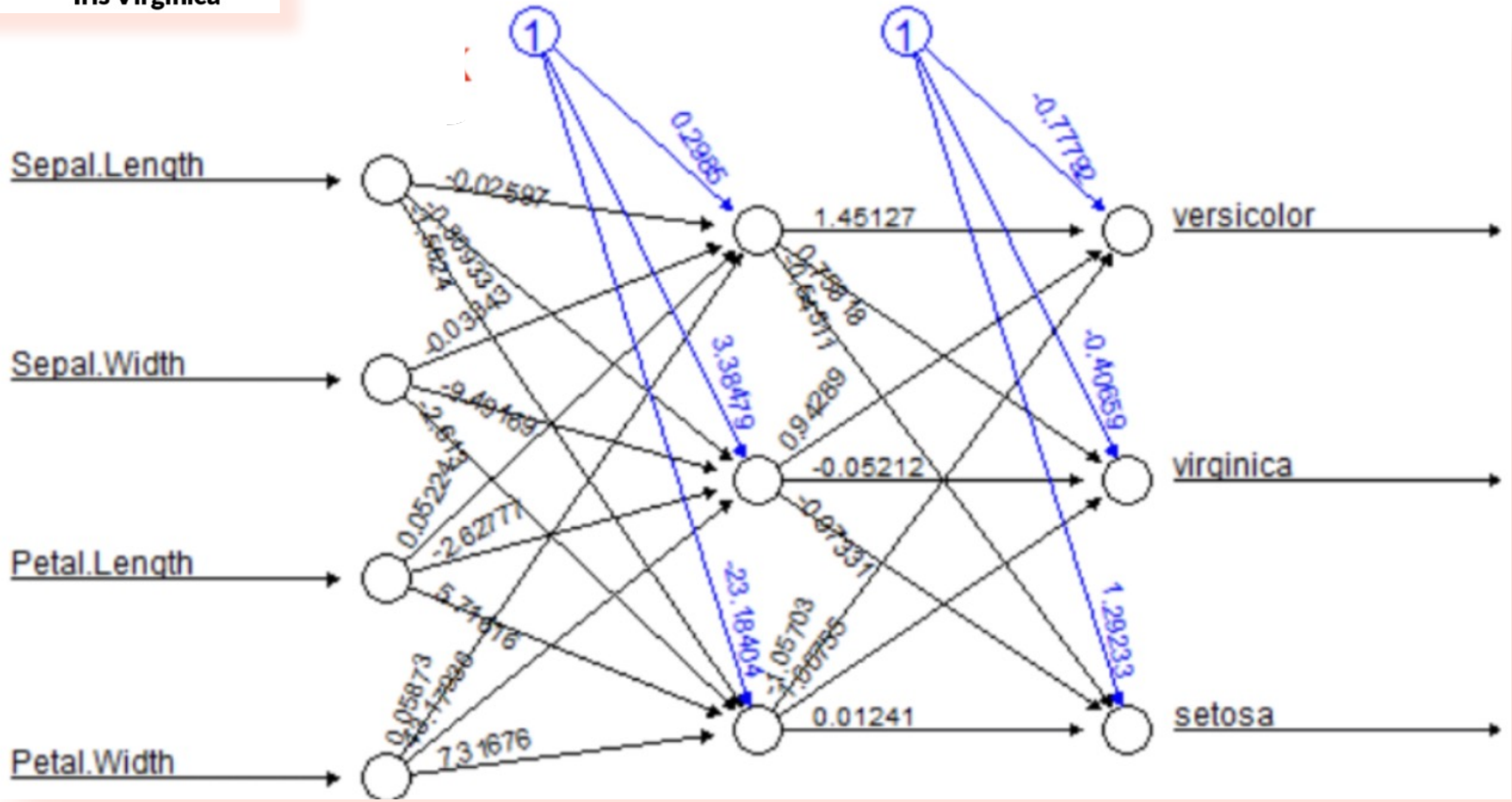
From Programming to Learning

Unboxing the Black Box

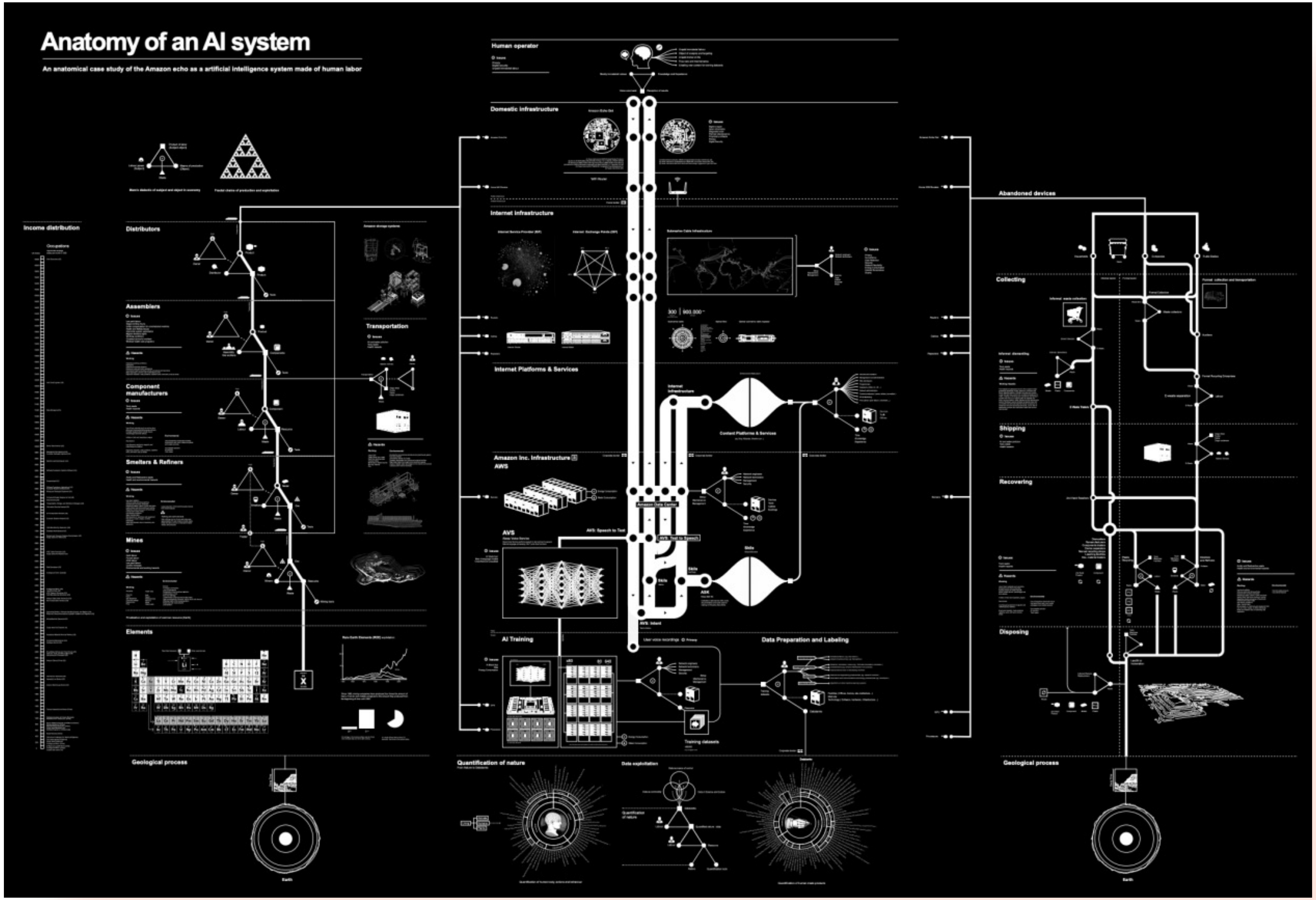
Iris Versicolor Iris Setosa Iris Virginica

Günther, Frauke, and Stefan Fritsch. "Neuralnet: training of neural networks." R J. 2.1 (2010): 30.





Günther, Frauke, and Stefan Fritsch. "Neuralnet: training of neural networks." R J. 2.1 (2010): 30.



“What does it mean to sociologists?”

Project Credits:
Maps and design: Vladan Joler and Kate Crawford
Published by: SHARE Lab, SHARE Foundation (https://labs.rs) and The AI Now Institute, NYU (https://ainowinstitute.org/)

npj | Digital Medicine

www.nature.com/npjdigitalmed

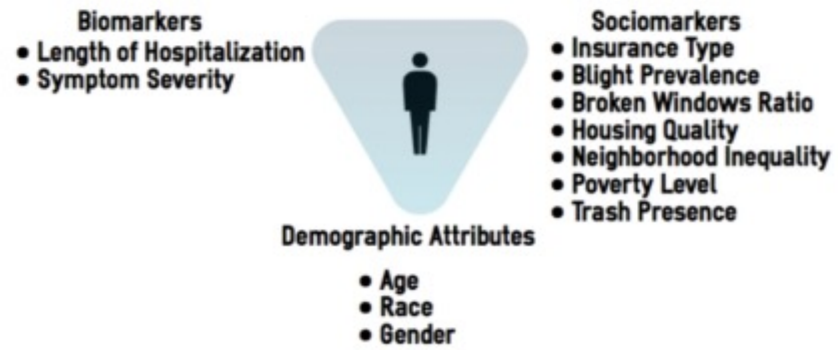
ARTICLE OPEN

Sociomarkers and biomarkers: predictive modeling in identifying pediatric asthma patients at risk of hospital revisits

Eun Kyong Shin[1], Ruhi Mahajan[1], Oguz Akbilgic[1,2] and Arash Shaban-Nejad[1]

The importance of social components of health has been emphasized both in epidemiology and public health. This paper highlights the significant impact of social components on health outcomes in a novel way. Introducing the concept of sociomarkers, which are measurable indicators of social conditions in which a patient is embedded, we employed a machine learning approach that uses both biomarkers and sociomarkers to identify asthma patients at risk of a hospital revisit after an initial visit with an accuracy of 66%. The analysis has been performed over an integrated dataset consisting of individual-level patient information such as gender, race, insurance type, and age, along with ZIP code-level sociomarkers such as poverty level, blight prevalence, and housing quality. Using this uniquely integrated database, we then compare the traditional biomarker-based risk model and the sociomarker-based risk model. A biomarker-based predictive model yields an accuracy of 65% and the sociomarker-based model predicts with an accuracy of 61%. Without knowing specific symptom-related features, the sociomarker-based model can correctly predict two out of three patients at risk. We systematically show that sociomarkers play an important role in predicting health outcomes at the individual level in pediatric asthma cases. Additionally, by merging multiple data sources with detailed neighborhood-level data, we directly measure the importance of residential conditions for predicting individual health outcomes.

npj Digital Medicine (2018)1:50 ; doi:10.1038/s41746-018-0056-y



Data: **Medical data** (Pediatric asthma encounter records collected from Jan 1st 2016 to Dec 31st 2016, 3,678 cases of the first time visit) + **Urban Housing Quality data** (Property Hub) + **US Census** (2010)

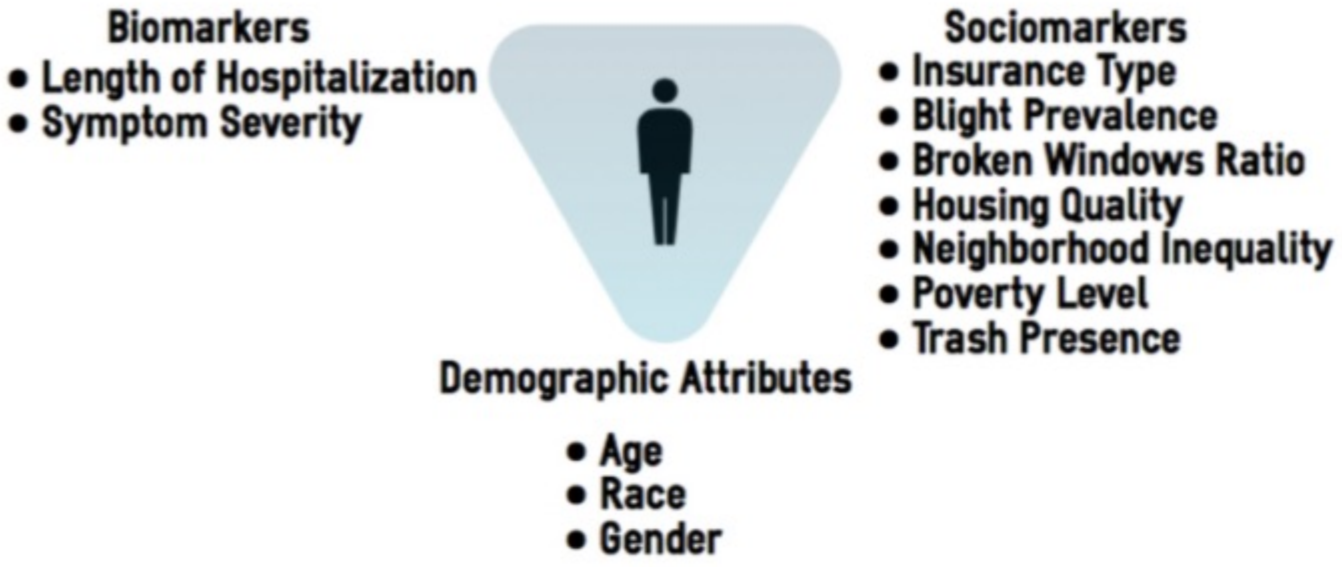
Random Forest Classifier and SVM Classifier

		Validation Set			Training Set		
		Accu	Sepc	Sens	Accu	Spec	Sens
RF	Model 1	66.05	67.63	64.82	66.11	67.67	64.82
	Model 2	65.39	67.11	64.07	65.48	67.12	64.14
	Model 3	61.17	62.59	60.11	61.28	62.70	60.16
SVM	Model 1	62.10	62.00	62.32	62.21	62.10	62.35
	Model 2	59.58	59.89	59.41	59.70	59.96	59.48
	Model 3	57.83	59.07	56.98	57.97	59.17	57.08

SOCIOMARKERS
Precision Medicine

Shin, Eun Kyong, Ruhi Mahajan, Oguz Akbilgic, and Arash Shaban-Nejad. "Sociomarkers and biomarkers: predictive modeling in identifying pediatric asthma patients at risk of hospital revisits." NPJ digital medicine 1, no. 1 (2018): 50.

SOCIOMARKERS

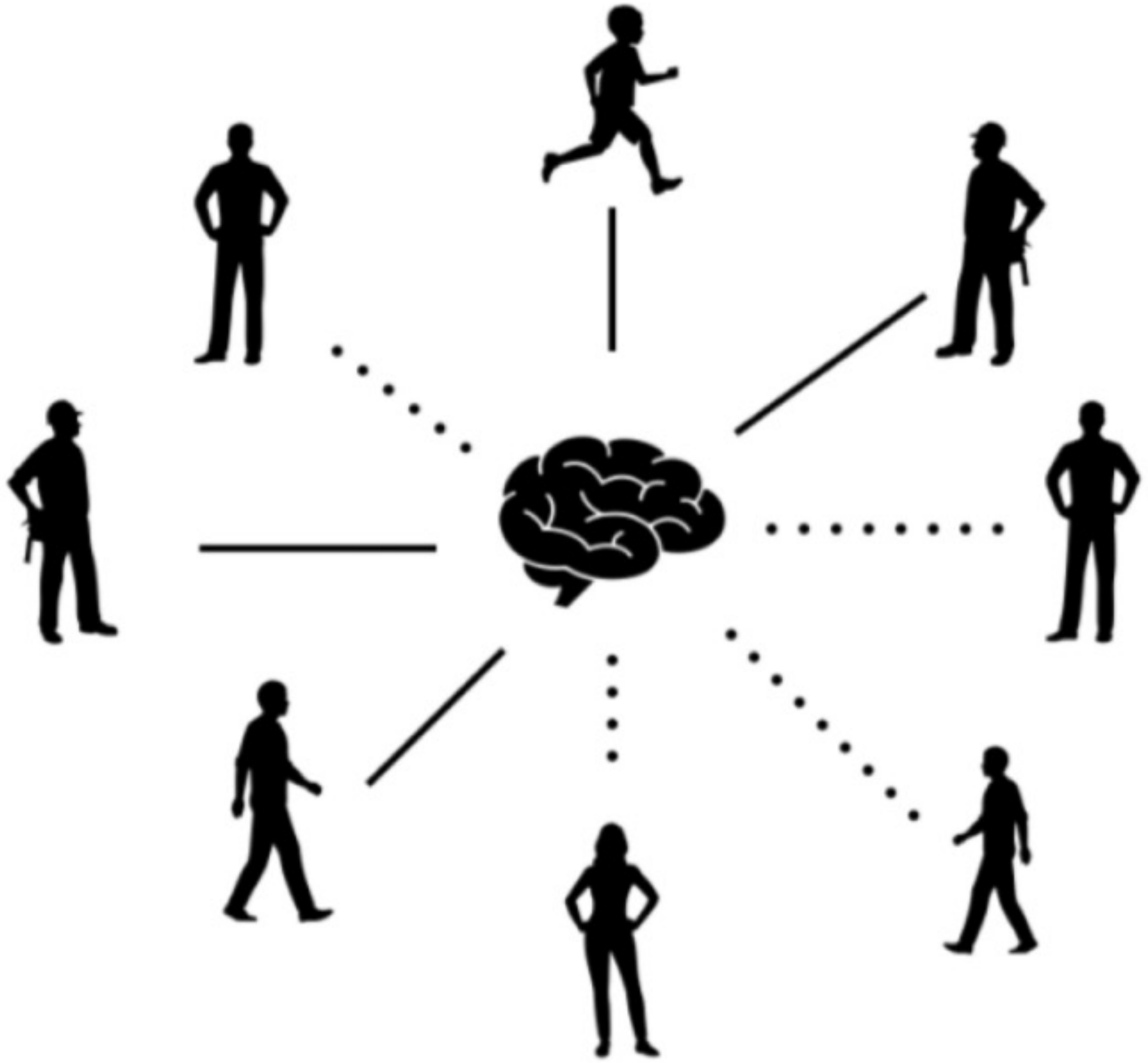


Data: **Medical data** (Pediatric asthma encounter records collected from Jan 1st 2016 to Dec 31st 2016, 3,678 cases of the first time visit) + **Urban Housing Quality data** (Property Hub) + **US Census** (2010)

Random Forest Classifier and SVM Classifier

		Validation Set			Training Set		
		Accu	Sepc	Sens	Accu	Spec	Sens
RF	Model 1	66.05	67.63	64.82	66.11	67.67	64.82
	Model 2	65.39	67.11	64.07	65.48	67.12	64.14
	Model 3	61.17	62.59	60.11	61.28	62.70	60.16
SVM	Model 1	62.10	62.00	62.32	62.21	62.10	62.35
	Model 2	59.58	59.89	59.41	59.70	59.96	59.48
	Model 3	57.83	59.07	56.98	57.97	59.17	57.08

DNA (geneticcally)
NEURON (empirically)
SOCIETY



NEUROSCIENCE

A dedicated network for social interaction processing in the primate brain

J. Sliwa* and W. A. Freiwald*

Primate cognition requires interaction processing. Interactions can reveal otherwise hidden properties of intentional agents, such as thoughts and feelings, and of inanimate objects, such as mass and material. Where and how interaction analyses are implemented in the brain is unknown. Using whole-brain functional magnetic resonance imaging in macaque monkeys, we discovered a network centered in the medial and ventrolateral prefrontal cortex that is exclusively engaged in social interaction analysis. Exclusivity of specialization was found for no other function anywhere in the brain. Two additional networks, a parieto-premotor and a temporal one, exhibited both social and physical interaction preference, which, in the temporal lobe, mapped onto a fine-grain pattern of object, body, and face selectivity. Extent and location of a dedicated system for social interaction analysis suggest that this function is an evolutionary forerunner of human mind-reading capabilities.

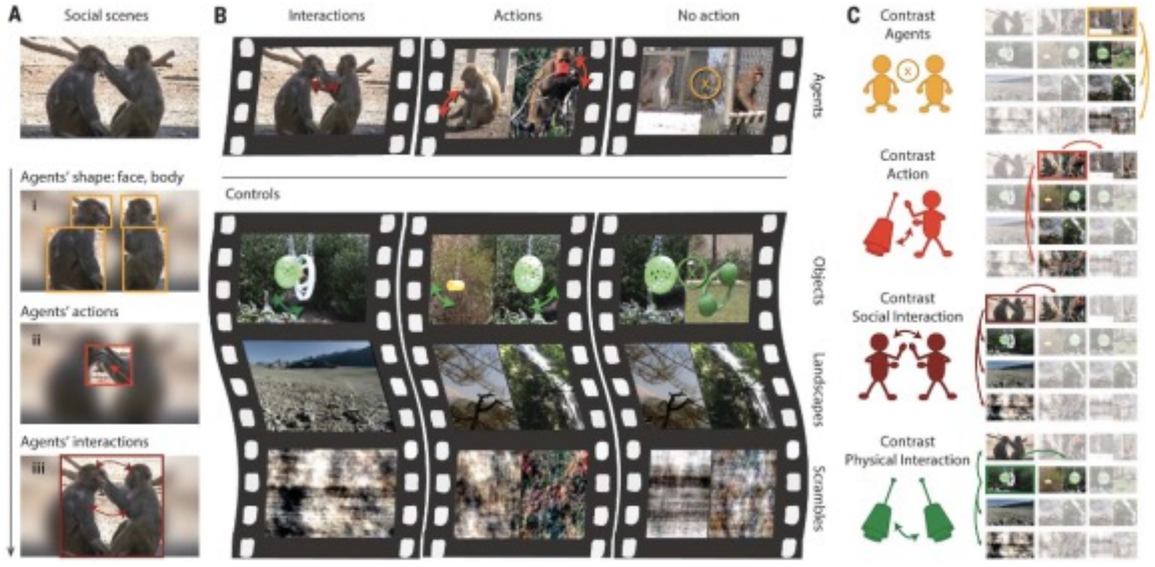
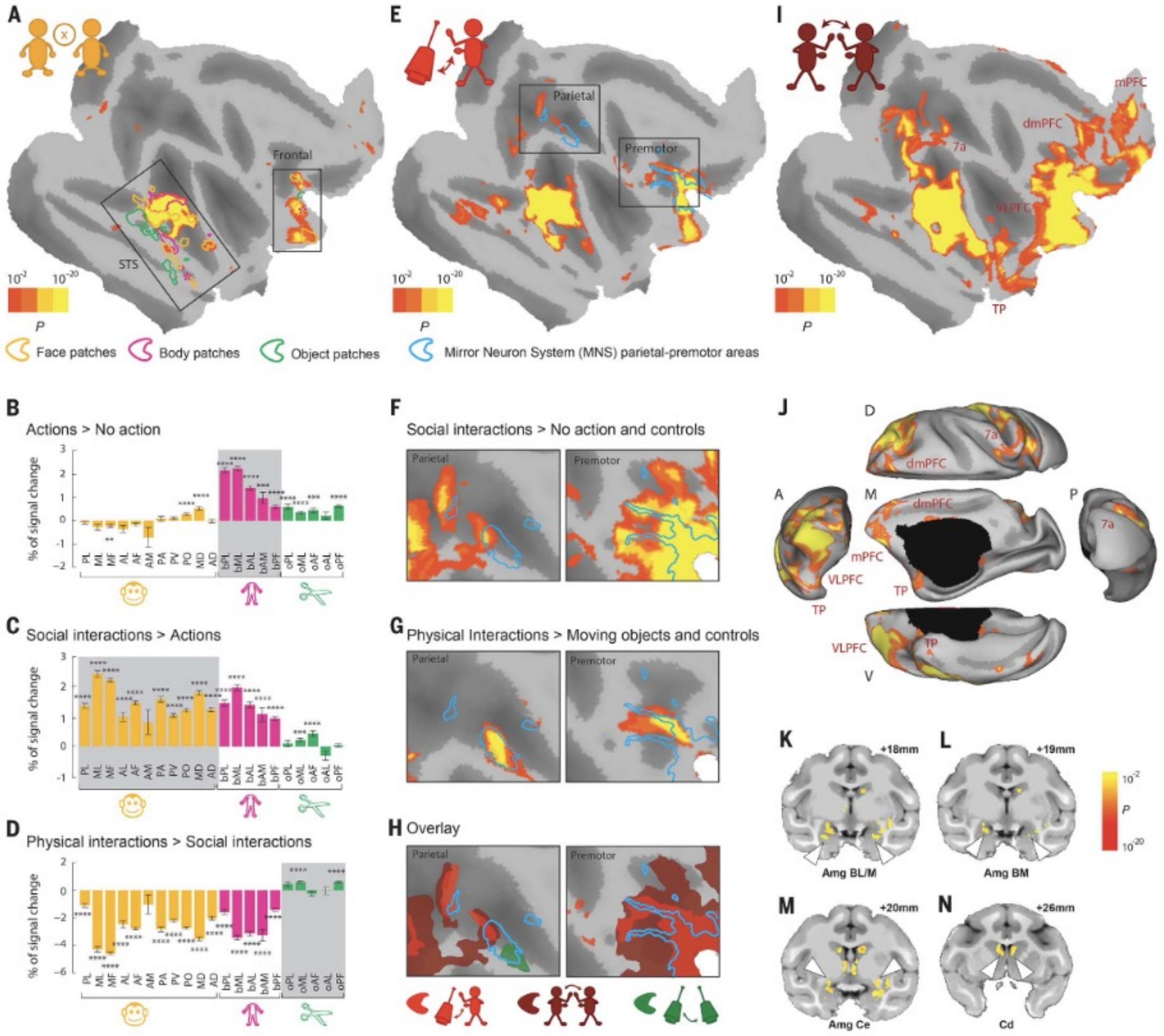
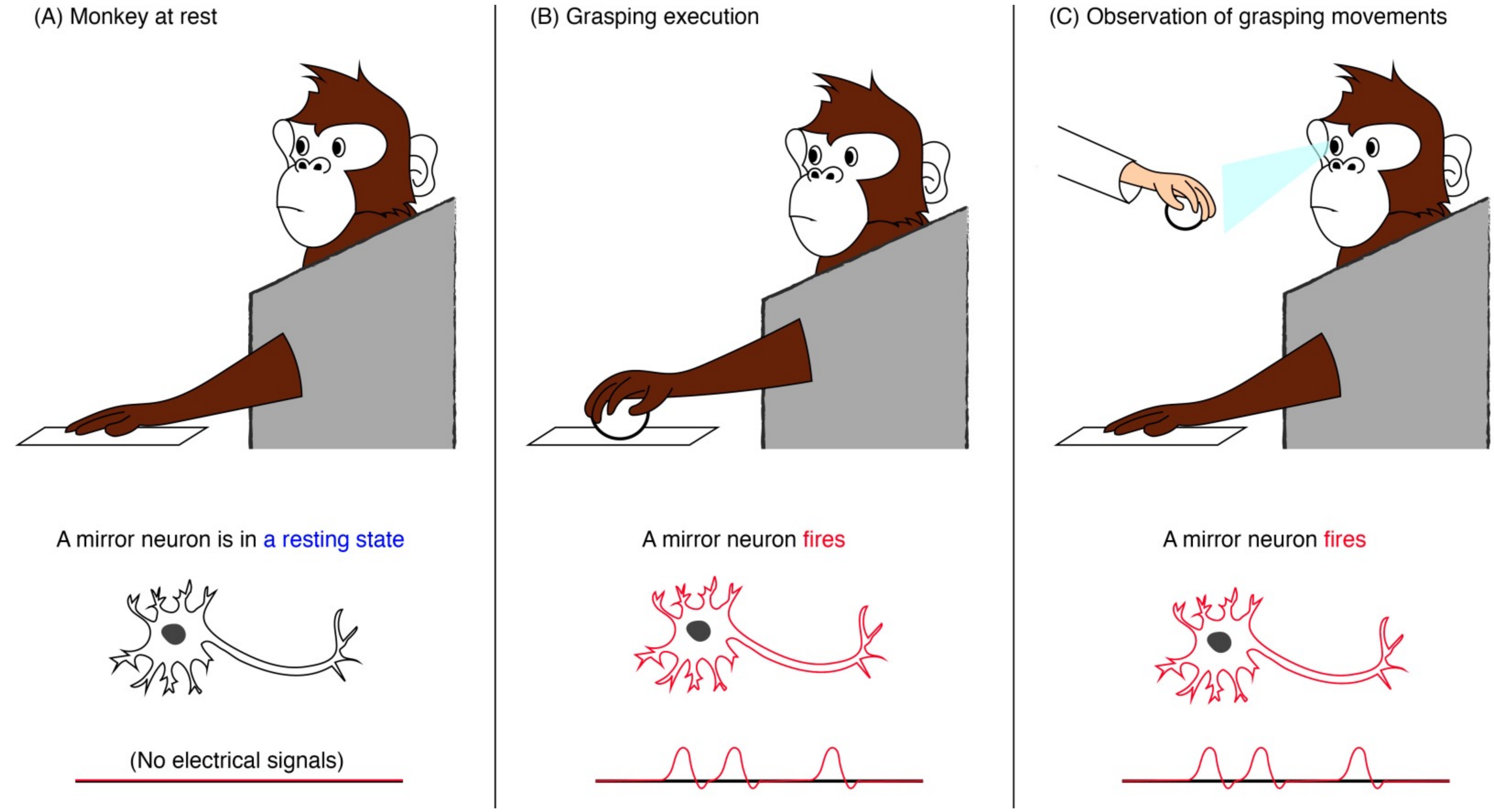


Fig. 1. Task design and hypothesis. (**A**) Hypothesized cognitive steps occurring when watching social interactions. (i) Processing agents' shape, (ii) processing action, (iii) processing interaction. (**B**) Classes of videos used for stimuli and controls that monkeys could freely watch during the experiment. (**C**) Schematics of the contrasts (arrows between nonblurred images) used in conjunction to identify brain activity related to four main conditions: agents, action, social interaction, and physical interaction. Pictures are presented in the same order as in (B).



Sliwa, J. and Freiwald, W.A., 2017. A dedicated network for social interaction processing in the primate brain. Science, 356(6339), pp.745-749.


(A) Monkey at rest
(B) Grasping execution
(C) Observation of grasping movements
A mirror neuron is in a resting state
(No electrical signals)
A mirror neuron fires
A mirror neuron fires

JAMA Network Open

Original Investigation | Pediatrics

Association of Maternal Social Relationships With Cognitive Development in Early Childhood

Eun Kyong Shin, PhD; Kaja LeWinn, ScD; Nicole Bush, PhD; Frances A. Tylavsky, DrPH; Robert Lowell Davis, MD, MPH; Arash Shaban-Nejad, PhD, MPH

Abstract

IMPORTANCE This study examines how different types of social network structures are associated with early cognitive development in children.

OBJECTIVES To assess how social relationships and structures are associated with early cognitive development and to elucidate whether variations in the mother's social networks alter a child's early cognitive development patterns.

DESIGN, SETTING, AND PARTICIPANTS This cohort study used data from 1082 mother-child pairs in the University of Tennessee Health Science Center–Conditions Affecting Neurocognitive Development and Learning and Early Childhood project to examine the association between networks of different levels of complexity (triad, family, and neighborhood) and child cognitive performance after adjustment for the mother's IQ, birth weight, and age, and the father's educational level. The final model was adjusted for the household poverty level. Data were collected from December 2006 through January 2014 and analyzed from October through November 2018.

EXPOSURES The child-mother relationship, child-mother-father triad, family setting, child's dwelling network, mother's social support network, and neighborhood networks.

MAIN OUTCOMES AND MEASURES Measure of cognitive development of the child using Bayley Scales of Infant Development (BSID) at 2 years of age.

Key Points

Question How are social relationships and structures, such as dyads, families, and neighborhoods, associated with early cognitive development in children?

Findings In this cohort study of 1082 mother-child pairs, the mother's social networks were significantly positively associated with early childhood cognitive development. Being in a large family network was significantly associated with lower cognitive performance.

Meaning The findings suggest that maternal social relationships are associated with cognitive development in children and that social relationships beyond the mother-child-father triad are significantly associated with cognitive development.

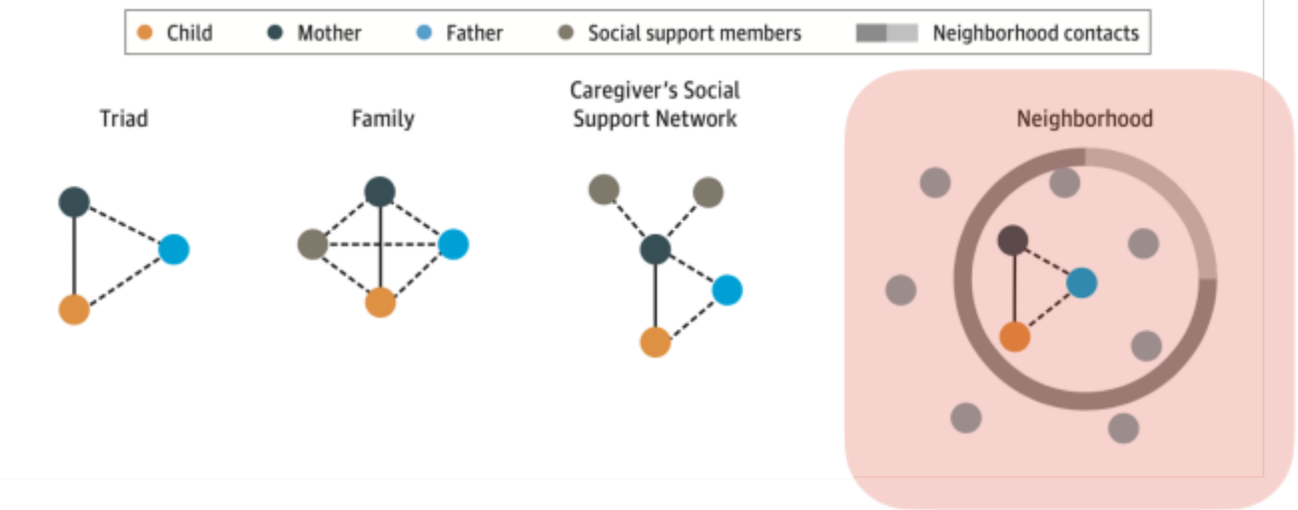
Table 1. Characteristics of 1082 Study Participants in CANDLE

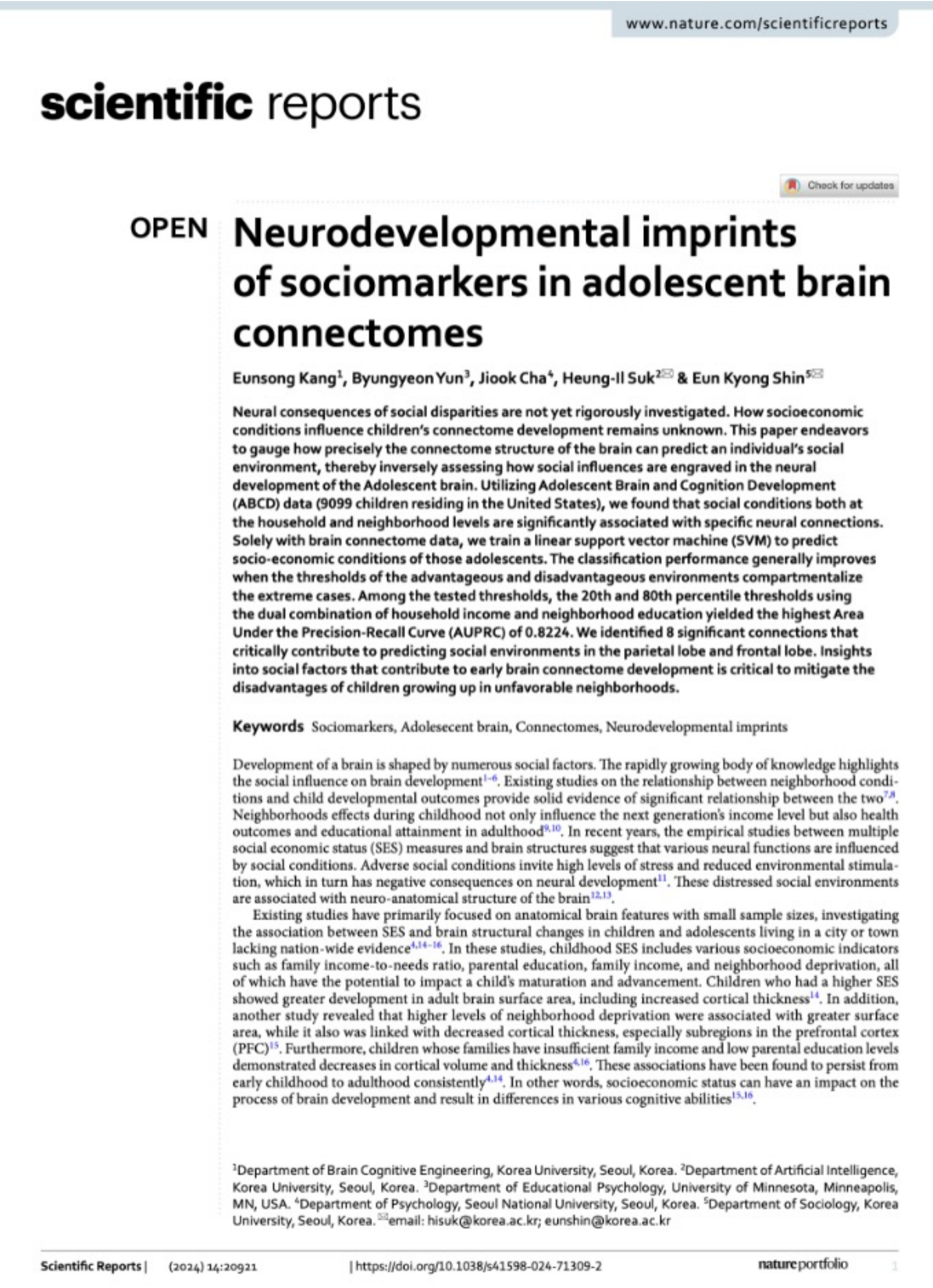
Variable	Value[a]	BSID Score, Mean (SD)
BSID categories	1082 (100)	97.74 (12.87)
Sex		
Male		
Female		
Race/ethnicity		
African American		
White		
Poverty level		
Below FPL		
At or above FPL		
Father's educational level		
Some high school or graduated high		
Some college		
College and graduate school		
Father's cohabitation		
Father not living at home	421 (42.0)	94.53 (11.38)
Father living at home	570 (58.0)	100.61 (13.47)
Family size		
<6 People	813 (75.1)	99.19 (13.23)
≥6 People	269 (24.9)	93.36 (10.63)
Mother's social network size, mean (SD)	3.49 (1.82)	NA
Neighborhood embeddedness		
Not knowing many people in the neighborhood	408 (41.2)	95.47 (11.46)
Knowing many people in the neighborhood	582 (58.8)	99.92 (13.65)
Mother's WASI IQ, mean (SD)	96.11 (16.25)	NA
Birth weight, mean (SD), g	3266.04 (537.10)	NA
Mother's age at child's birth, mean (SD), y	26.55 (5.50)	NA

Table 2. Stepwise Network Associations in Early Cognitive Development at 2 Years of Age

	Model 1[a]		Model 2[b]	
Characteristic	Coefficient Value (95% CI)[c]	P Value	Coefficient Value (95% CI)[c]	P Value
Father's cohabitation	0.40 (−1.22 to 2.01)	NA	0.07 (−1.58 to 1.73)	NA
Large family	−2.41 (−4.21 to −0.61)	.01	−2.21 (−4.02 to −0.40)	.01
Mother's social network	0.43 (0.03 to 0.83)	.04	0.40 (0.001 to 0.80)	.05
Neighborhood	1.44 (0.01 to 2.88)	.05	1.39 (−0.04 to 2.83)	.06
Below poverty level	NA	NA	−1.59 (−3.33 to 0.15)	NA

Figure. Stepwise Network Exposure Conditions





scientific reports

Check for updates

OPEN

Neurodevelopmental imprints of sociomarkers in adolescent brain connectomes

Eunsong Kang[1], Byungyeon Yun[3], Jiook Cha[4], Heung-Il Suk[2]✉ & Eun Kyong Shin[5]✉

Neural consequences of social disparities are not yet rigorously investigated. How socioeconomic conditions influence children's connectome development remains unknown. This paper endeavors to gauge how precisely the connectome structure of the brain can predict an individual's social environment, thereby inversely assessing how social influences are engraved in the neural development of the Adolescent brain. Utilizing Adolescent Brain and Cognition Development (ABCD) data (9099 children residing in the United States), we found that social conditions both at the household and neighborhood levels are significantly associated with specific neural connections. Solely with brain connectome data, we train a linear support vector machine (SVM) to predict socio-economic conditions of those adolescents. The classification performance generally improves when the thresholds of the advantageous and disadvantageous environments compartmentalize the extreme cases. Among the tested thresholds, the 20th and 80th percentile thresholds using the dual combination of household income and neighborhood education yielded the highest Area Under the Precision-Recall Curve (AUPRC) of 0.8224. We identified 8 significant connections that critically contribute to predicting social environments in the parietal lobe and frontal lobe. Insights into social factors that contribute to early brain connectome development is critical to mitigate the disadvantages of children growing up in unfavorable neighborhoods.

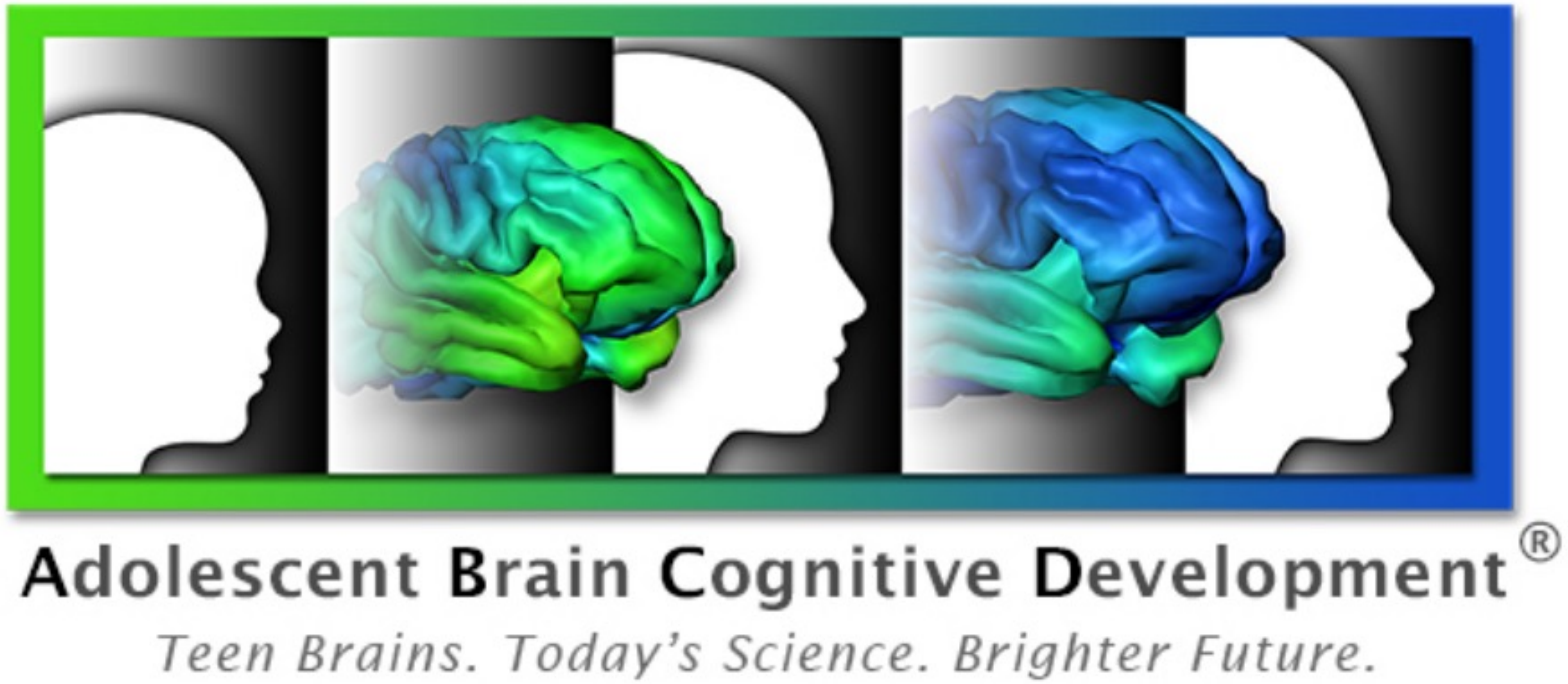
Keywords Sociomarkers, Adolesecent brain, Connectomes, Neurodevelopmental imprints

Development of a brain is shaped by numerous social factors. The rapidly growing body of knowledge highlights the social influence on brain development[1–6]. Existing studies on the relationship between neighborhood conditions and child developmental outcomes provide solid evidence of significant relationship between the two[7,8]. Neighborhoods effects during childhood not only influence the next generation's income level but also health outcomes and educational attainment in adulthood[9,10]. In recent years, the empirical studies between multiple social economic status (SES) measures and brain structures suggest that various neural functions are influenced by social conditions. Adverse social conditions invite high levels of stress and reduced environmental stimulation, which in turn has negative consequences on neural development[11]. These distressed social environments are associated with neuro-anatomical structure of the brain[12,13].

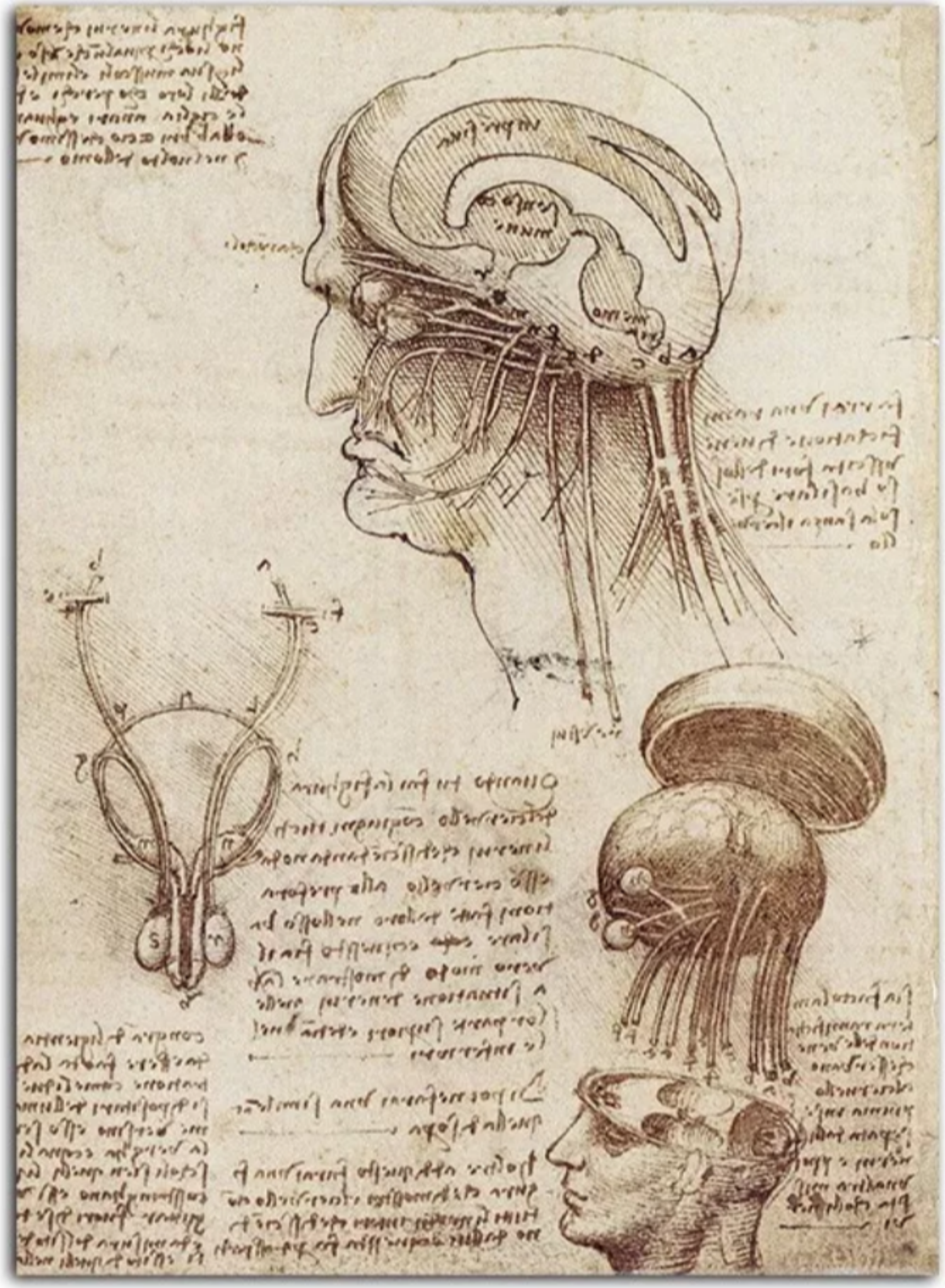
Existing studies have primarily focused on anatomical brain features with small sample sizes, investigating the association between SES and brain structural changes in children and adolescents living in a city or town lacking nation-wide evidence[4,14–16]. In these studies, childhood SES includes various socioeconomic indicators such as family income-to-needs ratio, parental education, family income, and neighborhood deprivation, all of which have the potential to impact a child's maturation and advancement. Children who had a higher SES showed greater development in adult brain surface area, including increased cortical thickness[14]. In addition, another study revealed that higher levels of neighborhood deprivation were associated with greater surface area, while it also was linked with decreased cortical thickness, especially subregions in the prefrontal cortex (PFC)[15]. Furthermore, children whose families have insufficient family income and low parental education levels demonstrated decreases in cortical volume and thickness[4,16]. These associations have been found to persist from early childhood to adulthood consistently[4,14]. In other words, socioeconomic status can have an impact on the process of brain development and result in differences in various cognitive abilities[15,16].

[1]Department of Brain Cognitive Engineering, Korea University, Seoul, Korea. [2]Department of Artificial Intelligence, Korea University, Seoul, Korea. [3]Department of Educational Psychology, University of Minnesota, Minneapolis, MN, USA. [4]Department of Psychology, Seoul National University, Seoul, Korea. [5]Department of Sociology, Korea University, Seoul, Korea. ✉email: hisuk@korea.ac.kr; eunshin@korea.ac.kr

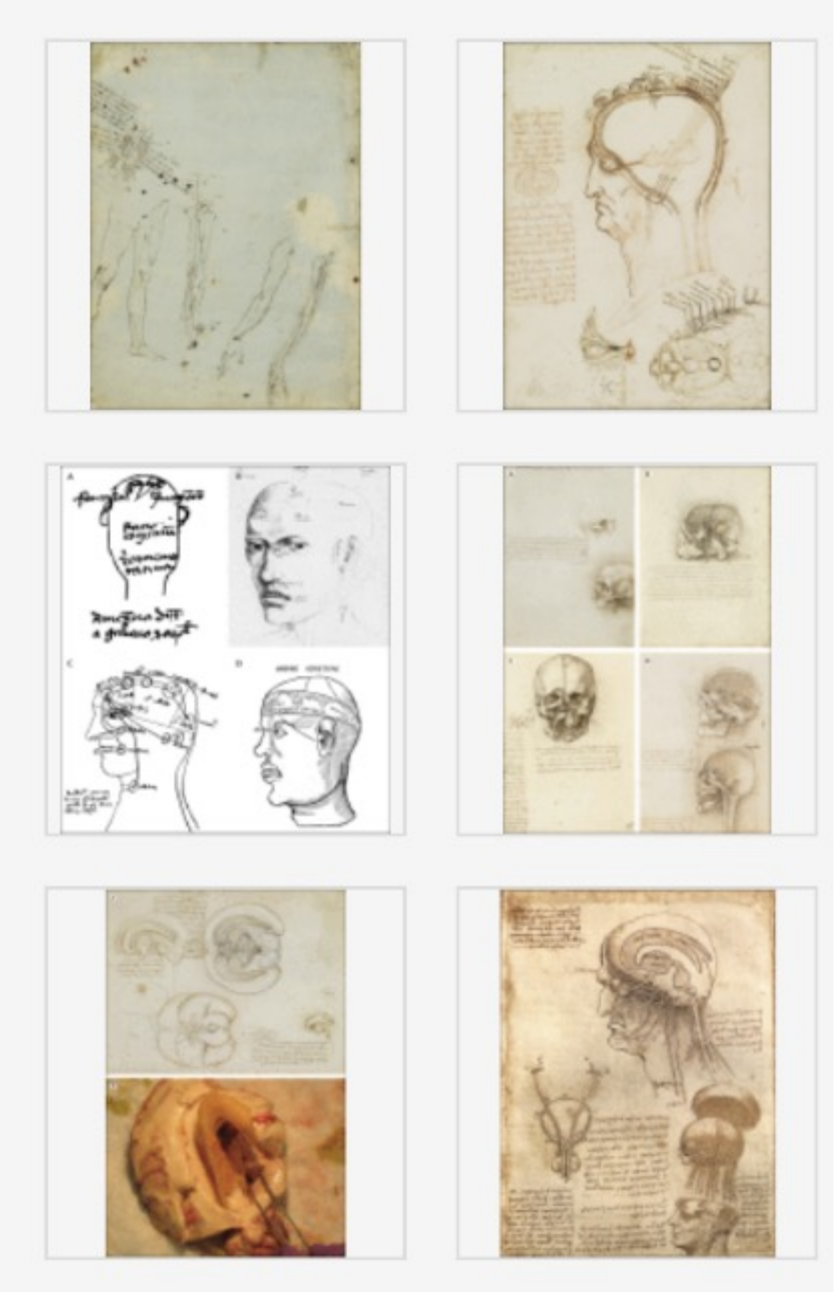
The ABCD Study® is the largest long-term study of brain development and child health in the United States.

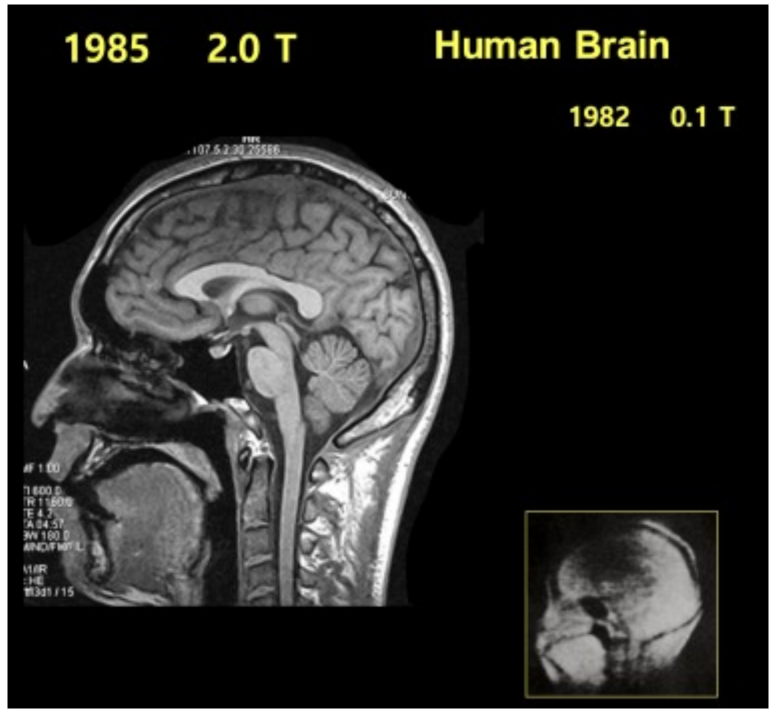


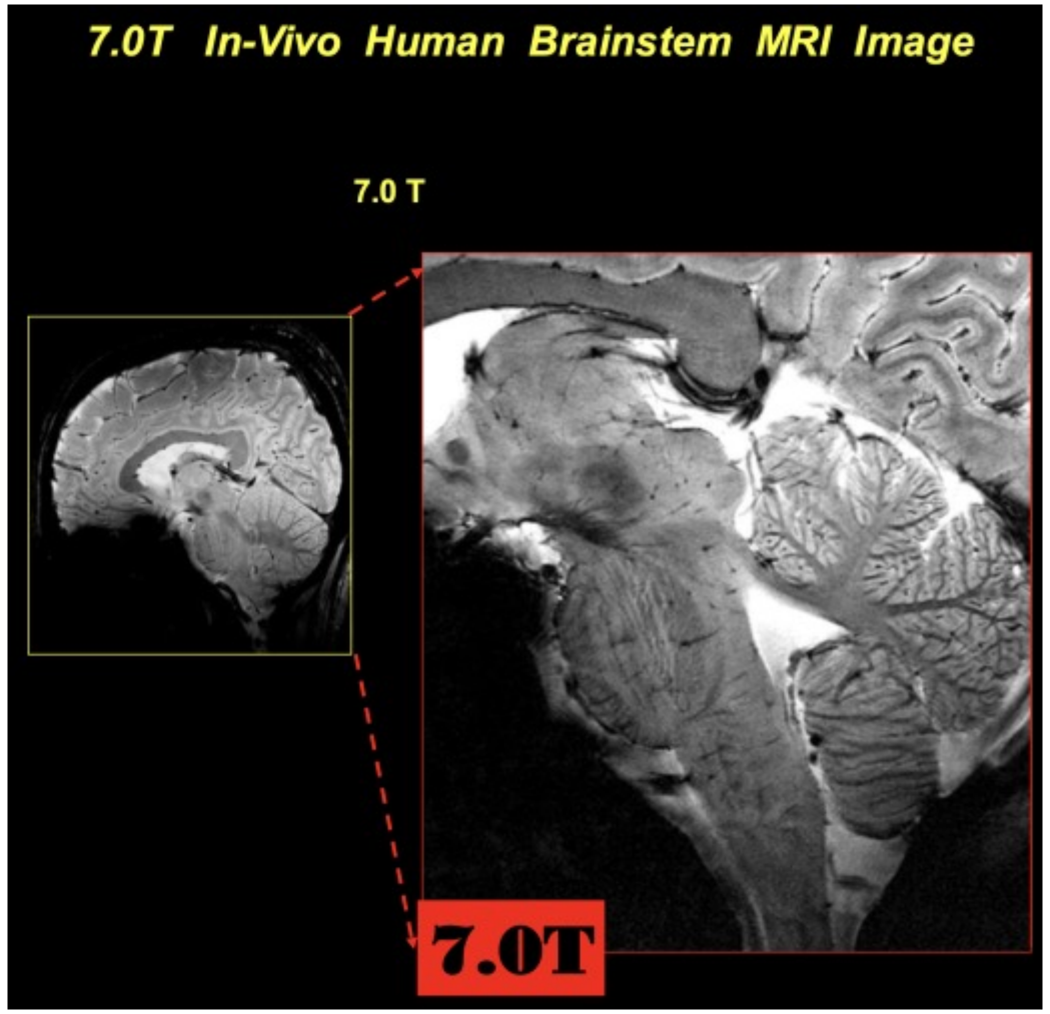
The ABCD Data Repository houses all data generated by the Adolescent Brain Cognitive DevelopmentSM, (ABCD) Study. The ABCD Study(R) is a prospective longitudinal study starting at the ages of 9–10 and following participants for 10 years. The study includes a diverse sample of nearly 12,000 youth enrolled at 21 research sites across the country.



(circa 1485–93)



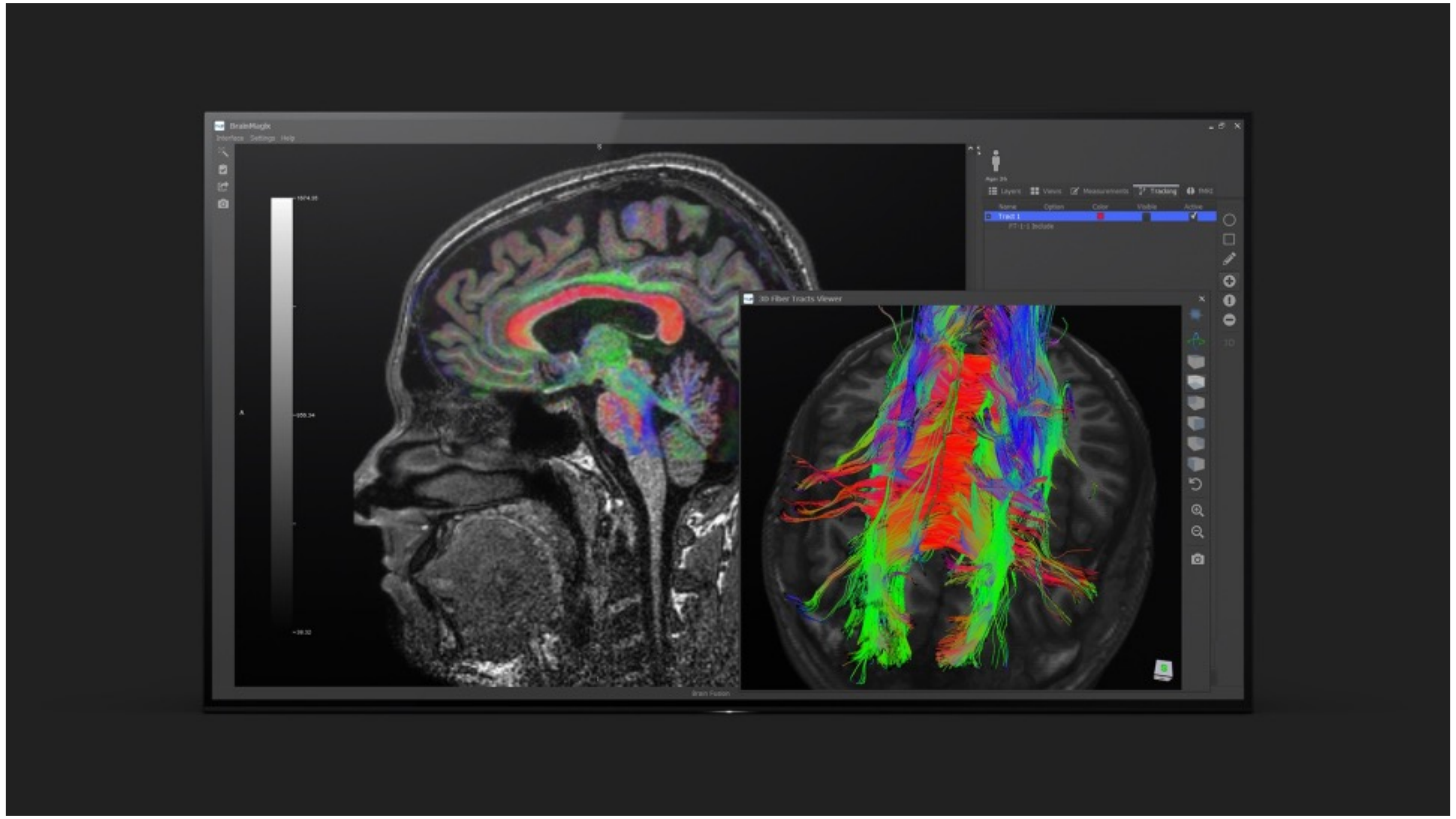


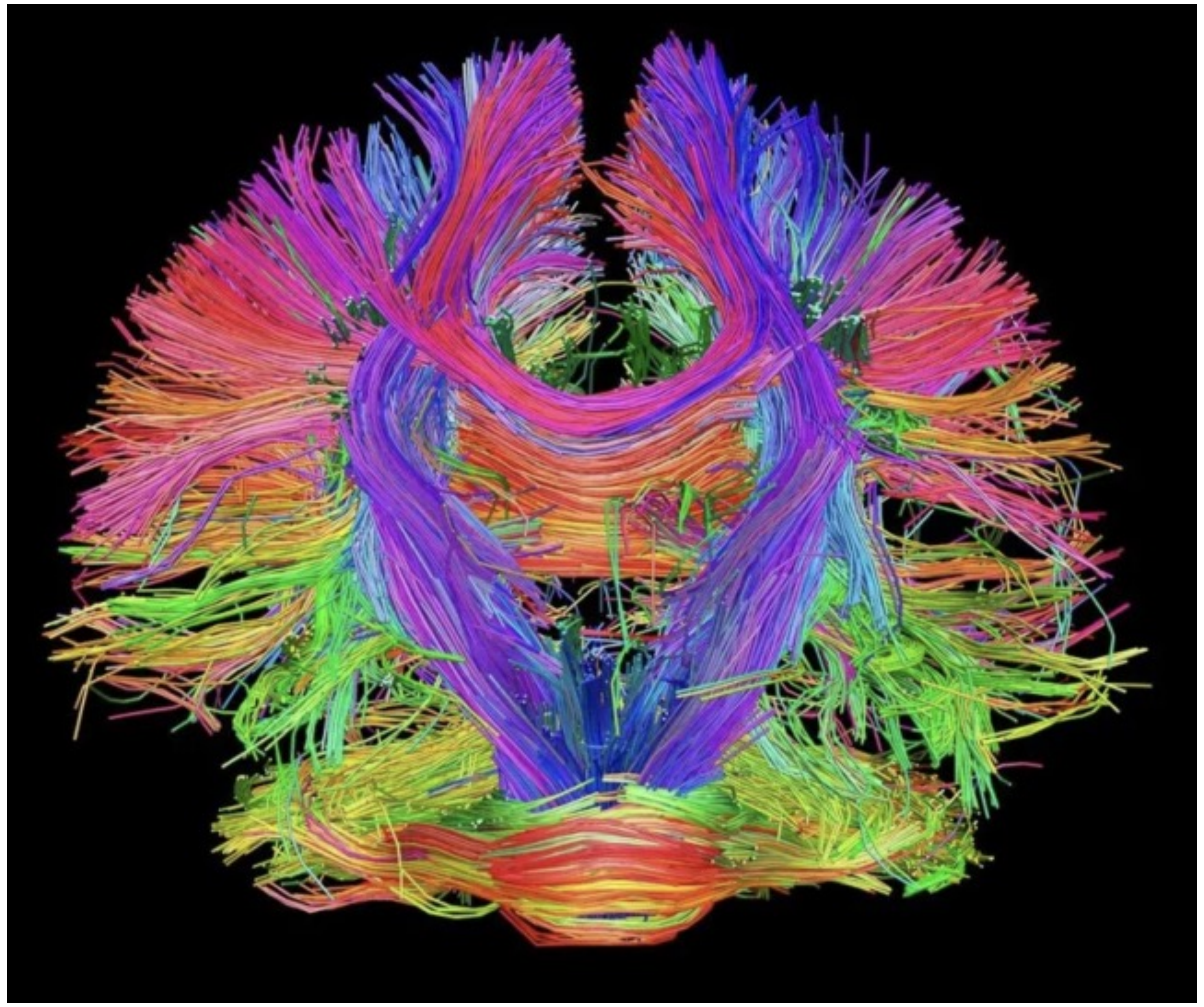


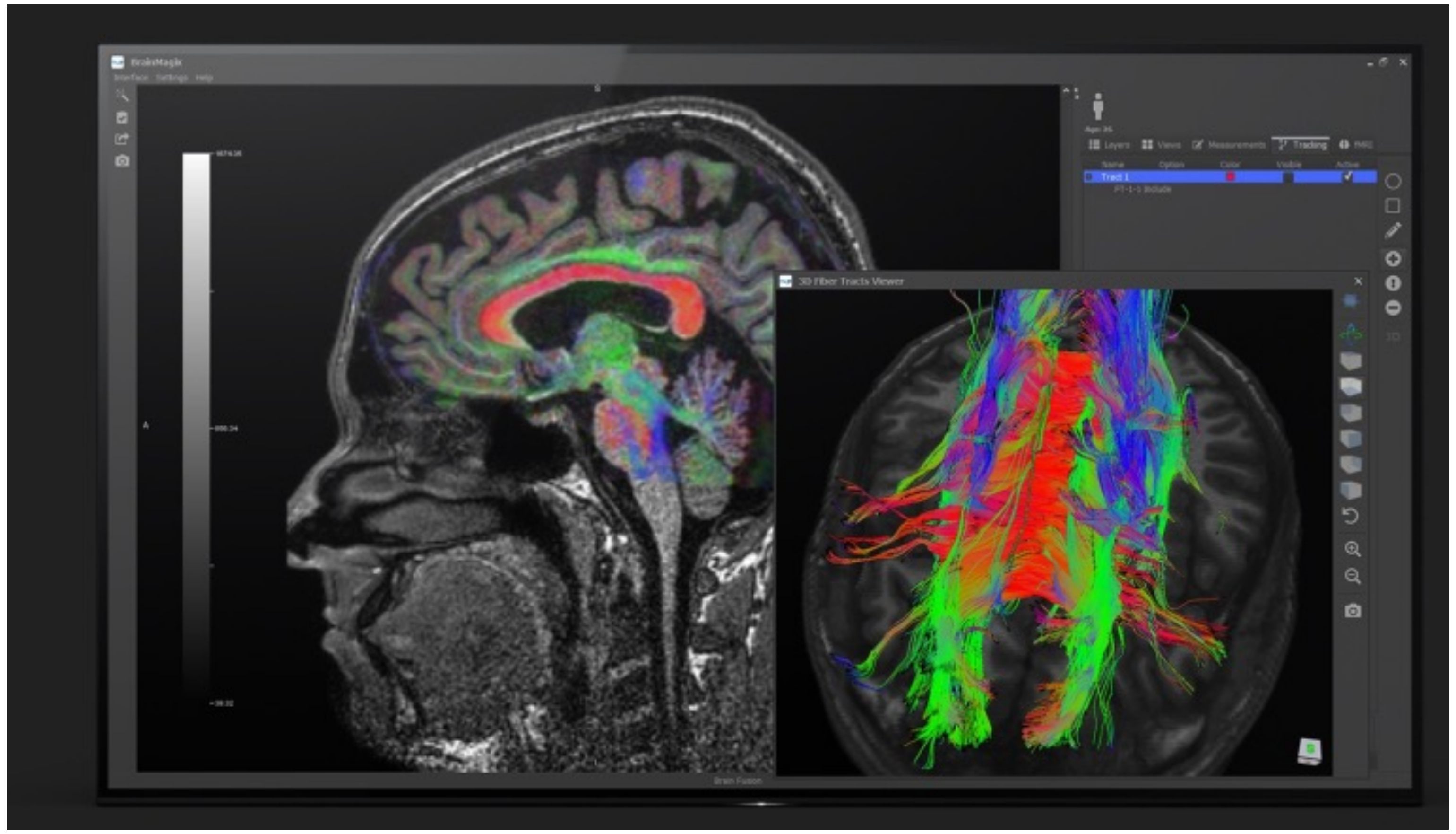
Leonardo da Vinci's studies of the brain
Pevsner, Jonathan (2019)
The Lancet, Volume 393, Issue 10179, 1465 – 1472

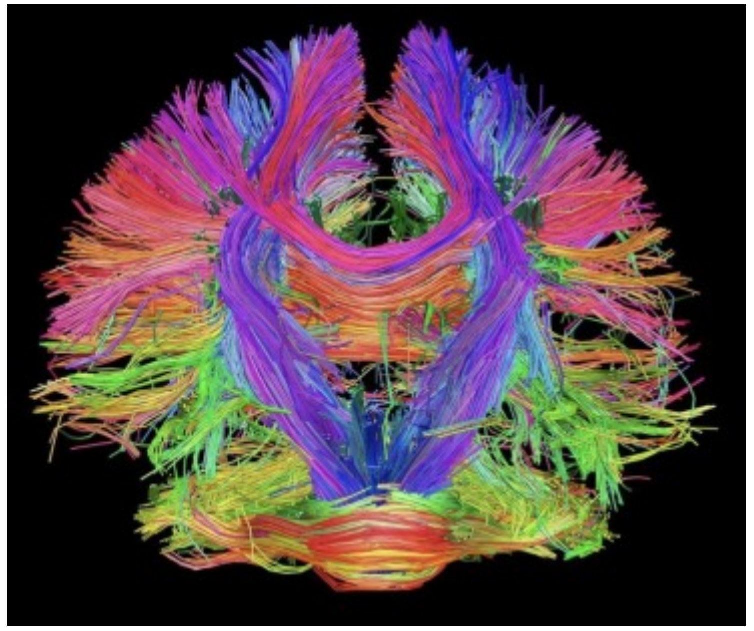
Diffusion Tensor Imaging

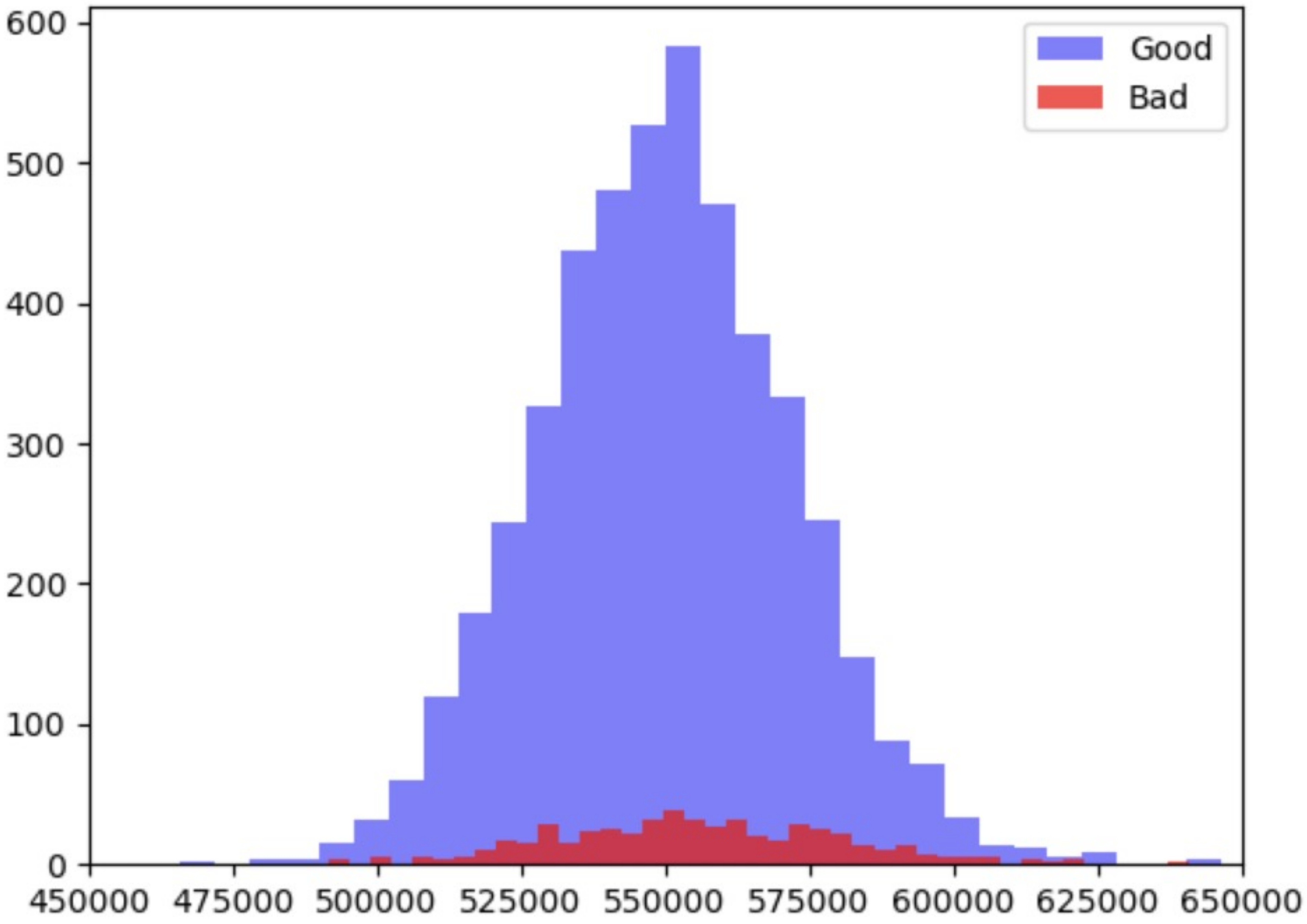
Diffusion Tensor Imaging (DTI) is an magnetic resonance imaging-based neuroimaging technique that makes it possible to estimate the location, orientation, and anisotropy of the white matter tracts of the brain



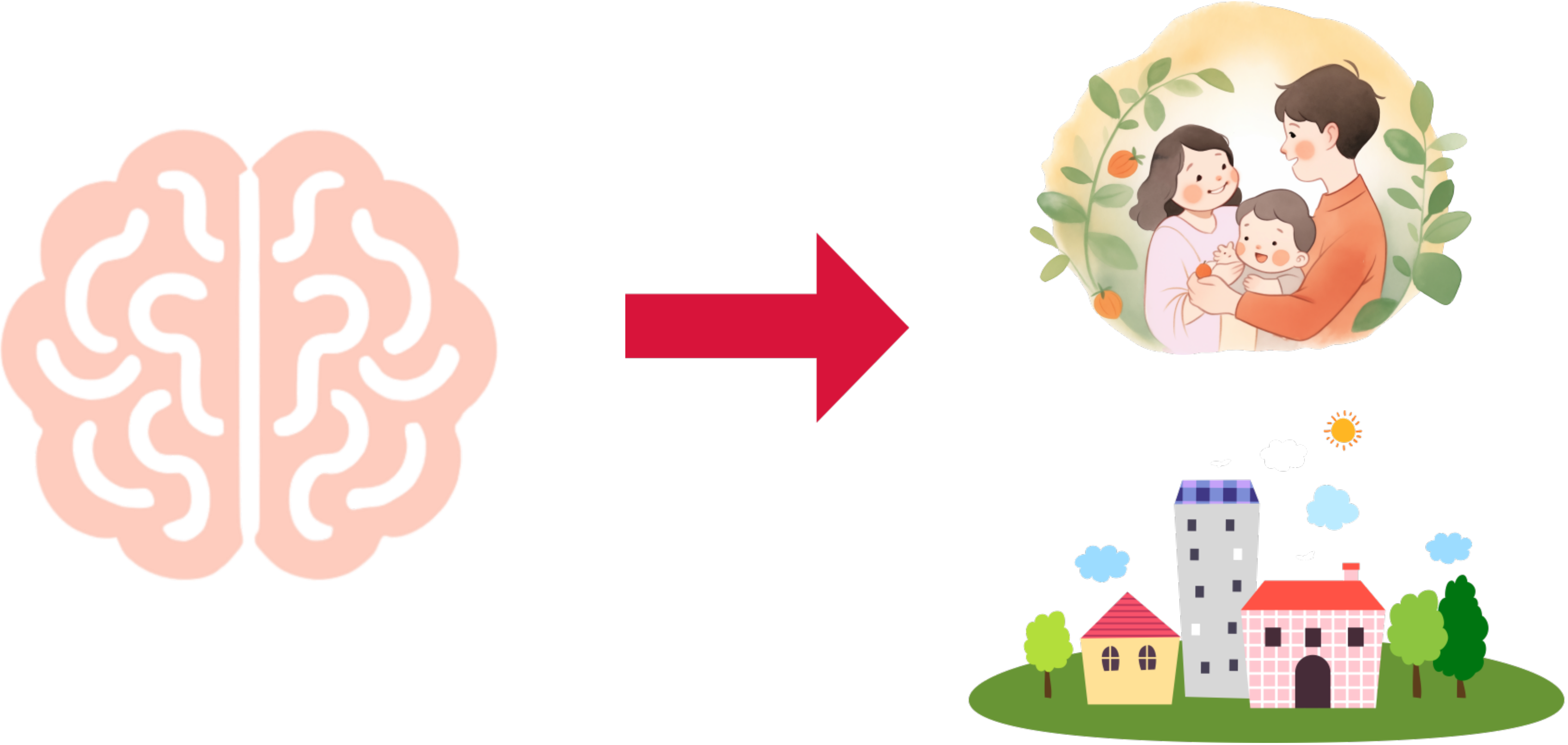








Diffusion Tensor Imaging



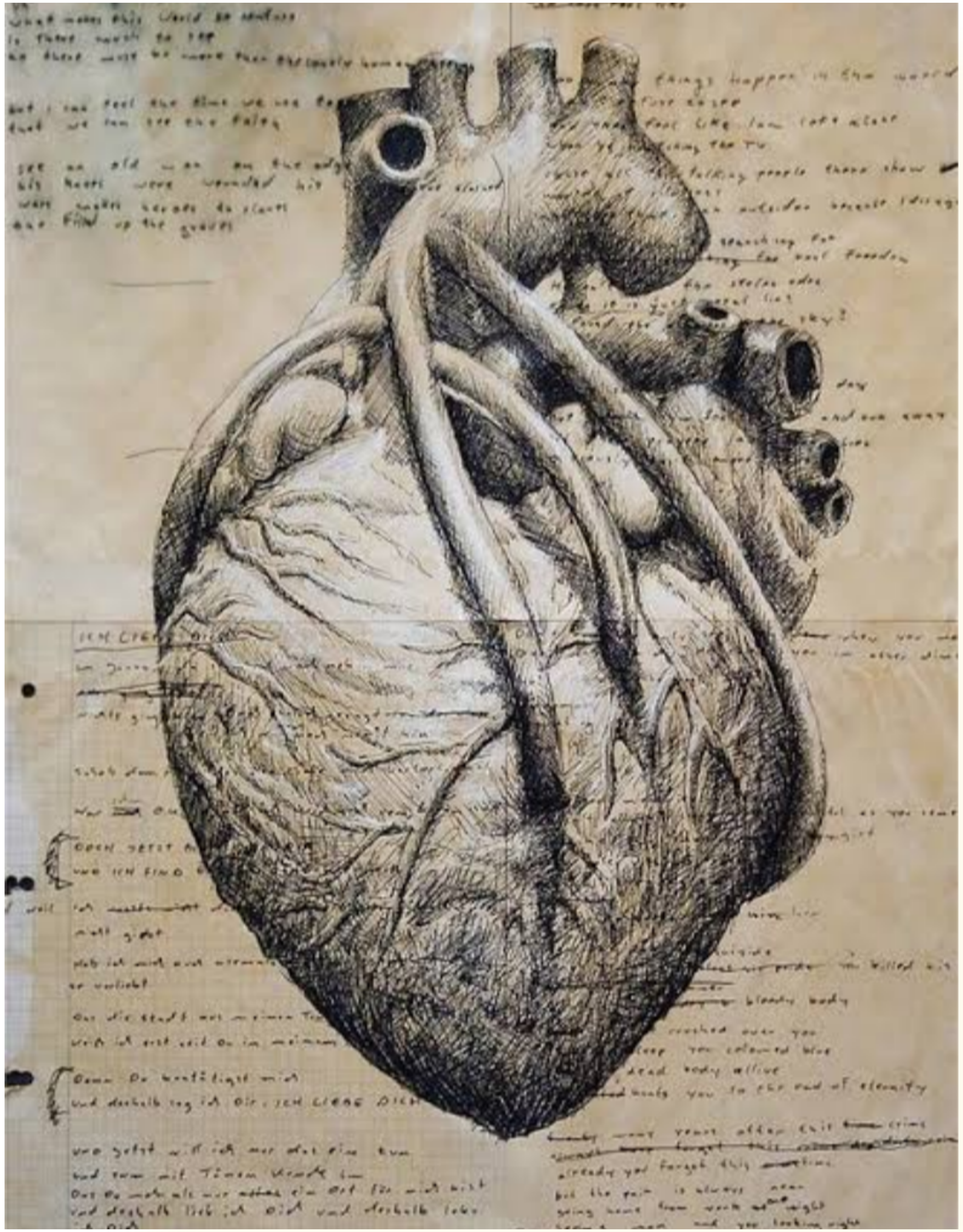
Sociomarkers & Connectomes

ROI	Lobe	Function
IPG	Parietal	Various cognitive processes, including attention, language, and spatial cognition
SPG	Parietal	Spatial attention and perception
RMFG	Frontal	Particularly for executive function (planning, problem-solving, and cognitive flexibility)
SFG	Frontal	A broader range of cognitive processes, including attention, working memory, and motor planning
PoCG	Parietal	Processing somatosensory information from the body (also known as primary somatosensory cortex)
PrCG	Frontal	Controlling voluntary movements in the body (also known as primary somatosensory cortex)
PCU	Parietal	Self-processing, attention, and memory

Table 2. The function of selected significant connections.

Anatomy of Algorithms

The main concern of sociologists lies not in how a single algorithm can explain individuals across diverse social boundaries but rather in ***dissecting how the metrics operate differently within the boundaries.*** *Even if an algorithm can predict the entire general population well, what interests sociologists is how each explanatory feature within the algorithm operates. Moving beyond merely attaining high predictive accuracy, sociological investigation delves into scrutinizing how the key explanatory feature exerts distinctive influence across different social categories.*



K-Depression Prediction

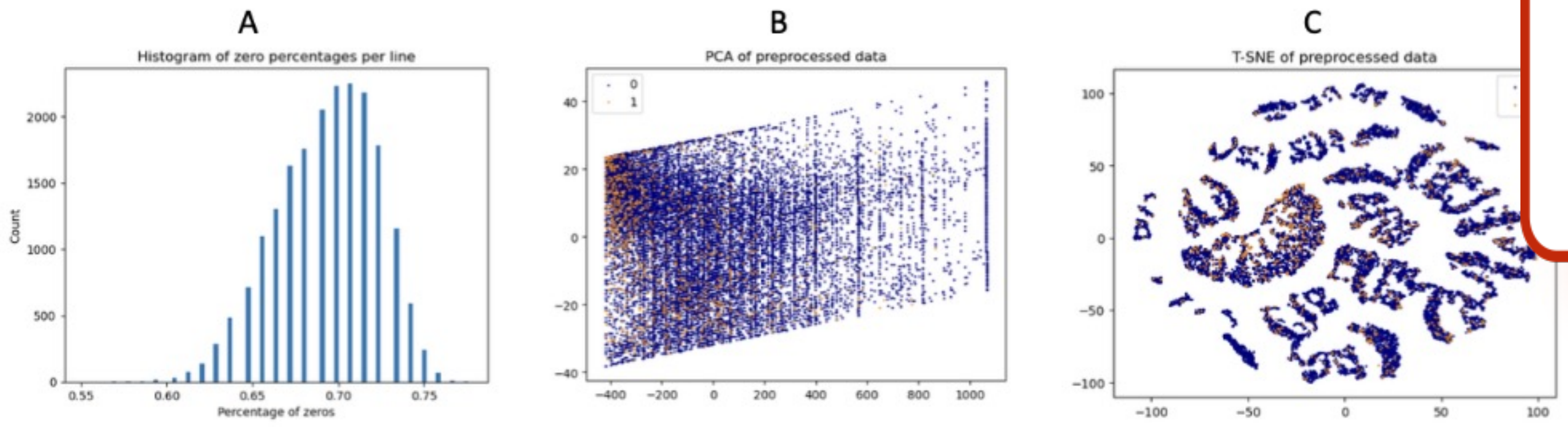


Figure 2. Data sparsity and dimension reduction

"more vulnerable people within the not-depressed who are well separated from the depressed. For instance, while 44.9% of the included 1,559 data points are elementary or junior high graduates, only 26.3% of the filtered-out population were.

Table 1. Five-fold cross validation results for hyperparameter tuning

Zero Percentage	Negative Labels	Positive Labels	CV AUROC	CV Precision	CV Recall
0.64	910	1088	0.9351	0.8663	0.8686
0.65	1559	1088	0.9296	0.8979	0.7806
0.66	2593	1088	0.9130	0.8377	0.7057

Table 2. FT-Transformer prediction model performance on test dataset

Model	AUROC	Precision	Recall
All-inclusive	0.9600	0.9155	0.8287
Sociological	0.7871	0.9800	0.1503
Medical	0.9424	0.9690	0.7669

Target Variable

Questioning PHQ 9

hetegogeneous inviduals are categrozed in a shell(the depressed)

<Table 2> Patient Health Questionnair

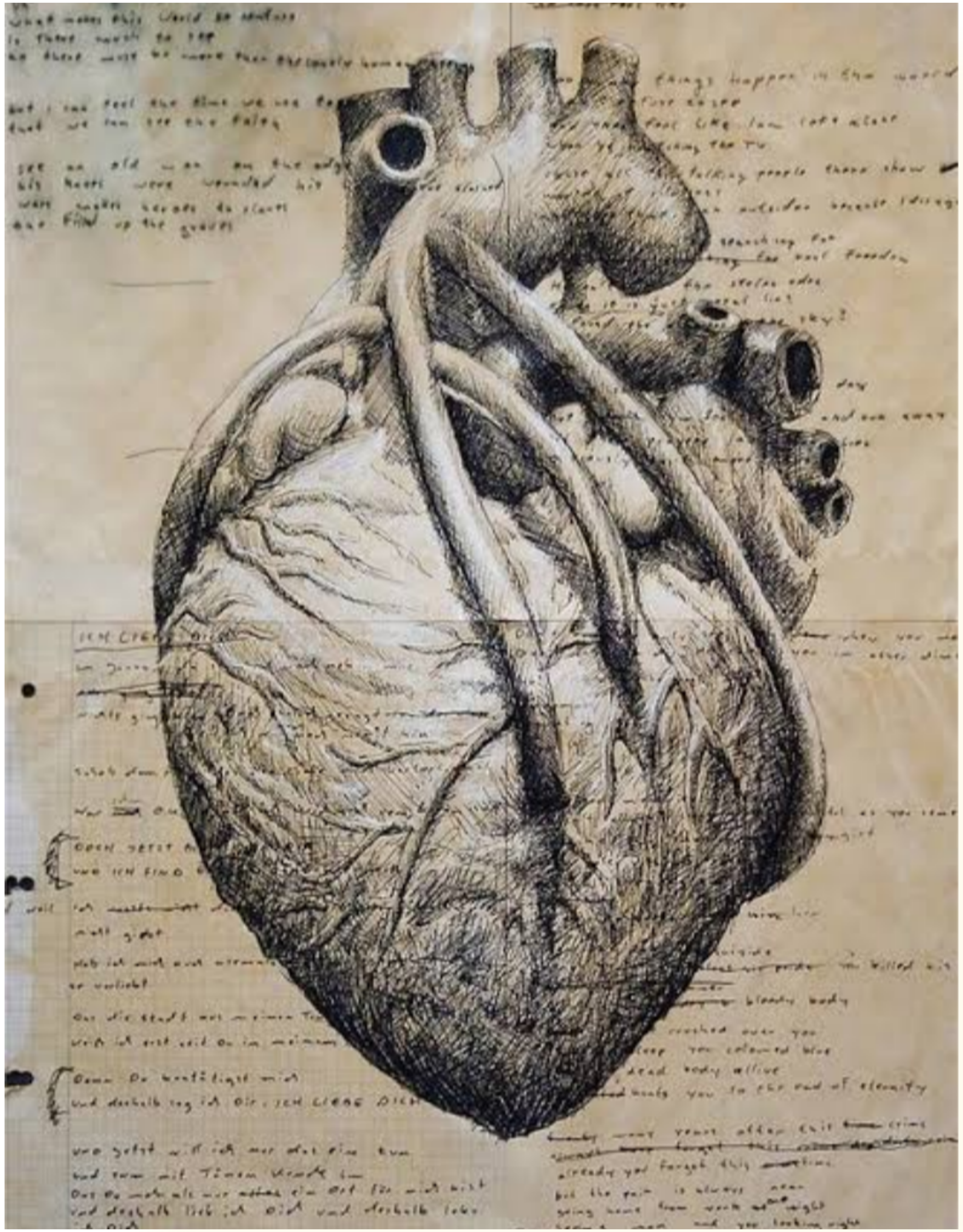
Questions
Little interest or pleasure in doing things?
Feeling down, depressed, or hopeless?
Trouble falling or staying asleep, or sleeping too much?
Feeling tired or having little energy?
Poor appetite or overeating?
Feeling bad about yourself - or that you are a failure or have let yourself or your family down?
Trouble concentrating on things, such as reading the newspaper or watching television?
Moving or speaking so slowly that other people could have noticed? Or so fidgety or restless that you have been moving a lot more than usual?
Thoughts that you would be better off dead, or thoughts of hurting yourself in some way?

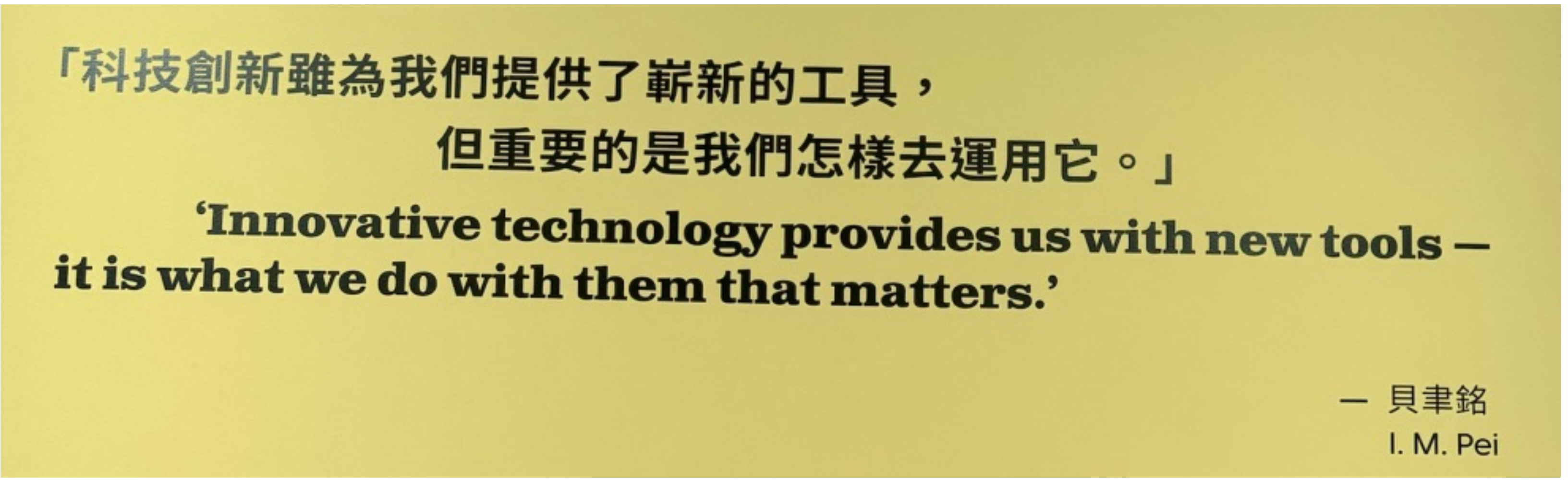
<Table 3> AUROC, Precision, and Recall by Target Variable

Target Variable	AUROC	Precision	Recall
Depression (PHQ-9 score same with or over 10)	0.7740	0.7143	0.0917
Lack of interest *and* depressive mood	0.9010	0.8083	0.6831
Lack of interest *or* depressive mood	0.9385	0.9634	0.7884

The Quest for Explainability and Interpretability

Through the anatomy of the fitted algorithms, we can debunk hidden social mechanisms in unprecedented ways.





科技創新雖為我們提供了嶄新的工具，
但重要的是我們怎樣去運用它。

Innovative technology provides us with new tools - it is what we do with them that matters.

THANK YOU!

eunkyongshin.com