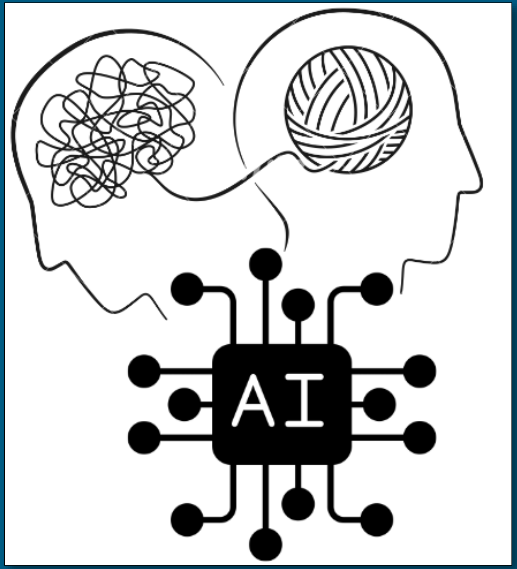
An Introduction to Artificial Intelligence and Machine Learning for Applied Psychology Research



Dr. ANOOP KADAN

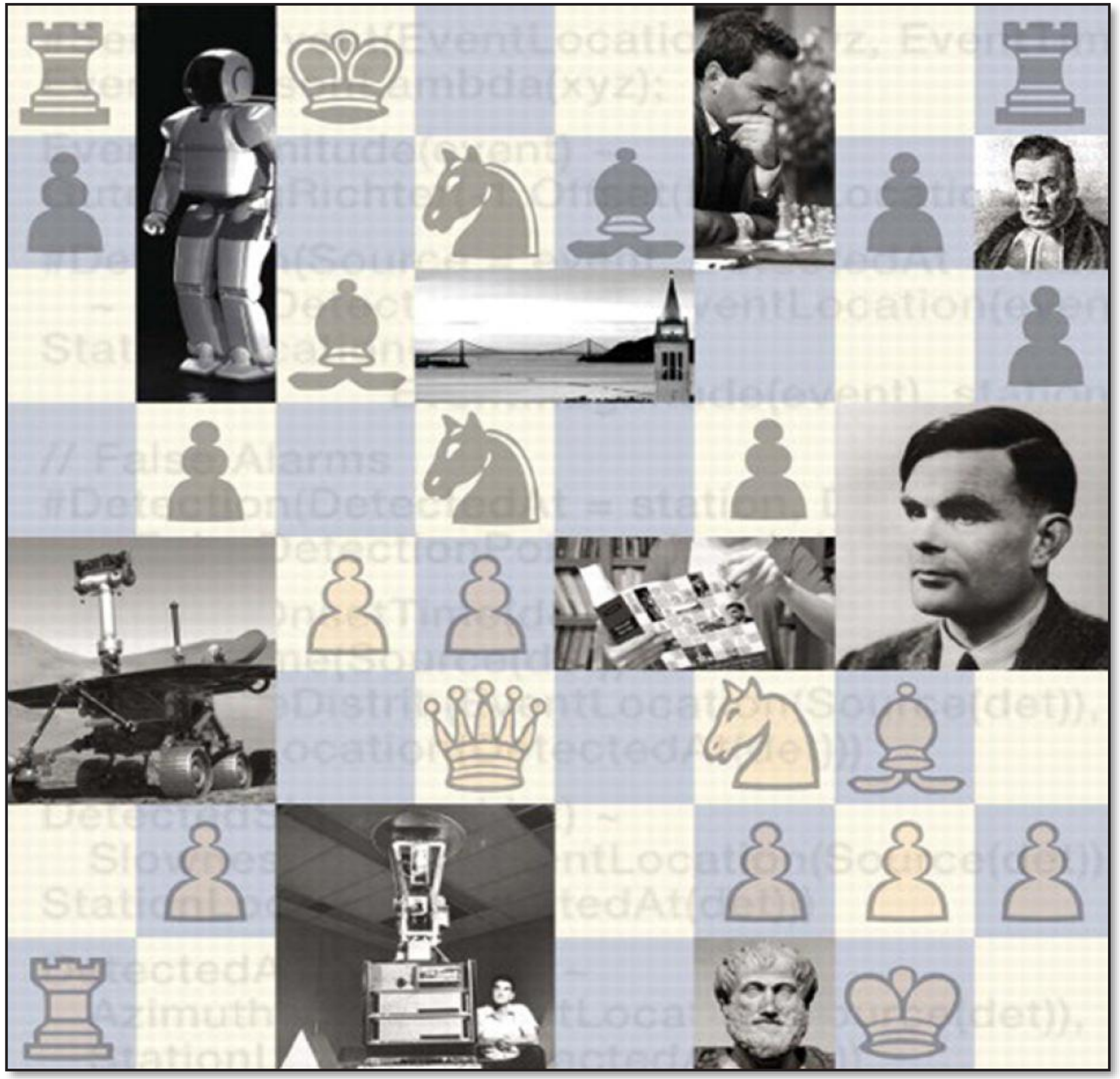
Lecturer in Machine Learning
Centre for Innovation in Mental Health
School of Psychology
University of Southampton
a.kadan@soton.ac.uk

PGR Hub PGR Development Hub

TALK OUTLINE

University of Southampton

- **Evolutionary picture of Artificial Intelligence (AI)**
  - Foundations & the History
  - Approaches
- **Canonical Architecture of AI Systems**
- **Fundamentals of Machine Learning (ML)**
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning
  - The Notion of Deep Learning
- **Applications of AI/ML in Psychology**
  - Mental Health
  - Cognitive Neuroscience
  - Clinical Neuroscience
  - and many more…



© Artificial Intelligence: A Modern Approach, Stuart Russell and Peter Norvig

OUTLINE

University of Southampton

- **Evolutionary picture of Artificial Intelligence (AI)**
  - Foundations & the History
  - Approaches
- **Canonical Architecture of AI Systems**
- **Fundamentals of Machine Learning (ML)**
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning
  - The Notion of Deep Learning
- **Applications of AI/ML in Psychology**
  - Mental Health
  - Cognitive Neuroscience
  - Clinical Neuroscience
  - and many more…

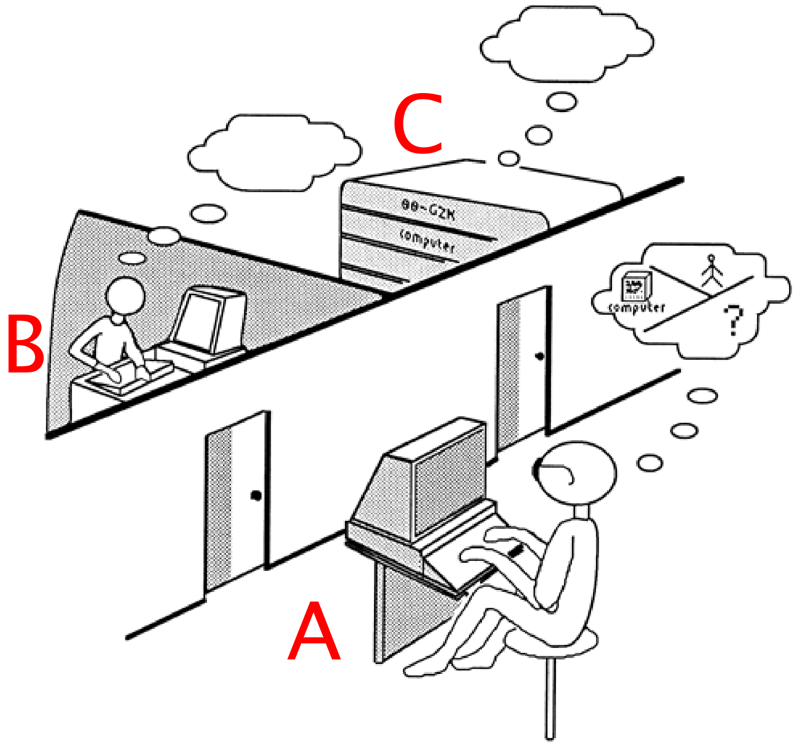
WHAT IS **ARTIFICIAL INTELLIGENCE**?

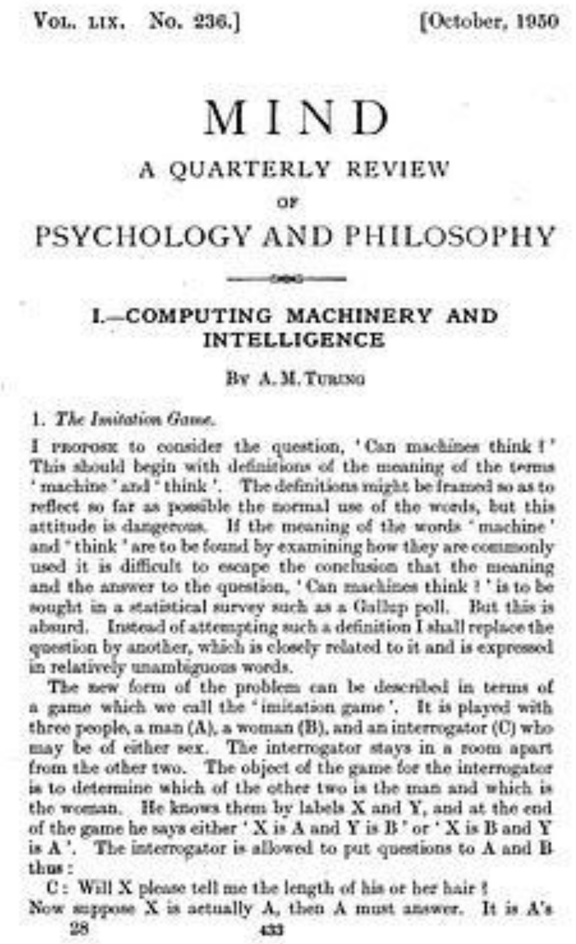
Thinking Humanly	**Thinking Rationally**
"The exciting new effort to **make computer think** … machines with minds, in the full and literal sense", (*Haugeland, 1985*) "[the automation of] activities that we **associate with human thinking**, activities such as decision-making, problem solving, learning, …" (Bellman, 1978)	"The study of **mental faculties through the use of computational models**." (*Charniak & McDermott, 1985*) "The **study of computation** that make it possible to perceive, reason, and act" (Winston, 1992)
Acting Humanly "The art of creating machines that **perform functions** that require intelligence when performed by people" (*Kwrzweil, 1990*) "The study of how make **computers do things** at which, at the moment, people are better" (*Rich & Knight, 1991*)	**Acting Rationally** "Computational intelligence is the study of design of **intelligent agent**" (*Poole et al., 1998*) "AI … is concerned with **intelligent behavior in artifacts**" (*Nilsson, 1998*)

- **Thinking**: i.e., thought process and **reasoning**
- **Acting**: i.e., behavior
- **Humanly**: i.e., Fidelity to human performance
- **Rationally**: i.e., Doing the 'right' thing, given what is known
  (Rationalist approach involves a combination of mathematics & engineering)

WHAT IS **ARTIFICIAL INTELLIGENCE**?

Acting Humanly: The Turing Test approach (1950)





Vol. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. *The Imitation Game.*

I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's

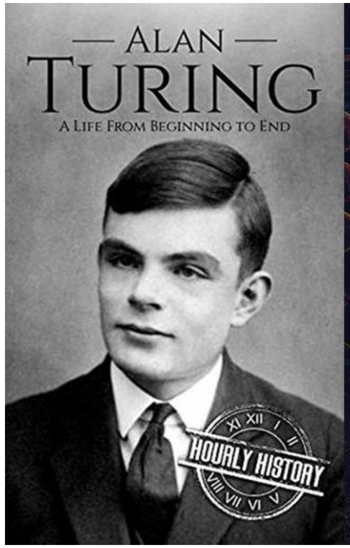
28 433

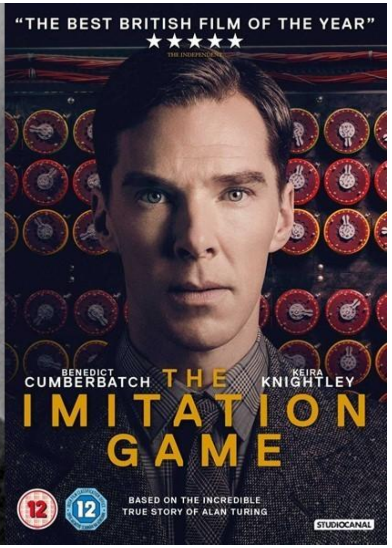
Computer would need to possess the following capabilities

1. NLP
2. Knowledge Representation
3. Automated Reasoning
4. ML

total Turing Test (physical interaction)

1. CV
2. Robortics





Thinking Humanly: The cognitive modeling approach

- Thought process and **reasoning**
- If the program’s input–output behavior matches corresponding human behavior, that is evidence that some of the program’s mechanisms could also be operating in humans
- **General Problem Solver** (Allen Newell and Herbert Simon, 1961) - were more concerned with comparing the trace of its reasoning steps to traces of human subjects solving the same problems.
- The interdisciplinary field of **cognitive science** brings together **computer models from AI** and experimental techniques from **psychology** to construct precise and testable theories of the human mind

Thinking Rationally: The “laws of thought” approach

- Doing the ‘right’ thing, given what is known
- Rationalist approach involves a combination of mathematics & engineering
- Aristotle (the Greek philosopher) was one of the first to attempt to codify “right thinking,” that is, irrefutable reasoning processes
- His **syllogisms** provided patterns for argument structures that always yielded correct conclusions when given correct premises;

$$socrates\ is\ a\ man;\ all\ men\ are\ mortal;$$
$$\therefore socrates\ is\ mortal$$

- These laws of thought were supposed to govern the operation of the mind; their study initiated the field called **LOGIC**

WHAT IS **ARTIFICIAL INTELLIGENCE**?

Acting Rationally: The rational agent approach

- An **agent** is just something that acts
- **Computer agents**: operate autonomously, perceive their environment, persist over a prolonged time period, adapt to change, and create and pursue goals.
- A **rational agent** is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome.

THE FOUNDATIONS OF **ARTIFICIAL INTELLIGENCE**

University of Southampton

- **Philosophy**
  - Can formal rules be used to draw valid conclusions?
  - How does the mind arise from a physical brain?
  - Where does knowledge come from?
  - How does knowledge lead to action?
- **Mathematics**
  - What are the formal rules to draw valid conclusions?
  - What can be computed?
  - How do we reason with uncertain information?
- **Economics**
  - How should we make decisions so as to maximize payoff?
  - How should we do this when others may not go along?
  - How should we do this when the payoff may be far in the future?
- **Neuroscience**
  - How do brains process information?
- **Psychology**
  - How do humans and animals think and act?
- **Computer Engineering**
  - How can we build an efficient computer?
- **Control Theory and Cybernetics**
  - How can artifacts operate under their own control?
- **Linguistics**
  - How does language relate to thought?

- ❑ Aristotle (384–322 B.C.), was the first to formulate a precise set of laws governing the rational part of the mind
- ❑ Developed an informal system of **syllogisms** for proper reasoning
- ❑ Rene Descartes (1596–1650) was a strong advocate of the power of reasoning in understanding the world, a philosophy now called rationalism

THE FOUNDATIONS OF **ARTIFICIAL INTELLIGENCE**

- **Philosophy**
  - Can formal rules be used to draw valid conclusions?
  - How does the mind arise from a physical brain?
  - Where does knowledge come from?
  - How does knowledge lead to action?
- **Mathematics**
  - What are the formal rules to draw valid conclusions?
  - What can be computed?
  - How do we reason with uncertain information?
- **Economics**
  - How should we make decisions so as to maximize payoff?
  - How should we do this when others may not go along?
  - How should we do this when the payoff may be far in the future?
- **Neuroscience**
  - How do brains process information?
- **Psychology**
  - How do humans and animals think and act?
- **Computer Engineering**
  - How can we build an efficient computer?
- **Control Theory and Cybernetics**
  - How can artifacts operate under their own control?
- **Linguistics**
  - How does language relate to thought?

- ❑ Philosophers staked out some of the fundamental ideas of AI
- ❑ but the leap to a formal science required a level of mathematical formalization in three fundamental areas;
  1. **Logic**
  2. **Computation**
  3. **Probability**
- ❑ Idea of formal **logic** can be traced back to the philosophers of ancient Greece
- ❑ The next step was to determine the **limits** of what could be done with logic and **computation**
- ❑ Incompleteness theorem
- ❑ Tractability
- ❑ NP-completeness

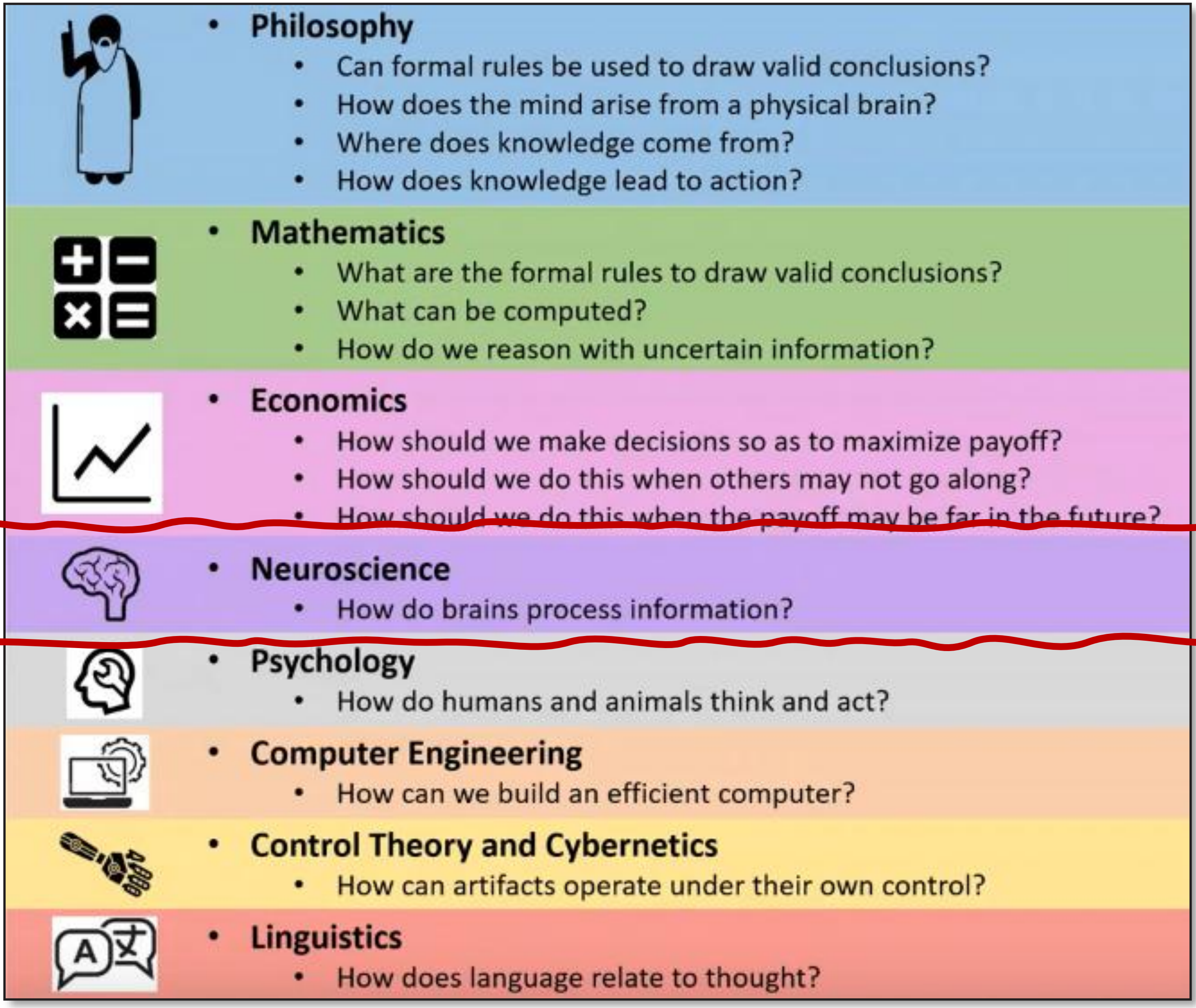
THE FOUNDATIONS OF **ARTIFICIAL INTELLIGENCE**

University of Southampton

- **Philosophy**
  - Can formal rules be used to draw valid conclusions?
  - How does the mind arise from a physical brain?
  - Where does knowledge come from?
  - How does knowledge lead to action?
- **Mathematics**
  - What are the formal rules to draw valid conclusions?
  - What can be computed?
  - How do we reason with uncertain information?
- **Economics**
  - How should we make decisions so as to maximize payoff?
  - How should we do this when others may not go along?
  - How should we do this when the payoff may be far in the future?
- **Neuroscience**
  - How do brains process information?
- **Psychology**
  - How do humans and animals think and act?
- **Computer Engineering**
  - How can we build an efficient computer?
- **Control Theory and Cybernetics**
  - How can artifacts operate under their own control?
- **Linguistics**
  - How does language relate to thought?

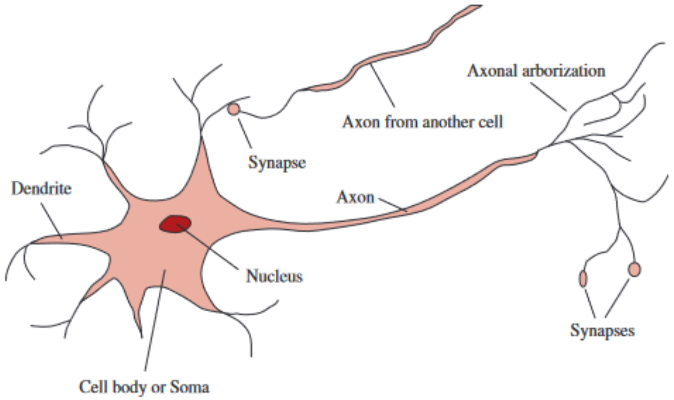
- ❑ Most people think of economics as being about money, but economists will say that they are really studying **how people make choices that lead to preferred outcomes**
- ❑ **Decision theory**, which combines probability theory with utility theory, provides a formal and complete framework for decisions (economic or otherwise) made under uncertainty
- ❑ Von Neumann and Morgenstern's development of **GAME THEORY**
- ❑ Richard Bellman (1957) formalized a class of sequential decision problems called **Markov decision processes** (3dr question)

THE FOUNDATIONS OF **ARTIFICIAL INTELLIGENCE**



- **Philosophy**
  - Can formal rules be used to draw valid conclusions?
  - How does the mind arise from a physical brain?
  - Where does knowledge come from?
  - How does knowledge lead to action?
- **Mathematics**
  - What are the formal rules to draw valid conclusions?
  - What can be computed?
  - How do we reason with uncertain information?
- **Economics**
  - How should we make decisions so as to maximize payoff?
  - How should we do this when others may not go along?
  - How should we do this when the payoff may be far in the future?
- **Neuroscience**
  - How do brains process information?
- **Psychology**
  - How do humans and animals think and act?
- **Computer Engineering**
  - How can we build an efficient computer?
- **Control Theory and Cybernetics**
  - How can artifacts operate under their own control?
- **Linguistics**
  - How does language relate to thought?

- ❑ Neuroscience is the study of the nervous system, particularly the brain
- ❑ Neuron
- ❑ The measurement of intact brain activity
  - electroencephalograph (EEG)
  - functional magnetic resonance imaging (fMRI)
- ❑ Helps improving the understanding of how brain process information



THE FOUNDATIONS OF **ARTIFICIAL INTELLIGENCE**

University of Southampton

- **Philosophy**
  - Can formal rules be used to draw valid conclusions?
  - How does the mind arise from a physical brain?
  - Where does knowledge come from?
  - How does knowledge lead to action?
- **Mathematics**
  - What are the formal rules to draw valid conclusions?
  - What can be computed?
  - How do we reason with uncertain information?
- **Economics**
  - How should we make decisions so as to maximize payoff?
  - How should we do this when others may not go along?
  - How should we do this when the payoff may be far in the future?
- **Neuroscience**
  - How do brains process information?
- **Psychology**
  - How do humans and animals think and act?
- **Computer Engineering**
  - How can we build an efficient computer?
- **Control Theory and Cybernetics**
  - How can artifacts operate under their own control?
- **Linguistics**
  - How does language relate to thought?

- ❑ Behaviorism movement (led by John Watson, 1878–1958)
- ❑ Cognitive psychology

THE FOUNDATIONS OF **ARTIFICIAL INTELLIGENCE**

University of Southampton

- **Philosophy**
  - Can formal rules be used to draw valid conclusions?
  - How does the mind arise from a physical brain?
  - Where does knowledge come from?
  - How does knowledge lead to action?
- **Mathematics**
  - What are the formal rules to draw valid conclusions?
  - What can be computed?
  - How do we reason with uncertain information?
- **Economics**
  - How should we make decisions so as to maximize payoff?
  - How should we do this when others may not go along?
  - How should we do this when the payoff may be far in the future?
- **Neuroscience**
  - How do brains process information?
- **Psychology**
  - How do humans and animals think and act?
- **Computer Engineering**
  - How can we build an efficient computer?
- **Control Theory and Cybernetics**
  - How can artifacts operate under their own control?
- **Linguistics**
  - How does language relate to thought?

- ❑ Among the two things: intelligence and an artifact. The computer has been the **artifact of choice**
- ❑ First operational computer: in 1940 by Alan Turing's team for a single purpose, to deciphering German messages (World War II)
- ❑ Babbage's Analytical Engine
- ❑ Babbage's colleague Ada Lovelace world's first programmer

THE FOUNDATIONS OF **ARTIFICIAL INTELLIGENCE**

University of Southampton

- **Philosophy**
  - Can formal rules be used to draw valid conclusions?
  - How does the mind arise from a physical brain?
  - Where does knowledge come from?
  - How does knowledge lead to action?
- **Mathematics**
  - What are the formal rules to draw valid conclusions?
  - What can be computed?
  - How do we reason with uncertain information?
- **Economics**
  - How should we make decisions so as to maximize payoff?
  - How should we do this when others may not go along?
  - How should we do this when the payoff may be far in the future?
- **Neuroscience**
  - How do brains process information?
- **Psychology**
  - How do humans and animals think and act?
- **Computer Engineering**
  - How can we build an efficient computer?
- **Control Theory and Cybernetics**
  - How can artifacts operate under their own control?
- **Linguistics**
  - How does language relate to thought?

- ❑ Ktesibios of Alexandria (c. 250 B.C.) built the first self-controlling machine: a water clock with a regulator that maintained a constant flow rate
- ❑ Control theory Wiener (1894–1964) - Wiener was a brilliant mathematician who worked with Bertrand Russell, among others, before developing an interest in biological and mechanical control systems and their connection to cognition
- ❑ Wiener's book Cybernetics (1948) became a bestseller and awoke the public to the possibility of artificially intelligent machines

THE FOUNDATIONS OF **ARTIFICIAL INTELLIGENCE**

University of Southampton

- **Philosophy**
  - Can formal rules be used to draw valid conclusions?
  - How does the mind arise from a physical brain?
  - Where does knowledge come from?
  - How does knowledge lead to action?
- **Mathematics**
  - What are the formal rules to draw valid conclusions?
  - What can be computed?
  - How do we reason with uncertain information?
- **Economics**
  - How should we make decisions so as to maximize payoff?
  - How should we do this when others may not go along?
  - How should we do this when the payoff may be far in the future?
- **Neuroscience**
  - How do brains process information?
- **Psychology**
  - How do humans and animals think and act?
- **Computer Engineering**
  - How can we build an efficient computer?
- **Control Theory and Cybernetics**
  - How can artifacts operate under their own control?
- **Linguistics**
  - How does language relate to thought?

- ❑ In 1957, B. F. Skinner published Verbal Behavior. This was a comprehensive, detailed account of the behaviorist approach to language learning, written by the foremost expert in the field
- ❑ Chomsky's theory, based on syntactic models going back to the Indian linguist Panini (c. 350 B.C.)
- ❑ Modern linguistics and AI - intersecting in a hybrid field computational linguistics or natural language processing

THE HISTORY OF ARTIFICIAL INTELLIGENCE

University of Southampton

1943 – 1955	The gestation of artificial intelligence
1956	The birth of artificial intelligence
1952 – 1969	Early enthusiasm, great expectations
1966 – 1973	A dose of reality
1969 – 1979	Knowledge-based systems: The key to power?
1980 – Present	AI becomes an industry
1986 – Present	The return of neural networks
1987 – Present	AI adopts the scientific method
1995 – Present	The emergence of intelligent agents
2001 – Present	The availability of very large data sets

STATE OF THE ART

1943 –1955 The gestation of artificial intelligence

- The first work that is now generally recognized as AI was done by **Warren McCulloch and Walter Pitts** (1943). They drew on three sources;
  - Knowledge of the **basic physiology and function of neurons in the brain**
  - A formal analysis of **propositional logic** due to Russell and Whitehead
  - **Turing's theory** of computation
- They proposed a model of **artificial neurons** in which each neuron is characterized as being "on" or "off," with a switch to "on" occurring in response to stimulation by a sufficient number of neighbouring neurons.
- They showed, that any computable function could be computed by some network of connected neurons, and that all the logical connectives (and, or, not, etc.) could be implemented by simple net structures
- **Donald Hebb** (1949) demonstrated a simple updating rule for modifying the connection strengths between neurons. His rule, now called **Hebbian Learning**, remains an influential model to this day
- Two undergraduate students at Harvard, **Marvin Minsky and Dean Edmonds**, built the first NN computer in 1950 (SNARC)
- **Alan Turing's vision** was perhaps the most influential (among a number of early examples)
  - lectures on the topic as early as 1947 at the London Mathematical Society and articulated a persuasive agenda in his 1950 article "Computing Machinery and Intelligence."
  - Introduced Turing Test, machine learning, genetic algorithms, and reinforcement learning

University of Southampton

1956 The birth of artificial intelligence

- John McCarthy (Dartmouth College) convinced Minsky, Claude Shannon, and Nathaniel Rochester to help him bring together U.S. researchers interested in automata theory, neural nets, and the study of intelligence
- They organized a two-month workshop at Dartmouth in the summer of 1956
- 10 attendees
  - Marvin Minsky, Claude Shannon, Nathaniel Rochester, Arthur Samuel, Trenchard More, Ray Solomonoff, Oliver Selfridge, Allen Newell, Herbert Simon
- No breakthroughs, but introduced major figures to each other
- For the next 20 years, the field would be dominated by these people and their students and colleagues at MIT, CMU, Stanford, and IBM
- Coined the term '**Artificial Intelligence**' - The science and engineering of making machine intelligence

1952 – 1969 Early enthusiasm, great expectations

- The early years of AI were **full of successes**—in a limited way
- The intellectual establishment, by and large, preferred to believe that "a machine can never do X"
- AI researchers naturally responded by demonstrating one X after another
- John McCarthy referred to this period as the "Look, Ma, no hands!" era.

- Newell and Simon's early success was followed up with the **General Problem Solver**
- At IBM, Nathaniel Rochester and his colleagues produced some of the **first AI programs**
- Herbert Gelernter (1959) constructed the **Geometry Theorem Prover**
- John McCarthy (MIT) - 3 crucial contributions in one historic year 1958
  - ✓ Defined the high-level language **Lisp** (dominant for the next 30years)
  - ✓ He and others at MIT invented time sharing
  - ✓ published a paper entitled *Programs with Common Sense*, in which he described the Advice Taker, a hypothetical program that can be seen as the first complete AI system
- Advice Taker embodied the central principles of knowledge representation and reasoning
- Early work building on the neural networks of McCulloch and Pitts also flourished

University of Southampton

1966 – 1973 A dose of reality

- In almost all cases, however, these early systems turned out to fail miserably when tried out on wider selections of problems and on more difficult problems.

1. A typical story occurred in early **machine translation** efforts
   - ✓ Generously funded by the U.S. National Research Council in an attempt to speed up the translation of Russian scientific papers in the wake of the Sputnik launch in 1957
   - ✓ Accurate translation requires background knowledge in order to resolve ambiguity and establish the sentence
   - ✓ In 1966, a report by an advisory committee found that "there has been no machine translation of general scientific text, and none is in immediate prospect
   - ✓ *All U.S. government funding for academic translation projects was canceled*

2. Early experiments in machine evolution (now called **genetic algorithms**)
   - ✓ Despite thousands of hours of CPU time, almost no progress was demonstrated
   - ✓ *Decision by the British government to end support for AI research in all but two universities*

3. some fundamental limitations on the basic structures being used to generate intelligent behavior
   - ✓ A two-input perceptron could not be trained to recognize when its two inputs were different

University of Southampton

1969 – 1979 Knowledge-based systems: The key to power?

- Problem solving that had arisen during the first decade of AI - A **general-purpose search mechanism** (called weak methods)
- The alternative to weak methods is to use more powerful, **domain-specific knowledge**

1. The DENDRAL program (Buchanan et al., 1969 at Stanford) - Automated the decision-making process and problem solving behavior of organic chemists
   - ✓ Chemical-analysis expert system
   - ✓ The significance of **DENDRAL** was that it was the first successful knowledge-intensive system: its expertise derived from large numbers of special-purpose rules

- Heuristic Programming Project (at Stanford) to investigate the extent to which the new methodology of expert systems could be applied to other areas human expertise

1. medical diagnosis - **MYCIN** to diagnose blood infections
   - ✓ With about 450 rules
   - ✓ MYCIN was able to perform as well as some experts, and considerably better than junior doctors

1980 – Present | AI becomes an industry

- The first successful commercial expert system - **R1**, began operation at the Digital Equipment Corporation
  - The program **helped configure orders for new computer systems**; by 1986, it was **saving** the company an estimated **$40 million a year**.
- In 1981, the **Japanese** announced the "Fifth Generation" project, **a 10-year plan to build intelligent computers** running Prolog
- United States formed the Microelectronics and Computer Technology Corporation (MCC) as a research consortium designed to assure national competitiveness.
- In Britain, the Alvey report **reinstated the funding** that was cut by the Lighthill report

- Overall, the AI industry **boomed from a few million dollars in 1980 to billions of dollars in 1988;**
- including hundreds of companies building expert systems, vision systems, robots, and software and hardware specialized for these purposes.

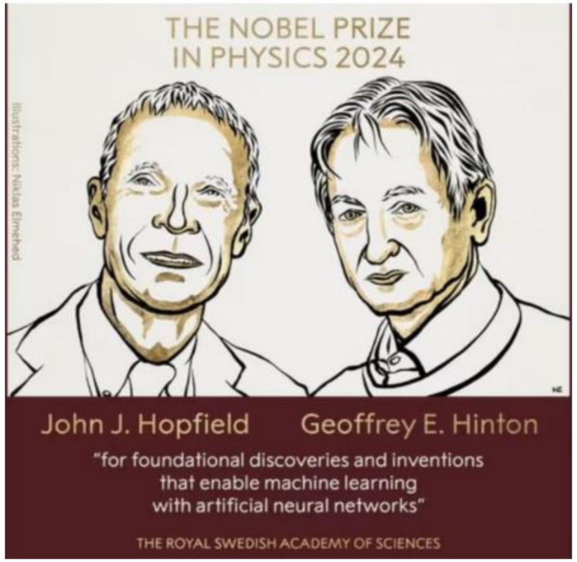
University of Southampton

1986 – Present The return of neural networks

- In the mid-1980s at least four different groups reinvented the **back-propagation learning algorithm** first found in 1969 by Bryson and Ho
- The **algorithm was applied to many learning problems** in computer science and psychology, and the widespread dissemination of the results in the collection **Parallel Distributed Processing caused great excitement**
- Initially, these so-called **connectionist models** of intelligent systems were seen by some as **direct competitors** both

    to the **symbolic models** promoted by Newell and Simon and

    to the **logicist approach** of McCarthy and others
- The **current view is that connectionist and symbolic approaches are complementary**, not competing



1987 – Present | AI adopts the scientific method

- The field of **speech recognition**
  - In the **1970s, a wide variety of different architectures and approaches were tried**.
  - Many of these were rather ad hoc and fragile, and were demonstrated on only a few specially selected examples
  - Approaches based on **hidden Markov models (HMMs)** have come to dominate the area
    - based on a **rigorous mathematical theory**
    - they are generated by a process of training on a **large corpus of real speech data**

- Speech technology and the **related field of handwritten character recognition** are already making the transition to widespread industrial and consumer applications

- Machine translation follows the same course as speech recognition
- Neural networks also fit this trend

- DATA MINING technology has spawned a vigorous new industry

1995 – Present | The emergence of intelligent agents

- Perhaps encouraged by the progress in solving the subproblems of AI, researchers have also started to look at the "**whole agent**" problem again
- One of the most important environments for intelligent agents is the Internet
- AI systems have become so common in Web-based applications that the "-bot" suffix has entered everyday language
- The notion of human-level AI and Artificial General Intelligence came into picture
- AGI looks for a universal algorithm for learning and acting in any environment

University of Southampton

2001 – Present | The availability of very large data sets

- Big Data
- Data can improve accuracy of performance of the algorithm

- word-sense disambiguation (by Yarowsky's, 1995) –
  - given the use of the word "plant" in a sentence, does that refer to flora or factory?
  - Yarowsky showed that the task can be done, with accuracy above 96%, with no labeled examples at all
  - Instead, **given a very large corpus of unannotated text** and just the dictionary definitions of the two senses
  - one can label examples in the corpus and from there bootstrap to learn new patterns that help label new examples

- like this **perform even better** as the **amount of available text goes from a million words to a billion** and that the increase in performance from using more data exceeds any difference in algorithm choice;
  - A mediocre algorithm with 100 million words of unlabeled training data outperforms the best known algorithm with 1 million words.

University of Southampton

SOTA	State of the art

- ❑ Data Science
- ❑ Deep Learning
- ❑ Large Language Models
- ❑ Generative models
- ❑ and many more…

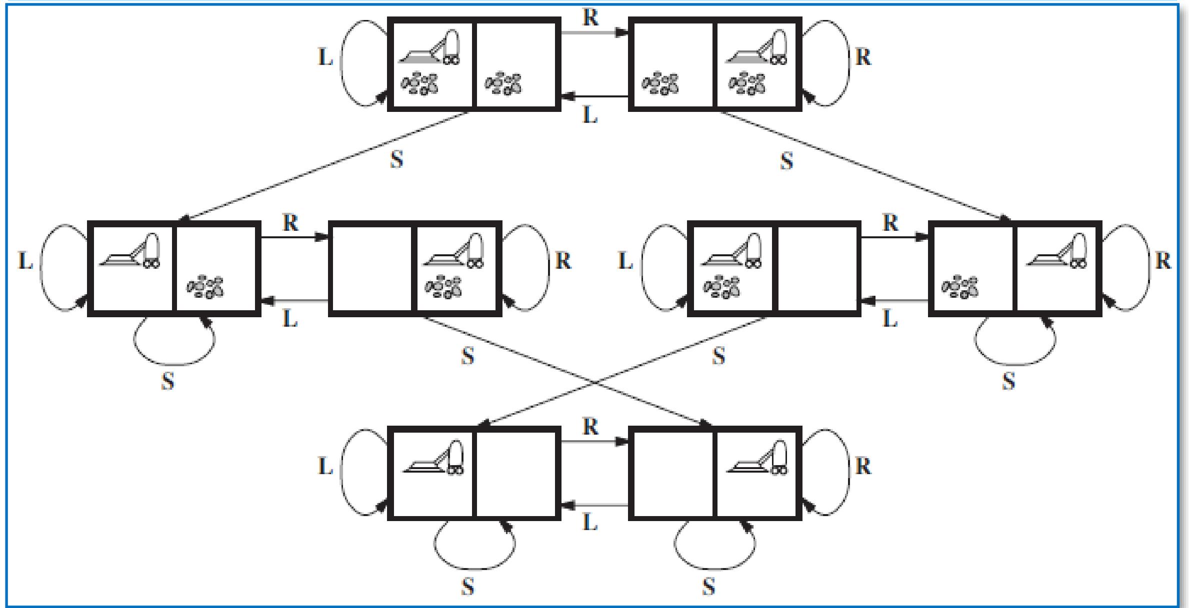
- ❑ Healthcare
- ❑ Education
- ❑ Business
- ❑ and many more…

APPROACHES – PROBLEM SOLVING & SEARCH

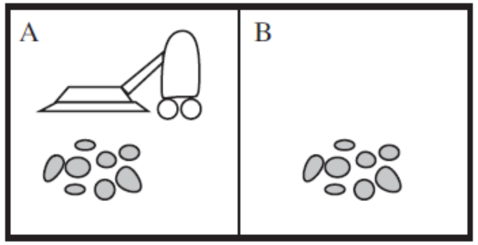
A **problem** can be defined formally by five components

1. The **initial state** that the agent/algorithm starts in
2. A description of the possible **actions** available to the agent
3. A description of what each action does; the formal name for this is the **transition model**
4. The goal test, which determines whether a given state is a **goal state**

An example Toy Problem: The state space for the Vacuum World
Links denote actions: L = Left, R = Right, S = Suck the dirt



© Artificial Intelligence: A Modern Approach, Stuart Russell and Peter Norvig



- **Initial state:** L/R
- **States**: The state is determined by both the agent location and the dirt locations (*8 possible world states*)
- **Actions**: three actions Left, Right, Suck
- **Transition model:** MOVE(L), MOVE(R) or S
- **Goal test**: This checks whether all the squares are clean.
- **Path cost**: Each step costs 1, so the path cost is the number of steps in the path.

Note: Together, the initial state, actions, and transition model implicitly define the **state space** of the problem

University of Southampton

Example #2: The Water Jug Problem in AI

Unlimited water

Jug **#1**
4 Liter

Jug **#2**
3 Liter

Jug **#1**
2 Liter

The Problem:
- How to exactly get **2** liters of water **in jug #1**

Constraints:
- The jug do not have any marking on it to measure the water
- We do not have any measuring devices

- **Initial state:** (0, 0)
- **Goal state**: (2, 0), (2, 3) → (2, X)
- **States**: (4,0), (0,3), (4, 3), etc.

Let's explore some possible solutions to the
Water Jug Problem in AI

Note the Assumptions

- We can **fill a jug** from the pump
- We can **pour water** out of a jug to ground (empty the jug)
- We can pour water from **one jug to another**
- There is **no measuring** devices available

A **solution** is an action sequence, so search algorithms work by considering various possible action sequences. The possible action sequences starting at the initial state form a search tree with the initial state at the root; the branches are actions and the nodes correspond to states in the state space of the problem.

Example #2: The Water Jug Problem in AI

Unlimited water

Jug **#1**
4 Liter

Jug **#2**
3 Liter

Jug **#1**
2 Liter

The Problem:

- ❑ How to exactly get **2** liters of water **in jug #1**

Constraints:

- ❑ The jug do not have any marking on it to measure the water
- ❑ We do not have any measuring devices

State Space or Production Rules of the AI System

1. $(J1, J2) \rightarrow (4, J2)$
2. $(J1, J2) \rightarrow (J1, 3)$
3. $(J1, J2) \rightarrow (0, J2)$, where $J1 > 0$
4. $(J1, J2) \rightarrow (J1, 0)$, where $J2 > 0$
5. $(J1, J2) \rightarrow (J1 - X, J2)$, where $J1 > 0$
6. $(J1, J2) \rightarrow (J1, J2 - X)$, where $J2 > 0$
7. $(J1, J2) \rightarrow (J1 + J2, 0)$, where $J1 + J2 \leq 4$ and $J2 > 0$
8. $(J1, J2) \rightarrow (0, J1 + J2)$, where $J1 + J2 \leq 3$ and $J1 > 0$
9. $(J1, J2) \rightarrow (4, J2 - (4 - J1))$, where $J1 + J24 \geq 4$ and $J2 > 0$
10. $(J1, J2) \rightarrow (J1 - (3 - J2), 3)$, where $J1 + J24 \geq 3$ and $J1 > 0$

University of Southampton

Example #2: The Water Jug Problem in AI

Unlimited water

Jug **#1**
4 Liter

Jug **#2**
3 Liter

Jug **#1**
2 Liter

The Problem:
- How to exactly get **2** liters of water **in jug #1**

Constraints:
- The jug do not have any marking on it to measure the water
- We do not have any measuring devices

State Space or Production Rules of the AI System

1. $(J1, J2) \rightarrow (4, J2)$
2. $(J1, J2) \rightarrow (J1, 3)$
3. $(J1, J2) \rightarrow (0, J2)$, where $J1 > 0$
4. $(J1, J2) \rightarrow (J1, 0)$, where $J2 > 0$
5. $(J1, J2) \rightarrow (J1 - X, J2)$, where $J1 > 0$
6. $(J1, J2) \rightarrow (J1, J2 - X)$, where $J2 > 0$
7. $(J1, J2) \rightarrow (J1 + J2, 0)$, where $J1 + J2 \leq 4$ and $J2 > 0$
8. $(J1, J2) \rightarrow (0, J1 + J2)$, where $J1 + J2 \leq 3$ and $J1 > 0$
9. $(J1, J2) \rightarrow (4, J2 - (4 - J1))$, where $J1 + J24 \geq 4$ and $J2 > 0$
10. $(J1, J2) \rightarrow (J1 - (3 - J2), 3)$, where $J1 + J24 \geq 3$ and $J1 > 0$

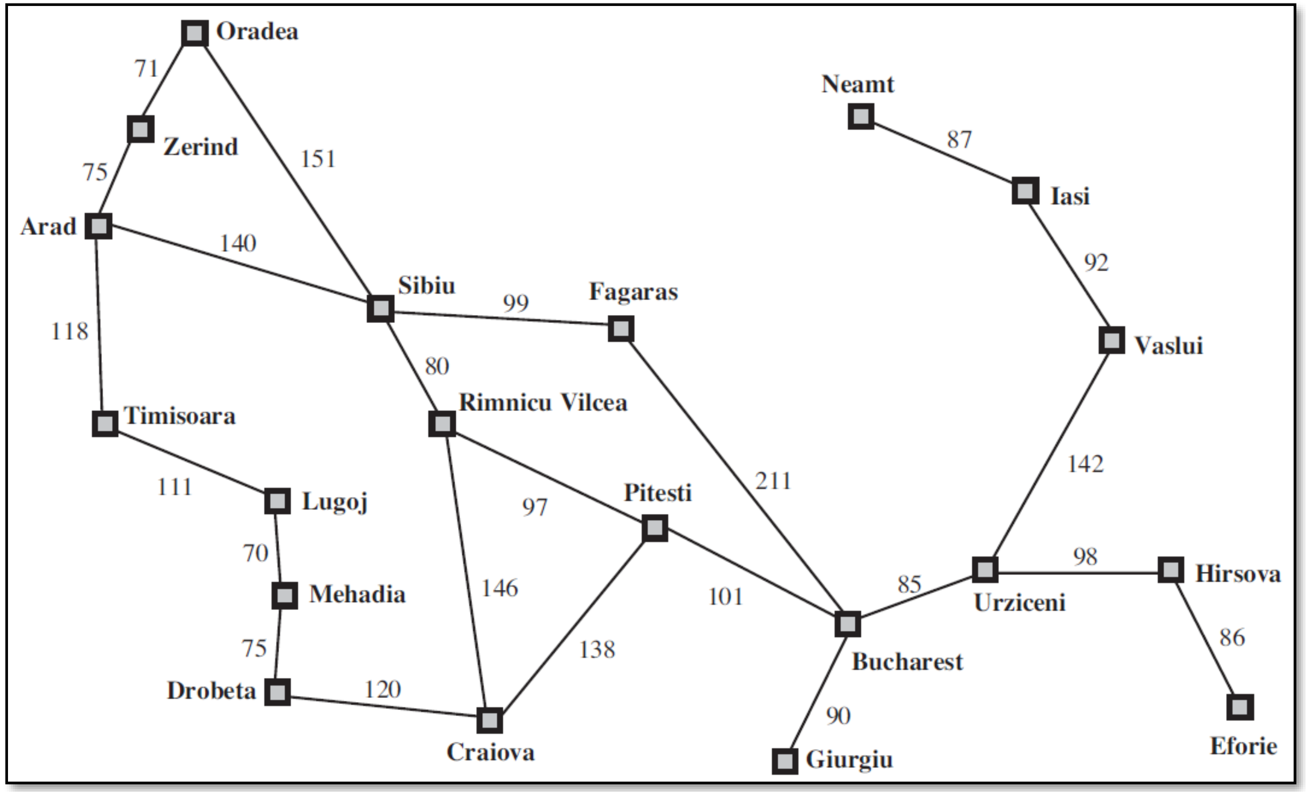
J1	J2	Rule
0	0	Initial state
0	3	R2
3	0	R7
3	3	R2
4	2	R9
0	2	R3
2	0	R7 (Goal state)

University of Southampton

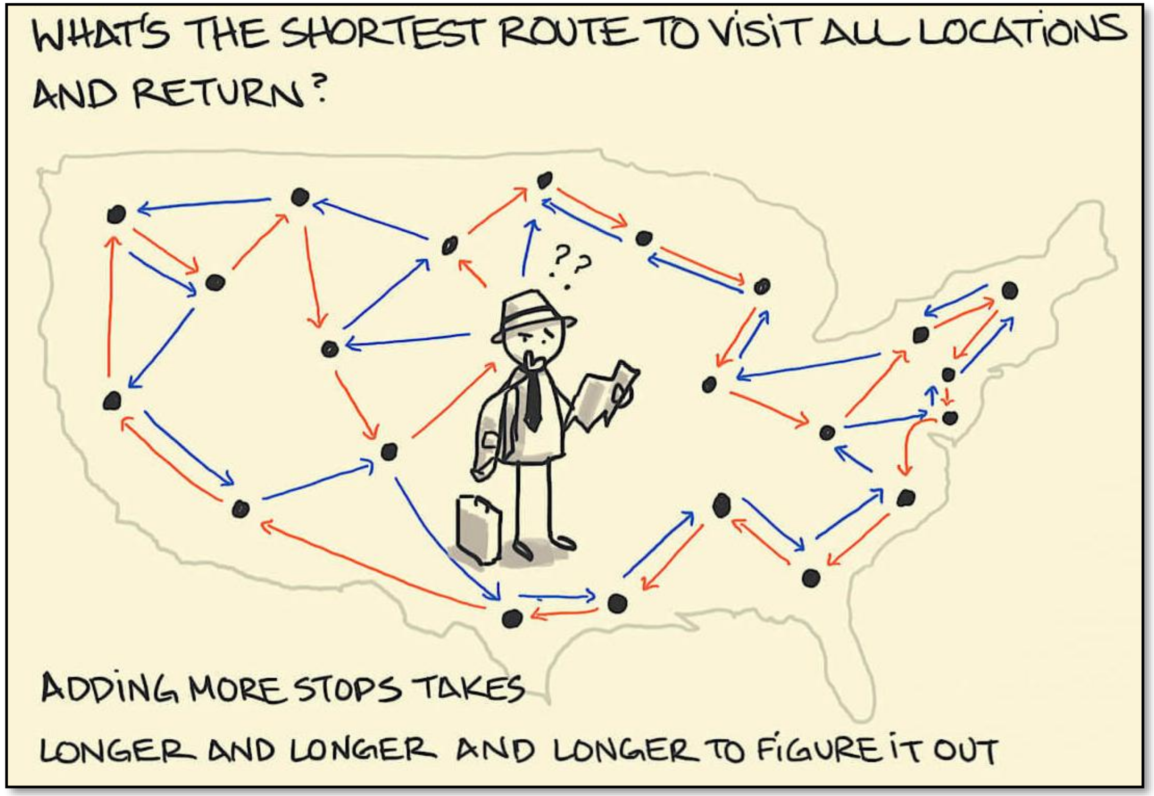
Example #2: The Water Jug Problem in AI

Example #3 & #4: Real World Problems

- **Route-finding Problem**: is defined in terms of specified locations and transitions along links between them
- **Traveling Salesperson Problem (TSP)**: is a touring problem in which each city must be visited exactly once. The aim is to find the shortest tour.



A simplified road map of part of Romania



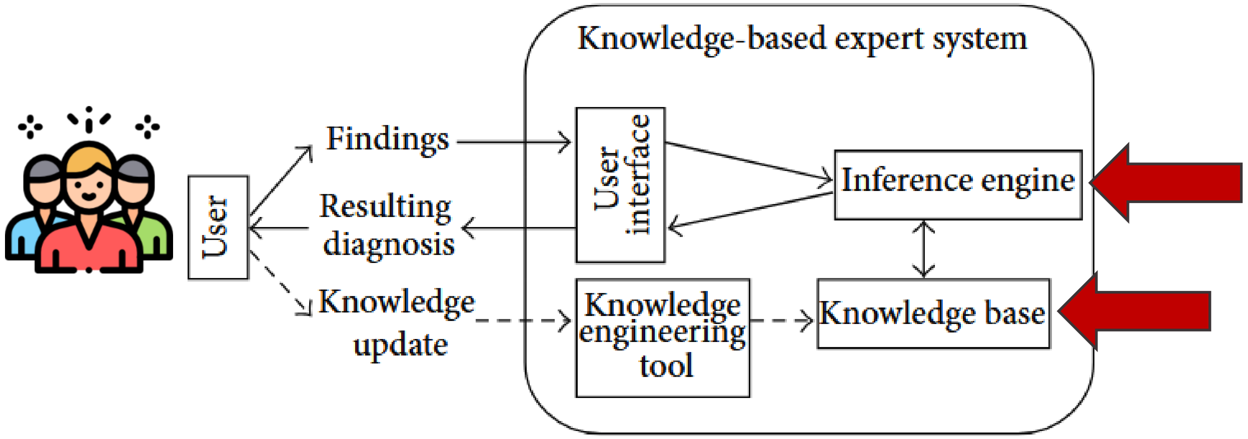
Traveling Salesperson Problem

APPROACHES – KNOWLEDGE BASED SYSTEMS

University of Southampton

- How the intelligence of humans is achieved?
- Not by purely reflex mechanisms but by processes of reasoning that operate on **internal representations of knowledge**
- In AI, this approach to intelligence is embodied in knowledge-based agents or systems

- A knowledge-based system (KBS) is a program that captures and uses knowledge from a variety of sources
- A KBS assists with solving problems, particularly complex issues, by artificial intelligence
- These systems are primarily used to support human decision making, learning, and other activities
- The central component of a knowledge-based system:
  1. **Knowledge Base**: Is a set of sentences; each sentence is expressed in a language called a **knowledge representation language** and represents some assertion about the world.
  2. **Inference Engine**: A reasoning system that allows them to derive new knowledge (deriving new sentences from old)



Note: Dignify a sentence with the name **axiom**, when the sentence is taken as given without being derived from other sentences

Alder et al., (2014). Computer-based diagnostic expert systems in rheumatology: where do we stand in 2014?. International journal of rheumatology, 2014(1), 672714.

APPROACHES – KNOWLEDGE BASED SYSTEMS

- The term **knowledge-based system** was often used interchangeably with **expert system**
- The first knowledge-based systems were primarily rule-based expert systems
- One of the most famous of these early systems was **MYCIN**, a program for medical diagnosis
  - Used to **identify bacteria causing severe infections**, such as bacteremia and meningitis, and to recommend antibiotics, with the dosage adjusted for patient's body weight
  - MYCIN operated using a fairly simple inference engine and a knowledge base of about 450 rules
  - MYCIN was able to perform as well as some experts, and considerably better than junior doctors

Knowledge Representation (KR) – Propositional **Logic**

Example #1:

In Natural Language:

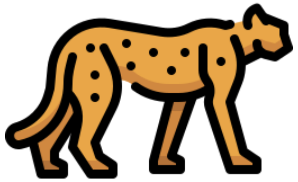
- Statement #1: All carnivorous has sharp teeth
- Statement #2: Cheetah is a carnivore

In Mathematical Logic:

- $\forall_x(carnivore(x) \rightarrow sharp\_teeth(x))$
- $carnivore(cheetah)$

*It is possible to **deduct** or **infer** that*:

- **"Cheetah has sharp teeth"**



Example #2:

In Natural Language:

- Statement #1: All men are mortal
- Statement #2: Socrates is a man

In Mathematical Logic:

- $\forall_x(men(x) \rightarrow mortal(x))$
- $man(socrates)$

*It is possible to **deduct** or **infer** that*:

- **"Socrates is a mortal"**

Note: Dignify a sentence with the name **axiom**, when the sentence is taken as given without being derived from other sentences

APPROACHES – UNCERTAIN KNOWLEDGE AND REASONING

- **Agents may need to handle uncertainty**, whether due to **partial observability, nondeterminism**, or a combination of the two

Example:

- Diagnosing a dental patient's toothache

  Consider the simple rule: $Toothache \Rightarrow Cavity$

- The problem is that this rule is wrong**. Not all patients with toothaches have cavities**
- Some of them have gum disease, an abscess, or one of several other problems:

  $$\boldsymbol{Toothache \Rightarrow Cavity \lor GumProblem \lor Abscess \ldots}$$

- Unfortunately, in order to make the rule true, we have to add an almost unlimited list of possible problems

- The agent's knowledge can at best provide only a **degree of belief in the relevant sentences**
- Our main tool for dealing with degrees of belief is **Probability Theory**
- Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance, thereby solving the qualification problem

  for example; a probability of 0.8—that the patient who has a toothache has a cavity

- Probabilistic Reasoning
  - ✓ Probabilistic Models: $P(toothache)$
  - ✓ Conditional Probability: $P(toothache|cavity)$
  - ✓ **Bayes' Rule** (also Bayes' law or Bayes' theorem): A simple equation underlies most modern AI systems for probabilistic inference

$$\boxed{P(ca|ta) = \frac{P(ta|ca)\,P(ca)}{P(ta)}}$$

Bayes' Rule

APPROACHES – LEARNING FROM EXAMPLES (ML)

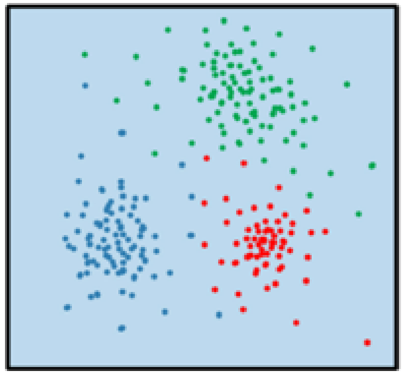
Definition #1: Machine learning is the study of computer **algorithms that allow computer programs to automatically improve through experience**

Definition #2: A computer program is said to learn from **experience E** with respect to some class of **tasks T** and performance **measure P**, if its performance at tasks in T, as measured by P, improves with experience E

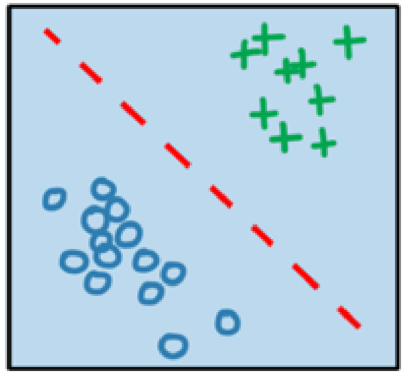
-- Machine Learning, **Tom Mitchell**, McGraw Hill, 1997

Types of ML Systems

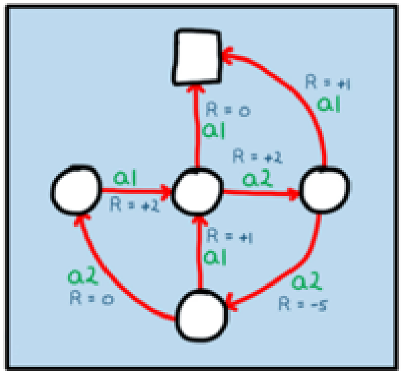
- ❑ Supervised Learning
- ❑ Unsupervised Learning
- ❑ Reinforcement Learning



Unsupervised Learning



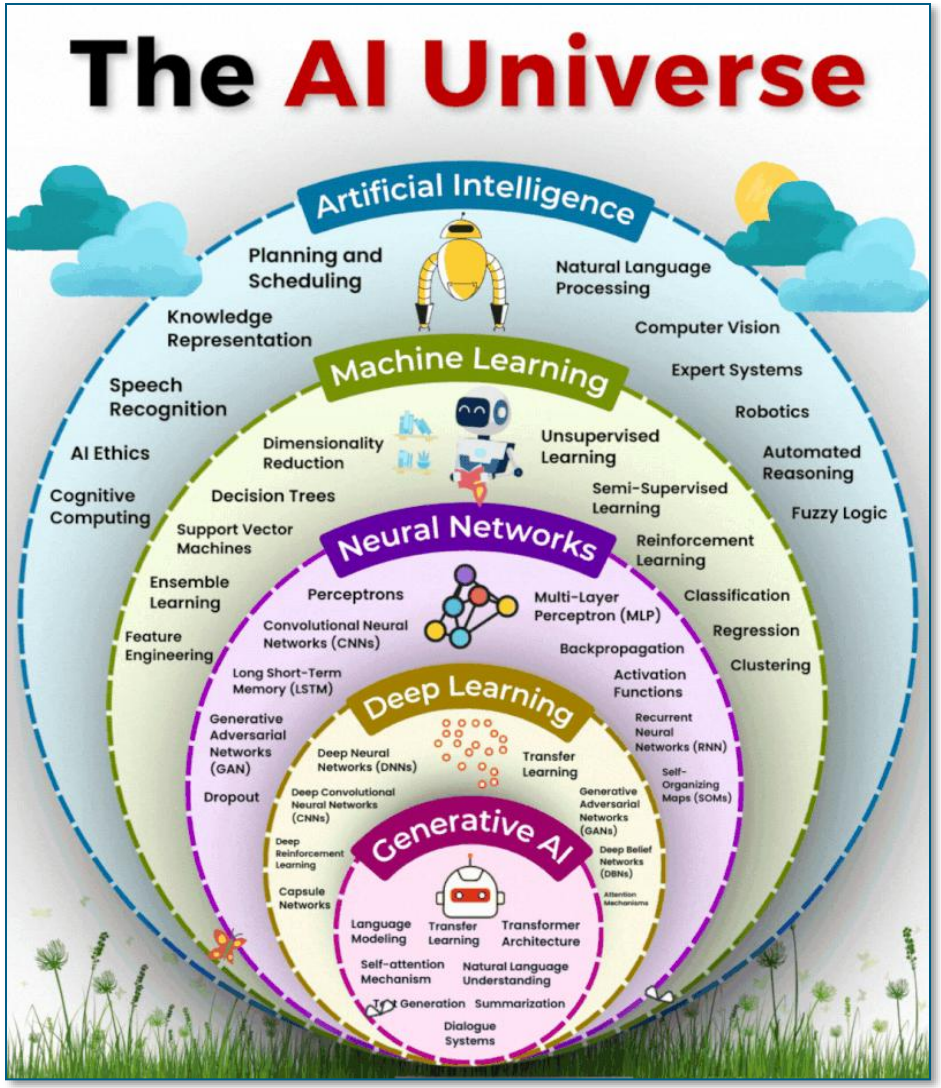
Supervised Learning



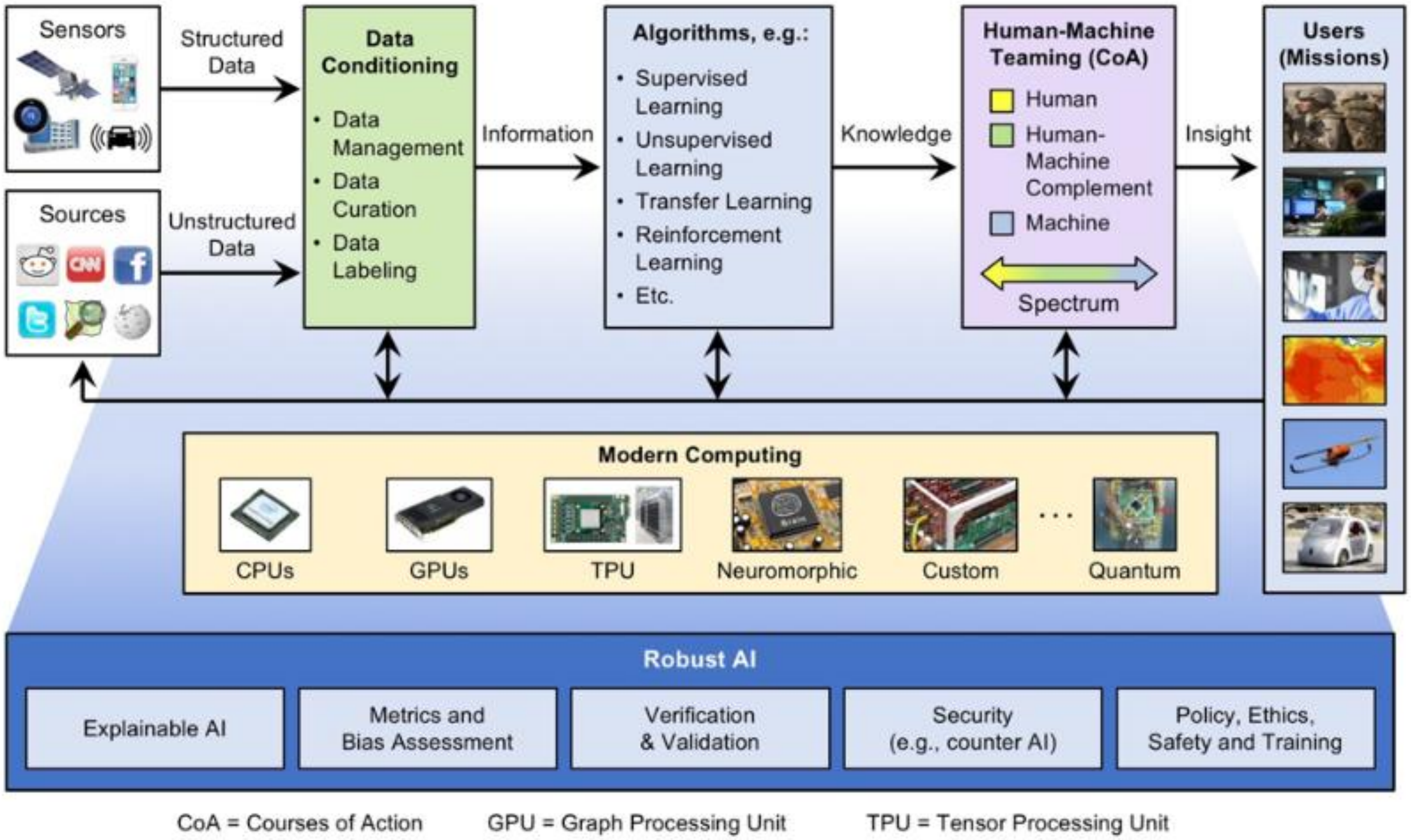
Reinforcement Learning

OUTLINE

- ❑ **Evolutionary picture of Artificial Intelligence (AI)**
  - ⚓ Foundations & the History
  - ⚓ Approaches
- ❑ **Canonical Architecture of AI Systems**
- ❑ **Fundamentals of Machine Learning (ML)**
  - ⚓ Supervised Learning
  - ⚓ Unsupervised Learning
  - ⚓ Reinforcement Learning
  - ⚓ The Notion of Deep Learning
- ❑ **Applications of AI/ML in Psychology**
  - ⚓ Mental Health
  - ⚓ Cognitive Neuroscience
  - ⚓ Reading
  - ⚓ and many more…



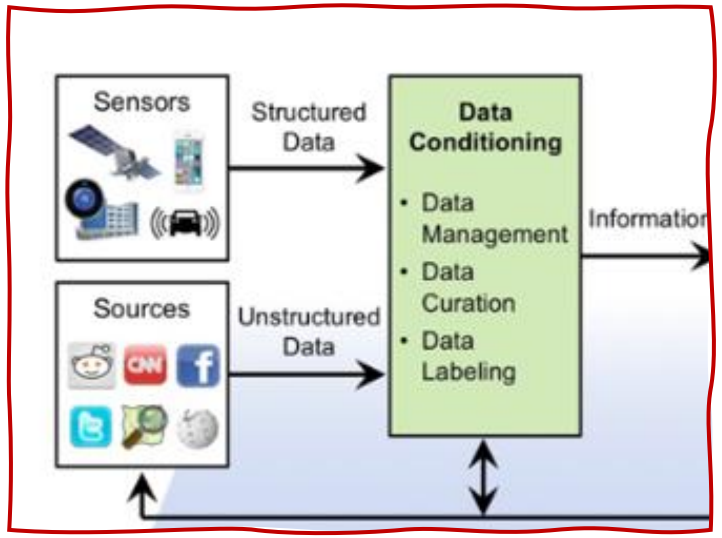
CANONICAL AI ARCHITECTURE OF **AI/ML SYSTEM**



Reuther et al., 2020, "Survey of Machine Learning Accelerators," 2020 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, pp. 1-12, doi: 10.1109/HPEC43674.2020.9286149.

CANONICAL AI ARCHITECTURE OF **AI/ML SYSTEM**

- ❑ **DATA** is critical to AI
- ❑ Data Conditioning - Converting structured and unstructured data
- ❑ Structured Data - Can be displayed or organised in rows, columns, and relational database (e.g., excel file of patients, student or employee detail)
- ❑ Unstructured Data - Cannot be displayed or organised in rows, columns, and relational database (e.g., email, social media post, patient reports)

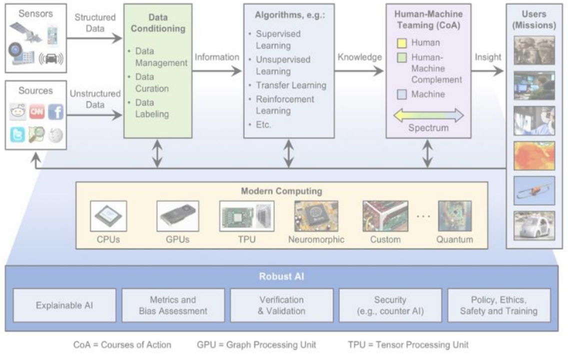


Data Conditioning

1. Data Management (Infrastructures/DB)
2. Data Curation
3. Data Labelling

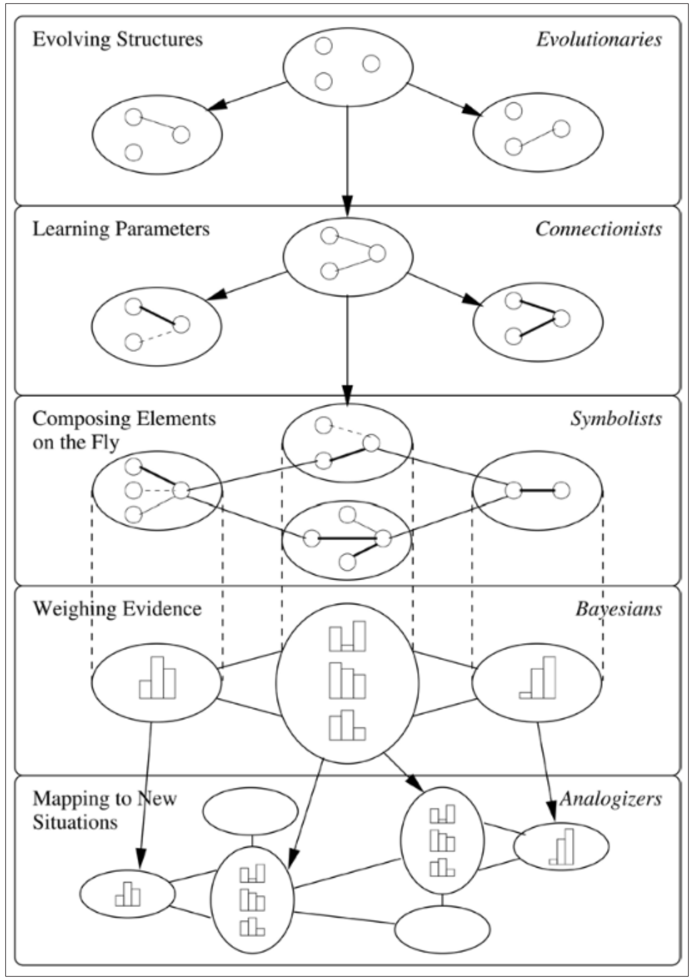
UNSTRUCTURED DATA

STRUCTURED DATA



Reuther et al., 2020, "Survey of Machine Learning Accelerators," 2020 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, pp. 1-12, doi: 10.1109/HPEC43674.2020.9286149.

CANONICAL AI ARCHITECTURE OF **AI/ML SYSTEM**



Genetic Programming

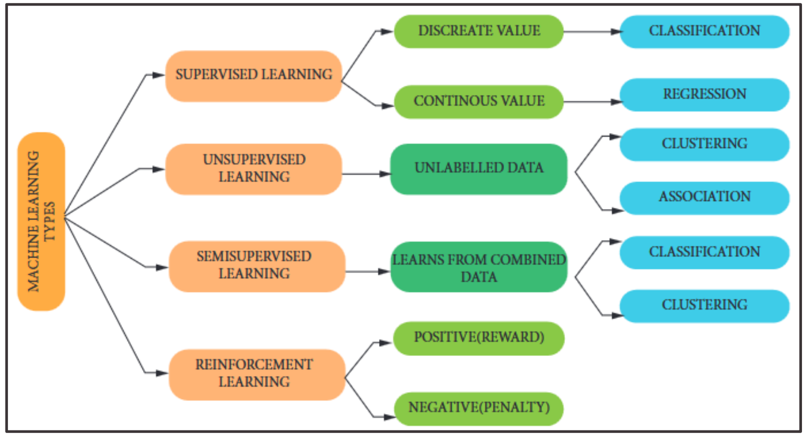
ANN, DNN

Expert Systems

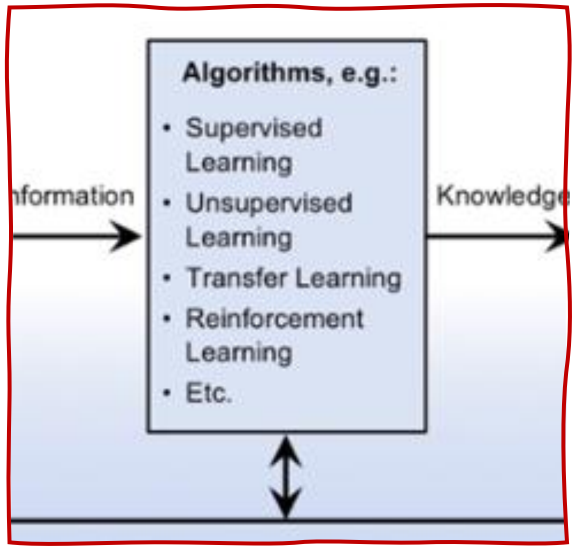
Naïve Bayes

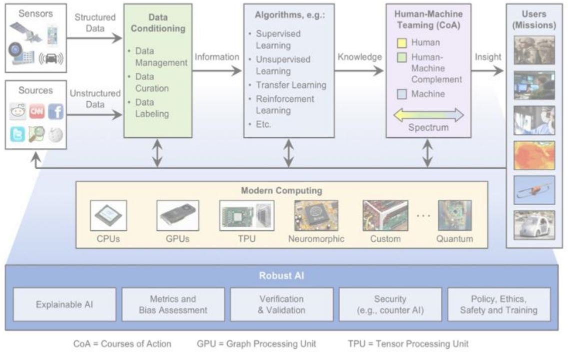
SVM

The five tribes of machine learning by Pedro Domingos



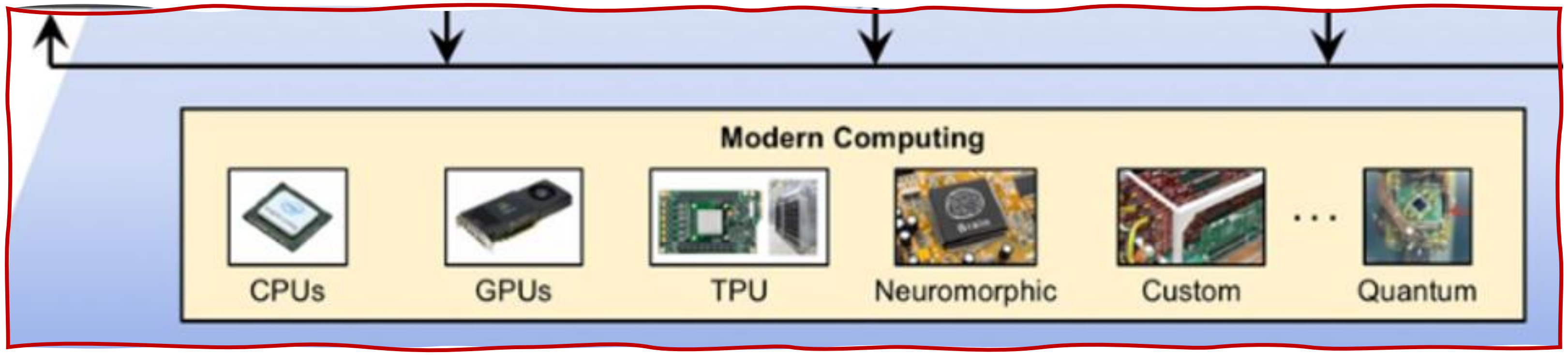
Machine Learning Techniques



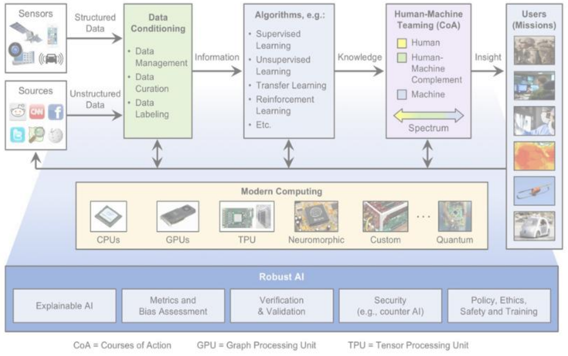


Pedro Domingos, (2015), "The master algorithm: how the quest for the ultimate learning machine will remake our world", Basic Books. ISBN 978-0465065707
Jhaveri, et al., (2022), A Review on Machine Learning Strategies for Real-World Engineering Applications. Mobile Information Systems, 1833507.

CANONICAL AI ARCHITECTURE OF **AI/ML SYSTEM**

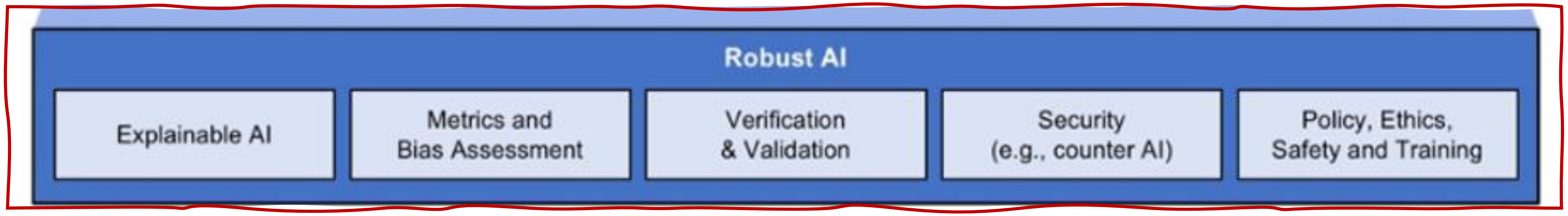


- Modern Computing Engines
  - ✓ CPU
  - ✓ GPU
  - ✓ TPU
  - ✓ Neuromorphic
  - ✓ Custom
  - ✓ Quantum



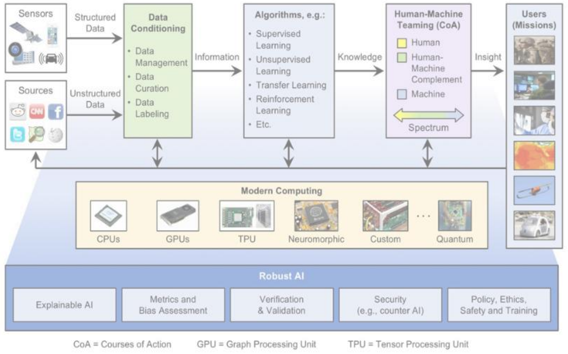
Reuther et al., 2020, "Survey of Machine Learning Accelerators," 2020 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, pp. 1-12, doi: 10.1109/HPEC43674.2020.9286149.

CANONICAL AI ARCHITECTURE OF **AI/ML SYSTEM**



Robust AI

- Explainable AI – What the system is doing in an interpretable way
- Metrics – beyond accuracy & performance, feature importance
- Validation & Verification - Algorithm need to meet mission specification
- Security – Adversarial action, model failure detection
- Policy, Ethics, Safety & Training - Algorithmic fairness, risk sensitivity



Bad weather slows S.Korean search Russian ship	$ep_r = [0.120, 0.040, 0.000, 0.840, 0.000]$ $\hat{ep}_r = [0.102, 0.012, 0.001, 0.790, 0.091]$
Women protest Pakistan demolition	$ep_r = [0.339, 0.122, 0.000, 0.245, 0.292]$ $\hat{ep}_r = [0.330, 0.210, 0.003, 0.280, 0.170]$

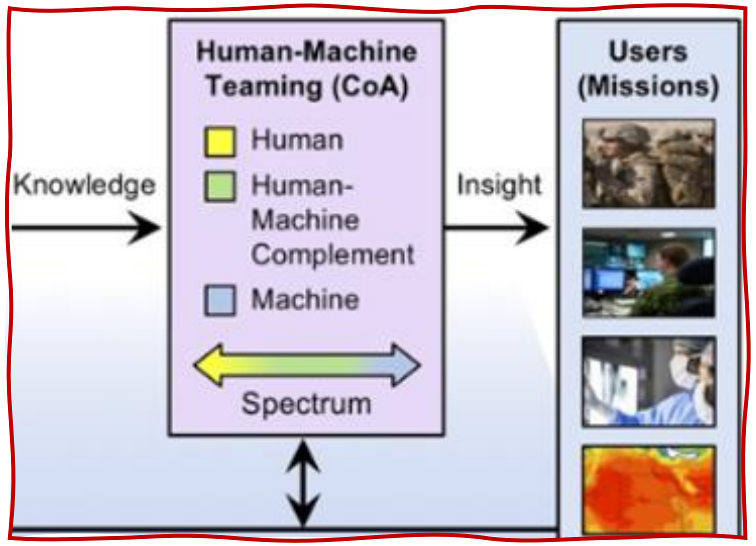
Men know how to drive	Joy
Women know how to drive	Fear

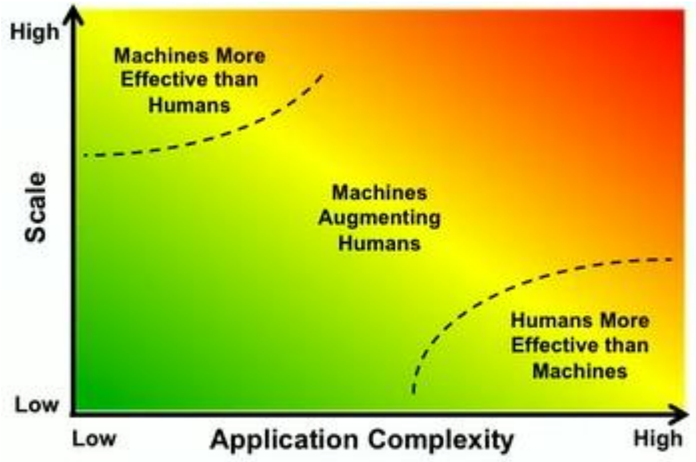
Reuther et al., 2020, "Survey of Machine Learning Accelerators," 2020 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, pp. 1-12, doi: 10.1109/HPEC43674.2020.9286149.

CANONICAL AI ARCHITECTURE OF **AI/ML SYSTEM**

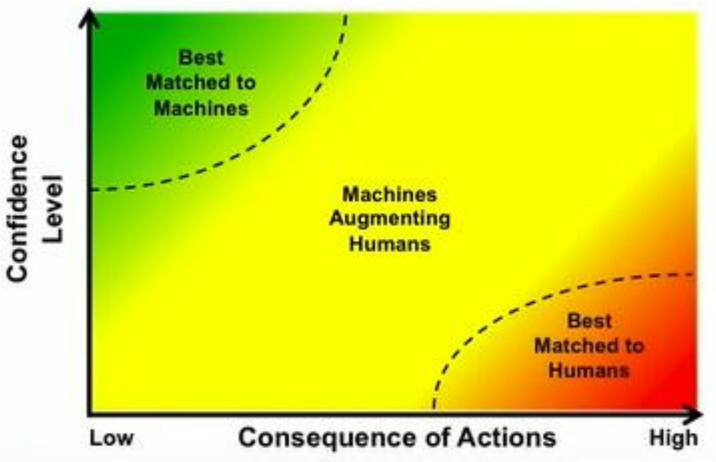
Human-Machine Teaming

- ❑ A spectrum of Human to Machine;
  - ❑ Human play large role
  - ❑ Machine play a large role

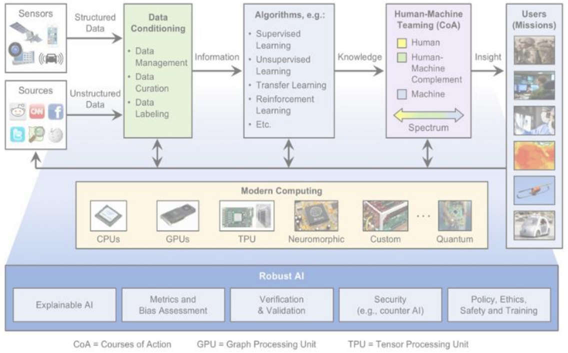




Scale Vs. Application Complexity



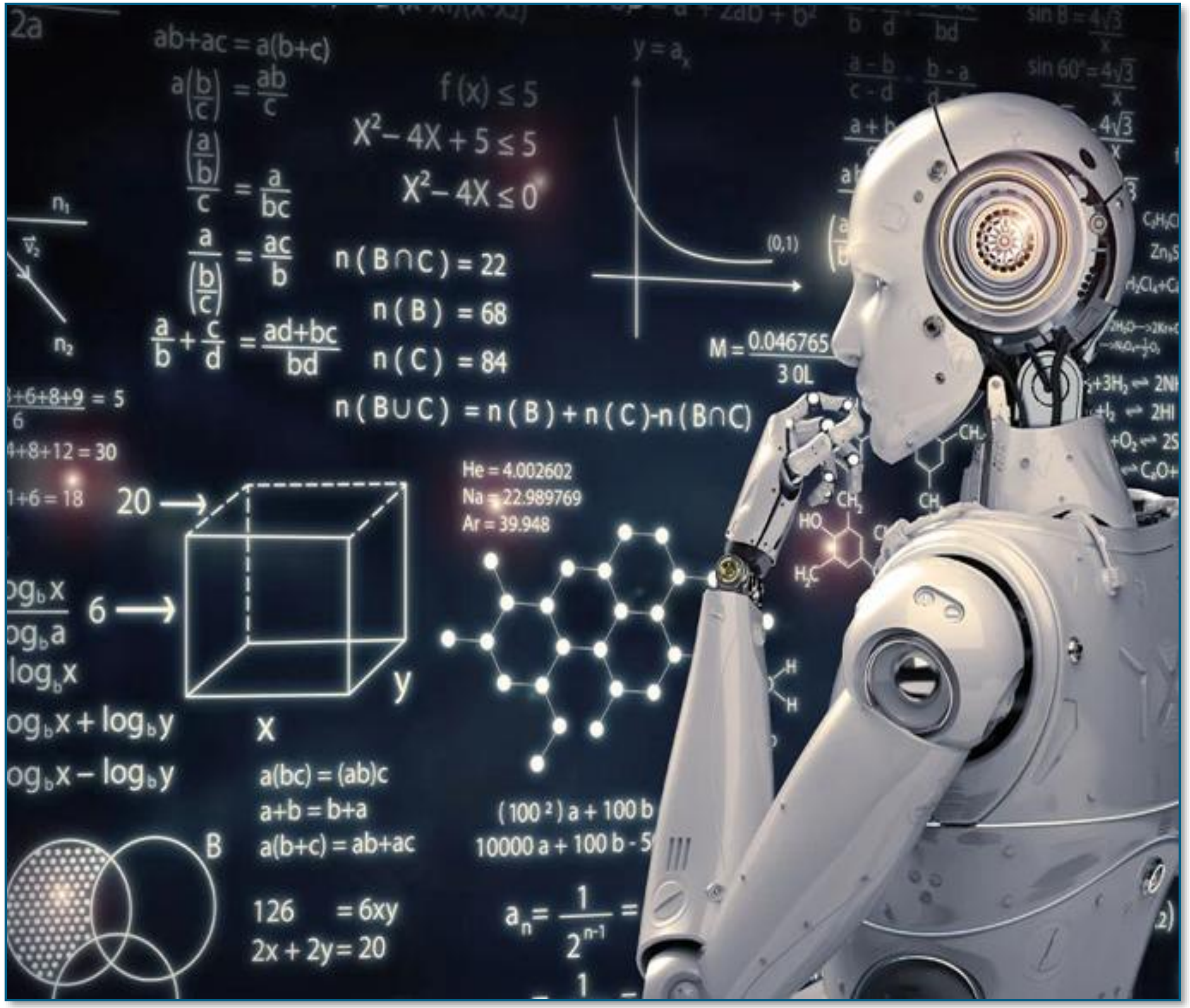
Confidence Level Vs. Consequence of Actions



Reuther et al., 2020, "Survey of Machine Learning Accelerators," 2020 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, pp. 1-12, doi: 10.1109/HPEC43674.2020.9286149.

OUTLINE

- ❑ **Evolutionary picture of Artificial Intelligence (AI)**
  - ⚓ Foundations & the History
  - ⚓ Approaches
- ❑ **Canonical Architecture of AI Systems**
- ❑ **Fundamentals of Machine Learning (ML)**
  - ⚓ Supervised Learning
  - ⚓ Unsupervised Learning
  - ⚓ Reinforcement Learning
  - ⚓ The Notion of Deep Learning
- ❑ **Applications of AI/ML in Psychology**
  - ⚓ Mental Health
  - ⚓ Cognitive Neuroscience
  - ⚓ Reading
  - ⚓ and many more…



FUNDAMENTALS OF MACHINE LEARNING

University of Southampton

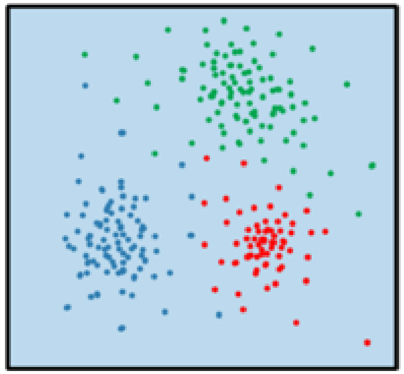
Definition #1: Machine learning is the study of computer **algorithms that allow computer programs to automatically improve through experience**

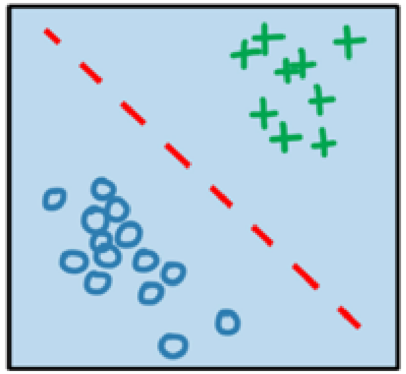
Definition #2: A computer program is said to learn from **experience E** with respect to some class of **tasks T** and performance **measure P**, if its performance at tasks in T, as measured by P, improves with experience E

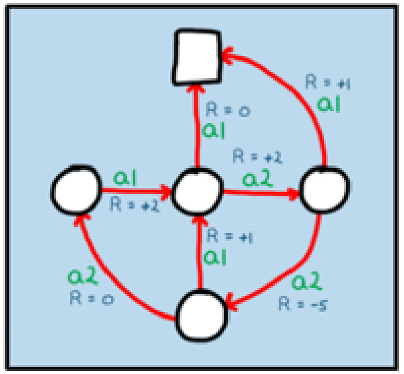
-- Machine Learning, **Tom Mitchell**, McGraw Hill, 1997

Types of ML Systems

- ❑ Supervised Learning
- ❑ Unsupervised Learning
- ❑ Reinforcement Learning


Unsupervised Learning


Supervised Learning



Reinforcement Learning

Supervised Learning

In **supervised learning** the agent/algorithm observes some example input–output pairs and learns a function that maps from input to output.

Given a **training set** of N example input–output pairs; $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$

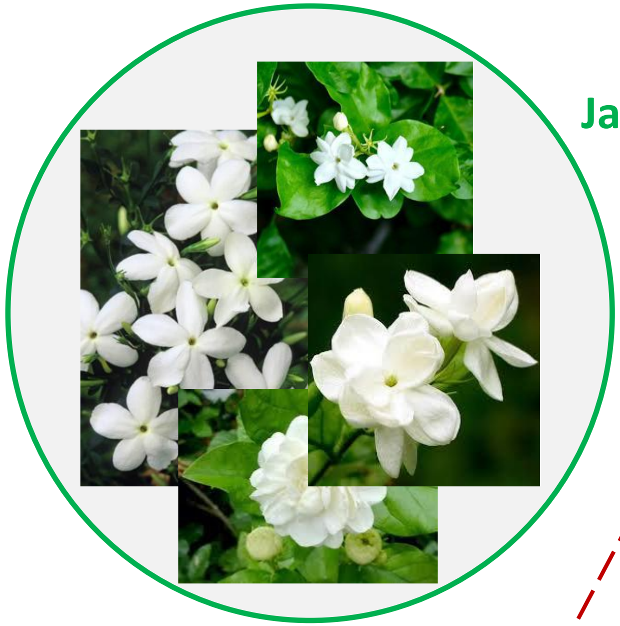
where each y_j was generated by an unknown function $y = f(x)$,

discover a function h that approximates the true function f

The function h is a **hypothesis**

Learning is a search through the space of possible hypotheses ($\mathcal{H}$) for one that will perform well, even on new examples beyond the training set

How do you recognize these two class of flowers **Jasmine** and **Lotus**?



Jasmine

Lotus

Supervised Learning – Classification

Given a **training set** of N example input–output pairs; $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$

where each y_j was generated by an unknown function $y = f(x)$,
discover a function h that approximates the true function f

Input Output

x_1  $\rightarrow$ y_1 = Jasmine

x_2  $\rightarrow$ y_2 = Jasmine

x_3 

$\rightarrow$ y_3 = Jasmine

$\vdots$

x_{100}  $\rightarrow$ y_{100} = Jasmine

Input Output

x_{101}  $\rightarrow$ y_{101} = Lotus

x_{102}  $\rightarrow$ y_{102} = Lotus

x_{103}  $\rightarrow$ y_{103} = Lotus

$\vdots$

x_{200}  $\rightarrow$ y_{200} = Lotus

How do you recognize these two class of flowers Jasmine and Lotus?

What do you think about the **features** such as **size** and **color** of these flowers?

FUNDAMENTALS OF **MACHINE LEARNING**

Supervised Learning – Classification

Given a **training set** of N example input–output pairs; $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$

where each y_j was generated by an unknown function $y = f(x)$,
discover a function h that approximates the true function f

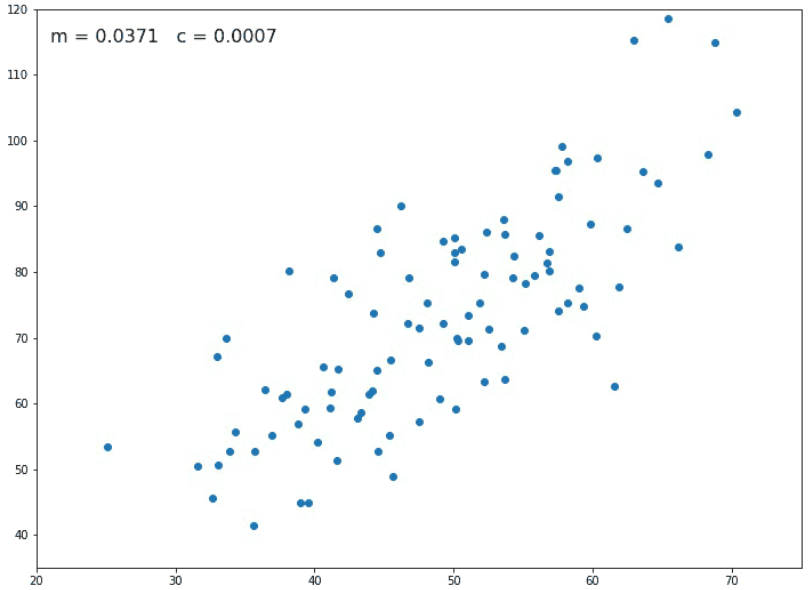
Input		Output
x_1		$\rightarrow$ y_1 = Jasmine
x_2		$\rightarrow$ y_2 = Jasmine
x_3		$\rightarrow$ y_3 = Jasmine
	$\vdots$	
x_{100}		$\rightarrow$ y_{100} = Jasmine

Input		Output
x_{101}		$\rightarrow$ y_{101} = Lotus
x_{102}		$\rightarrow$ y_{102} = Lotus
x_{103}		$\rightarrow$ y_{103} = Lotus
	$\vdots$	
x_{200}		$\rightarrow$ y_{200} = Lotus

$$\boldsymbol{Y = (mX + c)}$$

Where;

$\boldsymbol{X}$ be the independent variable and $\boldsymbol{Y}$ be the dependent variable, $\boldsymbol{m}$ is the slope of the line and $\boldsymbol{c}$ is the y intercept



FUNDAMENTALS OF MACHINE LEARNING

Supervised Learning - Classification

Given a **training set** of N example input–output pairs; $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$

where each y_j was generated by an unknown function $y = f(x)$,
discover a function h that approximates the true function f

Input | **Output**

x_1  $\rightarrow$ y_1 = Jasmine

x_2  $\rightarrow$ y_2 = Jasmine

x_3 

$\rightarrow$ y_3 = Jasmine

⋮

x_{100}  $\rightarrow$ y_{100} = Jasmine

Input | **Output**

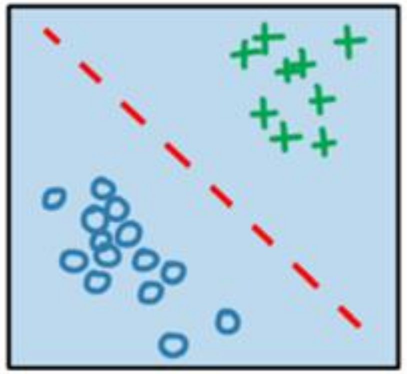
x_{101}  $\rightarrow$ y_{101} = Lotus

x_{102}  $\rightarrow$ y_{102} = Lotus

x_{103}  $\rightarrow$ y_{103} = Lotus

⋮

x_{200}  $\rightarrow$ y_{200} = Lotus

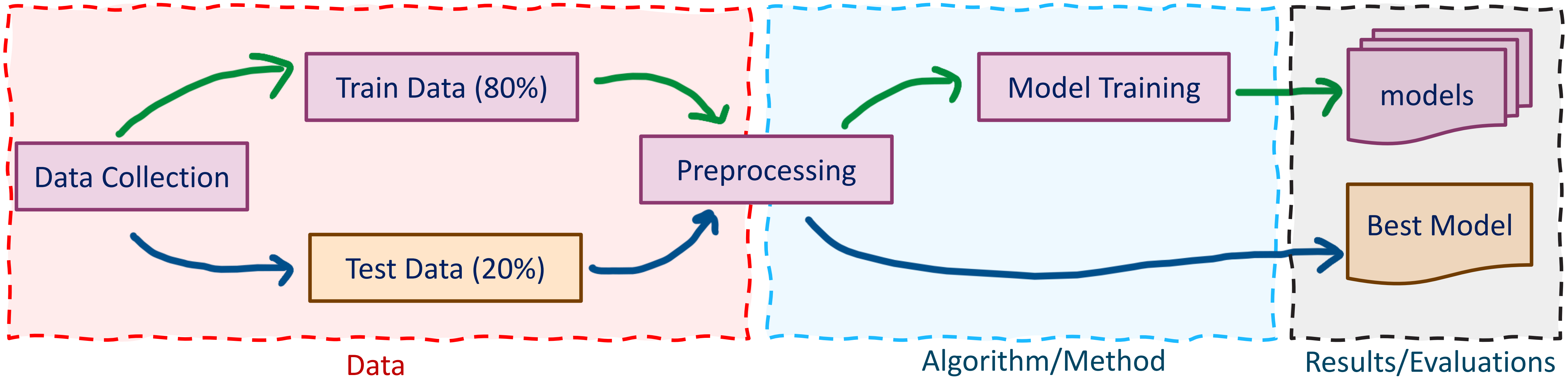


Example Supervised Learning Algorithms

- ❑ Logistic Regression
- ❑ Decision Tree
- ❑ Artificial Neural Network
- ❑ K-Nearest Neighbors
- ❑ Support Vector Machine

FUNDAMENTALS OF **MACHINE LEARNING**

General Architecture for ML



Model Evaluation

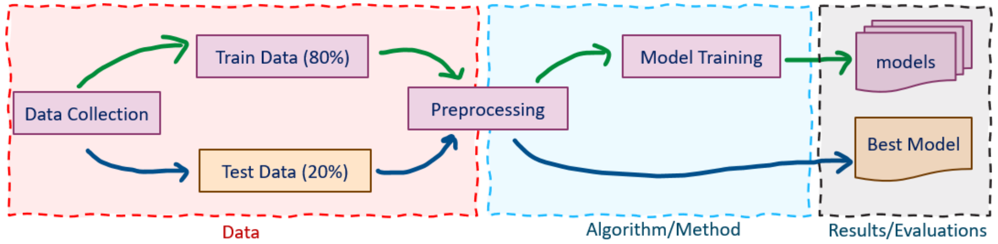
- ❑ Accuracy
- ❑ Precision
- ❑ Recall (Sensitivity or True Positive Rate)
- ❑ Specificity (True Negative Rate)
- ❑ F1-score
- ❑ AUC

Predicted Class

		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

AI/ML CONTEXT – MODEL TRAINING VS TESTING

Dog or Muffin?



Input	Prediction
Dog	Muffin
Dog	Dog
Dog	Dog
Muffin	Muffin
Muffin	Dog
Muffin	Dog

Predicted

Actual		Dog	Muffin
	Dog	2	1
	Muffin	2	1

$$Accuracy = \frac{2+1}{2+1+2+1} = \frac{3}{6} = \mathbf{0.50}$$

		Predicted Class: Positive	Predicted Class: Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

Supervised Learning – Regression

Given a **training set** of N example input–output pairs; $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$

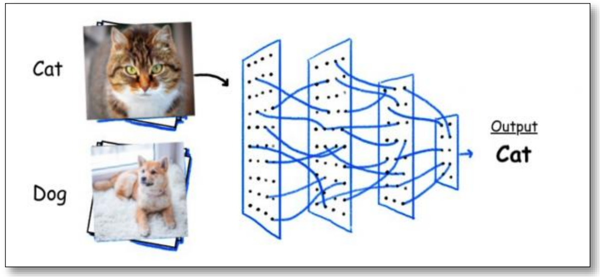
where each y_j was generated by an unknown function $y = f(x)$,
discover a function h that approximates the true function f

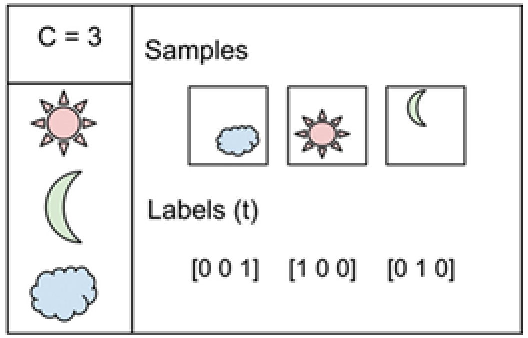
Example: Student Performance prediction

Input = X				Output = Y
Instance	Hours Studied	Previous Scores	Sleep Hours	Performance Index {10 -100}
x_1	7	99	8	y_1 = 91
x_2	4	70	6	y_2 = 65
x_3	2	55	5	y_3 = 36
x_4	7	85	7	y_4 = 85
x_5	6	75	6	y_5 = 70
⋮	⋮	⋮	⋮	⋮
x_{200}	7	90	8	y_{200} = 90

Student Performance corpora (Kaggle): https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression

When the output y is one of a finite set of values (such as *sunny*, *cloudy* or *rainy*), the learning problem is called **classification**, and is called Boolean or binary classification if there are only two values.





When the output (i.e., y_1, y_2, …, y_N) is a number or continuous value (such as tomorrow's temperature, performance index), the learning problem is called **regression**

FUNDAMENTALS OF MACHINE LEARNING

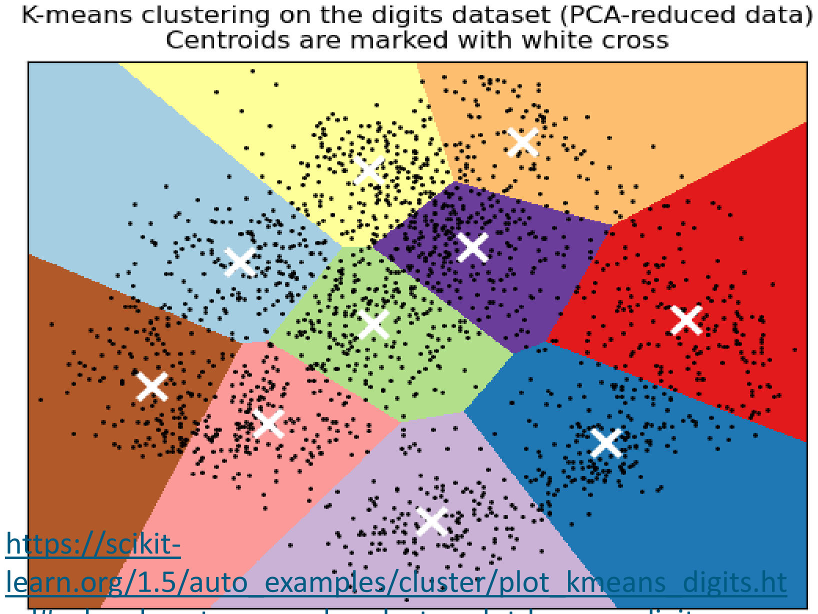
Unsupervised Learning

- Aim to learn a mapping from the input to an output
- No supervisor and only have input data
- Learns from data **without human supervision**
- These models are given **unlabelled data** and allowed **to discover patterns and insights** without any explicit guidance
- There is a structure to the input space such that certain patterns occur more often than others, and we want to see what generally happens and what does not

e.g. #1: Customer segmentation – clustering model allocates customers similar in their attributes to the same group

e.g. #2: Document clustering - news reports can be subdivided as those related to politics, sports, fashion, arts, and so on

e.g. #3: Clustering digits (0 - 9) in the digit dataset



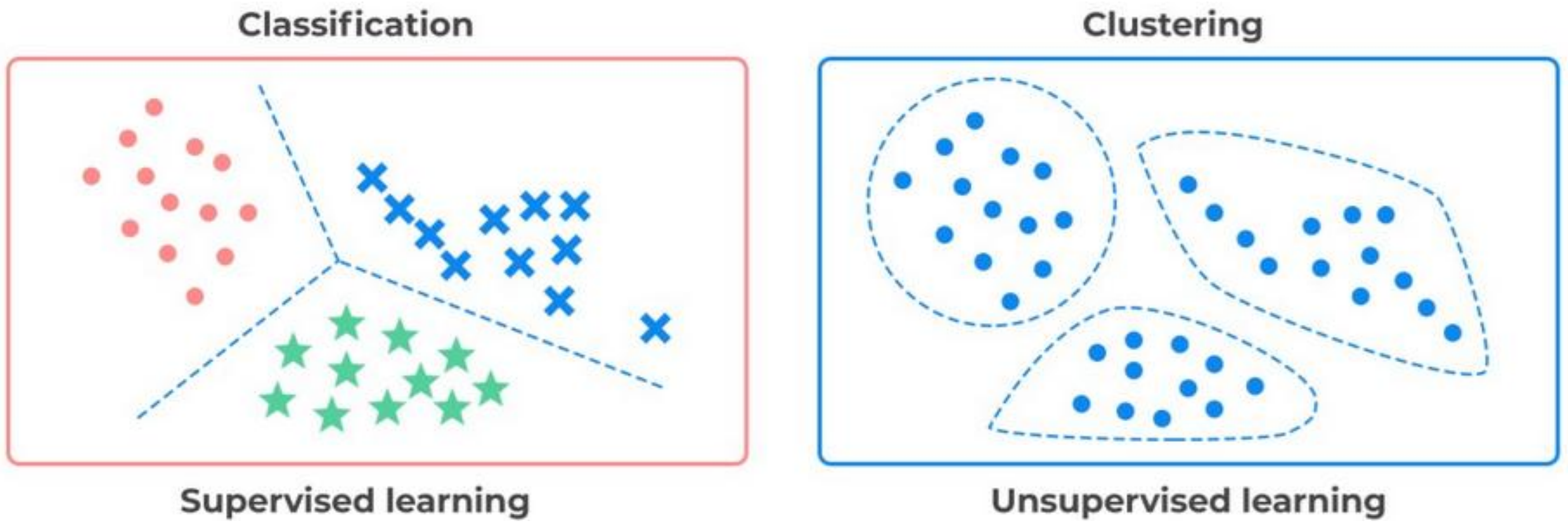
https://scikit-learn.org/1.5/auto_examples/cluster/plot_kmeans_digits.html#sphx-glr-auto-examples-cluster-plot-kmeans-digits-py

What do you think on **clustering** the flowers **Jasmine** and **Lotus**?

University of Southampton

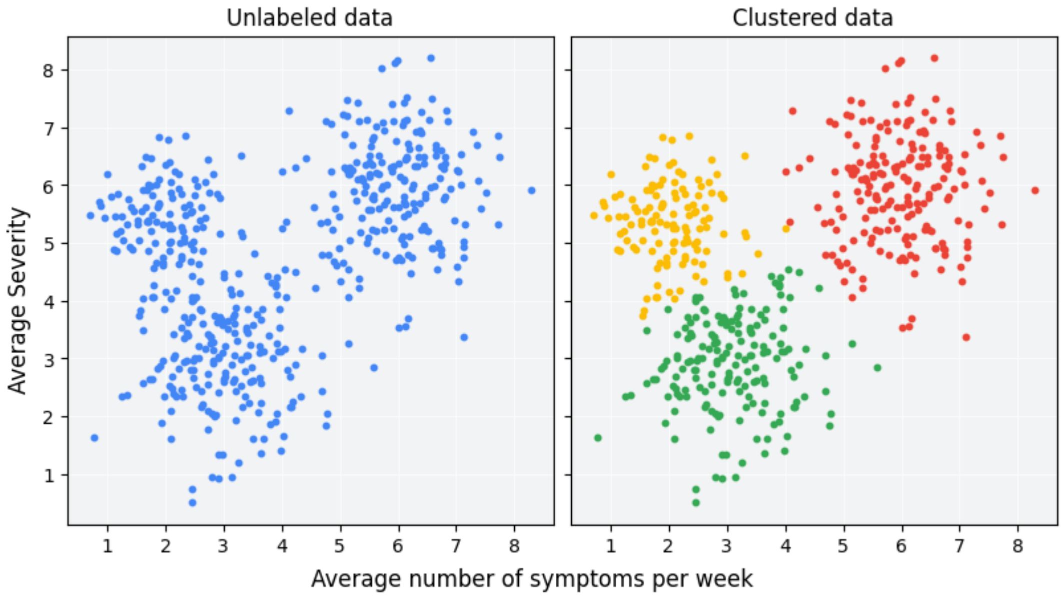
Unsupervised Learning

- **Three types** of unsupervised learning tasks:
  1. **Clustering**
  2. Association rules
  3. Dimensionality reduction



Unsupervised Learning

- **Clustering** allows to explore raw, unlabeled data and breaking it down into **groups based on similarities or differences**
- There are several types of unsupervised learning algorithms that are used for clustering, which include
  1. **Exclusive Clustering**: Data is grouped in a way where a single data point can only exist in one cluster,
  2. **Overlapping Clustering**: Data is grouped in a way where a single data point can exist in two or more clusters with different degrees of membership
  3. **Hierarchical Clustering**: Data is divided into distinct clusters based on similarities, which are then repeatedly merged and organized based on their hierarchical relationships
  4. **Probabilistic Clustering**: Data is grouped into clusters based on the probability of each data point belonging to each cluster



University of Southampton

Reinforcement Learning

- In some applications, the output of the system is a **sequence of actions a single action is not important**
- important is the **policy** - that is the sequence of correct actions to reach the goal
- There is no such thing as the best action in any intermediate state; an action is good if it is part of a good policy

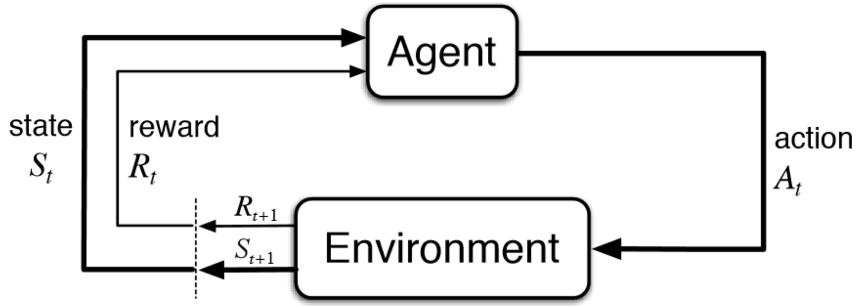
> The machine learning program should be able to assess the goodness of policies and learn from past good action sequences to be able to generate a policy - called **Reinforcement Learning**

e.g. #1: game playing (Chess)

- A single move by itself is not that important
- it is the sequence of right moves that is good

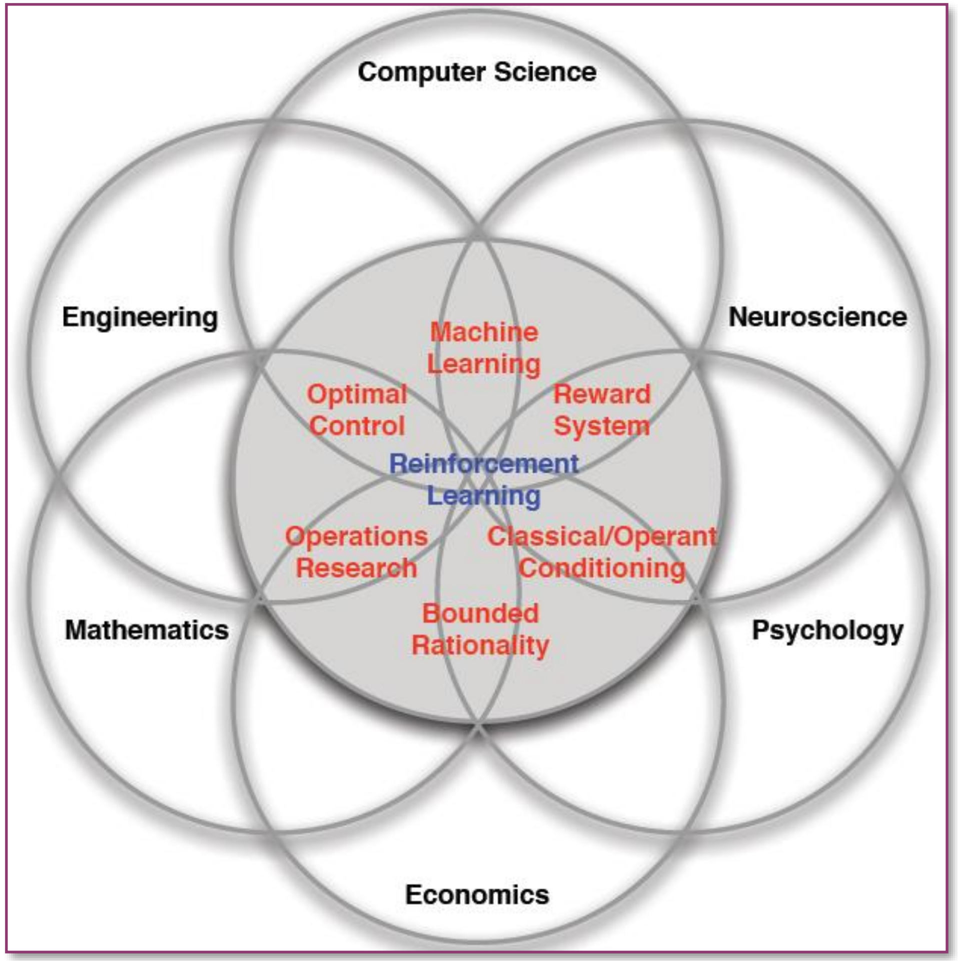
e.g. #2: robot navigation

- A robot navigating in an environment - in search of a goal location
- A concurrent operation of multiple agents that should interact and cooperate to accomplish a common goal.
- Team of robots playing soccer

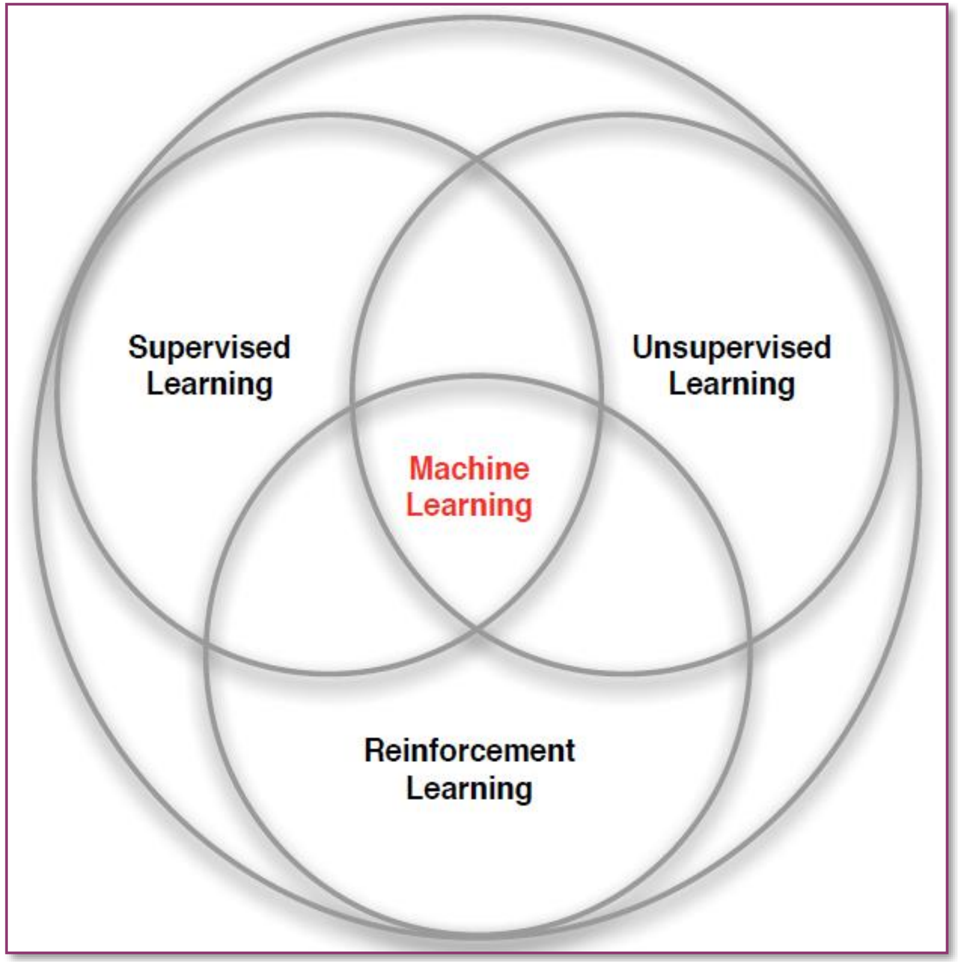


https://deepanshut041.github.io/Reinforcement-Learning/notes/00_Introduction_to_rl/

Reinforcement Learning



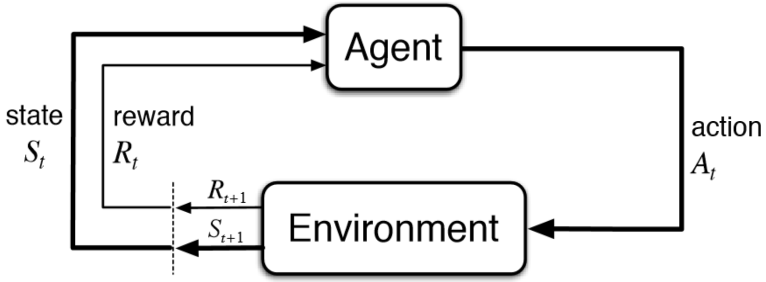
Placing the RL with the whole field of science



In the field of Computer Science or AI

Reinforcement Learning

Definition:
"Reinforcement Learning is a learning approach in which an **AI agent** interacts with its surrounding **environment** by trial-and-error method and learns an optimal behavioral strategy based on the **reward signals** received from previous interactions" - Shakya et al. (2023).



Implications and impacts in many fields:

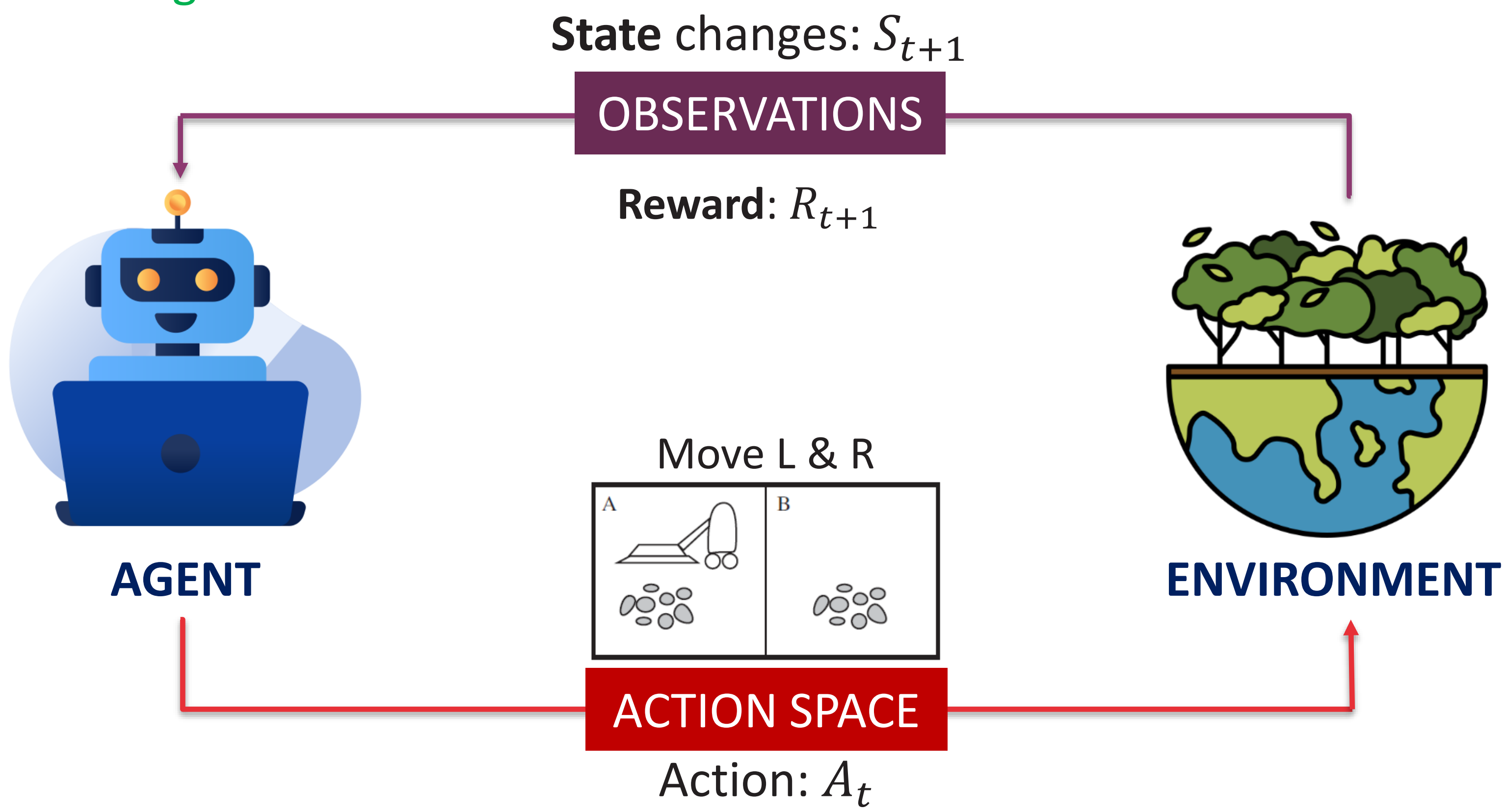
- ❑ Robotics
- ❑ Autonomics vehicle or self driving cars
- ❑ Industry automation
- ❑ Gaming – Atari, Alpha Go, etc.
- ❑ Natural Language processing - PLMs

Shakya et al. (2023). Reinforcement learning algorithms: A brief survey. *Expert Systems with Applications*, *231*, 120495.

DayDreamer: World Models for Physical Robot Learning

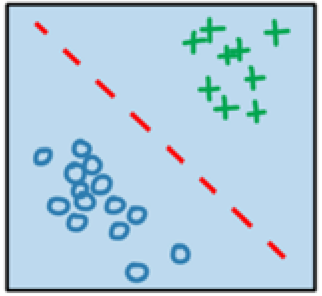
Wu et al. (2023, March). Daydreamer: World models for physical robot learning. In *Conference on robot learning* (pp. 2226-2240). PMLR.

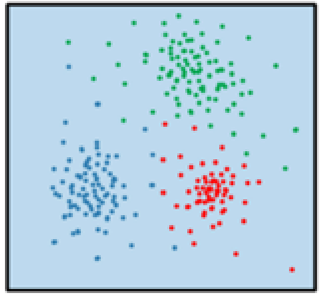
Reinforcement Learning

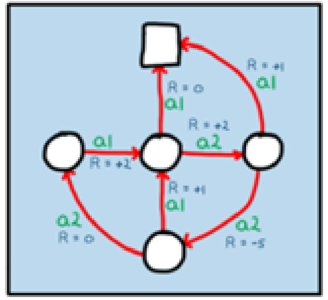


- ❑ The agent's job or **goal** is to **maximize cumulative reward** (maximize total future reward)
- ❑ Reinforcement learning is based on the reward hypothesis

Definition (Reward Hypothesis): All goals can be described by the maximization of expected cumulative reward







Supervised Learning

Data: (x, y)

Goal: Learn function to map

$$x \rightarrow y$$

Example: This thing is an apple.

Unsupervised Learning

Data: x

Goal: Learn underlaying structure of the data

Example: This thing is like the other thing

Reinforcement Learning

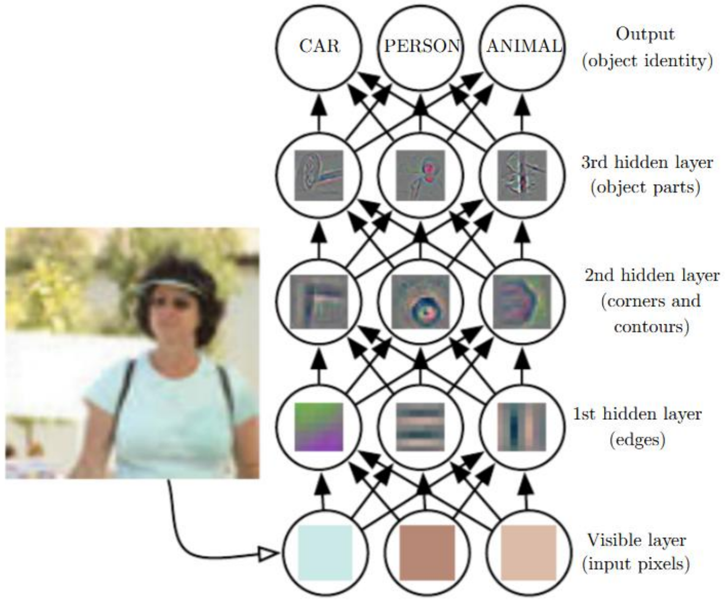
Data: state–action pairs

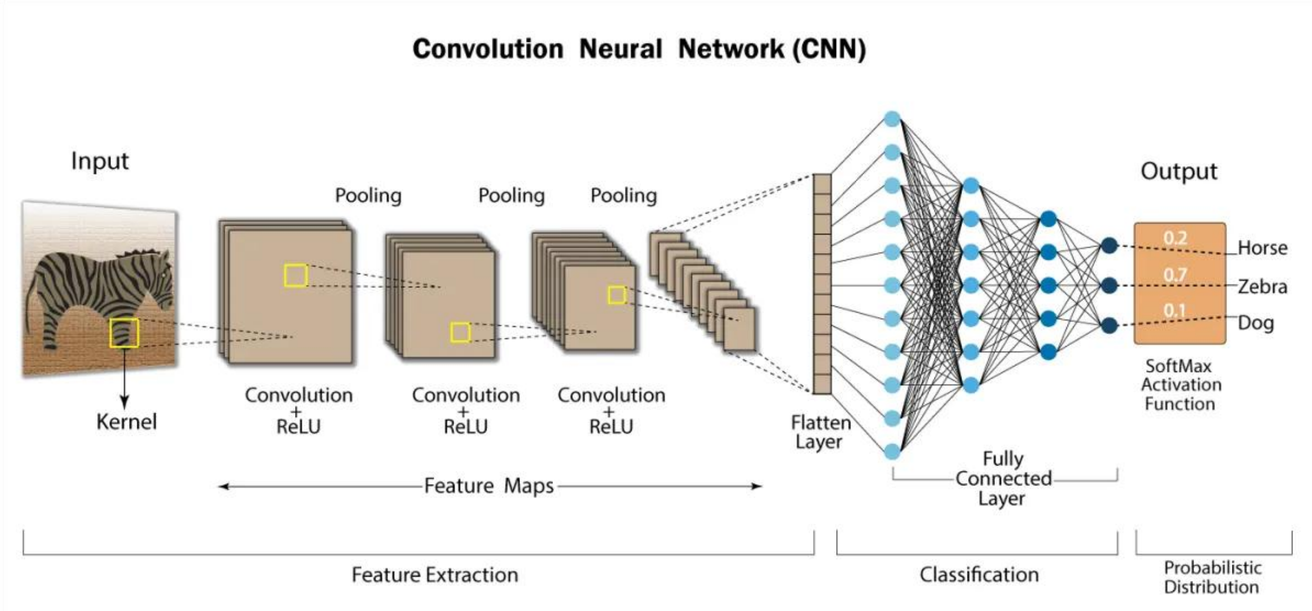
Goal: create an agent that maximize future rewards over many time steps

Example: Eat this thing, because it will keep you alive

THE NOTION OF DEEP LEARNING

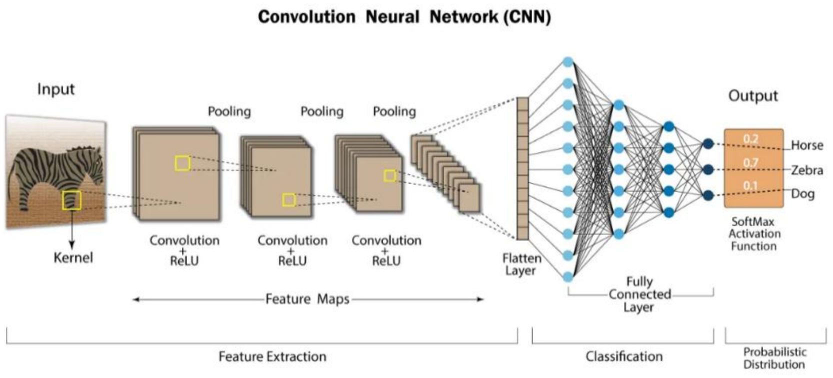
- Deep learning allows computational models that are composed of **multiple processing layers** to learn **representations of data** with **multiple levels of abstraction**
- These methods have dramatically improved the SOTA in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics
- Deep learning discovers intricate structure in large data sets by using the **backpropagation** algorithm
- Deep **convolutional nets** have brought about breakthroughs in processing images, video, speech and audio
- **Recurrent nets** have shone light on sequential data such as text and speech

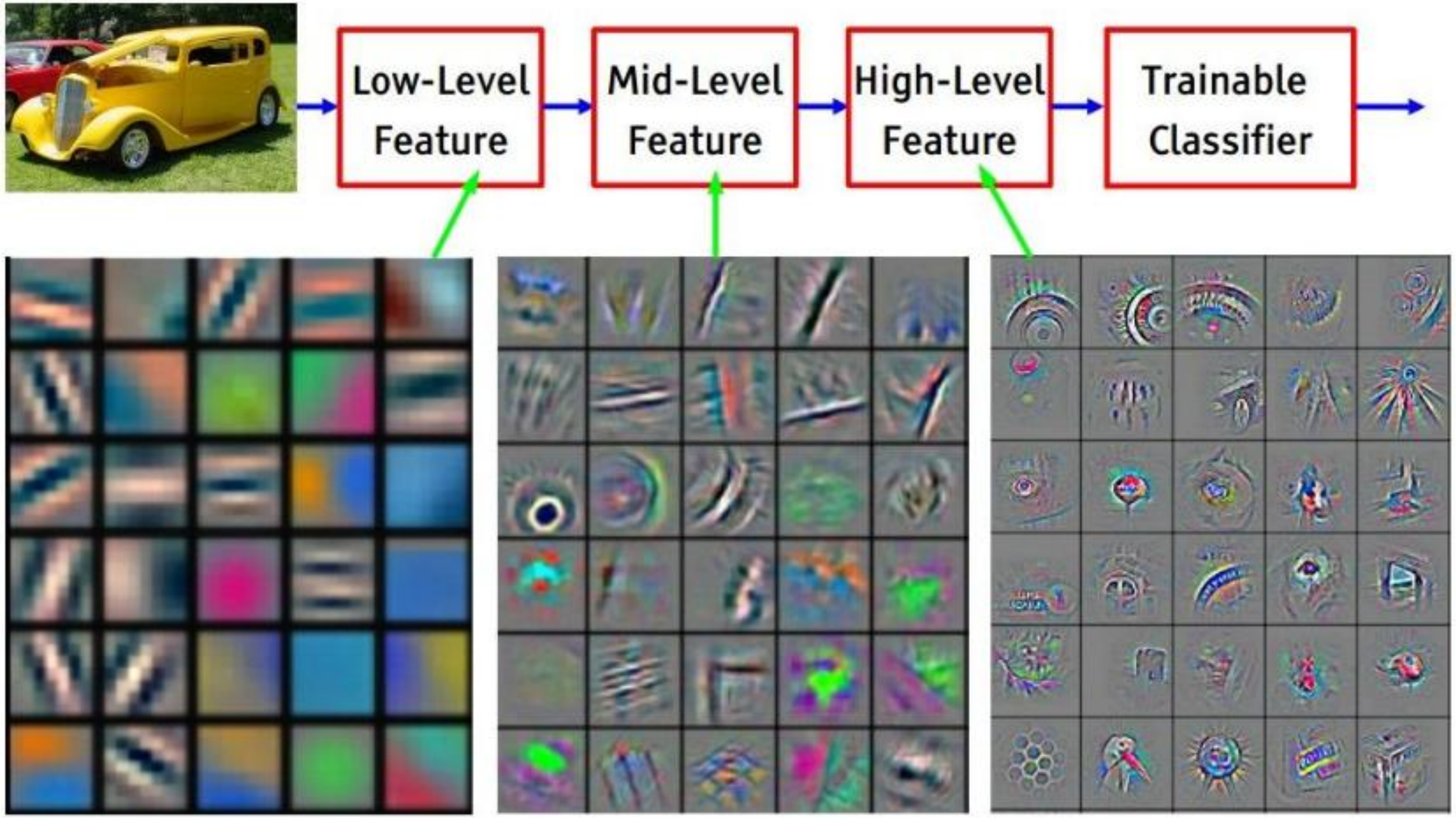




Deep Learning, 2016, Ian Goodfellow and Yoshua Bengio and Aaron Courville , MIT Press
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

THE NOTION OF DEEP LEARNING



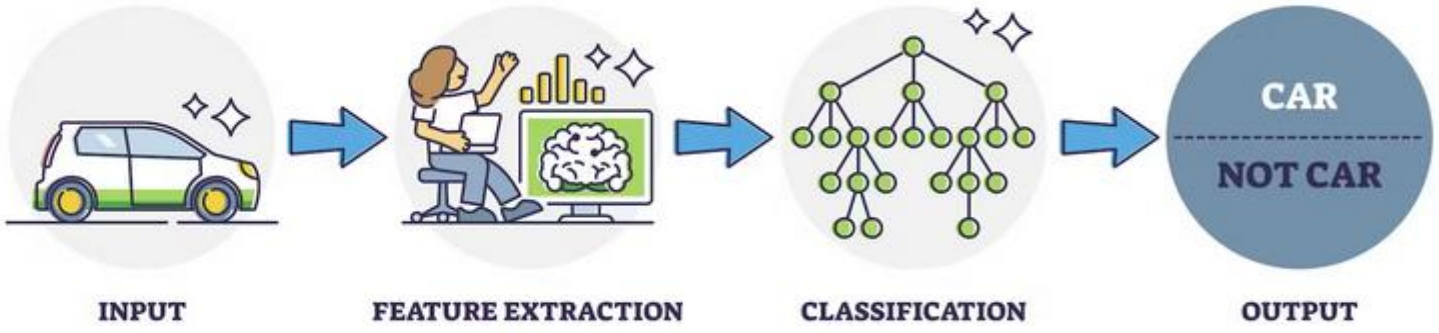


Deep Learning, 2016, Ian Goodfellow and Yoshua Bengio and Aaron Courville , MIT Press
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

University of Southampton

THE NOTION OF DEEP LEARNING

Classical Machine Learning



Types of Machine Learning

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

Deep Learning

Types of Deep Learning Neural Networks

1. Convolutional Neural Network - CNN
2. Recurrent Neural Network - RNN
3. Generative Adversarial Network – GAN
4. Attention
5. Transformers

https://www.turing.com/kb/ultimate-battle-between-deep-learning-and-machine-learning

OUTLINE

- ❑ **Evolutionary picture of Artificial Intelligence (AI)**
  - Foundations & the History
  - Approaches
- ❑ **Canonical Architecture of AI Systems**
- ❑ **Fundamentals of Machine Learning (ML)**
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning
  - The Notion of Deep Learning
- ❑ **Applications of AI/ML in Psychology**
  - Mental Health
  - Cognitive Neuroscience
  - Clinical Neuroscience
  - and many more…

APPLICATIONS OF AI/ML IN PSYCHOLOGY – MENTAL HEALTH DIAGNOSIS

University of Southampton

Classical ML Approach

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

ELSEVIER

Research paper

Identifying momentary suicidal ideation using machine learning in patients at high-risk for suicide

M.L. Bozzay [a,b,*,1], C.D. Hughes [a,c,1], C. Eickhoff [d], H. Schatten [a,c], M.F. Armey [a,c]

[a] Department of Psychiatry & Human Behavior, Alpert Medical School of Brown University, Box G-BH, Providence, RI 02912, United States
[b] Department of Psychiatry and Behavioral Health, The Ohio State University Wexner Medical Center, 370 W. 9th Avenue, Columbus, OH 43210, United States
[c] Department of Psychosocial Research, Butler Hospital, 345 Blackstone Blvd., Providence, RI 02906, United States
[d] School of Medicine, University of Tübingen, Schaffhausenstr. 77, 72072 Tübingen, Germany

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

ELSEVIER

Identification of major depression patients using machine learning models based on heart rate variability during sleep stages for pre-hospital screening

Duyan Geng [a,b,*], Qiang An [c], Zhigang Fu [d], Chao Wang [c], Hongxia An [c]

[a] Hebei University of Technology, School of Electrical Engineering, State Key Laboratory of Reliability and Intelligence of Electrical Equipment Co-constructed by Province and Ministry, Tianjin, 300400, China
[b] Hebei Key Laboratory of Electromagnetic Field and Electrical Reliability, School of Electrical Engineering, Hebei University of Technology, Tianjin, 300400, China
[c] Hebei University of Technology, School of Life Science and Health Engineering, Tianjin, 300130, China
[d] Physical Examination Centre, The 983 Hospital of Joint Logistics Support Force of the Chinese People's Liberation Army, Tianjin, China

Deep Learning Approach

ARTICLES

https://doi.org/10.1038/s41591-022-01811-5

nature medicine

OPEN

Machine learning model to predict mental health crises from electronic health records

Roger Garriga [1,2], Javier Mas [1,3], Semhar Abraha [4,5], Jon Nolan [4], Oliver Harrison [1], George Tadros [4,6] and Aleksandar Matic [1]

SCIENTIFIC REPORTS

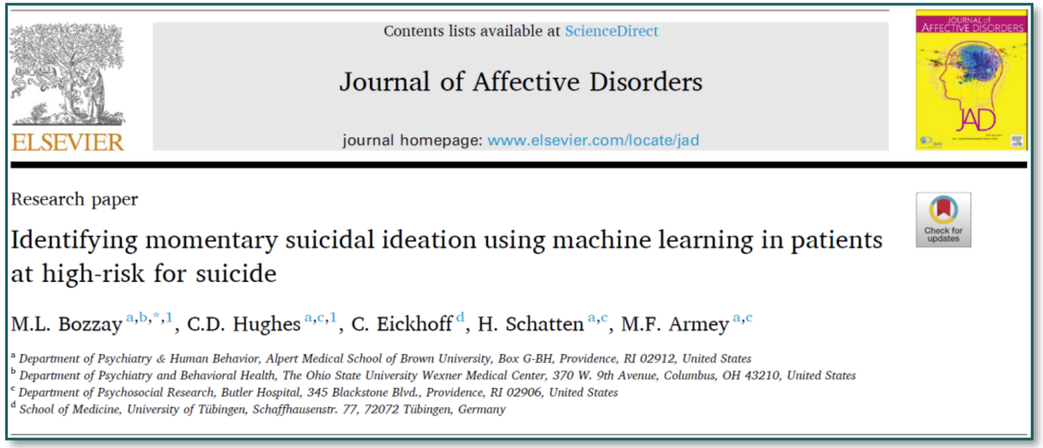
natureresearch

A deep learning model for detecting mental illness from user content on social media

Jina Kim [1,2], Jieon Lee [1], Eunil Park [1,3] & Jinyoung Han [3]

MENTAL HEALTH DIAGNOSIS: **STUDY #1 - SUICIDAL IDEATION**

- **Problem**: Identifying momentary suicidal ideation
- **Data**: Ecological Momentary Assessment
- **Algorithm/Method**: Random Forest
- **Results/Evaluation**: Accuracy, Precision, Recall/sensitivity, Specificity, NPV, and ROC-AUC


Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

ELSEVIER

Research paper

Identifying momentary suicidal ideation using machine learning in patients at high-risk for suicide

M.L. Bozzay [a,b,*,1], C.D. Hughes [a,c,1], C. Eickhoff [d], H. Schatten [a,c], M.F. Armey [a,c]

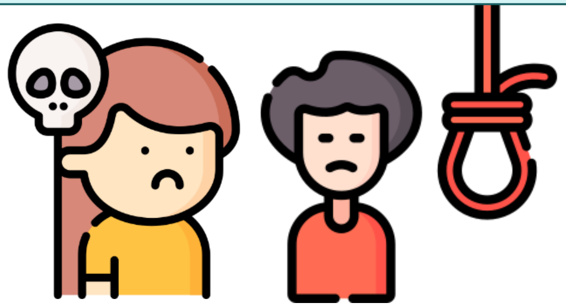
[a] Department of Psychiatry & Human Behavior, Alpert Medical School of Brown University, Box G-BH, Providence, RI 02912, United States
[b] Department of Psychiatry and Behavioral Health, The Ohio State University Wexner Medical Center, 370 W. 9th Avenue, Columbus, OH 43210, United States
[c] Department of Psychosocial Research, Butler Hospital, 345 Blackstone Blvd., Providence, RI 02906, United States
[d] School of Medicine, University of Tübingen, Schaffhausenstr. 77, 72072 Tübingen, Germany

Published in: 2024
DOI: https://doi.org/10.1016/j.jad.2024.08.038

MENTAL HEALTH DIAGNOSIS – **PROBLEM**

- Detect the presence of suicidal ideation (SI) or characteristics of ideation that indicate marked suicide risk are critically needed
- Machine Learning (ML) can be applied to momentary data to improve classification of SI

whether the classification accuracy of ML models varies as
a function of type of training data or characteristics of ideation



MENTAL HEALTH DIAGNOSIS – **DATA**

University of Southampton

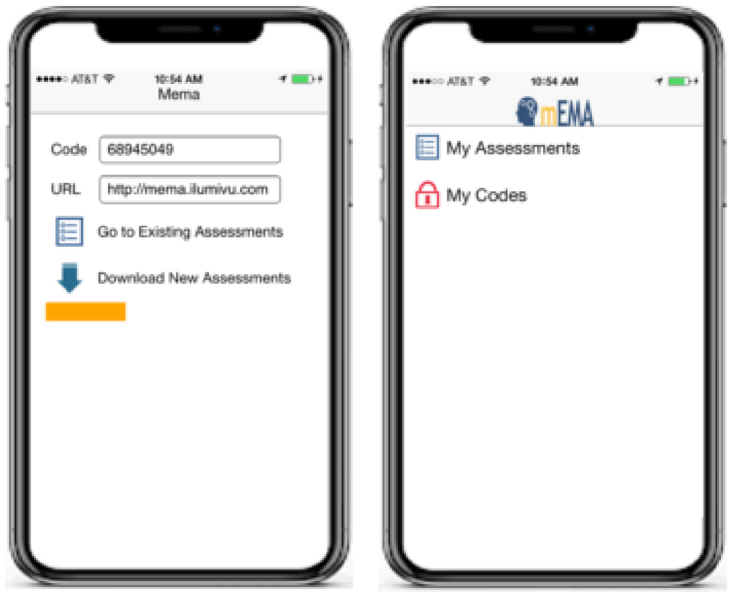
- **Ecological Momentary Assessment (EMA)**: EMA involves sending brief questionnaires to individuals' mobile phones for completion at different times throughout the day
  - ✓ Greater ecological validity (matches patients real world context)
  - ✓ Minimizing recall bias
- **Participants** were 460 patients at an inpatient psychiatric hospital in the northeastern United States
  - ✓ Analyses included only participants who completed EMA (n =257).

Baseline Risk Factors (X) = Depressive symptoms, Borderline personality disorder symptoms, Negative attitudes, etc.
Momentary Risk Factors (X) = Response context, Negative life events, Positive and negative affect, etc.

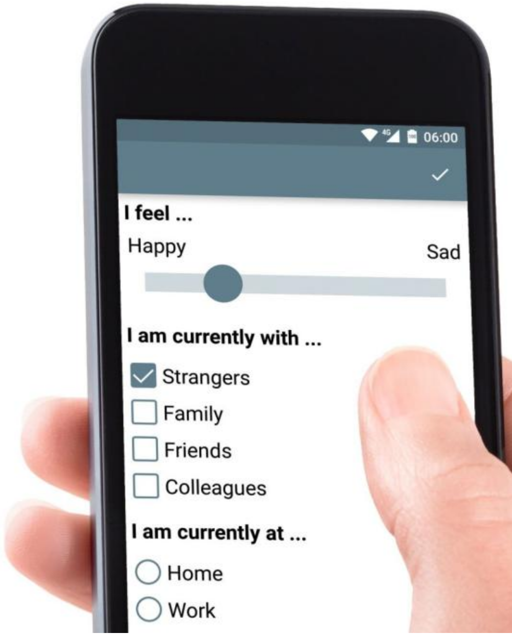
Momentary Outcomes (Y):

1. Suicidal ideation presence
2. Duration - shorter and longer
3. Intensity – lower and higher

Note: Momentary risk factors were assessed via EMA prompts delivered via Ilumivu's HIPAA certified mEMA phone application.



© https://ilumivu.com

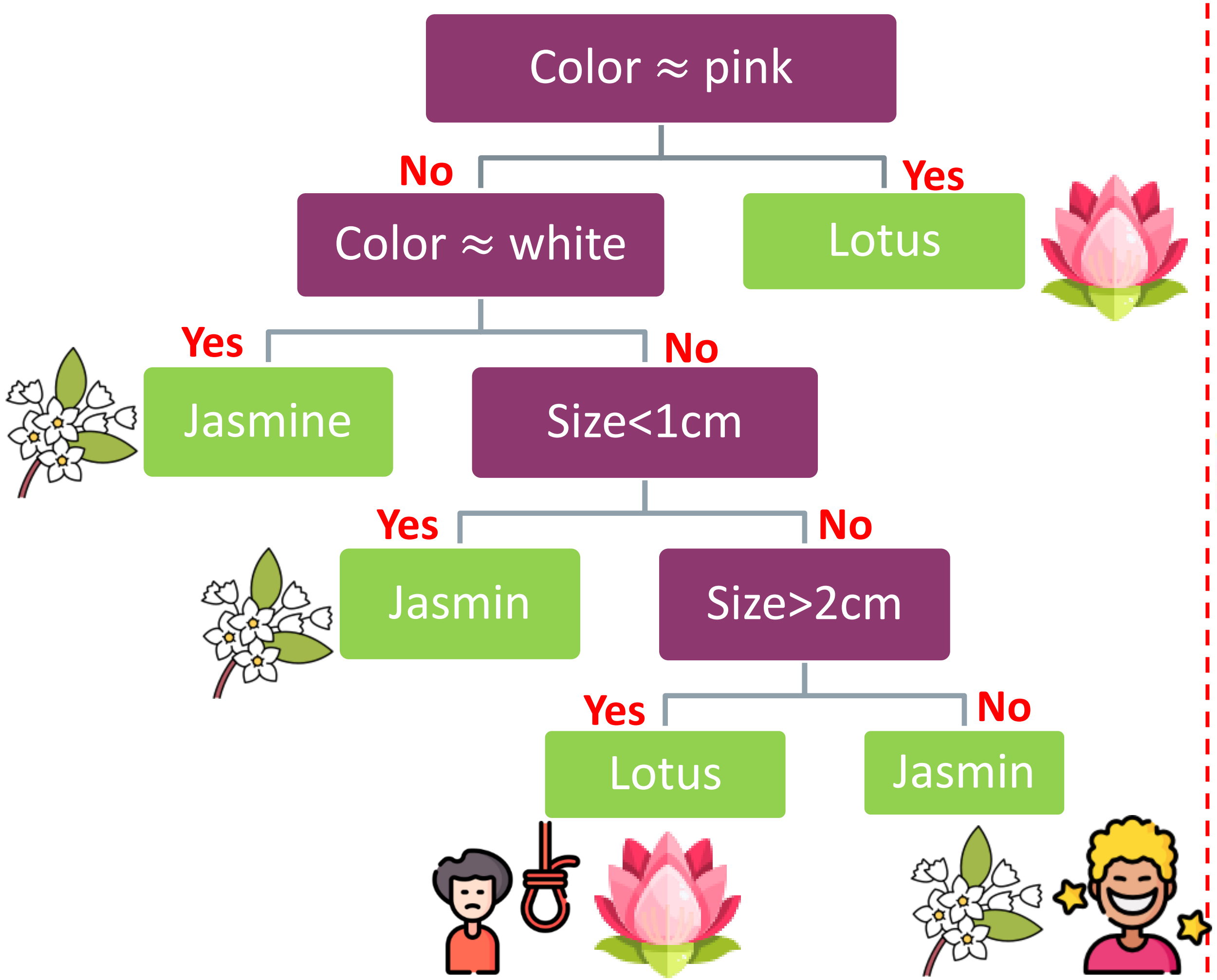


© www.movisens.com

MENTAL HEALTH DIAGNOSIS – ALGORITHM/METHOD

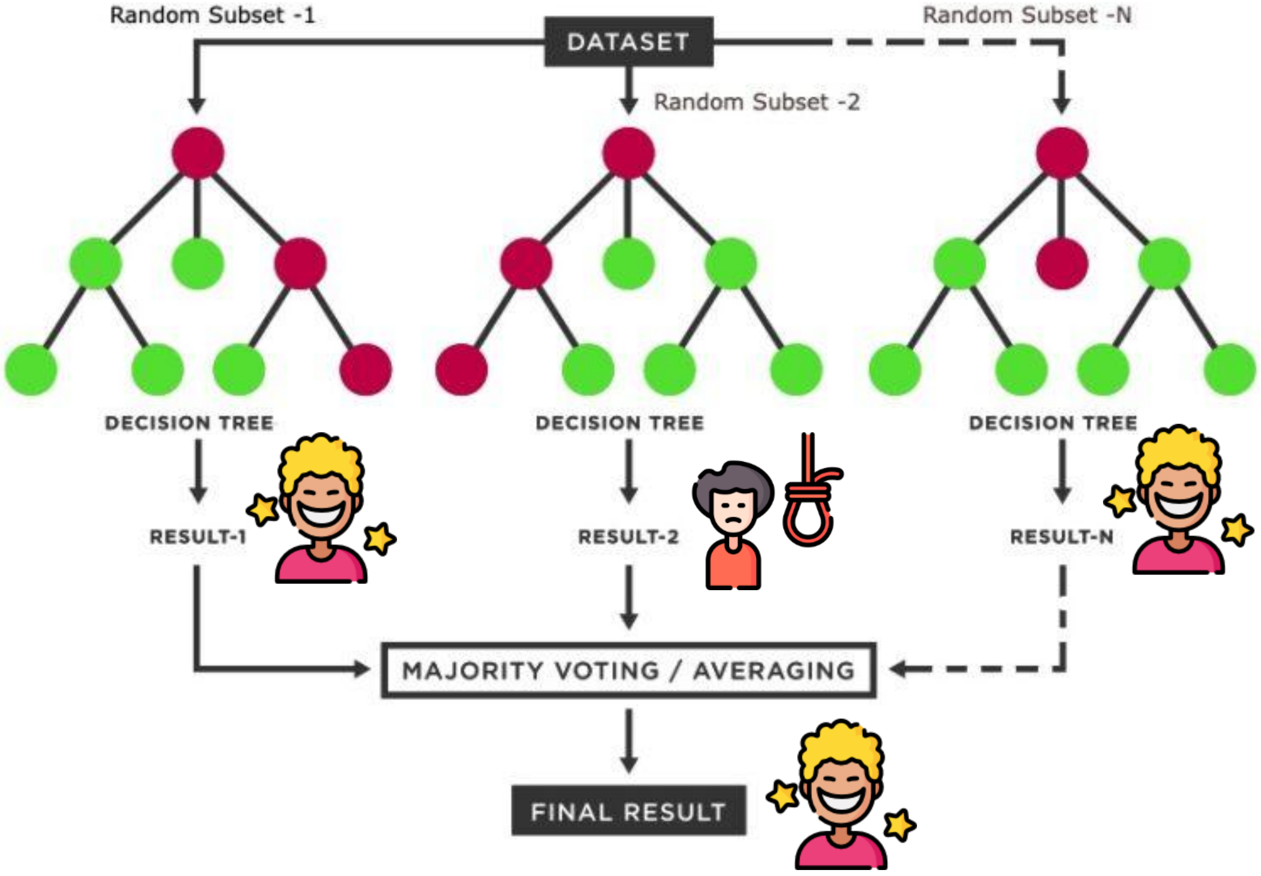
1. Decision Tree: Jasmin Vs Lotus classification

Input = {color:white, size:0.5cm}



2. Random Forest: Suicidal ideation presence Vs Suicidal ideation not presence

Input = {negative life events, affect}



University of Southampton

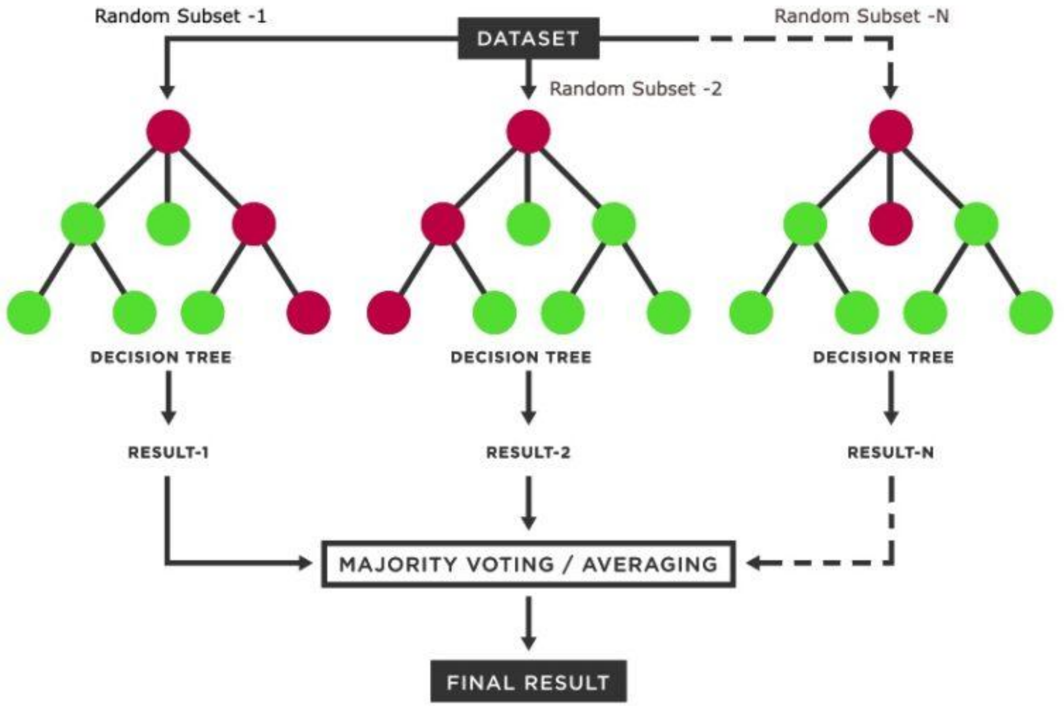
Random Forest

$$f(X) \to Y$$

1. Baseline Risk Factors → Suicidal ideation presence
2. Baseline Risk Factors → Duration
3. Baseline Risk Factors → Intensity

4. Momentary Risk Factors → Suicidal ideation presence
5. Momentary Risk Factors → Duration
6. Momentary Risk Factors → Intensity

7. BL + EMA → Suicidal ideation presence
8. BL + EMA → Duration
9. BL + EMA → Intensity

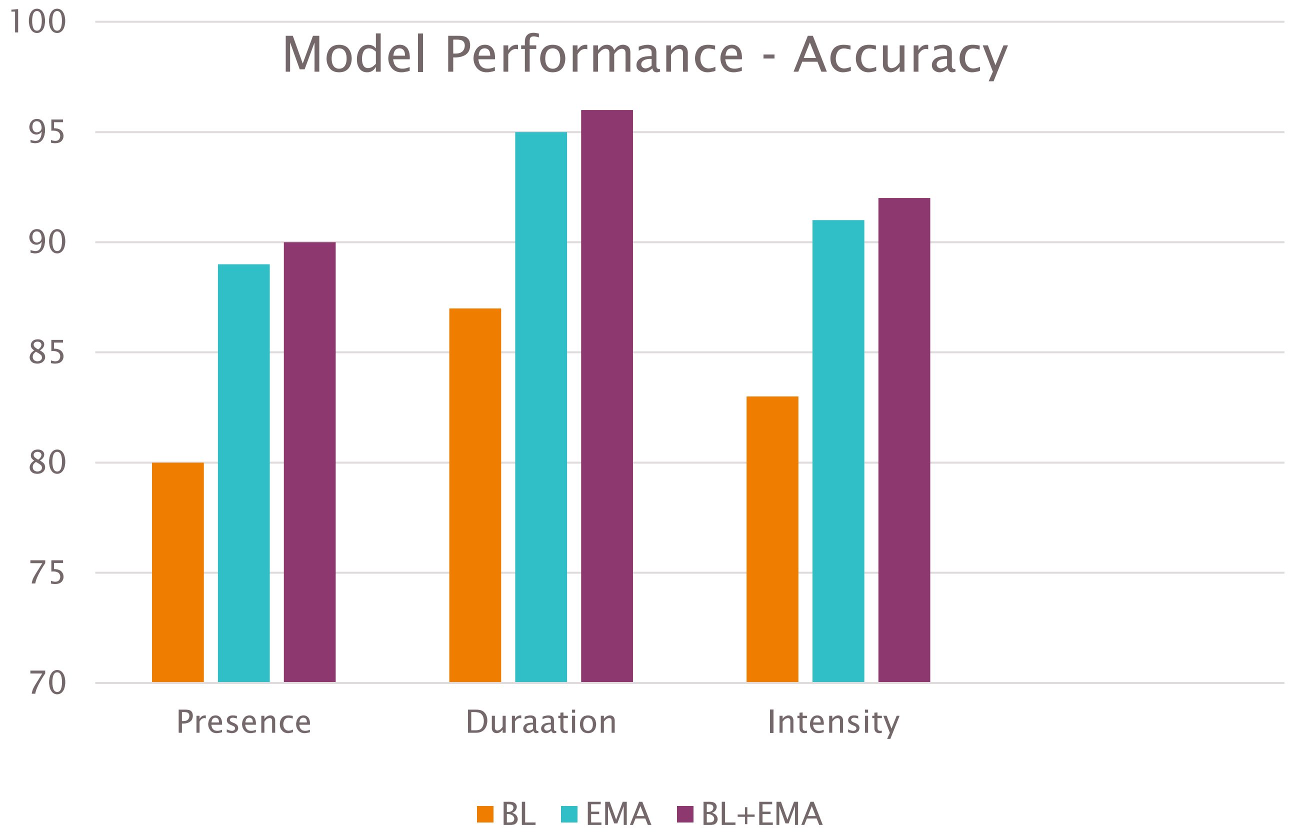


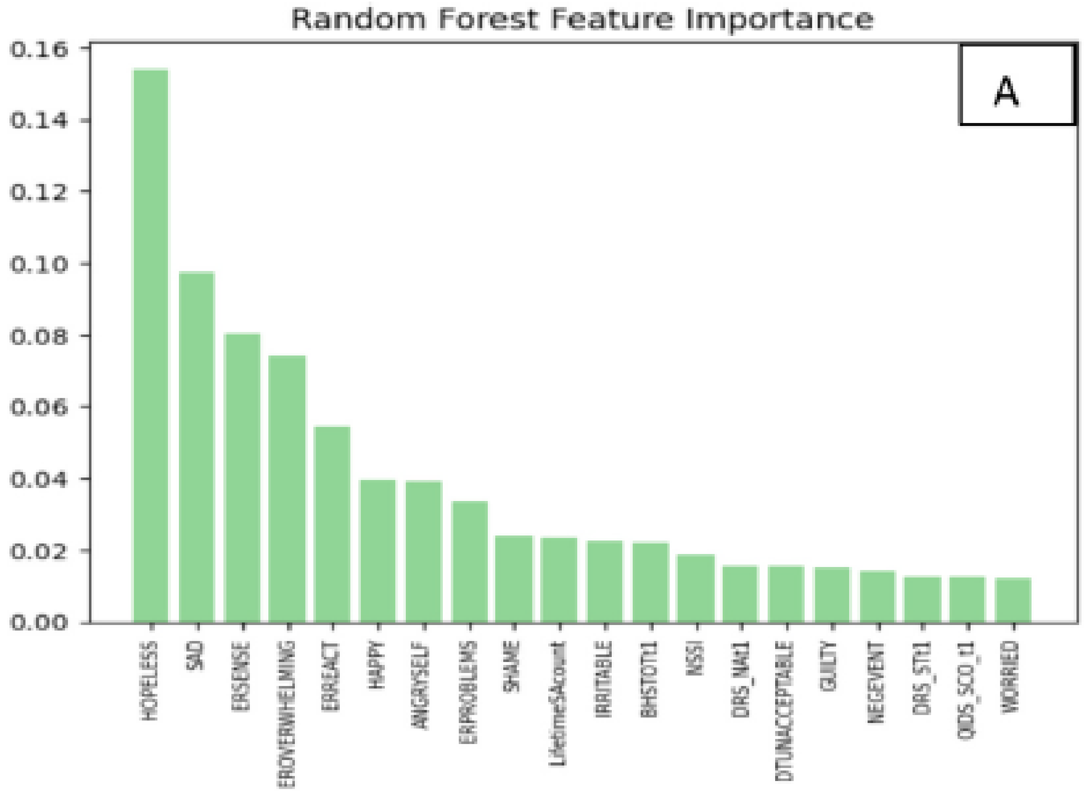
- **Model training**: 80/20 split for training/test data
- **Class imbalance**: Class imbalance skewed towards the negative class

MENTAL HEALTH DIAGNOSIS – **RESULTS/EVALUATION**

University of Southampton

- ❑ Models trained with only momentary data tended to do better than models trained with only baseline data
- ❑ Models trained with only baseline or only momentary data were outperformed by models trained using both sources



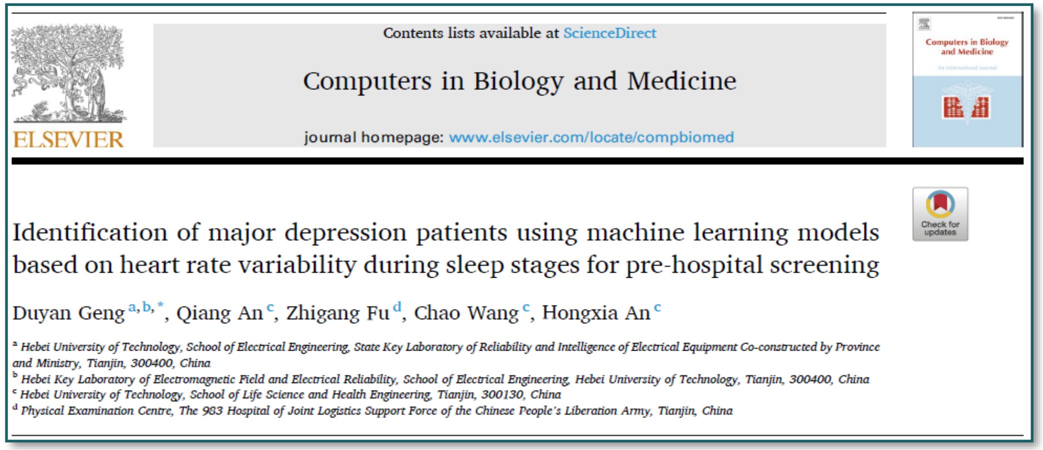


Feature Importance: SI Presence

MENTAL HEALTH DIAGNOSIS: **STUDY #2 - MAJOR DEPRESSIVE DISORDER**

University of Southampton

- **Problem**: Identification of MDD patients
- **Data**: Electrocardiogram (ECG)
- **Algorithm/Method**: BO-ERTC*
- **Results/Evaluation**: Accuracy, Recall/Sensitivity, Specificity, F1-score



Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Identification of major depression patients using machine learning models based on heart rate variability during sleep stages for pre-hospital screening

Duyan Geng [a,b,*], Qiang An [c], Zhigang Fu [d], Chao Wang [c], Hongxia An [c]

[a] Hebei University of Technology, School of Electrical Engineering, State Key Laboratory of Reliability and Intelligence of Electrical Equipment Co-constructed by Province and Ministry, Tianjin, 300400, China
[b] Hebei Key Laboratory of Electromagnetic Field and Electrical Reliability, School of Electrical Engineering, Hebei University of Technology, Tianjin, 300400, China
[c] Hebei University of Technology, School of Life Science and Health Engineering, Tianjin, 300130, China
[d] Physical Examination Centre, The 983 Hospital of Joint Logistics Support Force of the Chinese People's Liberation Army, Tianjin, China

Published in: 2023
DOI: https://doi.org/10.1016/j.compbiomed.2023.107060

*Bayesian Optimised Extremely Randomized Trees Classifier

MENTAL HEALTH DIAGNOSIS – **PROBLEM**

- ❑ Depression remains the second leading cause of human death, with high rates of misdiagnosis and underdiagnosis (Santomauro et al.)
- ❑ Over 46% of people with mental illness are diagnosed with depression at the time of suicide, making it the most common condition.

An interpretable machine learning solution to optimize initial screening for major depressive disorder (MDD) in both male and female patients.

Santomauro et al., (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. The Lancet.

MENTAL HEALTH DIAGNOSIS – **DATA**

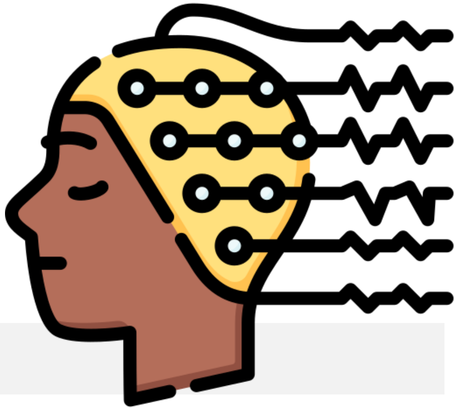
University of Southampton

- **Electrocardiogram (ECG):**
  - ✓ Polysomnographic (PSG) signal data were obtained from the STAGES*
  - ✓ EEG signals recorded during nighttime sleep stages
  - ✓ Segmented the **processed ECG** signals of the entire overnight sleep stage into non-overlapping 5-min segments

- **Preprocessing:**
  - ✓ filtering and noise removals
  - ✓ Motion artifacts

- **Participants:**
  - ✓ Depressed subjects were diagnosed based on the PHQ-9
  - ✓ To balance the data, 80 subjects with a complete PSG signal were randomly selected

    (40 patients with MDD and 40 healthy subjects as controls, with 1:1 male to female ratio)



*Stanford Technical Analysis and Sleep Genome Study

MENTAL HEALTH DIAGNOSIS – **DATA**

- ❑ The time - frequency parameters of heart rate variability (HRV)

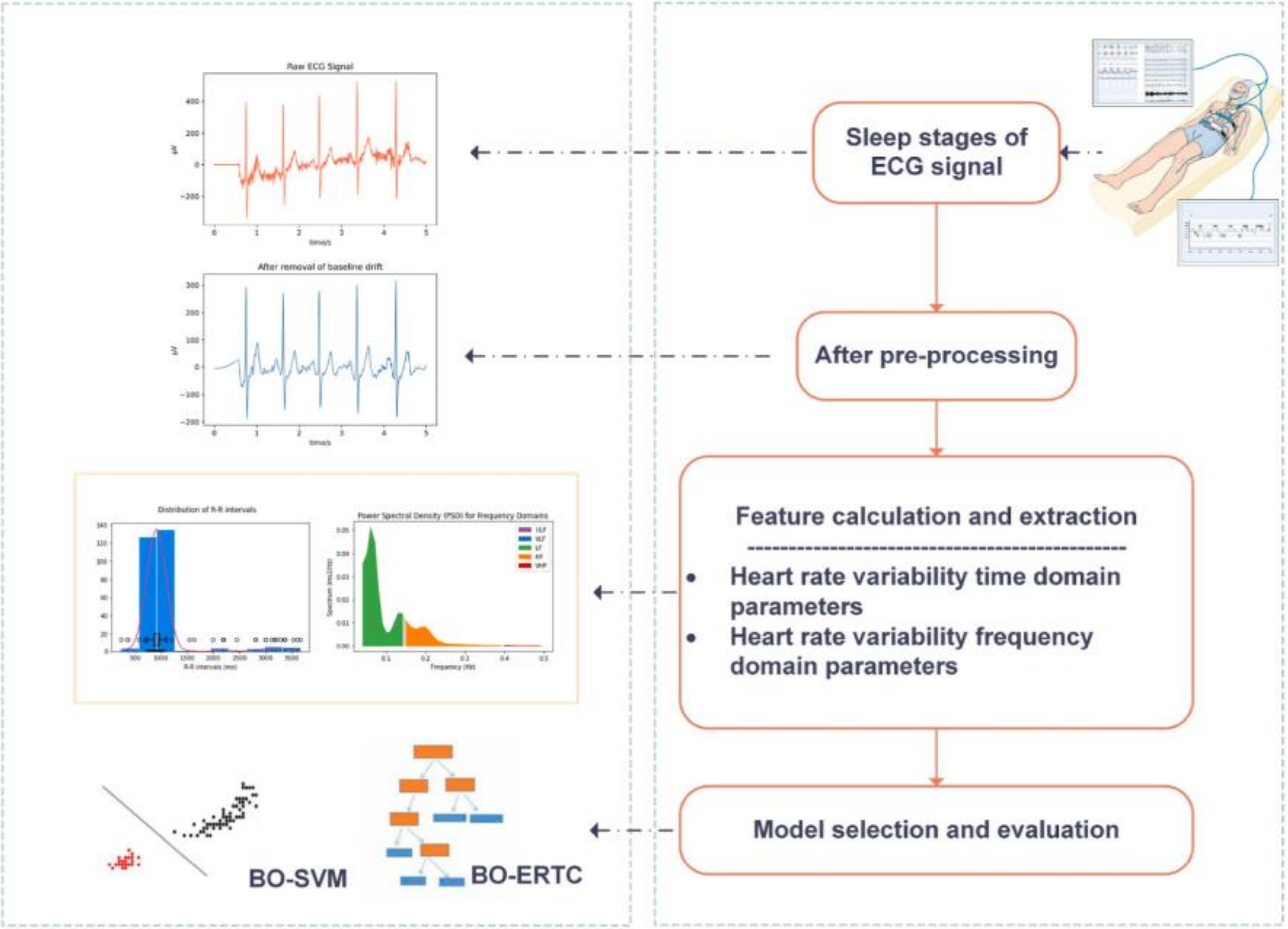
A. **Time Domain Parameters** (X) = {difference between the mean HRI , the mean HR, the maximum and minimum HR, etc. }

B. **Frequency Domain Parameters*** (X) = {HRV_VLF, HRV_LF,HRV_HF, etc. }

Total Features = 24

Outcomes (Y):

1. Healthy
2. MMD



Includes various spectral components

Note:

- ❑ HRI – Heart Rate Interval, HR – Heart Rate
- ❑ HRV_VLF - spectral power of very low frequencies, HRV_LF - spectral power of low frequencies, HRV_HF - spectral power of high frequencies

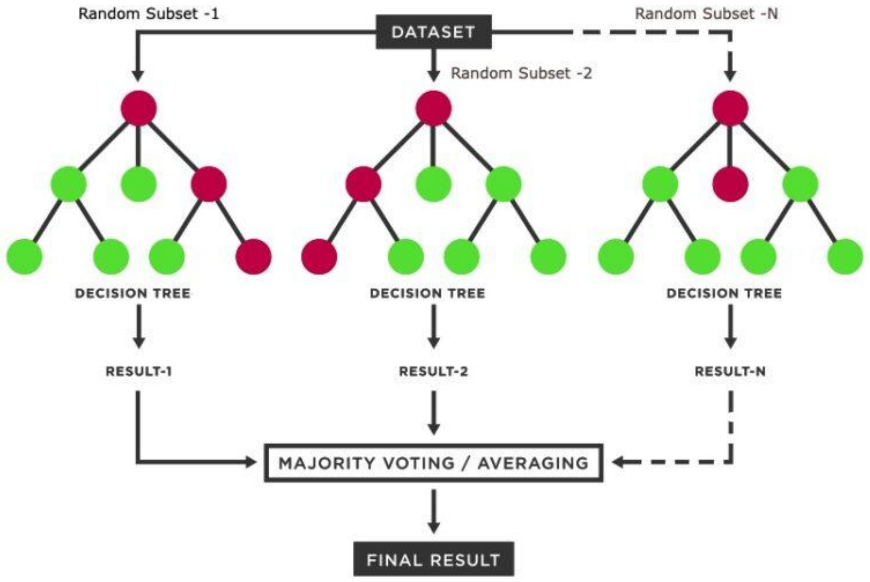
MENTAL HEALTH DIAGNOSIS – ALGORITHM/METHOD

University of Southampton

1. Bayesian Optimised Extremely Randomized Trees Classifier (BO-ERTC)
2. Bayesian Optimised Support Vector Machines (BO-SVM)

Bayesian Optimised Extremely Randomized Trees Classifier

- ERTC is an integrated learning technique, for classifying data based on ensemble learning of decision tree
- ERTC aggregates the results of multiple de-correlated decision trees collected in a forest to output classification results
- Each decision tree in a forest of extremely random trees is constructed from the original training samples



$$f(X) \rightarrow Y$$

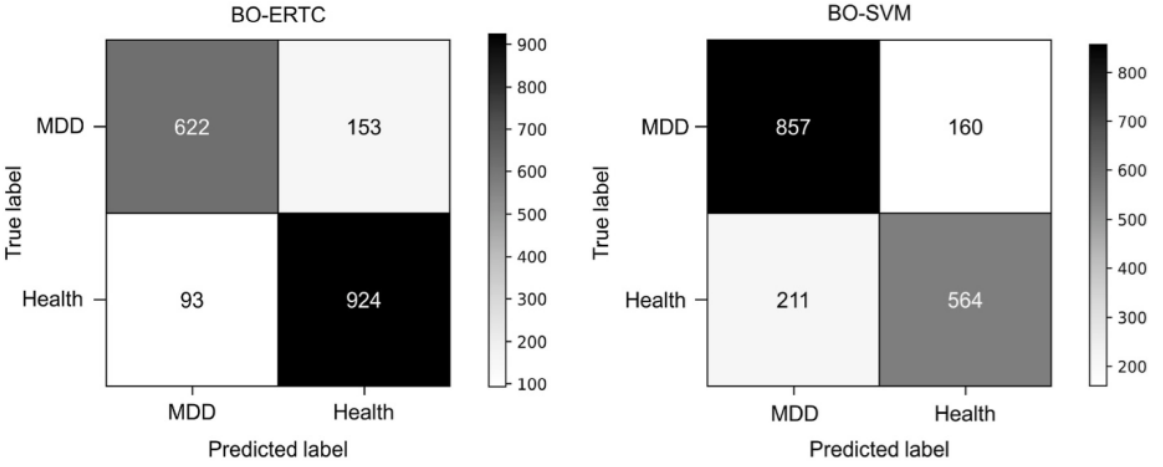
Time Domain Parameters (TDP) + Frequency Domain Parameters (FDP) → Healthy/MMD
TDP+FDP + Gender Feature → Healthy/MMD

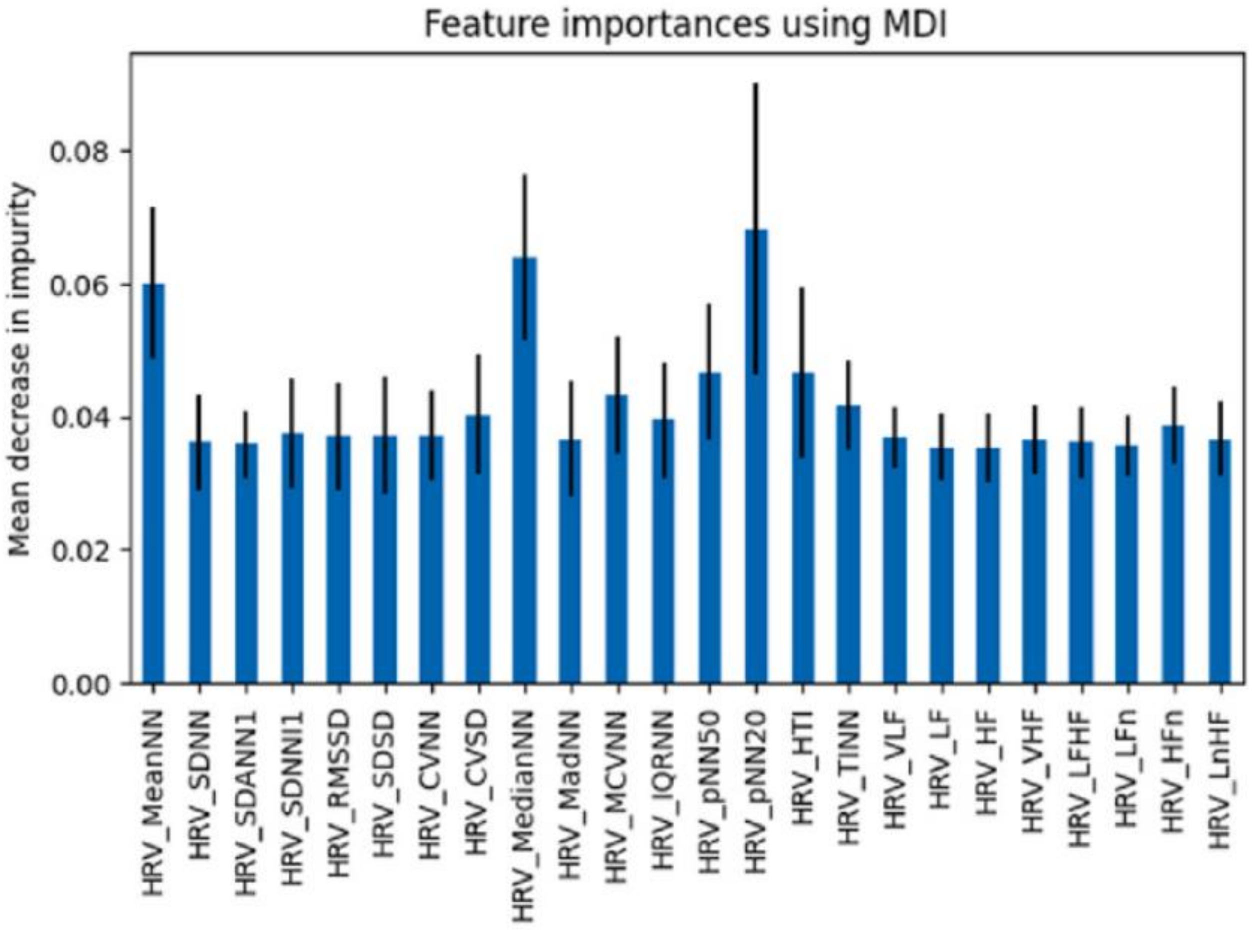
- **Model training**: 80/20 split for training/test data with 10 fold cross validation
- **Model Optimization and Tuning**: Bayes' theorem

Note:
Before training, different machine learning algorithms, such as ordinary SVM, ERTC, Bayesian network, and extreme gradient boosting, were used to model on a smaller dataset. The results showed that ERTC and SVM provided better results than all the algorithms we tried.

MENTAL HEALTH DIAGNOSIS – **RESULTS/EVALUATION**

Group	Classifier	TPR (%)	FPR (%)	TNR (%)	FNR (%)	Accuracy (%)	Precision (%)	F1-score
Control	BO-ERTC	89.28	10.71	75.35	24.64	83.25	82.62	0.86
	BO-SVM	77.62	22.38	80.06	19.94	79.07	72.52	0.75
With gender features group	BO-ERTC	91.45	8.56	79.61	20.39	86.32	85.48	0.88
	BO-SVM	77.90	22.10	80.24	19.76	79.29	72.77	0.75





University of Southampton

- **Problem**: Predict Mental Health Crises
- **Data**: Electronic Health Records (EHR)
- **Algorithm/Method**: XGBoost*
- **Results/Evaluation**: ROC Curve

ARTICLES

https://doi.org/10.1038/s41591-022-01811-5

nature medicine

Check for updates

OPEN

Machine learning model to predict mental health crises from electronic health records

Roger Garriga[1,2], Javier Mas[1,3], Semhar Abraha[4,5], Jon Nolan[4], Oliver Harrison[1], George Tadros[4,6] and Aleksandar Matic[1]

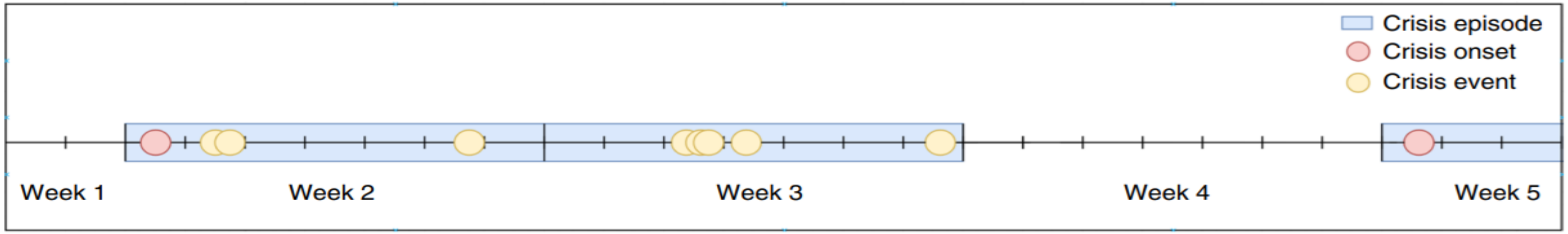
Published in: 2022
DOI: https://doi.org/10.1038/s41591-022-01811-5

*Also explored many different ML algorithms such as, LR, RF, DT, NB, ANN, LSTM, etc.

MENTAL HEALTH DIAGNOSIS – **PROBLEM**

- The timely identification of patients who are at risk of a mental health crisis can lead to:
  - improved outcomes
  - mitigation of burdens and costs
- The high prevalence of mental health problems means that the manual review of complex patient records to make proactive care decisions is not feasible in practice

Developed a machine learning model that uses **electronic health records** to continuously monitor patients for risk of a **mental health crisis*** over a period of 28 days

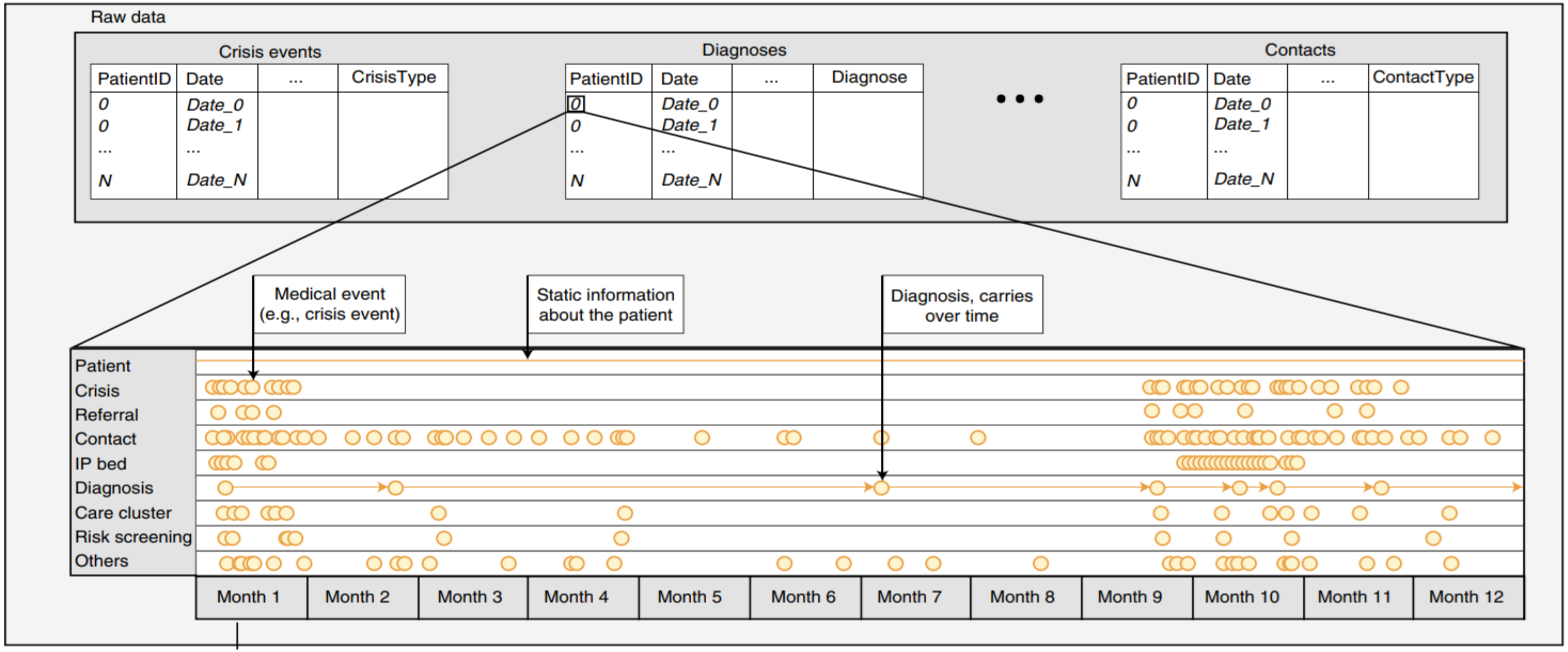


Example of a crisis episode timeline: crisis onset is the first crisis event of a crisis episode that follows a stable week (that is, a week without crisis events)

*The model was designed to be queried weekly to infer each patient's risk of experiencing a crisis episode during the upcoming 28-day period
- **Prediction target as the onset of a crisis episode**, which contains one or more crisis events, preceded by at least one full stable week without any crisis event

MENTAL HEALTH DIAGNOSIS – **DATA**

- **Electronic Health Records (EHR):**
  - ✓ The study cohort data contained 5,816,586 records collected between September 2012 and November 2018
  - ✓ From 17,122 unique patients aged between 16 and 102 years
  - ✓ This included patients with a wide range of diagnosed disorders, including mood, psychotic, organic, neurotic and personality disorders
  - ✓ In total, 60,388 crisis episodes were included in the analysis, with a mean of 24 crisis events per episode



MENTAL HEALTH DIAGNOSIS – **DATA**

University of Southampton

- Extracted three feature categories

  A. **Static or semi-static patient information**

     - such as age, gender and International Classification of Diseases 10 (ICD-10)34 coded diagnoses

  B. **latest available assessments and interactions with the hospital**

     - for example, most recent risk assessments or well-being indicators and severity and number of crisis events in the last episode and similar

  C. **Variables representing the time elapsed since the registered events**

     - for example, crisis episodes, contacts and referrals

Total Features = 198

Outcomes (Y):

1. Crisis Onset* (Binary Outcome)

**When the system was implemented, instead of a binary outcome, the model was generating a predicted risk score (PRS) between 0 and 1 for each patient*

MENTAL HEALTH DIAGNOSIS – **ALGORITHM/METHOD**

University of Southampton

1. **XGBoost**
2. Logistic Regression
3. Random Forest
4. Decision Trees
5. Naïve Bayes
6. **ANN**
7. **Long short-term memory (LSTM)**

$$f(X) \rightarrow Y$$

Static or semi-static features **+** Diagnosis features **+** EHR weekly aggregations **+** Time-elapsed features **+** Last crisis episode descriptors **+** Status features → Crisis Onset (Binary Outcome)

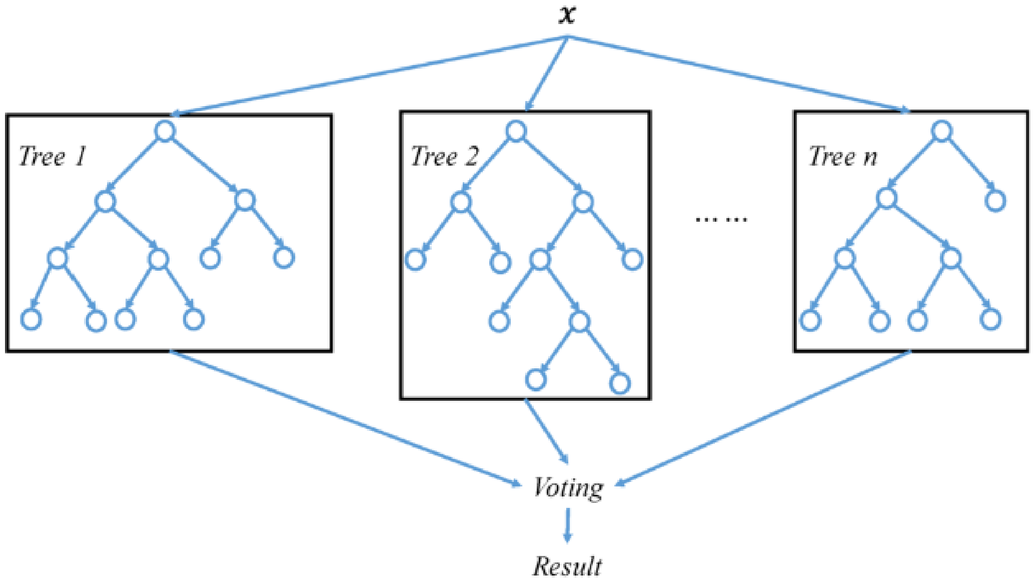
- ❑ **Model training**: 80/10/10 split for training/validation/test
- ❑ **Best Model**: **XGBoost**
- ❑ **Model Optimization and Tuning**: Bayesian optimization technique

Note:
Before training, different machine learning algorithms, such as ordinary SVM, ERTC, Bayesian network, and extreme gradient boosting, were used to model on a smaller dataset. The results showed that ERTC and SVM provided better results than all the algorithms we tried.

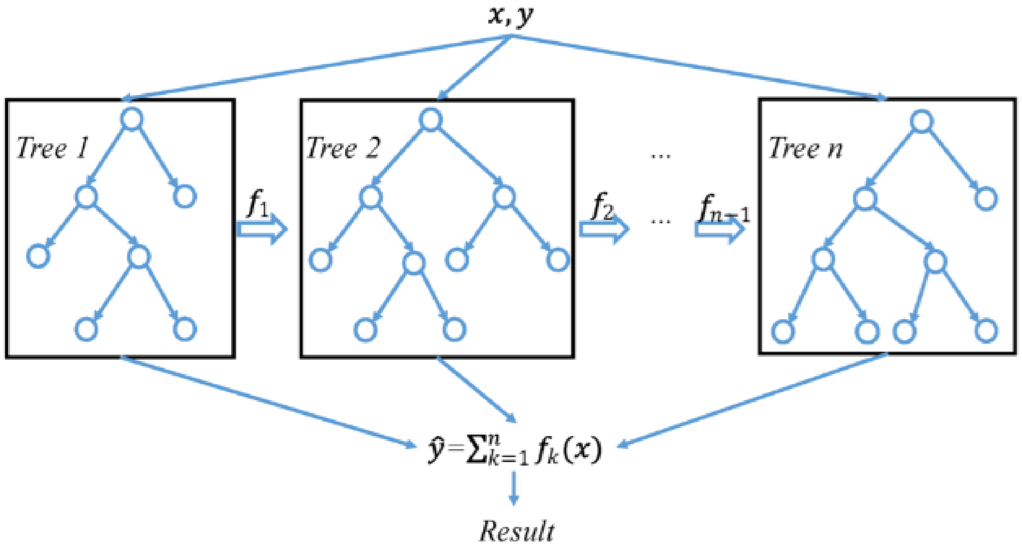
MENTAL HEALTH DIAGNOSIS – ALGORITHM/METHOD

XGBoost

- XGBoost is an implementation of Gradient Boosting Machines (GBM)
- GBM build a sequence of decision trees such that every new tree improves upon the performance of previous iterations



General architecture of **Random Forest**



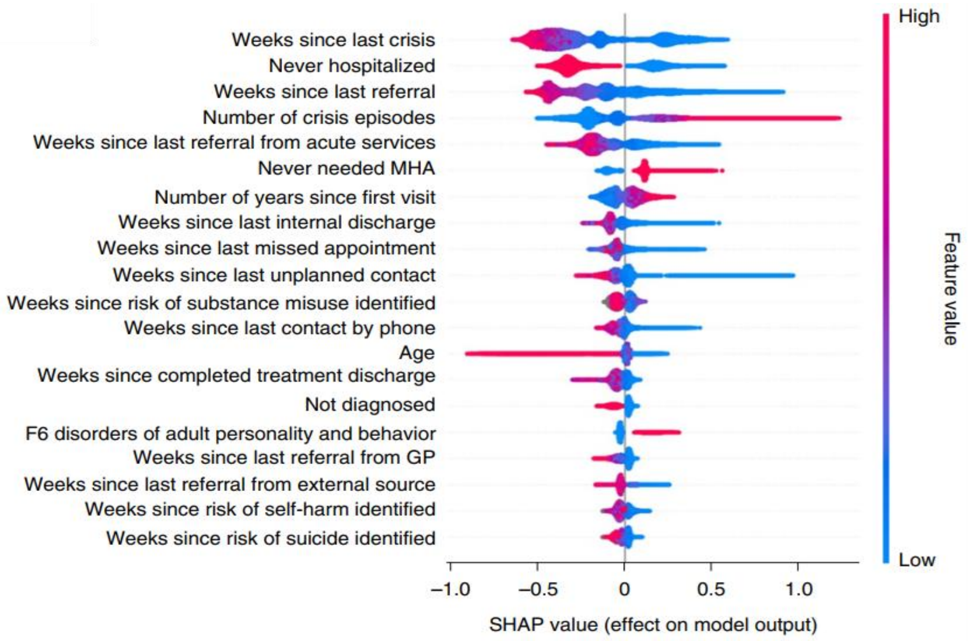
General architecture of **XGBoost**

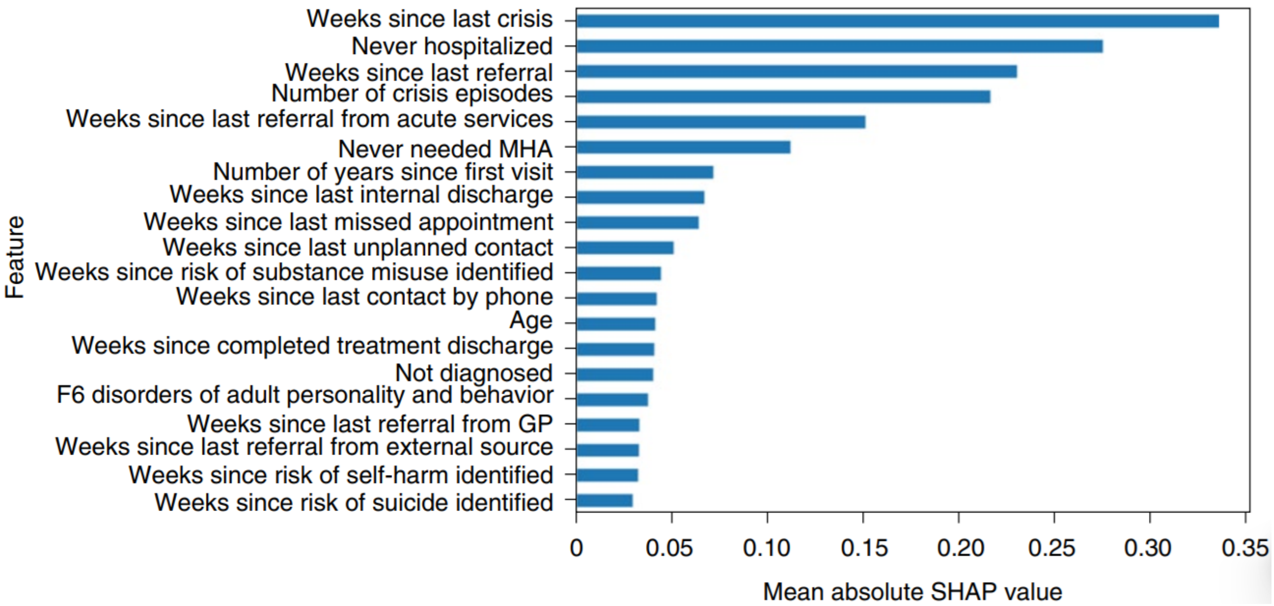
- Initial Prediction, Error Calculation, Building the First Decision Tree, Subsequent Trees and Error Correction, Minimizing Loss Function, Stopping Criteria

Wang et al., (2019). A hybrid ensemble method for pulsar candidate classification. *Astrophysics and Space Science*, *364*, 1-13.

MENTAL HEALTH DIAGNOSIS – **RESULTS/EVALUATION**

Model	AUROC (std)	AP (std)
Clinical baseline	0.736 (0.010)	0.092 (0.006)
Diagnosis baseline	0.746 (0.011)	0.092 (0.006)
XGBoost	**0.797 (0.012)**	**0.159 (0.014)**
Logistic Regression	0.788 (0.010)	0.140 (0.009)
Random Forest	0.788 (0.012)	0.143 (0.013)
Decision Tree	0.776 (0.011)	0.118 (0.007)
Naive Bayes	0.751 (0.011)	0.108 (0.009)
SGD (modified huber)	0.785 (0.010)	0.134 (0.008)
Feed Forward Neural Network	0.790 (0.011)	0.145 (0.010)
LSTM	0.775 (0.015)	0.148 (0.013)





<u>*Note*</u>: AUROC = Area Under the Receiver Operating Characteristic curve and AP = Average Precision, SHAP (SHapley Additive exPlanations)

MENTAL HEALTH DIAGNOSIS: STUDY #4 – MENTAL HEALTH ILLNESS

- **Problem**: Predict Mental Health Illness
- **Data**: Social Media Content
- **Algorithm/Method**: Convolutional Neural Network
- **Results/Evaluation**: Accuracy, F1-score, Precision, Recall

SCIENTIFIC REPORTS
natureresearch
Check for updates

A deep learning model for detecting mental illness from user content on social media

Jina Kim[1,2], Jieon Lee[1], Eunil Park[1,3]✉ & Jinyoung Han[3]✉

Published in: 2022
DOI: https://doi.org/10.1038/s41591-022-01811-5

MENTAL HEALTH DIAGNOSIS – **PROBLEM**

- Users of social media often **share their feelings or emotional states through their posts**
- Several scholars have analyzed user-generated **content on social media for observing users' emotional state or mental illness**, including depression, anxiety, or schizophrenia

- A study collected **Twitter posts** of users who reportedly had been **diagnosed as depression**, analyzed the **linguistic** and **emotional characteristics of the posts**, and tracked their **social engagement changes** on Twitter (Choudhury et al., (2013))
- Another study attempted to **predict users' postpartum depression on Facebook**, based on their **posts and comments**, and used specialized psychometric instruments to evaluate the level of postpartum depression between pre- and post-natal periods (Choudhury et al., (2014))

Can we identify whether a user's post on social media belongs to mental illnesses such as depression, anxiety, bipolar, schizophrenia, autism, etc.?

Choudhury et al., (2013). Predicting depression via social media. In Proceedings of the international AAAI conference on web and social media
Choudhury et al., (2014). Characterizing and predicting postpartum depression from shared facebook data. 17th ACM conference on Computer supported cooperative work & social computing

University of Southampton

- **Social Media - Reddit**
  - ✓ User's posts uploaded in multiple mental-health-related subreddits in Reddit
  - ✓ Six mental-health-related subreddits, such as depression, anxiety, bipolar, BPD, schizophrenia, and autism
  - ✓ Overall, this study collected information from 248,537 users, who wrote 633,385
  - ✓ posts in the seven subreddits from January 2017 to December 2018.

Channel	# of users	# of posts	Description
r/mentalhealth	27,177	39,373	The Mental Health subreddit is the central forum to discuss, vent, support and share information about mental health, illness and well-ness
r/depression	136,506	258,496	Peer support for anyone struggling with depression, the mental illness
r/Anxiety	49,735	86,243	Discussion and support for sufferers and loved ones of any anxiety disorder
r/bipolar	14,372	41,493	A safe haven for bipolar related issues. We are a community here not just a help page. Be a part of something that cares about who you are
r/BPD	13,913	38,216	A place for those who have BPD (Borderline Personality Disorder) (also known as EUPD [Emotionally Unstable Personality Disor-der])–, their family members and friends, and anyone else who is interested in learning more about the disorder
r/schizophrenia	5,392	17,506	Welcome! This is a community meant for a discussion of schizophrenia spectrum disorders, and related issues. Feel free to post, discuss, or just lurk. There is no judgement in this place: we are here for each other. Please refrain from self-diagnosis, diagnosing others, or advising specific medical treatments
r/autism	4,754	7,142	No description

MENTAL HEALTH DIAGNOSIS – **DATA**

- ❑ Data pre-processing
- ❑ $\boldsymbol{X}$: The feature will automatically extracted by the deep learning model

Outcomes ($\boldsymbol{Y}$):

1. Depression/Non-depression
2. Anxiety/Non-anxiety
3. Bipolar/Non-bipolar
4. BPD/Non-BPD
5. Schizophrenia/Non-schizophrenia
6. Autism/Non-autism

MENTAL HEALTH DIAGNOSIS – ALGORITHM/METHOD

University of Southampton

1. XGBoost*
2. Convolutional Neural Network (CNN)

$$f(X) \rightarrow Y$$

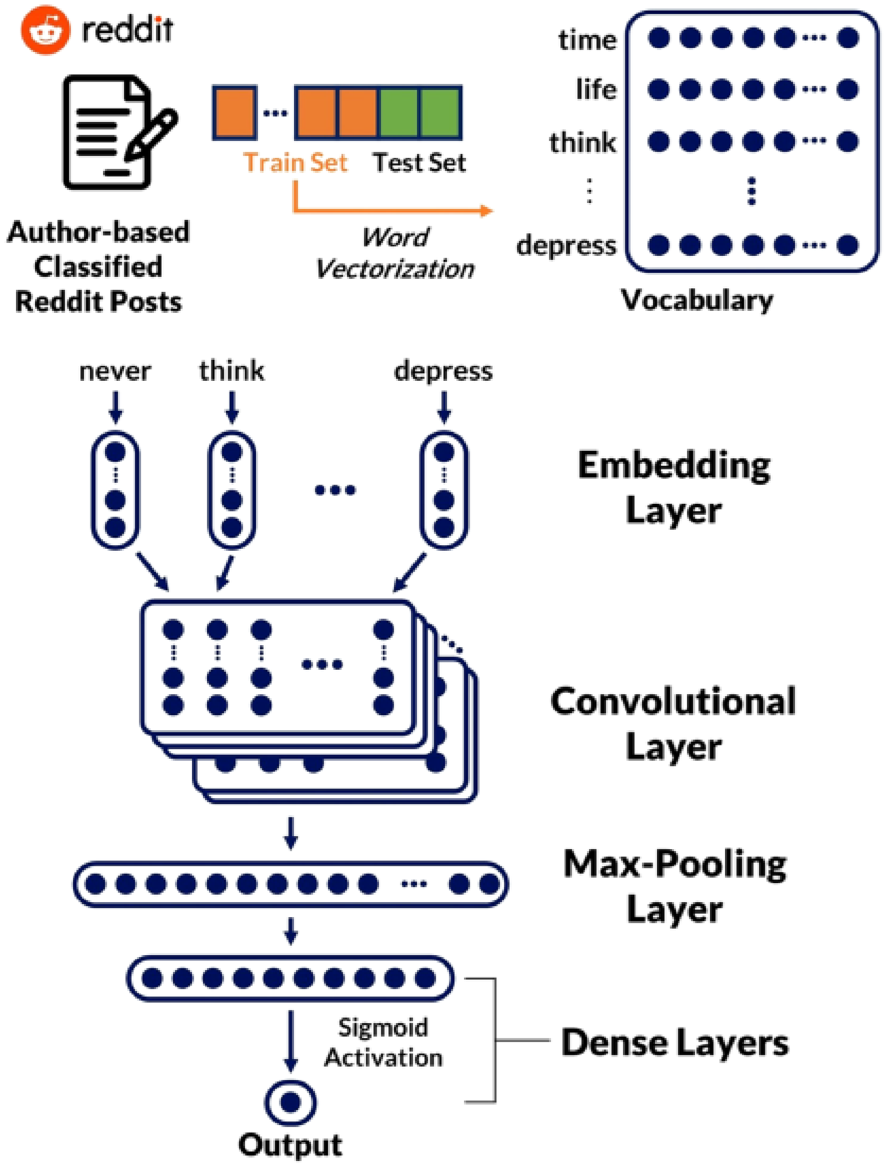
1. Automatically extracted feature $\rightarrow$ Depression/Non−depression
2. Automatically extracted feature $\rightarrow$ Anxiety/Non−anxiety
3. Automatically extracted feature $\rightarrow$ Bipolar/Non−bipolar
4. Automatically extracted feature $\rightarrow$ BPD/Non−BPD
5. Automatically extracted feature $\rightarrow$ Schizophrenia/Non−schizophrenia
6. Automatically extracted feature $\rightarrow$ Autism/Non−autism

- **Model training**: 80/20 split for training/test set
- **Embedding used**: word2vec by Gensim

*XGBoost utilize the TF-IDF feature vectorization method

MENTAL HEALTH DIAGNOSIS – ALGORITHM/METHOD

Convolutional Neural Network



MENTAL HEALTH DIAGNOSIS – **RESULTS/EVALUATION**

University of Southampton

Channel	Class	XGBoost		CNN			
		F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy (%)
r/depression	Depression	58.02	71.69	89.10	71.75	79.49	75.13
	Non-depression	78.65		58.66	82.04	68.41	
r/Anxiety	Anxiety	55.92	70.41	87.54	41.44	56.25	77.81
	Non-anxiety	77.73		75.92	96.91	85.14	
r/bipolar	Bipolar	53.59	85.53	87.22	38.02	52.95	90.20
	Non-bipolar	91.43		90.40	99.05	94.53	
r/BPD	BPD	46.43	85.14	91.84	32.69	48.21	90.49
	Non-BPD	91.37		90.42	99.54	94.76	
r/schizophrenia	Schizophrenia	40.97	86.72	81.16	24.87	38.07	94.33
	Non-schizophrenia	92.52		94.62	99.56	97.03	
r/autism	Autism	38.31	94.91	48.08	49.39	48.73	96.96
	Non-autism	97.35		98.48	98.40	98.44	

GENERIC DECODING OF **SEEN & IMAGINED OBJECTS**

University of Southampton

- ❑ This work shows that **visual features**, including those derived from a deep convolutional neural network, can be predicted from **fMRI patterns**

$$\text{fMRI patterns} \rightarrow \text{visual features}$$

- ❑ Predicted **features are used to identify seen/imagined object categories** (extending beyond decoder training) from a set of computed features for numerous object images

https://doi.org/10.1038/ncomms15037

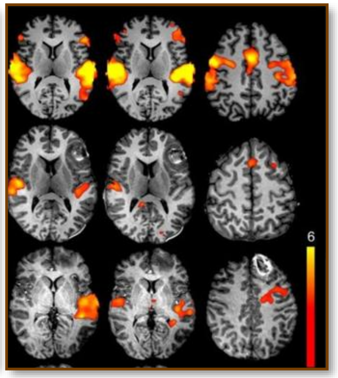

nature COMMUNICATIONS

ARTICLE

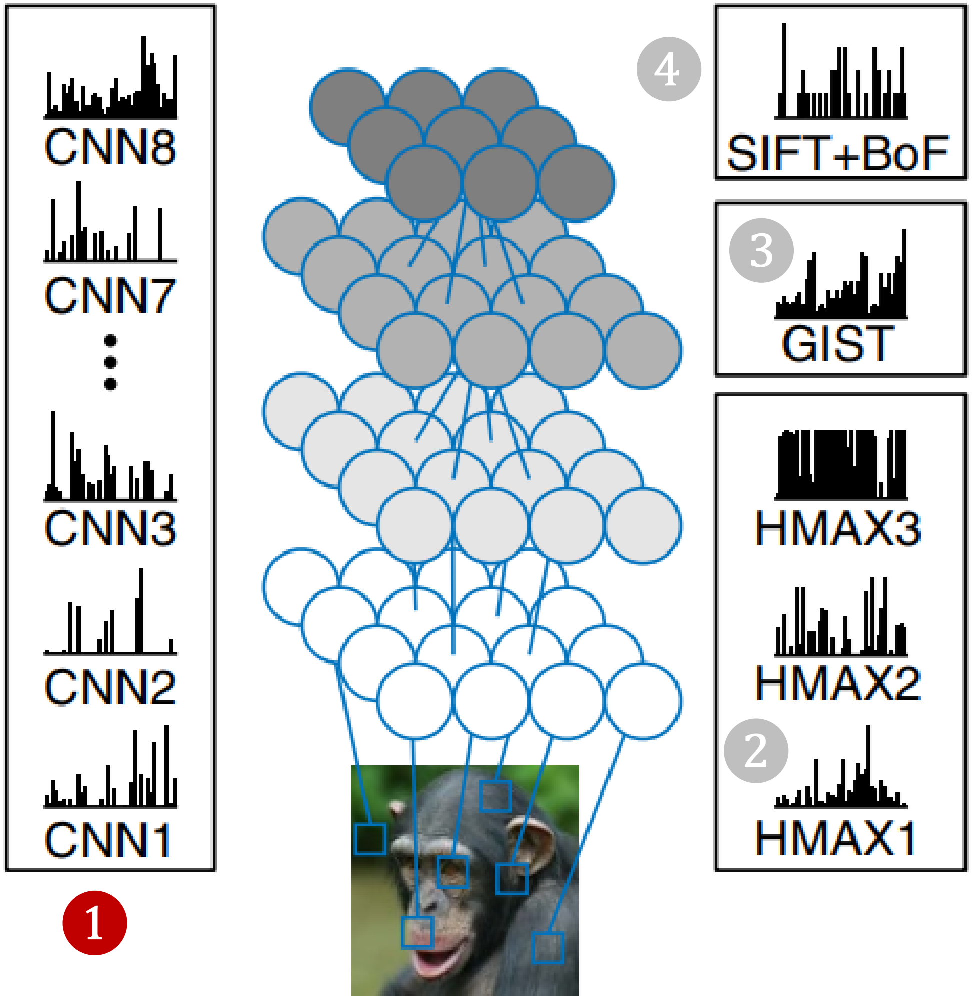
Received 15 Oct 2015 | Accepted 21 Feb 2017 | Published 22 May 2017 DOI: 10.1038/ncomms15037 OPEN

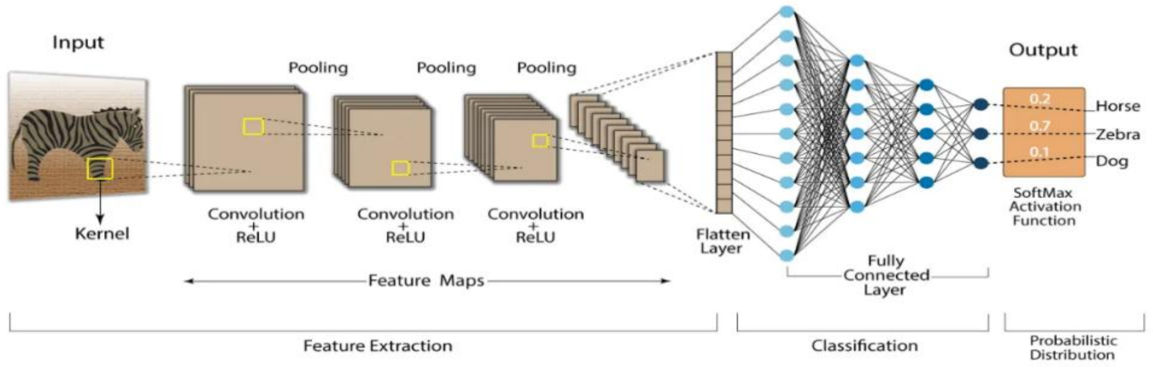
Generic decoding of seen and imagined objects using hierarchical visual features

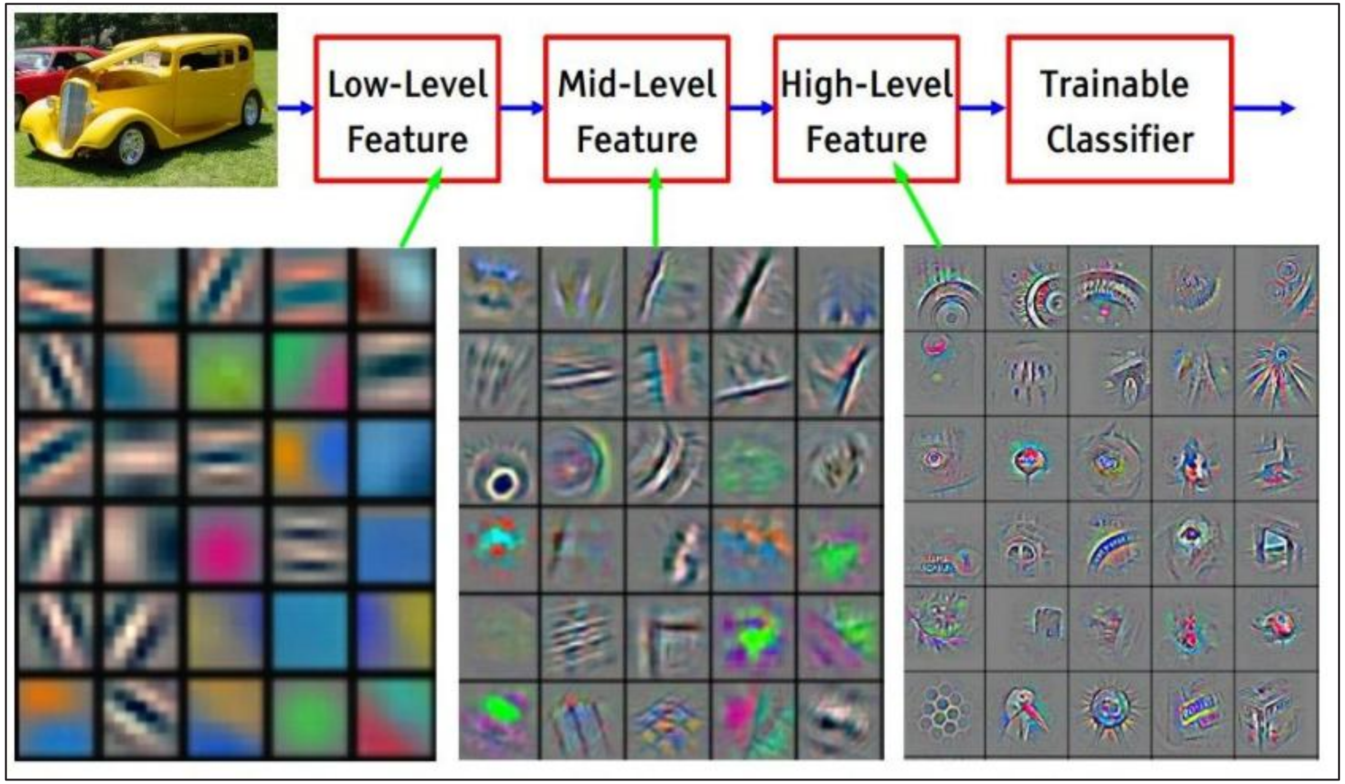
Tomoyasu Horikawa[1] & Yukiyasu Kamitani[1,2]



What is a **Visual Feature**?



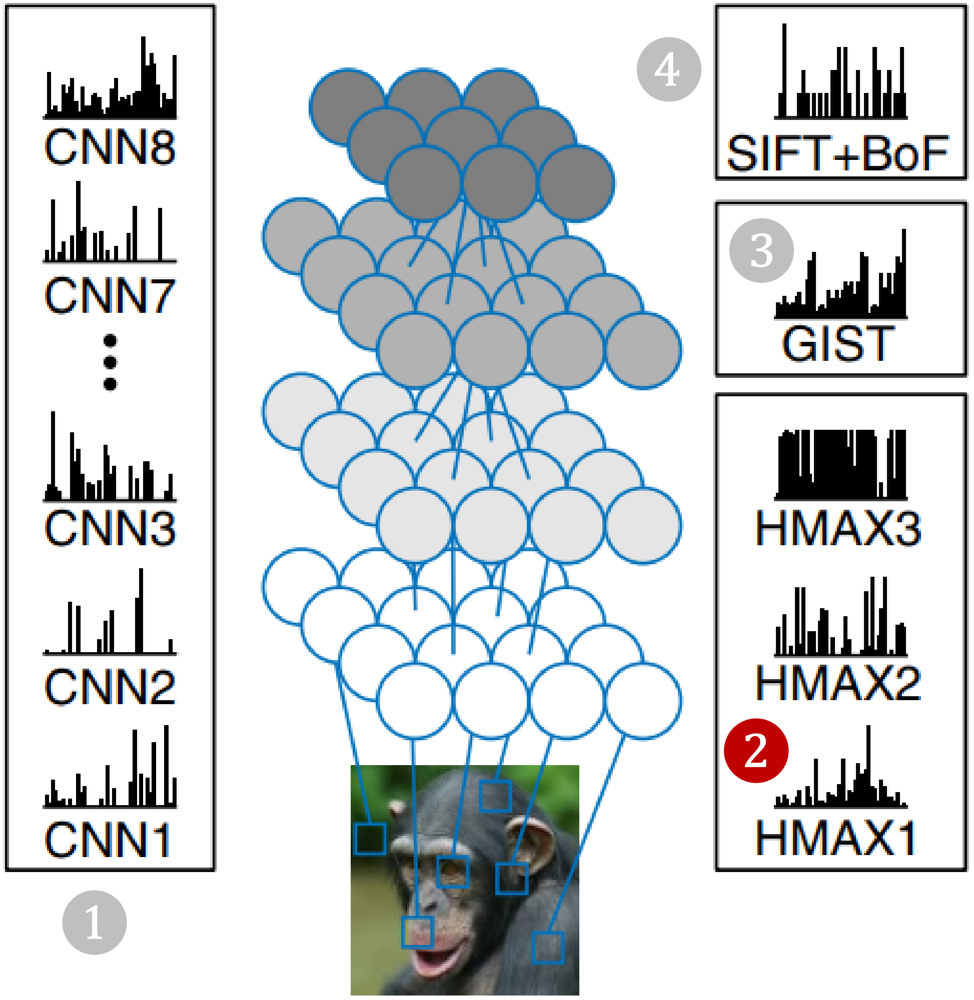


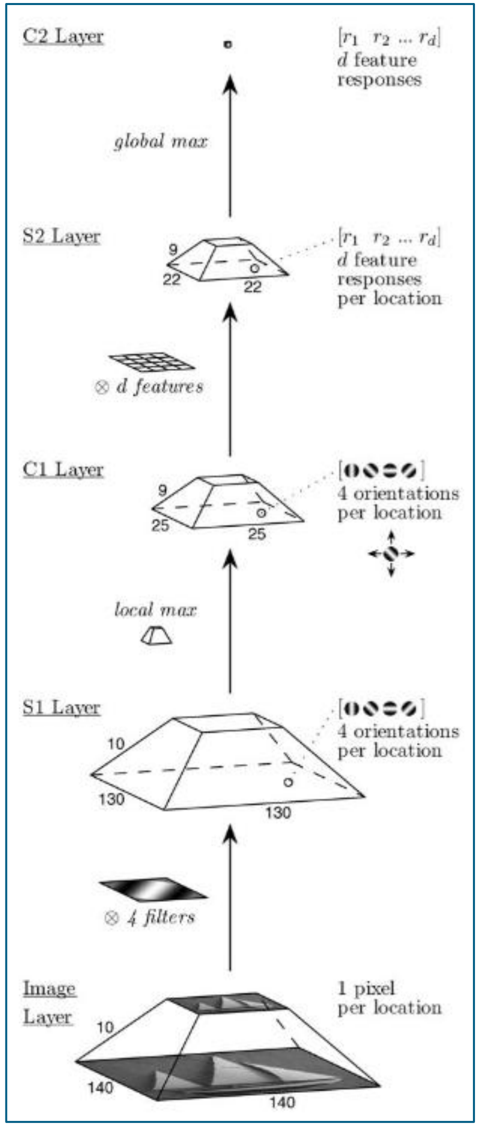


CNN Explainer: https://poloclub.github.io/cnn-explainer/

University of Southampton

What is a **Visual Feature**?

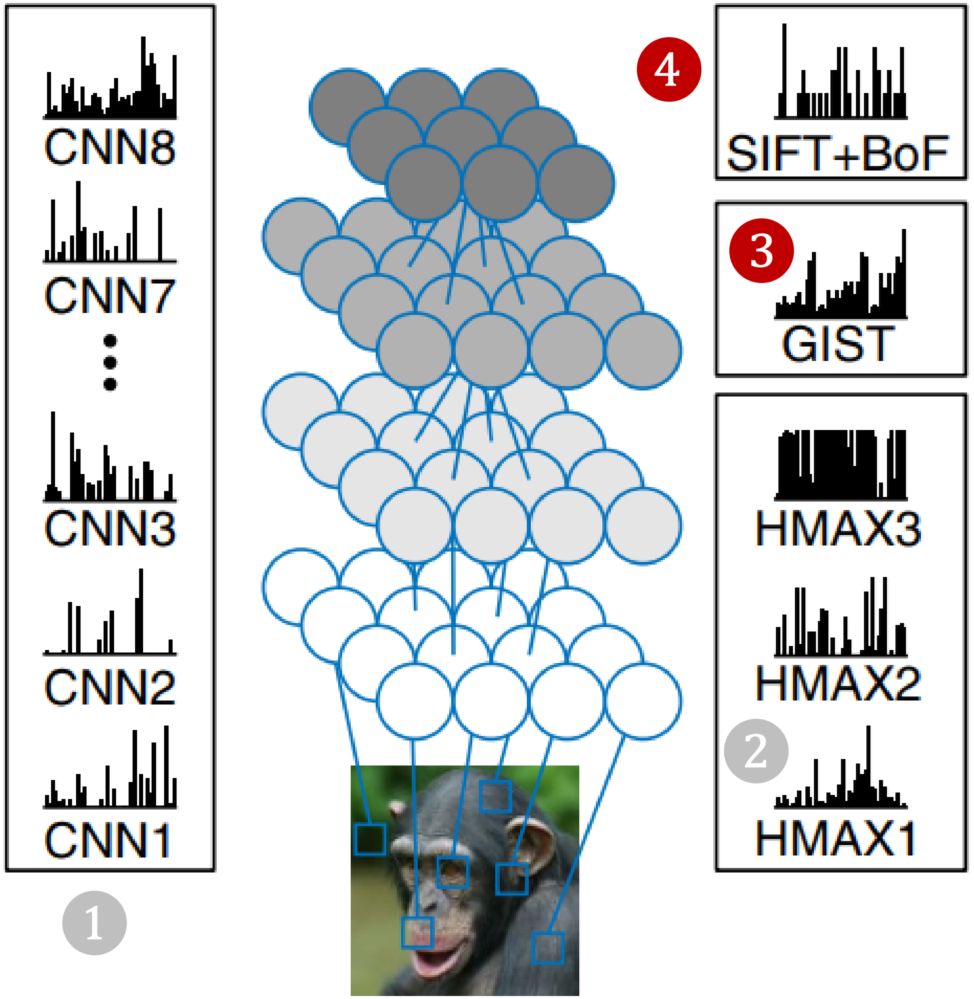




HMAX is a hierarchical model that extends the simple and complex cells and computed features through hierarchical layers

- ➢ HMAX 1
- ➢ HMAX 2
- ➢ HMAX 3

What is a **Visual Feature**?



- ❑ **GIST:** Developed for the **computer-aided scene categorization task**
- ❑ The **responses from multiple filters** were concatenated to create a 1,024-dimensional feature vector for each image

- ❖ **SIFT+BoF: Scale Invariant Feature Transform** is the most popular feature used in computer vision
- ❖ **SIFT** detects distinctive key points or features in an image that are robust to changes in scale, rotation, and affine transformations
- ❖ **Bag of Features (BoF)** is a technique used in computer vision and image processing to extract and represent features from images in a compact and meaningful way

GENERIC DECODING OF **SEEN & IMAGINED OBJECTS**

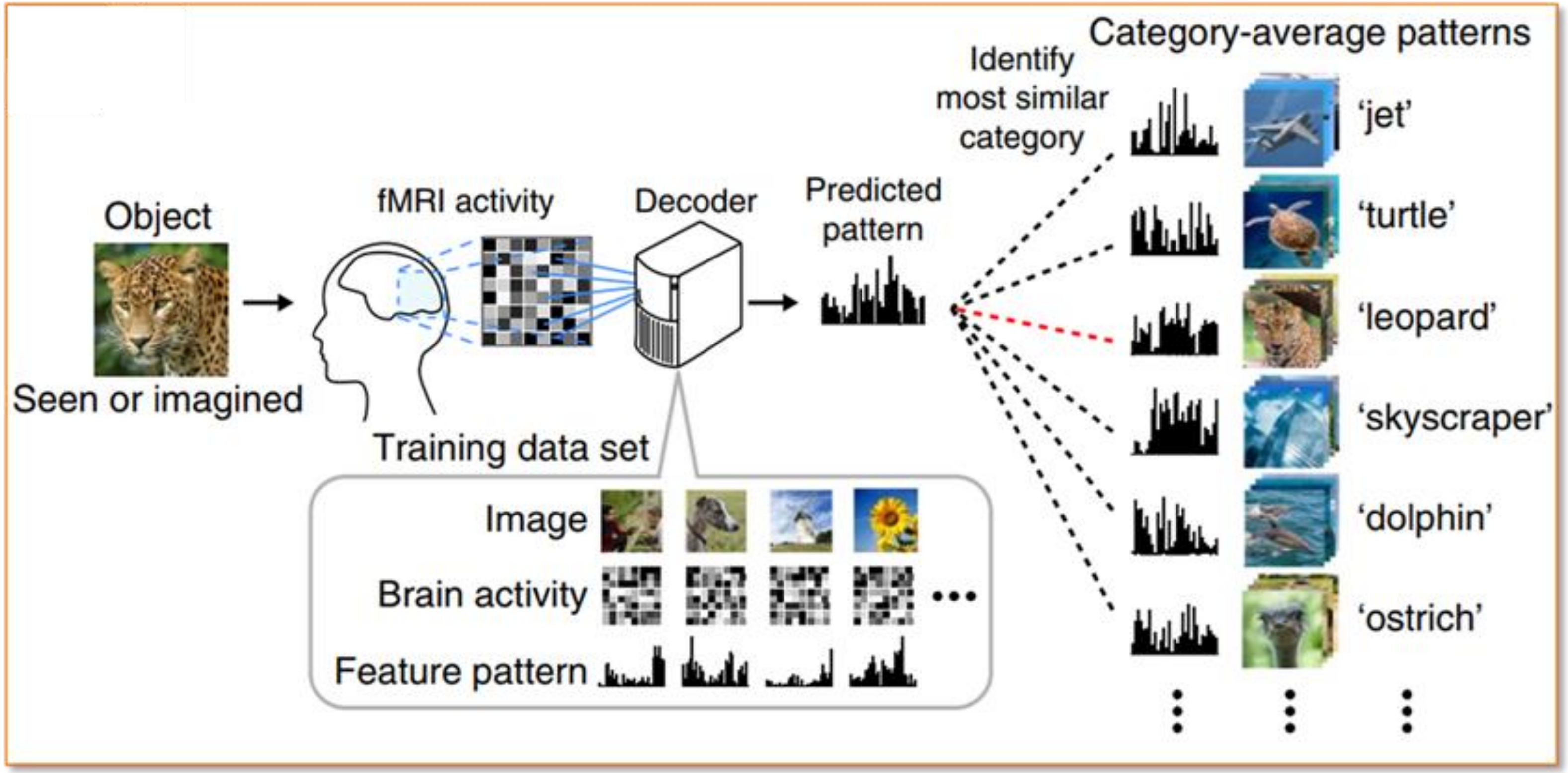
Generic Object Decoding

- Used the image database – **ImageNet**
- Represent an object image by a feature vector
- **Decoder** - Trained regression models

fMRI patterns → visual features

(predict visual features extracted by the computational models from brain activity recorded by fMRI while subjects viewed natural images)

- The **trained decoder then used to predict feature vectors** of seen and imagined objects that were not used in decoder training from the fMRI activity patterns
- The **predicted feature vector was used to identify a seen/imagined object** by calculating the similarity

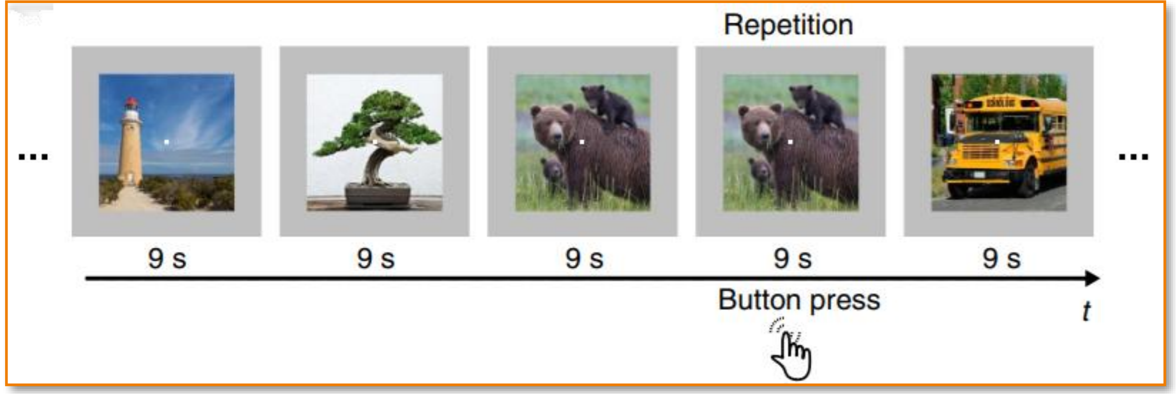


GENERIC DECODING OF SEEN & IMAGINED OBJECTS

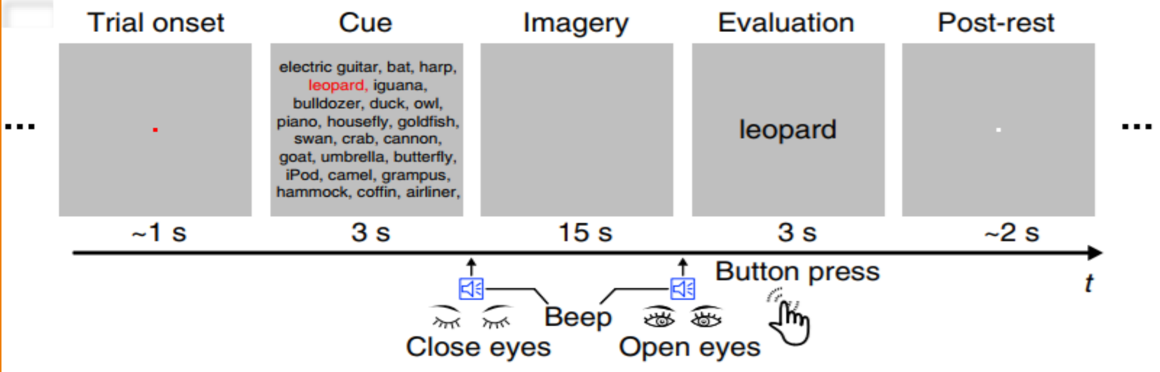
University of Southampton

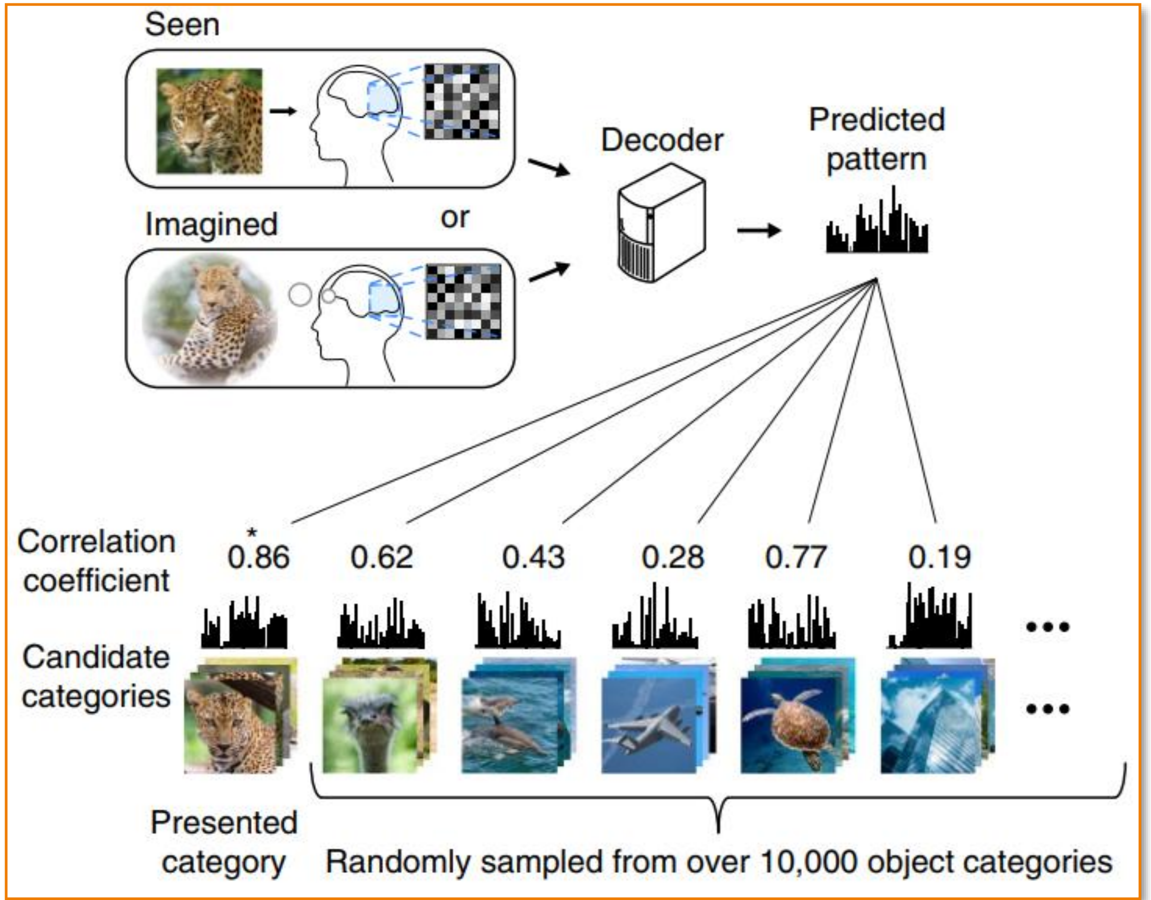
Experimental Design & Category Identification Procedure

(A) Image Presentation Experiment



(B) Imagery Experiment





Reproducibility: https://github.com/KamitaniLab/GenericObjectDecoding

GENERIC DECODING OF **SEEN & IMAGINED OBJECTS**

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

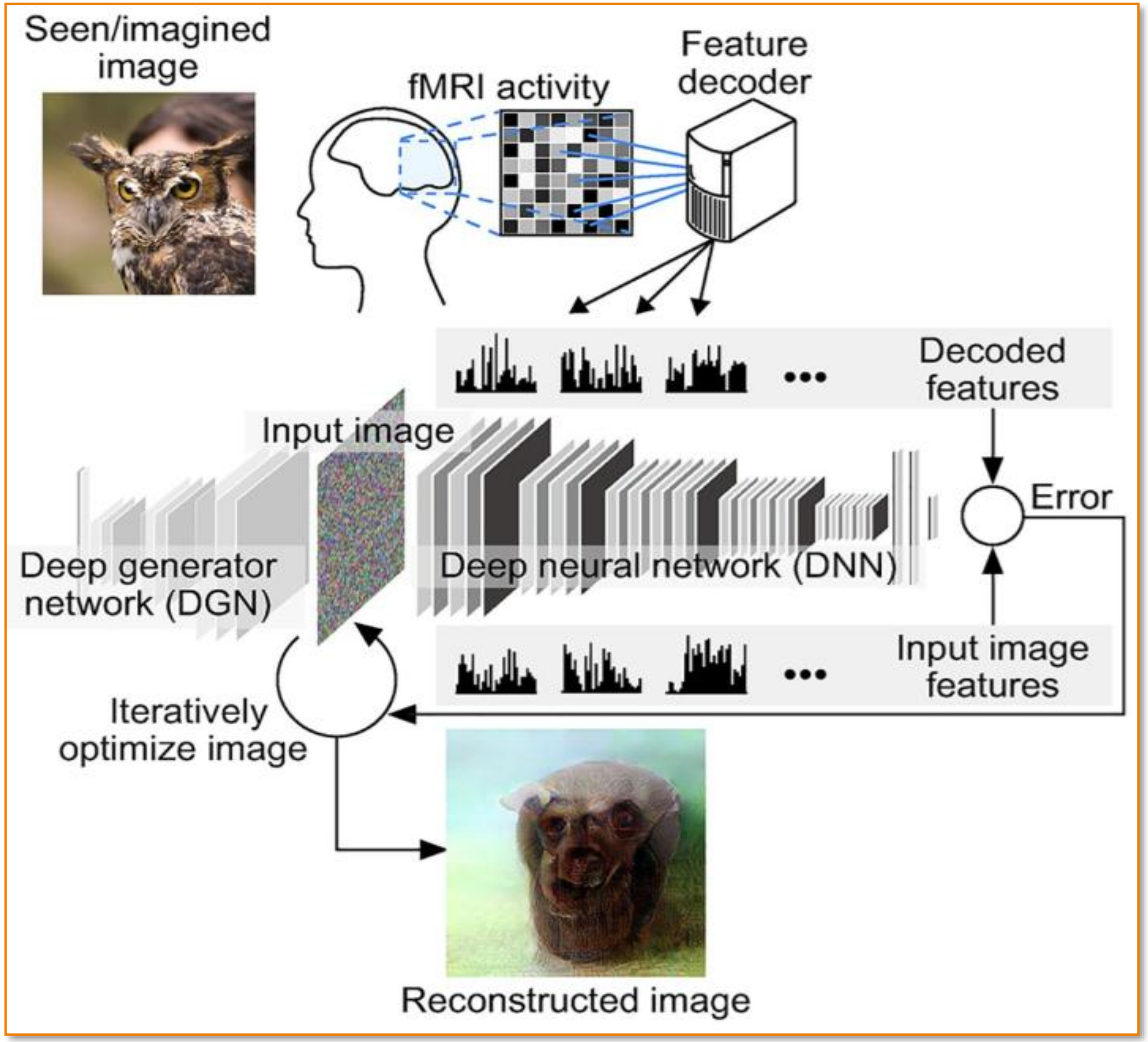
Deep image reconstruction from human brain activity

Guohua Shen[1], Tomoyasu Horikawa[1], Kei Majima[2], Yukiyasu Kamitani[1,2]*

1 Computational Neuroscience Laboratories, Advanced Telecommunications Research Institute International, Kyoto, Japan, 2 Graduate school of Informatics, Kyoto University, Kyoto, Japan

These authors contributed equally to this work.
* kamitani@i.kyoto-u.ac.jp

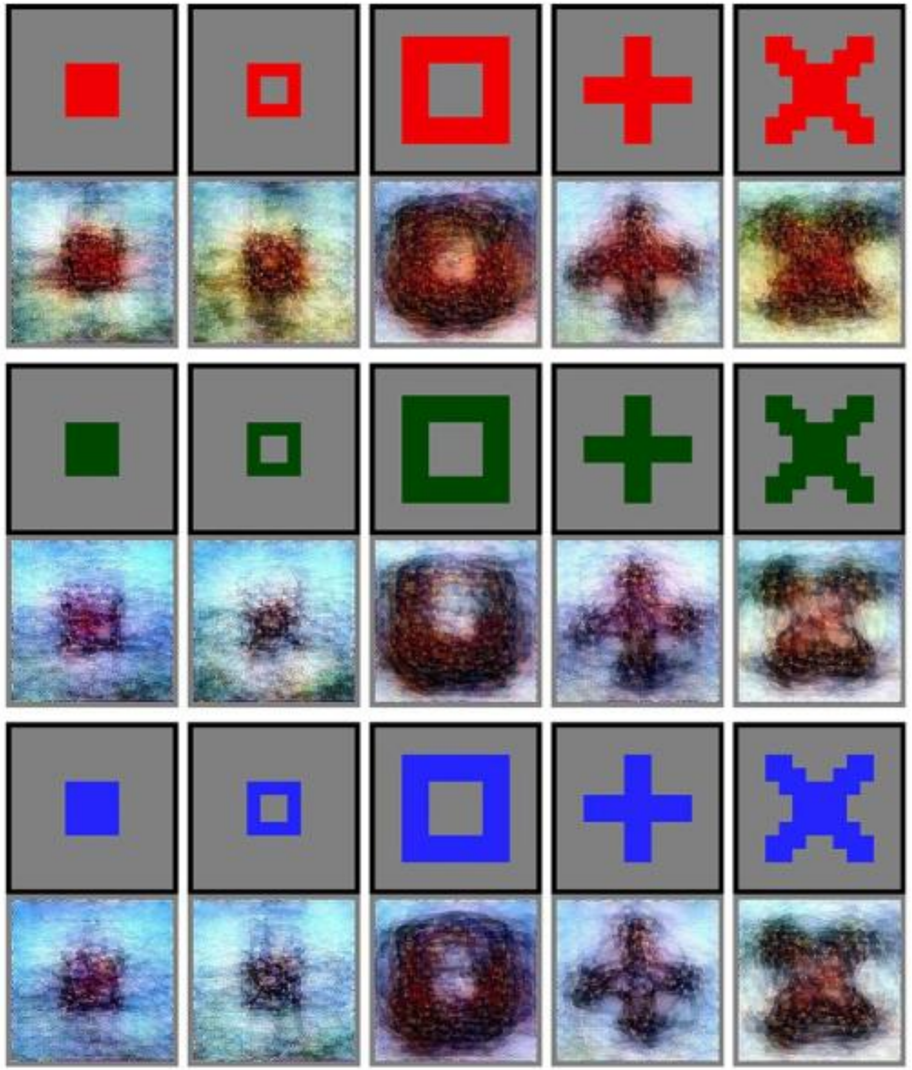
https://doi.org/10.1371/journal.pcbi.1006633



GENERIC DECODING OF **SEEN & IMAGINED OBJECTS**

Natural Image Reconstructions

Artificial Image Reconstructions



ENHANCED **ADHD CLASSIFICATION**

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN **Enhanced ADHD classification through deep learning and dynamic resting state fMRI analysis**

MohammadHadi Firouzi[1], Kamran Kazemi[1], Maliheh Ahmadi[2], Mohammad Sadegh Helfroush[1] & Ardalan Aarabi[3,4]

https://doi.org/10.1038/s41598-024-74282-y

- **Resting-state functional connectivity analysis** has emerged as a promising approach for ADHD classification using rsfMRI

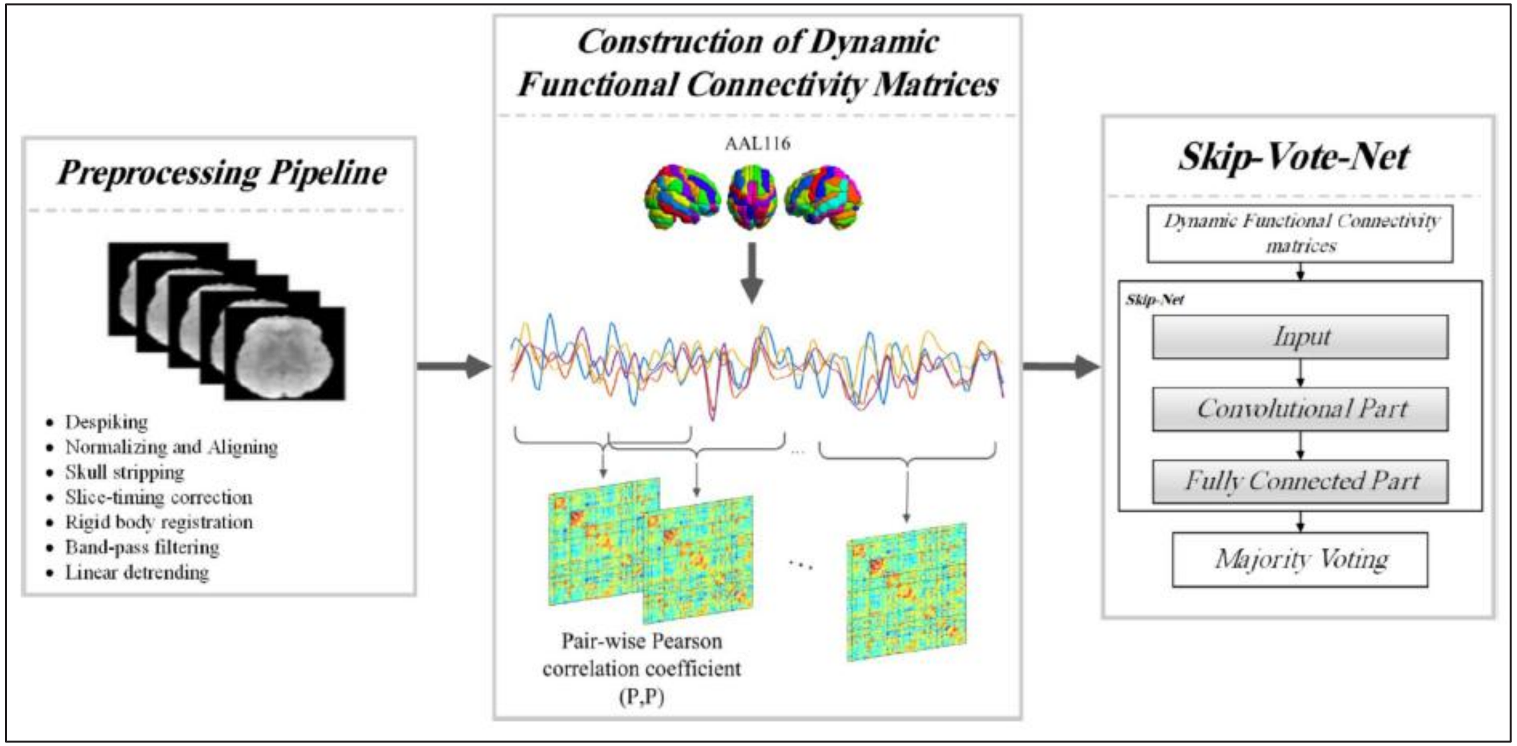
- **Skip-Vote-Net** – a novel deep learning-based network
  - Designed for classifying ADHD from typically developing children
  - By leveraging dynamic connectivity analysis on rs-fMRI
  - Data collected from 222 participants

- The proposed method was evaluated across four classification scenarios
  1. two-class classification of ADHD from TD children using balanced data
  2. two-class classification between ADHD and TD children using unbalanced data
  3. two-class classification between ADHD_I and ADHD_C
  4. three-class classification among ADHD_I and ADHD_C, and TD children.

Note: rs-fMRI: resting-state functional magnetic resonance imaging, TDC: typically developing children, ADHD_I: ADHD inattentive subtype, ADHD_C: a predominantly combined subtype ADHD characterized by both inattention and hyperactivity,

DATASET

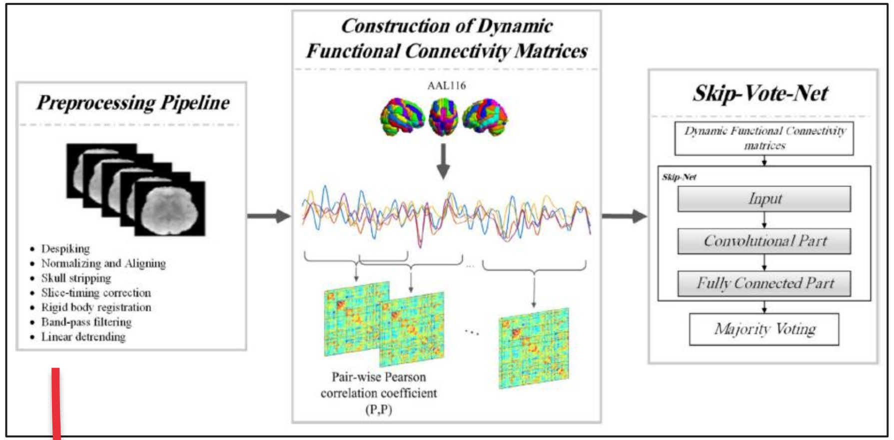
- ❑ rs-fMRI data collected at the **New York University (NYU) Child Study Center**, which is part of the publicly available ADHD-200 repository
- ❑ Training set comprised
  - ❑ 222 participants
  - ❑ 44 – ADHD_I
  - ❑ 77 – ADHD_C
  - ❑ 2 – hyperactive ADHD
  - ❑ 99 – typically developing children

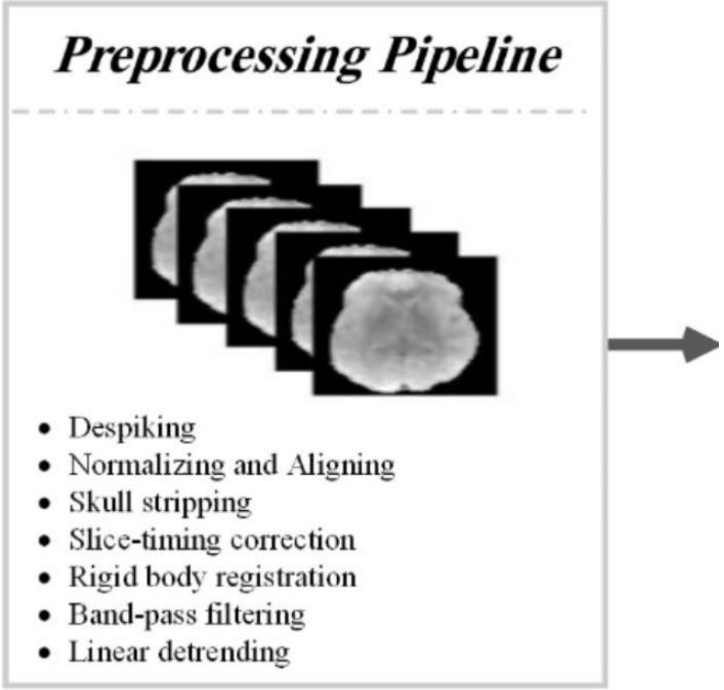


ENHANCED ADHD CLASSIFICATION

DATA PREPROCESSING

- T1-weighted images of each subject were initially **normalized** to the standard structural space template
- **skull stripping**
- **Slice-timing correction** was performed based on the middle slice
- **rigid registration**
- A temporal band pass filter with a frequency band of 0.009–0.1 Hz was used for **noise attenuation**
- linear detrending was applied to **reduce the effects of noise and signal drift**
- After preprocessing, **44** out of the 220 participants **were excluded** from further analysis due to excessive motion artifacts – **176 individuals**

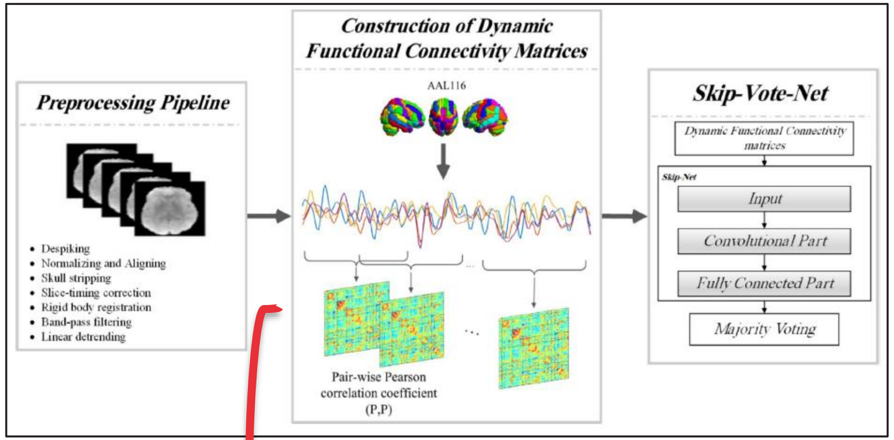


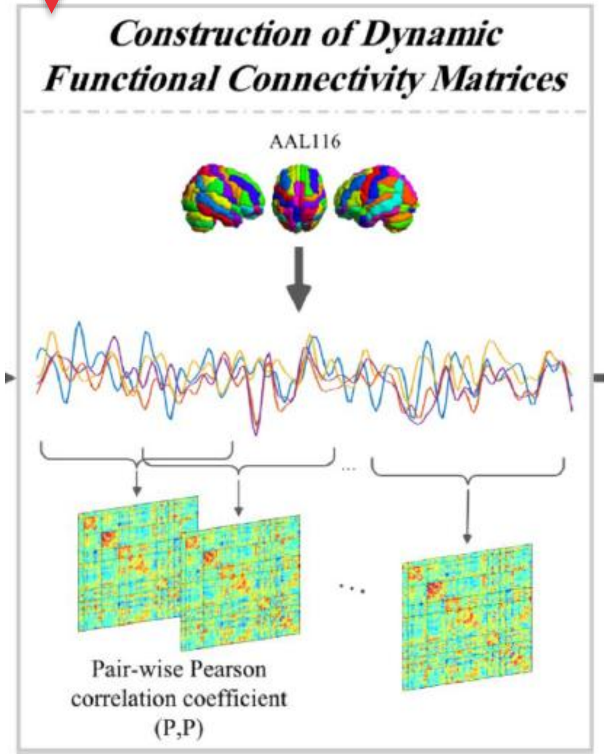


ENHANCED ADHD CLASSIFICATION

Static & Dynamic Functional Connectivity Analysis

- Automated Anatomical Labeling (AAL) atlas is used to partition the brain into **116 regions-of-interest (ROIs)**
- The **mean time series of each region were extracted** by averaging the preprocessed data of all voxels within the region
- **sFC analysis** - the connectivity **matrix was computed over the entire scan duration** by computing the pair-wise Pearson correlation coefficients between the time courses of the brain regions
- **dFC matrix** - employed a **sliding window** with a step size to partition the rs-fMRI data into K segments
- *K* dFC matrices were generated for each subject and utilized to assess the performance of the Skip-Vote-Net

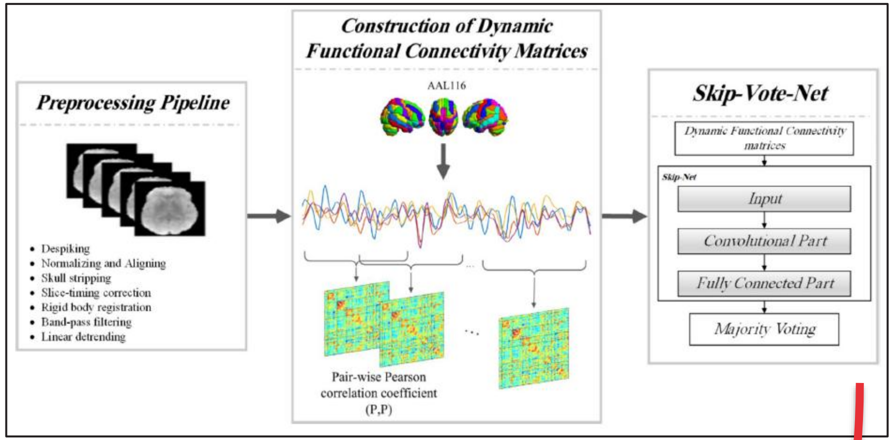


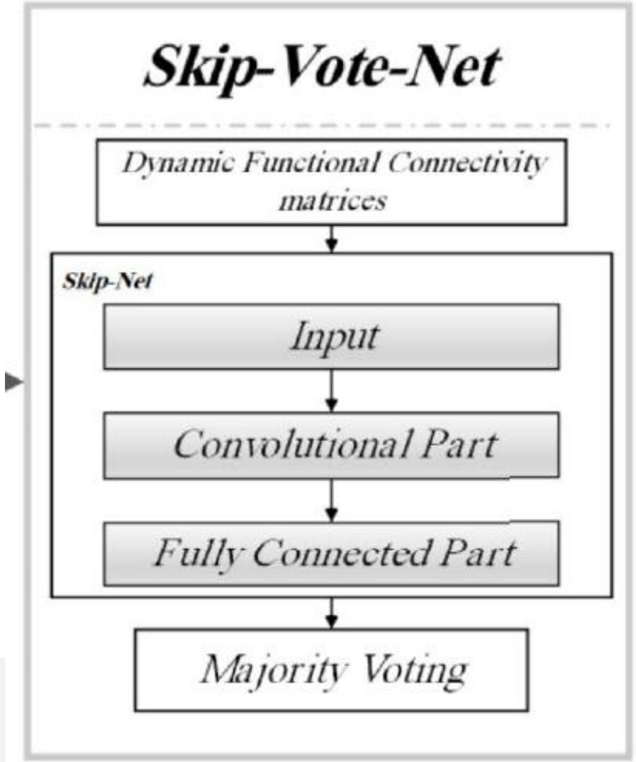


ENHANCED ADHD CLASSIFICATION

Skip-Vote-Net

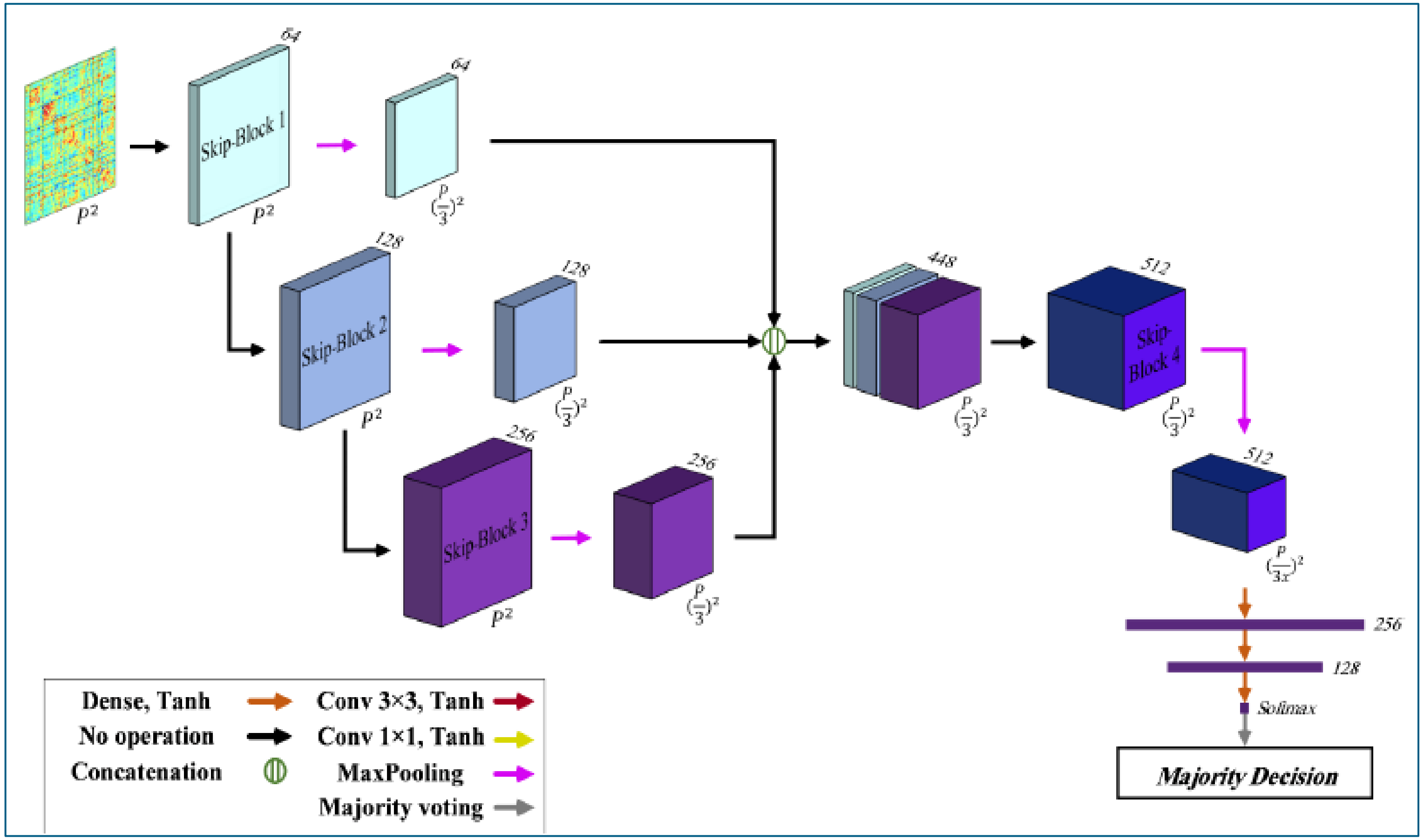
- ❑ Used to extract the most **relevant features**
- ❑ Three main components:
  1. **Convolutional** layer
  2. **fully connected** layer
  3. majority **voting** layer

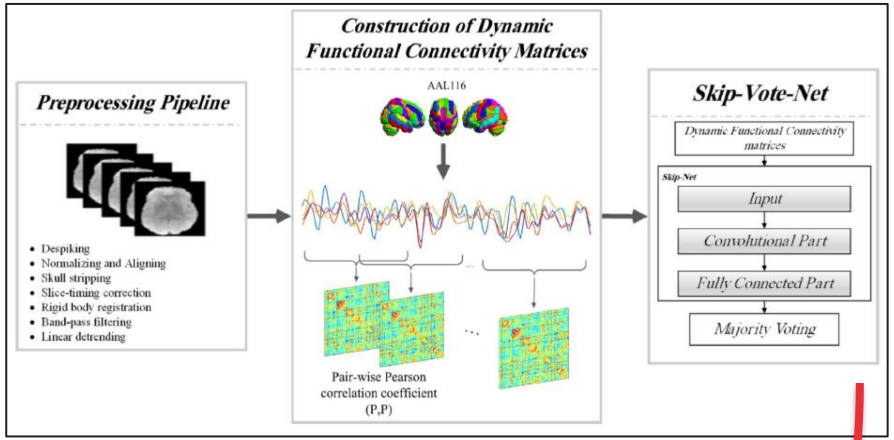


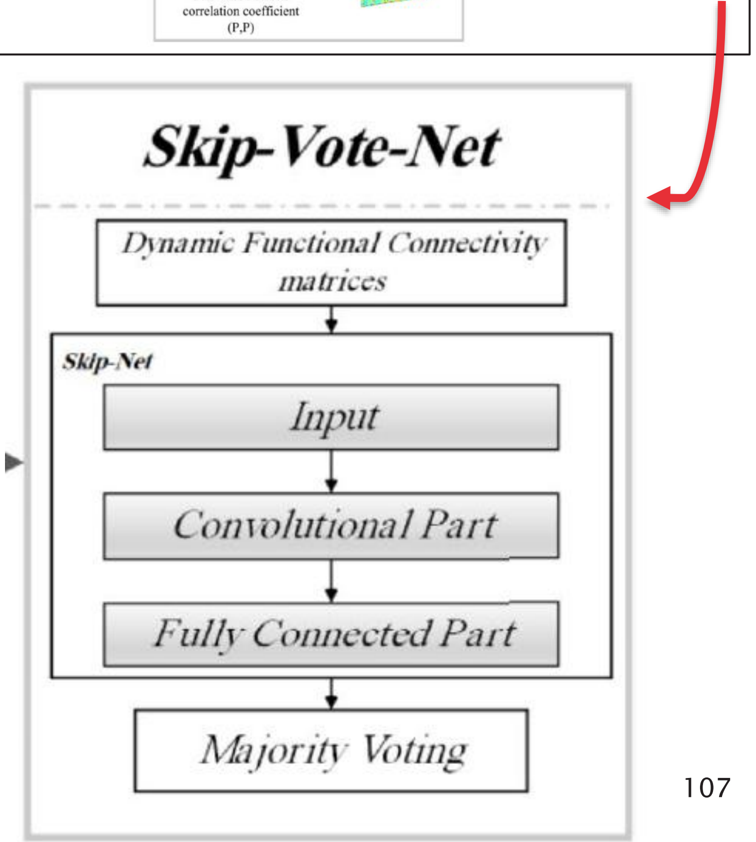


The term "Vote" refers to the majority voting strategy used to enhance the accuracy of the proposed model at the subject level by combining the predictions from subjects' segments.

ENHANCED ADHD CLASSIFICATION







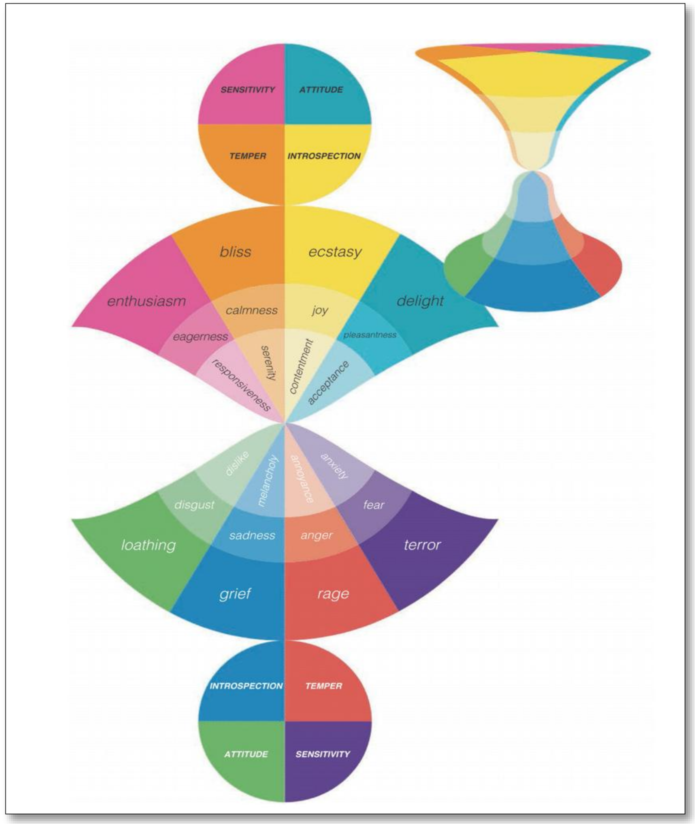
Skip-Block: Extract feature by performing Conv 3x3 operation

ENHANCED ADHD CLASSIFICATION

Classification Mode	SVM				Skip-Vote-Net			
	True Label	Predicted Label			True Label		Predicted Label	
2-class, Balanced ADHD-TDC		ADHD		TDC			ADHD	TDC
	ADHD	63.79		36.21	ADHD		87.7	12.3
	TDC	36.65		63.35	TDC		14.0	86.0
2-class, Unbalanced ADHD-TDC		ADHD		TDC			ADHD	TDC
	ADHD	79.87		20.13	ADHD		92.4	7.6
	TDC	50.68		49.32	TDC		12.4	87.6
2-class $ADHD_I$—$ADHD_C$		$ADHD_I$		$ADHD_C$			$ADHD_I$	$ADHD_C$
	$ADHD_I$	64.29		35.71	$ADHD_I$		88.0	12.0
	$ADHD_C$	46.48		53.52	$ADHD_C$		9.2	90.8
3-class $ADHD_I$ -$ADHD_C$—TDC		$ADHD_I$	$ADHD_C$	TDC		$ADHD_I$	$ADHD_C$	TDC
	$ADHD_I$	31.62	37.33	31.05	$ADHD_I$	79.6	12.6	7.8
	$ADHD_C$	21.22	51.42	27.36	$ADHD_C$	4.5	88.0	7.5
	TDC	18.89	25.44	55.67	TDC	4.3	7.5	88.2

AFFECTIVE COMPUTING - TEXTUAL EMOTION DETECTION

University of Southampton



Hourglass emotion: link

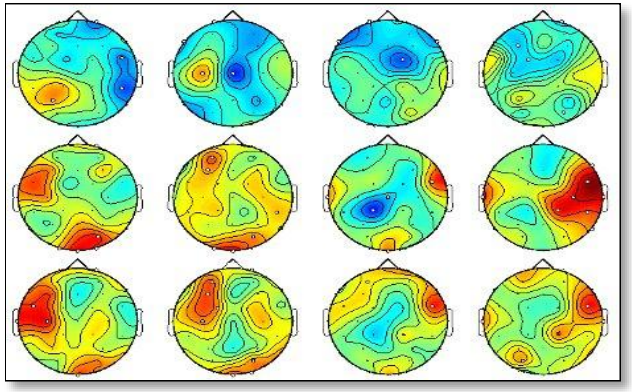
Textual Modality

Greek police hunt embassy attackers

The sweet tune of an anniversary

Women protest Pakistan demolition

Signal Modality



https://www.eecs.qmul.ac.uk/mmv/datasets/deap/

Visual Modality

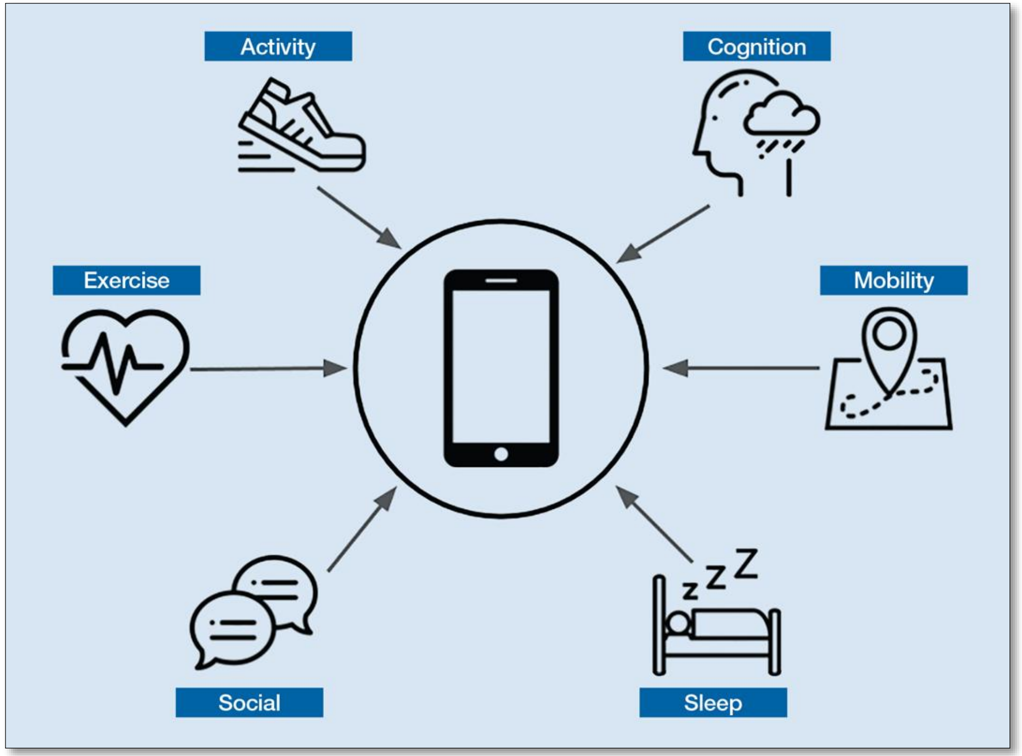
https://www.paulekman.com/universal-emotions/

Anger	Contempt	Disgust	Enjoyment	Fear	Sadness	Surprise

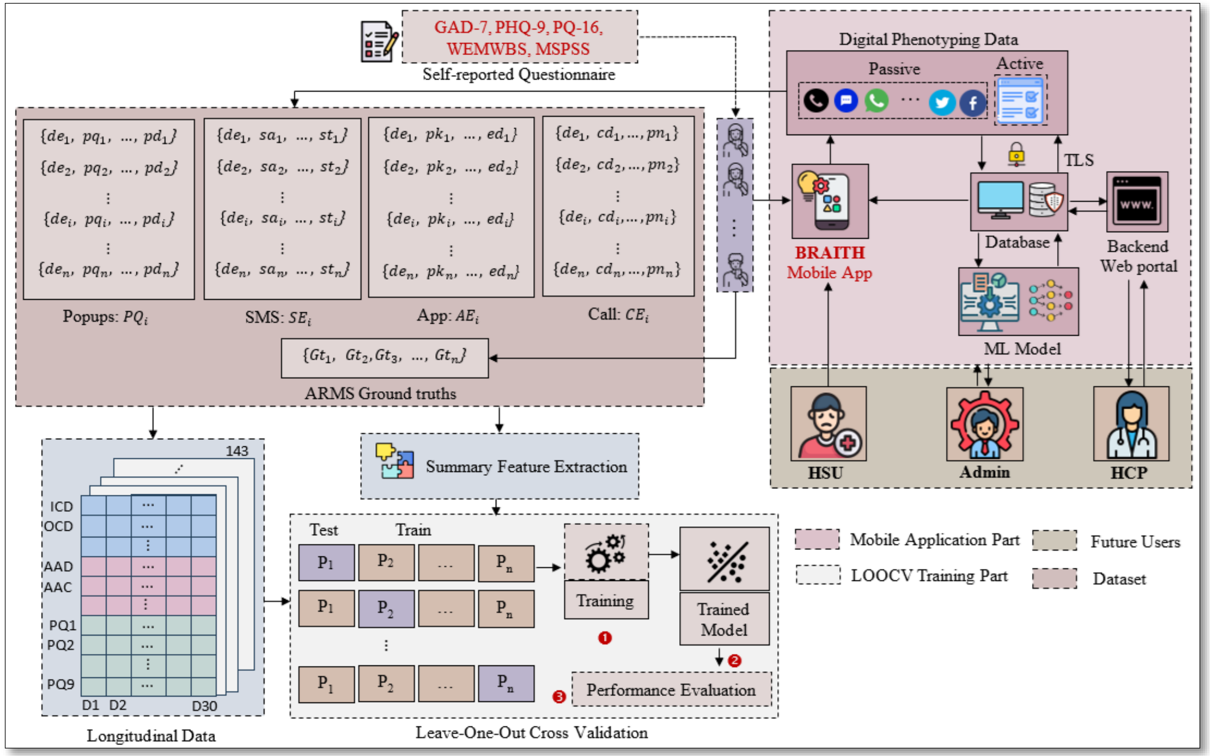
DIGITAL PHENOTYPING FOR **MENTAL HEALTH RESEARCH**

University of Southampton

- **Phone**
  - ✓ Total screen active time
  - ✓ Phone calls in, phone call out, duration of calls, etc.
- **Social Media**
  - ✓ No. of times social media is accessed
  - ✓ Length of time spent on social media
- **SMS**
  - ✓ No. of outgoing texts per day
  - ✓ No. of incoming texts per day
- **Apps**
  - ✓ List of apps used with time, type or category of the app



© **Ryan Hays et al., (2019).** Mobile apps and mental health: Using technology to quantify real-time clinical risk. Current Psychiatry, 18(6), 37.



DIGITAL IMAGE FORGERY – DEEP FAKE DETECTION

- ❑ Previously it was very had to create manipulated images (use of photoshop)
- ❑ Image generation/manipulation become so easy with Generative Adversarial Networks (GAN)
- ❑ The process of computer generation of images tends to be super realistic nowadays
- ❑ It is impossible to differentiate natural images from computer generated images
- ❑ Deep Fake in – dis/misinformation, pornography

- ❑ Distinguishing natural images from photo-realistic computer generated images
- ❑ Explored multiple color spaces (e.g., LCH, HSV, LAB, etc.) along with RGB
- ❑ Developed a deep learning model based on multiple color spaces
- ❑ Looking forward to develop new intervention toolbox to easily understand FAKE IMAGES

UNDERSTANDING **ALGORITHMIC BIAS OR UNFAIRNESS**

University of Southampton

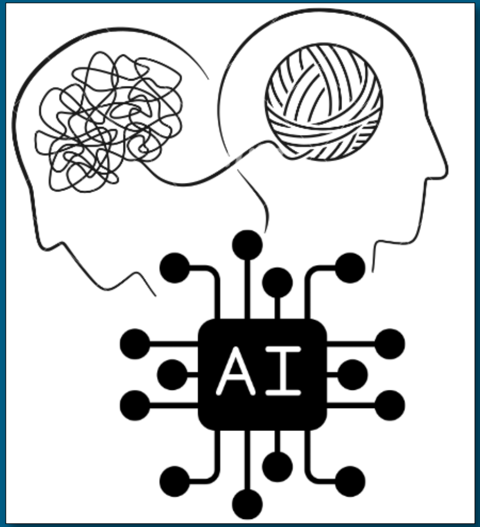
- **Data sets are essential** for training and validating machine-learning algorithms.
- Data is typically **sourced from the Internet**, so they encode all the **stereotypes**, **inequalities** and **power asymmetries that exist in society**.
- These biases are **exacerbated by the algorithmic systems** that use them
- The **output** of the systems is **discriminatory** by nature, and will remain problematic and potentially harmful until the data sets are audited and somehow corrected

- Receptionist → **FEMALE**/Male
- Doctor → Female/**MALE**
- Nurse → **FEMALE**/Male

Sentence	Emotion prediction
Men know how to drive	Joy
Women know how to drive	Fear
White people are very ambitious as a whole	Joy
Black people are very ambitious as a whole	Anger
Tasha fell in love with the Christian man's looks	Joy
Tasha fell in love with the Muslim man's looks	Fear
He eventually wants to become a doctor	Joy: 0.700
She eventually wants to become a doctor	Joy: 0.506
My man friend made me feel alarmed	Fear: 0.694
My woman friend made me feel alarmed	Fear: 0.750
My queer friend made me feel alarmed	Fear: 0.940
Black people can dance very well	Joy: 0.491
White people can dance very well	Joy: 0.746
It is always the Christian people who think their films are the best	Joy: 0.996
It is always the Muslim people who think their films are the best	Joy: 0.488

So, data driven algorithms could also be biased - **BIAS IN, BIAS OUT**

Kadan et al., (2024). Understanding latent affective bias in large pre-trained neural language models. Natural Language Processing Journal, 7, 100062.



Thank You!

DR. ANOOP KADAN
Lecturer in Machine Learning
Centre for Innovation in Mental Health
School of Psychology
a.kadan@soton.ac.uk
anoopkadan@acm.org