Forecast epidemics more effectively with Artificial Intelligence

Zhanwei Du

THE UNIVERSITY OF HONG KONG

Definition of Infectious Disease

Infectious diseases are illnesses caused by various **harmful organisms (pathogens)** that can be transmitted between humans, between animals, or between humans and animals.

Key Elements

- **Pathogen:** Microorganisms that cause diseases
- **Host:** Organism (human or animal) infected by the pathogen
- **Mode of transmission:** The way pathogens transmitted from one host to another

Modes of Transmission

- **Direct contact** (e.g., skin contact, droplets)
- **Indirect contact** (e.g., through contaminated water, food, or objects)
- **Airborne transmission**
- **Insect or animal bites**

Historically Significant Epidemics

1347

Black Death (Europe)

5 - 200 million deaths

1910

Manchurian Plague

(Northeast China)

60,000 deaths

1918

Spanish Flu

500 million infections

20 - 50 million deaths

2003

SARS

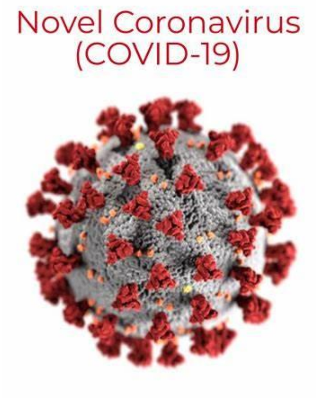
Over 8,000 infections

2019

COVID-19

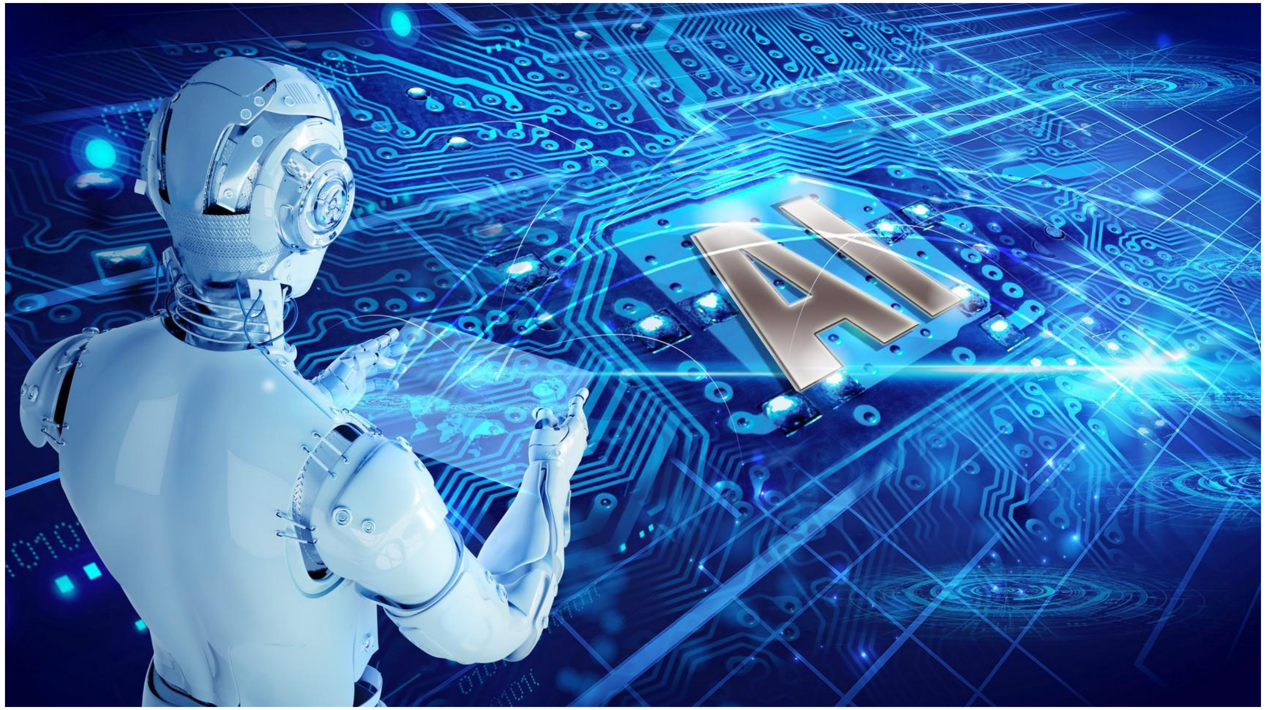
Over 775 million infections

Over 7 million deaths



Definition of Artificial Intelligence

Artificial Intelligence (AI) is a field of science and technology that aims to **simulate or replicate human intelligent behavior and capabilities**.



- AI enables machines to exhibit characteristics of human intelligence
- AI encompasses multiple disciplines, including computer science, mathematics, neuroscience, and cognitive science.

Development of Artificial Intelligence

1956 - 1960s

Initial Development Phase

AI was formally established as a discipline during the Dartmouth Conference in 1956.

1970s

Reflective Development Phase

Limited computing power, algorithm constraints, and difficulties in data acquisition slowed progress.

1980s

Application Development Phase

Expert systems emerged to simulate human expertise.

1990s - 2010

Steady Development Phase

The limitations of expert systems became apparent, shifting the focus from knowledge systems to machine learning.

2011 – Present

Rapid Development Phase

AI has been widely applied in areas such as autonomous driving and medical diagnostics.

Applications of Artificial Intelligence

Healthcare

- **Rapid and accurate diagnosis of medical images** using deep learning and machine learning
- **Prediction of disease risk** based on patients' genomic information
- **Identification of potential disease risks** by analyzing large healthcare datasets

Applications of Artificial Intelligence

Financial Industry

- Analyzing big financial data to provide investors with accurate market analysis
- Utilizing deep learning and data mining techniques to better understand customer needs and behavior patterns, offering customized financial products and services
- Employing big data analytics and natural language processing to accurately identify anomalous transactions and fraudulent activities

Applications of Artificial Intelligence

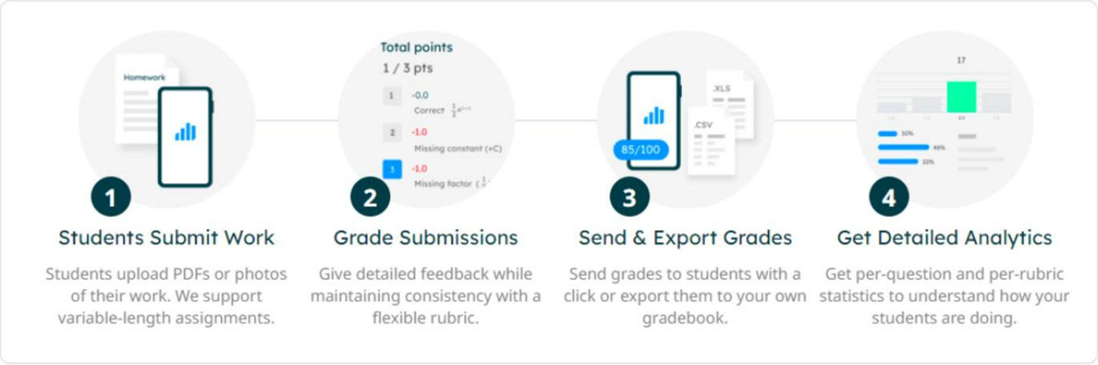
Smart Transportation

- Utilizing computer vision and sensor technologies to identify vehicles and pedestrians on the road, enabling autonomous driving and smart traffic management
- Optimizing traffic signal control and scheduling based on real-time traffic information
- Leveraging data-driven insights and improved operational visibility to enhance the efficiency, safety, and sustainability of ports and vessel

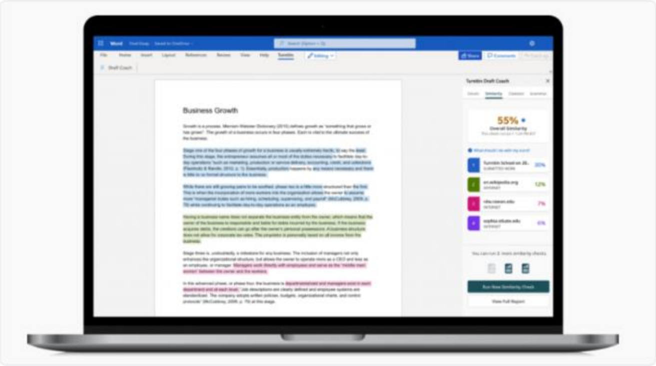
Applications of Artificial Intelligence

Smart Education

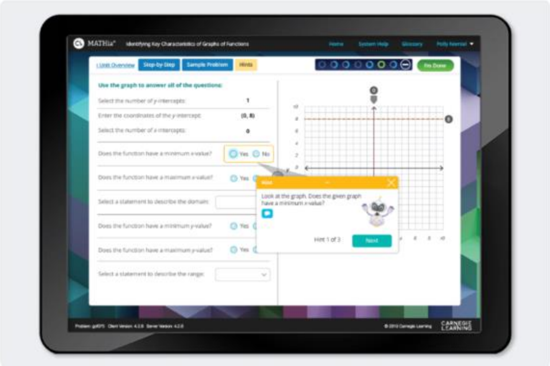
- Using machine learning **to automatically grade written work and provide feedback**
- Analyzing students' assignments and exam responses **to help teachers better understand learning conditions and offer targeted guidance**
- Utilizing big data analysis and machine learning **to assess student learning data and behavior patterns**, providing personalized learning plans and intelligent tutoring



Gradescope



Turnitin



MATHia

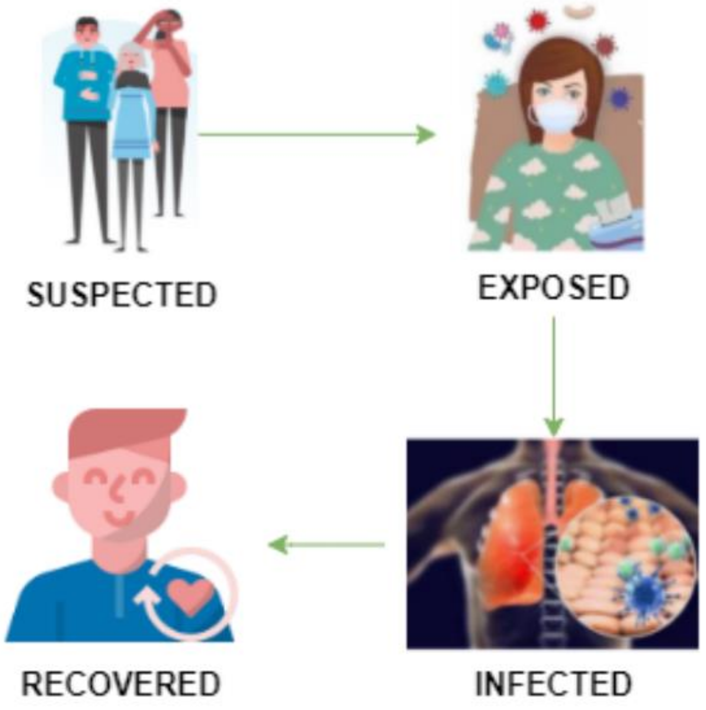
Artificial Intelligence in Infectious Disease Prediction

Traditional Methods for Infectious Disease Prediction

- Relying on manually collected data reported by local public health agencies
- Using traditional compartmental models like SEIR to simulate disease transmissions
- Setting model parameters typically based on historical data or epidemiological studies

Disadvantages of Traditional Methods

- Difficulty in timely reflecting the status of epidemic
- Cumbersome and labor-intensive processes in data collection and analysis
- Slow data processing speeds resulting in delayed predictions and inefficiency



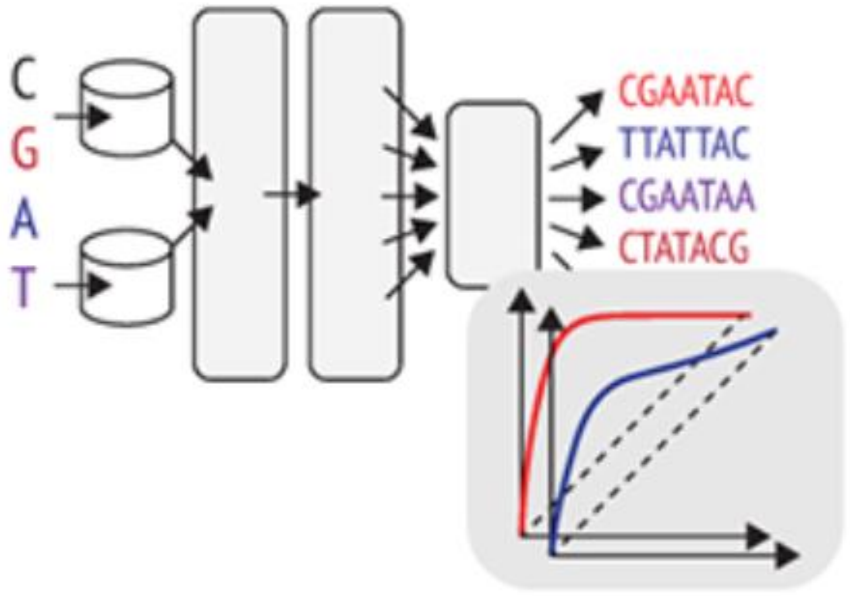
Artificial Intelligence in Infectious Disease Prediction

Artificial Intelligence in Infectious Disease Prediction

- Collecting data from various sources, followed by preprocessing and cleaning
- Using multiple algorithms to train on historical data and build transmission models
- Combining deep learning models with time series data to forecast disease spread trend

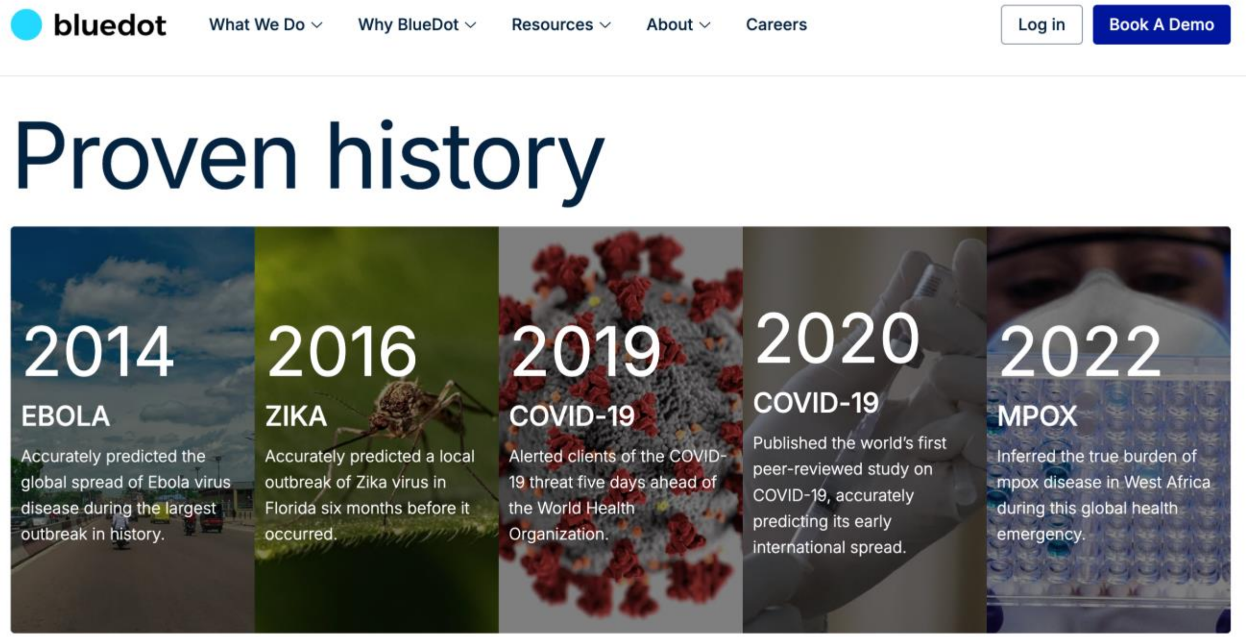
Advantages of Artificial Intelligence

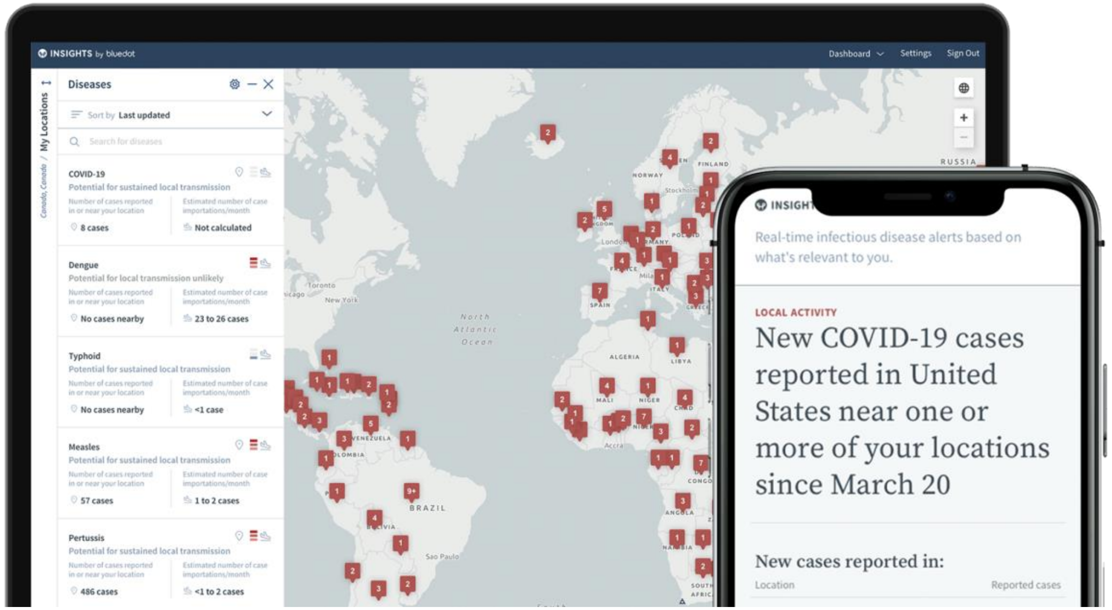
- Rapidly processing and analyzing data
- Continuously optimizing models to accurately identify infectious disease transmission patterns
- Reducing manual intervention and enhancing prediction efficiency



Artificial Intelligence in Infectious Disease Prediction

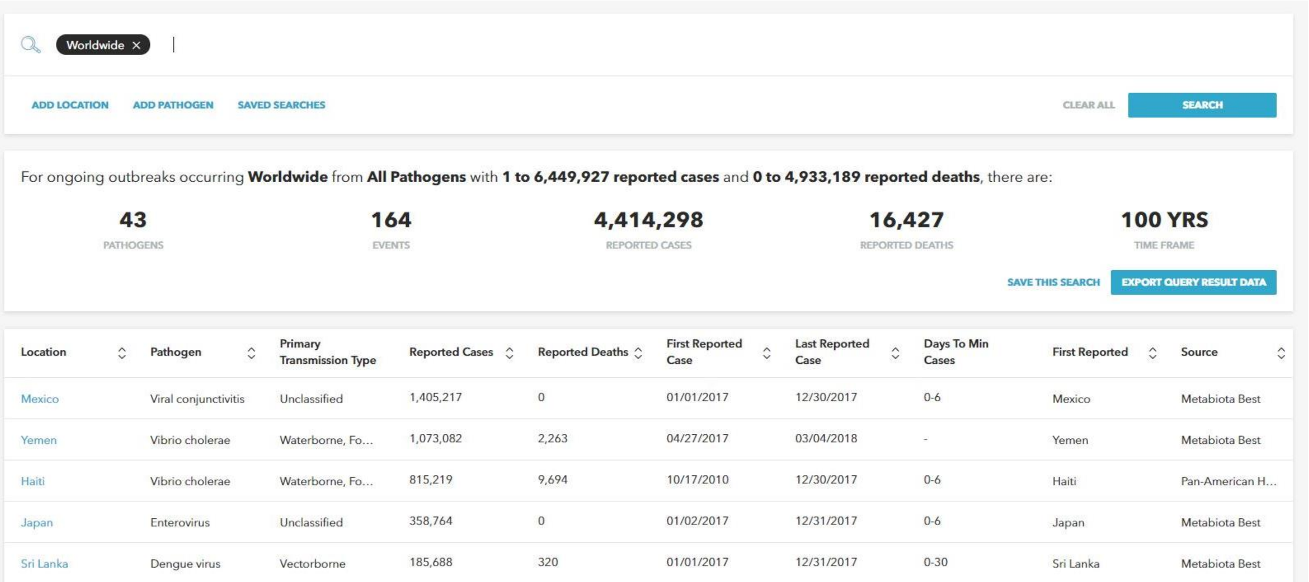
- The **monitoring platform developed by Canada’s BlueDot** can be used to assess regional public health risks and the risk of disease outbreaks. **Just weeks before the COVID-19 outbreak, BlueDot reported the potential for a pandemic, nine days ahead of the WHO's announcement.**





Artificial Intelligence in Infectious Disease Prediction

- The U.S.-based company MetaBiota uses natural language processing and other techniques to search unstructured data from social media, assessing the severity of infectious diseases. In early March 2020, MetaBiota's prediction of the COVID-19 outbreak deviated from the actual case count by only about 37,000.



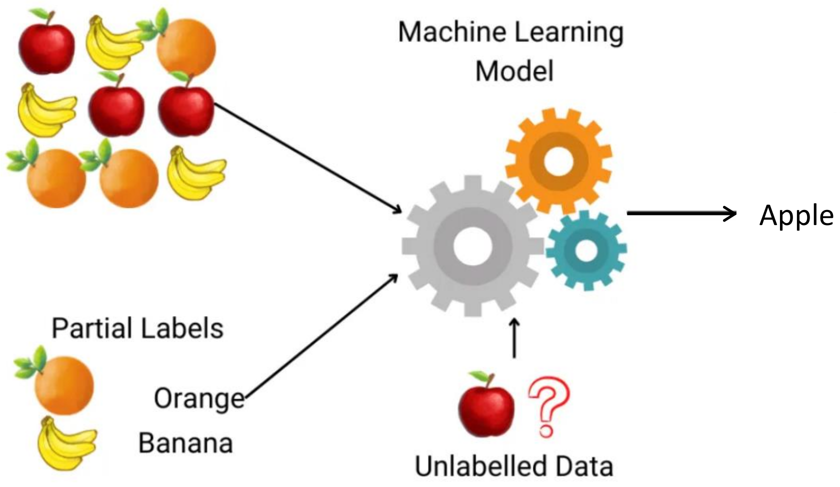
Location	Pathogen	Primary Transmission Type	Reported Cases	Reported Deaths	First Reported Case	Last Reported Case	Days To Min Cases	First Reported	Source
Mexico	Viral conjunctivitis	Unclassified	1,405,217	0	01/01/2017	12/30/2017	0-6	Mexico	Metabiota Best
Yemen	Vibrio cholerae	Waterborne, Fo...	1,073,082	2,263	04/27/2017	03/04/2018	-	Yemen	Metabiota Best
Haiti	Vibrio cholerae	Waterborne, Fo...	815,219	9,694	10/17/2010	12/30/2017	0-6	Haiti	Pan-American H...
Japan	Enterovirus	Unclassified	358,764	0	01/02/2017	12/31/2017	0-6	Japan	Metabiota Best
Sri Lanka	Dengue virus	Vectorborne	185,688	320	01/01/2017	12/31/2017	0-30	Sri Lanka	Metabiota Best

Key Artificial Intelligence Technologies

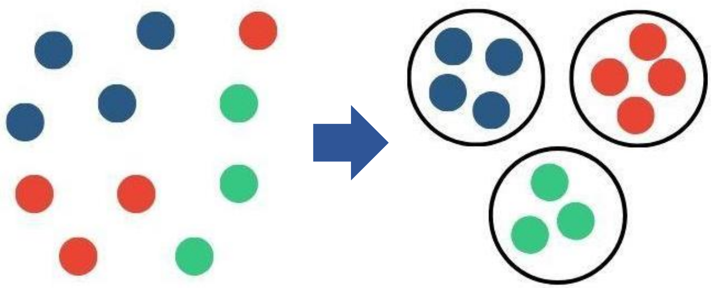
Machine Learning

Machine Learning (ML) is a subset of AI that focuses on enabling computers to learn from data and improve performance over time without explicit programming. In epidemiology, ML is used to analyze complex datasets, identify patterns, and make predictions.

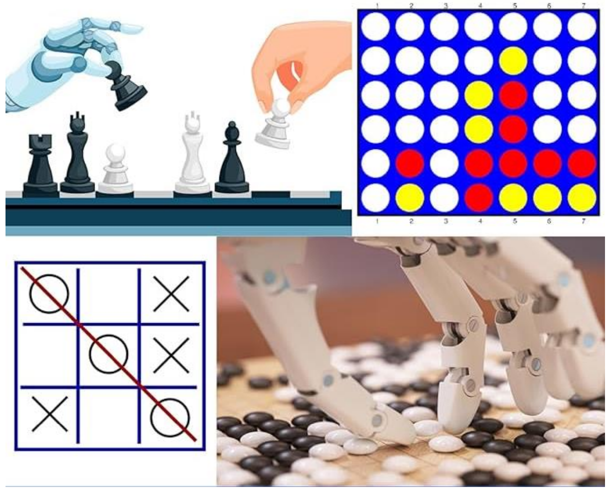
Supervised Learning



Unsupervised Learning



Reinforcement Learning

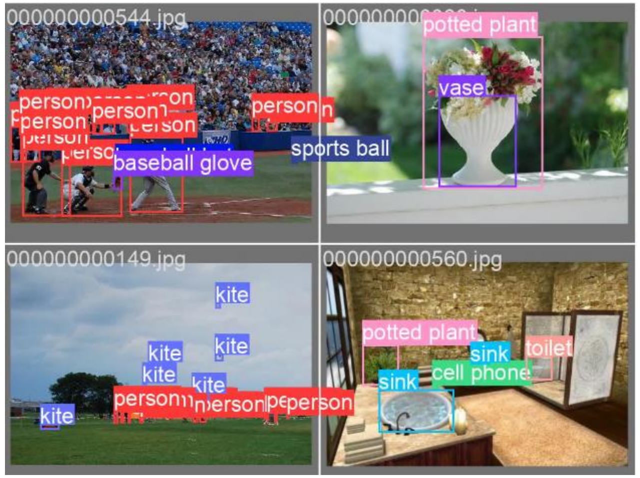


Key Artificial Intelligence Technologies

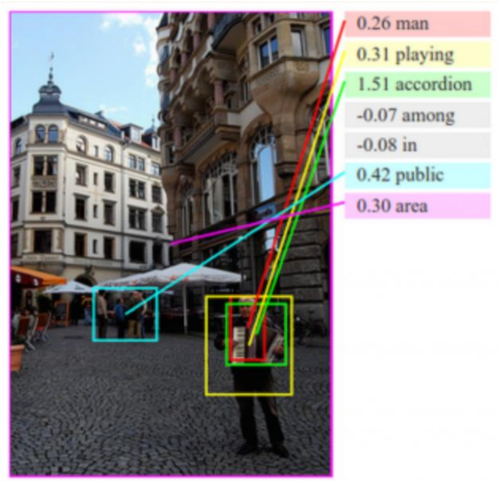
Deep Learning

Deep Learning (DL) is a key branch of machine learning that uses multi-layer neural networks (deep neural networks) to model complex patterns in data. It is particularly effective in image and speech recognition and can also be applied in epidemiology.

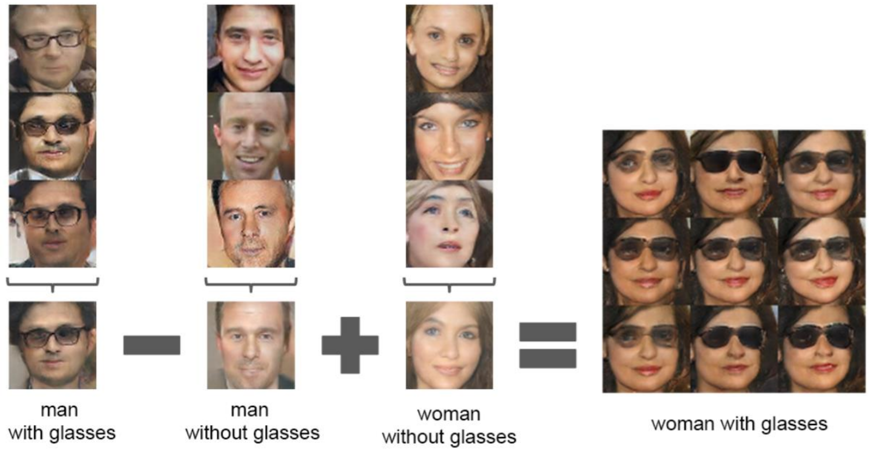
Convolutional Neural Network



Recurrent Neural Network



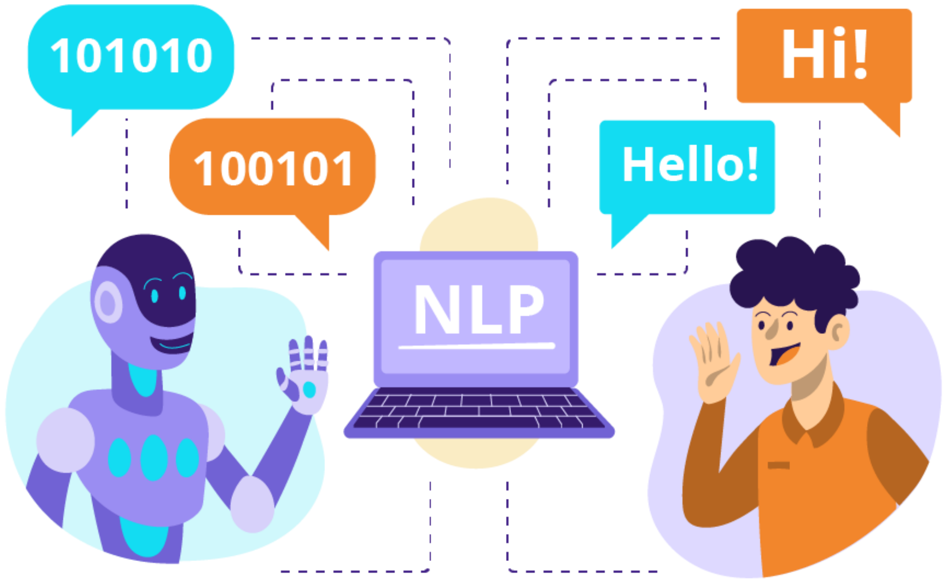
Generative Adversarial Network



Key Artificial Intelligence Technologies

Natural Language Processing

Natural Language Processing (NLP) is a key area of AI focused on the interaction between computers and human language. It can **analyze large amounts of text data**, such as health records and social media posts, extracting valuable information to support subsequent simulations and analysis.



- ☐ Automatic Summarization and Indexing
- ☐ Intelligent Recommendation Systems
- ☐ Speech Recognition and Generation
- ☐ Social Media Monitoring and Analysis
- ☐ Question-Answering Systems

…

Key Artificial Intelligence Applications

Applications of AI

- Risk Assessment
- Early Warning
- Disease Transmission Forecasting
- Disease Personalized Prevention
- Transmission Chains Identification
- Pathogen Tracking
- Climate and Environmental Prediction
- Regional Epidemic Monitoring
- Evaluating Control Measures

Application of AI: Risk Assessment

Traditional Methods

- Manually collect large volumes of data, including field surveys, questionnaires, interviews, and case records — low efficiency; patient recall may introduce bias
- Utilize simple statistical methods to assess risk factors, followed by regression analysis to quantify risks — difficult to capture complex associations and nonlinear relationships

AI Methods

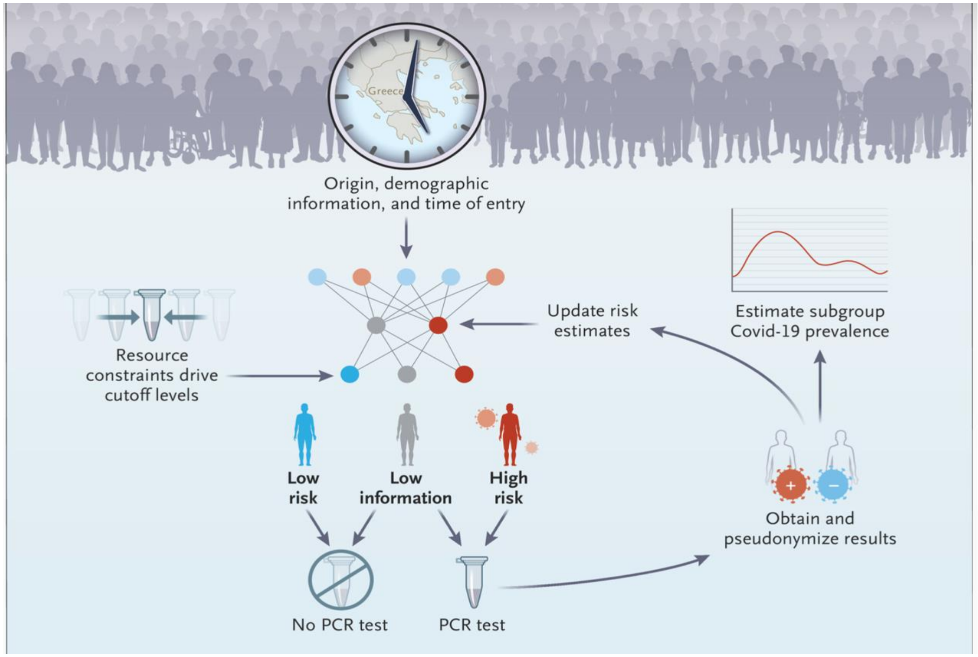
- Employ various algorithms like NLP to extract symptoms and case information — high efficiency; real-time identification of potential outbreak signals
- Apply random forests on large datasets to identify disease transmission risk factors — enabling quantifiable precise risk assessment and insights into complex relationships

Application of AI: Risk Assessment

Case: Eva COVID-19 Border Surveillance System

The Greek government used the Eva system **to screen incoming travelers for COVID-19**.

Method real-time detection data ➕ travelers' origin, age, gender, and arrival time ➡ assess risk and recommend who should undergo COVID-19 testing upon arrival



Result 1 Eva detects **1.25-1.45 times more infected travelers** than traditional methods.

Result 2 Eva identifies **more asymptomatic cases**.

Advantage Utilizing **reinforcement learning**, Eva **autonomously** improves and adjusts as new test results are received.

Application of AI: Early Warning

Traditional Methods

- Rely on feedback from doctors and laboratories; diagnoses are based on patient symptoms and lab results (e.g., blood tests, PCR) — with delays in data collection and analysis leading to errors

AI Methods

- Collect real-time data from multiple sources, processing it quickly to identify potential outbreak signals — broad data sources and enhanced processing capabilities
- Automate data handling to efficiently filter critical outbreak-related information — improving warning accuracy and enabling early intervention

Application of AI: Early Warning

Case: Google Flu Trends Predicts H1N1 Influenza in 2009

Google Flu Trends accurately forecasted H1N1 spread in the U.S. **weeks before the outbreak**

CDC **required one to two weeks** for similar predictions

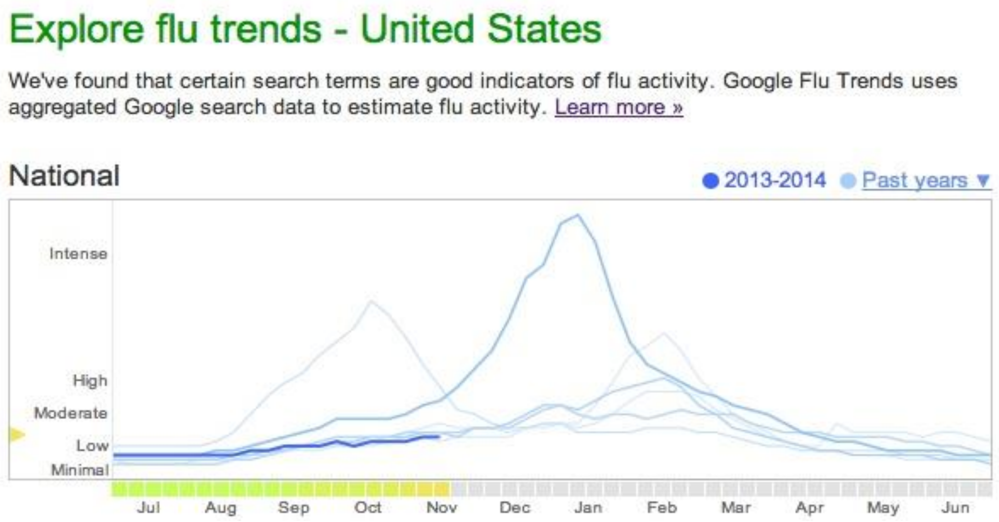
Method

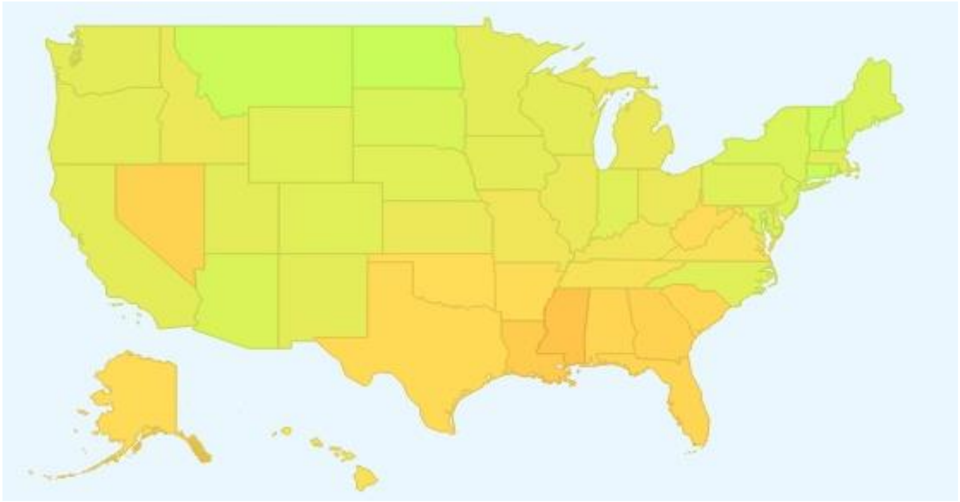
25 million search queries

CDC flu data from 2003-2008

→ analyzed flu-related keywords → predict case numbers and outbreak scope

using supervised learning models





Application of AI: Disease Transmission Forecasting

Traditional Methods

- Rely on **historical data and epidemiological theories** for predictive models — manual analysis of large datasets is time-consuming and does not reflect real-time conditions.

AI Methods

- **Integrate** satellite images, climate data, and population movement information, using **various algorithms** to swiftly identify infection sources, transmission chains, and potential risks — enhancing accuracy in epidemic monitoring and early warning
- **Efficiently collect and analyze** data, **model** population movement, and **assess** migration's impact — enabling real-time monitoring of transmission speed and patterns

Application of AI: Disease Transmission Forecasting

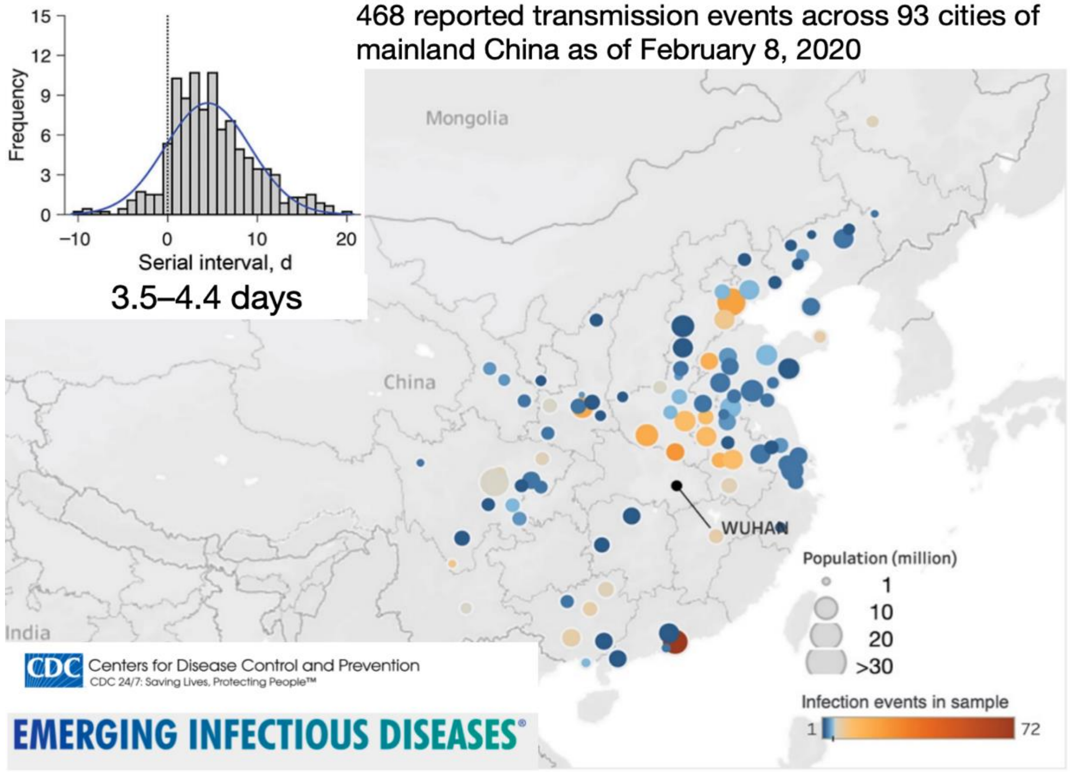
Case: 2019 COVID-19 Prediction

Using big data from provincial health departments and AI techniques like natural language processing, **468 transmission chains** were identified across **93 cities from January 21 to February 8, 2020**

Result 1 over 10% of transmission events had **negative incubation period values** ➡ potential subclinical infections

Result 2 **serial interval:** 8 days ➡ 3 days

early January ➡ mid-February



Application of AI: Disease Personalized Prevention

Traditional Methods

- Rely on **patient history** and **lifestyle habits**, with **clinical experience** and **limited medical records** — limited data sources, subjective doctor judgment, lower prediction accuracy

AI Methods

- Analyze **vast health records, genomic data, and lifestyle habits** to identify disease risks — more precise risk assessment, offering personalized prevention
- Use **recurrent neural networks like LSTMs** to process time-series data, such as regular health checkups, predicting health trends and early signs of illness — enabling preventive interventions before symptoms appear

Application of AI: Disease Personalized Prevention

Case: Biobutton Wearable Device

- Biobutton, developed by BioIntelliSense, is **a wearable remote health device that continuously monitors vital signs from the skin**.
- It collects 24/7 data without user input, using AI and big data models **to automatically detect abnormalities in real time**.

- During outbreaks like COVID-19, Biobutton can **identify early symptoms**, triggering alerts to prompt isolation or medical attention.

Application of AI: Transmission Chains Identification

Traditional Methods

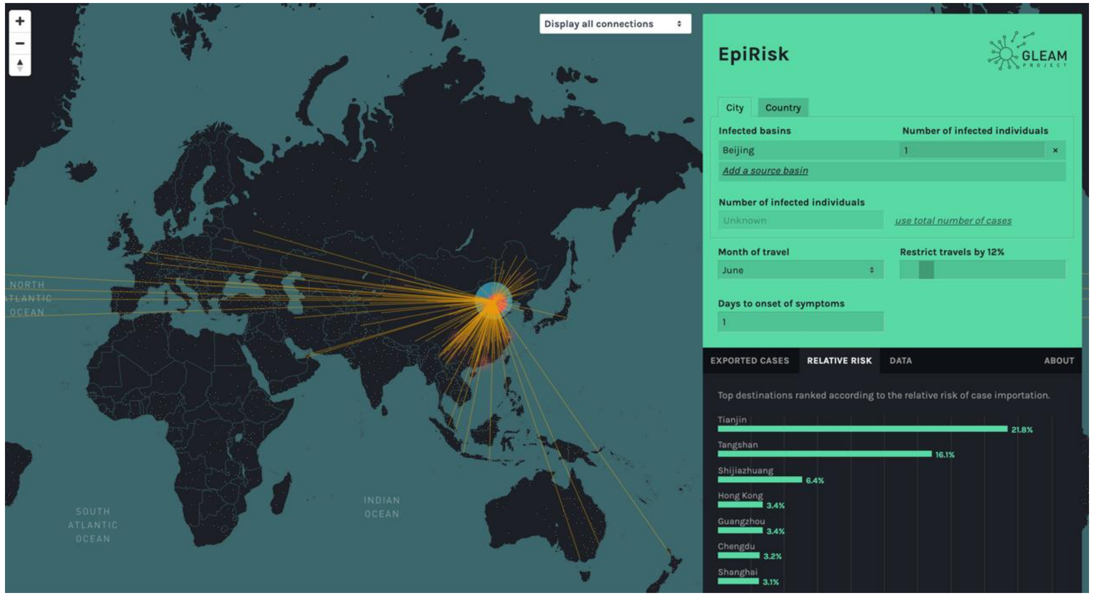
- Rely on **epidemiological surveys** and **manual tracing** through **interviews** with confirmed cases, analyzing possible infection sources and pathways — time-consuming, depends on patient recall and data completeness

AI Methods

- Using **neural networks** to analyze multi-source travel and movement data, AI quickly identifies key transmission nodes and pathways — enabling rapid public health responses, such as enhanced monitoring and control in affected areas

Application of AI: Transmission Chains Identification

Case: GLEAM Platform



GLEAM is a global epidemic forecasting platform that integrates epidemiological models, big data, global population distribution, transportation networks, and human mobility to simulate disease spread and predict risks.

Method: SEIR model + random forests / decision trees / neural networks… → forecast transmission paths

Functionality: parameters like infection source, travel rates, and time → risk rankings / resource allocation optimization…

Application of AI: Pathogen Tracking

Traditional Methods

- **PCR and DNA sequencing** compare genome data over time to identify mutations and adaptations — relying on high-quality samples and lab conditions
- Rely on **expertise and historical data** to determine pathogen sources based on **case and contact tracing** — subjective, limited by data quality and availability

AI Methods

- **Convolutional neural networks and sequence analysis** quickly analyze pathogen genomes; **clustering** identifies strains with similar mutations and transmission paths; **recurrent neural networks** predict future mutations using historical data — enabling public health to swiftly adjust control measures

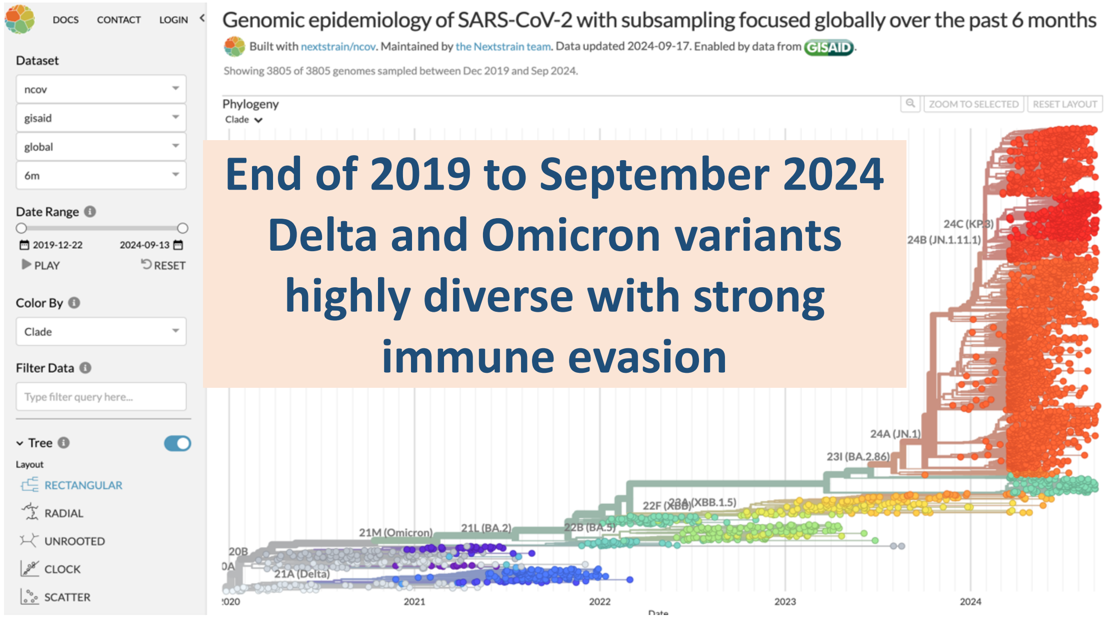
Application of AI: Pathogen Tracking

Case: Nextstrain Genome Variation Tracking and Prediction

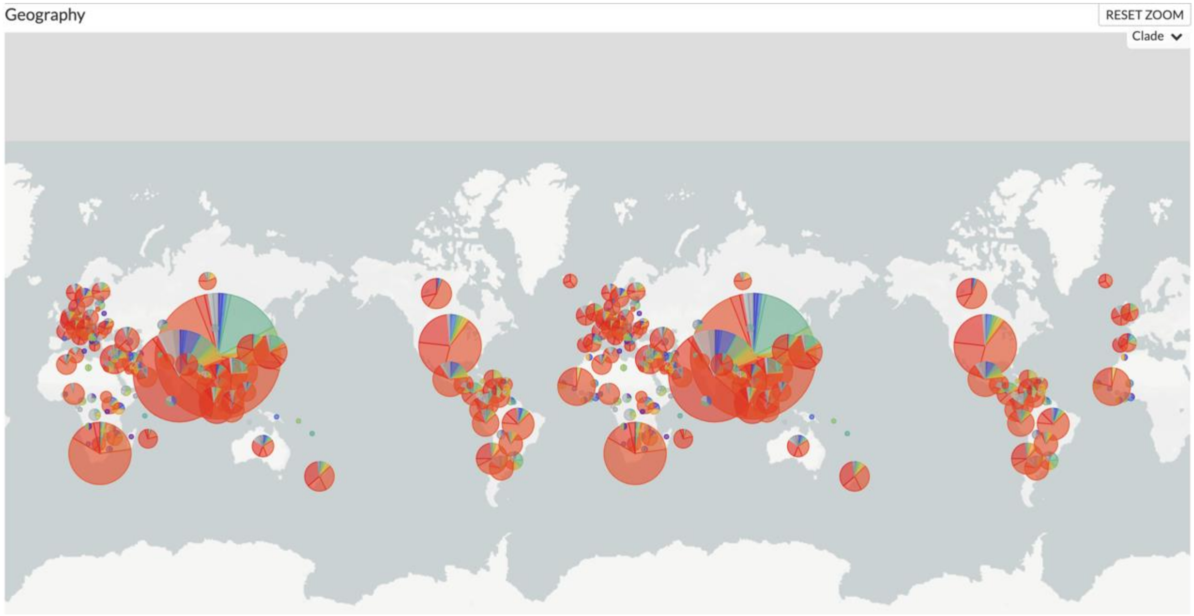
Nextstrain analyzes global pathogen genomes in real-time to track virus evolution.

Method genome sequences → (CNN) → multidimensional matrices → identify mutation patterns

probabilistic algorithms → continuously refine viral phylogenetic trees → predict evolutionary paths and origins



End of 2019 to September 2024
Delta and Omicron variants
highly diverse with strong immune evasion



Application of AI: Climate and Environmental Prediction

Traditional Methods

- Rely on static historical epidemic data and climate records for linear statistical modeling — failing to reflect current environmental changes and disease transmission patterns
- Rely on prior knowledge and assumptions, limiting the number of climate variables processed — low prediction accuracy and lack of real-time updates

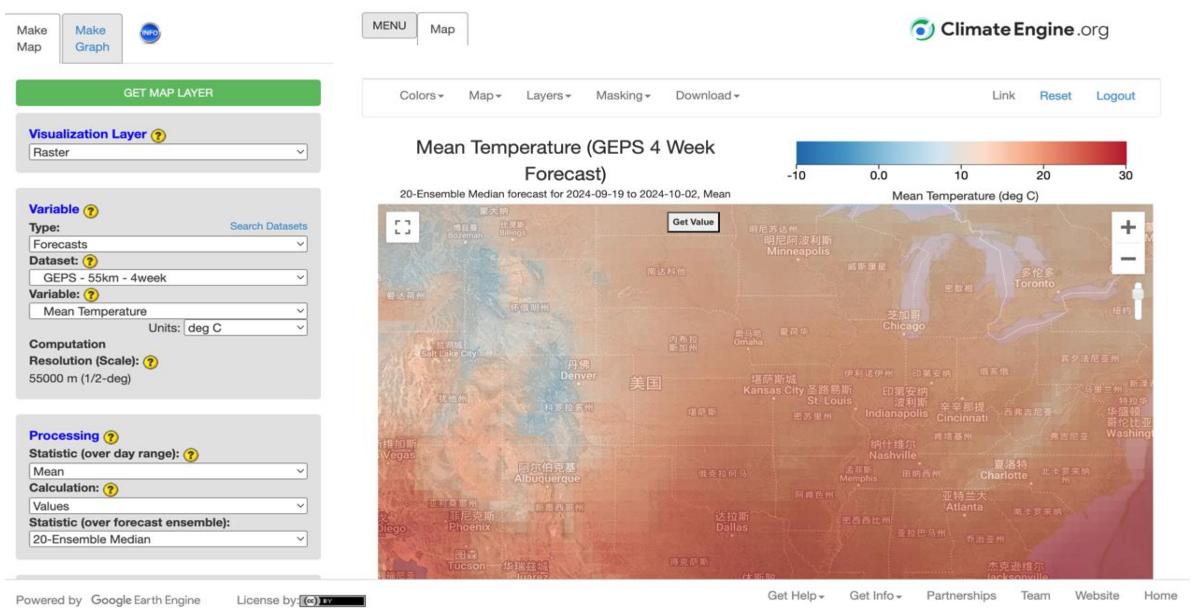
AI Methods

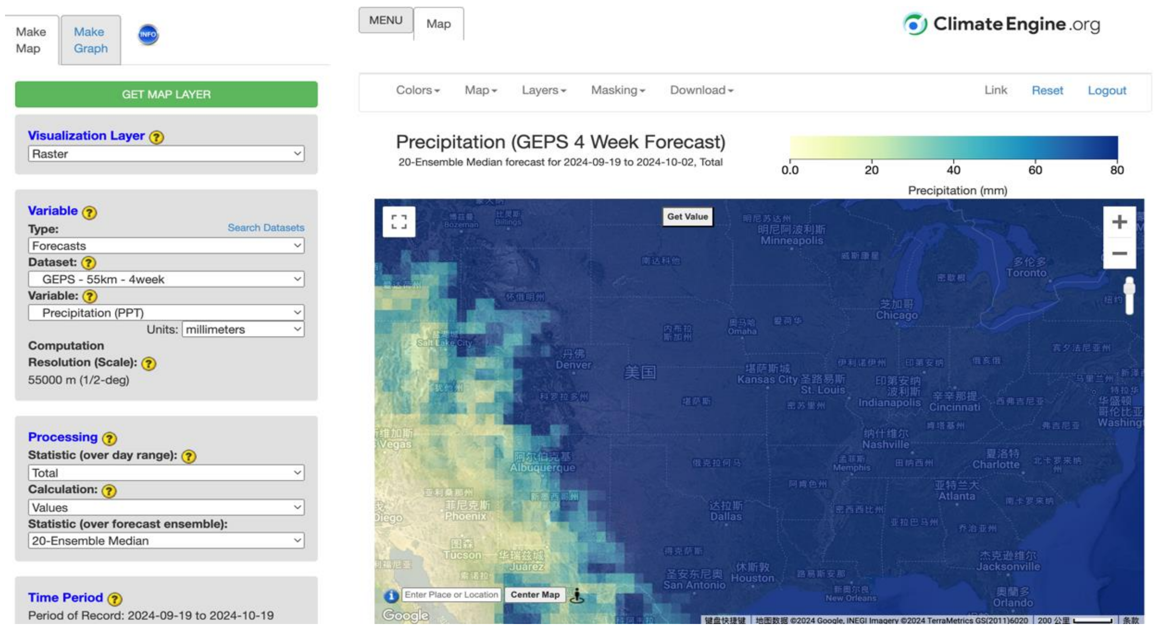
- Utilize DL algorithms for complex modeling to analyze multidimensional climate factors affecting vector breeding — capturing nonlinear relationships and interactions
- Feature self-learning and real-time updates, dynamically predicting outbreak risks of mosquito-borne diseases — ensuring high accuracy and timeliness

Application of AI: Climate and Environmental Prediction

Case: Climate Engine Environmental Monitoring and Prediction Platform

- Climate Engine **analyzes global climate data and environmental changes**, supporting agriculture, water resource management, and public health.
- Climate Engine utilizes satellite remote sensing, meteorological data, and climate models **to generate high-risk alerts for mosquito-borne disease outbreaks**.





Application of AI: Regional Epidemic Monitoring

Traditional Methods

- Rely on Geographic Information Systems to create epidemic hotspot maps for regional monitoring — challenges in ensuring accuracy and completeness of information
- Rely on predefined spatial smoothing assumptions, such as similar epidemic conditions in neighboring areas — limited spatial predictive capabilities

AI Methods

- Utilize big data models to integrate diverse sources to extract key information like case counts and geographic locations — real-time monitoring of global infectious diseases
- Employ ML to identify epidemic hotspots and transmission patterns for early warning — assisting public health departments in quickly obtaining the latest information

Application of AI: Regional Epidemic Monitoring

Case: HealthMap Infectious Disease Monitoring Platform

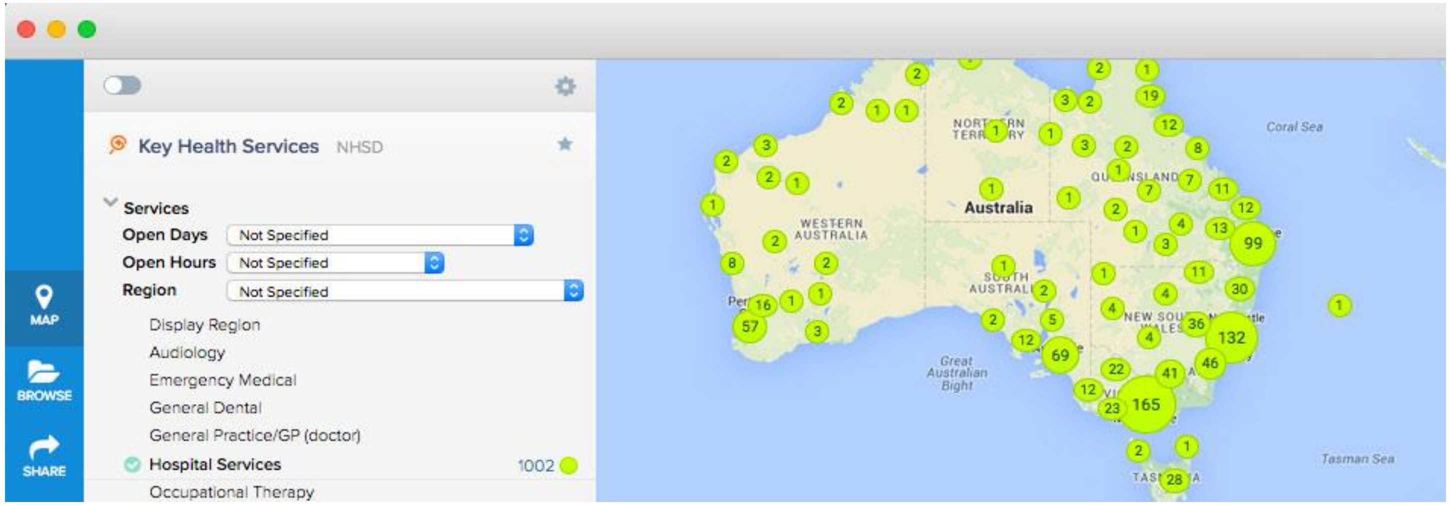
HealthMap **monitors global infectious disease outbreak information in real-time.**

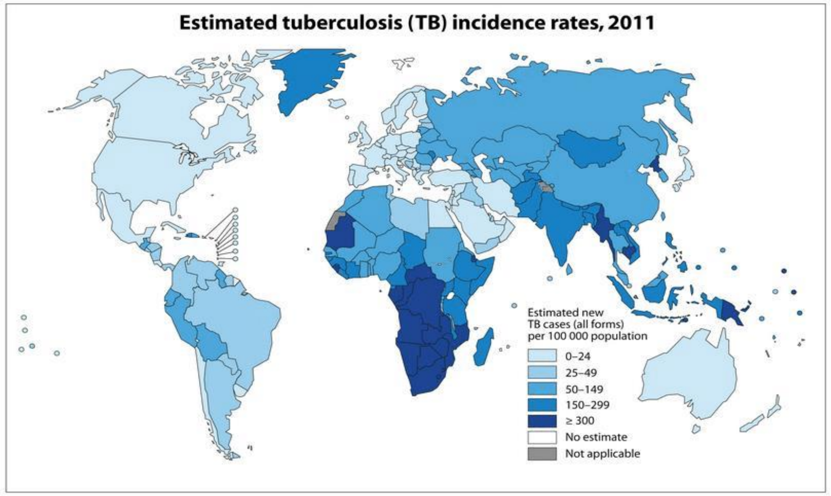
Method NLP → identify **key entities** and extract **keywords** related to outbreaks

frequency analysis → assess **disease outbreak risks** in specific areas

training classification models → assess the **link between symptoms and diseases**

cluster similar outbreak events → reveal **hidden transmission patterns**





Application of AI: Evaluating Control Measures

Traditional Methods

- Rely on surveys and polls to gather public opinion on vaccination and health policies — limited coverage, lengthy data collection, and slow reflection of public sentiment

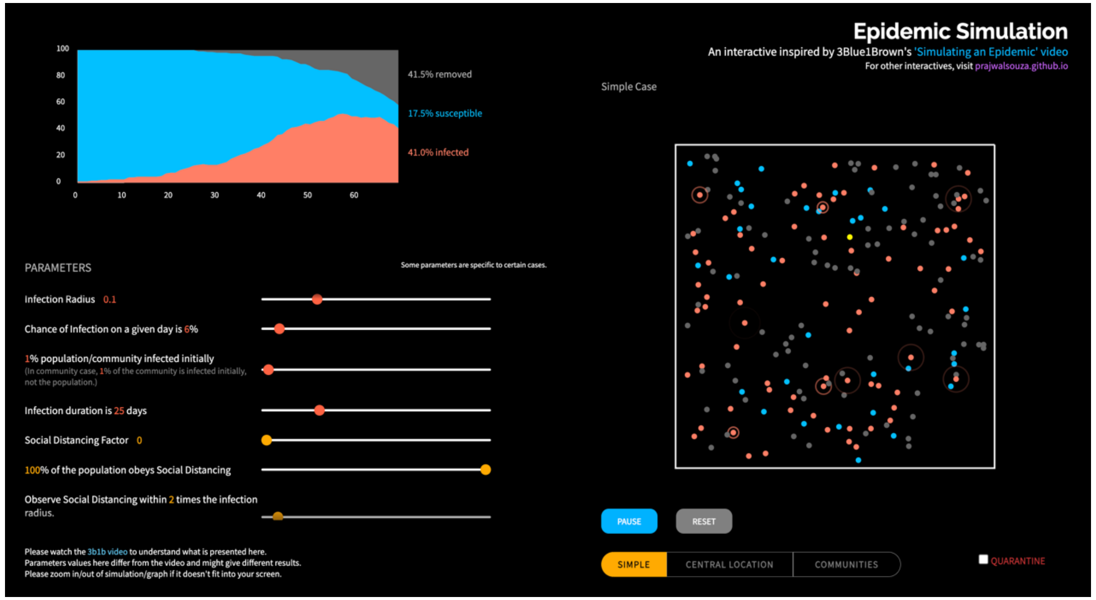
AI Methods

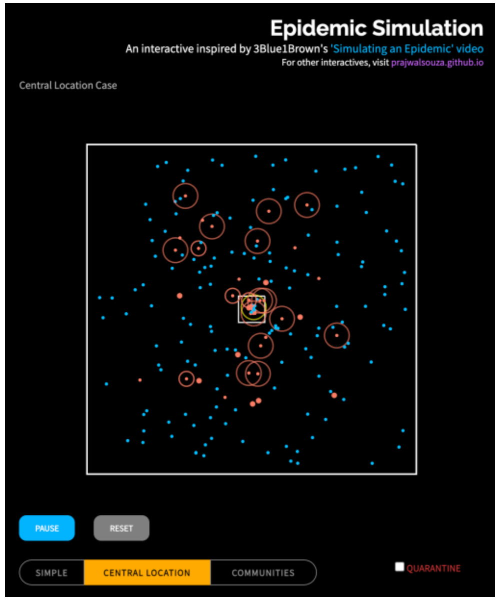
- Use NLP to gather real-time feedback from social media on control measures — shortening data collection time and offering timely insights into public sentiment
- Combine epidemiological models and ML to simulate disease spread under different scenarios, such as varying vaccine doses or no vaccination — evaluate disease burden and cost-effectiveness to dynamically adjust control measures

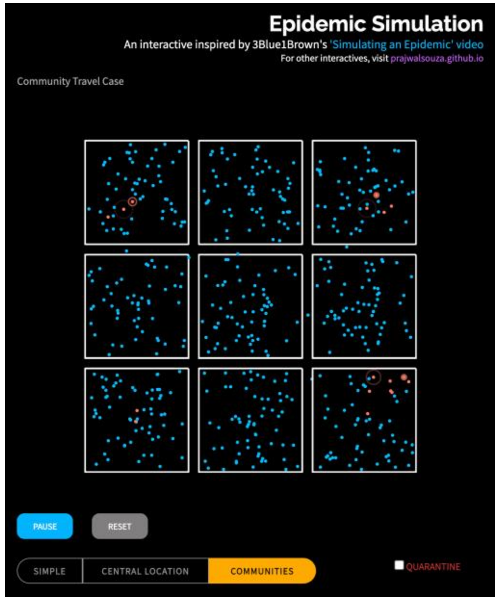
Application of AI: Evaluating Control Measures

Case: Epidemic Simulation Platform

- Use the **SIR model** to simulate disease spread under various scenarios, comparing infection numbers to quantify the effectiveness of control measures.
- Combine RL to optimize interventions through a **trial-and-error approach**, while **Bayesian inference** is used to update model parameters in real-time.







Process of Predicting Infectious Diseases Using AI

Data Acquisition	Collect structured and unstructured data.
Extraction of Epidemiological Features	Extract relevant epidemiological features, including the impact of climate change on disease transmission and patterns of population movement.
Determination of Prediction Model	Employ various machine learning techniques, such as CNN and SVM, to predict transmission trends and case numbers.
Identification of Training and Testing Sets	Analyze historical data to determine the most relevant training and testing datasets, ensuring a robust model training process.
Model Training	Process complex epidemiological data, integrating non-medical factors to predict potential outbreaks of infectious diseases and their spread patterns.
Valuation of Predictive Accuracy	Use cross-validation or error assessment to evaluate model accuracy and compare the model's performance with traditional or less effective algorithms.

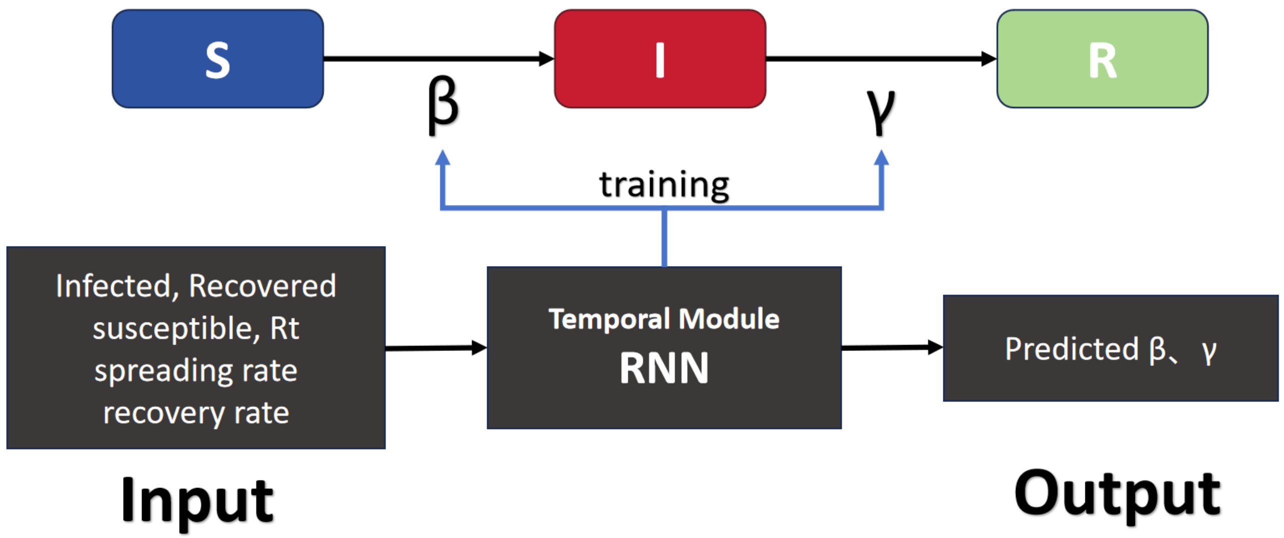
Process of Predicting Infectious Diseases Using AI

Training Epidemiological Parameters Based on AI Models

- **Traditional mechanistic** models rely on preset or fixed epidemiological parameters — struggle to adapt to the complex and dynamic real-world environment

- **AI-based models** use real data to train variables, allowing parameters to adapt to various scenarios — effectively addressing uncertainties and variations within the data

AI & SIR



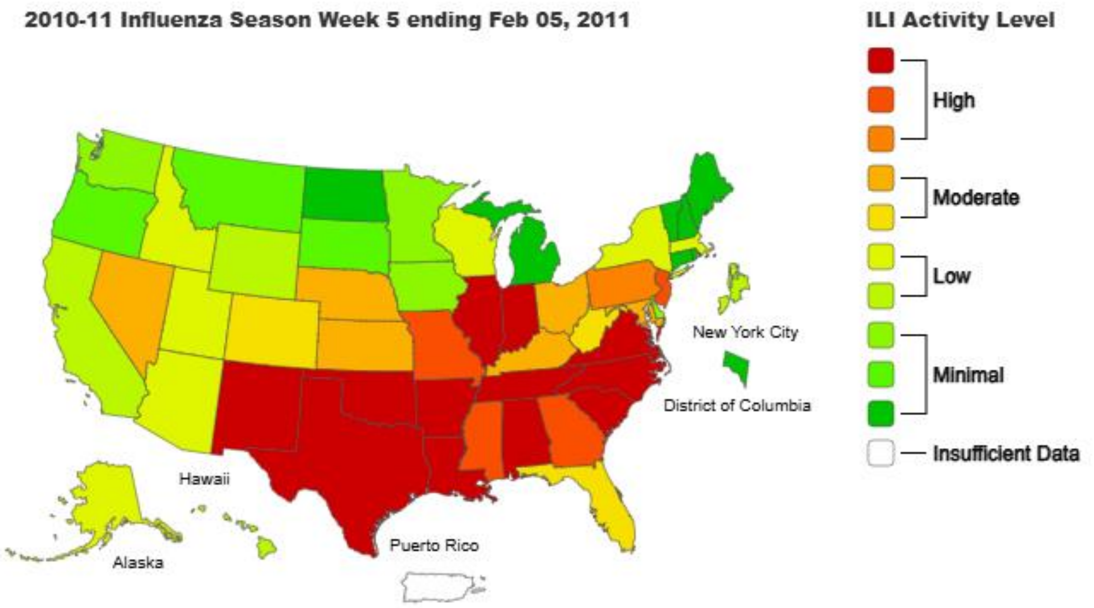
Reference: Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction

Process of Predicting Infectious Diseases Using AI

Paper: Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction

Data Acquisition → Extraction of Epidemiological Features → Determination of Prediction Model → Identification of Training and Testing Sets → Model Training → Evaluation of Predictive Accuracy

Utilizing a weekly dataset of influenza-like illness activity from ten U.S. regions, covering 364 weeks from Week 1 of 2010 to Week 52 of 2016.



Process of Predicting Infectious Diseases Using AI

Paper: Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction

Data Acquisition → **Extraction of Epidemiological Features** → Determination of Prediction Model → Identification of Training and Testing Sets → Model Training → Evaluation of Predictive Accuracy

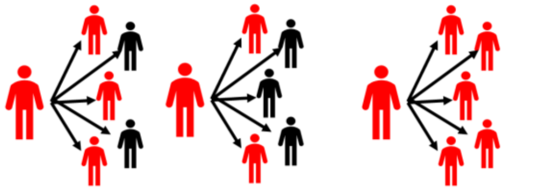
Input the historical data on epidemiological characteristics into the model:

S **I** **R**

including the number of susceptible, infected, and recovered individuals, as well as the transmission rate, recovery rate, and basic reproduction number

β γ

R0 = attack rate (AR) * number of contacts

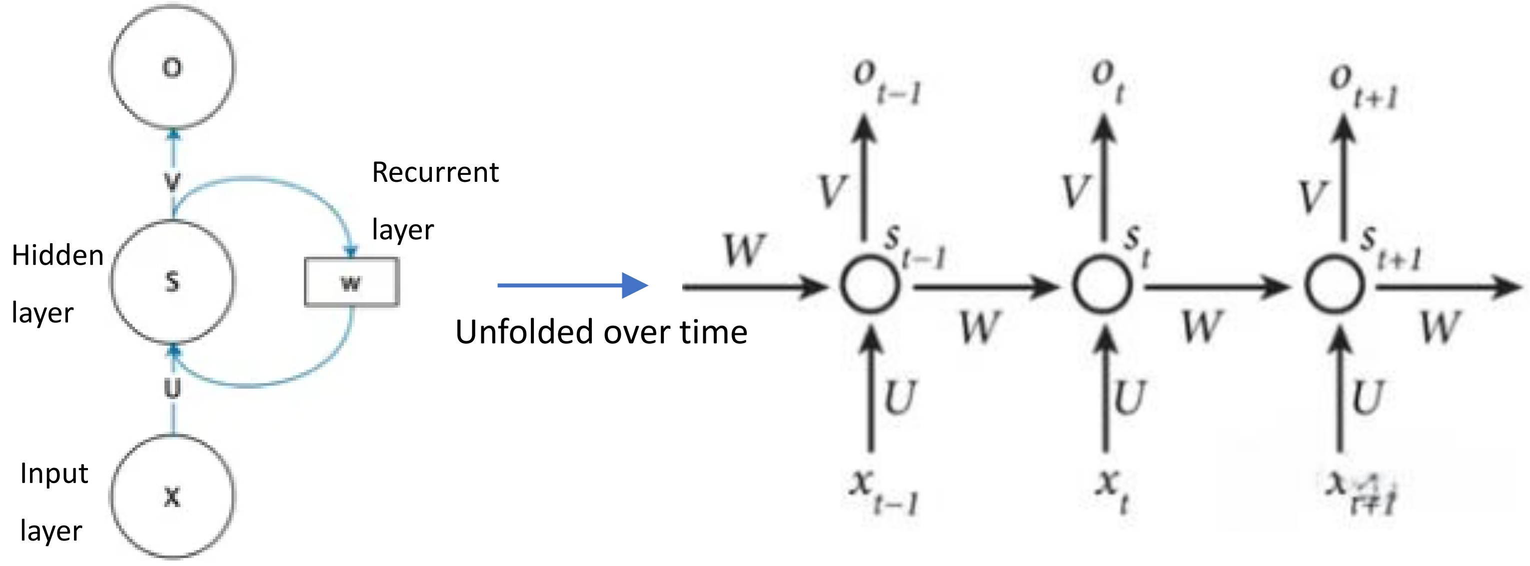


Process of Predicting Infectious Diseases Using AI

Paper: Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction

Data Acquisition → Extraction of Epidemiological Features → **Determination of Prediction Model** → Identification of Training and Testing Sets → Model Training → Evaluation of Predictive Accuracy

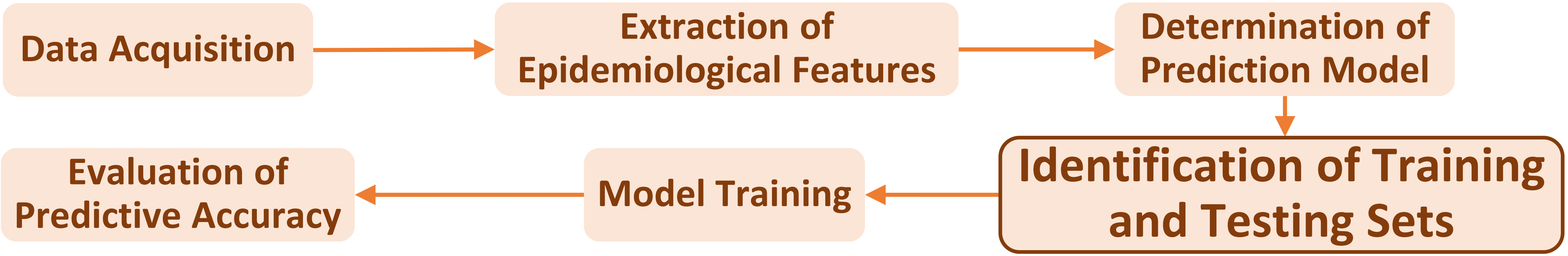
Utilize a deep learning-based recurrent neural network (RNN) model for epidemiological predictions.



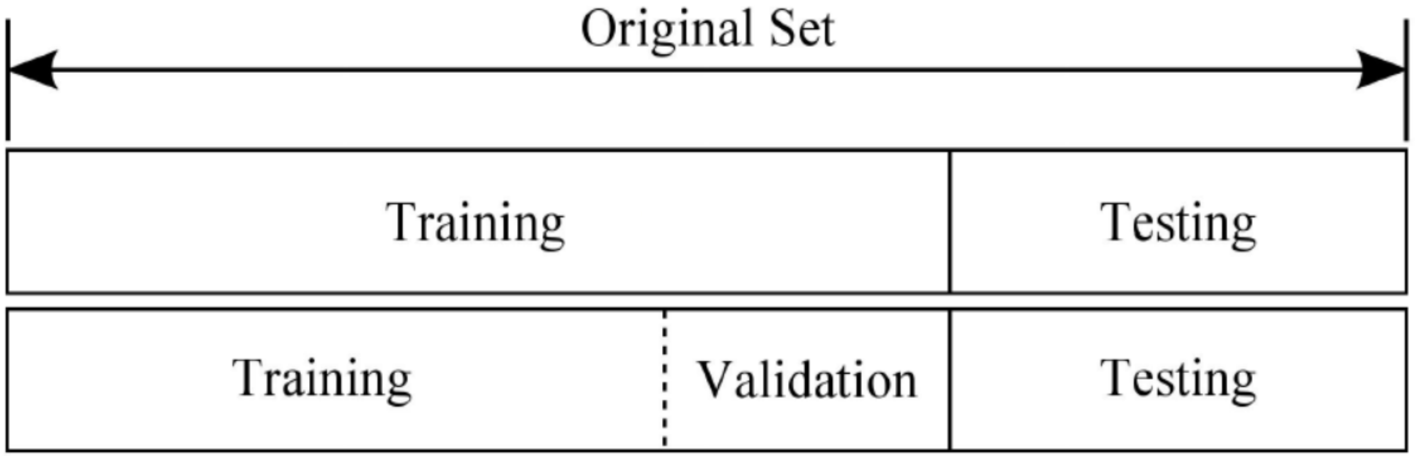
Principle of the RNN model

Process of Predicting Infectious Diseases Using AI

Paper: Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction

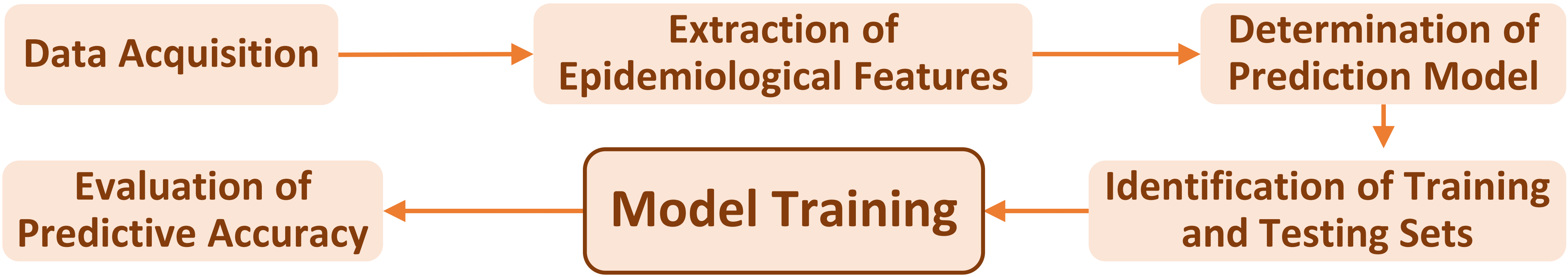


Divide historical data into training sets, testing sets, and validation sets in proportion to verify the accuracy of the prediction results.



Process of Predicting Infectious Diseases Using AI

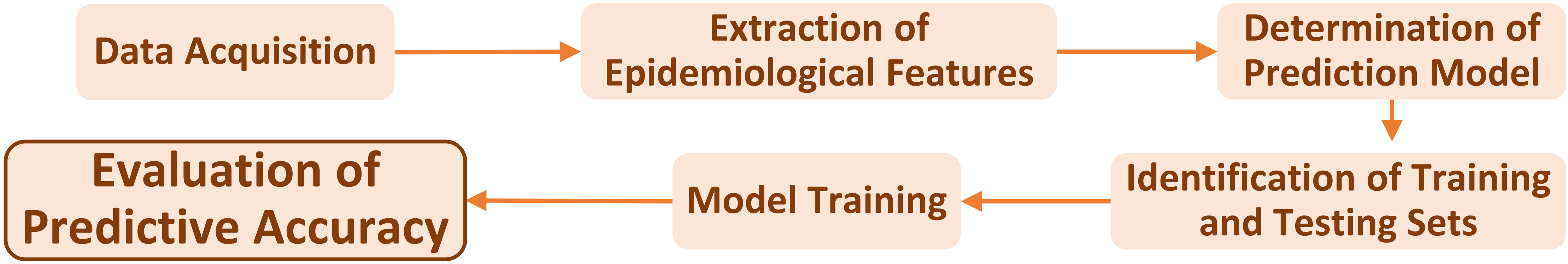
Paper: Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction



- **Input:** Model receives historical data on infected, recovered, susceptible counts, transmission rate, recovery rate, and basic reproduction number.
- **Processing:** Model learns time-dependent features and updates internal state.
- **Output:** Network outputs predicted transmission and recovery rates.
- **Loss Calculation:** Predicted parameters simulate flu activity; difference with actual data forms the loss.
- **Parameter Update:** Parameters are updated based on loss gradient.

Process of Predicting Infectious Diseases Using AI

Paper: Epidemiology-aware Deep Learning for Infectious Disease Dynamics Prediction



Evaluate the model's predicted case numbers against actual cases using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

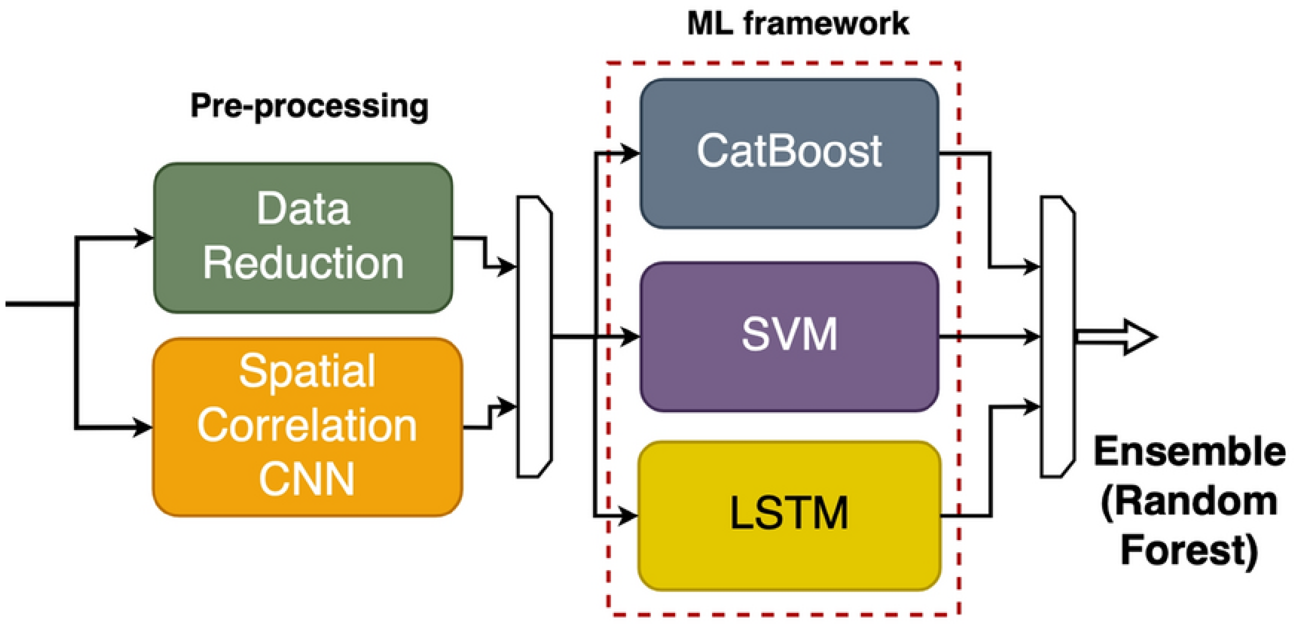
$$\text{MSE}=\frac{1}{m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2 \qquad \text{RMSE}=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2} \qquad \text{MAE}=\frac{1}{m}\sum_{i=1}^{m}\left|(y_i-\hat{y}_i)\right|$$

Process of Predicting Infectious Diseases Using AI

Dengue Prediction Based on Ensemble Models

- **Traditional** dengue forecasts rely on single models like time series or mechanistic models — lacking accuracy due to data noise and parameter sensitivity

- **Ensemble models** combine multiple approaches, enhancing robustness and accuracy — better handle epidemic complexity and data uncertainty



Reference: A reproducible ensemble machine learning approach to forecast dengue outbreaks

Process of Predicting Infectious Diseases Using AI

Paper: A reproducible ensemble machine learning approach to forecast dengue outbreaks

Data Acquisition → Features Extraction → Determination of Prediction Model → Identification of Training and Testing Sets → Model Training → Evaluation of Predictive Accuracy

- **Medical Data:** Monthly dengue incidence rate (DIR) in Brazil (2001–2019).
- **Meteorological Data:** Climate variables (temperature, rainfall, humidity, wind) from ERA5-Land and NDVI from MODIS.
- **Geographic Data:** Includes factors like geographic elevation.
- **Socioeconomic Data:** Includes population density, urbanization, road density, and social vulnerability index (SVI).

https://gisgeography.com/brazil-map/

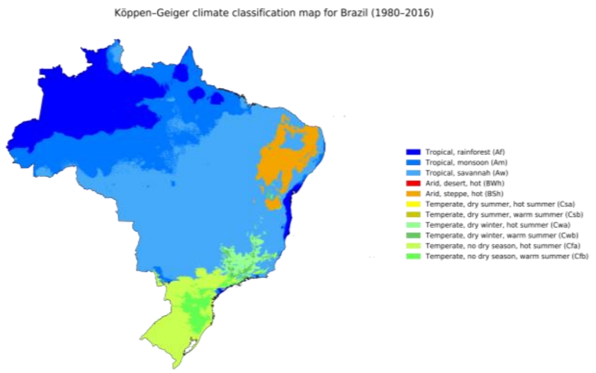
Process of Predicting Infectious Diseases Using AI

Paper: A reproducible ensemble machine learning approach to forecast dengue outbreaks

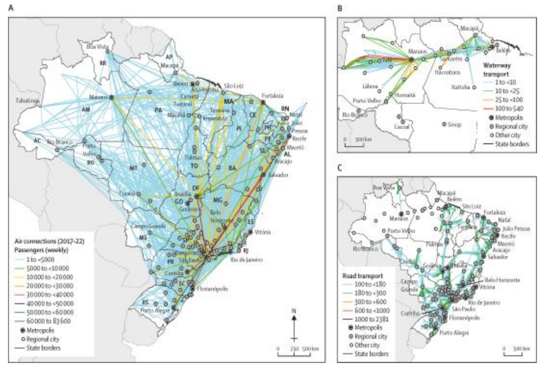
Data Acquisition → **Features Extraction** → Determination of Prediction Model → Identification of Training and Testing Sets → Model Training → Evaluation of Predictive Accuracy

- **Climate Change Impact:** Temperature (max/min/average), rainfall, relative humidity, wind speed.
- **Population Mobility Patterns:** Social vulnerability index, population density, urbanization rate, road density.
- **Geographic and Environmental Features:** Geographic elevation, vegetation index.

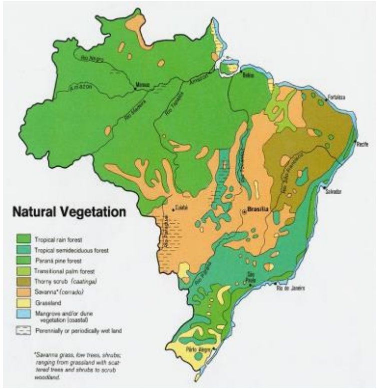
Climate Change



Population Mobility



Geography and Environment



Process of Predicting Infectious Diseases Using AI

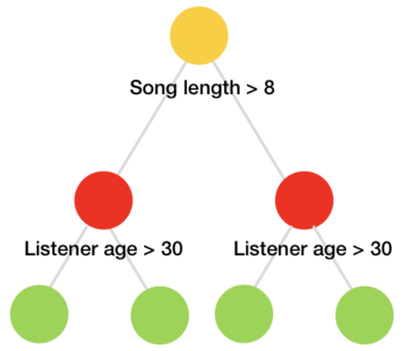
Paper: A reproducible ensemble machine learning approach to forecast dengue outbreaks

Data Acquisition → Features Extraction → **Determination of Prediction Model** → Identification of Training and Testing Sets → Model Training → Evaluation of Predictive Accuracy

- **Meteorological Data:** Temperature (max/min/average), rainfall, relative humidity, and wind speed.
- **Population Data:** Total population, 0-19 age ratio.
- **Socioeconomic Data:** Population density, urbanization level, and road density.
- **Historical Dengue Data:** Previous months' DIR.

Input

A robust decision tree model suited for complex nonlinear relationships



CatBoost Model

Next Month's DIR Prediction

Output

Process of Predicting Infectious Diseases Using AI

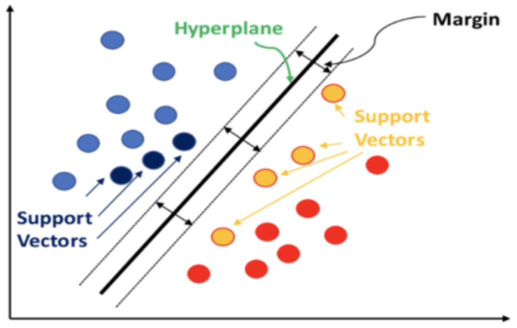
Paper: A reproducible ensemble machine learning approach to forecast dengue outbreaks

Data Acquisition → Features Extraction → **Determination of Prediction Model** → Identification of Training and Testing Sets → Model Training → Evaluation of Predictive Accuracy

- **Meteorological Data:** Temperature (max/min/average), rainfall, relative humidity, and wind speed.
- **Population Data:** Total population, 0-19 age ratio.
- **Socioeconomic Data:** Population density, urbanization level, and road density.
- **Historical Dengue Data:** Previous months' DIR.

Input

A mathematically optimized algorithm that extracts useful patterns from data



Support Vector Machine (SVM)

Next Month's DIR Prediction

Output

Process of Predicting Infectious Diseases Using AI

Paper: A reproducible ensemble machine learning approach to forecast dengue outbreaks

Data Acquisition → Features Extraction → **Determination of Prediction Model**

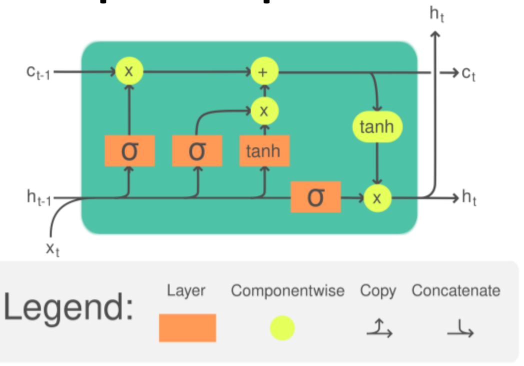
↓

Identification of Training and Testing Sets → Model Training → Evaluation of Predictive Accuracy

Time Series Data: Meteorological data (temperature, rainfall, humidity, wind speed), socioeconomic data, population data, and dengue incidence rate (DIR) **from previous months**.

Input

A deep learning algorithm for processing time series data that captures long-term temporal dependencies



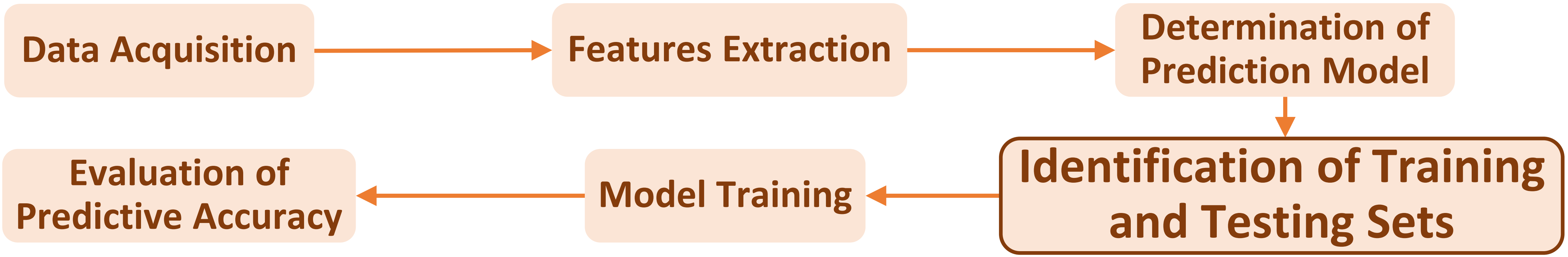
Long Short-Term Memory Network (LSTM)

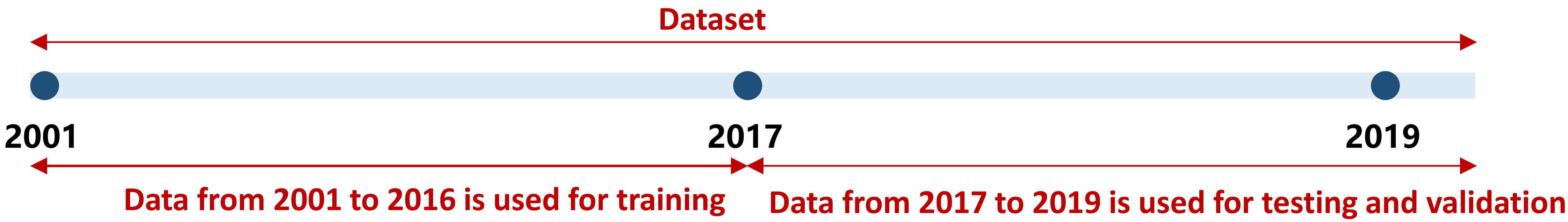
Output

Next Month's DIR Prediction

Process of Predicting Infectious Diseases Using AI

Paper: A reproducible ensemble machine learning approach to forecast dengue outbreaks

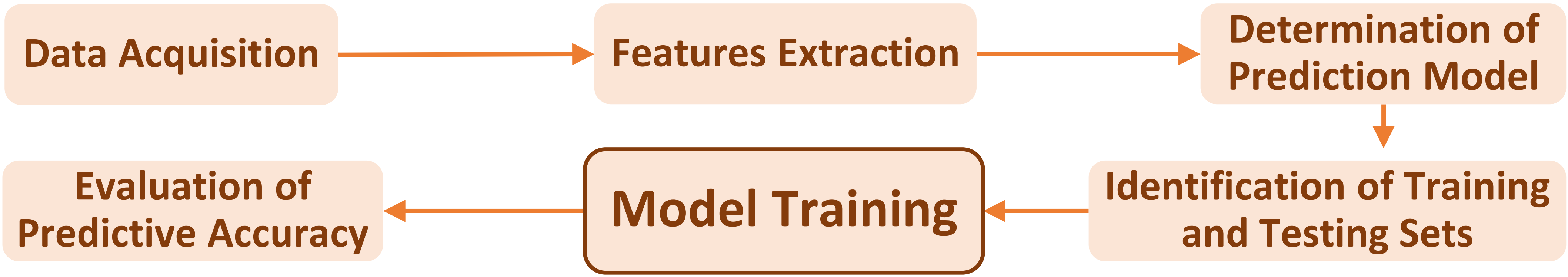




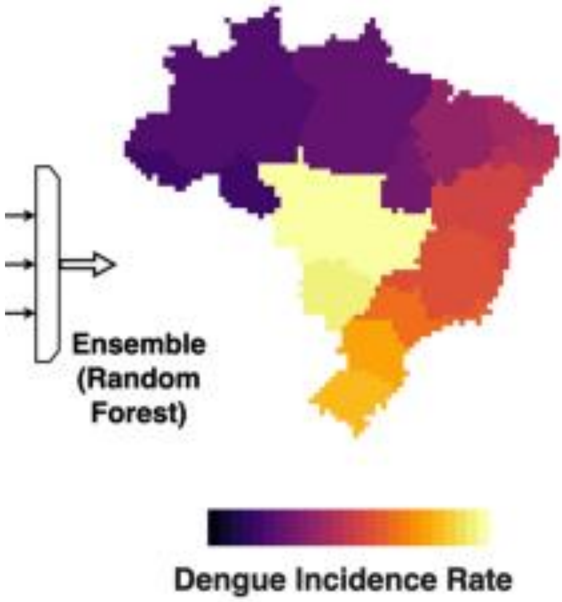
The model learns trends and correlations from historical data to accurately predict future dengue outbreaks.

Process of Predicting Infectious Diseases Using AI

Paper: A reproducible ensemble machine learning approach to forecast dengue outbreaks



- **Input:** Predictions from CatBoost, SVM, and LSTM serve as inputs for the random forest.
- **Decision Tree Construction:** Randomly select combinations of predictions to build multiple decision trees.
- **Majority Voting:** Each tree predicts based on its weight; the random forest combines these through majority voting or averaging for the final prediction.
- **Final Output:** Output the final predicted dengue incidence rate based on the votes or weighted averages from all decision trees.

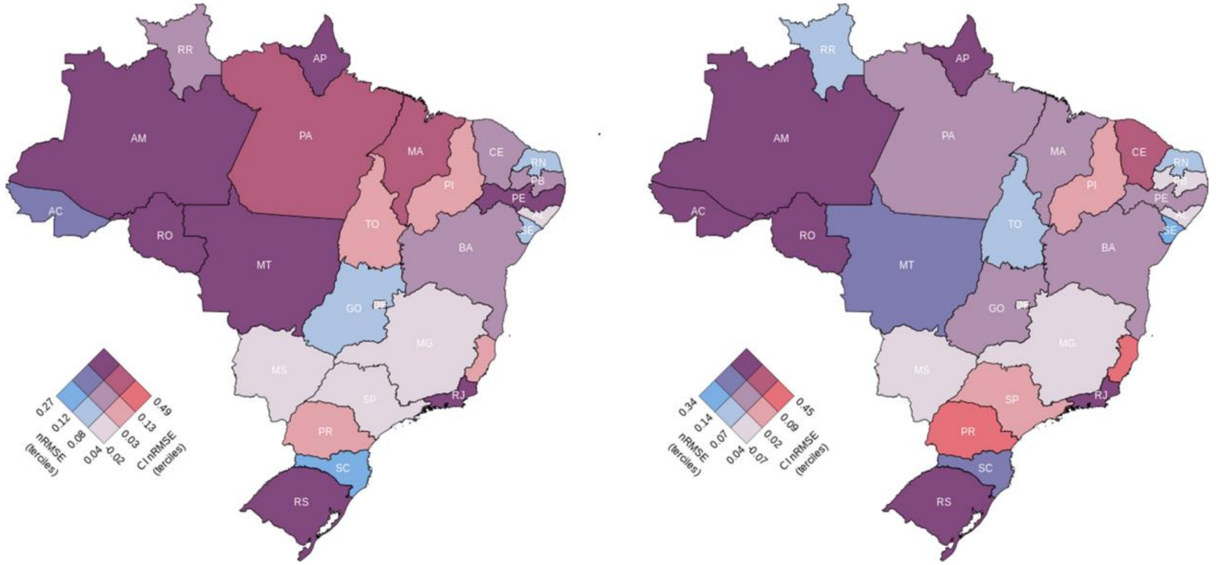


Process of Predicting Infectious Diseases Using AI

Paper: A reproducible ensemble machine learning approach to forecast dengue outbreaks

Data Acquisition → Features Extraction → Determination of Prediction Model → Identification of Training and Testing Sets → Model Training → **Evaluation of Predictive Accuracy**

- **Cross-Validation:** Conduct multiple tests to ensure model stability across different datasets.
- **Error Assessment:** Analyze the differences between predictions and actual data for accuracy.



Display a bivariate plot of the ensemble model's nRMSE: (Left) Total population; (Right) Population aged 0-19. Light and dark purple represent lower and higher nRMSE values and uncertainty; blue indicates high nRMSE with low uncertainty, and orange indicates the opposite.

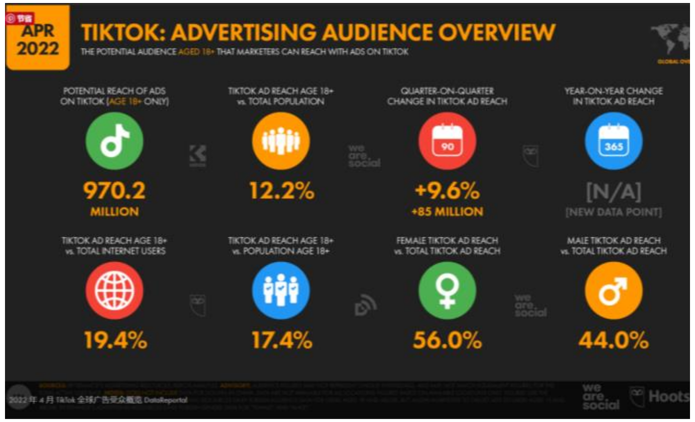
Development and Challenges of AI

Development

1. Enhanced Data Processing and Integration

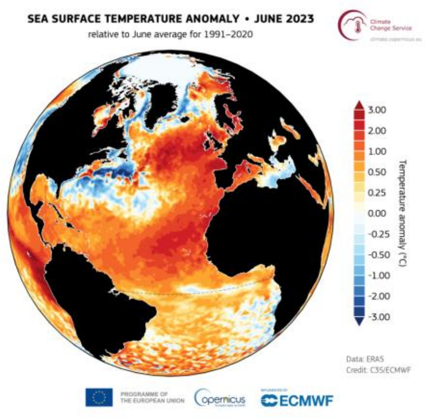
AI will integrate multisource data, including health records, social media, traffic data, and climate data. This will enable the processing of complex, real-time datasets to improve prediction accuracy and identify potential outbreak points from multiple dimensions.

Health records



Social data

Traffic data



Climate data

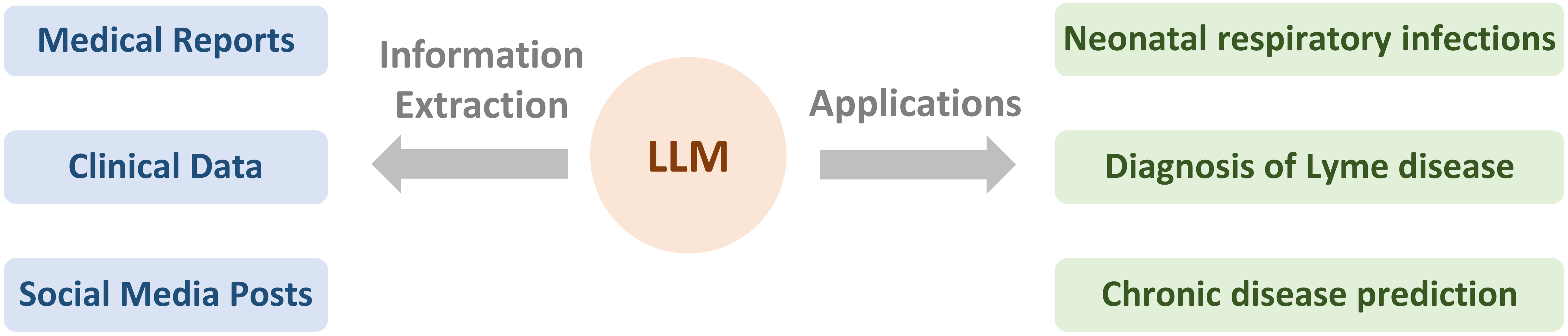
Development and Challenges of AI

Development

2. Improved Processing of Unstructured Data

Large language models like GPT-4 can analyze extensive unstructured data, uncovering hidden trends and key information while saving time and resources.

Large Language Model

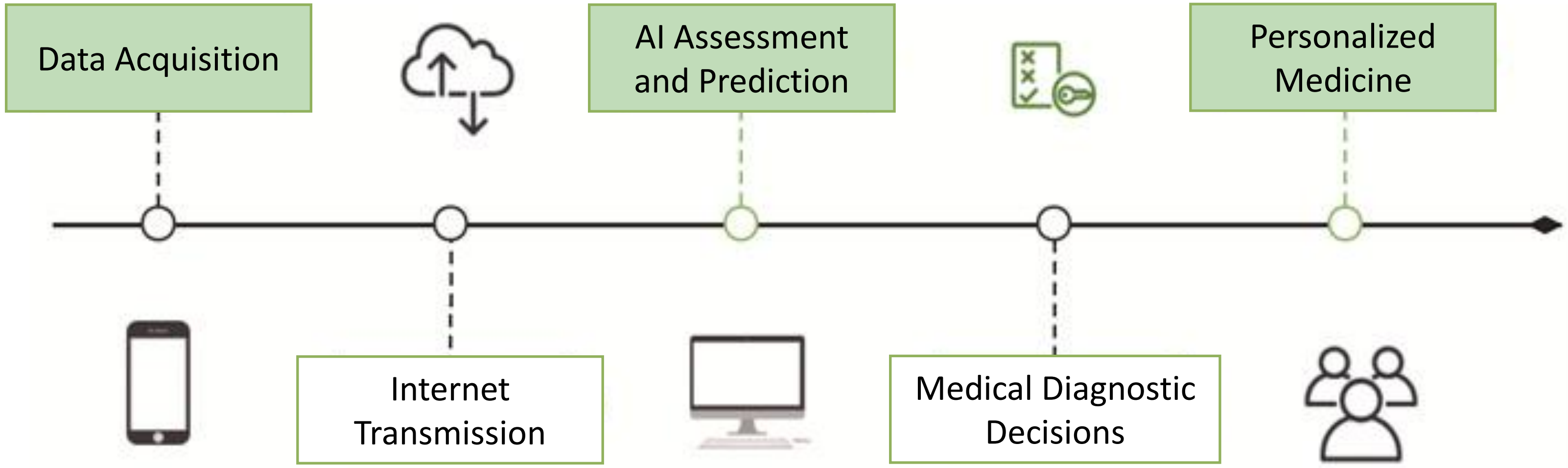


Development and Challenges of AI

Development

3. Personalized Infectious Disease Prediction

Future AI systems may offer tailored prediction models for various regions, age groups, ethnicities, and socioeconomic backgrounds.



The Geisinger Health System in the U.S. has already used AI to personalize patient health data analysis and predict cardiovascular disease risk.

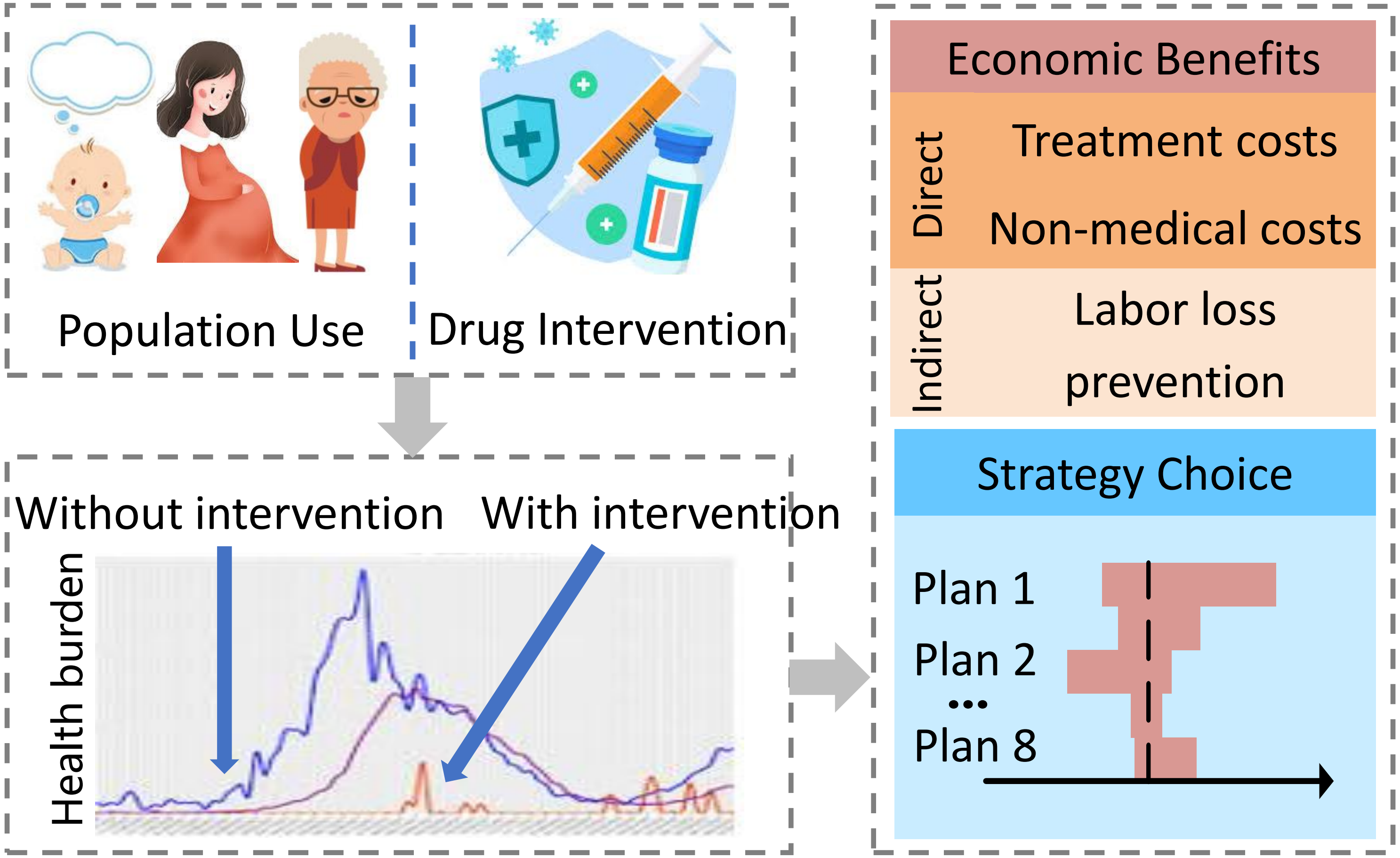
Development and Challenges of AI

Development

4. Multidisciplinary Integration

The future of AI in infectious disease prediction depends on integration with various.

Since 2023, RSV vaccines have been approved. Health economic analysis can support pricing by assessing cost-effectiveness.



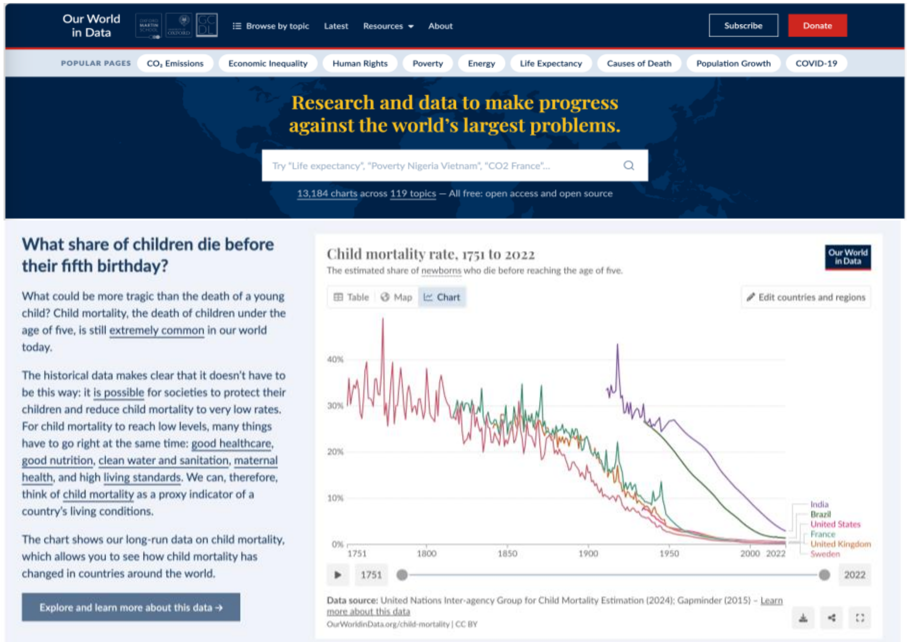
AI + Health Economics

Development and Challenges of AI

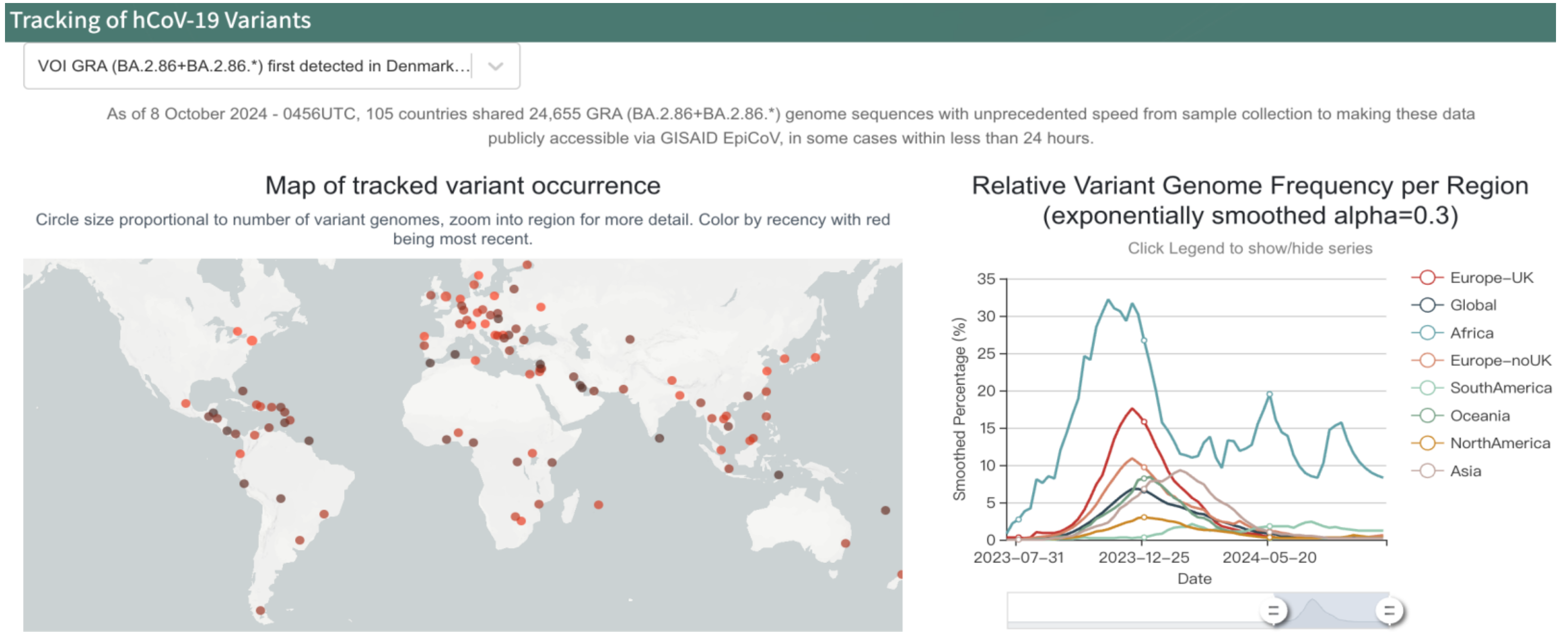
Development

5. Global Data Sharing and Collaboration

Regions will establish unified open data platforms, enabling AI to learn from global data, improving prediction accuracy and supporting disease control efforts.



Our World in Data



GISAID

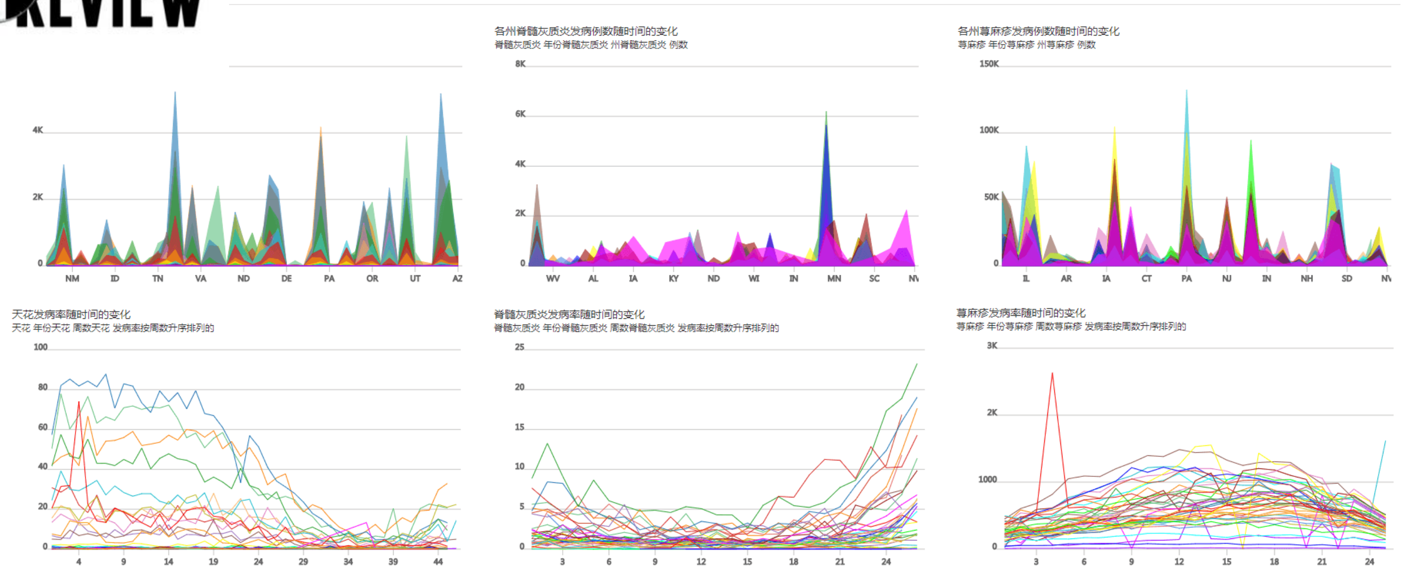
Development and Challenges of AI

Challenges

1. Insufficient Historical Data

AI models need extensive, quality data, but early-stage outbreaks often lack sufficient history and reliable databases, limiting tracking and prediction.

For COVID-19, initial data mostly came from small, potentially biased samples in China, with many studies unreviewed.



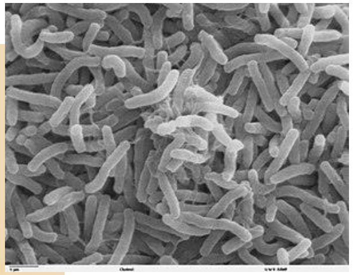
Development and Challenges of AI

Challenges

2. Data Quality Issues

While big data sources like social media are abundant, they often contain significant noise (e.g., misinformation), requiring filtering for accurate predictions.

In 2007, Google searches for "cholera" surged, not due to an outbreak, but because Oprah Winfrey selected the novel *Love in the Time of Cholera* for her book club.

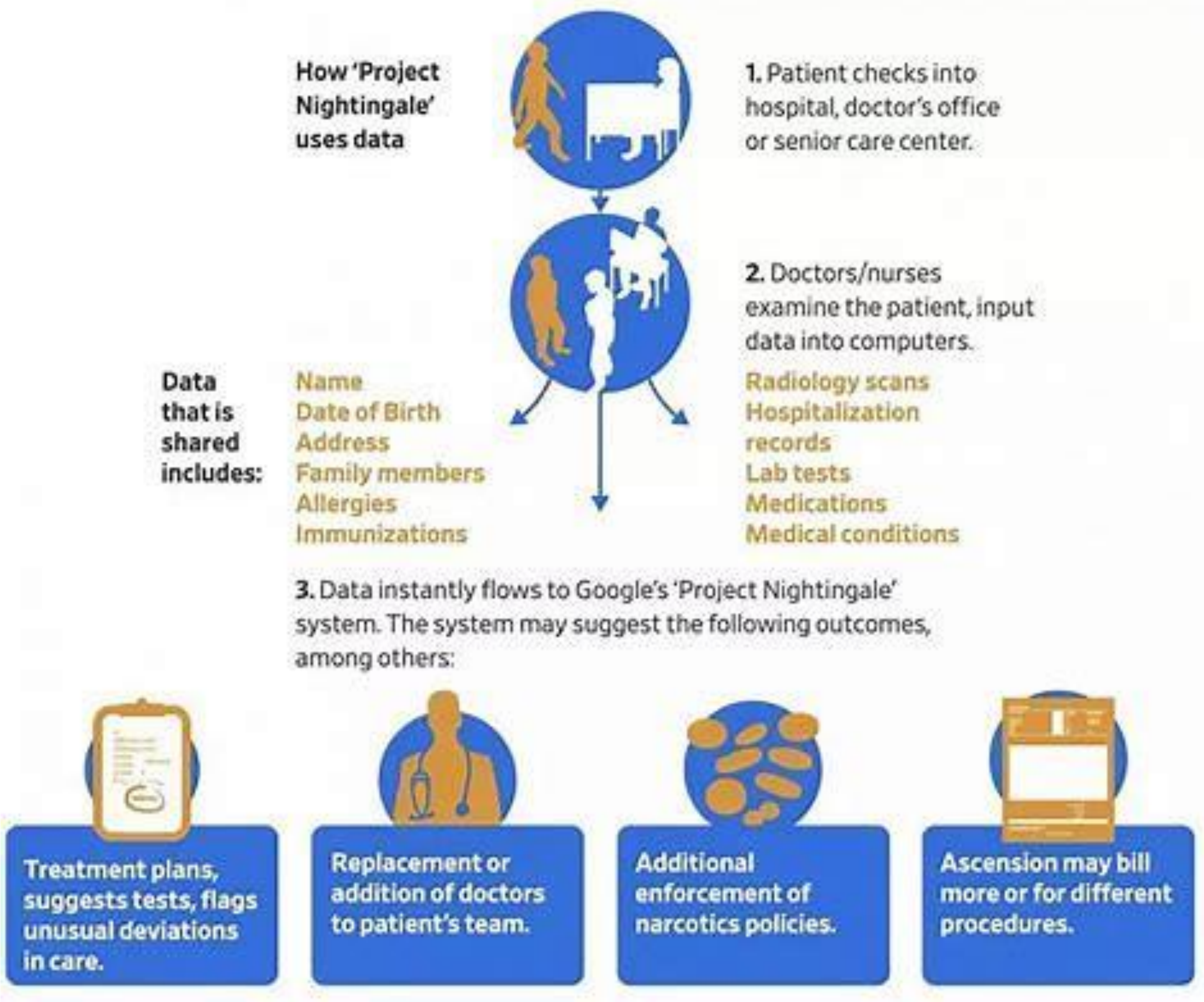


Challenges

3. Ethical and Privacy Issues

AI's use of personal data raises privacy issues, particularly with social media and mobile data, increasing breach risks.

Google's "Project Nightingale" gathered detailed health information from millions of unaware Americans in 21 states.



Development and Challenges of AI

Challenges

4. Algorithmic Bias Risks

AI models lacking diverse training data may worsen social inequalities.

During COVID-19, biased data caused U.S. algorithms to underestimate Black and Hispanic mortality differences by 60% and provide fewer care plans for non-White patients.


12 国际新闻

光明日报

算法黑箱强化偏见

数字技术加剧美国的种族歧视

世贸组织第十二届部长级会议对维护多边主义具有重要意义

——访中国常驻世贸组织代表李成钢

未来医学将如何保护我们的生命健康

Thanks!

Zhanwei Du

THE UNIVERSITY OF HONG KONG