



浙江大学  
ZHEJIANG UNIVERSITY

# 基于LLM和卫星图像的社会公共卫生数据发掘与建模

答辩人：冯明浩

指导老师：崔兆鹏

项目负责人：冯明浩

组员：陈君林、付泽阳



## 01 项目背景及意义

Project Background And Significance

## 02 项目主要内容

Main content of the project

## 03 项目优势和风险

Advantages and risks of the project

## 04 项目预期成果

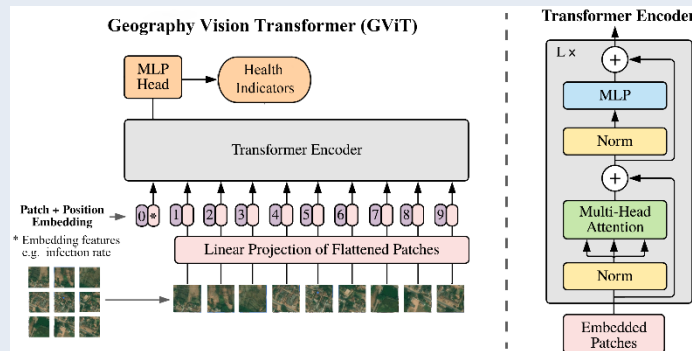
Expected result of the project

## 背景

目前亚非拉地区的一些国家仍面临着严重的**公共卫生挑战**。而对这些国家的公共卫生数据调研主要依赖于**实地考察**。由于地域广阔、人力物力投入巨大等问题，实地考察的方式**难以保证数据的全面性和准确性**。政府和国际组织在制定针对性的卫生资源分配和干预措施时也面临巨大困难。

## 目的

- 借助大语言模型挖掘和分析社会公共卫生数据
- 结合卫星图像数据，构建一种能**从卫星图像中预测相关统计数据的深度学习模型**。



## 意义

- 为制订更有效的公共卫生政策提供科学依据，进一步推动国际社会对第三世界国家的**关怀和援助**。
- 有助于拓展卫星图像在公共卫生领域的应用。通过**深度学习技术**，利用卫星图像的**全覆盖性和客观性**充分挖掘其中有价值的信息，为解决亚非拉地区的公共卫生问题提供新的思路和方法。

- (1) **大语言模型**已具备多模态数据挖掘、分析的能力，这使得我们能够大规模地发掘已有的社会公共卫生统计数据。
- (2) **深度学习模型**的评估和优化方法较为完善。
- (3) 已经有一些团队在基于卫星遥感图片的数据研究上取得成果。这些研究为我们结合卫星图像预测和分析公共卫生问题提供了范例和参考。







[www.nature.com/npjdigitalmed](http://www.nature.com/npjdigitalmed)

REVIEW ARTICLE

OPEN



# Digital public health surveillance: a systematic scoping review

Zahra Shakeri Hossein Abad <sup>1,2</sup>✉, Adrienne Kline<sup>1,3</sup>, Madeena Sultana<sup>1,2</sup>, Mohammad Noaeen<sup>4</sup>, Elvira Nurmambetova<sup>1</sup>,  
Filipe Lucini <sup>1,5</sup>, Majed Al-Jefri <sup>1,3</sup> and Joon Lee <sup>1,2,6</sup>

The ubiquitous and openly accessible information produced by the public on the Internet has sparked an increasing interest in developing digital public health surveillance (DPHS) systems. We conducted a systematic scoping review in accordance with the PRISMA extension for scoping reviews to consolidate and characterize the existing research on DPHS and identify areas for further research. We used Natural Language Processing and content analysis to define the search strings and searched Global Health, Web of Science, PubMed, and Google Scholar from 2005 to January 2020 for peer-reviewed articles on DPHS, with extensive hand searching. Seven hundred fifty-five articles were included in this review. The studies were from 54 countries and utilized 26 digital platforms to study 208 sub-categories of 49 categories associated with 16 public health surveillance (PHS) themes. Most studies were conducted by researchers from the United States (56%, 426) and dominated by communicable diseases-related topics (25%, 187), followed by behavioural risk factors (17%, 131). While this review discusses the potentials of using Internet-based data as an affordable and instantaneous resource for DPHS, it highlights the paucity of longitudinal studies and the methodological and

### **scientific** reports



**OPEN**

**Spatio-temporal analysis  
and prediction of malaria  
cases using remote sensing  
meteorological data in Diébougou  
health district, Burkina Faso,  
2016–2017**

# scientific reports

[www.nature.com/scientificreports](http://www.nature.com/scientificreports)

 Check for updates

OPEN

Investigation of the spatial and temporal variation of soil salinity using Google Earth Engine: a case study at Werigan–Kuqa Oasis, West China

## 02 项目主要内容——主要内容





## 02 项目主要内容——数据发掘



## 02 项目主要内容——数据发掘

使用 LLM 查询相

使用 LLM 直接收

人工整理现有的数



extract the specific number of U5MR of every city in India



### Ask PDF

The National Health Profile 2022 provides Under-5 Mortality Rate (U5MR) data for different states in India rather than for each city. Here are the U5MR figures for various states from 2011-2015 and projected rates up to 2035:

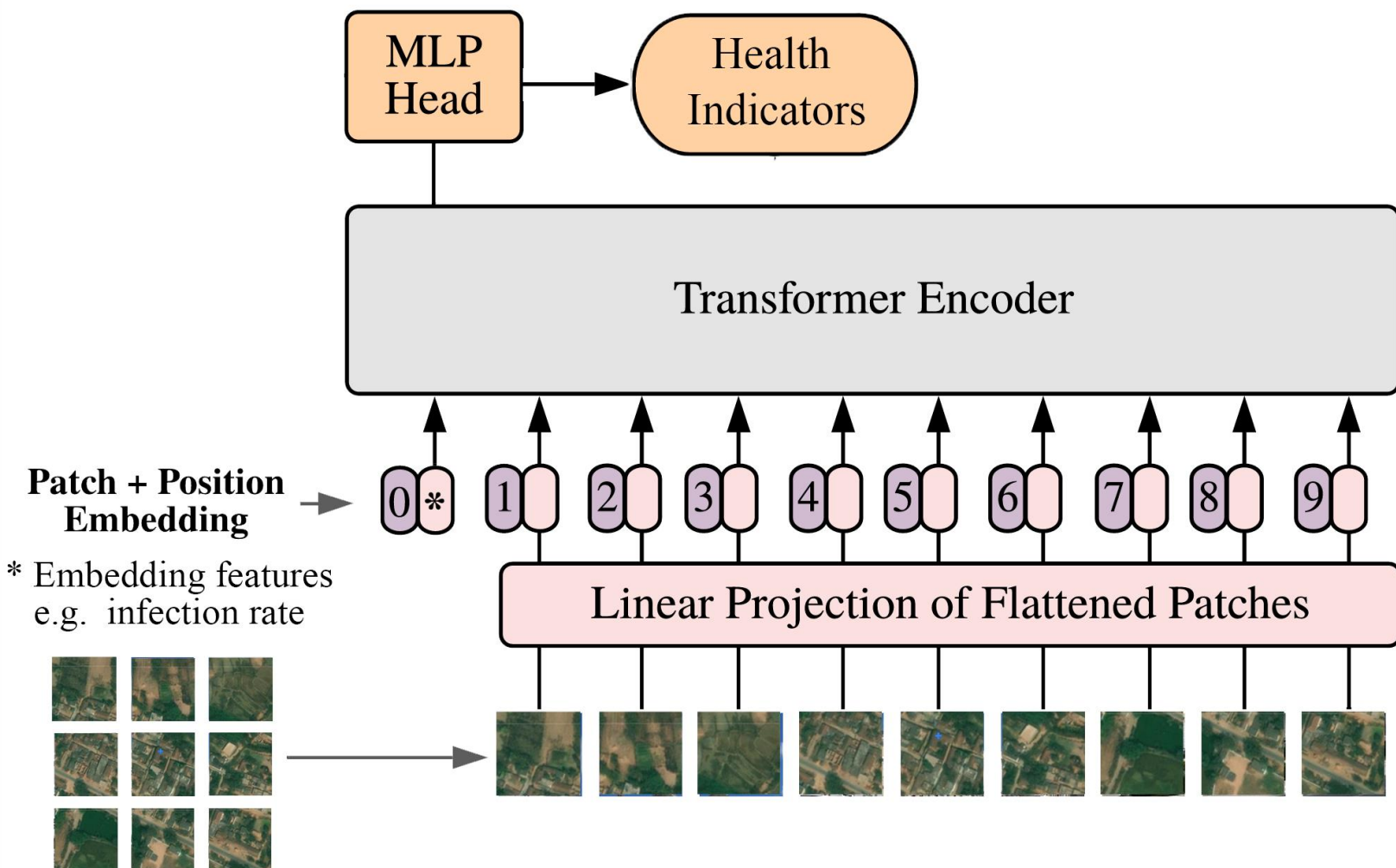
- Jammu & Kashmir (UT): 56.7 (2011-2015) to 39.4 (2031-2035)
- Himachal Pradesh: 41 to 30
- Punjab: 41 to 25
- Uttarakhand: 38 to 25
- Haryana: 52 to 37
- NCT of Delhi: 29 to 21
- Rajasthan: 73 to 50
- Uttar Pradesh: 84 to 56
- Bihar: 57.3 to 36.6
- Assam: 73.9 to 49.1
- West Bengal: 35 to 25
- Jharkhand: 49 to 35
- Odisha: 69 to 46
- Chhattisgarh: 63 to 40



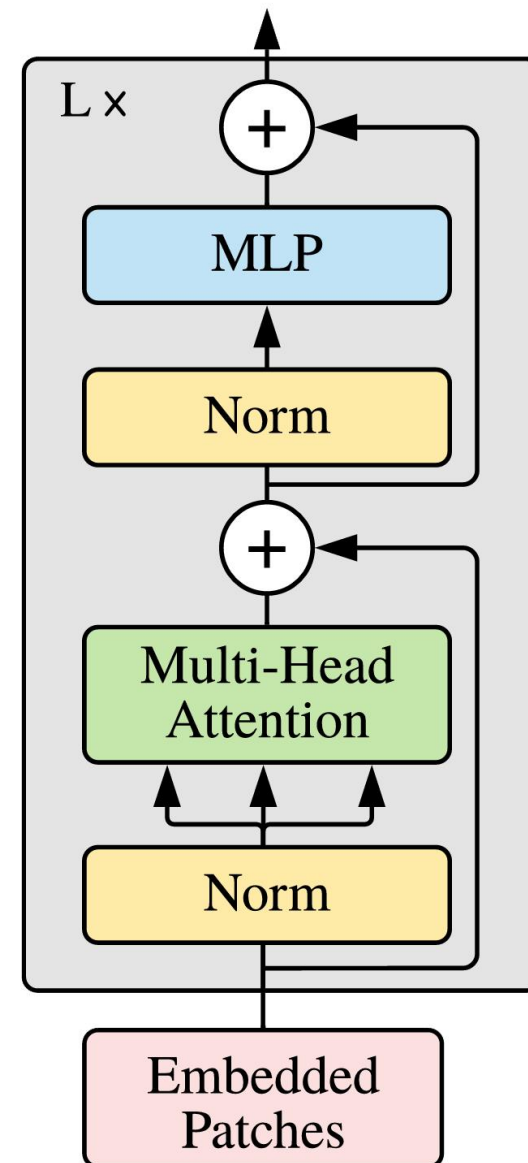
Database

## 02 项目主要内容——模型初步设计

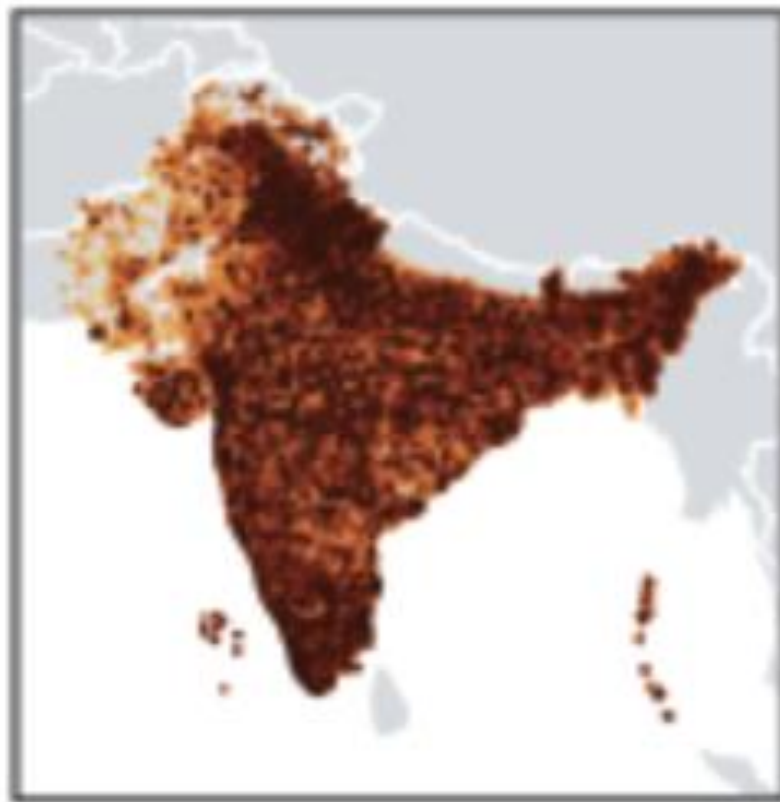
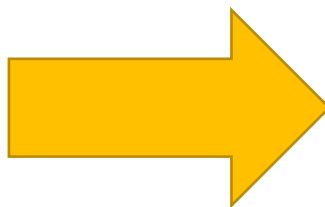
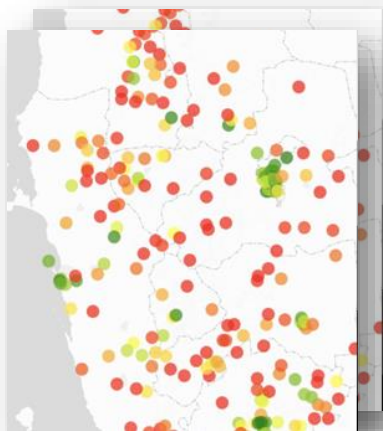
### Geography Vision Transformer (GViT)



### Transformer Encoder



## 02 项目主要内容——模型应用



### 学科交叉

公共卫生、地理信息、统计、计算机等多个领域交叉  
这些领域的交叉研究较为有限

### 数据支持

本项目与斯坦福大学的团队进行合作，获得印度地区9万多个村庄级采样数据



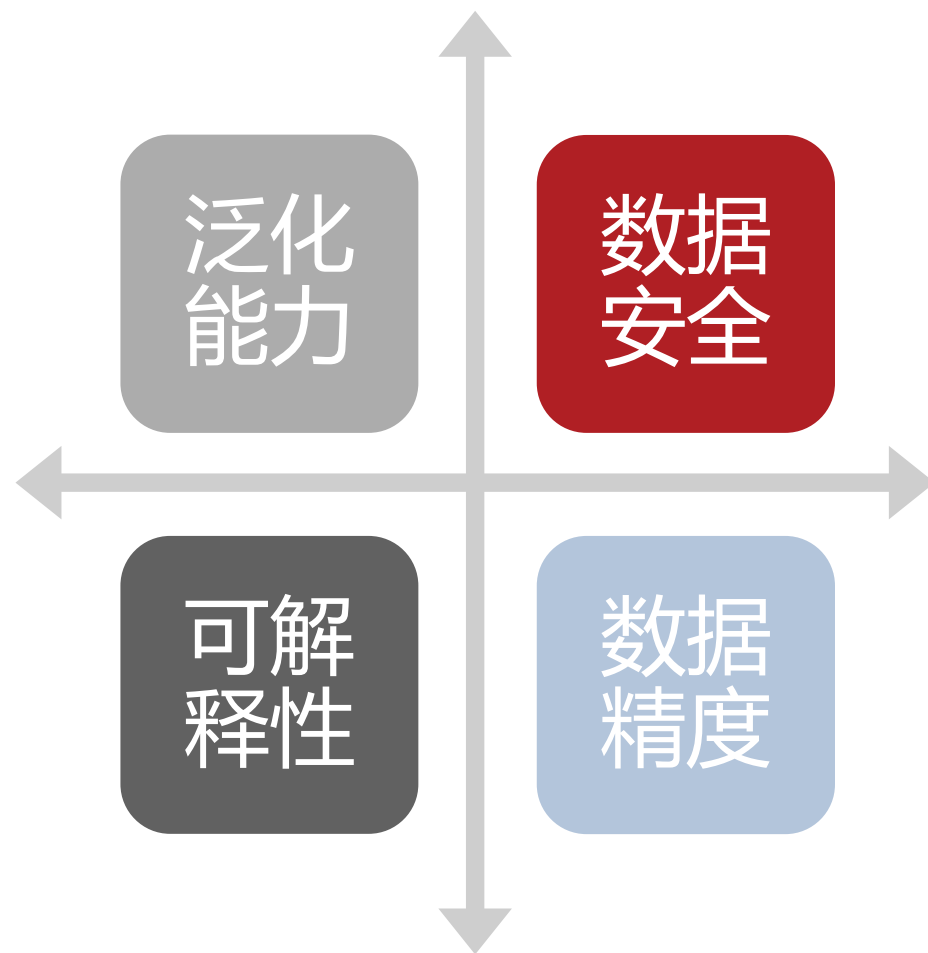
### 创新性

- 大语言模型数据挖掘
- 基于计算机视觉的社会学研究

### 应用前景

- 社会学研究新思路
- 公共卫生决策科学依据

### 03 项目优势和风险——风险



- (1) **论文成果**：总结项目中涉及的卫星图像处理技术、大语言模型应用技术和统计数据挖掘方法，为相关领域的研究和应用提供参考和借鉴。
- (2) **系统开发**：建立基于卫星图像的社会公共卫生数据预测模型，开发亚非拉等地区公共卫生指标的预测系统。
- (3) **社会效益**：实现对亚非拉等地区社会学数据的预测，为政府和国际组织制订针对性的政策提供科学依据，推动对第三世界国家的人文关怀。

- [1] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [2] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [3] Ma, S., He, B., Xie, B. *et al.* Investigation of the spatial and temporal variation of soil salinity using Google Earth Engine: a case study at Werigan–Kuqa Oasis, West China. *Sci Rep* **13**, 2754 (2023).
- [4] Bationo, C.S., Gaudart, J., Dieng, S. *et al.* Spatio-temporal analysis and prediction of malaria cases using remote sensing meteorological data in Diébougou health district, Burkina Faso, 2016–2017. *Sci Rep* **11**, 20027 (2021).

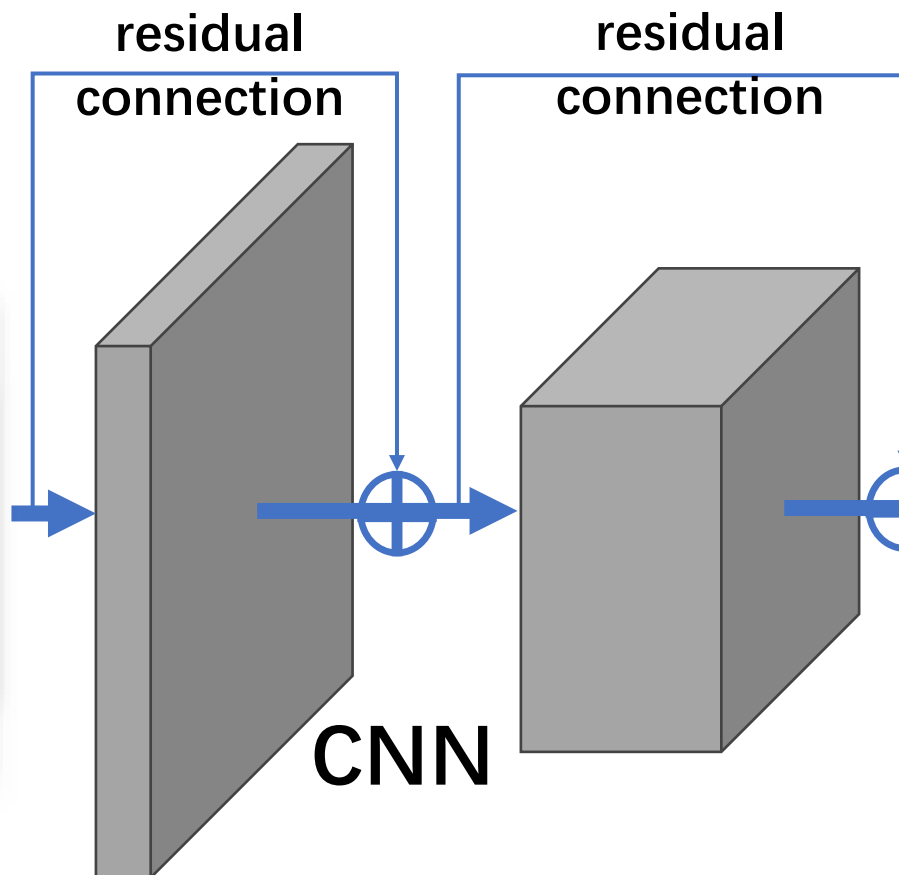


# 谢谢



- 专用材料费 6,000 元；用途：GPT PLUS API 2000, GPU服务器4000
- 印刷费与资料费 500元；用途：印刷和购买相关资料
- 交通与差旅费 500 元；用途：往返紫金港和玉泉校区等
- 出版费 1,000 元；用途：论文发表

Input images:  
同尺度卫星图片集



Embedding features:  
附加特征  
如当地传染病感染率等

