



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：

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

01

项目背景及意义

Project Background And Significance



近年来，亚非拉地区的一些国家因长期战乱、医疗资源匮乏及基础设施不完善等多重因素，面临着严重的**公共卫生挑战**。

同时，这些数据收集和监测体系也相对薄弱，目前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，实地考察数据采样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面临巨大困难。

因此，我们亟需一种新的、高效的数据收集和分析方法，以解决这些问题。

本研究的主要目标是运用大语言模型（LLM）强大的总结归纳能力，对已有的大量社会公共卫生数据进行深度挖掘和分析，并结合卫星图像数据，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。

该模型将有助于我们更准确地预测国家或地区的公共卫生指标，如新生儿死亡率、传染病感染率等。我们希望能够提供针对亚非拉地区国家的公共卫生情况的准确预测，为政府和国际组织制定针对性的卫生干预措施提供科学依据。

(1) 通过卫星图像数据，我们可以获取第三世界国家的各项关键指标，如人口分布、城市化程度、公共卫生设施分布等。该方法可以为政府和国际组织制定更有效的公共卫生政策提供科学依据，而且可以进一步推动国际社会对第三世界国家的**关怀和援助**。

(2) 本研究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。

02

项目主要内容

Main content of the project



- (1) **LLM大语言模型**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

REVIEW ARTICLE

OPEN



Digital public health surveillance: a systematic scoping review

Zahra Shakeri Hossein Abad ^{1,2}✉, Adrienne Kline^{1,3}, Madeena Sultana^{1,2}, Mohammad Noaeen⁴, Elvira Nurmambetova¹,
Filipe Lucini ^{1,5}, Majed Al-Jefri ^{1,3} and Joon Lee ^{1,2,6}

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

 Check for updates

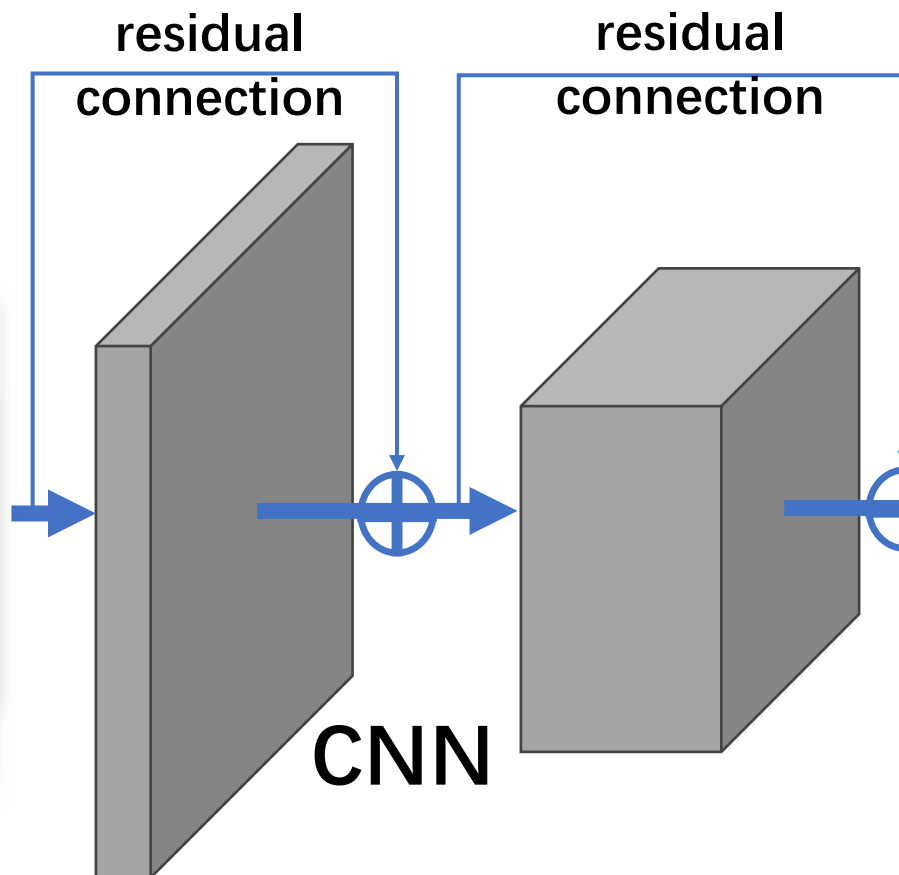
OPEN

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

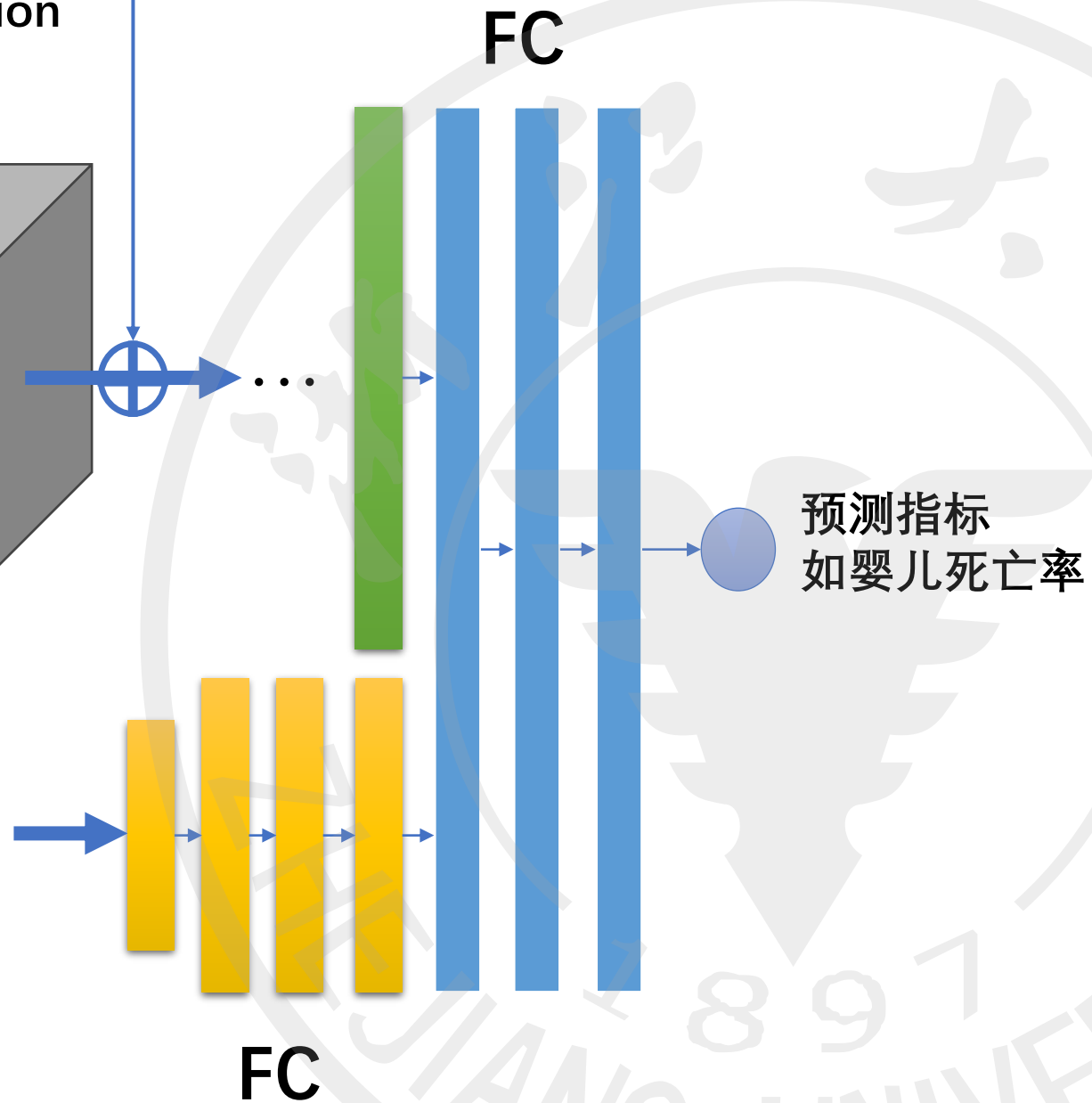
02 项目主要内容——主要内容



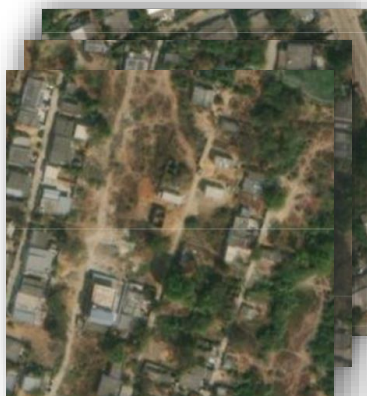
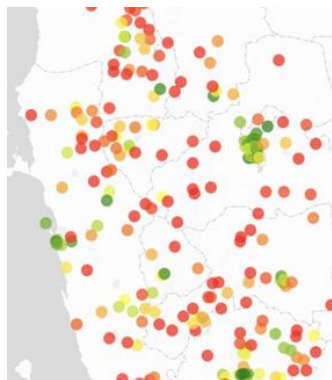
Input images:
同尺度卫星图片集



Embedding features:
附加特征
如当地传染病感染率等



02 项目主要内容——模型应用



03

项目优势和风险

Advantages and risks of the project



03 项目优势和风险——优势

学科交叉

结合公共卫生、地理信息、统计、计算机等多个领域

数据支持

本项目与斯坦福大学的团队进行合作



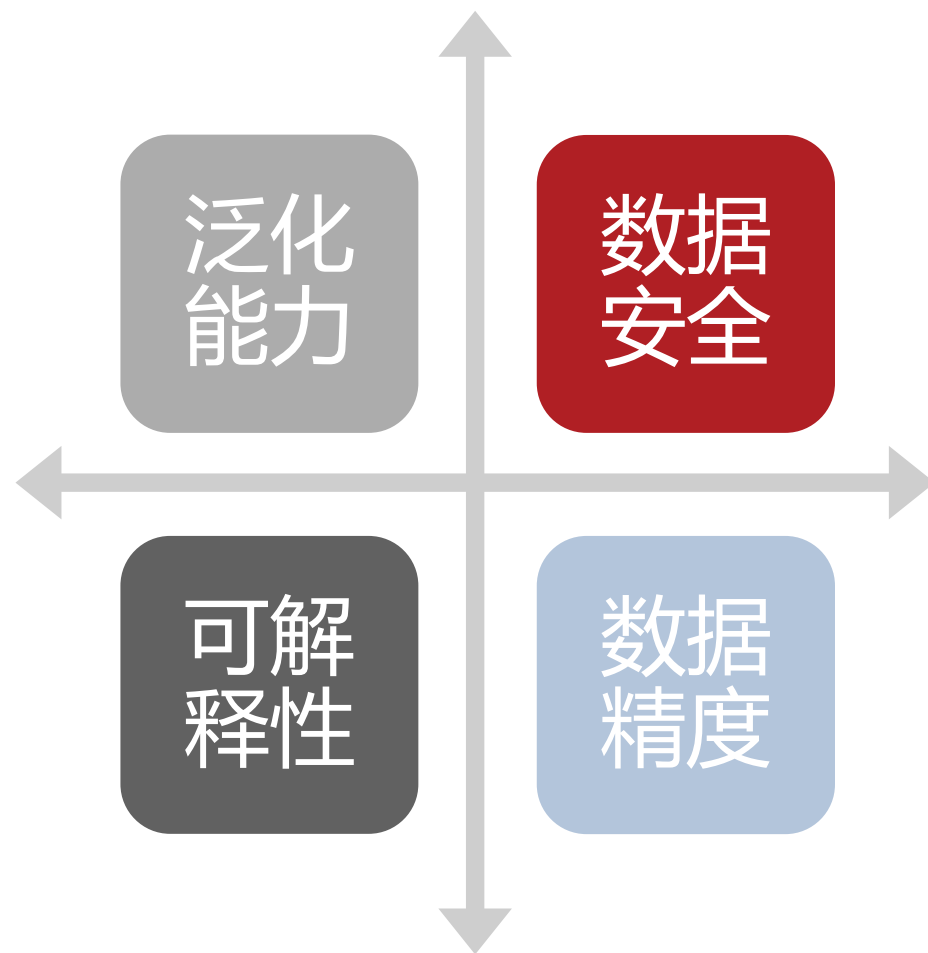
创新性

当前在这一领域的交叉研究较为有限

应用前景

对于亚非拉地区的公共卫生情况评估和决策具有广阔的应用前景和意义

03 项目优势和风险——风险



04

项目预期成果

Expected result of the project



(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) **验证评估**：验证该模型的准确度和泛化能力，展示项目成果在社会学数据统计领域的应用潜力和价值。

(4) **推广成果**：将模式推广到社会学研究领域，为该领域研究开辟新的道路。

(5) **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。

谢谢

