**1. Project Idea:** Forecasting Disease Outbreaks Using Integrated Mobility and Health Data

**Problem/Goal:**  Rapid and unpredictable outbreaks of infectious diseases pose major public health challenges. The project aims to predict spikes in disease cases by integrating real-time population mobility data with health statistics. The goal is to develop an early warning system for identifying and mitigating potential outbreaks.

**2. Relevance to Sustainable Development Goals (SDGs)**

This project directly addresses **SDG 3: Good Health and Well-Being**, which aims to combat epidemics and promote health security. Predictive tools can improve governments' and health agencies’ readiness, enabling timely interventions, optimizing resource allocation, and saving lives during outbreaks.

**3. Literature Examples**

* [**Kraemer et al. (2020). “The effect of human mobility and control measures on the COVID-19 epidemic in China.”**](https://pubmed.ncbi.nlm.nih.gov/32213647/)This study demonstrated that changes in human mobility, derived from mobile data, strongly predicted the spread of COVID-19, providing evidence that mobility restrictions effectively suppress outbreaks.
* [**Pepe et al. (2021). “COVID-19 outbreak response: a first assessment of mobility changes in Italy following national lockdown.”**](https://www.researchgate.net/publication/340242855_COVID-19_outbreak_response_a_first_assessment_of_mobility_changes_in_Italy_following_national_lockdown)The authors analyzed Google mobility data and linked it with COVID-19 epidemic curves, showing a significant association between reduced mobility and slowed disease transmission.

**4. Describe Your Data**

* **Sources:** [WHO Global Health Observatory](https://www.who.int/data/gho) (CSV/Excel): Daily and weekly country-level infectious disease statistics, [Google Community Mobility Reports](https://www.google.com/covid19/mobility/) (CSV): Daily location/mobility trends by country and region, [Global Human Epidemic Database (Nature paper)](https://www.nature.com/articles/s41597-025-04663-z): Outbreak catalogue, event-level
* **Preprocessing:** Align data spatially (region/country) and temporally (date/week), Handle missing values, normalize metrics (percent change, per capita), Engineer lag features (mobility values from prior weeks)

**5. Approach (Machine Learning or Deep Learning)**

We propose using a **deep learning approach** (such as LSTM time-series neural networks) because the problem requires capturing non-linear and delayed (lagged) effects between population movement and outbreak timing. Data is high-frequency, sequential, and multi-dimensional, making deep learning especially effective for pattern discovery and forecasting.

**Justification:**Deep learning, particularly sequence models (LSTMs), can model complex temporal dependencies and interactions between mobility shifts and health outcomes far better than basic regression or classical forecasting.