Working title: Implementing Machine Learning Techniques in Epidemic Modelling

Overview:

This literature review will critically assess the use of machine learning (ML) techniques in epidemic modelling, with a particular focus on their implementation, methodological robustness, performance, and interpretability. Drawing on recent peer-reviewed research, it will explore how ML compares with traditional models, such as SEIR and agent-based simulations, and identify key challenges, gaps and future directions in this emerging field.

Planned Structure:

1. **Introduction**

- Defining epidemic modelling and explaining its relevance to public health and policy.
- Outlining the growing role of ML in epidemic forecasting and surveillance.
- Contrasting ML approaches with traditional models (e.g., SEIR, agent-based).
- Stating the aim of the review: evaluating the implementation of ML techniques in epidemic modelling, focusing on methodological rigour, performance, and interpretability.
- Clarifying the scope: a secondary literature review based on peer-reviewed academic sources.

2. **Overview of Machine Learning Techniques in Epidemic Modelling**

- Describing key ML approaches used in this context:
  – Supervised models (e.g., Random Forest, SVM, XGBoost)
  – Time-series models (e.g., LSTM, GRU)
  – Hybrid models (e.g., ML combined with compartmental models like SEIR)
- Identifying use cases across diseases (e.g., COVID-19, influenza, dengue).

- Summarising types of data used for modelling (e.g., case numbers, mobility, weather, social media).

3. **Implementation and Methodological Approaches**

- Discussing typical data processing techniques (e.g., cleaning, feature selection, time lags).
- Evaluating model training and validation practices (e.g., cross-validation, hold-out sets, external testing).
- Reviewing performance evaluation metrics (e.g., RMSE, MAE, AUC) and their strengths and limitations.
- Considering reproducibility and transparency (e.g., code availability, open data, documentation).
- Comparing methodological quality across studies, highlighting good and poor practices.

4. **Interpretability and Practical Use**

- Exploring the challenge of interpretability in ML models, especially in health applications.
- Introducing popular tools and techniques (e.g., SHAP, LIME, feature importance).
- Explaining why interpretability matters in public health decision-making and risk communication.
- Critically reviewing how different studies address or neglect interpretability.
- Considering the extent to which ML models are being adopted in practice, and discussing barriers to real-world implementation.

5. **Challenges, Gaps and Opportunities**

- Identifying common limitations in the literature: data quality, generalisability, model bias, and ethical concerns.

- Highlighting the lack of standardisation in ML model development and reporting.
- Discussing promising directions, such as integrating ML with epidemiological models, advancing explainability, and using real-time data streams.
- Suggesting areas where research and practice could converge to improve public health outcomes.

6. **Conclusion and Future Research Directions**

- Summarising the strengths and weaknesses of ML approaches in epidemic modelling.
- Emphasising the need for methodological consistency, interpretability, and transparency.
- Recommending that future research focus on standard benchmarks, clearer reporting, and closer collaboration with epidemiologists and public health stakeholders.