**Paper Introduction**

**Introduction**

This is an Introduction to the paper by Niksirat (2023), entitled Changes in research ethics, openness, and transparency in empirical studies between chi 2017 and chi 2022.

**Background**

Vaccine hesitancy (VH) poses a significant challenge to global health initiatives, particularly in the context of mass immunization campaigns such as those aimed at combating COVID-19. VH not only jeopardizes efforts to achieve herd immunity against infectious diseases like COVID-19 but also threatens progress made in eradicating vaccine-preventable diseases such as measles and rubella. Several individual-level factors have been identified as associated with VH, including socioeconomic status, trust in conventional medicine, and risk perception. However, collecting such data on a large scale is challenging, especially during time-sensitive vaccination campaigns. Therefore, leveraging readily available area-level indicators to predict communities at high risk of VH could provide a valuable alternative for policymakers.

**Research Question**

The paper addresses the question of whether machine learning algorithms can effectively predict vaccine hesitancy at the community level using area-level indicators. Specifically, the study seeks to determine which machine learning model performs best in predicting VH risk and which area-level indicators are most influential in this prediction.

**Application Scenarios**

The findings of this research have significant implications for public health policymaking, particularly in the context of planning and implementing vaccination campaigns. By identifying communities at high risk of vaccine hesitancy, policymakers can target interventions more effectively, such as tailored awareness campaigns or increased access to vaccination services.

**Methodology**

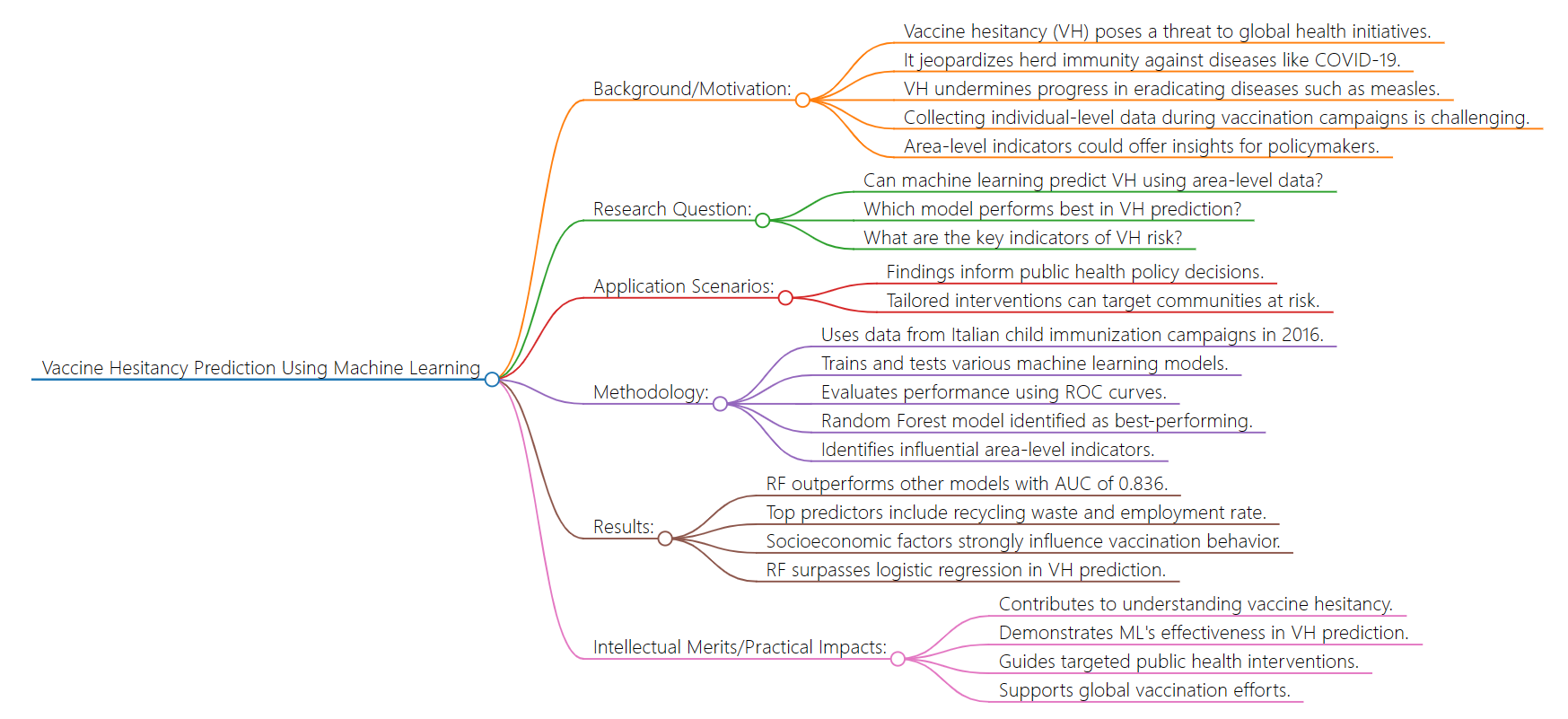
The study utilizes data from child immunization campaigns conducted in Italian municipalities in 2016, encompassing seven vaccine-preventable diseases. Various machine learning models, including LASSO, Random Forest (RF), neural network (NN), and gradient boosted machine (GBM), are trained and tested on a dataset comprising municipality-level indicators. The performance of these models is evaluated using receiver operating characteristic (ROC) curves, with the RF model identified as the best-performing algorithm. Feature importance analysis is conducted to identify the most influential area-level indicators in predicting VH risk.

**Results**

The RF model outperforms other machine learning algorithms, achieving an area under the ROC curve (AUC) of 0.836. Key predictors of VH risk identified by the RF model include the proportion of recycling waste and the employment rate. These findings are consistent with those of other studies and underscore the importance of socioeconomic and environmental factors in shaping vaccination behavior. Additionally, the study confirms the significance of RF in predicting VH risk, with superior performance compared to logistic regression.

**Intellectual Merits/Practical Impacts**

The research contributes to the growing body of literature on vaccine hesitancy by demonstrating the efficacy of machine learning techniques in predicting VH risk at the community level. By leveraging area-level indicators, policymakers can develop targeted strategies to address vaccine hesitancy and improve vaccination coverage. The findings have practical implications for public health interventions, especially in the context of global efforts to combat infectious diseases and achieve immunization targets.



**Fig 1. Paper Summary**

**Reference**

Niksirat, K. S., Goswami, L., Tyler, J., Silacci, A., Aliyu, S., Aebli, A., ... & Cherubini, M. (2023). Changes in research ethics, openness, and transparency in empirical studies between chi 2017 and chi 2022.