Breathe In, Breathe Out, combining machine learning with data analysis, fusion and data assimilation for incomplete, noisy air pollution data.

Project motivation: ~90% of the global population breathe polluted air resulting in ~7 million annual deaths. Pollution is a serious threat to public health with only cancer, obesity and heart disease causing more illness. Health risks associated with fine/ultrafine particulate matter ($PM_{2.5}$/$PM_{0.1}$) and gases (NOx/SO_2/O_3) include development/exacerbation of lung disease.

Allergic fungal airway diseases including severe asthma with fungal sensitization and allergic bronchopulmonary aspergillosis affect 11.3million people globally. Treatment options are toxic, ineffective, and lead to antimicrobial resistance. Studies show a temporal association and immunogenicity of airborne fungal load with air pollutants, however the geographic susceptibility related to air pollution is undetermined.

Project proposal: This project aims to accurately predict air pollution propagation at city scale and its effects on fungal allergenicity and antifungal resistance developing novel data-driven models. The student will have access to a big data set to develop novel data-driven models combining machine learning with data analysis, fusion and data assimilation for incomplete, noisy big-data problems, including uncertainty quantifications and minimizations.

Available data for a strong case study: In this study, we propose to integrate, through data science tools (data assimilation and data fusion), comprehensive air sampling and low-cost wearable data across London collected during the case studies of the projects MAGIC and INHALE and used to develop environmental prediction models of air pollution, with longitudinal electronic health records (EHR) including allergenicity, microbiology and hospital admission data. This will be ideally suited to the DISCOVER-NOW data, however, existing EHR datasets at 3 tertiary care centres can also be used. Further data integration will be performed with ongoing Wellcome and MRC funded longitudinal air and soil fungal environmental sampling (FREAL study) and existing UK environmental *Aspergillus sp* resistance datasets.

How does this project advance AI beyond the state-of-the-art: In the past few years, ML models have shown great capability in making forecasts. However, the training data provided to ML technologies, is noisy and heterogenous introducing errors which are trained in the ML model. This project integrates ML with Data Science tools such as Data Assimilation (DA). DA reduces errors in the ML model by including data with higher accuracy and medical meaning, merging and fusing several types of observed data and increasing the reliability of prediction. The resulting cohesion of ML and DA is then blended in a future state-of-the-art generation of fast and more accurate predictive models.

Expected impact that the project will have within 3.5 years: At the end of the project, the student will have developed an accurate ML forecasting model to predict the effects of air pollution on fungal sensitisation and fungal disease burden. This will be validated through existing collaborations at global leading centres. The technology developed by the student at the end of the projects are completely general and can be used to model and predict other heterogenous diseases and real-world problems outside a healthcare setting. Expected data outputs will be used to inform national pollution mitigation strategy and commercial pharmaceutical interest in novel allergy biologic therapy and antifungal compounds.

Student backgrounds for this project: Any motivated student with a background in computational science.

The candidate will be part of the DataLearning working group at Earth Science and Engineering & Data Science Institute. In the course of the proposal development and project, new competences (fundamental for career progression) will be acquired by the student into the context of mathematical method for Data Assimilation and Machine Learning.

Contact:

Dr Rossella Arcucci (r.arcucci@imperial.ac.uk) for more information. Details for how to apply can be found at (www.imperial.ac.uk/study/pg/apply/how-to-apply)