

INTERVIEW OF COORDINATOR GIANNA TSAKOU

1. What is INCISIVE?

(Please explain the full title of the project and mention the funding agency)

INCISIVE is a 42-month project, funded by the EC under the Horizon 2020 Programme, specifically the call DT-TDS-05-2020, focusing on AI for Health Images. Our project is *exploring* and *supporting* in different ways the potential of Artificial Intelligence (AI) tools for enhancing state-of-the-art health imaging solutions for four types of cancer: breast, colorectal, lung and prostate cancer.

INCISIVE is *exploring* the potential of AI tools by developing an AI-based toolbox consisting of novel AI models, combined with a set of predictive, descriptive and prescriptive analytics, that aim at improving the sensitivity and specificity of cancer imaging methods, thereby, enabling better-informed decision-making of clinicians in charge of diagnosing and treating the four types of cancer addressed by the project. INCISIVE not only explores the potential of AI tools in health imaging but it is also *supporting* the potential of AI tools by developing an interoperable federated health imaging repository for the afore-mentioned types of cancer. This repository will be used for training the models of the project's AI toolbox but will also be made available to the wider AI research community to facilitate further AI training and experimentation far beyond the project's lifetime.

2. How can Artificial Intelligence and Big Data help fight cancer?

The increasing amount and availability of cancer imaging and other health data in combination with advances in Artificial Intelligence (AI) and Machine Learning (ML), provide unprecedented opportunities for fighting cancer, provided that all appropriate ethical and legal measures are taken for ensuring health data privacy and security.

Using Machine Learning, AI models can be trained to detect patterns in cancer imaging data that may not be detectable by the human eye, thus, increasing the interpretability of imaging data in terms of cancer detection and classification, treatment optimization and, eventually, supporting more effective decision-making for healthcare professionals. Importantly, the effectiveness of any AI model is tightly linked to the amount and diversity of the data that has been used for its training. And this is where Big Data comes into the play: we need large volumes of high-quality data for AI training and validation to ensure the robustness, fairness and trustworthiness of the developed decision-support solutions.

We could say that AI and Big Data are two sides of the same coin –two big fronts– when aiming at enhancing current imaging methods. And INCISIVE has the ambition of making major contributions to both fronts.

3. What are the main expected results in the project?

(Please explain the algorithms and the federate repository)

As briefly mentioned above, INCISIVE targets two main results: A) an AI-based toolbox and B) an interoperable pan-European federated repository of health images.

The first major project result, the AI-based toolbox, will comprise AI models supporting the following: semantic and instance segmentation of medical images; deep machine learning methods for improving image-based diagnosis of cancer in terms of sensitivity and specificity; prognostic models via spatio-temporal simulation of cancer progression; descriptive, predictive and prescriptive data analysis pipelines for risk assessment and decision support. Our project aims at communicating the outputs of the AI models to healthcare professionals in a way that is as intuitive and transparent as possible, so our partners are conducting work on transforming medical images to medical reports and using Augmented Reality (AR) for enhancing visualization of AI outputs, as well as establishing an easy-to-understand recommendation mechanism. We are also working on implementing a framework for explainable AI so that the logic behind the AI models' outputs becomes transparent to the end user, addressing the "black box" challenge. Last but not least, we plan to implement a feedback mechanism for healthcare professionals and AI re-training framework so that the performance of the different AI models is improved over time.

The second major project result, the federated repository of health images, will allow the sharing and donation of health data by hospitals and other potential data providers with the wider scientific community working on AI, in compliance with legal, ethical, privacy and security requirements. It will be possible to use the shared data for AI-related training and experimentation. The INCISIVE data repository will store de-identified data following a federated approach, namely the data will not have to leave the data provider's site; accordingly, it will become accessible to interested AI researchers via federated learning mechanisms and, where necessary, will be supported by High Performance Computing (HPC) as-a-service. The data will comprise mainly cancer imaging data but also other relevant health data (a volume of around 30 TB), for the four types of cancer that our project is targeting. The repository will be initially created based on major contributions from our 9 data providers in the INCISIVE consortium. However, during and notably after the project, after our data sharing scheme will have been fully defined, implemented and piloted, we hope to attract other interested organisations willing to share data.

4. What benefits will INCISIVE bring to society?

(e.g. health and ICT professionals and academics, wider public, etc.)

With our two main results, and all side project results that are associated to these two main results, we aspire to create impact on multiple fronts.

With the AI toolkit we want to benefit healthcare professionals and, indirectly impact on the current patient care pathways, by supporting and better informing clinical decisions related to the diagnosis, treatment and follow-up of cancer through our work on improving the accuracy, specificity, sensitivity, interpretability and cost-effectiveness of existing cancer imaging

methods. Specifically, in lung cancer, we aim at supporting more accurate detection of malignant, non-small-cell lung cancer lesions, with a possible extension of the diagnostic capabilities of our AI models to solitary lung nodules. In breast cancer, we aim at optimization of treatment selection by clinicians through inference of histopathological information in non-invasive mpMRI imaging screenings, while in colorectal cancer we want to support improved risk stratification of patients, prediction of metastasis risk and prediction of the patient's response to treatment to positively impact the colorectal cancer clinical workflow. In prostate cancer, we focus on supporting early diagnosis using low-cost non-invasive diagnostic solutions, which can in turn facilitate reliable and affordable periodic screening, thus, leading to increased survival rate.

The federated data repository will create benefits for the wider community developing AI tools and experimenting with machine learning by making a large volume of data possible to share and reuse, fully respecting legal and ethical requirements. Besides increasing the availability of annotated, validated, high quality data, INCISIVE will also introduce a structured data donation mechanism that ensures that the data providers maintain full control of their data and that these are aligned with regulations and security aspects.

5. The first year of the project is over. What are next steps in the year to come?

During the first year, we focused on designing and architecting our two main results, the AI toolbox and the data repository. To do this, we consulted extensively with healthcare professionals and other relevant stakeholders within and outside of our consortium in order to understand their requirements and define how our AI tools could support them in improving the current care pathways for cancer patients. We also worked a lot on defining and addressing all legal and ethical issues related to the planned extensive data sharing. We now have a clearer idea of which are the cancer imaging challenges that our AI work can help to address and we have already kicked off most of our AI implementation work. We have also started the data annotation, quality checking and data sharing process for the retrospective data of the participating data providers. And we are almost half-way through our planning work for the collection and sharing of the prospective data that will result from our pilot activities.

In terms of technical work, we are close to completing the first version of the overall INCISIVE system technical specifications and architecture, which will give us a good basis for starting implementation work towards our first integrated prototype.

So the next steps in the year to come will continue all of the above activities and conclude some of them. One of the major foci for our consortium team will be the implementation of the initial version of the INCISIVE integrated prototype, which will enable basic federated data storage, data search and federated learning functionalities. We will continue with the implementation of all our AI work towards an initial prototype of the AI toolkit. We will also focus on ensuring data interoperability since our data is coming from multiple clinical centres and data sources. We still have a lot of work ahead of us but it is exciting.

6. What excites you the most about the project?

Many things! I find the great potential of the data repository and the huge opportunities that it will create for many researchers to be exciting. There are still huge challenges in AI development and validation, as well as in health data federation and sharing. E.g. how to ensure data interoperability, or how to facilitate and speed up data annotation and de-identification by (semi-) automating as much of the work as possible; how to train robust, fair and trustworthy AI tools? how to ensure full respect for patients' privacy and GDPR requirements; how to design a data sharing scheme that meets the diversity of as many data providers' requirements as possible. There are so many open issues linked to our project that the wider R&D community is striving to address for several years now in the EU and worldwide.

INCISIVE is operating at the forefront of innovation, striving to deliver cutting-edge scientific approaches and technologies. This is what makes it so difficult but, at the same time, so interesting and exciting. We know that we will not be able to address all these open issues in a conclusive way. But we *will* take smaller or bigger steps towards the right direction and try to make our contribution significant!