



AI4ES

RED DE EXCELENCIA EN
TECNOLOGÍAS HABILITADORAS
BASADAS EN EL DATO

RULES FOR THE DATATHON 2022 “AI4ES- EXCELLENCE NETWORK IN DATA-BASED ENABLING TECHNOLOGIES”

The Cervera Excellence Network AI4ES (excellence network in data-based enabling technologies), formed by the technology centres CTIC, EURECAT, ITI and TECNALIA, in collaboration with the Health Department from the Catalanian Government (Generalitat de Catalunya) and the company BASF, announces the 2022 AI4ES Datathon Awards according to the following RULES:

FIRST. – REASON FOR THE DATATHON AND SUGGESTED CHALLENGES

The main objective of the AI4ES Datathon is to encourage the use of artificial intelligence (AI) and data analytic technologies to solve challenges of business and social importance, which involve the processing of large volumes of data and recognize the best projects presented for the resolution of these challenges. To this end, the 2022 AI4ES Datathon Awards are organised to recognise the best students and professionals in the use of AI technologies for data processing.

In its 2022 edition, the AI4ES Datathon proposes two different challenges:

- ***Challenge 1: Disease detection and quantification based on drone images for crop protection***
- ***Challenge 2: Incidence and evolution of viral diseases***

The detailed description of these challenges can be found in FIFTH section of this document.

The projects proposed by the participants of the AI4ES Datathon must focus on solving **one** of the two challenges defined above.

SECOND. – PARTICIPANTS

Participation is open to all individuals of legal age, both individually and as a team. Each prize will be awarded to each project regardless of the number of participants in the project.

THIRD. – SUBMISSION DEADLINE

The deadline for submission of applications will begin the day of publication of this document on the website and will end at **23.59 pm on the 25th November 2022**, through the application form available on the website <https://www.ai4es.com/ai4es-datathon>. The organisation reserves the right to extend the closing date for applications to encourage a greater participation.

FORTH. - CATEGORIES AND PRIZES

There are three categories of participation.

- a) **Starters category ("Future Promises")**, for the presentation of work by students and recent graduates (all participants in the project must be either currently studying for a degree or have obtained their degree after 31/12/2018).
- b) **Professionals' category ("Specialists")**, for the submission of work by graduates and professionals from companies with more than three years of professional experience (i.e. when the conditions of the initiation category are not met).
- c) **AI4ES category ("AI4ES Members")**, for the submission of work by researchers from the AI4ES network member centres (when the researcher or a member of the team is part of one of the AI4ES network member centres).

A prize of **1,500 euros** will be awarded to the best project in each challenge in the "Future Promises" category, and **2,000 euros** to the best project in each challenge in the "Specialists" category.

Each of the prizes will be equally shared among all the members of the winning project group. Each of the members of the winning project teams must provide the personal and fiscal documentation necessary to the AI4ES network to be able to award the corresponding prizes. Income withholdings will be applied, where applicable, as established by the current Spanish legislation.

All projects, as well as their presentations, regardless of whether they have been awarded a prize or not, may be communicated by any means that the AI4ES network deems appropriate to achieve a maximum dissemination.

FIFTH. - CHALLENGES SUGGESTED

CHALLENGE 1 - DISEASE DETECTION AND QUANTIFICATION BASED ON DRONE IMAGING FOR CROP PROTECTION

BASF's Agricultural Solutions division conducts hundreds of field trials each year to evaluate the biological performance of new crop protection products. Currently, most of the data is generated through visual assessments by experienced field research agronomists. For example, in fungicide trials, the field agronomist assesses the percentage of crops infected by each disease in each plot.

BASF is digitising and modernising this assessment process to become more efficient and improve data quality, e.g. by using handheld devices (smartphones) and drones.

The aim of this challenge is to train a robust image regression model, which processes images of individual field plots and predicts their assessment value. For this, the data used will be multi-spectral images captured by drones in different years and field trials in wheat crops at various locations (with different diseases and disease levels) and the ground-truth will be provided by manual assessments by field agronomists.

Aspects to be taken into account in this challenge are:

- 1) Depending on the test and its location, different sensors with varying image resolution are used. Some tests are captured with 10-band sensors (including infrared channels and red edge band), others are captured with 5-band sensors, and for some tests only standard RGB images are available.
- 2) The labels, which represent the "ground-truth", are in fact noisy human estimates of the actual disease state. Therefore, there may be outliers or errors to varying degrees in the assessment values provided.
- 3) The data provided are not balanced, the most common assessment ratings are close to 0 and the highest rating values from 50 to 100 are less frequent in the dataset.
- 4) In addition, some plots are affected by one disease, and some plots are affected by a combination of diseases.

For this challenge, the following will be provided:

- The training data, including RGB, 5-band and 10-band images for field plots and their target assessments (labels) in CSV format.
- An example of a training process, based on Python and XGBoost, which serves as a starting point for experimentation and development, although each participant is free to use it or start from scratch.

Participants who register for this challenge will receive a link to download and run the dataset and the example of a training process. Participants may only use the data and code provided to participate in the Datathon and will delete the data at the end of the

Datathon. BASF may provide broader permission to use the data under a specific agreement.

CHALLENGE 2 - INCIDENCE AND EVOLUTION OF VIRAL DISEASES

With the open data on the evolution of the Covid-19 pandemic in Catalonia, obtained through the Sivic website, the aim of this challenge is to make progress in improving the creation of predictive models capable of adjusting/predicting the future progression of Covid-19.

To this end, the set available until week 22 of 2022 is proposed to be used as training data, and the rest of the current year, as test data. This test period (between approximately weeks 22 and 38) is of particular interest, as this is when the last spike of covid-19 cases (the sixth wave) occurred, from week 22 onwards.

Therefore, the aim of this challenge is to create models capable of providing four types of forecasts for the Catalonian data:

- Incidence,
- Hospital admissions,
- ICU,
- Deaths.

The metrics to be reported in the test set must include at least MAE and MAPE. These will be computed 1, 3, and 5 weeks ahead.

Since all data (including test data) are publicly available, five aspects will be assessed in addition to the effectiveness of the models in the test period, such as: the soundness of the methodology used, the novelty and quality of the solution, the interpretability of the models (justification of the reason for the results reported by the model), the visualisation (variety and quality) of the results, and finally, the clarity and objectivity of the presentation.

Datasets

More information on open data in Catalonia and Sivic in:

<https://sivic.salut.gencat.cat/>

<https://analisi.transparenciacatalunya.cat/>

SIXTH - PRESENTATION OF PROJECTS

Participants must send a summary report of the project carried out in PDF format by e-mail to datathon@ai4es.es , with a maximum of five pages, and no later than **23:59 pm on the 17th January 2023**. This summary report must explain: the approach adopted to solve the challenge, the characteristics of its development, and the main results and conclusions.

In addition, each participant must make an oral presentation of the solution developed with a maximum duration of 10 minutes, during an online evaluation session that will be convened between the 18th and/or the 20th January 2023, between 9:00 am and 15:00 pm.

SEVENTH - JURY, DECISION AND PUBLICATION OF THE AWARDS

The jury will be composed of one or more evaluation teams, made up of members of the AI4ES network and members of the AI4ES Advisory Committee. Each evaluation team will be made up of five people: one from each AI4ES network member technology centre (CTIC, EURECAT, ITI and TECNALIA), and one representative of the AI4ES Advisory Committee. For the categories of challenge 1 "Disease detection and quantification based on drone images for crop protection", there will also be a representative of the company BASF.

The winners will be announced on the **24th January 2023**, after the evaluation sessions and consensus meetings among Jury members have been completed. The winning projects will be made public through the communication channels of the AI4ES network. The decision will also be communicated to the winners by e-mail.

The jury may decide, in the event that the projects submitted are of insufficient quality or there is low participation, that one or more of the prizes will be declared void.

EIGHTH - EVALUATION CRITERIA

The winning projects will be chosen based on the following criteria:

1. The impact and importance of the challenge solved.
2. Creativity and innovation.
3. The use of data.
4. The degree of completion achieved.
5. The clarity of the oral presentation.

NINTH - PARTICIPATION CONDITIONS

Participation in these prizes is subject to the following conditions:

- It is the responsibility of the participants to ensure that the proposal is original and that it solves one of the challenges proposed in section FIFTH.
- The AI4ES network may publish the results of the projects developed and submitted to the 2022 AI4ES Datathon in any media it deems appropriate. In addition, they may disseminate any information related to the results of each project and the people who proposed it to make it known by any means they deem appropriate, as well as images and videos of the participants for the purpose of dissemination and promotion of the event.

TENTH. - DATA PROTECTION



In accordance with the provisions of the current Spanish legislation on data protection, the AI4ES network will process the data provided in the application, as well as those contained in the accompanying documentation, to manage the registration, participation and resolution of the call for the 2022 AI4ES Datathon Awards.

The legitimacy of this processing is based on the public interest of the call and on the consent of the person concerned in their application to participate in it. The data relating to the winners of the awards may be published on the AI4ES network website and on social networks.

The details of the prize winners in the categories that include a financial benefit will be communicated to bank entities to make the payment, to the State Tax Administration Agency, to the General State Comptroller's Office and to the Court of Auditors for accounting control.

The data will be kept for the duration of the awards procedure, with the Spanish archives and documentary heritage regulations being applicable. The data of the prize winners in the categories that include a financial benefit will be kept in accordance with the provisions of Law 58/2003, of 17 December, General Taxation.

ELEVENTH. – RULES ACCEPTANCE

Participation in this competition implies the acceptance of these rules, which may be interpreted by the jury and the organising body in those aspects not initially foreseen. More information at <https://www.ai4es.com/ai4es-datathon/>