HR EXCELLENCE IN RESEARCH

UNIVERSIDADE DA CORUÑA

citic

3-i ICT – Guide for applicants

Revision history

Version	Date	Description
1.0	25/03/2022	1st version

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement Nº 101034261.

Table of contents

1 CITIC at the UDC

CITIC, the Centre for Information and Communications Technology Research at the University of A Coruña, is a unique research centre with over 150 affiliated researchers and more than 50 active research projects with competitive funding from regional, national and European calls. The annual scientific output of CITIC's research includes approximately 175 articles in JCR journals and over 200 international conference papers.

ICT covers a remarkable spectrum of scientific inquiry, from pure basic to highly applied research. Research at CITIC is organised around five areas: artificial intelligence, data science and engineering, high performance computing, intelligent networks and services, and cybersecurity.

- Artificial intelligence focuses on the design and programming of machines capable of performing tasks that require intelligence, and offers a wide range of cross-disciplinary applications.
- Data science and engineering is an inherently multidisciplinary field, and has become an increasingly necessary tool in the era of big data.
- High performance computing is essential for processing the large data sets needed to understand and meet social, scientific and industrial challenges across a wide range of fields.
- Intelligent Networks and Services is a highly transferable, intersectoral area, especially in the new Industry 4.0 environment.
- Cybersecurity is a cross-disciplinary field that draws on and feeds into the other key areas of research at CITIC, including data processing and management, artificial intelligence systems, computation systems, and online services and communications.

2 About 3-i ICT

The 3-i ICT International, Interdisciplinary and Intersectoral Information and Communications Technology PhD Programme is a 60-month H2020 COFUND project, co-funded by the European Union under a Marie Skłodowska-Curie grant agreement (101034261). The overall project is worth €1,301,760, half of which is financed by the EU. The project will run from 1 January 2022 to 31 December 2026.

The aim of 3-i ICT is to provide eight Early-Stage Researchers (ESR) with unique opportunities for basic and applied interdisciplinary research, training and career development. To ensure the 3-i dimension, during the fellowship period, ESRs will:

- Acquire real experience and practice on interdisciplinary projects that bridge the gap between different fields of science, under the supervision of one expert from the field of ICT and one expert from a different field.
- Carry out secondments in non-academic institutions and/or companies to broaden and deepen their skills.
- Obtain International Doctorate certification, in accordance with UDC regulations.

3-i ICT is a catalyst programme, designed to transform young, talented PhDs into the next generation of independent ICT research leaders by widening and developing their competences across a range of areas and experiences. The programme also aims to boost ESRs' emerging careers by enhancing their employability in both academic and non-academic sectors.

The ESRs selected for the programme will complete a PhD degree with a 36-month contract of employment under the co-supervision of a CITIC researcher and an expert from another academic field. ESRs will be enrolled on one of six academic PhD programmes operated by or in partnership with CITIC:

- Computational Science.
- Information and communications technology.
- Information technology and mobile network communication.
- Information technology research.
- Mathematical modelling and numerical simulation in engineering and applied science.
- Statistics and operational research.

The programme will have a significant impact on CITIC's ability to attract, build and retain talent, and the interdisciplinary PhD projects undertaken by the ESRs will open up new lines of research and collaborations within CITIC and with other entities around the world.

The evaluation and selection process for candidates will be open, transparent and merit-based, to ensure the highest standards of integrity, fairness and transparency, in accordance with the principles of the European Charter for Researchers and the Code of Conduct for the Recruitment of Researchers. In recognition of its implementation of these principles across the whole university community, UDC was granted the HRS4R award for "HR Excellence in Research" in 2017. For more information about the evaluation and selection process, see Section 6.

3 PhD projects

All PhD projects will have a **supervision team** that will comprise:

- one main supervisor: a CITIC researcher qualified to guide a PhD thesis through to completion and provide the candidate with training and support of the highest standard.
- one secondary supervisor: an experienced researcher from another academic discipline from any national or international academic institution, research organisation or private company.
- one non-academic supervisor to monitor intersectoral secondments and help prepare the candidate for life outside of academia.

Each fellow and their/her/his supervision team will create a **personalised career development plan** (PCDP) to help them to map out their goals, interests and needs. The PCDP will comprise a description of the PhD project and the main research objectives, potential risks, timeline, contingency plans, ethical considerations, a secondment plan, and a plan for dissemination, exploitation and public engagement.

The PCDP will also include details of the fellow's plan for **research and non-academic transferable skills training** to enhance their research abilities and potential. The training programme for 3-i ICT will combine different types of activities, courses and methodologies and will be adapted to the individual needs of each fellow. All activities will be funded by 3-i ICT and conducted in English.

Applicants may select up to four PhD projects from the ten offered in the first call for candidates. Each of the PhD projects is linked to one of CITIC's five research areas and to one of the five External Panels of experts (for more details, see section 6). Only four positions will be funded in this first call. For a full description of the PhD projects, see Annex I.

CODE	TITLE	CITIC RESEARCH AREA EXTERNAL PANEL	INTERDISICIPLINARY RESEARCH AREA
2022-C1-001	Addressing challenging optimization problems in cell signalling networks with High Performance Computing and Cloud-based approaches	High Performance Computing	Biochemistry and Molecular Biology
2022-C1-002	Modelling complex biological phenomena via inverse optimal control and inverse reinforcement learning	Data Science and Engineering	Chemical Engineering Biochemistry and Molecular Biology
2022-C1-003	Distributed and parallel algorithms for inference of cell lineage trees	High Performance Computing	Genetics
2022-C1-004	Linking Linguistics to Low-resource NLP Neural Models	Artificial Intelligence	General Linguistics
2022-C1-005	Sequence Labelling Parsing for Applied Natural Language Processing	Artificial Intelligence	General Linguistics
2022-C1-006	New microbiomics algorithms and data analytics in colorectal cancer	Data Science and Engineering	Microbiology
2022-C1-007	Automatic animal behaviour analysis from video data	Artificial Intelligence	Ecology
2022-C1-008	Advances on Age-related Macular Degeneration treatment response prediction by means of ocular preclinical image analysis	Artificial Intelligence	Pharmacology
2022-C1-009	Wireless virtual sensing for control applications	Intelligence Services and Networks	Mechanical Engineering
2022-C1-010	Flexible cure models in data science to predict sustained remission in rheumatoid arthritis	Data Science and Engineering	Medicine

4 Eligibility criteria

The eligibility criteria are:

- **Early-Stage Researchers**: According to the MSCA definition, candidates should have less than four years' research experience, measured as full-time equivalent research experience, and not yet hold a PhD degree. Full-time equivalent research experience is measured from the date when a researcher obtained the degree entitling her/him/them to embark on a PhD, either in the country in which the degree was obtained or in the country in which the researcher is recruited or seconded, even if a PhD was never started or envisaged.

- **MSCA mobility rule**: Candidates must not have resided or carried out their main activity in Spain for more than twelve months in the three years immediately prior to the deadline of each call. Short stays such as holidays will be not taken into account.
- **Be entitled to enrol in a PhD at CITIC/Universidade da Coruña**: In accordance with UDC internal regulations, candidates must hold a Bachelor's degree and a Master's degree, equivalent to 300 ECTS in total, 60 of which should correspond to a Master's degree (or equivalent credit system for degrees awarded by universities outside the EU).[1] Experts from the UDC International Doctoral School (EIDUDC) will perform a pre-admission eligibility check on all applications.
- **Be of any nationality**. **No age restrictions apply.**

Academic and research break periods due to maternity or parental leave, compulsory national military service, sick or family care leave, or procedures for obtaining refugee status will not be taken into account and will not be considered to measure research experience.

5 Application process

Applications will be made online through the "Apply" section on the 3-i ICT website. The 3-i ICT website will also contain all of the information and documents necessary to complete the application process. The Programme Office (PO) will assist applicants throughout the application period with any queries they may have.

All queries and communications should be addressed directly to the Programme Office (3i.ict@citic-research.org) and not to the individual PhD project supervisors. Supervisors have been instructed to redirect all queries to the PO and to avoid all contact with applicants during the selection phase. Failure to do so will result in the exclusion of both parties and the corresponding PhD project.

Applicants will be required to complete an application form in English with personal and academic information, and to upload the following support documentation:

- Copy of national ID card or passport.
- Bachelor's degree certificate and academic record.
- Master's degree certificate and academic record.
- CV.
- Cover letter.
- Support documents: employment records, personal documents such as birth certificates of children, disability certificates, etc., and documents to show compliance with the MSCA mobility rule.

All forms must be completed in English and uploaded as PDFs. Candidates may apply for up to four different positions. Submissions will include a declaration of honour.

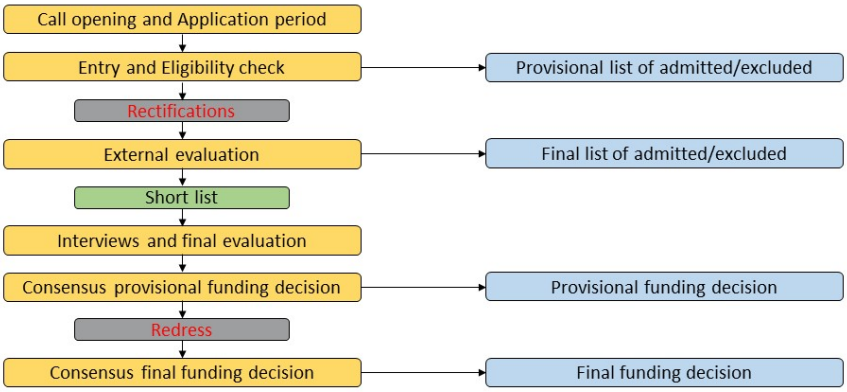
Applicants will receive a confirmation email upon submission. Selected candidates may be asked to provide documentary proof of any of the merits claimed and/or compliance with the entry and eligibility criteria.

Following the deadline, a provisional list of admitted and excluded candidates will be published and candidates will have ten working days to submit any additional information or corrections required.

[1] https://www.udc.es/en/eid/admision/

6 Evaluation and selection process

The selection process is designed to ensure the integrity, fairness and transparency of the programme. The process will comprise six main stages and an additional redress stage to deal with any complaints or appeals following communication of the provisional decision of the Selection Committee. The process workflow is presented below.



External evaluation

The external evaluation will be conducted by five external panels, based on the five main areas of research at CITIC. Each external panel will consist of three external experts (independent evaluators with no connection or conflict of interest with CITIC or the PhD projects offered).

The evaluation criteria for the external evaluation are as follows:

Dimension	Criteria	Scoring
Academic background	Level and duration of academic training (number of ECTS, number of BSc and MSc degrees). Multidisciplinary nature of degrees and subjects. Academic excellence (record, prizes, participation in international programmes, e.g. Erasmus).	50%
Research experience	Publications. Full and short papers at international conferences. Prizes for research results.	15%
Professional experience	Experience in research environments (academic and non-academic). Multidisciplinary nature of professional experience.	15%
Training	Professional training and qualifications. Other non-academic training.	10%
Personal ambition	Quality of education and PhD project. Personal motivation.	10%

Scores for each candidate and for each external panel will be published in rank order on the 3-i ICT website. Candidates will be notified of the publication of scores by email and receive a short personal evaluation report. Following completion of all evaluations, the External Panels will propose a shortlist of the best (i.e., highest scoring) applicants for each PhD project. At least the three highest scoring candidates for each position will be invited to participate in the next stage of the selection process, subject to a minimum overall threshold of 75 %.

Since candidates may apply for up to four different positions, different scores may be awarded for each of their applications. Candidates applying for different positions may be shortlisted for multiple interviews.

Interviews and final evaluation

Shortlisted candidates will be invited to participate in the interview stage of the selection process. Interviews will be conducted by a mixed committee, comprising three external experts and the two supervisors of each PhD project, and will take place via videoconference.

Candidates will be informed of the structure of the interview and content at least one week in advance. All interviews will have the same structure:

- Candidate presentation.
- Theoretical-practical questions.
- General questions.

Interviews will be assessed based on the following criteria:

Dimension	Subcriteria	Scoring
Domain background	Knowledge of the disciplines involved in the chosen PhD project.	25%
Communication	Clear presentation. Ability to communicate during interview. Ability to talk about different topics.	25%
Creative problem-solving	Ability to solve theoretical-practical questions set by Mixed Committees. Creativity of proposed solution.	20%
3-i potential	Potential and willingness to work in a 3-i context.	20%
Career prospects	Suitability of profile to the research project.	10%

Candidates must obtain at least 60 % in the interview to be eligible for a 3-i ICT position. The final evaluation score will be the direct sum of the scores from the external evaluation and the interview.

Consensus funding decision

The **selection committee** for the 3-i ICT first call for candidates will consist of the chairperson of each external panel and a member of the programme governing committee.

Following completion of all interviews, the selection committee will propose the four candidates to be selected based on the scores obtained in the previous stages of the selection process. A provisional list of proposed beneficiaries and a reserve list of candidates will be published on 3-i ICT website. Candidates will be notified of the decision by email and receive an Evaluation Summary Report of their application.

Each of the selected candidates will be offered the position in which they/she/he have achieved the highest score. If two or more of the highest scores correspond to the same candidate, they/she/he will

be allowed to choose the position they wish to be hired for. This process will be repeated until all four positions have been filled.

Where two or more candidates obtain the same score, the following tiebreak criteria will apply (in order of precedence):

- Highest score in interview.
- Highest score in academic background.
- Gender, giving preference to women as they are under-represented in the ICT sector.

All selected candidates will be asked to provide evidence of their compliance with the eligibility criteria.

Redress

Candidates will be given ten days to appeal the funding decision. Appeals should be sent by email to the Redress Committee, who will deal with all appeals within one month as from publication of the provisional funding decision. The final funding decision will be published on the 3-i ICT website and, once again, all candidates will be notified of the decision by email.

Following publication of the final funding decision, successful candidates will be guided through their arrival, enrolment and induction by the PO: signing of contracts, registration on PhD programme, initiation of residence permit procedures, support with VISA, travel and accommodation arrangements, etc.

7 Working conditions

The selected ESRs will be employed by UDC under a 36-month contract of employment, and will enjoy the same treatment, benefits, opportunities and standards of safety and occupational health as other CITIC researchers of an equivalent standing. Their status as MSCA fellows will be acknowledged in their contracts.

3-i ICT fellows will receive full social security coverage in compliance with Spanish legislation, including unemployment benefits, full social healthcare coverage, maternity, paternity, fostering and adoption leave, and paid holidays. In accordance with CITIC internal regulations, employment and working conditions will ensure researchers a good work-life balance. Researchers will also have access to a range of general UDC facilities, training and services.

The gross annual salary of each ESR will be €24,516.12, corresponding to a total annual remuneration cost of €32,508.00 (€2,709.00/month). ESR salaries will include personal income tax arrangements, social security coverage and social benefits (parental leave, pension fund contribution, severance payment, unemployment benefits, and health and accident insurance). For fellows with disabilities, UDC will apply for a MSCA special needs allowance to cover the cost of any additional assistance, adaptations or acquisitions they may require.

A variable research budget will be allocated to each fellow to cover any cost that may arise from their/her/his research (conference fees, travel and accommodation expenses for conferences and short research visits, open access fees, secondment mobility allowances).

ESR employment contracts will be incompatible with any other grant or contract, payment that implies a contractual relationship of a similar nature, or activities that may prevent the fellow from devoting them/her/himself exclusively to the work that is the object of the contract. By signing the contract, fellows undertake to carry out the PhD project described in their application and conduct their research in person.

8 About UDC

UDC is a young public university with an extensive catalogue of courses, including 39 undergraduate, 57 Master's and 36 PhD degree programmes. UDC's activity is coordinated around its three specialised campuses: Industry, Sustainable Development, and Innovation.

UDC and CITIC offer tailored specific courses and training to PhD students, based on their personal needs and their PhD project. The UDC International Doctoral School (EIDUDC) currently offers a selection of training and development activities specifically aimed at PhD students, in collaboration with the UDC Centre for Educational Training and Innovation (CUFIE), the Office of Research and Knowledge Transfer and the UDC Library. The EIDUDC/CUFIE programme also includes a Support Plan for PhD Researchers, which offers cross-disciplinary training courses throughout the academic year, covering general skills such as research methodology, gender analysis in research, research ethics and public speaking.

A Coruña and Galicia

CITIC and the UDC are located at the heart of A Coruña, a city that offers great quality of life and professional growth opportunities, thanks to its modern network of services and a wide range of activities to suit all tastes. Galicia, the region where A Coruña is located, boasts a rich landscape of forests, beaches, farmland and mountains, as well as centuries of historical and cultural heritage, extraordinary food, and a full programme of festivals and events throughout the year. There are also numerous public and private sports facilities available, including golf, tennis, sailing and rugby. This short video offers a taste of just some of what Galicia has to offer: https://www.youtube.com/watch?v=ZtUqecPNXgI

Annex I – List of selected PhD projects

CODE	2022-C1-001
Title	Addressing challenging optimization problems in cell signalling networks with High Performance Computing and Cloud-based approaches
CITIC supervisor	Dra. Patricia González Gómez https://pdi.udc.es/es/File/Pdi/Z399E
Research lines	High Performance Computing / Biochemistry and Molecular Biology
Secondary supervisor	Dr. Julio Sáez Rodríguez Institute for Computational Biomedicine, Heidelberg University
Academic PhD programme	Information technology research https://estudos.udc.es/en/study/start/5023V01
Summary	Computational models have become very popular to analyze the functioning of complex biochemical networks such as those involved in cell signaling networks. Successful models build predictive logic models of signaling pathways by training a prior knowledge network to biochemical data obtained from perturbation experiments. This training shows up as an optimization problem that require efficient and robust solution methods. The use of High-Performance Computing (HPC) techniques may represent an effective strategy to speed up the time-to-solution. However, most of these methods, handled as algorithms, may have limited parallelism, while if they are tackled as problem solving methods, they offer other opportunities for large-scale parallel computing. In this project we will explore the use of HPC and Cloud-based techniques in the context of multimethod global optimization, in which multiple different search algorithms are performed concurrently and cooperate between them through information exchange. These algorithms will also be adapted to the specific problem structure of training models of signaling networks. To address this challenge, it is required to combine the complementary knowledge in the research areas of both supervisors of the Ph.D. project. This research will be addressed in a context of an already established international collaboration, having also in mind the potential technology transfer of the research results. More precisely, the methods developed here can be used in the pharmaceutical industry to construct signaling networks of diseases and use these models to find new targets for therapy as well as to better characterize the effect of existing drugs.
Foreseen secondments	Two research visits of at least three months each to the Institute for Computational Biomedicine. In addition, possible visits to IIM-CSIC (Marine Research Institute of the Spanish Council for Scientific Research), startups such as ProtAvio (https://www.multiplex-assays.com/) and Insilico Biotechnology (https://www.insilico-biotechnology.com/) could be carried out.

CODE	2022-C1-002
Title	Modelling complex biological phenomena via inverse optimal control and inverse reinforcement learning
CITIC supervisor	Dr. Carlos Vázquez Cendón https://pdi.udc.es/es/File/Pdi/DB58E
Research lines	Applied Mathematics / Chemical Engineering / Biochemistry and Molecular Biology
Secondary supervisor	Dr. Julio Rodríguez Banga Marine Research Institute – Spanish National Research Council (IIM-CSIC)
Academic PhD programme	Mathematical modelling and numerical simulation in engineering and applied sciences https://estudos.udc.es/en/study/start/5026V01
Summary	In recent years, the areas of molecular biology and biochemistry are witnessing a "data-deluge" due to major technological advances in genomics, proteomics and metabolomics. However, systematic analysis of these huge new experimental data sets from a mechanistic point of view remains an open question. In this thesis, we will develop novel methods to identify optimality principles from data in order to reverse engineer complex biological systems. In particular, these methods will be used to generate mechanistic understanding of the dynamics of biochemical pathways at the cellular level. The main idea is to bridge concepts and methods from numerical and mathematical optimization (inverse optimal control) and artificial intelligence (inverse reinforcement learning) to facilitate the dynamic modelling of these biological systems. These methods will be used to automatically infer the optimality principles that can explain the observed dynamic behaviour. The research will require tight interdisciplinary collaboration (involving the areas of applied mathematics, optimization, machine learning and cellular biology) to develop and apply novel methods and tools to fundamental problems in computational systems biology. These developments will be tested with case studies involving the metabolic response of microorganisms and human cells to environmental changes. As a result, a better mechanistic understanding of these bio-systems will be achieved. This new knowledge can have a major societal impact, improving intervention strategies in biomedicine (e.g. human metabolism and cancer) and industrial biotechnology (e.g. microbial fermentation bioprocesses).
Foreseen secondments	For this PhD project, the supervisors have carefully designed a list of high-quality international and intersectoral secondments tailored to complement different aspects of the research plan. In particular, research stays have been jointly agreed at the following research groups: - Mathematical Algorithmic Optimization group, leaded by Prof. Sebastian Sager (Otto-von-Guericke University Magdeburg, Germany), world-class researcher in optimal control theory and methods (more information at https://mathopt.de/). - Microcosme group at INRIA Grenoble (France), where Prof. Hidde de Jong is a leader in computational systems biology of microorganisms and its applications in industrial biotechnology (more information at https://team.inria.fr/microcosme/). - Systems and Data Analysis group at Fraunhofer-Chalmers Research Centre for Industrial Mathematics (Gothenburg, Sweden), leaded by Prof. Mats Jirstrand (Goteborg), a top-class group developing computational tools and techniques (including machine learning) for systems and data

CODE	2022-C1-003
Title	Distributed and parallel algorithms for inference of cell lineage trees
CITIC supervisor	Dr. Diego Darriba López https://pdi.udc.es/es/File/Pdi/4Y4NH
Research lines	High Performance Computing / Genetics
Secondary supervisor	Dr. David Posada González Universidade de Vigo
Academic PhD programme	Information technology research https://estudos.udc.es/en/study/start/5023V01
Summary	Recent advances in DNA/RNA sequencing technologies have allowed the generation of genomic data at single-cell resolution, promoting an unprecedented opportunity to better understand how somatic evolution works within our bodies and the evolutionary mechanisms behind diseases like cancer. Reconstructing cell lineage trees is essential to achieve this goal, but current algorithmic approaches have shortcomings: they cannot handle large data sets and/or they use simplistic models of somatic evolution. The aim of this project is to overcome the limitations of methods for inferring cell genealogies by developing new algorithms, tools and more biologically-realistic models. The massive amount of data currently generated thanks to Next-Generation Sequencing technologies brings up the importance of using High-Performance Computing techniques and environments. The results of this project will be publicly available as Open Access publications and Open Source software and will help researchers worldwide by increasing the power of hypothesis testing in the field of single-cell phylogenetics, thus increasing the knowledge about somatic evolution and its effects.
Foreseen secondments	**Torusware** is a technology company, spin-off of the Computer Architecture Group of the University of A Coruña, and it is specialized in Big Data and DevOps services. Collaborating with Torusware is interesting because of two aspects: on the one hand it may help provide additional parallel approaches, capable of handling larger data sets. On the other hand, their DevOps experience may help during the development and release of software to increase its scope (e.g., making it available to more platforms). The **Computational Molecular Evolution** (CME) research group belongs to the **Heidelberg Institute of Theoretical Studies** (H-ITS), a private, non-profit research institute in Germany. CME focuses on large-scale evolutionary biology data analysis and High-Performance Computing. Their experience in these fields can be extremely helpful for the development of this project.

CODE	2022-C1-004
Title	Linking Linguistics to Low-resource NLP Neural Models
CITIC supervisor	Dr. David Vilares Calvo https://pdi.udc.es/es/File/Pdi/UX4BF
Research lines	Computational Science and Artificial Intelligence / General Linguistics
Secondary supervisor	Dr. Marcos García González Universidade de Santiago de Compostela
Academic PhD programme	Computational science https://estudos.udc.es/en/study/start/5009V01
Summary	Neural networks and contextualized word vectors are leading the advances for most of natural language processing (NLP) tasks, such as machine translation, question answering and information extraction. Still, these methods are often data-hungry, requiring vast amounts of data to be trained. Therefore, its effectiveness is limited in practice to a few rich-resource languages, with a prevalence of English. These issues have been exacerbated even more in the era of large language models, which are extensively demanded both in academia and industry. This leaves out of the language technology democratization process thousands of languages and billions of speakers. In this context, low-resource modelling is referred as one of the main open problems in NLP. Yet, there is little understanding on how to exploit linguistic factors as an alternative to boost the performance, under the assumption of data-scarce scenarios, and how to use them to create better technologies. The cornerstone of this thesis is to explore and improve neural models for general-purpose NLP technologies, with linguistics involved in the loop. More particularly, we will explore three research lines: (i) explaining linguistics factors that might be playing a role in the deficient performance of existing neural models for certain languages, (ii) designing linguistically-motivated methods to improve the performance of large language models for low-resource models, and (iii) explore data-augmentation approaches for supervised downstream NLP tasks that lack of annotated data. Overall, this project will contribute to develop better multilingual technologies and help democratize the access to these tools by minority populations.
Foreseen secondments	The selected applicant will be encouraged to do from 1 to 3 international stays (of approximately three months each) and we will help them to find the most suitable host. Some possible collaborators are: – University of Copenhagen (2nd best ranked NLP group in Europe) and their group, Core Natural Language Processing led by Anders Søgaard. – Yulan He, Turing AI Fellow from Warwick University, working on unsupervised machine learning. – Aline Villavicencio, Chair of NLP of the University of Sheffield, and Program Chair of ACL 2022.

CODE	2022-C1-005
Title	Sequence Labelling Parsing for Applied Natural Language Processing
CITIC supervisor	Dr. Carlos Gómez Rodríguez https://pdi.udc.es/es/File/Pdi/ZU9BH
Research lines	Computational Science and Artificial Intelligence / General Linguistics
Secondary supervisor	Dra. Margarita Alonso Ramos Universidade da Coruña – https://pdi.udc.es/es/File/Pdi/HF9AF
Academic PhD programme	Computational science https://estudos.udc.es/en/study/start/5009V01
Summary	Syntactic parsing, the process of automatically obtaining the internal structure of a sentence, is a key task for natural language processing (NLP) applications that benefit from extracting meaning from texts. Unfortunately, its potential has so far been underutilized, to a large extent because until recently, parsing algorithms were slow, ad hoc and difficult to integrate into downstream NLP tasks. However, the recent breakthrough of reducing syntactic parsing to a sequence labelling task is beginning to produce very efficient parsing models that provide their output as a sequence of discrete tags, opening possibilities for integration with other models. The objective of this thesis is to exploit this potential by exploring how sequence labelling parsing can be used to improve practical NLP tasks in large-scale setups. In this line, we will explore different means of integrating information from both constituency parsing and dependency parsing into downstream tasks like named entity recognition, aspect-based sentiment analysis or text summarization. These will include using parsing information as tag features and designing multitask learning architectures to perform parsing and the downstream tasks jointly, and will involve adapting the encodings and representations used by all involved tasks in such a way to make the integration of linguistic information as effective as possible. In addition, we will extend the results to more advanced parsing formalisms, such as semantic parsing, that can provide extra useful information. The end goal is to improve the accuracy and quality of downstream NLP tasks using linguistic information, without a significant cost to efficiency.
Foreseen secondments	We have we have regular contact with several companies, mainly Galician or Spanish (Classora Technologies, Imaxin Software, NLPgo, Zendal, MeaningCloud, or the DataLife Digital Innovation Hub) but also international (Proxem), within the framework of a national government-funded project (SCANNER) on efficient named entity recognition; as well as a collaboration with a software company with which we have submitted a project proposal related with an industrial application of NLP (more details cannot be given at this point due to confidentiality agreements). A secondment in one of these companies would be helpful, both to improve the training of the ESR by familiarising them with industrial NLP, and to improve the project itself by keeping its goals and execution in line with the requirements that industry demands. In addition, we expect the ESR to make an international secondment at the CoAStaL group at the University of Copenhagen, one of the leading NLP groups in Europe, led by Prof. Anders Søgaard. The secondment would be centered on the application of multitask learning techniques for integrating syntactic information into downstream NLP tasks.

CODE	2022-C1-006
Title	New microbiomics algorithms and data analytics in colorectal cancer
CITIC supervisor	Dra. Susana Ladra González https://pdi.udc.es/es/File/Pdi/WT6MH
Research lines	Computer Languages and Systems / Microbiology
Secondary supervisor	Dra. Margarita Poza Domínguez Galician Health Service
Academic PhD programme	Computational science https://estudos.udc.es/en/study/start/5009V01
Summary	Colorectal cancer is the third most common cancer diagnosed worldwide and the first in Spain. Risk factors, such as diet or lifestyle, have a great influence on the microbiota, which is the set of microbial species that inhabit our organism. It has recently been determined that the role of microbiome in the development and evolution of pathologies such as cancer is greater than previously established. Also, it has been described that microbiome is capable of modulating the efficacy and side effects of a drug, such as chemotherapy or immunotherapy. Therefore, the motivation of this work is to identify the beneficial and harmful bacterial populations (biomarkers) present in the gastrointestinal cavity that affect the development and treatment of colorectal cancer, with the aim of promoting prevention as well as personalised therapies. In this thesis, new bioinformatics tools and algorithms will be developed for the management and analysis of heterogeneous samples and clinical data to study the role of microbiome in the development and treatment of colorectal cancer, covering all the steps of the pipeline. Different samples (faecal, oral and tissue samples, including paraffin samples) from patients and healthy relatives will be analysed using metagenomic, whole-genome or target-gene sequencing. For paraffin embedding tissue samples, new tools will be developed to address the highly degraded nucleic acids. The obtained results will be further analysed to compare the presence (or non-presence) of different microorganisms throughout the evolution of the disease (including response
Foreseen secondments	We planned a 3-month stay at the Department of Computer Science of the University of Helsinki in the second year of the PhD thesis. This will allow the PhD candidate to acquire advanced skills on bioinformatics algorithms. In particular, the research stay will be supervised by Leena Salmela, who is a collaborator of Susana Ladra. We planned one 3-month secondment to IGA Technology Services Srl (IGATEch) during the third year of the PhD thesis for the PhD candidate, such that he/she can acquire both the skills and the knowledge from this different non-academic environment. IGATech is an Italian SME and is the largest lab in Italy offering genomic research services on a wide range of organisms, from humans and other animals, to plants and microorganisms. Other international secondments can be planned to institutions that have been partners in the H2020 MSCA RISE project led by Susana Ladra in Bioinformatics and Information Retrieval Data Structures Analysis and Design (BIRDS), including the University of Chile or University of Melbourne.

CODE	2022-C1-007
Title	Automatic animal behavior analysis from video data
CITIC supervisor	Dr. Álvaro Rodríguez Tajes https://pdi.udc.es/es/File/Pdi/TX3BH
Research lines	Computational Science and Artificial Intelligence / Ecology
Secondary supervisor	Dr. Cristiano Venícius de Matos Araújo ICMAN – Institute of Marine Sciences of Andalusia – Spanish National Research Council
Academic PhD programme	Information and communications technology https://estudos.udc.es/en/study/start/5032V01
Summary	Study of animal behavior in laboratory experiments is a key factor in ethology, ecotoxicology, drug discovery, drug testing, neuroscience and other fields. These experiments are of major importance to assess animal health and wellbeing in nature and in human facilities such as fish farms and can tell us how climate change and human pollution affect animal species. The main objective of this PhD will be to develop a solution based on Computer Vision and Machine Learning, to model and extract complex behavioral traits from video recordings of animals in laboratory experiments. This PhD will expand on the previous findings of a collaboration between Umeå University (Sweden) and The University of A Coruña (Spain), resulting in the most versatile and used animal tracking solution in the world. From this framework, we will address the current limitations on the field by using Computer Vision algorithms to improve detection and tracking, Deep Learning models to identify multiple animals in social experiments, and Machine Learning models to detect high-level behaviors from tracking data. By achieving this outcome, we will be pioneering the field of behavioral computation analysis. To achieve this goal, we will collaborate with the Animal Learning and Behavior Laboratory (Distance University UNED, Spain), and the Institute of Marine Sciences of Andalusia (National Research Council CSIC, Spain), the Biophysics and Biophotonics group of the Dept. of Physics (Umeå University, Sweden), and the Dept. of Wildlife, Fish and Environmental Studies (SLU Umeå, Sweden).
Foreseen secondments	The student will perform a 3 month research stay in the Biophysics and Biophotonics group, from the Department of Physics; and The Dept. of Wildlife, Fish and Environmental Studies from Umeå, Sweden. This partner was a member of the original Toxtrac project. They will provide image datasets from experiments with fish and insects and assist with the definition of behavioral traits.

CODE	2022-C1-008
Title	Advances on Age-related Macular Degeneration treatment response prediction by means of ocular preclinical image analysis
CITIC supervisor	Dr. Marcos Ortega Hortas https://pdi.udc.es/es/File/Pdi/ML53G
Research lines	Computational Science and Artificial Intelligence / Pharmacology
Secondary supervisor	Dr. Anxo Fernández Ferreiro Galician Health Service
Academic PhD programme	Computational science https://estudos.udc.es/en/study/start/5009V01
Summary	Age-related Macular Degeneration (AMD) is a disease with a high prevalence in modern society and expected to increase due to population aging. Being the main cause of blindness among adult citizens in the EU, it greatly diminishes patient capabilities yielding comorbility. Also, it seriously damages life quality of the patient having a deeper impact, as stated by the own patients perception, than other serious pathologies such as acute myocardial infarction, cerebrovascular diseases or AIDS implying a great self-esteem loss and isolation. There is a variant of the disease called wet AMD generally caused by abnormal blood vessels that leak fluid or blood into the macula. Anti-VEGF treatments are a group of medicines which reduce new blood vessel growth (neovascularisation) or oedema (swelling) and are indicated for AMD treatment. These medicines are antiangiogenics supplied by intravitreal injections which have a high cost. The response of the patient to this treatment can be total, partial or none depending on numerous variables that make currently impossible to predict treatment response. AMD in its different forms can be assessed using retinal imaging, being Optical Coherence Tomography (OCT) one of the most used nowadays as it allows to evaluate thickness of different retinal layers associated to AMD. In this proposal, the aim is to use OCT preclinical images for automatic extraction of anatomic variables of the retina in order to develop a machine learning model for treatment response stratification based on the extracted variables.
Foreseen secondments	Regular visits to the hospital are expected for meetings with the supervisor and his team. This interaction with the health sector will provide a unique and first-hand experience to the PhD candidate who will be able to learn to propose and successfully develop research activities in the sector. It is planned to do an international secondment at the INESC TEC in Porto under supervision of Prof. Jaime Cardoso alo Professor in University of Porto. As an institution operating at the interface of the academic and business worlds, bringing closer together academia, companies, public administration, and society, INESC TEC typically applies the knowledge and results generated as part of its research in technology transfer projects, seeking value creation and immediate social relevance.

CODE	2022-C1-009
Title	Wireless virtual sensing for control applications
CITIC supervisor	Dr. Luis Castedo Ribas https://pdi.udc.es/es/File/Pdi/J729E
Research lines	Signal theory and communications / Mechanical Engineering
Secondary supervisor	Dr. Francisco Javier Cuadrado Aranda Universidade da Coruña – https://pdi.udc.es/es/File/Pdi/DF59E
Academic PhD programme	Information technology and mobile network communication https://estudos.udc.es/en/study/start/5029V01
Summary	Virtual sensing is an advanced sensing paradigm that combines real sensors with models to produce estimates of magnitudes that cannot be sensed directly. Virtual sensing is an active area of research in mechanical engineering as it enables high performance control systems. The applicability of virtual sensing, however, is still severely restricted to the consideration of only a few sensors and simple but inaccurate models due to the real-time performance limitations of the on-board hardware elements where virtual sensors are embedded. In this PhD project we propose to overcome current VS bottleneck by introducing the innovative concept of wireless virtual sensing. The idea is to leverage the extraordinary recent advances in wireless vehicular connectivity and edge computing to develop a wireless network infrastructure suitable for virtual sensing. Major work in this thesis will focus on the design of this specific wireless network infrastructure and on the optimization of its transmission and computing resources to effectively carry out the following actions: (i) real-time collection and uploading of sensor data from individual vehicles, (ii) consideration of VS models with unprecedented complexity and accuracy, (iii) seamlessly integration of side information coming from the environment and other vehicles to enable cooperative VS strategies, (iv) real-time delivering of actuating data meeting ultra-reliable and low latency requirements. This project is strongly interdisciplinary as it shares concepts, methods and technologies mainly from the areas of mechanical engineering and wireless communications engineering. The results will be useful in many different sectors like automotive industry, ICT and smart industry.
Foreseen secondments	Intersectoral secondments: - **Navantia**. Both GTEC and LIM have participated in the UMI Navantia-UDC recently finished. Negotiations are now well advanced for a new phase of the UMI that will start in 2022 and will support the work in this PhD project. - **CAF Signalling** (CAFS) is a Spanish company working on railway signalling. The PhD student will be encouraged to spend short stays at foreign research institutions to complement their training and allow his/her doctoral thesis to receive the international mention distinction (e.g. Technical University of Munich (TUM). Prof. Wolfgang Utschick; Technical University of Vienna (TUW), Prof. Markus Rupp; Tongji University, Shanghai, Prof. José Rodríguez-Piñeiro; Universitá degli Studi di Padova. Prof. Alberto Trevisani).

CODE	2022-C1-010
Title	Flexible cure models in data science to predict sustained remission in rheumatoid arthritis
CITIC supervisor	Dr. Ricardo Cao Abad https://pdi.udc.es/es/File/Pdi/2X29E
Research lines	Statistics and Operations Research / Medicine
Secondary supervisor	Dr. Francisco Javier Blanco García Galician Health Service
Academic PhD programme	Statistics and operations research https://estudos.udc.es/en/study/start/5017V01
Summary	Reliable assessment of remission is important for the optimal management of rheumatoid arthritis (RA) patients. The search of biomarkers able to predict response to treatment in Rheumatoid Arthritis (RA) has been very fruitful. Many studies with different approaches (genetic, immunologic, proteomic, functional genomics) found biomarkers associated with response to treatment. In this sense, at least 15 single nucleotide polymorphisms (SNPs) and several proteins have been proposed as candidate genetic biomarkers associated with the therapeutic response in RA. However, these biomarkers have not been fully validated due to the lack of consistent studies. A critical type of evidence, such as the validation of this potential genetic biomarkers in randomized clinical trials (RCT), is missing. In this PhD project, flexible statistical data analysis techniques will be proposed, studied and efficiently implemented to predict remission of RA after treatment. Cure models, nonparametric and semiparametric methods for survival analysis, and functional data analysis techniques will be used to define biomarker scores for predicting point remission and sustained remission. The approach by nonparametric and semiparametric cure models will be also useful to predict the time to recurrence (latency) of those patients who did not experience sustained remission.
Foreseen secondments	The PhD student to be recruited will carry out a research stay in the Biostatistics Department of one of the following pharmaceutical companies: Pfizer, Janssen of Johnson and Johnson, Novartis, Galapagos and GSK. International stays will be planned with any of the following collaborators: - Prof. Ingrid Van Keilegom https://www.kuleuven.be/wieiswie/en/person/00062045 - Prof. Valentin Patilea https://ensai.fr/en/equipe/valentin-patilea/ - Prof. Yingwei (Paul) Peng https://www.queensu.ca/academia/pypeng/ - Prof. Martin Lotz https://www.scripps.edu/faculty/lotz/ - Prof. Rik Lories http://arthritisheal.eu/rik-lories/ - Prof. Francis Berenbaum https://www.researchgate.net/profile/Francis-Berenbaum - Prof. Floris Lafeber https://www.umcutrecht.nl/en/research/researchers/lafeber-floris-fpjg - Prof. Peter Nilsson https://www.kth.se/profile/nipe

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement Nº 101034261.