Project protocol

1. **Title of the proposal:** Parallel genetic programming for the generation of efficient approximate models.

2. **Identification of the subject and state of knowledge.**

Machine learning (ML) has become one of the most active research areas in the field of computational science, because it offers flexible and very powerful tools for modeling complex processes. ML applications are varied, covering areas as diverse as the study of climate change [1] and medicine [2].

In particular, our working group focuses on the development and analysis of the paradigm known as Genetic Programming (GP) [3], a learning approach based on neo-Darwinian evolutionary principles. Unlike most ML methods, GP generates symbolic models that are highly portable and potentially interpretable by human experts [4]. Our group has developed several variants of GP, for example to generate simple and compact models for complex problems [5], that transform the representation of a learning problem [6,7], or that are robust to the presence of outliers [8]. In addition, we have applied these techniques to solve problems related to wastewater treatment [9], brain signal classification [10], traffic applications [11], problems in civil engineering [12], energy efficiency [13], and automatic source code repair [14], to mention some prominent cases.

One of the benefits of GP is its ability to effectively model complex dynamics [20], and the relative simplicity of the models it can generate, compared to neural networks, for example. For these reasons, one of the advantages of applying GP is that it can generate designs that approximate complex processes or algorithms, sometimes requiring a fraction of the execution time [15].

However, one of the shortcomings of GP is its high computational cost of the learning process. While other techniques, such as neural networks and ensemble models, transform the learning problem into a parametric model fitting problem, GP performs the optimization process mainly in syntax or symbolic space. For this reason, our group is developing new variants and implementations of GP that exploit the massive parallelism offered by GPUs (Graphical Processing Units) and FPGAs (Field Programmable Gate Arrays). In particular, our group is currently developing versions of GP and Geometric Semantic GP (GSGP) [16], which run natively on GPUs and FPGAs. For example, we released the first worldwide version of GSGP on GPUs this year [16], while we have started with the development of basic modules to run these algorithms natively on FPGAs [17].

In the present project our previous work converges in two notable ways. First, from the algorithmic point of view, we are interested in developing new versions of GP and GSGP that exploit the highly parallel processing of architectures with GPUs and FPGAs. In particular, our intention is to further develop the parallel GSGP system originally presented in [16], incorporating important elements to improve its effectiveness and efficiency, such as local search [12,13], random sampling [8] and handling multi-

program representations [6,7]. Also, developing the first complete implementation of a GSGP algorithm that runs natively in hardware through FPGAs.

The second line of work that converges in this project has to do with the resolution of complex problems, which require efficient methods to maximize their field of application. In this sense, we propose to address two particular case studies. The first one has to do with the generation of models capable of approximating the numerical indicator known as Hypervolume (HV), which is widely used in the field of multi-objective optimization. HV allows us to transform a problem with two or more objectives into a single-objective problem, and to guarantee convergence to the Pareto optimal front. A shortcoming of HV, however, is that it is expensive, particularly for more than 3 objectives, being almost impossible to apply it to five or more in a practical way. In previous work we developed the first HV approximation model using GP, which offers similar convergence while being orders of magnitude more efficient [15]. Although the generated models generalize to different types of problems, they are not general for different numbers of targets. In this project we aim to generate models that are independent of the number of targets in the problem, using state-of-the-art GP algorithms, taking advantage of the efficient implementations we have developed and propose to develop in the present project. The second case study is focused on the area of automatic control, where sliding mode control is a widely used option in electromechanical systems. These controllers are robust to disturbances and manage to deal with uncertainty in measurements. The control actions for a sliding mode controller are discontinuous functions that are implemented by means of power elements that allow on-off switching. They have good convergence properties in finite time. However, these controllers are characterized by fast switching, known as the chattering effect, which has been tried to be solved by several methods that unfortunately produce an increase in the control complexity and the accuracy in the solution of the problem is reduced [18]. In this project we will address the problem of generating control models that approximate a sliding mode control, but reducing the energy consumption of the control and eliminating unwanted oscillations. Our group has achieved preliminary results in this domain, using basic GP systems [19]. In the present project we intend to apply the parallel GSGP variants that we propose to develop, increasing the accuracy of the generated models and the efficiency of the learning process.

In summary, the present project proposes to develop new GP variants, implemented on highly efficient platforms. Additionally, to apply these algorithms for the resolution of two complex case studies, multiobjective optimization and automatic control of electromechanical systems.

REFERENCES

1. David Rolnick, et al., 2022. Tackling Climate Change with Machine Learning. ACM Comput. Surv. 55, 2, Article 42 (March 2023), 96 pages.
2. Arunim Garg, Vijay Mago, Role of machine learning in medical research: A survey, Computer Science Review, Volume 40, 2021, 100370
3. J. Koza, Genetic Programming, MIT Press, 1992.
4. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1, 206–215 (2019).

5. Perla Juárez-Smith, **Leonardo Trujillo**, Mario García-Valdez, Francisco Fernández de Vega, Francisco Chávez. Local Search in Speciation-based Bloat Control for Genetic Programming, Genetic Programming and Evolvable Machines, 20:3, 351–384, 2019.
6. Muñoz**, L., Trujillo**, L. & Silva, S. Transfer learning in constructive induction with Genetic Programming. Genetic **Programming and Evolvable Machines, 1–41, 2019.**
7. **Muñoz, L., Trujillo**, L., Silva, S., Castelli, M., Vanneschi, L. Evolving multidimensional transformations for symbolic regression with M3GP. Memetic Comp., vol. 11, no. 2, pp. 111-126, 2019.
8. **Leonardo Trujillo**, Uriel López, **Pierrick Legrand**. SOAP: Semantic Outliers Automatic Preprocessing, Information Sciences, Volume 526, 85-101, 2020.
9. Emigdio Z-Flores, Mohamed Abatal, Ali Bassam, **Leonardo Trujillo**, Perla Juárez-Smith, Youness El Hamzaoui. Modeling the Adsorption of Phenols and Nitrophenols by Activated Carbon using Genetic Programming, Journal of Cleaner Production, 161:, 860-870, 2017.
10. Z-Flores, E.; **Trujillo, L.; Legrand, P**.; Faïta-Aïnseba, F. EEG Feature Extraction Using Genetic Programming for the Classification of Mental States. Algorithms 2020, 13, 221.
11. J. R. López, L. C. González, J. Wahlström, M. Montes y Gómez, **L. Trujillo** and G. Ramírez-Alonso, A Genetic Programming Approach for Driving Score Calculation in the Context of Intelligent Transportation Systems, in IEEE Sensors Journal, vol. 18, no. 17, pp. 7183-7192, 1 Sept.1, 2018.
12. Mauro Castelli, **Leonardo Trujillo**, Ivo Gonçalves and Aleš Popovič. An evolutionary system for the prediction of high performance concrete strength based on semantic genetic programming, Computers and Concrete, 19(6):651-658 (2017).
13. Mauro Castelli, **Leonardo Trujillo**, Leonardo Vanneschi, Ales Popovic. Prediction of energy performance of residential buildings: a genetic programming approach, Energy and Buildings 102:67-74 (2015).
14. **Trujillo, L.,** Villanueva, O.M. and **Hernandez, D.E.** 2021. A Novel Approach for Search-Based Program Repair. IEEE Software. 38, 4 (Jul. 2021), 36–42.
15. **Cristian Sandoval**, Oliver Cuate, Luis C. González, **Leonardo Trujillo**, Oliver Schütze, Towards fast approximations for the hypervolume indicator for multi-objective optimization problems by Genetic Programming, Applied Soft Computing, 2022, 109103, ISSN 1568-4946.
16. **Leonardo Trujillo, Jose Manuel Muñoz Contreras, Daniel E.Hernandez**, Mauro Castelli, Juan J.Tapia. GSGP-CUDA — A CUDA framework for Geometric Semantic Genetic Programming in SoftwareX, Volume 18, June 2022, 101085.
17. C. Goribar-Jimenez**, Y. Maldonado, L. Trujillo**, M. Castelli, I. Gonçalves and L. Vanneschi, Towards the development of a complete GP system on an FPGA using geometric semantic operators, 2017 IEEE Congress on Evolutionary Computation (CEC), San Sebastian, 2017, pp. 1932-1939.
18. S. Kharuf-Gutierrez, A. F. de Loza, L. T. Aguilar and **L. N. Coria**, "Robust stabilization of an inverted cart-pendulum: A second order continuous sliding mode approach," 2021 XXIII Robotics Mexican Congress (ComRob), 2021, pp. 112-117, doi: 10.1109/ComRob53312.2021.9628593.
19. **Luis N. Coria**, Ramon Ramirez-Villalobos, **Christian Aldrete-Maldonado**, Efrain Sanchez-Martinez, Control del péndulo de Furuta aplicando Programación Genética inspirado en un controlador por modos deslizantes, Revista Aristas: Investigación Básica y Aplicada, 2022.
20. **Leonardo Trujillo, Pierrick Legrand**, Gustavo Olague and Jacques Levy-Vehel. Evolving estimators of the pointwise Hölder exponent with Genetic Programming. Information Sciences, 209:61-79, 2012.

3. Objectives

General Objective

Develop efficient semantic geometric genetic programming (GSGP) systems using graphics processing units (GPUs) and FPGAs for solving problems requiring approximate models of complex and inefficient processes, particularly in the areas of multi-objective optimization and automatic control.

Specific Objectives

1. To develop a state-of-the-art GSGP system that runs directly on modern GPU architectures for solving supervised learning problems.
2. Develop a GSGP system with state-of-the-art features that runs directly on modern FPGA architectures for solving supervised learning problems.
3. Generate models capable of efficiently approximating the hypervolume of a Pareto front in a multi-objective optimization problem using GP and GSGP, which are independent of the number of objectives in the problems.
4. Generate models capable of efficiently approximating a sliding mode control strategy using GP and GSGP, which eliminate the chattering problem and do not affect the robustness of the control.

5. Methodology

The project is organized around the four specific objectives of the project, which will be developed as follows throughout the 4 Stages (years) of the project.

4.1 Specific Objective 1

Specific Objective 1 focuses on extending the parallel GSGP system, programmed in CUDA for execution on GPUs, that we developed in [16]. We intend to extend the system by considering optimization or local search mechanisms, such as those we integrated in sequential GSGP algorithms in [12,13], as well as using random sampling processes of the feature space to make the search more efficient and more robust to the modeling process. Furthermore, since the system proposed in [16] uses a stack-based representation, it is possible to extract more than one valid model from each individual in the evolutionary GSGP process, which could be exploited in an intelligent way. The main activities will be:

- Develop local search modules in CUDA for GSGP, to properly target search operators.
- Develop methods for interpreting individuals in the system proposed in [16], to exploit all valid models that can be extracted from a single individual.
- Integrate mechanisms to sample the feature space, as well as the instances, of the dataset of a supervised learning problem.
- Incorporate the developed system into an open-source licensed software library that will be made available with an interface compatible with the Scikit-learn standard.
- Rigorously evaluate proposed algorithms using state-of-the-art benchmarks.

- Publishing at least one article in proceedings of a prestigious international congress, such as GECCO, EvoStar or PPSN, and at least one journal article with JCR index.

Objective 1 will be developed mainly during the first two phases of this project.

4.2 Specific Objective 2

Recently, our group has started to develop a GSGP system that runs natively on hardware making use of FPGA platforms, with preliminary results in [17]. However, our work in this line has been halted in recent years due to issues arising from the COVID-19 global pandemic, so we propose to resume this development in this objective. Given the complexity of this objective, it will be developed mainly along the 4 stages of the project, considering the following main activities:

- Independently develop all the main modules of the system using VHDL (Hardware Description Language and Very High Speed Integrated Circuit) code, such as the model interpretation, evaluation, mutation and survival module, to name a few.
- Rigorously evaluate the proposed system, using state-of-the-art problems and high-profile FPGA platforms for Intel and AMD.
- It is proposed to publish the proposed system in a JCR indexed journal.

4.3 Specific Objective 3 y Specific Objective 4

The last two objectives refer to the two application case studies for the developed algorithms. Specific Objective 3 focuses on the problem of generating models to approximate the calculation of the Hypervolume of a Pareto front that generalize to any number of targets; i.e., that the models can be applied to problems with different number of targets, something that the models generated in [15] are unable to do. To this end, the following activities are proposed:

- Propose a Pareto fronts representation that allows a feature extraction process that is independent of the number of targets.
- Implement a feature extraction process appropriate for the proposed representation.
- Generate a database to pose a supervised learning problem, using the feature extraction process and the actual computation of the Hypervolume, with problems of different number of targets.
- Apply GP and GSGP methods for model generation and evaluate their performance to approximate the real Hypervolume, considering the accuracy and complexity of the models, and evaluate them when using them to guide an indicator-based algorithm for the optimization of multi-objective problems.

Specific Objective 4 addresses the second applied case study, to generate models capable of approximating a sliding mode control system. For this purpose:

- Electromechanical systems must be defined to validate the proposed approach. For the moment, a Furuta pendulum has been identified, which can be analyzed from Quanser's ROTPEN experimental platform.
- Design a sliding mode controller for the different electromechanical systems to be studied.

- Generate training databases from the controllers to be designed, to establish supervised learning problems, considering different operating conditions, e.g. with and without disturbances.
- Apply GP and GSGP to generate the models that approximate the sliding mode controllers that were designed, and evaluate them in terms of efficiency and robustness.

Both Objective 3 and Objective 4 will be developed over the four years of the project, with the most significant results planned for the last phase.

6. Expected results and overall project benefits.

The expected results and benefits of this project are organized into the following categories: 1) Scientific Products; 2) Technological Products; 3) Human Resources Training; and 4) Institutional Consolidation.

Scientific Products:

- Submit and publish at least four articles in high impact journals (ICR Q1 or Q2), each one derived from each specific objective of the project.
- Submit and publish at least four articles in conference proceedings at internationally recognized events, with their respective papers. Potential conferences include ACM GECCO, LEEE CEC, PPSN and EvoStar.
- Contribute two book chapters in edited collections (or digests) related to the areas of learning and genetic programming.

Technological Products:

- The software tools derived from this project will be released as software tools with book code licenses.
- At least one registration with INDAUTOR of the software tools derived from the project.

Human Resources:

- The participation of four doctoral students from the PhD program in Engineering Sciences is proposed, who will develop part of their thesis in the development of specific objectives of the project. It is proposed that at least three of these students defend their thesis at the latest in the last stage of the project. In addition, it is proposed to integrate a Ph.D. student on the French side, under the co-direction of the Technical Co-Principals, on the Mexican side and on the French side.

Institutional Consolidation

On the Mexican side, the present project is developed in the Graduate Program in Engineering Sciences, in particular with students and researchers of the Doctorate in Engineering Sciences (DCI) of the Tecnológico Nacional de México/Instituto Tecnológico de Tijuana, with participation of the lines of Cybernetics and Dynamic Systems and Control. The DCI has extensive experience of collaborations and joint work with European institutions, such as the University of Dublin in Ireland, the University of Extremadura in Spain, the University of Lisbon in Portugal, and in particular with the University of

Bordeaux in France. The project will allow the graduate program to consolidate its research work in the aforementioned lines of research, and will allow the graduate program to continue to have an international profile and projection. In addition, with the participation of three members in this project, it is intended to consolidate the PRODEP Academic Body called "Control systems and signal processing".

7. Institutional and social impact of the results achieved in the project.

Since its formation, the DCI has had an important international collaboration and presence, achieving previous co-directions with researchers abroad (from institutions such as Stevens Institute of Technology in the USA, University of Lisbon in Portugal, University of Dublin in Ireland, and University of Bordeaux in France), as well as the development of projects with international collaboration, for example the project of the European Commission - 7th Framework Programme for Research, technological Development and Demonstration - Marie Curie Actions - IRSES, entitled Analysis and classification of mental states of vigilance with evolutionary computation, with identifier ACoBSEC, which was developed from 2013 to 2016. This project counted with the collaboration of national and international institutions, such as the Tecnológico Nacional de México/Instituto Tecnológico de Tijuana and CICESE in Mexico, the University of Lisbon in Portugal, University of Extremadura in Spain, and the University of Bordeaux in France. The present project will allow these collaborative relationships, especially with the University of Bordeaux, to become even closer, boosting the international projection of the Graduate Program in Engineering Sciences.

8. Specify the participation of each researcher and the product of their work.

On the Mexican side, four researchers from the Graduate Program in Engineering Sciences of the Tecnológico Nacional de México/Instituto Tecnológico de Tijuana participated.

- Dr. Leonardo Trujillo Reyes – Sistema Nacional de Investigadores, Nivel II
  - Rol: Principal Investigator, on the Mexican Side
  - Dr. Trujillo will be responsible for organizing and coordinating the scientific work of the project. He is an internationally recognized expert in the area of Genetic Programming, Artificial Intelligence and Machine Learning, having directed national and international research projects. He will also serve as Thesis Director of three of the doctoral students who will participate in this project. Dr. Trujillo will also work with the students in the writing of scientific articles derived from the project.
- Dr. Daniel Eduardo Hernández Morales – Sistema Nacional de Investigadores, Nivel I
  - Rol: Participant
  - Dr. Hernández is a researcher in the areas of Genetic Programming, Computer Vision and Machine Learning, and will serve as Thesis Co-Director of students participating in the project, focusing mainly on the activities derived from Specific Objectives 1.
- Dra. Yazmin Maldonado Robles – Sistema Nacional de Investigadores, Nivel I
  - Rol: Participant

- Dr. Maldonado is an international expert in the development of digital systems with FPGA platforms, embedded systems and intelligent systems, and will participate in the direction of theses of participating doctoral students, collaborating mainly in the development of Specific Objectives 2 and 4.

- Dr. Luis Néstor Coria de los Ríos – Sistema Nacional de Investigadores, Nivel I
  - Rol: Participant
  - Dr. Coria is an expert in the analysis and modeling of nonlinear dynamic systems, especially for applications in medicine and electromechanical systems, with the design of robust controllers. He will participate in the direction of thesis of participating students and publications in conferences and JCR journals. His contributions are concentrated in the activities derived from Specific Objectives 4.

On the French side, there is the participation of three researchers from the University of Bordeaux, with the participation of the following researchers:

- Dr. Pierrick Legrand
  - Rol: Principal Investigator, on the French Side
  - Dr. Pierrick is attached to the Institute of Mathematics of Bordeaux and the ASTRAL group of INRIA, developing cutting-edge research in the areas of signal processing, applied mathematics, fractal systems and evolutionary computation. He will participate in the development of the four specific objectives of the project, and as a host advisor for the students' stays.
- Professor Jérôme Saracco
  - Rol: Participant
  - Professor Saracco is an international expert in statistical modeling, regression analysis, multidimensional statistics, dimensionality reduction and statistical learning. He will be a key participant in the project, supporting in the statistical analysis of the experimental results, particularly in the activities derived from Objectives 3 and 4.
- Professor Eric Grivel
  - Rol: Participant
  - Professor Grivel: is an international expert in the fields of signal processing and image analysis, and will participate in development and application of feature extraction methods for the datasets generated during the activities related to Objectives 3 and 4.

In addition, the participation of a doctoral student from France, under the co-direction of Dr. Pierrick Legrand, Professor Jérôme Saracco, and Professor Leonardo Trujillo, is proposed to be integrated in the activities derived from Specific Objectives 3 and 4.

9. **Activities of the doctoral student. Describe the activities that the doctoral student will perform in the project and clearly specify the work plan of the project.**

On the Mexican side, the following doctoral students from the PhD program in Engineering Sciences will participate in this project:

- Jose Manuel Muñoz Contreras
  - CVU CONACYT: 771416
  - RESEARCH AREA: Semantic Geometric Genetic Programming with GPUs
  - ACTIVITIES: Juan Manuel will work on the development of Objectives 1 and 2 of this project.
- Cristian Sandoval Reyes
  - CVU CONACYT: 846250
  - RESEARCH AREA: Approximation of the Hypervolume Indicator with Genetic Programming
  - ACTIVITIES: Cristian will participate in the activities derived from Objectives 1 and 3 of the project.
- Joel Lee Nation Morales
  - CVU CONACYT: 951303
  - RESEARCH AREA: Transfer learning with Genetic Programming and feature space transformations of learning problems.
  - ACTIVITIES: Joel Lee will participate in the activities derived from Objectives 1, 3 and 4 of the project.
- Joel Antonio Quevedo Felix
  - CVU CONACYT: 1014860
  - RESEARCH AREA: Development of embedded systems on FPGAs
  - ACTIVITIES: Joel will participate in the activities derived from Objectives 2 and 4 of the project.

10. **Supporting background. If pertinent, attach references of publications co-authored, theses, equipment, joint projects, patents, etc.**

The participating institutions, and the Technical Leads, of the present project have a long history of scientific collaboration. The collaboration has been in the areas of evolutionary computation and genetic programming, computer vision and pattern recognition. The following is a summary of each of the areas of research collaboration that has been achieved, documenting the scientific work done in collaboration:

1. <u>International Awards</u>
   a. Two Best Paper awards at the ACM GECCO conference in 2010, in the Genetic Programming track and in the Real World Applications track, with the following papers:
      i. Leonardo Trujillo, Pierrick Legrand and Jacques Lévy-Véhel. The Estimation of Hölderian Regularity using Genetic Programming. In Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation (Portland, Oregon, USA, July 7 - 11, 2010). GECCO 2010. ACM, New York, NY. (Best Paper Award in the track of Genetic Programming)

ii. Leonardo Trujillo, Pierrick Legrand, Gustavo Olague and Cynthia Pérez. Optimization of the Hölder Image Descriptor using a Genetic Algorithm. In Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation (Portland, Oregon, USA, July 7 - 11, 2010). GECCO 2010. ACM, New York, NY.

2. 13 Publications in High Impact Journals (JCR)
   a. **Leonardo Trujillo**, Gustavo Olague, *Pierrick Legrand*, Evelyne Lutton. Regularity based descriptor computer from local image oscillations. Optics Express. 15(10): 6140-6145, 2007.
   b. **Yuliana Martínez**, **Leonardo Trujillo**, Edgar Galván-López and **Pierrick Legrand**, A comparison of predictive measures of problem difficulty for classification with Genetic Programming, Research in Computing Science, 59, 159-170.
   c. **Leonardo Trujillo, Pierrick Legrand**, Gustavo Olague and Jacques Levy-Vehel. Evolving estimators of the pointwise Hölder exponent with Genetic Programming. Information Sciences, 209:61-79, 2012.
   d. **Laurent Vezard**, **Pierrick Legrand**, Marie Chavent, Frédérique Faïta-Aïnseba, **Leonardo Trujillo**. EEG classification for the detection of mental states, Applied Soft Computing 32:113-131 (2015).
   e. **Yuliana Martínez, Leonardo Trujillo, Pierrick Legrand** and Edgar Galván-López. Prediction of Expected Performance for a Genetic Programming Classifier, Genet Program Evolvable Mach (2016) 17: 409-449 (2016).
   f. **Emigdio Z-Flores, Leonardo Trujillo**, Arturo Sotelo, **Pierrick Legrand** and **Luis Coria.** Regularity and Matching Pursuit Feature Extraction for the Detection of Epileptic Seizures, Journal of Neuroscience Methods, 266:107–125 (2016)
   g. **Enrique Naredo, Leonardo Trujillo, Pierrick Legrand**, Sara Silva, **Luis Muñoz**. Evolving Genetic Programming Classifiers with Novelty Search, Information Sciences, 369:347-367 (2016).
   h. **Víctor R. López-López, Leonardo Trujillo, Pierrick Legrand**, Victor H. Díaz-Ramírez and Gustavo Olague. A comparison of local feature extraction paradigms applied to visual SLAM, Computación y Sistemas, 20(4):565-588 (2016).
   i. **Yuliana Martínez, Enrique Naredo, Leonardo Trujillo, Pierrick Legrand** and **Uriel López**. A comparison of fitness-case sampling methods for genetic programming, Journal of Experimental & Theoretical Artificial Intelligence, 29(6):1203-1224, 2017.
   j. José Enrique Hernández, Víctor Díaz-Ramírez, **Leonardo Trujillo, Pierrick Legrand**. Design of estimators for restoration of images degraded by haze using genetic programming, Swarm and Evolutionary Computation, 44:49-63, 2019.
   k. **López-López, V.R., Trujillo, L. & Legrand, P.** Applying genetic improvement to a genetic programming library in C++. Soft Comput 23, 11593–11609 (2019)
   l. **Leonardo Trujillo, Uriel López, Pierrick Legrand**. SOAP: Semantic Outliers Automatic Preprocessing, Information Sciences, Volume 526, 85-101, 2020.

m. **Z-Flores, E.; Trujillo, L.; Legrand,** P.; Faïta-Aïnseba, F. EEG Feature Extraction Using Genetic Programming for the Classification of Mental States. Algorithms 2020, 13, 221.

This collaboration is documented in 13 publications in high-impact journals, most of them Q1 and Q2.

3. Conference proceedings papers (17 in total)
   a. **Leonardo Trujillo, Pierrick Legrand** and Jacques Lévy-Véhel. The Estimation of Hölderian Regularity using Genetic Programming. In Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation (Portland, Oregon, USA, July 7 - 11, 2010). GECCO 2010. ACM, New York, NY.
   b. **Leonardo Trujillo, Pierrick Legrand,** Gustavo Olague and Cynthia Pérez. Optimization of the Hölder Image Descriptor using a Genetic Algorithm. In Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation (Portland, Oregon, USA, July 7 - 11, 2010). GECCO 2010. ACM, New York, NY.
   c. **Leonardo Trujillo, Yuliana Martínez**, Edgar Galvan-Lopez, **Pierrick Legrand**. Predicting problem difficulty for genetic programming applied to data classification. In GECCO '11: Proceedings of the 13th annual conference on Genetic and evolutionary computation, pages 1355-1362, Dublin, Ireland, 2011. ACM.
   d. **Leonardo Trujillo** and Sara Silva and **Pierrick Legrand** and Leonardo Vanneschi. An empirical study of functional complexity as an indicator of overfitting in Genetic Programming. In Proceedings of the 14th European Conference on Genetic Programming, EuroGP 2011, volume 6621, pages 263-274, Turin, Italy, 2011. Springer
   e. **Leonardo Trujillo, Yuliana Martínez**, Edgar Galván López, and **Pierrick Legrand**. 2012. A comparative study of an evolvability indicator and a predictor of expected performance for genetic programming. In Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference companion (GECCO Companion '12), Terence Soule (Ed.). ACM, New York, NY, USA, 1489-1490
   f. **L. Vezard**, Marie Chavent, **P. Legrand**, F. Faita-Ainseba and **L. Trujillo**, Detecting mental states of alertness with genetic algorithm variable selection. IEEE Congress on Evolutionary Computation 2013, Cancun, Mexico, 20 - 23 June, 2013. IEEE Press, pp 1247-1254.
   g. **Yuliana Martinez** and **Leonardo Trujillo** and **Enrique Naredo** and **Pierrick Legrand**. A Comparison of Fitness-Case Sampling Methods for Symbolic Regression with Genetic Programming. In Alexandru-Adrian Tantar et al. editors, EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V, volume 288, pages 201-212, Peking, 2014. Springer.
   h. **Emigdio Z-Flores** and **Leonardo Trujillo** and Oliver Schuetze and **Pierrick Legrand**. Evaluating the Effects of Local Search in Genetic Programming. In Alexandru-Adrian Tantar et al. editors, EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V, volume 288, pages 213-228, Peking, 2014. Springer
   i. **Victor R. López-López, Leonardo Trujillo, Pierrick Legrand**, Victor H. Díaz-Ramírez. Evaluation of Local Feature Extraction Methods Generated through Genetic Programming on Visual SLAM,

accepted to Proceedings of the 2014 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC 2014), Ixtapa, México, November 5-7, 2014.

j. **Emigdio Z-Flores, Leonardo Trujillo**, Oliver Schütze, and **Pierrick Legrand**. 2015. A Local Search Approach to Genetic Programming for Binary Classification. In Proceedings of the 2015 on Genetic and Evolutionary Computation Conference (GECCO '15), Sara Silva (Ed.). ACM, New York, NY, USA, 1151-1158.

k. Mauro Castelli, **Leonardo Trujillo**, Leonardo Vanneschi, Sara Silva, **Emigdio Z-Flores**, and **Pierrick Legrand**. 2015. Geometric Semantic Genetic Programming with Local Search. In Proceedings of the 2015 on Genetic and Evolutionary Computation Conference (GECCO '15), Sara Silva (Ed.). ACM, New York, NY, USA, 999-1006

l. **Víctor R. López-López, Leonardo Trujillo, Pierrick Legrand**, and Gustavo Olague. 2016. Genetic Programming: From Design to Improved Implementation. In Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion (GECCO '16 Companion), Tobias Friedrich (Ed.). ACM, New York, NY, USA, 1147-1154.

m. José Enrique Hernández-Beltrán; Víctor H. Díaz-Ramírez; **Leonardo Trujillo** and **Pierrick Legrand**. Restoration of degraded images using genetic programming, Proc. SPIE 9970, Optics and Photonics for Information Processing X, 99700K (September 14, 2016)

n. **Uriel Lopez, Leonardo Trujillo, Yuliana Martinez, Pierrick Legrand, Enrique Naredo** and Sara Silva. RANSAC-GP: Dealing with Outliers in Symbolic Regression with Genetic Programming. In Mauro Castelli et al. editors, EuroGP 2017: Proceedings of the 20th European Conference on Genetic Programming, volume 10196, pages 114-130, Amsterdam, 2017, Springer Verlag.

o. **López U., Trujillo L., Legrand P**. (2018) Filtering Outliers in One Step with Genetic Programming. In: Auger A., Fonseca C., Lourenço N., Machado P., Paquete L., Whitley D. (eds) Parallel Problem Solving from Nature – PPSN XV. PPSN 2018. Lecture Notes in Computer Science, vol 11101. Springer, Cham

p. **Víctor R. López-López, Leonardo Trujillo, and Pierrick Legrand**. 2018. Novelty search for software improvement of a SLAM system. In Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '18), Hernan Aguirre (Ed.). ACM, New York, NY, USA, 1598-1605.

q. **Uriel López, Leonardo Trujillo**, Sara Silva, Leonardo Vanneschi, and **Pierrick Legrand**. 2020. Unlabeled multi-target regression with genetic programming. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference (GECCO '20). Association for Computing Machinery, New York, NY, USA, 976–984.

4. Student and Thesis Supervisions

In the list of products presented in item (2), the participation of 8 doctoral students is highlighted. The students from Mexico are Emigdio Z. Flores, Uriel López Islas, Víctor López López López, Yuliana Martínez Ramos, Enrique Naredo García, Luis Muñoz Delgado, all from the Graduate Programs in Engineering Sciences and the Doctorate in Engineering Sciences. On the French side, Laurent Vezard participated. In addition to the collaboration in the

execution of research and publication of articles, we also had a formal collaboration in the co-direction of the doctoral thesis of the student Emigdio Z. Flores of the Doctorate in Engineering Sciences of the Instituto Tecnológico de Tijuana, as well as co-advising Yuliana Martínez Ramos and Enrique Naredo García.

5. Research Projects

From 2013 to 2016, the international collaboration project entitled Analysis and classification of mental states of vigilance with evolutionary computation with key PIRSES-GA-2013-612689, led by Pierrick Legrand and Leonardo Trujillo, was executed, with the inter-institutional and international collaboration of Tecnológico Nacional de México/Instituto Tecnológico de Tijuana and CICESE in Mexico, the University of Lisbon in Portugal, the University of Extremadura in Spain, the University of Bordeaux in France and INRIA in France, which was funded for a total of 130,200 EUROS to finance mobility and scientific collaboration activities.