

# Machine Learning-Based System for the Availability and Reliability Assessment and Management of Critical Infrastructures (CASO)

---

**Luigi Serio**

*CERN*

*Engineering Department*

*1211 Geneva 23, Switzerland*

*E-mail: [luigi.serio@cern.ch](mailto:luigi.serio@cern.ch)*

A critical infrastructure is a complex interconnected system of systems providing basic and essential services to support the operation of particle accelerators but also industries and households for which they must guarantee high reliability of critical functions.

Model-based approaches are usually adopted to provide an early identification of failures and to reveal hidden dependencies among subsystems. System models are complex and require constant updating to be reactive to system changes and real operating conditions, wear and aging. The interconnections between the different systems and the functional dependencies between their components are in many cases modified at both physical and functional levels while their degraded performances impact the overall system availability and reliability.

A novel approach is proposed which combines model-based and Big Data analytics by machine learning techniques to extract descriptive and predictive models directly from data. The objective is to foresee and react in time to failures to reduce downtimes as well as to optimize maintenance and operation costs.

The Computer-Aided System for critical infrastructure Operation (CASO) is designed to significantly and efficiently enhance the quality, safety, reliability and availability of critical infrastructures.

We report on the design of CASO, its implementation and on the preliminary results inferred on historical and live stream data recorded from CERN's technical infrastructure. Proposal for the full deployment and expected long-term capabilities will also be discussed.

*Artificial Intelligence for Science, Industry and Society, AISIS2019  
October 21-25, 2019  
Universidad Nacional Autónoma de México, Mexico City, México*

## 1. Introduction

Critical infrastructures [1, 2] provide essential services for our modern society requiring high reliability and availability. The early identification of failures is essential to prevent any disruption of service.

A smart framework [3, 4] for the availability and reliability assessment and management of accelerators technical facilities has been proposed to support the analysis, to guide the operation and improve the overall efficiency of the process. The framework provides an integrated environment to collect data from heterogeneous sources (sensors, alarms, logbooks), implements mining and machine learning techniques to infer functional dependency models and fault logic models, and dynamically updates them to follow the evolution of CERN [5] Technical Infrastructure [6].

## 2. System framework concept

A Computer-Aided System for critical infrastructure Operation (CASO) is being developed as a primary tool for the extraction and analysis of critical infrastructure functional and physical dependencies from alarms and measured physical parameters. CASO will not only smartly guide the operation of critical infrastructures but also identify critical components requiring maintenance or consolidation based on algorithms modelling the nominal operation and extracting features to anticipate failure, wear and tear, and ageing. The developed techniques and framework will significantly and efficiently enhance the quality, safety and availability of critical infrastructures serving society, industrial manufacturing and production, as well as research. Nowadays, Big Data represents one of the most valuable resources for the management of large, medium and small enterprises. The expression Big Data usually refers to information with a volume, velocity and variety that cannot be processed using classical database management technologies and data analysis techniques. This huge volume of data is not the only characteristic that makes data Big. By definition, Big Data also includes the concept of unstructured data, which is nowadays increasing due to the wide diffusion and high availability of digital information.

The use of machine learning for the analysis and interpretation of Big Data is an observed trend applied in several heterogeneous domains. It can be used in healthcare by considering huge amounts of medical data, symptoms, and clinical reports for supporting medical decisions. Machine learning is widely applied for the failure analysis and remaining useful life estimation of automotive and aircraft engines. NASA, for example, defined several statistical approaches for the anomaly detection of cruise flight data, based on historical data of previous flights. In this case, thanks to the statistical model, they can identify anomalies in cruise data, simply by looking for outliers produced by the model.

The critical component to be analysed (the engine, the valve, the pump, the bearing, etc.) must be first identified by experts and operators. In fact, each analysis provides information on the specific component but not on the way a failure can affect and/or propagate to the rest of the infrastructure in which the component is present. Furthermore, the applied machine learning algorithms are mostly supervised and therefore driven by knowledge that guide and sometime limit the outcomes that can be obtained.

CASO innovatively proposes the ambitious goal of exploiting all the available data sources at the infrastructure level, and in particular alarms, reports, logs, and physical data and, without

any apriori knowledge, to infer components, sub-systems and systems availability, reliability and dependencies models and to identify autonomously critical components of the infrastructure in need for deeper analysis, maintenance and consolidation.

The identification of intersystem dependencies is a key task for the efficient operation of critical infrastructures. The evolution of such dependencies, due to changing operational requirements and consolidation activities, makes it difficult to accurately maintain the information up to date and ensure the coherence of the information. In general, the interconnections among the different systems and the functional dependencies between their components are in many cases modified with respect to the initial design, both at the physical and functional levels.

Given the importance of dependent failures propagation on the overall system risk, reliability, and availability analysis, CASO's ability to identify the intersystem dependencies among components of different systems and of sets of interconnected components will bring several benefits:

- An accurate estimation of the critical infrastructure reliability and dependability, which is sometimes not well-estimated if the dependencies are hidden;
- The ability to anticipate or, at least, to react in time to failures, with a significant reduction of downtimes in the provided services;
- The identification of critical components and locations of the infrastructure to improve maintenance planning and guide consolidation strategies.

### 3. System design methodology

A critical infrastructure is a complex interconnected system of systems providing basic and essential services such as electrical power distribution, cooling, heating and ventilation, access and safety systems, and IT.

The availability of economical, high-efficient and smart sensors brings Big Data but their volume and their heterogeneity make them difficult to analyse and to integrate in the modelling process.

The data collected and stored undoubtedly contains critical knowledge on dependencies. In this project a novel approach is proposed which combines model-based and Big Data analytics by machine learning techniques to extract descriptive and predictive models directly from data, to foresee and react in time to failures thus reducing the downtimes of components and systems.

The Computer-Aided System for critical infrastructures Operation (CASO), can infer and interpret data coming from different and heterogeneous sources and systems, to extract descriptive and predictive models and to suggest mitigation and consolidation strategies for the performance enhancement of complex systems and infrastructures.

CASO can discover and track hidden dependencies of components and subsystems, to prevent failures, to analyze and guide maintenance and operation activities, to reduce cost and downtimes, and to increase quality, safety, availability and reliability of services and processes.

According to the high-level schema of the approach shown in Figure 1, the critical infrastructure chosen for the design and the development of the framework is CERN technical infrastructure.

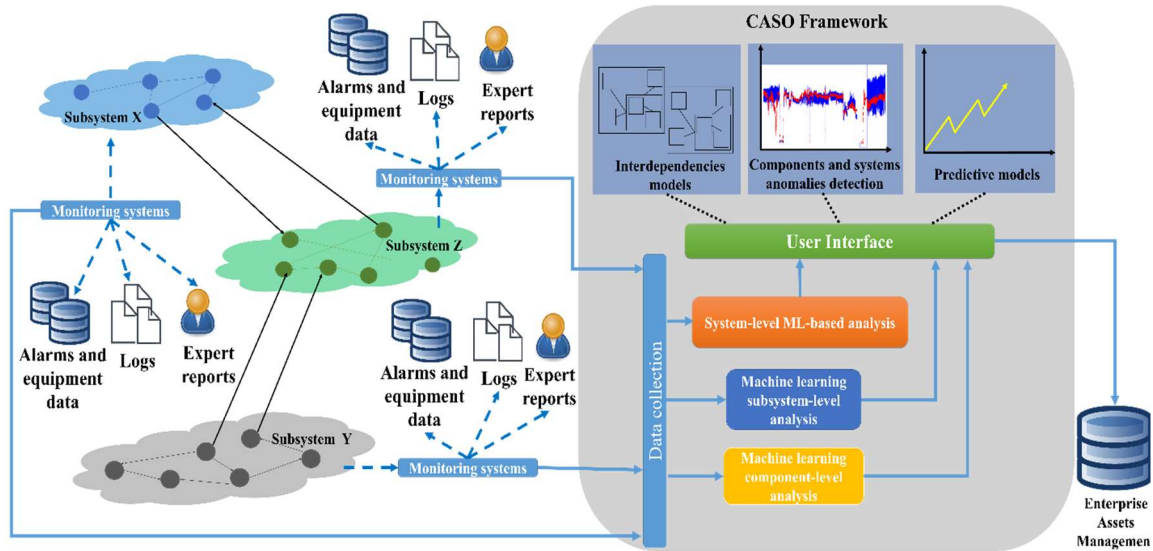


Figure 1 - Reference architecture of the Smart Framework

CERN technical infrastructure is a representative example of a critical infrastructure composed of several subsystems, each one responsible for a given critical functionality (e.g. distribute electrical power, cooling fluids, provide essential communication systems, etc.). Each subsystem is composed of thousands of components and is managed by a different monitoring system. A monitoring system collects all data related to the subsystem: physical signals of the equipment, reports, and logs related to failures, expert evaluations, maintenance interventions, installation dates of equipment and more.

The engine of CASO is a machine-learning core, composed of algorithms developed for the component-level and high-level system dependencies analysis.

Furthermore, CASO will be dynamically linked to an enterprise asset management system to select the most appropriate and efficient maintenance program and to guide and implement the required consolidation strategies to tackle aging problems of components, whilst maintaining the highest level of performance at a reasonable cost.

#### 4. Envisioned innovation potential and applications

Machine learning is nowadays an enabling technology for the scientific advancement of small and medium sized enterprises, as well as large companies and organizations dealing with huge amount of available operational data.

CASO, as by design, will become an industrial-driven tool to be integrated into the monitoring and control systems of critical infrastructures, system of systems and even combinations of infrastructures directly or indirectly linked. It will consist of a user-friendly interface, adaptable and self-tuning algorithms, and a dependencies detector, to perform components and system-level analysis.

Once operational, CASO will progressively provide continuous and updated information on dependencies, performances and reliability at components and systems level, to smartly guide the operation, maintenance and consolidation of critical infrastructures. All the above will positively and efficiently affect and improve the operation of research facilities, industry, public utilities, and therefore society as a whole.

A non-exhaustive list of possible application are:

- Smart cities exploit ICT (Information and Communication Technologies) to provide city essential services such as energy and fluid distribution, transportation and medical services. CASO would make it possible to consider the dependencies among the different public services and to increase the quality and efficiency of the services provided to the citizens. For example, by knowing the dependencies between the energy grid, transportation, and medical services, it would be possible to elaborate emergency plans to be activated timely in case of failure of one or more components or systems. Moreover, it would be possible to efficiently consolidate and maintain several interconnected energy grids ensuring, at lower capital and operation cost, the same level of services or even their improvement without additional investments;
- Public healthcare infrastructures are essential for the management of medical services as well as for the continuous monitoring of the diffusion of diseases within several communities. Data collected from hospitals, public and private clinics and other medical offices, allow several organizations to define epidemic disease diffusion scenarios and strategies to counteract. CASO would enhance the analysis of the diffusion scenarios and definition of strategies considering dependencies among healthcare, transport, utilities and other infrastructures or systems;
- Manufacturing industry and supply chains, which are essential drivers of the individual countries and world economies. Exploiting the scalability of CASO to extract dependency models among heterogeneous infrastructures and manufacturing processes, would allow identifying and tracking manufacturing flaws and inefficient processes. This will result in the anticipation and mitigation of potential impacts on energy consumption, production, services and processes as well as guaranteeing the safe and efficient operation of manufactured products (e.g. failure of components in power plants, aircraft engines, and transport sectors).

## 5. Proof of concept and case studies

The proof of concept of CASO's Smart Framework is being developed, implemented and tested on CERN technical infrastructure, which can be compared to a small city in size and a nuclear power plant in complexity.

CERN technical infrastructure is a large and complex critical infrastructure consisting in several systems of systems and components that provide a significant number of heterogeneous data. Existing maintenance and operation databases will ease the implementation of a comprehensive tool to store dependencies and models.

CERN technical infrastructure has been analysed and three heterogeneous subsystems have been identified to develop the proof of concept algorithms for the detection and analysis of hidden patterns and interdependencies. Representative components have also been chosen, to implement algorithms for component level analyses.

The data pipeline has been defined with proper connectors and software tools to extract and organize data collected by each monitoring system and required supporting hardware.

A scalable prototype of CASO based on the chosen subset of components and subsystems on a representative portion of the LHC accelerator has been developed, including:

- An analysis module for each of the selected representative components;
- The methodology and algorithms for unsupervised inter-system dependencies detection;
- The user interfaces with the data connector, to integrate the Smart Framework in the existing infrastructure monitoring environment.

Equipment responsables and system experts provide the validation of the outcomes of component-level and system-level analysis algorithms, each expert owning a deep knowledge of its respective components and subsystems.

### 5.1 Systems and components dependencies: LHC point 8 infrastructure hidden dependencies analysis

CERN’s Technical Infrastructure is a complex systems of systems (figure 2), with numerous sub-systems and thousands of components. These are tightly interdependent and interconnected physically, functionally, distributed over a large geographical area. Furthermore they are represented by heterogeneous data set composed of physical measurements, alarms and text reporting. Local malfunctions and perturbations quickly and easily propagate through highly dependent group of components and can cause a cascade of failures with major consequences for the reliability and availability of the accelerator complex.

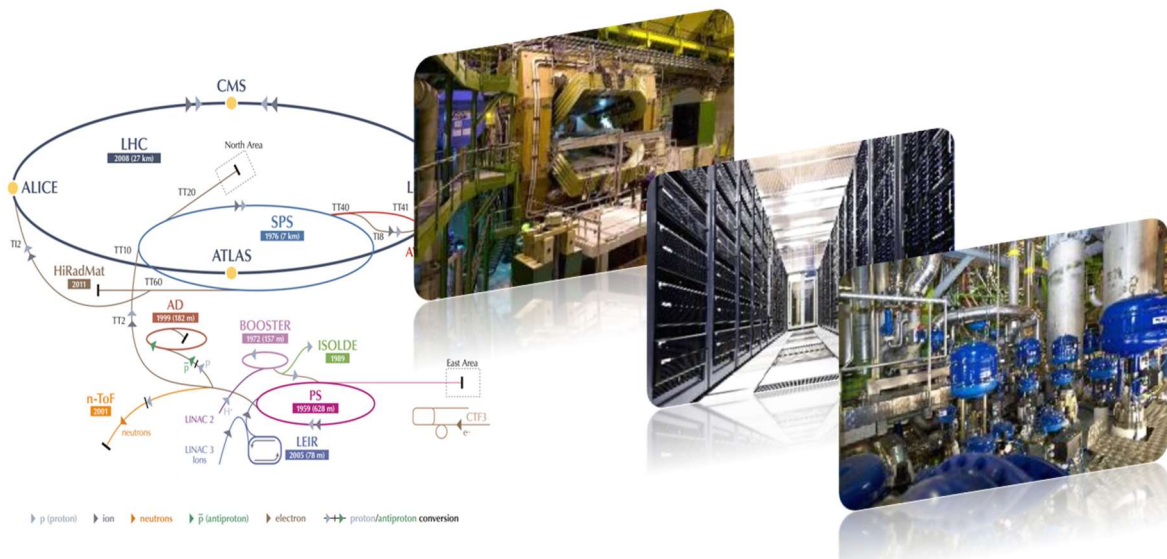


Figure 2: CERN’s Technical Infrastructure

It is not easy to performe vulnerability and resilience modelling with classical methods analysing functional dependencies [7, 8] both from the complexity point of view and the continuous evolution of the installations.

A novel methodology [9] has been proposed to analyse large databases of alarms generated by the monitoring system and infer dependencies between systems and components to identify critical systems, mitigate and eliminate causes of downtime.

The methodology to identify system and components dependencies is based on the representation of the alarm database using a binary matrix and the extraction of association rules using the Apriori algorithm.

The proposed methodology has been first tested on an artificial test dataset and then successfully applied to a real large scale database of alarms from 1/8 of the CERN accelerator infrastructure (the LHC Point 8 technical infrastructure), comprising components and subsystems of three of the major infrastructure systems: the LHC cryogenic, cooling and ventilation and electrical systems. All systems are deeply interconnected and present significant risks of cascade propagation of failures.

More than 10'000 malfunctions have been identified from a database of almost 20 million of alarms collected during one year of operation. The Apriori algorithm was able to extract several association rules identifying abnormal behaviour of groups of components across the three different systems. More than 1'000 rules were extracted, of which more than 100 across the three different systems and 8 new functionally dependent malfunctions identified and confirmed by systems and equipment experts.

	Electric system	Cryogenic system	Cooling and ventilation (CV) systems	Total LHC Zone 8
No. of malfunctions types	9,655	1,472	2,324	13,451
No. of alarms	16,034,050	2,607,030	70,657	18,711,737

Table 1: Number of malfunctions and alarms during one year of operation of CERN Technical Infrastructure in Point 8

	Rules
Total No. of rules extracted	1031
No. Causal cross systems rules	147
No. of groups of functionally dependent malfunctions	8

Table 2: Rules extracted and number of functionally dependent malfunctions identified

## 5.2 Critical components identification: the electrical network perturbations

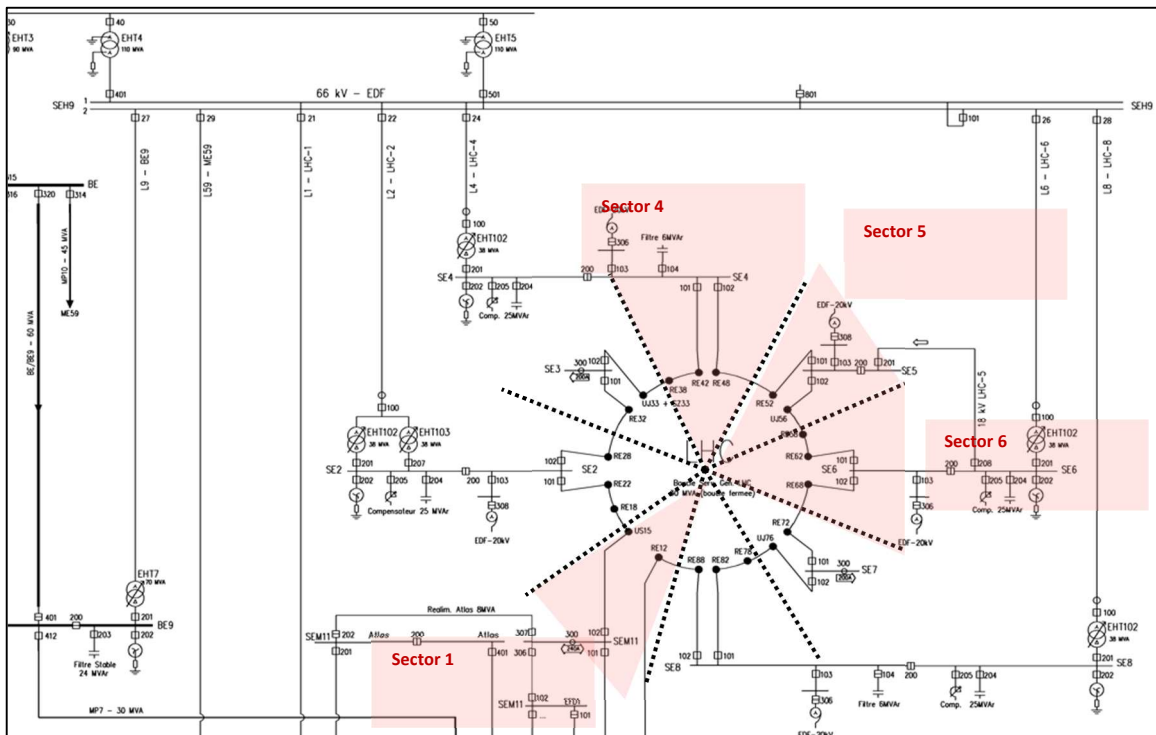
CERN's electrical network distributes electrical power to the whole accelerator complex and impact directly (power converter controlling the beam trajectory via superconducting magnets) and indirectly (cooling and cryogenic systems providing the required temperature conditions for the superconducting magnets nominal operation) the operation of the colliding beams.

Electrical disturbances are generated inside the CERN electrical network (e.g. components short circuits or failures) or outside coming from the French and Swiss feeding grids (e.g. thunderstorms, network failures). They can easily propagate and impact the operation of sensitive components such as the sensitive power converters controlling the beam trajectory or the high power compressors of the cryogenic systems. When the electrical disturbance is detected, the impacted components are shut down to avoid damages or loss of control. The propagation of perturbations through the network cannot easily be modelled as it depends on several features, layout and operating modes.

A significant amount of the overall LHC machine downtime has been caused in the past by electrical disturbances. The identification of the sensitive components and the identification of the propagation path of the disturbances can significantly reduce or mitigate the downtime.

A dataset of electrical disturbances (several thousands) and affected components (few hundreds) with (48) or without beam dump has been analysed using a novel method [10] based on binary classifiers, which associates monitored signal values to nominal operation and failed states. The identification of critical components has been performed using a wrapper feature selection approach based on a binary differential evolution and cost sensitive support vector machine classification. More than 5'000 components (e.g. electrical transformers, UPS, distribution switchboards) and more than 10'000 signals (e.g. power, voltage, current) have been analysed.

The results showed that the 18 kV power distribution level is the most sensitive to the perturbations coming from the 400 kV level. In particular the power converter system connected to the 18 kV transformers are the most sensitive to the perturbations. 22 selected features and 19 critical components have been identified in specific zones of the LHC machine around the 27-km accelerator ring. The false detection rate is 7 % and the missed detection rate is 25 %.





### 5.3 Anomaly detection: beam collimators temperatures sensors failure

Early detection and diagnosis of the failures of critical components is paramount to ensure high reliability and availability. An estimate of the Remaining Useful Life of equipment can provide critical information to improve the reliability and availability of the system as well as optimize the maintenance and consolidation interventions.

A proof of concept based on deep learning and random forest algorithms has been developed and tested on the temperature sensor of beam intercepting devices, the LHC collimators. The models provide the estimated nominal behaviour of the temperature sensors signals and their time evolution. The modeled signal is compared with the actual measured values to detect and anticipate failures of the sensors due to normal and abnormal (e.g. radiation, environmental conditions) ageing. Anticipated detection of the failures provides the possibility of performing predictive maintenance and prevent a critical component failure during beam operation. The detection threshold is optimised and tuned by sweeping through a set of values to maximise the f-score as show in figure 4.

The results in figure 4 show the reached accuracy of 95 % in detecting failing sensors and the capability of anticipating by several hours their breakdown.

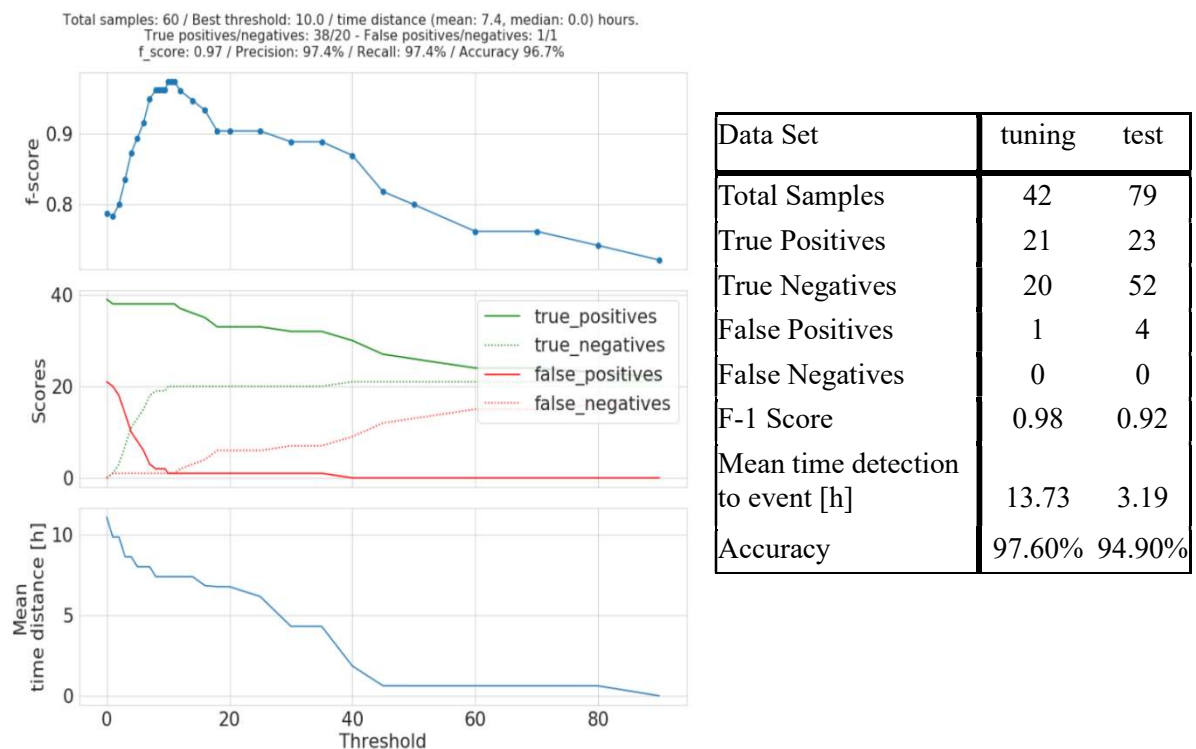


Figure 4: F-score maximisation curves and results table

### 6. Further development

Big Data analytics, the developed machine learning techniques and the preliminary results obtained in the proof of concept of the CASO framework have proven to be successful on several use cases. Descriptive and predictive models were successfully extracted directly from data.

The objective to detect critical components and predict anomalies to react ahead of time to failures in order to reduce downtimes and optimize maintenance and operation costs has been demonstrated.

CASO will be extended to the whole LHC Technical Infrastructure and linked to the Enterprise Asset and Maintenance Management framework to provide operation and maintenance analytics and guidance to anticipate maintenance and consolidation.

## 7. Acknowledgments

The work reported is the results of a long-term collaboration with Politecnico di Milano and the team and students of Prof. E. Zio and Prof. Baraldi. We acknowledge the valuable support and contribution of F. Antonello, P. Baraldi, A. Castellano, U. Gentile, R. Picatoste.

## References

- [1] S. M. Rinaldi and J. P. Peerenboom and T. K. Kelly, *Identifying, understanding, and analyzing critical infrastructure interdependencies*, IEEE Control Systems 21.6 (2001).
- [2] V. Rosato and L. Issacharoff and F. Tiriticco and S. Meloni and S. Porcellinis and R. Setola, *Modelling interdependent infrastructures using interacting dynamical models*, International Journal of Critical Infrastructures, 4(1-2), (2008).
- [3] Gentile U., Serio L. 2018. *A machine-learning based methodology for performance analysis in particles accelerator facilities*, European Conference on Electrical Engineering and Computer Science (EECS), Bern, 2017, pp. 90-95.
- [4] Serio, L., Antonello, F., Baraldi, P., Castellano, A., Gentile, U., Zio, E. 2018, *Smart framework for the availability and reliability assessment and management of accelerators technical facilities*, 9th International Particle Accelerator Conference, IPAC 2018.
- [5] The CERN accelerator complex, <https://cds.cern.ch/record/2197559?ln=it>, 08-07-2016.
- [6] J. Nielsen and L. Serio, *Technical Services: Unavailability Root Causes, Strategy and Limitations*, in Proc. 7th Evian Workshop on LHC beam operation, Evian Les Bains, France, December 2016.
- [7] Rinaldi, S.M.; Peerenboom, J.P.; Kelly, T.K. , *Understanding, and analyzing critical infrastructure interdependencies*, IEEE Control Syst. 2001, 21, 11–25.
- [8] Zio, E. 2016, *Challenges in the vulnerability and risk analysis of critical infrastructures*, (2016) Reliability Engineering and System Safety, 152, pp. 137-150.
- [9] Antonello, F., Baraldi, P., Shokry, A., Zio, E., Gentile, U., Serio, L., *Data-Driven Extraction of Association Rules of Dependent Abnormal Behaviour Groups*, Proceedings of the 29th International European Safety and Reliability Conference, ESREL 2019.
- [10] A. Castellano, *A Feature Selection-Based Approach for the Identification of Critical Components in Complex Technical Infrastructures: Application to the CERN Large Hadron Collider*, MSc Thesis, Politecnico di Milano, 2018.