





# KoopaML, a Machine Learning platform for medical data analysis


Alicia García-Holgado  [ GRIAL Research Group, University of Salamanca | aliciagh@usal.es ]

Andrea Vázquez-Ingelmo  [ GRIAL Research Group, University of Salamanca | andreavazquez@usal.es ]


Julia Alonso-Sánchez  [ GRIAL Research Group, University of Salamanca | juliaalonso@usal.es ]


Francisco José García-Peñalvo  [ GRIAL Research Group, University of Salamanca | fgarcia@usal.es ]

Roberto Therón  [ GRIAL Research Group, University of Salamanca | theron@usal.es ]

Jesús Sampedro-Gómez  [ Cardiology Department, Hospital Universitario de Salamanca, SACyL. IBSAL, University of Salamanca, and CIBERCV (ISCiii) | jmsampedro@saludcastillayleon.es ]

Antonio Sánchez-Puente  [ Cardiology Department, Hospital Universitario de Salamanca, SACyL. IBSAL, University of Salamanca, and CIBERCV (ISCiii) | asanchezpu@saludcastillayleon.es ]

Víctor Vicente-Palacios  [ Philips Healthcare, Spain | victor.vicente.palacios@philips.com ]

P. Ignacio Dorado-Díaz  [ Cardiology Department, Hospital Universitario de Salamanca, SACyL. IBSAL, University of Salamanca, and CIBERCV (ISCiii) | acho@usal.es ]

Pedro L. Sánchez  [ Cardiology Department, Hospital Universitario de Salamanca, SACyL. IBSAL, University of Salamanca, and CIBERCV (ISCiii) | plsanchez@usal.es ]

## Abstract

Machine Learning allows facing complex tasks related to data analysis with big datasets. This Artificial Intelligence branch allows not technical contexts to get benefits related to data processing and analysis. In particular, in medicine, medical professionals are increasingly interested in Machine Learning to identify patterns in clinical cases and make predictions regarding health issues. However, many do not have the necessary programming or technological skills to perform these tasks. Many different tools focus on developing Machine Learning pipelines, from libraries for developers and data scientists to visual tools for experts or platforms to learn. However, we have identified some requirements in the medical context that raise the need to create a customized platform adapted to end-user found in this context. This work describes the design process and the first version of KoopaML, an ML platform to bridge the data science gaps of physicians while automatizing Machine Learning pipelines. The platform is focused on enhanced interactivity to improve the engagement of physicians while still providing all the benefits derived from the introduction of Machine Learning pipelines in medical departments, as well as integrated ongoing training during the use of the tool's features.

**Keywords:** Machine Learning, Data Analysis, Machine Learning Pipelines, Learning Platform, Health.

## 1 Introduction

Machine Learning (ML) applications are continuously growing in different fields. The application of ML algorithms enables the possibility of providing support to automate the undertaking of complex tasks.

Moreover, datasets are constantly being generated through several sources, providing enough data to apply these kinds of algorithms. However, the application of ML is not trivial; it is necessary to rely on data science and programming skills to select the proper algorithm and preprocess the datasets accordingly.

This discipline has permeated many areas, one of those is medicine, with applications in a wide range of fields, from diagnosis and prediction of risk factors to drug development (Nature Materials, 2019; Spruit & Lytras, 2018); although the accuracy of the results is directly linked to the quality of the data provided (Rajkomar et al., 2019). The health context involves several tasks, including diagnosis, classification, disease detection, segmentation, assessment of organ functions, etc. (González Izard et al., 2020; Izard et al., 2018; Litjens et al., 2017), that can be partly automated and enhanced through artificial intelligence support.

More and more medical professionals want to learn about Machine Learning, train their first ML model, or expand their knowledge and skills on preparing or processing their data for training. However, applying ML is not an easy task without technical knowledge. If we focus on health domain, using datasets and ML algorithms in a sensitive task. It could be dangerous that medical professionals do not understand how to use a model or how to analyze the data.

Given the benefits that can be derived from the application of ML in the medical domain, it is crucial to democratize knowledge regarding algorithms and pipelines, as it could enable non-expert users to support their decision-making processes with Artificial Intelligence (AI).

In this study, we present KoopaML, an ML platform to assist non-expert users in the definition and application of ML pipelines. The main challenge of this platform is to provide a proper user interface to ease the understanding of the outcomes and processes involved in ML pipelines. For these reasons, we followed a user-centered design approach to capture requirements and needs from a variety of users profiles, including AI experts and non-experts. This paper extends the work presented initially at the International Conference VII Iberoamerican Conference on Human-Computer Interaction (García-Holgado et al., 2021).

The rest of the paper is structured as follows. Section 2 outlines related applications and works to assist users in ML and data science tasks. Section 3 describes the ML platform proposal. Section 4 details the user study carried out to validate the conceptual application. Section 5 summarizes the focus group results and presents the first version of the application. Section 6 provides technical details about the development and deployment. Finally, section 7 presents the conclusions derived from this work.

## 2 Related works

Several tools are already developed for supporting ML processes. We can organize them into the following categories:

- Tools that require programming skills (for developers and data scientists).
- Tools that do not require programming skills. This category is also divided into two subcategories:
  - Tools for target experts and non-specialized users.
  - Tools for educational purpose.

First, tools for developers and data scientists provide libraries for creating ML applications. For example, TensorFlow, which is an ML system that operates at large scale and in heterogeneous environments (Abadi et al., 2016), helps researchers push the state-of-the-art in ML and developers quickly build and deploy ML-powered applications. Moreover, Kubeflow (Bisong, 2019b) is dedicated to deploying ML workflows in Kubernetes, an open source orchestrator for deploying containerized applications (Burns et al., 2017). This toolkit started as a simpler way to run TensorFlow jobs on Kubernetes, but it has become a framework for running end-to-end ML pipelines.

Another example is Apache Mahout, a library for scalable ML on distributed dataflow systems (Anil et al., 2020). Also, Apache has Spark MLlib, a library for Apache Spark (Zaharia et al., 2016) with “fast and scalable implementations of standard learning algorithms for common learning settings including classification, regression, collaborative filtering, clustering, and dimensionality reduction” (Meng et al., 2016). In this category, we can also include libraries for Python such as PyTorch, Scikit-learn, or Keras.io, and cloud services such as Google Colab, which is a serverless Jupyter notebook environment (Bisong, 2019a).

Second, there are applications for target experts and at the same time provide tools for non-specialist users. Several applications provide visual environments that support the visual definition of ML models.

For example, Weka, a collection of ML algorithms for data mining tasks. It has four environments: Simple CLI, a simple interface based on commands; Explorer, to explore the dataset; Experimenter, systematic comparison of a run of Weka's predictive algorithms on a collection of datasets; and Knowledge Flow, a visual interface that enables users to specify a data stream by graphically connecting components representing data sources, preprocessing tools, learning

algorithms, evaluation methods, and visualization tools (Frank et al., 2009; Hall et al., 2009).

RapidMiner Studio provides tools for building ML workflows in a comprehensive data science platform. It has the Visual Workflow Designer tools to create ML workflows, and each step is documented for complete transparency. This part of the tool allows connecting the data source, automated in-database processing, data visualization as well as the Model Validation process (Bjaoui et al., 2020).

SPSS by IBM includes a product for supporting visual data science and ML. The main work area in SPSS Modeler is the stream canvas, an interface to build ML streams connecting nodes (McCormick & Salcedo, 2017).

Another example is the KNIME Analytics Platform. It provides tools for creating visual workflows for data analytics with a graphical interface without the need for coding. KNIME is a modular environment, which enables easy visual assembly and interactive execution of a data pipeline (Berthold et al., 2009).

On the other hand, the ML has started to introduce in primary and secondary education. This has prompted the development of tools to help non-expert users, such as children, to perform simple ML processes using a visual interface. In particular, two tools are noteworthy. LearningML (Rodríguez-García et al., 2021), which is a tool to foster computational thinking skills through practical AI Projects, and Machine Learning for Kids (<https://machinelearningforkids.co.uk>). Both tools are based on a simple pipeline used to train models and integration with other tools to use the trained model. In particular, both are integrated with Scratch, a visual programming language based on blocks developed by the Media Lab of the Massachusetts Institute of Technology.

There are plenty of powerful applications focused on easing the application of ML algorithms as well as educational tools to understand these complex workflows. However, our application context asks for a customized tool with specific requirements related to the health sector and more emphasis on providing an educational experience to those unskilled users while using the platform. The following sections will outline these requirements and our approach to address the needs regarding the automatization of ML pipelines in this domain.

## 3 Platform definition

### 3.1 The problem

The health sector is taking advantage of AI algorithms to enhance the decision-making processes and support complex common activities such as image segmentation, disease detection, identification of risk factors, etc (González Izard et al., 2020; Izard et al., 2018; Litjens et al., 2017). However, it is important to rely on robust data science skills to benefit from these algorithms. In fact, in the health sector, it is necessary to rely both on data science skills as well as on domain knowledge to get the most out of the application of AI pipelines in the health sector and to avoid biases (C.

Weyerer & F. Langer, 2019; Ferrer et al., 2021; Hoffman, 2021; Wachter et al., 2021).

Having both data science skills and domain knowledge is very powerful, but it is also challenging, as physicians are not usually (deeply) trained in data science, and data scientists might not be specialized in specific domains. For these reasons, it is necessary to support them to acquire these skills and fill the domain knowledge gaps, which are usually addressed by having multidisciplinary teams.

However, bringing AI and data science concepts closer to physicians would be more efficient, as they would be less dependent on data scientists (Miller, 2019). The presented tool lays its foundations on this specific issue: to support domain experts to fill data science gaps while automatizing ML pipelines.

Section 2 proved that there are already plenty of tools that address the automatization of AI and data science pipelines, but the necessity of adapting them to the user needs identifying in the medical sector asks for a customized tool. Although these tools are mostly generic and powerful, we want to focus on enhanced interactivity to improve the engagement of physicians while still providing all the benefits derived from the introduction of ML pipelines in medical departments, as well as an integrated ongoing training during the use of the tool's features.

On the other hand, the necessity of developing KoopaML was also motivated by the specific requirements of AI team from the cardiology department of the University Hospital of Salamanca, as they have developed different ML pipelines they wanted to make reusable and accessible to other teams (Sampedro-Gómez et al., 2020).

Moreover, by developing this customized platform, we can also focus on the integration of these services with other developed applications for the cardiology department of the University Hospital of Salamanca, such as the CARTIER-IA platform (García-Peñalvo, Vázquez-Ingelmo, et al., 2021; Vázquez-Ingelmo et al., 2021), fostering the creation of a robust technological ecosystem (García-Holgado & García-Peñalvo, 2017, 2019).

### 3.2 User needs

The development of the tool follows a user-center approach. The first step was defining the audience to develop a solution appropriate to users' needs, objectives and situations. Therefore, it is essential to understand the users that make up the target audience. We have used Alan Cooper's Persona method (Cooper, 1999) for identifying and defining the primary and secondary users. We followed the 10-step guide by Lene Nielsen (Nielsen, 2003, 2004, 2013a, 2013b):

1. Collect data about who the users are, how many, what they will do with the system.
2. Form a hypothesis. To identify the differences among the users.
3. Ensure everyone accepts the hypothesis.
4. Establish a number of personas.
5. Construct and describe your personas.
6. Prepare situations for your personas to identify needs and situations.

7. Get acceptance from your organization.
8. Disseminate knowledge.
9. Create scenarios for your personas.
10. Make ongoing adjustments.

The users' analysis was conducted in collaboration with the cardiology department of the University Hospital of Salamanca. In particular, we have a meeting to discuss about their needs. After the meeting, we proposed a set of user profiles and domain experts provided feedback. Based on the analysis of the users, the Persona technique was applied to be able to deepen our understanding of the users' needs. The users have been categorized into two target groups. The main user group is physicians. They have knowledge or interest in AI/ML and have data. Moreover, they do not have enough knowledge of programming or using AI algorithms. These users are also characterized by a wide age range and different levels of digital competence, which also influence the user needs.

The main target group was divided into several User Persona representing the whole group due to the wide age range and the many different personalities. For each User Persona the following information was collected:

- Body, including name, age, and picture.
- Psyche, to indicate whether the person is introvert or extrovert.
- Background from the point of view of occupation.
- Emotions and attitudes about technology, relationships, and information.
- Personal characteristics.

The main Users Persona are shown in text format:

- **Genoveva**
  - Body: Genoveva, 60 years old.
  - Psyche: Extrovert.
  - Background: Veteran physician.
  - Emotions and attitudes: She is reluctant to change their routine, has limited knowledge of technology, but is very interested in tools that help in his profession.
  - Personal characteristics: She has a daily routine that she follows with great diligence, trying to avoid stress and unexpected situations. She loves her profession and appreciates teamwork.
- **Javier**
  - Body: Javier, 40 years old.
  - Psyche: Introvert.
  - Background: Physician.
  - Emotions and attitudes: He handles a large amount of data, limited knowledge about ML, has zero knowledge about programming, and good use of technology.
  - Personal characteristics: He is methodical, organized, has an eidetic memory, which helps him to deal with the multiple data present in his work.
- **Aarón**
  - Body: Aarón, 50 years old.

- Psyche: Extrovert.
- Background: Physician.
- Emotions and attitudes: Neither programming nor ML knowledge.
- Personal characteristics: He is curious about everything around him; he likes to learn.
- **Sheila**
  - Body: Sheila, 45 years old.
  - Psyche: Extrovert.
  - Background: Physician.
  - Emotions and attitudes: Without ML or programming knowledge, and good use of technology.
  - Personal characteristics: She is observant, has great intuition, and is responsible.
- **Leticia**
  - Body: Leticia, 30 years old.
  - Psyche: Introvert.
  - Background: Internal Medicine Resident Physician.
  - Emotions and attitudes: Very basic knowledge about ML, zero programming knowledge.
  - Personal characteristics: She is hard-working, responsible, insecure, nervous, tenacious, curious, and pragmatic; she likes to draw her own conclusions.

The secondary user group is students and ML experts. Medical students neither do they have knowledge of programming, but they are mostly young, and it is supposed that they have more experience with digital tools. Regarding ML experts, they have knowledge of AI/ML as well as programming skills; however, they are not domain experts. We defined two User Persona, one to represent students and one to represent ML experts:

- **Laura**
  - Body: Laura, 23 years old.
  - Psyche: Introvert.
  - Background: Student of medicine.
  - Emotions and attitudes: Fundamental knowledge of ML, zero programming knowledge.
  - Personal characteristics: She is organized, restless, practical, and gets nervous with chaos and disorder.
- **Juan**
  - Body: Juan, 49 years old.
  - Psyche: Introvert.
  - Background: Machine Learning Expert.
  - Emotions and attitudes: Expert knowledge of ML.
  - Personal characteristics: He is meticulous, likes to feel in control, non-conformist, perfectionist, insecure.

The needs of physicians and medical students are mainly focused on:

- Get a tool to apply AI/ML in medical datasets without technical expertise and with limited knowledge of ML.
- Visualize and analyze medical data.
- Learn about data analysis and its usefulness and benefits in medicine.
- Learn about ML algorithms in a practical way.
- Be able to use an ML application that allows detailed data visualization and helps in interpreting the data.

On the other hand, the needs of ML experts lie in facilitating the integration of algorithms in ML pipelines. Furthermore, we also identify customization as a need. They need to modify the parameters and heuristics of an ML algorithm and learn from the results of their tests.

### 3.3 Main scenarios

User scenarios are stories about people and their activities (Carroll, 2000). In particular, for each User Persona, their main need and the scenario where each one would occur have been identified (Table 1). Through this analysis, we identified the main scenarios. We have identified five scenarios that cover the goals and questions to be achieved through the tool. The scenarios are mainly focused on supplying the needs of the physicians, the main user group.

**Table 1.** User Persona definition.

Name	Main needs	Situation
Genoveva	Train Machine Learning algorithms without programming or ML knowledge.	A veteran doctor, her hospital wants to incorporate a Machine Learning tool. She is interested and considers it an advantage but does not know how to use it.
Javier	Visualize and analyze medical data.	A physician who often deals with a large amount of information that he wants to be able to organize, analyze and visualize.
Aarón	Learn about data analysis and its usefulness and benefits in medicine.	A physician who wants to learn new techniques and tools to help him in his profession.
Sheila	Be able to use an ML application that allows detailed data visualization and helps in interpreting the data.	A physician who attaches great importance to the visualization and interpretation of data.
Leticia	Learn about ML algorithms in a practical way.	A resident intern who has observed her mentor using image recognition with ML at

		work and wonders whether it might be better to use a different algorithm and when it is more appropriate to use one or the other.
Laura	Learn the basics and operation of ML to understand the theory that she is studying quickly.	Fifth-year medical student studying ML applied to medicine.
Juan	Be able to customize the parameters and heuristics of an ML application and learn from the results of his tests.	An ML expert who wants to be able to customize the parameters and heuristics of a machine learning application.

First scenario is the pipelines definition. The platform interface will allow the definition of ML pipelines through graphical elements and interactions to materialize the tasks defined using a programming approach such as those developed using SciLuigi (<https://github.com/pharmbio/sciluigi>), a wrapper for Spotify's Luigi library (<https://github.com/spotify/luigi>), the Python module to build complex pipelines of batch jobs. Users will be able to customize the pipeline and access the intermediate results, as well as save configurations for sharing or later use.

The second scenario covers the algorithm training. Users of the platform will be able to train various algorithms by providing some input data. The platform will allow the user to choose the type of algorithm and configure the parameters associated with it. It is proposed that the algorithm and parameter selection process will be guided by a series of heuristics based on existing literature (although it is also proposed that expert users can add or modify these heuristics) depending on the scheme, the problem to be solved and the volume of data entered.

The third scenario is focused on visualization and interpretation of the results. Various metrics and results will be obtained after training is completed depending on the algorithm or ML model used. These results can be ROC curve plots, cut-off points with specificity/sensitivity values, variable importance, etc. The platform will assist users in interpreting these results through visualizations, annotations, and explanations.

The next identified scenario will support the data checking and visualization. Users will also be able to check and visualize the input data to explore it. Moreover, the platform will provide feedback regarding the potential problems of the dataset and the feasibility of using different algorithms on the input data.

Finally, the last scenario is related to the use of heuristics. This scenario addresses two objectives, one focused on medical students without ML knowledge and the other for ML experts. First, using heuristics through a rule-based

recommender will be a functionality to use the platform as a didactic tool. The feature will provide a guided process during the definition of pipelines, selection, and training of algorithms, interpretation of the results, etc., to provide an educational component in the platform itself for those users who are not experts in AI or programming. This recommender system will help to reduce incorrect use of the algorithm, for example, when the user has a dataset that cannot use with a particular algorithm regarding the distribution of the data or the size.

Secondly, the heuristics are also for ML expert users, who may be physicians, data scientists, or developers. Particularly, we have identified customization as the main need for ML experts. For this reason, the platform will allow the modification of heuristics. The basic heuristics will be based on existing literature (Buitinck et al., 2013; Scikit-Learn, 2020), but expert users will be able to modify them through XML documents or a graphical interface, being able to store several versions of the heuristics used in the platform.

## 4 Concept validation

The test phase was focused on validating the application to make sure it works well for the primary and secondary users, physicians, and ML experts. In particular, we focused on identifying the design requirements through a focus group with real users.

The concept validation included the development of a high-fidelity prototype (Pernice, 2016) using Adobe XD together a remote focus group (Krueger & Casey, 2014) with final users testing the digital prototype remotely and answering questions related to usability to identify problems and discover new requirements.

### 4.1 The prototype

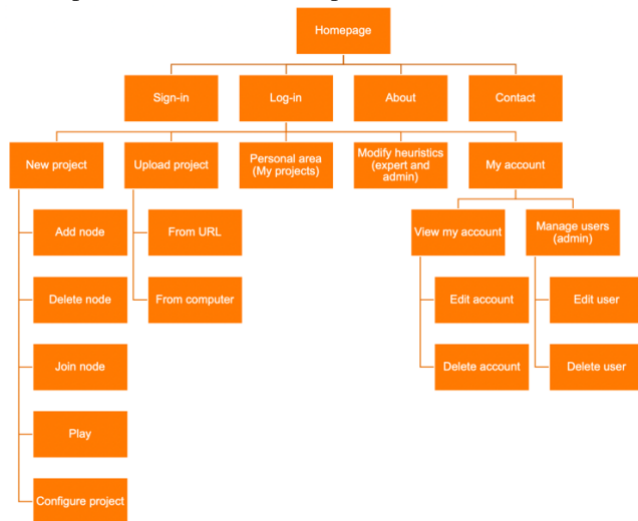
KoopamL aims to assist non-expert users in the definition and application of ML pipelines. The platform will also support the integration with other developed tools.

There are three types of user roles to represent the main target groups. First, "regular user" represents users without advanced knowledge of ML. Usually, physicians and students have this role when they register on the platform. Regular users can use the main functionality of the application. The second user role is "expert", which represents users with advanced knowledge of ML. The expert role has access to the regular user's functionality and can manage the application's heuristics. Finally, the admin role has permissions to perform all the tasks available in the platform, in addition to managing user accounts (validation, creation, deletion, and modification).

The tool is divided into different screens that bring together all the functionality (Figure 1). We can organize the functionality into three groups: the public pages for anonymous users, the functionality related to creating and managing projects, and the functionality to manage users. First, the homepage provides information about the platform and user access. The platform is fully private, so access is protected by user registration and login. In addition, the user

sign-in involves a validation process, so users apply for a new account and administrators must approve it.

The next set of functionality focuses on project management within the platform. A project represents an ML pipeline with one or more data inputs and one or more outputs. Users (both regular and experts) will create ML projects. The platform allows saving the projects in a personal area (Figure 2), so users can reuse the ML pipelines as many times as needed. However, the datasets and the results are not saved in the platform; only the pipeline and the configuration of each node inside the pipeline are saved. This decision was made to avoid problems managing sensitivity medical information. The responsibility of uploading non anonymous medical data fall in the users, although in medical context, and in cardiology, the medical software provides datasets already pseudoanonymized. Users can also download projects to have a copy or send them to another user using any other media (email, shared folder, etc.). Users can upload those projects to their accounts through their personal area using an URL to a shared space or a zip file located in the computer.



**Figure 1.** The main site flows.

Most of the functionality is in the project editor (Figure 3). The editor allows creating a new pipeline or editing an existing one. This interface is divided into three main parts. First, a top bar provides access to the personal area and user account. It also has a button to reset the project.

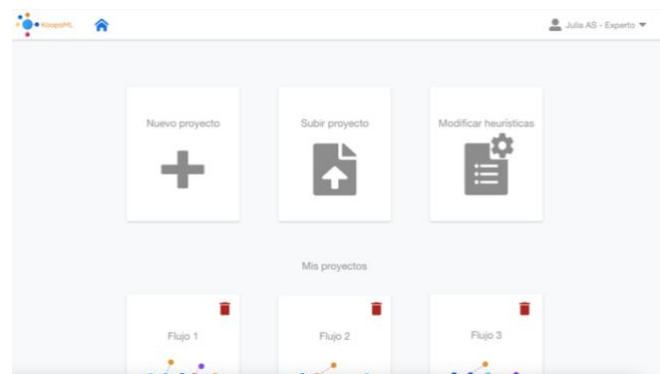
Second, a toolbar containing all the nodes that can be added to the pipeline grouped by categories: data, data preparation, data processing, visualization, ML algorithms, and evaluation. The design is conceived to be able to add new nodes over time. The data preparation and data processing nodes will ensure the data is in the correct format to be used in the tool. In particular, the first version of the tool will support CSV files due to it is the format provided by the medical tools used in the cardiology department.

Moreover, the toolbar provides tools for joining nodes, configuring, executing, and saving the project. The toolbar has two views, a simplified view for expert users and regular users with enough experience using the platform and a detailed view showing the name of all nodes and categories.

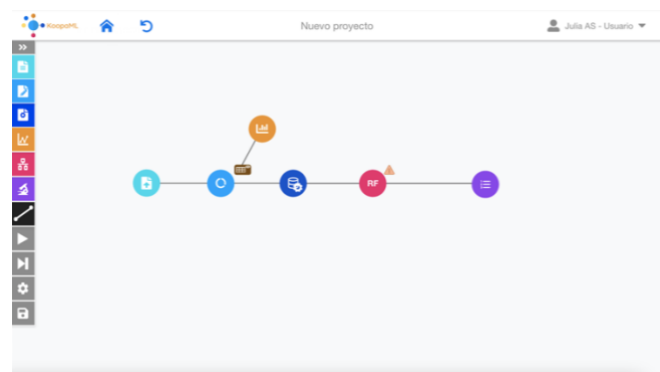
The last and principal part of the project editor is the workspace, a blank area to build the pipeline connecting different nodes (Figure 3). Each node has inputs and outputs; it produces an output, a set of results, which are the input to the next node. A node can be connected with several nodes, for example, a visualization node to get a visual analysis of the intermediate results and a node to apply an algorithm to that dataset. The nodes also have different visual marks to provide useful information:

- A warning mark in those cases in which there is missing information, a problem in the execution of the node or a recommendation to improve the configuration of the node.
- A result mark to indicate that this node produces as output a set of intermediate results that can be retrieved when the pipeline is executed.
- A suggestion mark provides additional information and supports regular users without ML knowledge.

Furthermore, the functionality associated with project management is completed with an interface to manage the heuristics. This tool is accessible only for expert users and administrations through the personal area. Experts can modify the parameters and heuristics of an ML algorithm and learn from the results of their tests. It is possible to create new heuristics, delete and modify existing heuristics, except for the base heuristic.



**Figure 2.** Digital prototype of the personal area for expert users.



**Figure 3.** Digital prototype of a pipeline with different tasks (nodes).

## 4.2 The focus group

The focus group is a qualitative technique that involves a group of participants discussing a topic directed by the researcher. Through a focus group, we can learn about users' attitudes, beliefs, desires, and reactions to concepts (Cope,

2020). It provides valuable assistance to the specification of the interaction and visual design concept of the product under consideration (Kuhn, 2000).

The objective of the focus group is to identify the design requirements and test the digital prototype. The focus group was organized fully online to facilitate user participation and to meet the social distancing measures established by the COVID-19 health crisis (Fardoun et al., 2020; García-Peñalvo et al., 2020; García-Peñalvo, Corell, et al., 2021). We used Zoom as a communication tool because it not only supports grid view, but also allows remote control. During the focus group we gave remote control to some participants to perform some tasks within the digital prototype in Adobe XD.

The participants in the focus group represented the User Persona previously identified:

- CE1: A cardiologist with some knowledge of ML.
- CE2: A cardiologist and researcher involved in ML projects.
- CE3: A cardiologist with a positive attitude towards ML.
- DE1: A physics and data scientist.
- DE2: An industrial engineer and data scientist.
- DE3: A software engineer and data scientist.
- S1: A computer science student working on an ML project related to cardiology.

The focus group was in Spanish, it took 60 minutes, and it was moderated by three women researchers related to software engineering and human-computer interaction. The discussion was divided into two phases. A first phase in which we introduced the platform and explained the user access and roles; and a second phase in which participants answered different questions related to the main parts of the digital prototype (the personal area and the workspace).

There were common questions for each screen:

- Can you identify briefly what can be done on this screen?
- Can you describe it?
- What is your opinion regarding the visual aesthetics of the screen?

After these questions, the moderators shared some specific questions for each part. First, some questions focused on the personal area (Figure 2):

- How would you delete a project?
- How would you download a project?
- How would you create a project?

Second, questions focused on the workspace of a new project:

- How would you start an ML pipeline?
- Do you know how to continue the workflow from this first node?
- How would you modify a previous/already created project?

Later, we showed a workspace with an ML pipeline example (Figure 3):

- Could you describe what you see on this screen?
- How would you get the results once you have the ML pipeline defined?
- How would you explore the results of the ML pipeline?

Finally, some general questions:

- Have you been able to see how to navigate between screens?
- What positive aspects of the platform have caught your attention?
- What negative aspects of the platform have caught your attention?
- What tasks would you like to perform with the platform?

Some questions also involved the interaction of a volunteer with the digital prototype using the remote-control tool provided by Zoom.

## 5 Results

The analysis of the focus group was focused on identifying positive and negative aspects across the different screens. Moreover, we identified a set of suggestions.

### 5.1 Positive comments

The ML experts (DE1, DE2, DE3) and cardiologists (CE1, CE2, CE3) have similar opinions about the workspace. Both indicate that it is an intuitive interface (CE3), notably the first step for uploading data to begin a pipeline (CE2, DE3) and the interaction to add nodes into the pipeline (CE2). Furthermore, CE1 highlights the general order of steps suggested by the category organization in the toolbar. All user groups have underlined the tools to facilitate the use of the platform for users with basic knowledge of ML, such as the tooltips (DE3) and the information about errors and recommendations associated to a pipeline (CE2).

Likewise, domain and ML experts emphasize the design. CE1 indicates that the look & feel is good, and DE3 likes the appearance with a few well-organized buttons and plenty of white space to get started.

Cardiologists pay more attention to pipeline execution. CE1 and CE2 comment on the option of executing each node step by step and seeing results from the initial stages without waiting to get the final results of the pipeline.

On the other hand, ML experts pay more attention to flexibility. The platform allows creating pipelines with few (or many nodes) connecting them in any order (DE1, DE3). Moreover, DE2 remarks that the platform allows adding different ML models and visualizing many interim results.

Finally, the participants highlighted several general positive aspects of the digital prototype. Both ML experts and cardiologists emphasized the interface design, navigation, and simplicity of the tool. Moreover, the ML experts point out that the tool is quite scalable.

### 5.2 Negative comments

Regarding the negative aspects, we have identified the following:

- The way the toolbar is expanded is not intuitive. (DE1).
- The “save data” category of nodes confuses their function. (CE1, DE1, DE3).
- Not fully self-explanatory. It is unclear how to proceed to create and join nodes the first time you use the tool if you have not used a similar tool before. Invites you to follow a trial-and-error strategy. (CE1).
- It is not clear which tasks are in progress and which have been completed during the pipeline execution. (DE3).

### 5.3 Suggestions

Throughout the focus group, the participants suggested several improvements to reduce the identified negative aspects:

- Support new users with the platform indicating which are the first elements of an ML pipeline (DE3) and a set of guidelines or instructions to start the definition of the pipeline (DE3). According to CE1, it would be useful as an example (initial simulation) showing how the platform works.
- Improve the toolbar with tooltips (CE1) and use the expanded view as default (CE2).
- Provide a set of short video tutorials on using the tool, illustrating basic operation and more concrete things (CE1, DE3).
- Restrict the options for defining the pipeline to beginners and allow total flexibility for experts (DE3). For example, a default pipeline structure for people with basic ML skills, so they can start a project with two clicks and learn how to do it (DE2).
- Changing the name of the subcategory “save data” to “download or export data” (DE3).
- Add a visual help to know how to join the nodes within the pipeline (CE1).
- Enable execution not only from the toolbar, but also using the right mouse button because this is a common interaction with this type of tool (DE3).
- Include an execution mark to indicate whether the node's execution within the pipeline was successful or there was a problem (DE1).

### 5.4 First version

The feedback received during the focus group was used to solve the negative comments and introduce the suggested improvements in a second version of the digital prototype. This new version was validated by the domain and ML experts involved in the focus group. This second validation was done via email. The validated prototype was used to develop the first version of the platform.

The main improvements have been the following:

- When a user starts a new project, a list of video tutorials is displayed, showing from first steps to more advanced tips on the functionality of the tool (Figure 4).
- The menu bar is expanded when a user starts a new project (Figure 4).

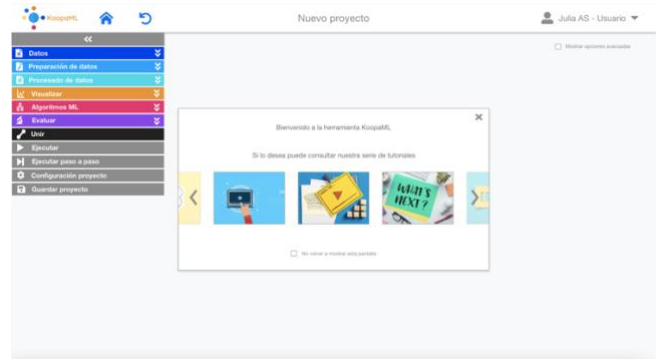


Figure 4. Second digital prototype of the project editor when a user starts a new project.

- Tooltips indicating the name of each category when the menu is collapsed.
- During the execution of a pipeline, indicate which tasks have been completed and which are still in progress (Figure 5).

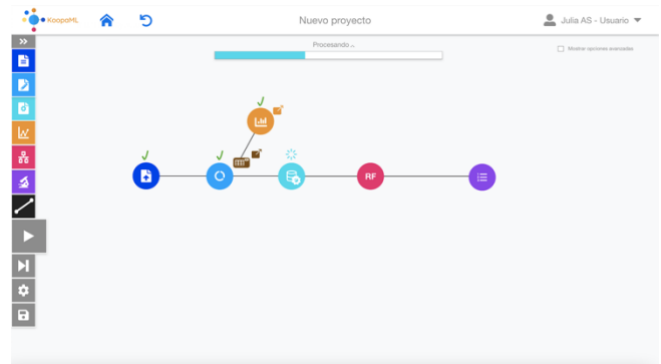


Figure 5. Second digital prototype of the execution of the pipeline.

- In the first version implementation, the nodes become squares to clearly see the joints and the input and output parameters of the nodes (Figure 6).
- Removal of the join node option in the side menu; the joins are made by clicking on the inputs or outputs of the nodes and drawing a line, so this option was useless in the menu.

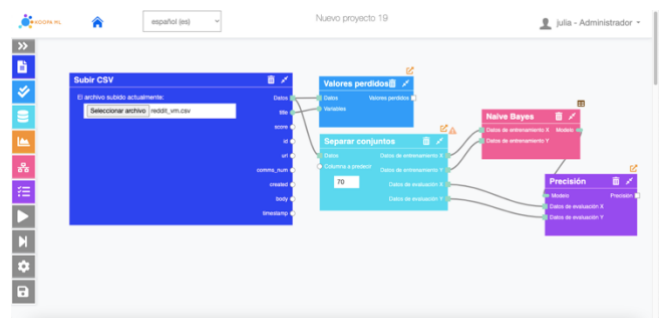


Figure 6. Square nodes in the first version.

- The heuristic codification is represented through a flowchart (Figure 7). This codification will be based on



pseudocode to represent the variables and the connections between them. Moreover, heuristics can be updated, so the flowchart associated will show the changes and the system will change the recommendations based on the updated heuristic.

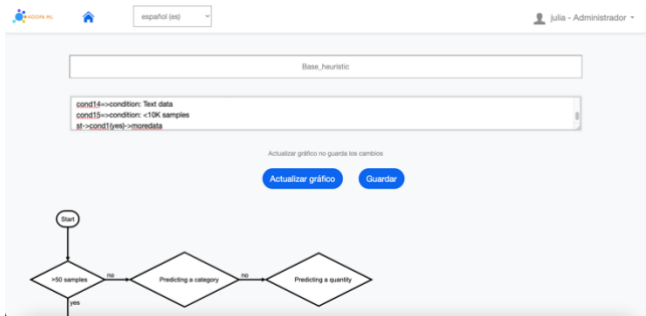


Figure 7. Heuristic editor.

## 6 Deployment

The application was developed using Django, a Python web development framework that allows applications to be scalable thanks to its loosely coupled philosophy [7], due to each component of the application architecture is independent of other components, which facilitates its modification. Regarding the interface, we used Bootstrap combined with jQuery.

The development has also integrated a set of libraries. First, Rete.js has been integrated to create and modify the nodes that make up the pipelines. Rete.js is a modular JavaScript framework oriented to visual programming.

The FlowChart.js library has also been used for the flowchart representation of the heuristics defined in the application. This allows users with the role of administrator or expert to code heuristics in a language similar to pseudocode and graphically visualize them. For the coding of the base heuristic, the Scikit-learn decision tree was used (Scikit-Learn, 2020).

On the other hand, we have used Pandas, a Python data analysis library. In particular, we applied Pandas for data analysis and manipulation tool.

Finally, we also integrated Scikit-learn, an ML library for Python. This library contains already implemented algorithms for Machine Learning and statistical modeling, including classification, regression, clustering and dimensionality reduction.

We deployed the application on a server with Ubuntu 18.04 LTS. In particular, we used the Gunicorn server, a Web Server Gateway Interface (WSGI) HTTP server to deploy the application together with the Nginx proxy server, a virtual environment and a database based on MongoDB 4.4.6 Community Edition (Figure 8).

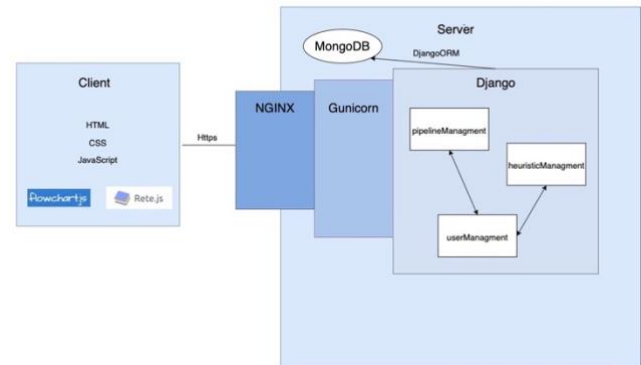


Figure 8. KoopaML deployment.

## 7 Conclusion

The health sector produces many datasets with useful information to support decision-making processes and complex common activities. However, physicians are not deeply trained in data science, although they have the domain knowledge. On the other hand, ML experts are not usually experts in the domain. This work describes the design process of KoopaML, an ML platform to bridge the data science gaps of physicians while automatizing ML pipelines.

There are different tools focused on developing ML pipelines, such as RapidMiner, KNIME Analytics Platform, or SPSS Modeler. However, we have identified some requirements in the medical context that raise the need to create a customized platform adapted to end-user found in this context: physicians with a lack of ML and programming skills that are interested in taking advantage of the application of these algorithms.

One of the novel features of KoopaML that differentiates it from other tools is the flexibility regarding the definition of heuristics. Thanks to the heuristics module, AI experts can tune recommendations and adapt them to different scenarios or audiences.

In addition, the development of a customized tool opens the path for the integration of other already developed tools for the cardiology department at the University Hospital of Salamanca. By implementing communication mechanisms, it is possible to connect different platforms to foster the creation of a technological ecosystem with data science features specifically adapted to the medical context requirements. Further research might explore the integration of DICOM images into the pipelines to apply deep learning. Moreover, the interactive modification of heuristics could serve to make dynamic and automatic suggestions to users on which algorithm to use depending on the input data, task to be performed, etc.

The focus group session provided a considerable amount of data concerning usability functions and their validation by domain and ML experts. It has provided information to understand users' current situation and needs, getting the perspective of multiple users discussing the same requirement and functionality. On the other hand, we have collected the reactions to the digital prototype.

The focus group results have served as an input to develop a new version of the digital prototype to solve the

main detected problems and improve it, including some suggestions. The second version was validated by the participants in the focus group and was the basis for developing the first version of the platform.

## Acknowledgments

This research work has been supported by the Spanish Ministry of Education and Vocational Training under an FPU fellowship (FPU17/03276). This work was also supported by national (PI14/00695, PIE14/00066, PI17/00145, DTS19/00098, PI19/00658, PI19/00656 Institute of Health Carlos III, Spanish Ministry of Economy and Competitiveness and co-funded by ERDF/ESF, “Investing in your future”) and community (GRS 2033/A/19, GRS 2030/A/19, GRS 2031/A/19, GRS 2032/A/19, SACYL, Junta Castilla y León) competitive grants.

## References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., . . . Zheng, X. (2016). TensorFlow: A System for Large-Scale Machine Learning. In *12th USENIX Symposium on Operating Systems Design and Implementation OSDI 16* (pp. 265-283). USENIX Association. <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>
- Anil, R., Capan, G., Drost-Fromm, I., Dunning, T., Friedman, E., Grant, T., Quinn, S., Ranjan, P., Schelter, S., & Yilmazel, Ö. (2020). Apache Mahout: Machine Learning on Distributed Dataflow Systems. *Journal of Machine Learning Research*, 21(127), 1-6. <https://jmlr.org/papers/v21/18-800.html>
- Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., & Wiswedel, B. (2009). KNIME - the Konstanz information miner: version 2.0 and beyond. *SIGKDD Explor. Newsl.*, 11(1), 26-31. <https://doi.org/10.1145/1656274.1656280>
- Bisong, E. (2019a). Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners* (pp. 59-64). Apress. [https://doi.org/10.1007/978-1-4842-4470-8\\_7](https://doi.org/10.1007/978-1-4842-4470-8_7)
- Bisong, E. (2019b). Kubeflow and Kubeflow Pipelines. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (pp. 671-685). Apress. [https://doi.org/10.1007/978-1-4842-4470-8\\_46](https://doi.org/10.1007/978-1-4842-4470-8_46)
- Bjaoui, M., Sakly, H., Said, M., Kraiem, N., & Bouhlel, M. S. (2020). *Depth insight for data scientist with RapidMiner « an innovative tool for AI and big data towards medical applications»* Proceedings of the 2nd International Conference on Digital Tools & Uses Congress, Virtual Event, Tunisia. <https://doi.org/10.1145/3423603.3424059>
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., & Varoquaux, G. e. (2013). *API design for machine learning software: experiences from the scikit-learn project* ECML PKDD Workshop: Languages for Data Mining and Machine Learning.
- Burns, B., Beda, J., & Hightower, K. (2017). *Kubernetes: Up & Running. Dive into the Future of Infrastructure*. O'Really Media.
- C. Weyerer, J., & F. Langer, P. (2019). Garbage in, garbage out: The vicious cycle of ai-based discrimination in the public sector. Proceedings of the 20th Annual International Conference on Digital Government Research, Dubai, United Arab Emirates.
- Carroll, J. (2000). *Making use: Scenario-based design of human-computer interactions*. The MIT Press.
- Cooper, A. (1999). *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity*. Sams.
- Cope, S. (2020). *Focus Groups: Are They Right for You?* Digital.gov. Retrieved March 12 from <https://digital.gov/2015/04/17/focus-groups-are-they-right-for-you/>
- Fardoun, H., González-González, C. S., Collazos, C. A., & Yousef, M. (2020). Exploratory Study in Iberoamerica on the Teaching-Learning Process and Assessment Proposal in the Pandemic Times. *Education in the Knowledge Society* 21. <https://doi.org/10.14201/eks.23437>
- Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2021). Bias and Discrimination in AI: a cross-disciplinary perspective. *IEEE Technology and Society Magazine*, 40(2), 72-80. <https://doi.org/10.1109/MTS.2021.3056293>
- Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., & Witten, I. H. (2009). Weka-A Machine Learning Workbench for Data Mining. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook*. Springer. [https://doi.org/10.1007/978-0-387-09823-4\\_66](https://doi.org/10.1007/978-0-387-09823-4_66)
- García-Holgado, A., & García-Peñalvo, F. J. (2017). A Metamodel Proposal for Developing Learning Ecosystems. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Novel Learning Ecosystems. 4th International Conference, LCT 2017. Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017. Proceedings, Part I* (Vol. 10295, pp. 100-109). Springer International Publishing. [https://doi.org/10.1007/978-3-319-58509-3\\_10](https://doi.org/10.1007/978-3-319-58509-3_10)
- García-Holgado, A., & García-Peñalvo, F. J. (2019). Validation of the learning ecosystem metamodel using transformation rules. *Future Generation Computer Systems*, 91, 300-310. <https://doi.org/10.1016/j.future.2018.09.011>
- García-Holgado, A., Vázquez-Ingelmo, A., Alonso, J., García-Peñalvo, F. J., Sampedro-Gómez, J., Sánchez-Puente, A., Vicente-Palacios, V., Dorado-Díaz, P. I., & Sánchez, P. L. (2021). User-centered design approach for a machine learning platform

- for medical purpose. In P. Ruiz, V. Agredo Delgado, & A. Kawamoto (Eds.), *7th Iberoamerican Workshop, HCI-COLLAB 2021, Sao Paulo, Brazil (September 8–10, 2021)* (Vol. 1478, pp. 237-249). Springer. [https://doi.org/10.1007/978-3-030-92325-9\\_18](https://doi.org/10.1007/978-3-030-92325-9_18)
- García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande-de-Prado, M. (2020). Online Assessment in Higher Education in the Time of COVID-19. *Education in the Knowledge Society*, 21. <https://doi.org/10.14201/eks.23086>
- García-Peñalvo, F. J., Corell, A., Rivero-Ortega, R., Rodríguez-Conde, M. J., & Rodríguez-García, N. (2021). Impact of the COVID-19 on Higher Education: An Experience-Based Approach. In F. J. García-Peñalvo (Ed.), *Information Technology Trends for a Global and Interdisciplinary Research Community* (pp. 1-18). IGI Global.
- García-Peñalvo, F. J., Vázquez-Ingelmo, A., García-Holgado, A., Sampedro-Gómez, J., Sánchez-Puente, A., Vicente-Palacios, V., Dorado-Díaz, P. I., & Sánchez, P. L. (2021). Application of Artificial Intelligence algorithms within the medical context for non-specialized users: the CARTIER-IA platform. *International Journal of Interactive Multimedia and Artificial Intelligence*, 6(6), 46-53. <https://doi.org/10.9781/ijimai.2021.05.005>
- González Izard, S., Sánchez Torres, R., Alonso Plaza, Ó., Juanes Méndez, J. A., & García-Peñalvo, F. J. (2020). Nextmed: Automatic Imaging Segmentation, 3D Reconstruction, and 3D Model Visualization Platform Using Augmented and Virtual Reality. *Sensors (Basel, Switzerland)*, 20(10), 2962. <https://doi.org/10.3390/s20102962>
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1), 10–18. <https://doi.org/10.1145/1656274.1656278>
- Hoffman, S. (2021). The Emerging Hazard of AI-Related Health Care Discrimination. *Hastings Center Report*, 51(1), 8-9. <https://doi.org/10.1002/hast.1203>
- Izard, S. G., Juanes, J. A., García Peñalvo, F. J., Estella, J. M. G., Ledesma, M. J. S., & Ruisoto, P. (2018). Virtual Reality as an Educational and Training Tool for Medicine. *Journal of Medical Systems*, 42(3), 50. <https://doi.org/10.1007/s10916-018-0900-2>
- Krueger, R. A., & Casey, M. A. (2014). *Focus Groups: A Practical Guide for Applied Research*. Sage publications.
- Kuhn, K. (2000). Problems and Benefits of Requirements Gathering With Focus Groups: A Case Study. *International Journal of Human-Computer Interaction*, 12(3-4), 309-325. <https://doi.org/10.1080/10447318.2000.9669061>
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Med Image Anal*, 42, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- McCormick, K., & Salcedo, J. (2017). *IBM SPSS Modeler essentials: Effective techniques for building powerful data mining and predictive analytics solutions*. Packt Publishing Ltd.
- Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D., Amde, M., & Owen, S. (2016). MLlib: Machine Learning in Apache Spark. *The Journal of Machine Learning Research*, 17(1), 1235-1241.
- Miller, D. D. (2019). The medical AI insurgency: what physicians must know about data to practice with intelligent machines. *npj Digital Medicine*, 2(1), 62. <https://doi.org/10.1038/s41746-019-0138-5>
- Nature Materials. (2019). Ascent of machine learning in medicine. *Nature Materials*, 18(5), 407-407. <https://doi.org/10.1038/s41563-019-0360-1>
- Nielsen, L. (2003). Constructing the user. In C. Stephanidis & J. Jacko (Eds.), *Human-computer interaction: theory and practice (Part 2)* (Vol. 2, pp. 430-434). CRC Press.
- Nielsen, L. (2004). *Engaging Personas and Narrative Scenarios*. Samfundslitteratur.
- Nielsen, L. (2013a). Personas. In *The Encyclopedia of Human-Computer Interaction*. The Interaction Design Foundation. <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/personas>
- Nielsen, L. (2013b). *Personas - User Focused Design*. Springer. <https://doi.org/10.1007/978-1-4471-4084-9>
- Pernice, K. (2016). UX Prototypes: Low Fidelity vs. High Fidelity. <https://www.nngroup.com/articles/ux-prototype-hi-lo-fidelity/>
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, 380(14), 1347-1358. <https://doi.org/10.1056/NEJMra1814259>
- Rodríguez-García, J. D., Moreno-León, J., Román-González, M., & Robles, G. (2021). Evaluation of an Online Intervention to Teach Artificial Intelligence with LearningML to 10-16-Year-Old Students. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 177–183). Association for Computing Machinery. <https://doi.org/10.1145/3408877.3432393>
- Sampedro-Gómez, J., Dorado-Díaz, P. I., Vicente-Palacios, V., Sánchez-Puente, A., Jiménez-Navarro, M., San Roman, J. A., Galindo-Villardón, P., Sanchez, P. L., & Fernández-Avilés, F. (2020). Machine Learning to Predict Stent Restenosis Based on Daily Demographic, Clinical, and Angiographic Characteristics. *Canadian Journal of Cardiology*, 36(10), 1624-1632. <https://doi.org/10.1016/j.cjca.2020.01.027>
- Scikit-Learn. (2020). Choosing the right estimator - Scikit-Learn documentation. [https://scikit-learn.org/stable/tutorial/machine\\_learning\\_map/](https://scikit-learn.org/stable/tutorial/machine_learning_map/)
- Spruit, M., & Lytras, M. (2018). Applied data science in patient-centric healthcare: Adaptive analytic

- systems for empowering physicians and patients. *Telematics and Informatics*, 35(4), 643-653. <https://doi.org/https://doi.org/10.1016/j.tele.2018.04.002>
- Vázquez-Ingelmo, A., Alonso, J., García-Holgado, A., García-Peñalvo, F. J., Sampedro-Gómez, J., Sánchez-Puente, A., Vicente-Palacios, V., Dorado-Díaz, P. I., & Sánchez, P. L. (2021). Usability Study of CARTIER-IA: A Platform for Medical Data and Imaging Management. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies: New Challenges and Learning Experiences. 8th International Conference, LCT 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I* (pp. 374-384). Springer. [https://doi.org/10.1007/978-3-030-77889-7\\_26](https://doi.org/10.1007/978-3-030-77889-7_26)
- Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, 41, 105567. <https://doi.org/10.2139/ssrn.3547922>
- Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, S., & Stoica, I. (2016). Apache Spark: a unified engine for big data processing. *Commun. ACM*, 59(11), 56–65. <https://doi.org/10.1145/2934664>