Vázquez-Ingelmo, A., García-Holgado, A., García-Peñalvo, F. J., Fraile-Sanchón, R., Andrés-Fraile, E., Pérez-Sánchez, P., Antúnez-Muiños, P., Sánchez-Puente, A., Vicente-Palacios, V., Dorado-Díaz, P. I., Cruz-González, I., & Sánchez, P. L. (2023). Testing and Improvements of KoopaML: A Platform to Ease the Development of Machine Learning Pipelines in the Medical Domain. In F. J. García-Peñalvo & A. García-Holgado (Eds.), Proceedings TEEM 2022: Tenth International Conference on Technological Ecosystems for Enhancing Multiculturality. Salamanca, Spain, October 19–21, 2022 (pp. 210-217). Springer Nature. https://doi.org/10.1007/978-981-99-0942-1_21

or
s

Testing and Improvements of KoopaML: A Platform to Ease the Development of Machine Learning Pipelines in the Medical Domain

Andrea Vázquez-Ingelmo[1(✉)], Alicia García-Holgado[1], Francisco José García-Peñalvo[1], Rubén Fraile-Sanchón[1], Esther Andrés-Fraile[1], Pablo Pérez-Sánchez[2], Pablo Antúnez-Muiños[2,3], Antonio Sánchez-Puente[3], Víctor Vicente-Palacios[4], Pedro Ignacio Dorado-Díaz[2], Ignacio Cruz-González[2,3], and Pedro Luis Sánchez[2,3]

[1] GRIAL Research Group, University of Salamanca, Salamanca, Spain
{andreavazquez,aliciagh,fgarcia,rubenfs2000,esther.andres}@usal.es

[2] Biomedical Research Institute of Salamanca (IBSAL), Salamanca, Spain
{pperezsanc,pidorado}@saludcastillayleon.es

[3] CIBERCV, University Hospital of Salamanca, Salamanca, Spain
asanchezpu@saludcastillayleon.es

[4] Philips Ibérica, Madrid, Spain
victor.vicente.palacios@philips.com

Abstract. Machine Learning (ML) applications in complex domains, such as the medical domain, can be highly beneficial, but also hazardous if some concepts are overlooked. In this context, however, health professionals denote expertise in their domain, but they often lack skills in terms of ML. In this sense, to leverage ML applications in the medical domain, it is important to combine both domain expertise and ML-related skills. In previous works, we tackled this challenge in the health context through a visual platform (KoopaML) that enables lay users to build ML pipelines. The present work describes the challenges derived from the first version of the platform and the prototypes for the new features designed to address them. The prototypes have been validated by two experts, obtaining highly valuable feedback.

Keywords: Machine Learning · Human-Computer Interaction · Health · Artificial Intelligence · Medical data management

1 Introduction

Machine Learning (ML) is widely employed in several fields. However, the complexity behind applying this kind of approach should not be overlooked. Several domains benefit from the application of Machine Learning models, but this process can also be sensitive. One of these domains is the medical domain, where sensitive data is processed to ease complex tasks such as diagnoses, disease detection, segmentation, and assessment of organ functions, among others [1–3].

For these reasons, it is important to train physicians regarding ML applications to reach two goals; (1) to allow them to apply ML pipelines without the necessity of external Artificial Intelligence (AI) experts, and (2) to avoid wrong conclusions, losses, discrimination, and even negligence due to the misunderstanding of the ML models' outputs [4–7].

In previous works, we introduced KoopaML, a platform to ease the application of machine learning in the medical domain without the necessity of having programming skills [8–10]. The first version of the platform provided main features such as a graphical user interface to design and execute ML pipelines visually.

This paper presents the improvements and validations of the KoopaML platform based on new necessities derived from the user-centered approach followed for its development.

The rest of this paper is organized as follows. Section 2 provides an overview of the first version of KoopaML. Section 3 describes the methodology followed to gather new requirements and validate the platform. Section 4 presents the prototypes developed to explore the new features of KoopaML. Section 5 presents the results of the prototype evaluations. Finally, Sect. 6 discusses the results and presents the conclusions derived from this process.

2 First Version of the Platform

The first version of the platform provided a set of functionalities related to the definition and execution of ML pipelines, input data validation, visualization, and the possibility of defining heuristics to drive and ease the development of ML pipelines.

To create a new pipeline, the system provides an empty workspace with a toolbar containing the potential tasks that can be included in the ML workflow. These tasks are connected by sockets related to their inputs and outputs. These connections need to be compatible to work, i.e., if a task requires a dataset as an input, only those tasks that output a dataset can be connected. Figure 1 displays an ML pipeline in which a Random Forest classifier is trained with an input dataset with 6 variables.

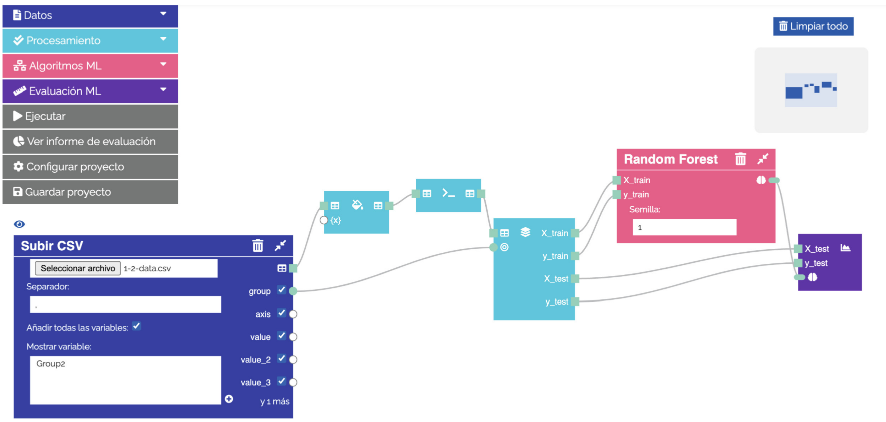

Fig. 1. Machine Learning pipeline example with KoopaML.

Once the pipeline is executed, the intermediate results from each step can be consulted individually, allowing better traceability of errors (if any) and other issues related to the resulting models.

KoopaML also provides the possibility of visually exploring the input dataset and the evaluation results of the models (Fig. 2). This feature takes advantage of the automatic generation of information dashboards through domain engineering and meta-modeling [11–14].

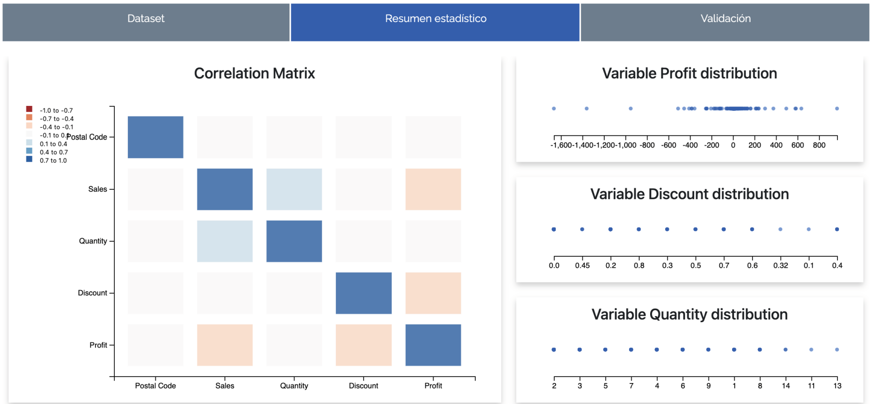

Fig. 2. A dashboard displaying summary statistics of an input dataset.

However, during the platform evaluation, users detected some features that could be included in the platform to improve the user experience and to extend the application of AI to DICOM images [15]. These new features have been explored and evaluated through prototypes before including them in production.

3 Methodology

The two prototypes with KoopaML were developed with Adobe XD. These prototypes were evaluated by two experts within the medical domain. Specifically, the links to the Adobe XD prototypes were shared with two medicine Ph.D. students who specialized in using AI to predict cardiological conditions.

Both experts received instructions to navigate through the prototypes and write down any thoughts, issues, or recommendations regarding the design and functionalities provided by the tool.

These comments have been analyzed to refine and improve the prototypes before the introduction of the new features in the current version of KoopaML.

4 Prototypes

Two different prototypes have been developed to explore new KoopaML features. One of the prototypes focuses on the support of DICOM images as input data for AI algorithms.

The other prototype focuses on improving user experience by displaying recommendations during the creation of ML pipelines, helping non-expert users to choose the best configuration for their data, as well as providing a learning experience while developing them.

Regarding the first prototype, a new workspace was devised. Separating the ML workspace from the DICOM workspace was crucial to avoid a cluttered interface. Also, the substantial differences between managing structured data and DICOM data require different treatments for these two kinds of inputs.

For these reasons, when creating a new project, users can choose between creating a new ML pipeline or a new DICOM project. In this case, it is necessary to add a configuration section for configuring picture archiving and communication systems (PACS) to enable KoopaML to request images from the configured servers.

When a new DICOM project is created, users are asked to select which images from the PACS they want to retrieve. Finally, different algorithms can be applied to the DICOM images depending on their modality. These algorithms are uploaded into the platform by authorized AI experts, making them accessible to non-expert users (Fig. 3).

In addition, the DICOM workspace also allows users to modify images using different edition tools to measure, annotate, crop, zoom, pan, and segmentate, among others.

Regarding the second prototype, the initial workspace of the platform was improved with new features, such as the "recommendation mode." This modality yields recommendations based on the current configuration of the workflow and the potential nodes that can be included to train the best model (Fig. 4). These recommendations will be driven by expert advice and heuristics.

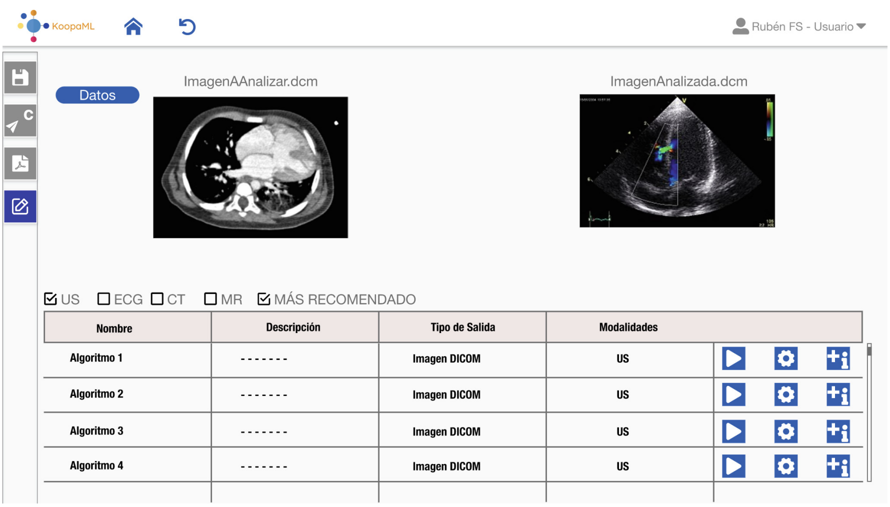

Fig. 3. Application of AI algorithms to DICOM images using a graphical interface.

On the other hand, the workspace has been redesigned to improve user experience. In this sense, the configuration tools (clear the workspace, execute the project, save the project, etc.) have been moved to the bottom part of the screen. This way, there is a

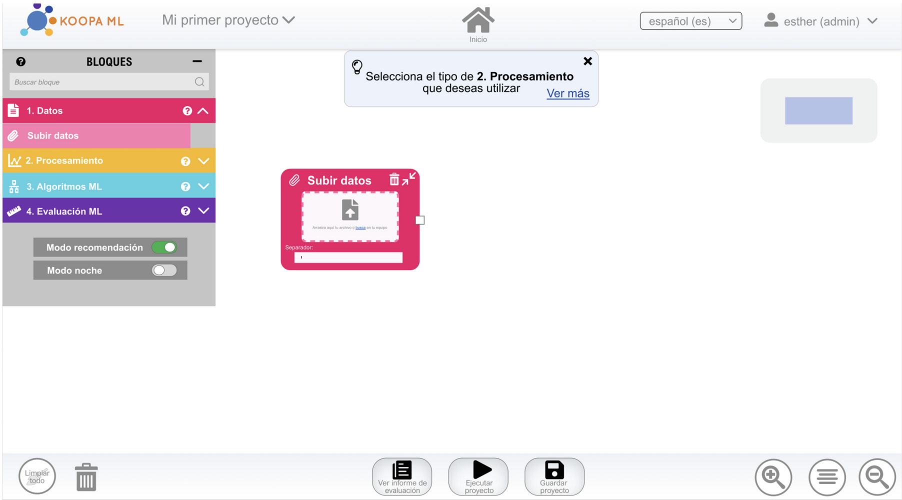

Fig. 4. Prototype for the new version of the KoopaML workspace.

differentiation between the configuration tools and the ML tasks that can be introduced into the pipeline.

5 Prototype Evaluation Results

5.1 DICOM Prototype

The following feedback was obtained for the prototyped DICOM functionalities. In general, both experts highlighted the simplicity and intuitiveness of the platform and the fact that these new features could be very useful for them.

Expert #1

- "PACS are easy to configure, and the configuration parameters are correct and sufficient"
- "DICOM images are easy to search by using patient IDs, studies, or existing projects"
- "DICOM edition tools seem good. Maybe the segmentation tool won't be employed by experts because they use other tools"
- "The interface to execute AI models is simple and intuitive"

Expert #2

- "Intuitive and easy-to-use platform, even for non-expert users"
- "It would be helpful to have some information and recommendations related to the available AI algorithms before applying them to the DICOM images"
- "It is likely that the segmentation tool won't be used. Maybe it would be interesting to upload already segmented images"

5.2 Redesign and Recommendation Mode

Regarding the new design of the ML workspace and the recommendation mode, both experts pointed out that yielding recommendations while developing the ML pipelines could be extremely useful for non-experts.

Expert #1

- "I think the new workspace design is better. I like the differentiation between the configuration tools and the ML nodes, it is very intuitive and understandable"
- "I think that recommendations are helpful for non-expert users"
- "I like all the improvements; the new placement and shape of the buttons are very intuitive"

Expert #2

- "I have doubts about the data formats to be accepted. Excel spreadsheets are widely used, but on some occasions, other databases are employed."
- "In addition, I don't know if it could be useful to allow the definition of datasets in the platform itself by filling forms. Anyway, I understand that it may not make sense to integrate it here on this platform"
- "As for the processing or coding of the data, I am not clear when it is necessary. Sometimes the data will be uploaded already coded and cleaned. I wonder if it is possible to already use that coding and not duplicate the work."
- "As for the recommendation option for the project, I find it very interesting; even some help or brief description of the different options would be nice"
- "Finally, once the project is completed (with its trained algorithm and ML evaluation), it may be interesting to add new datasets to the sample. I am considering the possibility of adding new data to the previously used data."

6 Discussion and Conclusions

This work presents the continuous improvement of a platform for creating AI-driven pipelines in the medical domain, namely KoopaML. The first version of the platform obtained good feedback [8–10], but some new features were required to support other complex tasks in the medical context.

In this sense, two prototypes were developed to evaluate and explore these new features and to make design decisions before including them in production. The prototypes were evaluated by two experts in the medical domain.

The feedback was positive and useful. The two experts pointed out that both prototypes were intuitive, simple, and easy to use, which were crucial goals because the platform aims to promote the use of AI among non-expert users.

Both experts also indicated that the segmentation tool could not be as useful as we thought because physicians and researchers usually employ their own tools to perform this task. However, it could be useful to support segmentation for other goals, like correcting the segmentation derived from applying an AI model to DICOM images.

Regarding the negative feedback, it is mostly related to the misunderstanding of some ML nodes, such as the "Encoding" node. This node is not mandatory, so the already-codified dataset could be employed without the necessity of including the encoding task in the ML pipeline.

In fact, this kind of issue can be avoided by improving the information related to the goal, inputs, and outputs of each ML node, which is another goal of the ML workspace redesign.

This work has provided enough feedback to refine the features and new designs before implementing the improvements in a new version of KoopaML. Future research lines will involve the evaluation of the implemented features, as well as user evaluations related to the usefulness of the displayed heuristics and recommendations during the development of ML pipelines (comparing experts with non-experts).

Acknowledgments. This research work has been supported by the Spanish Ministry of Education and Vocational Training under an FPU fellowship (FPU17/03276). This research was partially funded by the Spanish Government Ministry of Economy and Competitiveness through the DEFINES project grant number (TIN2016-80172-R) and the Ministry of Science and Innovation through the AVisSA project grant number (PID2020-118345RB-I00). 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

1. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
2. González Izard, S., Sánchez Torres, R., Alonso Plaza, Ó., Juanes Méndez, J.A., García-Peñalvo, F.J.: NextMed: automatic imaging segmentation, 3D reconstruction, and 3D model visualization platform using augmented and virtual reality. Sensors (Basel) **20**, 2962 (2020)
3. Izard, S.G., Juanes, J.A., García Peñalvo, F.J., Estella, J.M.G., Ledesma, M.J.S., Ruisoto, P.: Virtual reality as an educational and training tool for medicine. J. Med. Syst. **42**(3), 1–5 (2018). https://doi.org/10.1007/s10916-018-0900-2
4. Weyerer, J.C., Langer, P.F.: Garbage in, garbage out: the vicious cycle of ai-based discrimination in the public sector. In: Proceedings of the 20th Annual International Conference on Digital Government Research, pp. 509–511
5. Ferrer, X., van Nuenen, T., Such, J.M., Coté, M., Criado, N.: Bias and discrimination in AI: a cross-disciplinary perspective. IEEE Technol. Soc. Mag. **40**, 72–80 (2021)
6. Hoffman, S.: The emerging hazard of AI-related health care discrimination. Hastings Cent. Rep. **51**, 8–9 (2021)
7. Wachter, S., Mittelstadt, B., Russell, C.: Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI. Comput. Law Secur. Rev. **41**, 105567 (2021)
8. García-Holgado, A., et al.: User-centered design approach for a machine learning platform for medical purpose. In: Ruiz, P.H., Agredo-Delgado, V., Kawamoto, A.L.S. (eds.) HCI-COLLAB 2021. CCIS, vol. 1478, pp. 237–249. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-92325-9_18

9. Vázquez-Ingelmo, A., et al.: Bringing machine learning closer to non-experts: proposal of a user-friendly machine learning tool in the healthcare domain. In: Proceedings TEEM 2021. Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality, Barcelona, Spain, 27th–29th October 2021, pp. 324–329. Association for Computing Machinery, New York
10. García-Peñalvo, F.J., et al.: KoopaML: a graphical platform for building machine learning pipelines adapted to health professionals. Int. J. Interact. Multimed. Artif. Intell. (in press)
11. Vázquez-Ingelmo, A., García-Holgado, A., García-Peñalvo, F.J., Therón, R.: Proof-of-concept of an information visualization classification approach based on their fine-grained features. Expert Syst. e12872 (2022)
12. Vázquez-Ingelmo, A., García-Peñalvo, F.J., Therón, R.: Towards a technological ecosystem to provide information dashboards as a service: a dynamic proposal for supplying dashboards adapted to specific scenarios. Appl. Sci. **11** (2021)
13. Vázquez-Ingelmo, A., García-Peñalvo, F.J., Therón, R., Amo Filvà, D., Fonseca Escudero, D.: Connecting domain-specific features to source code: towards the automatization of dashboard generation. Clust. Comput. **23**(3), 1803–1816 (2019). https://doi.org/10.1007/s10586-019-03012-1
14. Vázquez-Ingelmo, A., García-Peñalvo, F.J., Therón, R.: Taking advantage of the software product line paradigm to generate customized user interfaces for decision-making processes: a case study on university employability. PeerJ Comput. Sci. **5** (2019)
15. Mildenberger, P., Eichelberg, M., Martin, E.: Introduction to the DICOM standard. Eur. Radiol. **12**(4), 920–927 (2001). https://doi.org/10.1007/s003300101100