

Implementation of eXplainable Artificial Intelligence

Case Study on the Assessment of Movements to Support Neuromotor Rehabilitation

Luiz Felipe de Camargo[1(✉)], Diego Roberto Colombo Dias[2], and José Remo Ferreira Brega[1]

[1] Universidade Estadual Paulista "Júlio de Mesquita Filho", Bauru, SP, Brazil
{luiz.felipe,remo.brega}@unesp.br
[2] Universidade Federal de São João del-Rei, São João del-Rei, Brazil
diegodias@ufsj.edu.br

Abstract. Solutions based on Artificial Intelligence are being used to solve problems in various domains. However, many people feel uncomfortable with this type of solution because they must understand how it works. In the face of this, the so-called eXplainable Artificial Intelligence arises, seeking not only to provide the answers produced by Artificial Intelligence but also to offer aspects of explainability, detailing the decision process and generating confidence. In this context, a literature review on eXplainable Artificial Intelligence has presented a brief comparative study between the most popular libraries for this implementation and a deepening of the theme of explainability evaluation and the comprehension process. A proposal for the implementation and evaluation of eXplainable Artificial Intelligence in the context of movement classification to support neuromotor rehabilitation was built from the results obtained. The first experiments performed showed to be promising. The proposal is expected to be relevant for addressing a growing theme in a context, the health area, that demands explicability and transparency in decisions.

Keywords: artificial intelligence · explainable artificial intelligence · implementation · rehabilitation · motion assessment

1 Introduction

Artificial Intelligence (AI) can perform the most diverse activities, almost always superior and more reliable than human beings. However, the world press spreads information that often leads the population to fear this type of technology, imagining that there will be a significant reduction in the number of jobs available. This dissemination of incorrect information about the use of AI hampers the adoption of this type of technology, ignoring that there will not necessarily be a reduction in jobs. However, a change like the work to be performed, often going

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023, LNCS 13956, pp. 564–580, 2023.
https://doi.org/10.1007/978-3-031-36805-9_37

from hard work, with significant physical wear and tear, to equipment control work, based on intellectual effort and requiring greater prior training [1].

The process of using AI almost always consists of inputting data into a computational model in order to train it to solve a particular problem. This model learns the patterns from this training data and can identify and predict them in other masses of data, identifying situations and making recommendations. However, the internal structure of the model, which leads to learning and making decisions, is often complex and challenging to understand, especially for the average user [14].

This complexity, added to the stance of media on the subject, generates great fear and discomfort for average users, creating barriers to adopting this type of technology. Life-threatening situations, such as using AI for medical diagnosis, tend to be even more frightening. That is the current situation of this technology, AI can make decisions superior many times to humans, but it cannot explain and justify its decision.

In this scenario, the concept of eXplainable Artificial Intelligence (XAI) arises, seeking to make AI more accessible, producing, in addition to the expected result, mechanisms that aid in understanding the results and the path taken to them. The first initiative to develop XAI came from the Defense Advanced Research Projects Agency (DARPA), an American governmental research agency focused on defense projects, where the foundations of the Internet were also developed. However, several other initiatives exploring XAI have emerged in recent years, including from giants such as Google [14,16].

The main objective of this article is to present a Systematic Literature Review (SLR) that seeks to explore the research field of XAI implementation and thereby obtain information about the implementation and application of XAI techniques to improve the understanding of users of the operation of AI methods. From the review, develop a proposal for implementing XAI in evaluating the movement to support neuromotor rehabilitation.

2 Background

2.1 Artificial Intelligence

The term was coined in the 1950s s by professor and researcher John McCarthy. He defined it as: “the science and engineering of making intelligent machines, especially intelligent computer programs” [9].

Several researchers in the area discuss the definition of AI, and with this, a definition considered quite enlightening is the one by authors Rich and Knight: “AI is the study of how to make computers perform tasks that, so far, men do better” [12]. Author Laurière adds: “every problem for which no algorithmic solution is known is an AI problem” [6].

2.2 eXplainable Artificial Intelligence

XAI is a strand of AI that aims to explain its purpose and provide results. This explanation should include its purpose and decision process, widely recognized

as a crucial feature for the practical implementation of AI models [2]. Discussions about XAI are becoming increasingly present in the technology world.

When the model is considered complete, without the need to improve it through explainability, and it is not necessary to question its functioning, understanding the problem can be considered complete. If this is the situation, the recommendation would be not to use machine learning since it should be used when it is impossible to define an algorithm with specific rules to address the problem [8].

3 Systematic Literature Review

This Section presents the methods and data about the review process performed in this study, describing mainly the RSL method.

3.1 Review Objectives and Research Questions

The executed review aims to get an overview of the implementation of XAI techniques and, through this, to deeply understand the research field.

After several prior discussions and research, the main research question was: How does one implement XAI? This main question can be expanded into the following questions: What are the existing XAI solutions (techniques, methods, or frameworks)? What are the main areas that XAI serves? How does XAI integrate with existing AI platforms? What visual capabilities are used by these XAI solutions? How is the XAI research area evolving?

3.2 Search Strategy

For the present work, the search string used was: (explainable AND ("artificial intelligence" OR "machine learning" OR "deep learning")) AND (framework OR technique) AND (trust OR interpretability). The initial searches were performed using search strings in ACM Digital Library, IEEE Digital Library, ISI Web of Science, Science@Direct and Scopus.

3.3 Inclusion and Exclusion Criteria

For the present review, the following criteria were defined:

- Inclusion - "Affinity of the study to the desired topics" and "Presents a practical application";
- Exclusion - "Study is not available in full in the English language", "Study has no access granted to the institution", "Lack of affinity of the study with the desired topics", "It is not an article", "It does not deal with implementation, being limited only to XAI concepts", "Review or survey on the subject" and "It deals with a tool already analyzed".

3.4 Data Extraction

For the present study, the following data composed the extraction form:

- Name of the technique, method, or framework;
- What is the objective?;
- AI technique developed with explainability or further explanation?;
- Which methods or AI platforms are supported?;
- What is the visualization technique employed?; and
- Has there been an evaluation or validation study with users?

3.5 Analysis and Discussion of Results

An overview of the review and its steps can be seen in Fig. 1.

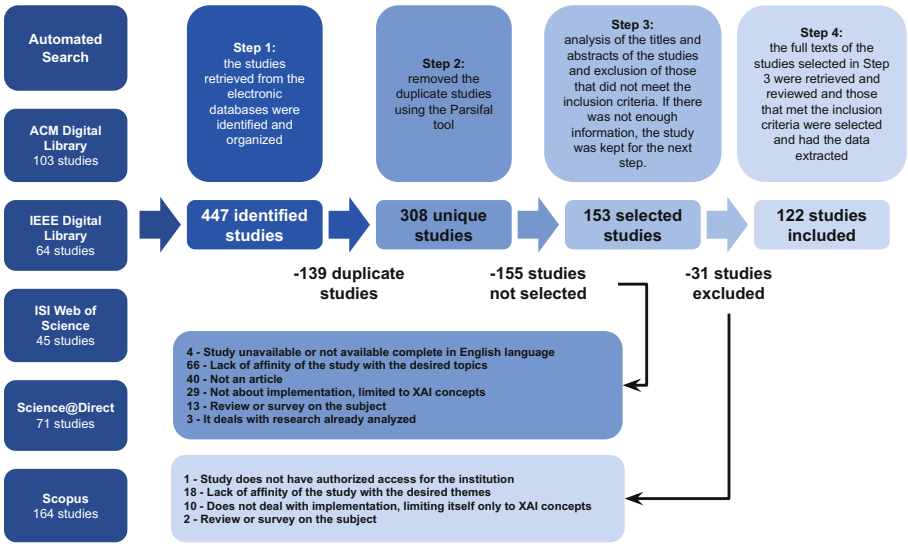

Fig. 1. Review steps.

Year of Studies. Figure 2 shows the increase in the number of studies dealing with the subject of the present review over the past five years, demonstrating the growing relevance of the subject, considering that the year 2021 was not fully analyzed.

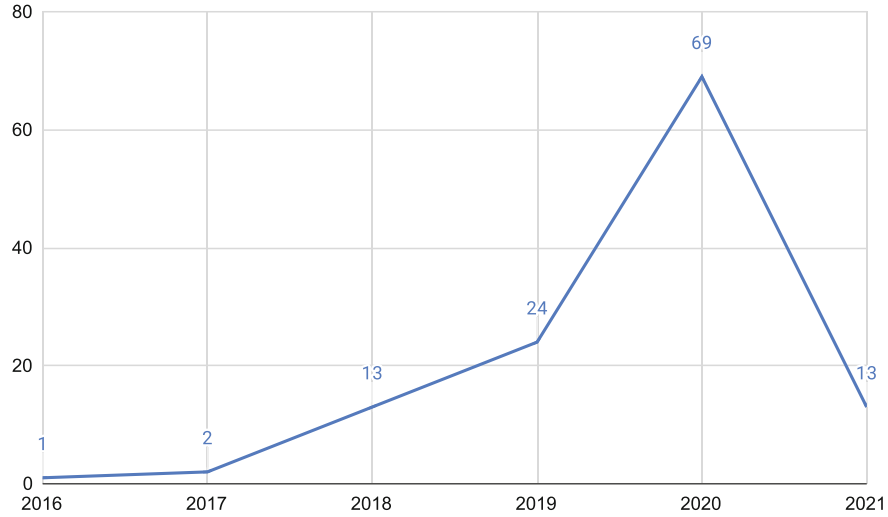

Fig. 2. Evolution of the amount of studies over the years.

Study Objectives. Through question 2, "What is the objective?", we sought to identify the main motivation that led to the development of each study. Studies with similar objectives were grouped into categories to organize the data better. Figure 3 represents the division by objective category, with the less relevant categories (1 study only) being grouped as "Other".

Explanation Generation Moment. Through question 3, it was sought to identify the moment of explanation generation, which can be a posterior explanation, using model input and output data, or during the execution of the AI model, integrating to the model and making it transparent. The number of studies that demonstrated using posterior explanation (63) and the number of studies that combined the AI model with explainability (60) are pretty similar, showing a balance between the two approaches. This balance can also be visualized in Fig. 4.

Artificial Intelligence Models. Question 4 sought to identify the families of AI models with which the developed explainability technique is compatible. The most diverse AI models were cited in the studies; thus they were grouped into broader families, which aggregate several more specific models.

In Fig. 5 the cited AI model families can be compared.

The use of neural networks stands out, cited in 64 studies, currently widely used models and tend to be complex, especially in the so-called deep neural networks, which have several layers, thus increasing the need for explainability. The tools that propose to be model-agnostic were identified in 22 studies.

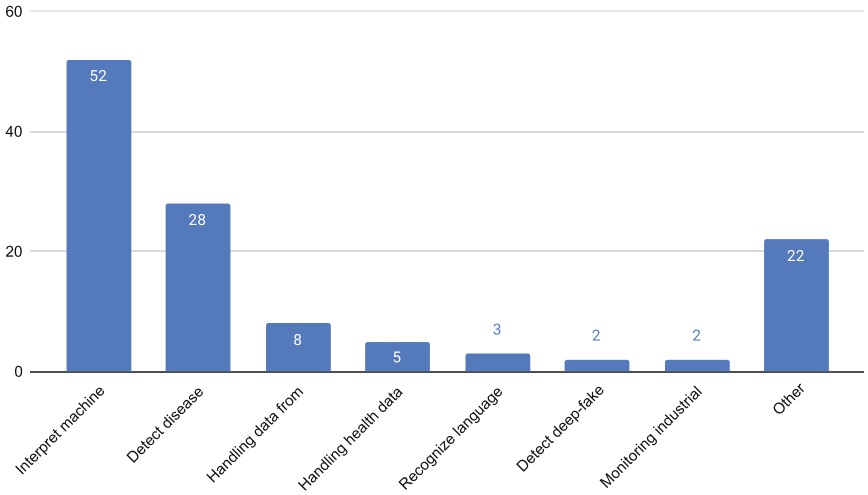

Fig. 3. Categories of study objectives.

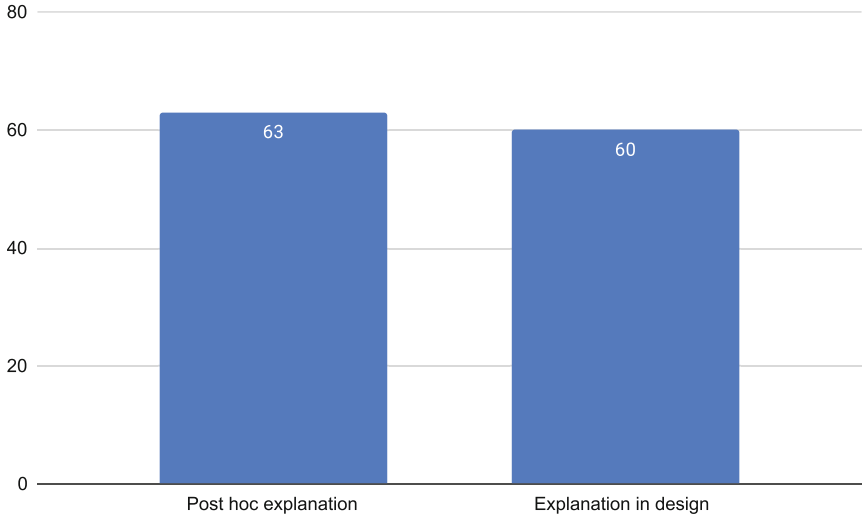

Fig. 4. Studies separated by a moment of explainability generation.

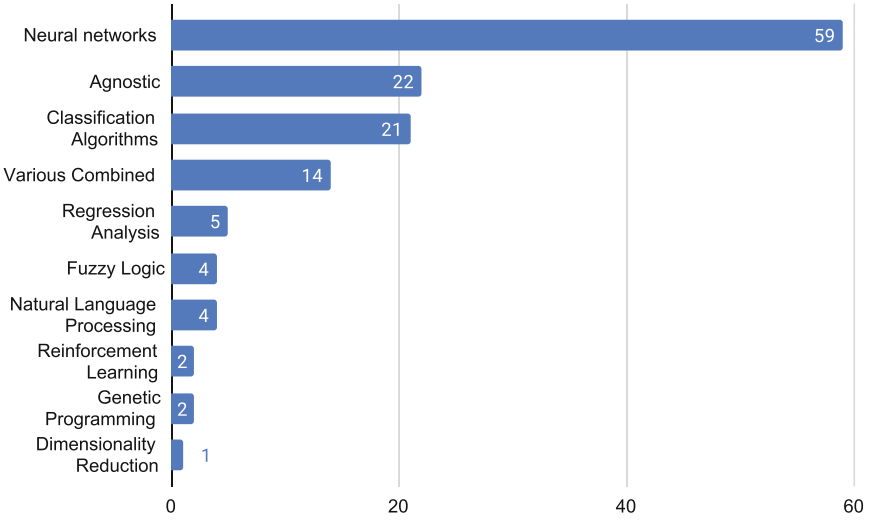

Fig. 5. Families of AI models.

These tools work using input and output data without the need to know the structure of a specific model. Next are decision-tree-based models with 14 studies and the combination of several different models, cited in 11 studies..

Support Libraries. Several studies cited the use of existing libraries for XAI implementation, such as LIME [11] and SHAP [7]. In Table 1, the quantities of studies citing involvement with LIME and SHAP in various ways can be seen.

Table 1. Using the LIME and SHAP libraries.

Frameworks	Studies	%
Uses Lime	16	13,11
Uses SHAP	11	9,01
Compare with Lime	8	6,55
Compare with SHAP	3	2,46
Uses Lime and SHAP	2	1,64
Compare with Lime and SHAP	1	0,81
Total	41	33,60

It can be observed that the relevant quantity (41 studies, 33.60%) of researchers that use somehow the two most relevant libraries in the implementation of XAI to obtain a speed gain by using already implemented solutions

instead of choosing to perform the full development without any support. Model-independent methods, such as SHAP and LIME, are widely used, as they offer explainability in the most diverse situations, be it for processing tabular data, text, and images, including deep learning models [2].

3.6 Information Visualization

Through question 5, it was sought to identify which visualization technique was used. To help identify the techniques, a taxonomy created by [5] was used, which classifies visualization techniques into the following families:

- Comparing categories
- Assessing hierarchies and part-to-whole relationships
- Show changes over time;
- Plotting connections and relationships; and
- Mapping geo-spatial data.

Besides the categories based on the taxonomy presented, two other categories were created: "Others", indicating differentiated techniques, often created together with the tool and which are not classified within the taxonomy; and "Does not use directly", where the studies that do not use Information Visualization in the process of generating explainability were allocated.

Once the categories were defined, the studies were distributed as follows, according to Fig. 6.

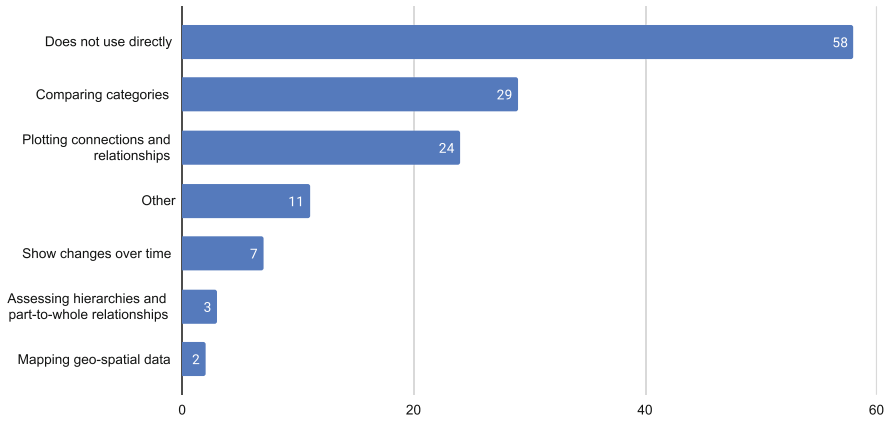

Fig. 6. Families of Visualization techniques.

It can be observed that most of the studies (58) do not directly use visualization techniques during the explanation generation process, among the studies that do use visual techniques, techniques that seek to compare categories (29) and demonstrate connections and relationships (24) can be highlighted.

3.7 Evaluation or Validation with Users

According to question 6, we tried to identify if the analyzed study described the execution of an evaluation or validation process of the proposed explainability technique with users through tests, interviews, or evaluation forms, thus seeking to guarantee the positive results of the proposal regarding ease of understanding and use. In Fig. 7, the results of this question can be observed.

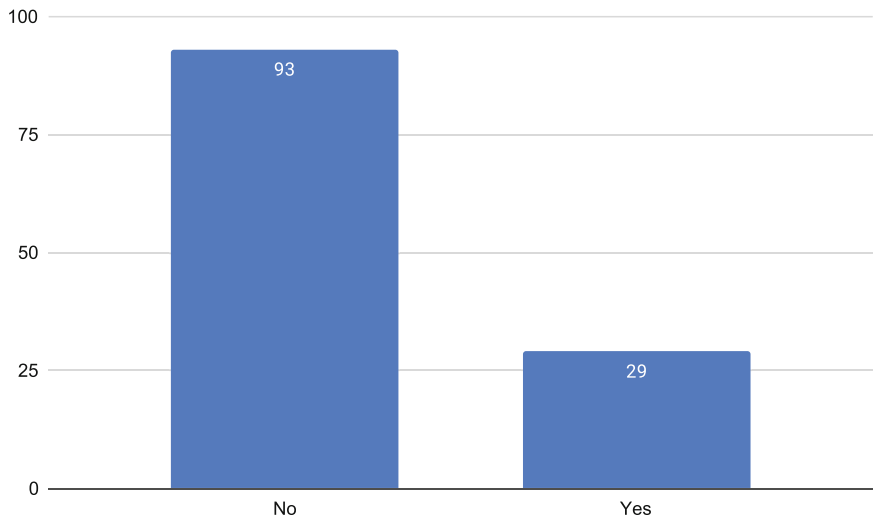

Fig. 7. Evaluation or validation with users.

It can be seen that most of the studies (93) did not seek to perform an evaluation or validation with real users, limiting themselves to tests with developers and simulations. Although the user testing process is quite labor-intensive, the lack of it weakens the proposed conclusions. However, studies (29) proposed to perform tests, usually with specialists in the context of applying the tools, but also with a lay audience, seeking to assess the level of understanding achieved.

3.8 Discussion

Next, the results obtained are discussed to extract new information and trends from them, exposing the review's relevance and limitations. In order to obtain information that composes an answer to the main research question presented, "How does one implement XAI?", the secondary research questions will be addressed, presenting information obtained that represents an answer for each of them.

What Are the Existing eXplainable Artificial Intelligence Solutions (techniques, Methods or Frameworks)? It was possible to ascertain many techniques based on libraries such as LIME and SHAP, but with different approaches and methods. The use of libraries for support indicates an interesting starting point for the continuation of the current project, seeking to understand how a working standard can be defined in the research area, which is still very recent.

What Are the Main Areas that eXplainable Artificial Intelligence Serves? A good part of the analyzed studies demonstrates its objective, such as improving interpretability issues, without depending on a closed context, showing that XAI tends to evolve as a tool to be applied in the most diverse areas. A point of emphasis is the application of XAI in Healthcare, especially disease detection, since this area deals with data and prudent decisions to be made, highlighting the importance of explainability in decisions made in an automated way. Several other areas were cited in the other studies, showing that the application of XAI has been expanding to other fields.

How Does eXplainable Artificial Intelligence Integrate with Existing Artificial Intelligence Platforms? Regarding integrating XAI tools into AI solutions, the two ways of integration (explainability generated during the AI process and explainability afterward) were equally cited, indicating that both approaches are relevant, presenting advantages and disadvantages but being appropriate depending on the situation to be addressed.

What Visual Resources Are Used by These eXplainable Artificial Intelligence Solutions? Analyzing the use of visual resources and Information Visualization, most studies do not directly use these resources. In the analyzed articles, graphs are often presented to expose the results, but graphical tools are not used during the explanation process.

How Is the Research Area of eXplainable Artificial Intelligence Evolving? Through the information of the studies' publication years, it can be seen that they deal with a new area, still in expansion, which has been gaining strength over the years. However, it was found that many studies need to present the evaluation of techniques and tools with real users, restricting themselves to simulated tests. That shows that there is room for evolution, even though the evaluation process with users is a complex process that demands considerable effort.

Relevance and Limitations of the Study. The present study is relevant because of the number of studies analyzed (122), obtained after several previous searches with various search strings to perform a comprehensive search that

returned studies relevant to the topic in question. As a limitation, we can cite the restriction of the automated search method used, with no recourse to manual searches or other search forms. Another limitation is the need for clear information in many studies, not informing the names of the tools developed, among other missing or unclear information.

4 Proposal

In this section, the XAI application proposal that composes this work is presented, initially presenting the context where the proposal is inserted, then detailing the AI solutions that are being studied for use, and finally, the proposal for continuity addressing XAI implementation.

4.1 Scenario

The scenario where explainability will be applied is the estimation and evaluation of movements obtained through body tracking to use data in rehabilitating patients, especially those who have suffered a stroke. The authors [13] have been working in this context with the Brazilian Institute for Neuroscience and Neurotechnology - CEPID BRAINN, using AI in motion classification, but still without involving aspects of XAI. Thus, one of the goals of this work is to add this approach to collaborate in the development of the solution and add explainability at various points, providing new information for different audiences, including users of the solution, domain experts, AI developers, thus obtaining very comprehensive aspects of explainability, with several possibilities.

4.2 Tool Structure

Several authors sought to apply AI in motion estimation, classification, and evaluation. In the work of [3], the Pose Trainer tool was proposed, using convolutional neural networks and heuristic techniques to detect a user's pose and then evaluate the vector geometry of the pose. Pose Trainer runs on four common exercises and supports any Windows or Linux computer with a GPU. [4] developed Pose Tutor, an AI-based explainable pose recognition and correction system. This system uses several AI models to estimate the pose, classify the motion and indicate by explainability tools the points to be corrected in the pose. For pose estimation, the tool uses the TransPose model, based on a neural network solution with Transformers [18]. Pose Tutor shows good results given pose data from Yoga, Pilates, and Kung Fu.

Based on these two studies, we propose creating a tool that can estimate poses in rehabilitation exercises and indicate through explainability the points to be corrected in the pose, to improve the estimation, thus improving the quality of the exercise. As a proof of concept, we initially sought to implement separately two AI solutions, one for pose estimation and the other for evaluation of the positions of key points, with explainability.

4.3 Pose Estimation

For the pose estimation studies, experiments were performed with the solutions proposed by [18] and [10]. Since the goal is to estimate using videos, it was decided to continue the study using the VideoPose3D solution by [10]. It does pose estimation on video data through the Detectron2 library, which uses a convolutional neural network model for several visual recognition tasks, among them pose estimation [17]. Both solutions, VideoPose3D and Detectron2, were developed and are maintained by the Meta Research team.

In Figs. 8 and 9, results of the VideoPose3D application on two selected videos can be seen, thus obtaining the joint position estimation, pose, and 3D reconstruction of the pose.

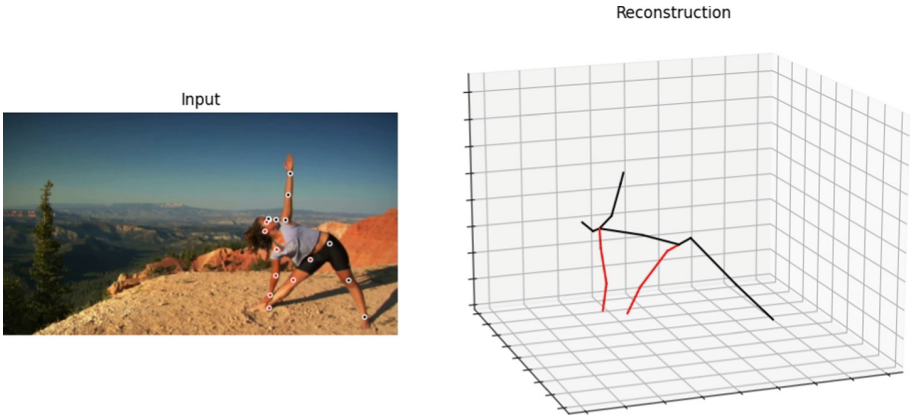

Fig. 8. Example of testing the 3D pose estimation model.

4.4 Motion Evaluation

Studies were started on measuring the angles of the joints detected in a pose to use this information to evaluate the movement quality and develop the motion evaluation solution. Tests were conducted with a prototype based on OpenCV and MediaPipe technologies, used respectively for computer vision, and detection and tracking in video inputs [15].

The proof of concept created allows the input of video information, and after pose estimation, it provides the value of the flexion angle of the trunk of an individual, as can be seen in Figs. 10 and 11.

4.5 Use of Explainability

Considering the use of explainability as a tool to support the end user of the application, as in the examples presented in the works of [3] and [4], the goal

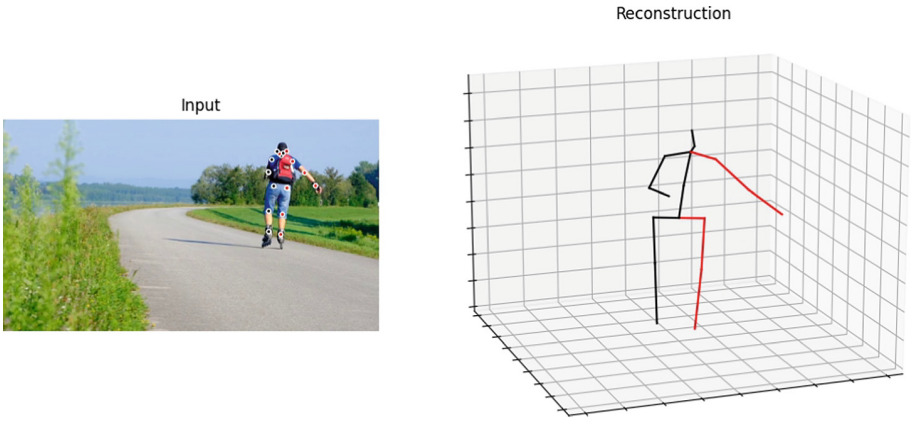

Fig. 9. Example of testing the 3D pose estimation model.



Fig. 10. Example of angle evaluation model test.

of the proposal is to use explainability as a tool to evaluate the movements presented in rehabilitation activities, allowing the evaluation of the movement, comparing it with the proposal of the health professional.

For the development and implementation of this solution, the solutions of pose estimation and angle evaluations will be combined with aspects of explainability, seeking to obtain visual feedback with the evaluation of the executed movement. Figures 12 and 13 depict tests performed with the concept tests presented in a video showing footage of a rehabilitation session.

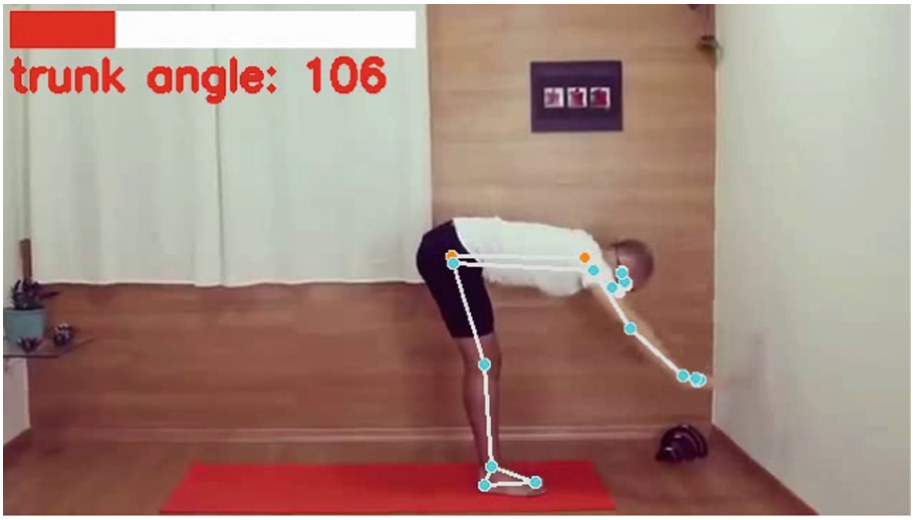

Fig. 11. Example of angle evaluation model test.

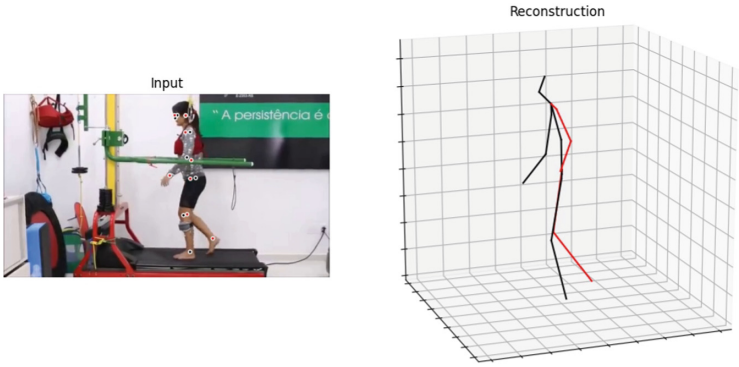

Fig. 12. Application of the proof of concept of pose estimation in rehabilitation imaging.

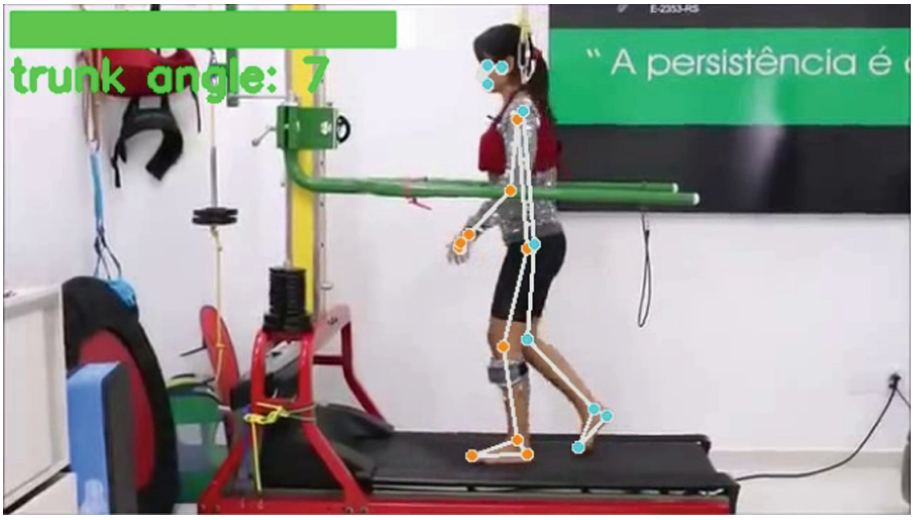

Fig. 13. Application of the proof of concept of angle evaluation in rehabilitation imaging.

5 Conclusion

In the present section, seeking to conclude this work, a brief discussion will be presented regarding the relevant points and the limitations of the present study.

5.1 Relevance

This study is relevant because it deals with a new area, XAI, which has been gaining strength in recent years, along with discussions about the fair use of AI in classification and decision processes. Collaborating with this relevance is the case study in the health area, one area that most demand an increase in confidence in computer-aided decisions.

Another point that contributes to the relevance of this work is its consistency, performing a literature review and proposing a case study for implementing XAI.

5.2 Limitations

Since XAI is a new field of research, there is the limitation that the implementation process needs to be better defined in the available literature, requiring a deepening of this theme and empirically seeking positive results.

The possibilities of the application of XAI are enormous. However, we focused efforts on the case study involving the evaluation of movements to aid neuromotor rehabilitation without seeking new areas for the development of practical applications.

Acknowledgements. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

References

1. Abuselidze, G., Mamaladze, L.: The impact of artificial intelligence on employment before and during pandemic: a comparative analysis. J. Phys. Conf. Ser. **1840**(1), 012040 (2021). https://doi.org/10.1088/1742-6596/1840/1/012040. https://iopscience.iop.org/article/10.1088/1742-6596/1840/1/012040
2. Barredo Arrieta, A., et al.: Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fus. **58**, 82–115 (2020). https://doi.org/10.1016/j.inffus.2019.12.012. https://www.sciencedirect.com/science/article/pii/S1566253519308103
3. Chen, S., Yang, R.R.: Pose trainer: Correcting exercise posture using pose estimation. arXiv; Computer Vision and Pattern Recognition, June 2020. http://arxiv.org/abs/2006.11718
4. Dittakavi, B., et al.: Pose tutor: an explainable system for pose correction in the wild. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3539–3548. IEEE, June 2022. https://doi.org/10.1109/CVPRW56347.2022.00398
5. Kirk, A.: Data Visualization: A Successful Design Process; A Structured Design Approach to Equip You with the Knowledge of how to Successfully Accomplish Any Data Visualization Challenge Efficiently and Effectively. Packt Publishing (2012)
6. Lauriere, J.L.: Problem-Solving and Artificial Intelligence. Prentice Hall (1990)
7. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. arXiv 1(Section 2), 1–10 (2017)
8. Masís, S.: Interpretable Machine Learning with Python: Learn to Build Interpretable High-Performance Models with Hands-on Real-world Examples. Packt Publishing (2021)
9. Mccarthy, J.: What is artificial intelligence? (2007). http://www-formal.stanford.edu/jmc/
10. Pavllo, D., Feichtenhofer, C., Grangier, D., Auli, M.: 3d human pose estimation in video with temporal convolutions and semi-supervised training. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7745–7754. IEEE (June 2019). https://doi.org/10.1109/CVPR.2019.00794
11. Ribeiro, M.T., Singh, S., Guestrin, C.: "why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13–17-August, pp. 1135–1144 (2016). https://doi.org/10.1145/2939672.2939778
12. Rich, E., Knight, K.: Inteligencia Artificial. McGraw-Hill (1994)
13. Rodrigues, L.G.S., et al.: Classification of human movements with motion capture data in a motor rehabilitation context. In: Anais do XXIII Simpósio de Realidade Virtual e Aumentada, pp. 55–62. SBC, Porto Alegre, RS, Brasil (2021). https://sol.sbc.org.br/index.php/svr/article/view/17519
14. Sarkar, T.: Google's new "explainable AI" (XAI) service - towards data science (2019). https://towardsdatascience.com/googles-new-explainable-ai-xai-service-83a7bc823773

15. Editorial Team: Detecting bad posture with machine learning, January 2023. https://towardsai.net/p/machine-learning/detecting-bad-posture-with-machine-learning
16. Turek, M.: Explainable Artificial Intelligence (2018). https://www.darpa.mil/program/explainable-artificial-intelligence
17. Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: Detectron2 (2019)
18. Yang, S., Quan, Z., Nie, M., Yang, W.: TransPose: keypoint localization via transformer. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 11782–11792, December 2020. http://arxiv.org/abs/2012.14214