XH-KaaS (eXplanable Health-Knowledge as a Service)

Thiago C. Montenegro[1], **Natasha C. Q. Lino**[2]

[1]Centro de Informática – Universidade Federal da Paraíba (UFPB) – João Pessoa – PB – Brazil

thiago.montenegro@academico.ufpb.br, natasha@ci.ufpb.br

Abstract. *Clinical Decision Support Systems (CDSS) and Artificial Intelligence techniques, especially machine learning, are highly effective and precise in healthcare. However, their lack of transparency and interpretability poses significant challenges. To overcome this issue, the article proposes a knowledge-as-a-service architecture in healthcare. This approach centralizes services, incorporates explainability techniques, and establishes reference architectures to minimize risks associated with a lack of transparency.*

Resumo. *Os Sistemas de Suporte à Decisão Clínica (SSDC) e as técnicas de inteligência artificial, especialmente o aprendizado de máquina, tornaram-se verdadeiros aliados devido à sua precisão e eficácia. No entanto, a falta de transparência e interpretabilidade desses sistemas representa desafios para sua aplicação prática. Para mitigar tal problemática, o artigo propõe uma arquitetura de conhecimento como serviço ao domínio da saúde. Essa abordagem busca centralizar serviços e incorporar técnicas de explicabilidade, visando aprimorar a compreensão do processo decisório dos modelos de aprendizado de máquina pelos usuários e estabelecer arquiteturas de referência que minimizem os riscos associados à falta de transparência.*

1. Introdução

Nas últimas décadas testemunhamos o avanço da Inteligência Artificial (IA). Tal tecnologia teve impacto significativo em diversas áreas do conhecimento, desde o marketing à área da saúde, proporcionando melhorias e métodos de apoio à decisão. O alto volume de dados em conjunto com o contínuo desenvolvimento de hardware, tem atuado como verdadeiros catalisadores para o avanço da IA.

No âmbito da saúde, as aplicações impulsionadas pela IA, especificamente por técnicas de aprendizado de máquina, são consideradas como soluções promissoras, devido à sua notável precisão, eficácia e capacidade de escalabilidade [Merjulah and Chandra 2019], auxiliando no processo de suporte à decisão clínica proporcionando os cuidados à saúde e bem estar do paciente [Greenes 2007].

Entretanto, é fundamental que, ao serem implementadas, as técnicas de aprendizado de máquina ofereçam métodos de compreensão que possam ser efetivamente aplicados no contexto clínico. Esse aspecto é essencial para permitir que profissionais de saúde compreendam o funcionamento e as decisões geradas por tais técnicas.

Em dezembro de 2023, a legislação da União Europeia [ACT 2023] estabeleceu medidas para regulamentar a aplicação das técnicas de IA. A legislação categoriza os

sistemas de IA em quatro níveis de risco: inaceitável, alto, limitado e mínimo. Sistemas de alto risco, com potencial significativo ou efeitos adversos, estão sujeitos a requisitos legais e obrigações de transparência.

Quando aplicada na prática clínica, a XAI (do inglês, *eXplainable Artificial Intelligence*, ou inteligência artificial explicável em português) aumenta a aceitação, a confiabilidade e a responsabilização dos modelos de aprendizado de máquina. XAI é um conjunto de métodos e técnicas que permitem que os resultados da solução sejam justificados e explicados. Essas técnicas são capazes de estabelecer conexões causais, o que torna mais fácil entender como os modelos fazem decisões [Jung et al. 2023].

Visando reduzir os erros médicos e melhorar a qualidade e eficiência do tratamento clínico oferecido, os Sistemas de Suporte à Decisão Clínica (SSDC) e técnicas de aprendizado máquina podem ser úteis como assistentes. A combinação de conhecimento médico prático com literatura especializada em uma base de conhecimento compartilhado pode melhorar significativamente o processo de tomada de decisão [Sim et al. 2001].

A partir dos SSDC, surge a necessidade de arquiteturas de referência para determinado domínio de aplicação, tais arquiteturas são definidas como modelos generalistas, que definem componentes fundamentais e suas relações [Bass and Kazman 2003].

As arquiteturas orientadas à serviço (*Service-oriented architecture* - SOA), segundo [McGovern et al. 2006], são definidas como um estilo arquitetônico baseado em componentes, na quais os módulos provêm serviços a outros módulos, permitindo a integração de diferentes serviços tanto internos quanto externos à organização.

Originado do modelo SOA, surgiram outros paradigmas, incluindo o software como serviço (SaaS) e a arquitetura baseada no conhecimento como serviço (KaaS). Esta última busca centralizar e fornecer conhecimento, extraído de diversas fontes de dados, por meio de serviços bem definidos, permitindo acesso facilitado para os consumidores.

O presente estudo apresenta uma arquitetura de referência para explicabilidade em IA, baseada na proposta de [Barreto 2016], denominada aqui de XH-KaaS (*eXplanable Health-Knowledge as a Service*). Nesta pesquisa, propomos uma arquitetura de conhecimento como serviço direcionada ao domínio da saúde, que visa complementar os serviços proporcionados pelo paradigma H-KaaS [Barreto 2016], através de serviços de explicabilidade, buscando aprimorar a compreensão de como o servidor de conhecimento fornece o seu conhecimento ao usuário.

2. Fundamentação Teórica

Segundo [Xu and Zhang 2016], o paradigma de conhecimento como serviço (do inglês, *knowledge as a service*), é uma abordagem que visa fornecer o conhecimento ao usuário (consumidor do conhecimento), através de um serviço, sendo possível encontrar três componentes principais: fontes de dados, serviço provedor de conhecimento e consumidores do conhecimento. Onde o provedor de serviços extrai o conhecimento a partir dos conjuntos de dados, fornecendo *insights* valiosos e o consumidor do conhecimento extrair este conhecimento por meio de consultas.

A partir da ideia de um serviço centralizador [Barreto 2016], propõe o H-KaaS (*Health Knowledge as a Service*), um serviço de KaaS dedicado ao domínio da saúde, adaptado a partir da definição de [Xu and Zhang 2016], que define os principais com-

ponentes da arquitetura dedicada a saúde, que são eles: detentores de dados, serviço provedor de conhecimento e consumidores de conhecimento.

XAI se refere ao conjunto de técnicas que visa prover ao usuário um melhor entendimento do processo decisório dos modelos e de como foram obtidos os resultados e as conclusões [Oblizanov et al. 2023]. A ausência de transparência, confiança e interpretabilidade é uma das principais barreiras para a adoção de modelos de aprendizado de máquina na área médica. Para os usuários, é fundamental compreender como os modelos funcionam e como chegaram a suas conclusões.

Nos últimos anos, tem havido um foco considerável no desenvolvimento de modelos de Aprendizado de Máquina na área da saúde, visando identificar áreas afetadas por tumores cerebrais em imagens de ressonância magnética [Xie et al. 2020] e detectar casos de covid [Brunese et al. 2020]. Embora esses modelos frequentemente superem especialistas humanos em desempenho, sua implementação prática é desafiadora devido a fatores como padronização dos dados, compreensão limitada dos algoritmos e presença de vieses. Essas questões impactam a eficácia e a confiabilidade desses modelos na prática clínica.

Com o intuito de proporcionar uma perspectiva estruturada dos métodos e conceitos relacionados a XAI, [Speith 2022] oferece uma visão ampla e coesa da área, apresentando características relevantes para a modelagem por meio da explicabilidade.

3. XH-KaaS (eXplanable Health Knowledge as a Service)

O paradigma KaaS, promove a centralização do acesso ao conhecimento através de serviços para diferentes domínios, quando combinada com técnicas de explicabilidade se tornam verdadeiros aliados, pois tais técnicas oferecem um embasamento maior acerca do conhecimento extraído, aprimorando a sua aplicabilidade e confiança.

A seguir propomos a arquitetura de referência para oferecer explicabilidade como serviço no domínio da saúde. A figura 1 exibe a arquitetura conceitual proposta XH-KaaS, onde observa-se a existência de dois componentes principais, o extrator de explicabilidade (*Explanability Extractor*) e o provedor de explicabilidade (*Explanability Provider*).

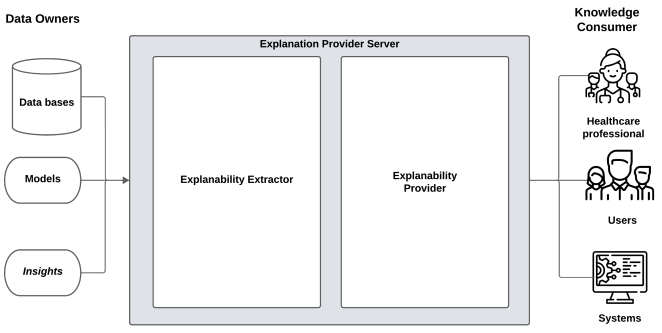

Figura 1. Arquitetura conceitual inicial, a partir do paradigma KaaS.

Desta forma, para o domínio da saúde, podemos considerar como fontes de dados, fontes que ainda não alimentaram processos de aprendizado de máquina, como bases de dados de domínio médico, prontuários médicos eletrônicos, planilhas, imagens, diretrizes clínicas, etc. Como também modelos gerados por métodos de aprendizagem de máquina, ou ainda *insights* que possam aprimorar os métodos de aprendizagem de máquina.

Já os consumidores do conhecimento seriam tanto profissionais da área da saúde quanto pacientes, podendo também o conhecimento ser consumido por outros serviços. Sendo possível a criação de soluções específicas e direcionadas a cada parte de interesse, como por exemplo, prover explicações direcionadas e específicas a determinados usuários.

A figura 2 exibe a arquitetura XH-KaaS em mais detalhes e seus respectivos componentes instanciados. Para a construção e definição de tais componentes presentes tanto no *explanibility extractor* quanto no *explability provider*, a taxonomia de explicabilidade proposta por [Speith 2022] serviu de embasamento, onde é apresentada uma visão ampla e estruturada das definições e métodos de explicabilidade.

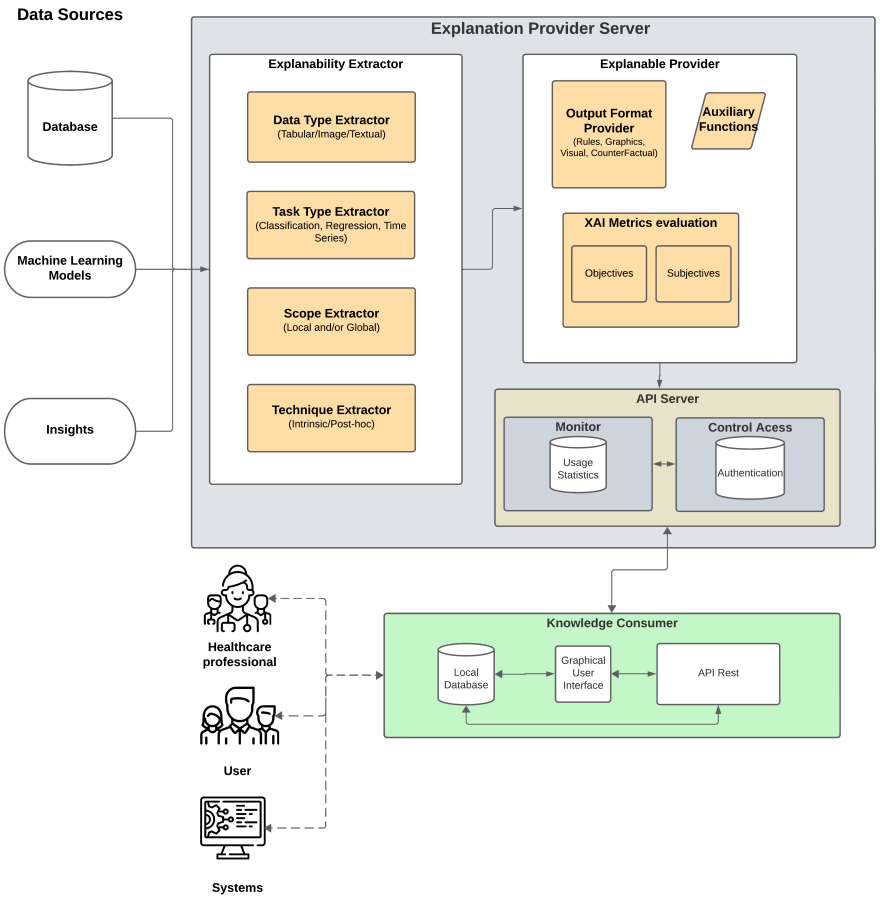

Figura 2. Arquitetura Proposta XH-KaaS

3.1. Extrator de Explicabilidade *(Explanability Extractor)*

O *Explanability Extractor* é responsável pelo processamento e a comunicação dos dados fornecidos como entrada ao servidor provedor de explicabilidade, processo realizado por meio de consultas realizadas pelos consumidores da informação, através de inferências e/ou modelos de aprendizado de máquina.

O *Data Type Extractor* é crucial para identificar o tipo de dado a ser explicado, a partir de quais as técnicas necessárias para garantir a explicabilidade adequada. Por

exemplo, imagens médicas requerem abordagens específicas, como mapas de calor, enquanto dados tabulares exigem métodos distintos. Este componente se relaciona com o *Task Type Extractor* para selecionar as melhores técnicas com base na tarefa preditiva a ser realizada. O *Scope Extractor* define o tipo de explicação a ser fornecida, dividindo-se em local (para uma instância) e global (para o comportamento do modelo em toda a base de treinamento).

Por fim, o componente *Technique Extractor*, método que define qual método de explicabilidade a ser usado, pois, dependendo do modelo fornecido, será possível extrair a explicabilidade do próprio modelo (*intrinsic*), por exemplo, ao utilizar o algoritmo florestas aleatórias (do inglês, *random forest*) podemos utilizar o coefieciente de gini para prover a explicabilidade [Petkovic et al. 2018]. Caso contrário, será aplicado uma técnica *pós-hoc*, métodos que permite extrair a explicabilidade de qualquer modelo utilizado.

3.2. Provedor de Explicabilidade *(Explanability Provider)*

Componente responsável por responder as consultas realizadas pelos consumidores do conhecimento, de acordo com o conhecimento extraído pelo extrator de explicabilidade, a comunicação é realizada através de um servidor de comunicação via API (*Application Programming Interface*).

O componente responsável por fornecer o conhecimento extraído é *Output Format Provider*, pois de acordo com os extrator de explicabilidade e a consulta realizada pelo consumidor do conhecimento, será possível gerar a melhor explicação de acordo com a consulta realizada, o *Output Format Provider*, oferece um conjunto de técnicas de explicação, como técnicas dedicadas a gráficos, mapas de calor, explicações contra factuais etc.

Um outro componente essencial é o *XAI Metrics Evaluation*, as métricas em XAI são projetadas para avaliar a qualidade, eficácia e interpretabilidade das explicações fornecidas por tais modelos [Coroamă 2022]. Segundo [Doshi-Velez 2017], as métricas dividem a avaliação de XAI em métricas subjetivas e objetivas, nas quais as métricas subjetivas são referentes ao *feedback* e avaliação dos usuários e suas percepções ao tema, e as métricas objetivas são métricas que estão relacionadas ao modelo e as técnicas utilizadas.

3.3. Servidor de API de Comunicação

O servidor de comunicação API é uma interface de comunicação compartilhada aos aplicativos dos consumidores, sendo um intermediador entre os servidores e os consumidores do conhecimento, além do monitoramento e autenticação das consultas realizadas.

4. Conclusões e Trabalhos Futuros

Nesse artigo foi proposta a arquitetura XH-KaaS, como uma arquitetura de referência para explicabilidade em saúde, segundo o paradigma kaas. Esta proposta foi baseada a partir do estudo das arquiteturas de KaaS [Xu and Zhang 2016] e H-KaaS [Barreto 2016], além da inclusão das propriedades presentes na taxonomia de [Speith 2022]. A partir da arquitetura proposta, será possível realizar o desenvolvimento e a implementação de instâncias de serviços de explicabilidade segundo a XH-KaaS. Para validar a arquitetura proposta serão desenvolvidos dois estudos de caso para área da saúde, o primeiro envolvendo um conjunto de dados estruturados (tabulares) e o segundo um conjunto de dados de imagens médicas, com a finalidade de identificar os componentes instanciados.

Referências

ACT, E. A. (2023). Eu ai act. https://artificialintelligenceact.com/the-act/. Accessed on March 03, 2024.

Barreto, R. (2016). Nefroservice: Plataforma baseada em conhecimento como serviço no domínio da nefrologia.

Bass, L.; Clements, P. and Kazman, R. (2003). Software architecture in practice. Addison-Wesley Professional.

Brunese, L., Mercaldo, F., Reginelli, A., and Santone, A. (2020). Explainable deep learning for pulmonary disease and coronavirus covid-19 detection from x-ray. In *Computer Methods and Programs in Biomedicine*.

Coroamă, L.; Groza, A. (2022). *Evaluation Metrics in Explainable Artificial Intelligence (XAI)*, pages 401–413.

Doshi-Velez, F.; Kim, B. (2017). Towards a rigorous science of interpretable machine learning.

Greenes, R. A. (2007). Clinical decision support: the road ahead. Elsevier.

Jung, J., Lee, H., Jung, H., and Kim, H. (2023). Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. Heliyon.

McGovern, J., Sims, O., Jain, A., and Little, M. (2006). Enterprise service oriented architectures: Concepts, challenges, recommendations.

Merjulah, R. and Chandra, J. (2019). Classification of myocardial ischemia in delayed contrast enhancement using machine learning. In *Intelligent Data Analysis for Biomedical Applications. pages 209–235*.

Oblizanov, A.; Shevskaya, N., Kazak, A., Rudenko, M., and Dorofeeva, A. (2023). Evaluation metrics research for explainable artificial intelligence global methods using synthetic data. In *Appl. Syst. Innov. 2023*.

Petkovic, D., Altman, R., Wong, M., and Vigil, A. (2018). Improving the explainability of random forest classifier - user centered approach. In *Pacific Symposium on Biocomputing,*.

Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., and Tang, P. C. (2001). Clinical decision support systems for the practice of evidence-based medicine. In *J Am Med Inform Assoc*.

Speith, T. (2022). A review of taxonomies of explainable artificial intelligence (xai) methods. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*.

Xie, B., Lei, T., Wang, N., Cai, H., Xian, J., He, M., Zhang, L., and Xie, H. (2020). Computer-aided diagnosis for fetal brain ultrasound images using deep convolutional neural networks. In *International journal of computer assisted radiology and surgery*.

Xu, S. and Zhang, W. (2016). Knowledge as a service and knowledge breaching. In *2005 IEEE International Conference on Services Computing (SCC'05)*.