UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Detecção de anomalias em prescrições médicas com aprendizagem federada e gerenciamento de armazenamento em *blockchain*

Gabriel Augusto Zutião

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

ICMC USP SÃO CARLOS

Gabriel Augusto Zutião

Detecção de anomalias em prescrições médicas com aprendizagem federada e gerenciamento de armazenamento em *blockchain*

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Jo Ueyama

USP – São Carlos
Outubro de 2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

Z93d

Zutião, Gabriel Augusto
Detecção de anomalias em prescrições médicas com aprendizagem federada e gerenciamento de armazenamento em blockchain / Gabriel Augusto Zutião; orientador Jó Ueyama. -- São Carlos, 2023.
95 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023.

1. Blockchain. 2. Federated Learning. 3. Registros médicos. 4. Detecção de anomalias. I. Ueyama, Jó, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Gabriel Augusto Zutião

Anomaly detection in medical prescriptions with federated learning and model management in blockchain

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Jo Ueyama

USP – São Carlos
October 2023

Este trabalho é dedicado a todos que passaram no meu caminho, seja por alguns dias ou por anos, mas que fizeram a diferença.

AGRADECIMENTOS

Gostaria de agradecer a todos que tanto me ajudaram até aqui, em especial meus pais, minha irmã e meu cunhado, minha noiva, meus sogros e todos nossos familiares, meus colegas do ICMC, todos aqueles que compartilharam momentos comigo em reuniões, ao meu professor orientador e a todos os pesquisadores mais experientes que me ajudaram em minha carreira. Por fim, gostaria de em um destaque especial agradecer a Deus por tudo.

“Nada como procurar quando se quer achar alguma coisa. Quando se procura geralmente se encontra alguma coisa, sem dúvida, mas nem sempre o que estávamos procurando.”

(J.R.R. Tolkien)

RESUMO

ZUTIÃO, G. A. **Detecção de anomalias em prescrições médicas com aprendizagem federada e gerenciamento de armazenamento em *blockchain***. 2023. 95 p. Dissertação (Mestrado – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Com os avanços nas capacidades de processamento e armazenamento de dados em sistemas de registros médicos eletrônicos, evidencia-se a relevância da discussão sobre a existência de um ponto de falha único nos sistemas tradicionais, nos quais todo o tratamento dos dados é feito por uma autoridade central suscetível a falhas e ataques. Os dados de registros médicos, como prescrição de medicamentos, são considerados sensíveis pois tratam de informações pessoais e por isso devem estar seguros e serem privados contra acessos indevidos. No caso de prescrições, podem ocorrer problemas como fraudes e anomalias, tais como dosagens e frequências incorretas ou maliciosas. Entre essas últimas, cita-se as feitas para adquirir medicamentos de mais difícil obtenção para revenda e a compra de medicamentos controlados sem a devida permissão de um médico autorizado para fins de uso abusivo. Algumas soluções presentes na literatura para os problemas apresentados se utilizam de redes descentralizadas para solucionar o problema do ponto único de falha. Outras se utilizam de algoritmos de aprendizado de máquina para a análise de fraudes incluindo a aprendizagem federada, que separa o treinamento do modelo entre os clientes tornando assim o processo descentralizado. Todavia, faz-se necessária a elaboração de um modelo que seja eficaz contra os dois grupos de problemas citados voltado à área de prescrições médicas e que seja eficiente, eficaz, que possa preservar a privacidade dos dados e que seja independente das tecnologias utilizadas e adaptável. Sendo assim, o presente trabalho propõe uma arquitetura de rede *blockchain* associada a uma rede de aprendizagem federada para o processamento de registros de prescrições médicas, utilizando regressão logística para detecção de anomalias na quantidade e na frequência da prescrição de medicamentos. Os experimentos relacionados à rede foram realizados em redes *Ethereum* locais criadas na ferramenta *Hyperledger Besu* integradas a redes de aprendizagem federada criadas com a ferramenta *Flower*. Os resultados obtidos nos experimentos provaram que a arquitetura foi capaz de ser escalável e os seus aspectos qualitativos justificam o aumento do tempo entre as rodadas da aprendizagem federada quando integrada à rede *blockchain*. A solução apresentada é independente de tecnologia, adaptável em relação ao âmbito e também à sua implementação e foi capaz de cumprir com seus propósitos, obtendo uma acurácia de 98,37% na detecção de anomalias e um tempo médio de aproximadamente 10s em cada rodada da aprendizagem em uma rede com 5 nós e aproximadamente 15s para 11 nós, o que demonstrou um aumento menos que linear do tempo.

Palavras-chave: *Federated learning*, Aprendizado de máquina, *Blockchain*, Sistema de saúde,

Detecção de fraude.

ABSTRACT

ZUTIÃO, G. A. **Anomaly detection in medical prescriptions with federated learning and model management in blockchain**. 2023. 95 p. Dissertação (Mestrado – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

With advancements in data processing and storage capabilities within electronic medical record systems, the relevance of discussing the existence of a single point of failure in traditional systems becomes evident. In these systems, all data processing is carried out by a central authority that is susceptible to failures and attacks. Medical record data, such as medication prescriptions, is considered sensitive due to its personal nature, necessitating security and privacy measures against unauthorized access. In the case of prescriptions, issues like fraud and anomalies can arise, including incorrect or malicious dosages and frequencies.

Among the latter, instances are noted where prescriptions are issued to acquire more difficult-to-obtain medications for resale, and the purchase of controlled substances without proper authorization from a licensed physician for abusive purposes. Several solutions found in the literature utilize decentralized networks to address the single point of failure issue. Others employ machine learning algorithms for fraud analysis, including federated learning, which separates model training among clients, thereby decentralizing the process.

Nevertheless, there's a need for an effective model that addresses both sets of issues specific to medical prescriptions. This model should be efficient, capable of preserving data privacy, and technology-agnostic. This study proposes a blockchain network architecture combined with a federated learning network for processing medical prescription records. It employs logistic regression for anomaly detection in medication prescription quantity and frequency. Experiments related to the network were conducted on local Ethereum networks created using the Hyperledger Besu tool, integrated with federated learning networks established using the Flower tool.

The experimental results demonstrated the scalability of the architecture. Qualitative aspects supported the extension of time intervals between rounds of federated learning when integrated with the blockchain network. The presented solution is technology-independent, adaptable in scope and implementation, and effectively fulfilled its objectives. It achieved a 98.37% accuracy in anomaly detection and an average time of around 10 seconds per round of learning in a network with 5 nodes. For 11 nodes, the average time was approximately 15 seconds, demonstrating a less-than-linear increase in time.

Keywords: Federated learning, Machine learning, Blockchain, Healthcare system, Fraud detection.

LISTA DE ILUSTRAÇÕES

LISTA DE QUADROS

LISTA DE ALGORITMOS

LISTA DE TABELAS

SUMÁRIO

CAPÍTULO

1

INTRODUÇÃO

Os avanços nas mais diversas áreas da computação, como a Web 3.0 (MARKOFF, 2006), o aumento na capacidade de processamento de dados e a maior preocupação com a privacidade e a proteção de dados, como tem-se observado com novas legislações como a Lei Geral de Proteção de Dados (LGPD) (BRASIL, 2018) e as implicações da utilização de novos modelos de inteligência artificial com maior capacidade de tratamento de dados e abrangendo novos horizontes, o que tem feito com que o termo seja mais popularizado.

Tendo em vista esse contexto, pode-se enfatizar a área dos registros médicos eletrônicos (ANGST; AGARWAL, 2009; JENSEN; JENSEN; BRUNAK, 2012), que requer atenção devido à sensibilidade dos dados tratados, como históricos médicos, que podem ser utilizados de maneira abusiva. Além disso, é importante destacar a possibilidade de fraudes e anomalias nas entradas dos registros, como, por exemplo, prescrições de medicamentos ilícitos ou venda de medicamentos controlados sem prescrição. Um dos casos que pode-se citar, é a adulteração de prescrições médicas para fins de comercialização em mercado paralelo (WHITE; READY; KATZ, 2016).

Tradicionalmente, após a transição dos registros para os meios eletrônicos, os sistemas de prescrição médica moldaram-se de forma a serem centralizados, onde uma autoridade central é responsável pela manutenção de toda a rede, como o gerenciamento do manuseio e da transmissão dos dados. Assim, é evidenciado o problema do ponto único de falha no modelo da rede, visto que o seu funcionamento depende integralmente dessa autoridade, que é suscetível a falhas e ataques, além de utilização maliciosa.

Tendo em vista os problemas das redes tradicionais de registros eletrônicos médicos, surgiram propostas de redes alternativas que focam na descentralização da autoridade para resolver a existência de um ponto único de falha (AZARIA *et al.*, 2016; ZYSKIND; NATHAN; PENTLAND, 2015; ABBAS *et al.*, 2020; MACKEY *et al.*, 2020), sendo, portanto, mais resistentes a ataques externos. Além disso, foram propostos avanços na identificação de fraudes

e anomalias a fim de tornar as mesmas mais confiáveis e privadas, em conformidade com as crescentes expectativas na proteção de dados sensíveis. É nesse contexto que pode-se citar a utilização de redes *blockchain* (NAKAMOTO, 2008) e de aprendizado de máquina, mais especificamente a abordagem da aprendizagem federada (MCMAHAN *et al.*, 2016) a fim de solucionar os problemas apresentados.

A utilização de redes *blockchain* tem sido expoente em diversas áreas, como cripto moedas, setores industriais, tecnologias de pagamento, entre outras. Trata-se de redes que favorecem a descentralização dos dados pela inexistência de uma autoridade central e a garantia de segurança por meio de protocolos de consenso entre os pares e algoritmos de criptografia, além da possibilidade de utilização conjunta de outros métodos, como assinaturas digitais, adição de ruído e ofuscação dos dados.

Além da utilização de redes *blockchain*, abordagens com aprendizado de máquina podem solucionar problemas de segurança com as redes tradicionais, principalmente com a detecção de dados anômalos. A aprendizagem federada, mais especificamente, tem algumas similaridades em sua proposta com esse tipo de rede, como a maior descentralização e distribuição das responsabilidades entre os pares.

Alguns estudos foram propostos resolver os problemas acima citados e demais preocupações com segurança e/ou privacidade de dados (LI *et al.*, 2021; SHAYAN *et al.*, 2021; ZHANG *et al.*, 2021; CHANG; FANG; SUN, 2021; SINGH *et al.*, 2022; ASHRAF *et al.*, 2022) em sua parcialidade. Apesar disso, não foi possível encontrar uma solução que estudasse a integração de aprendizagem federada assistida por redes *blockchain* para a detecção de anomalias em registros médicos de maneira efetiva e que se mostrasse eficiente na substituição das atuais redes. De maneira geral, os trabalhos presentes na literatura ou tiveram foco diferente (PENG *et al.*, 2022; LI *et al.*, 2021; SHAYAN *et al.*, 2021); ou tiveram foco em algum outro problema das redes na utilização de aprendizado de máquina e/ou *blockchain* (LI *et al.*, 2020; MCMAHAN *et al.*, 2016; BAO *et al.*, 2019); ou ainda possuíram limitações em relação a seu desempenho ou inexistência de testes/implementação (LI *et al.*, 2020; BAO *et al.*, 2019; SINGH *et al.*, 2022).

Tendo em vista as limitações apresentadas, o presente estudo propõe a utilização de uma rede *blockchain* unida de aprendizagem federada a fim de solucionar os problemas apresentados pelas redes tradicionais de distribuição de prescrições médicas eletrônicas. Para isso, proporcionou-se uma estrutura por camadas que separa a coleta, o processamento e a persistência dos dados com a detecção de dados anômalos a fim de identificar erros e fraudes. Assim, é provido um ambiente descentralizado, confiável, íntegro e portátil para assistir as partes, uma vez que pode ser integrado a outros tipos de sistemas e é adaptável conforme a necessidade, focando-se na construção de um modelo robusto para solucionar esses problemas.

Os testes da estrutura proposta foram feitos em uma rede *Ethereum* (BUTERIN, 2014) concebida com a ferramenta *Hyperledger Besu* (FOUNDATION, 2023), por sua vez integrada com uma rede de aprendizagem federada desenvolvida com a biblioteca *Flower* (BEUTEL *et al.*,

2022). Foram estudadas três soluções a fim de comparação, uma que confia maior quantidade de dados do treinamento à rede *blockchain*, outra que confia menos e uma terceira com a integração de um banco de dados paralelo à rede para o armazenamento dos modelos do treinamento com acesso via chaves que são armazenadas no livro razão.

Para alcançar a finalidade da detecção de anomalias são analisadas as disparidades na quantidade e na frequência de uma medicação nas prescrições e caracterizando uma possível fraude uma quantidade/frequência anormal de prescrições de um determinado medicamento em um determinado registro, por exemplo. O conjunto de dados explorados é relativo ao Hospital Nossa Senhora da Conceição (SANTOS *et al.*, 2019) e possui informações anônimas sobre as prescrições de Janeiro a Setembro de 2017, indicando quais foram anômalas. Apesar da base escolhida, o sistema desenvolvido possa ser adaptado para outros programas e sistemas de saúde. Os experimentos foram realizados em um ambiente de testes local e comparados entre as soluções.

1.1 Motivação

Tendo em vista o apresentado, nota-se, baseando-se na literatura, que existem ataques aos originais propósitos dos sistemas de prescrições, como abuso de medicamentos controlados (PHILLIPS, 2013) e venda em mercados paralelos (GOLDMAN, 1998), além da ocorrência de erros nos processos (entrada de dados anômalos de origem não maliciosa). Sendo assim, faz-se conveniente a proposta de uma arquitetura que possa detectar anomalias e ajudar na auditoria dos modelos dos pares que seja flexível, adaptável e independa de tecnologias específicas e que forneça bases para a proteção e privacidade de dados.

O uso de rede *blockchain* favorece a maior independência dos nós da rede de um sistema centralizado, além de prover proteção aos dados. Além disso, a utilização da tecnologia proporciona uma maior adaptabilidade devido ao uso de contratos inteligentes em caso de surgimento de novas lógicas de negócio.

A utilização de aprendizado de máquina, por sua vez, garante a segurança do sistema contra ataques de introdução de dados anômalos com intuitos maliciosos, como apontado. Com a identificação de pares maliciosos pode-se tomar medidas cabíveis. Além disso, pode-se utilizar futuramente a classificação para determinar quais pares da rede mais beneficiam o treinamento. Com a introdução de aprendizagem federada, garante-se um maior aproveitamento do treinamento e a melhora na classificação em eventuais mudanças nas informações das prescrições causadas por avanços no ramo médico.

1.2 Objetivos gerais e específicos

Esta pesquisa tem por objetivo propor um modelo que se utiliza de aprendizagem federada para a análise de prescrições médicas, buscando classificá-las em anômalas ou não baseando-se na quantidade e na frequência das prescrições em uma rede descentralizada baseada em *blockchain*, com o intuito de promover a detecção de fraudes em um ambiente descentralizado de compartilhamento de dados. Com isso, propõe-se uma rede capaz de auxiliar no compartilhamento de modelos para a detecção de anomalias e assim auxiliar no desenvolvimento de redes e sistemas médicos mais seguros.

Para alcançar esse objetivo primário, são definidos os seguintes objetivos específicos:

- Descobrir quais são os padrões relevantes na análise de prescrições médicas, a fim de poder detectar fraudes;
- Desenvolver um modelo de classificação que possa analisar os dados presentes a fim de separar os anômalos dos comuns;
- Elaborar uma arquitetura de rede *blockchain*, realizando testes por simulações, e integrar o modelo de classificação (aprendizagem federada) como uma das camadas a fim de poder analisar os dados que transitam nela;
- Avaliar o desempenho do modelo de classificação de maneira quantitativa e a rede por meio de testes de capacidade e custo computacional.

1.3 Organização da Monografia

O presente trabalho é composto por 5 capítulos, sendo os posteriores apresentados conforme descrito a seguir:

- Referencial teórico apresenta a base teórica sobre os tópicos abordados na presente pesquisa, como gestão de dados médicos, *blockchain* e suas implementações utilizadas, aprendizado de máquina e aprendizagem federada;
- Trabalhos relacionados expõe os trabalhos correlatos com os temas de apresentados de redes *blockchain* assistidas por aprendizagem federada, dando mais relevância aos que envolvem a área de registros médicos eletrônicos, descrevendo o estado-da-arte;
- Modelo busca descrever ao leitor os métodos utilizados na pesquisa em todas as suas etapas, desde a pesquisa na literatura até o desenvolvimento e obtenção dos resultados; apresenta o desenvolvimento do trabalho, desde a parte da aprendizagem federada à integração com uma rede *blockchain* e como a estrutura é proposta e, por fim, busca demonstrar os

resultados obtidos nos testes apresentados anteriormente e elaborar uma discussão sobre eles;

- Considerações finais apresenta as conclusões obtidas pela análise dos resultados e a discussão final.

CAPÍTULO

2

REFERENCIAL TEÓRICO

No presente capítulo é apresentado o referencial teórico da pesquisa. Na Seção 2.1 é explicada a gestão de dados de registros médicos físicos e eletrônicos e os problemas dos sistemas atuais; na Seção 2.2 são apresentadas as redes *blockchain*, sua definição e características, além das plataformas *Hyperledger Fabric* e *Ethereum* e o cliente *Hyperledger Besu*, que serão mencionados posteriormente no presente trabalho; na Seção 2.3 são explicados os contratos inteligentes e sua aplicação em redes *blockchain*. Na Seção 2.4 é explicado o conceito de aprendizado de máquina, especificando as diferenças entre supervisionado e não supervisionado e caracterizando os algoritmos utilizados, tanto na implementação do início das pesquisas quanto na final. Na Seção 2.5 é apresentada a aprendizagem federada, suas características e usos; na Seção 2.6 serão apresentadas as características da integração de redes *blockchain* e aprendizagem federada e por fim na Seção 2.7 é apresentada uma breve discussão sobre as informações previamente apresentadas.

2.1 Gestão de dados de registros médicos

Os registros médicos, como as informações de consultas médicas, diagnósticos e tratamentos eram inicialmente feitas de maneira manual e posteriormente foram introduzidos os registros médicos eletrônicos tratados de forma centralizada, ou seja, cada organização possui seu próprio sistema que requer um tipo de integração com os de outras organizações (ISMAIL; MATERWALA, 2020).

Os sistemas tradicionais de gerenciamento de registros médicos eletrônicos possuem dados sensíveis dos pacientes, como prescrições e informações dos atendimentos, assim requirindo uma atenção especial a falhas e ataques. Como eles possuem uma autoridade central que cuida do gerenciamento, é evidenciada a existência de um problema de centralidade e dependência da rede à autoridade, que pode ser comprometida tanto por ataques quanto por falhas e erros (TANWAR; PAREKH; EVANS, 2020), prejudicando sua confiabilidade.

2.2 Redes *blockchain*

As redes *blockchain* foram propostas para resolver problemas que as redes tradicionais apresentavam. As redes tradicionais eram dependentes de uma parte confiável para garantir seu funcionamento, como é apresentado por Nakamoto (2008), que introduziu a primeira implementação de uma rede *blockchain*, chamada de *bitcoin*. para utilizá-la como uma moeda descentralizada e inteiramente virtual. As características dessas redes são a existência de uma estrutura descentralizada, que independe de um terceiro confiável para garantir a segurança e integridade dos dados (NAKAMOTO, 2008).

Como afirma Christidis e Devetsikiotis (2016), uma rede *blockchain* opera com uma quantidade de nós que compartilham o mesmo livro-razão (*ledger*) e operam de forma par-a-par, onde os nós, por meio de criptografia, asseguram a autenticidade das transações por eles realizadas e validam as transações de outros nós. Essas transações, assim que validadas, são ordenadas e posteriormente adicionadas ao livro-razão.

As redes *blockchain* podem ser divididas em públicas, em que qualquer indivíduo ou organização pode ingressar; privadas, em que elas são controladas por uma organização e permissionadas; e de consórcio, em que existe uma associação de várias partes de maneira também permissionada (CHAN *et al.*, 2019).

2.2.1 Tecnologias e plataformas

Entre as plataformas *blockchain* e de registros distribuídos, destacam-se as redes *Ethereum* e a plataforma *Hyperledger Fabric*. A seguir serão descritas ambas tecnologias, além do cliente *Ethereum Hyperledger Besu*, que será posteriormente referenciado no trabalho.

2.2.1.1 Hyperledger Fabric

Hyperledger Fabric é um sistema de implementação de redes *blockchain* permissionadas que pode ser estendido. O consenso é alterável de acordo com a implementação da rede, sendo o endosso das transações enviado a outros nós responsáveis por essa validação, que por sua vez executam a validação com os dados presentes no livro-razão e retornam os dados de saída para o cliente. Este, por sua vez, envia as informações de endosso ao serviço de ordenação. A plataforma permite a escrita de contratos inteligentes utilizando linguagens de programação como Go, JavaScript, Python e Java e provê um sistema opcional de distribuição dos dados de maneira *peer-to-peer* (ANDROULAKI *et al.*, 2018).

2.2.1.2 Ethereum e Hyperledger Besu

As redes *Ethereum* foram propostas em 2014 por Vitalik Buterin para resolver limitações de aplicações *blockchain*, como a falta de uma linguagem de programação que possuísse com-

pletude de Turing, a impossibilidade de controle de estado nos contratos e a inacessibilidade do livro-razão nos mesmos (BUTERIN, 2014).

Ainda segundo Buterin (2014), devido a suas características, a plataforma permite que os desenvolvedores tenham mais liberdade para desenvolver contratos que permitam definir regras para transações, controle de estado e propriedade dos recursos, além de possuir um conceito de mensagens (semelhantes às transações da *Bitcoin*), uma moeda própria (*wei*/*ether*), um esquema de mineração próprio e uma máquina virtual para execução dos contratos.

Hyperledger Besu é um cliente Ethereum capaz de prover transações privadas sendo integrável com redes públicas Ethereum, podendo usar tokens ERC20 e a moeda Ether. Ele possui código aberto, é mantido pela Hyperledger Foundation e fornece uma API para gerenciar os nós e executar transações. Pela possibilidade de manter privacidade e permissionamento dos dados, pode alcançar casos de uso que outros clientes *Ethereum* não podem. Mesmo com essas vantagens, o cliente continua sendo flexível e possui um bom desempenho em relação de vazão das transações. A utilização de gerenciadores de transações privadas faz com que a lógica de negócios de contratos inteligentes seja privada a uma seção dos usuários (UDDIN *et al.*, 2021).

2.3 Contratos inteligentes

Somada à implementação de algumas redes *blockchain* está a implementação de contratos inteligentes. Eles se caracterizam pela existência de cláusulas implementadas em lógicas de programação que realizam operações nos dados que são salvos nas redes *blockchain*, assim permitindo a implementação de regras de negócio que podem ser alteradas com a necessidade no desenvolvimento da rede (ZHENG *et al.*, 2020). Suas vantagens são a redução de riscos, a diminuição de custos para operações de administração e serviços e aumento na eficiência dos processos.

Os contratos inteligentes são úteis em diversas aplicações, como sistemas de eleições, leilões, pagamento e distribuição de energia (KEMMOE *et al.*, 2020), entre outras.

2.4 Aprendizado de máquina

Aprendizado de máquina é uma subárea da inteligência artificial que tem se popularizado há décadas, em que algoritmos são desenvolvidos para aprender a partir de experiências adquiridas, geralmente tirando conclusões a partir de exemplos, o que é chamado de indução. As aplicações baseadas em aprendizado de máquina buscam modelos que possam representar um conhecimento adquirido com base em um conjunto de dados (DOMINGOS, 2012; MITCHELL, 2006; FACELI *et al.*, 2011).

Com o aumento do volume dos dados produzidos, assim como sua variedade de formatos (HALEVY; NORVIG; PEREIRA, 2009) nas mais diferentes áreas, além de avanços gerais na

área de aprendizado de máquina e outras subáreas de inteligência artificial fizeram com que seu uso se tornasse cada vez mais viável e capaz de realizar tarefas que antes eram consideradas dificultosas (FACELI *et al.*, 2011).

Entre áreas que utilizam aprendizado de máquina, Faceli *et al.* (2011) cita as áreas de agronegócios, análises comportamentais, bioinformática, energia, finanças, robótica, sistemas de recomendação, saúde, entre outras. Os conjuntos de dados possuem atributos chamados preditivos, que representam características das entradas que são utilizadas para realizar predições ou descrições a partir do modelo utilizado, e podem possuir atributo alvo, também chamado de rótulo, que indica a classe do objeto.

Os algoritmos de aprendizado de máquina podem ser separados em supervisionados e não supervisionados, conceitos que serão tratados nas subseções a seguir.

2.4.1 Aprendizado de máquina supervisionado e não supervisionado

Algoritmos supervisionados de aprendizado de máquina se baseiam na construção de um estimador a partir de um conjunto de dados que na maioria dos casos deve ser rotulado. Trata-se de uma tarefa de rotular os objetos a partir de classes conhecidas. Entre esses algoritmos, ainda pode-se separar entre os de classificação e os de regressão. Os primeiros buscam, como o nome indica, classificar as entradas baseando-se em um domínio finito de valores nominais, enquanto os algoritmos de regressão buscam estimar os valores em um conjunto infinito como domínio (DIETTERICH, 1998 apud FACELI *et al.*, 2011).

Os algoritmos de aprendizado de máquina não supervisionados são voltados à categorização (JAIN; DUIN; MAO, 2000) ou descrição dos dados, onde o aprendizado é baseado na identificação de padrões ou tendências no conjunto de dados a fim de auxiliar no propósito no qual eles estão sendo empregados, como descobrir maneiras de agrupar os objetos e fazer a associação entre eles a partir de padrões (SOUTO *et al.*, 2003 apud FACELI *et al.*, 2011).

2.4.2 Regressão logística

A regressão logística é uma técnica de aprendizado de máquina supervisionada utilizada em problemas que envolvem uma saída binária. Seu funcionamento é exposto na Equação 2.1, onde $p(x)$ representa a probabilidade da variável dependente Y seja 1, dado um conjunto de variáveis independentes X. Na expressão, $\beta_0, \beta_1, \ldots, \beta_p$ representam os coeficientes obtidos por meio do treinamento e $x_1, x_2, \ldots, x_p$ representam os valores das variáveis independentes (sendo p o seu número) (KLEINBAUM; KLEIN, 2010).

Como observa-se, nela são utilizadas variáveis independentes que preveem a probabilidade de um evento ocorrer por meio de um método chamado máxima verossimilhança, que calcula a probabilidade de se observar os valores das variáveis dependentes a partir das variáveis independentes, procurando coeficientes que maximizem essa probabilidade. O resultado desse

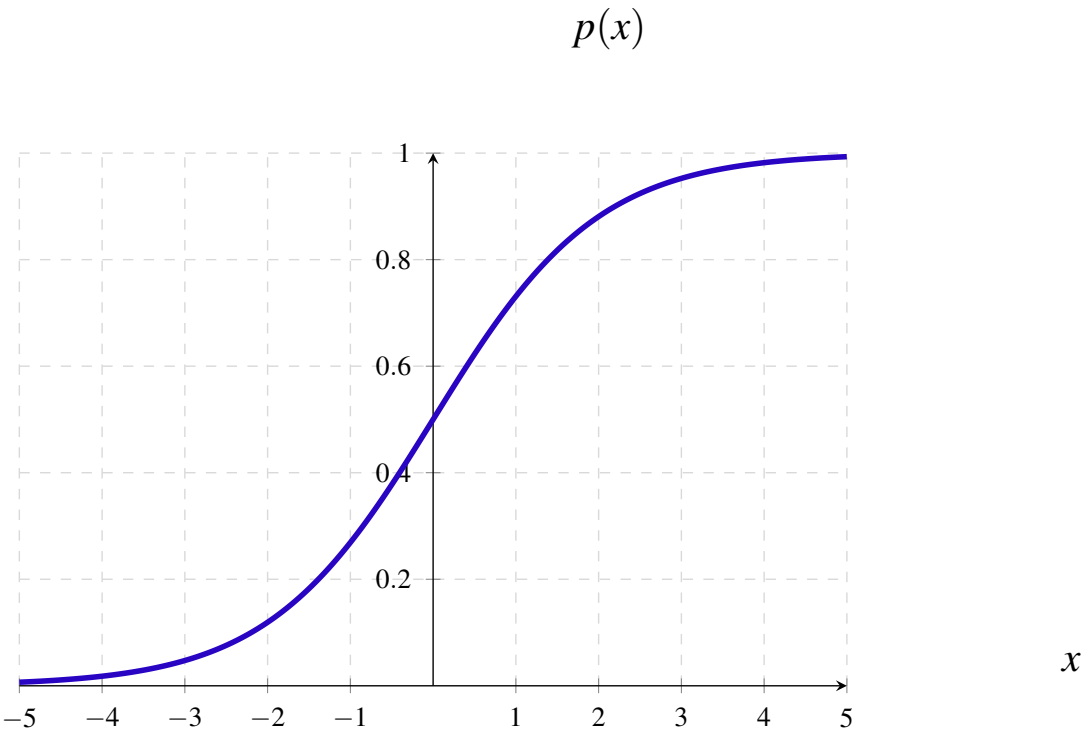

Figura 1 – Representação dos resultados gráficos da regressão logística

cálculo é uma combinação linear das variáveis e seus coeficientes, que necessita ser transformada em uma probabilidade entre 0 e 1 utilizando uma função sigmoide. No fim do processo, o modelo associa aos dados de teste um valor entre 0 e 1 (KLEINBAUM; KLEIN, 2010).

Conforme apresentado por Kleinbaum e Klein (2010), casos favorecidos pela utilização do algoritmo são aqueles que requerem uma classificação binária. Além disso, o modelo é capaz de contemplar os que possuem um problema de múltiplas variáveis independentes, uma vez que ele busca os coeficientes que mais adéquam a estimação dos valores das variáveis dependentes.

A representação gráfica dos resultados é uma curva, onde a variável independente é representada pelo eixo X e a probabilidade de pertencer à classe 1, representando a curva a relação entre essa probabilidade e o aumento da variável independente 1.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{2.1}$$

2.5 Aprendizagem federada

O termo *federated learning* (ou aprendizagem federada) foi apresentado por McMahan *et al.* (2016) para definir um novo método de aprendizado de máquina onde os dispositivos treinam modelos separadamente com dados locais e esses modelos são enviados a um servidor central, que realiza o processo de agregação e gera um novo modelo global.

Durante o processo de aprendizagem federada, o algoritmo mais comumente visto é a média federada (conhecida como FedAvg). Nesse algoritmo, é executado em paralelo certo número de passos do algoritmo Gradiente Descendente Estocástico (GDE, ou *Stochastic Gradient Descent*), cujo funcionamento é descrito abaixo (LI *et al.*, 2019).

O algoritmo Gradiente Descendente Estocástico é responsável por iterar um modelo a fim de otimizar a função de perda, buscando a convergência do modelo utilizando pequenas porções aleatórias por vez (BOTTOU, 2012).

A média federada executa o GDE em paralelo em um número específico de dispositivos e então realiza a média dos modelos com certa periodicidade, tendo seu foco mais na computação dos dados do que na comunicação entre os pares da rede (LI *et al.*, 2019).

A seleção de nós para as rodadas é, em geral, feita randomicamente, o que se caracteriza uma seleção sem viés. São selecionados uma quantidade predeterminada de clientes para contribuir com o modelo global, embora existam propostas de métodos diferentes, chamados de enviesados, por possuírem um critério específico para a seleção de nós, como a estratégia proposta por (CHO; WANG; JOSHI, 2020), por exemplo.

Um exemplo citado por McMahan *et al.* (2016) é o caso de melhoria em sistema de predição de próximas palavras para sugestão em teclados de *smartphones*. A utilização de aprendizagem federada nesse caso favoreceria a privacidade (pois os dados individualizados não seriam enviados ao servidor para serem agregados, mas somente os modelos).

Existem alguns paradigmas que são motivos de estudos visando melhorias nas arquiteturas de aprendizagem federada tradicionais, como métodos para tornar a comunicação entre os pares da rede mais eficiente por meio de um menor uso de banda nas atualizações de modelo enviadas pelos clientes ao servidor (KONEčNý *et al.*, 2016) e também seleção de clientes para a submissão dos modelos atualizados (XU *et al.*, 2021).

O uso de aprendizagem federada pode melhorar sistemas das mais diversas áreas, sendo uma delas de certo destaque a área médica, devido à preservação da privacidade dos dados devido ao compartilhamento do modelo e também à falta de um modelo que tenha um bom desempenho, visto que cada instituição possui uma quantidade de dados específica que pode ser enviesada, enquanto um modelo "global"teria mais diversidade de entradas, o que soluciona problemas causados pelo viés (XU *et al.*, 2021).

Além disso, Yang *et al.* (2019) separa as redes de aprendizagem federada em três: as horizontais, onde existe diferença nas entradas de dados porém com as mesmas seleções de informações; as verticais, onde existe diferença das informações selecionadas porém provindas das mesmas entradas de dados e a aprendizagem transferida federada, onde pouco se compartilha de semelhança entre as entradas e o tipo das informações tratadas.

2.6 Aprendizagem federada assistida por *blockchain*

As redes *blockchain* integradas às redes de aprendizagem federada podem solucionar problemas presentes nas duas tecnologias caso utilizadas separadamente, tais como a centralização das redes de aprendizagem federada, como também os problemas de incentivo para que clientes possam se juntar ao treinamento (uma vez que ele tem um custo de processamento), a falta de confiabilidade dos dados inseridos das redes *blockchain* e a vulnerabilidade à sabotagem presente nas redes tradicionais (WANG; HU, 2021).

A utilização de redes *Blockchain* aliada à aprendizagem federada pode fornecer rastreabilidade, transparência, decentralização e segurança aos dados, podendo os contratos inteligentes serem utilizados para realizar as operações da aprendizagem federada que anteriormente seriam realizadas em um servidor central (NGUYEN *et al.*, 2021; WANG; HU, 2021; RAMANAN; NAKAYAMA, 2020).

Wang e Hu (2021) separam as redes que unem *blockchain* e aprendizagem federada em três categorias:

- ***Fully coupled* (ou completamente acopladas)**, são as redes em que os próprios clientes da aprendizagem federada são responsáveis pelos processos dos nós da *blockchain*, isso é, além de realizar o treinamento dos dados localmente, também fazem as operações de gerar novos blocos, o que torna a rede independente de um servidor central mas aumenta os desafios relacionados ao uso de processamento e banda entre nos pares;

- ***flexibly coupled* (ou flexivelmente acopladas)**, onde as funções de cliente da aprendizagem federada e nós mineradores da *blockchain* são feitas em ambientes separados, um cliente não necessariamente é um nó. Isso ajuda com os desafios das redes completamente acopladas, porém possui limitação em sua descentralização devido ao processo de agregação e requer que as duas redes estejam funcionando de maneira ordenada;

- ***loosely coupled* (ou vagamente acopladas)**, onde as redes *blockchain* e de aprendizagem federada são independentes, porém a primeira é utilizada para guardar informações de atualização do modelo e para a avaliação da reputação dos clientes, o que ajuda no gerenciamento dos clientes que participam da rede porém sendo menos eficiente na utilização dos recursos e tendo o modelo da aprendizagem de máquina sendo centralizado.

2.7 Discussão final

Sistemas tradicionais de prescrições médicas eletrônicas, como visto, enfrentam desafios na segurança e privacidade de dados. Por tratarem dados sensíveis e que podem ser utilizados contra os pacientes e/ou contra a saúde pública em geral, faz-se conveniente a proposta de

sistemas alternativos que permitam privacidade, segurança e interoperabilidade de maneira eficiente.

Tendo em vista esses problemas e necessidades, o presente trabalho busca propor uma arquitetura que integra aprendizagem federada com aprendizado de máquina supervisionado e redes *blockchain* a fim de prestar auxílio a sistemas de prescrições médicas eletrônicas para garantir a segurança das informações e detecção de fraudes e anomalias, de maneira privativa e ajustável.

CAPÍTULO

3

TRABALHOS RELACIONADOS

O presente capítulo apresenta os trabalhos relacionados ao tópico de aprendizagem federada assistida por redes *blockchain*, dando ênfase mas não se limitando aos que possuem como foco a área dos sistemas de registros médicos eletrônicos para fins de comparação. Na seção 3.1 é definido o problema que será abordado nesta pesquisa e como foi feita a pesquisa bibliográfica. Na seção 3.2 são apresentados e discutidos de maneira descritiva os trabalhos encontrados na pesquisa bibliográfica e posteriormente comparados com o presente trabalho. Por fim, na seção 3.3 é feita a apresentação da lacuna que ele busca solucionar.

3.1 Definição do problema

Como apresentado pelo Capítulo 1, os sistemas eletrônicos de registros médicos tradicionais apresentam um ponto único de falha (KEMMOE *et al.*, 2020; CLIM; ZOTA; CONSTANTINESCU, 2019), sendo dependentes de um provedor específico que controla os dados. Além disso, como abordado, é conveniente ressaltar a possibilidade de fraude e/ou utilização indevida das prescrições, seja por falsificação (PRASAD; JENA, 2013), utilização de drogas ilícitas e abuso de drogas (STRANG *et al.*, 2012), entre outros.

Segundo Lapeyre-Mestre *et al.* (2014), que realizaram um estudo agregando e analisando dados de alguns países da Europa, os ataques mais comuns são os de modificação de prescrições, não seguimento de regras determinadas e alterações em geral na receita. Sendo assim, propõe-se a seguir o modelo de ataque em que os sistemas de prescrição médica estão sujeitos:

Dada uma rede de prescrições médicas (tradicional ou não), é determinado como ataque tanto a utilização de meios convencionais de maneira ilícita (como prescrever medicamentos além da quantia devida) quanto a utilização ilícita dos meios em si, como, por exemplo, o extravio de formulários médicos nos sistemas tradicionais e a introdução de entradas de dados por partes não autorizadas nas redes eletrônicas.

Este trabalho tem por objetivo propor uma solução ao problema de ataque descrito de maneira privativa, apresentando uma arquitetura baseada em uma rede *blockchain* com uma camada para a validação dos dados de prescrições eletrônicas utilizando aprendizagem federada, a fim de identificar fraudes e prover um sistema sem um ponto de falha único, assim sendo objetivamente menos suscetível a casos de instabilidade e falhas.

3.1.1 Definição da pesquisa bibliográfica

Na primeira parte da pesquisa, para elaboração da solução, foi feita uma pesquisa bibliográfica sobre a utilização de aprendizagem federada assistida por redes *blockchain*, posteriormente focando nas propostas para a área dos registros médicos eletrônicos. Isso se deu por meio da combinação das palavras chave "*blockchain*", "*federated learning*"e "*healthcare*"em pesquisas nas bases de dados *Scopus* (B.V., 2023) e *Google Acadêmico* (GOOGLE, 2023). Foram selecionados manualmente, por meio de análise das apresentações, trabalhos relevantes e que expressivamente mais se assemelhavam ao proposto na utilização de aprendizagem federada e redes *blockchain*, dando ênfase àqueles que possuíam relação com a área dos registros médicos. Após a primeira etapa de seleção dos trabalhos, foram por fim escolhidos aqueles que seriam relevantes à presente pesquisa.

3.2 Discussão dos Trabalhos Relacionados

3.2.1 Definição dos trabalhos

3.2.1.1 Trabalhos relacionados no uso de aprendizado de máquina

- *Identifying frauds and anomalies in Medicare-B dataset*

  O trabalho de Seo e Mendelevitch (2017) propôs a utilização de um algoritmo baseado no *PageRank* (onde os dados são organizados em grafos e é calculada a probabilidade de se chegar em um nó a partir de outro) para identificar fraudes na base de dados *Medicare-B* (uma base de dados com informações públicas de seguro médico dos Estados Unidos) e foi capaz de identificar pedidos de seguro de vida que eram anômalos quando comparados com outros da mesma especialidade médica.

- *Identifying medicare provider fraud with unsupervised machine learning*

  Bauder, Da Rosa e Khoshgoftaar (2018) utilizaram os algoritmos *Isolation Forest*, *Unsupervised Random Forest*, *Local Outlier Factor*, autocodificadores e *k-Nearest Neighbors* para identificar fraudes na base de dados *Medicare-B*, validando a performance com classificações do banco de indivíduos e entidades excluídos. Os autores concluíram que o algoritmo *Local Outlier Factor* teve a melhor performance quando analisada a curva de Característica de Operação do Receptor.

- *Medicare fraud detection using neural networks*

  Johnson e Khoshgoftaar (2019) utilizaram a mesma base de dados (*Medicare-B*) e base de dados para a caracterização das anomalias, medindo a performance dos algoritmos de redes neurais. Os autores concluíram que as melhores performances obtidas foram com o tratamento do desbalanceamento das classes, utilizando-se tanto de métodos em nível de dados (alterando a distribuição dos dados de treinamento) quanto em nível de algoritmo (alterando a função de perda para que as classes minoritárias tivessem mais relevância).

- *Comparative Study of Using Various Machine Learning and Deep Learning-Based Fraud Detection Models For Universal Health Coverage Schemes*

  O trabalho de Yashraj *et al.* (2021) buscou a detecção de fraudes nos dados do sistema de saúde indiano *Ayushman Bharat* utilizando variantes de redes neurais e modelos de classificação para detectar fraudes em valores como os montantes de pagamentos e o tempo do tratamento. O estudo pôde concluir que as redes neurais treinadas em bases em que o desbalanceamento foi tratado com a subamostragem (para balancear as entradas fraudulentas das legítimas) obtiveram uma melhor performance (medida com a medida F) do que os algoritmos de aprendizado de máquina convencionais.

- *Medical fraud and abuse detection system based on machine learning*

  Zhang, Xiao e Wu (2020) utilizaram uma rede neural para identificar anomalias no relacionamento entre informações de pagamentos, frequência, hospital, informação pessoal e tratamento de dados do sistema de saúde da província de Zhejiang (China). Por não obter um resultado satisfatório na classificação dos dados devido ao pequeno tamanho da base de dados de treinamento, os autores sugeriram a utilização de seu estudo apenas para fins de analisar a eficiência dos métodos empregados, por causa de limitações na base de dados e da possibilidade de melhora nos resultados utilizando outras técnicas de balanceamento dos dados, como a superamostragem de minorias sintéticas.

3.2.1.2 Trabalhos relacionados no uso de redes blockchain

- *Blockchain paradigm for healthcare: Performance evaluation*

  Quanto à utilização de redes *blockchain* na área da saúde, Ismail e Materwala (2020) demonstraram por meio de comparações a eficiência da utilização das redes descentralizadas ao invés da arquitetura comum cliente-servidor por meio da avaliação do tempo de execução e da quantidade de dados transferidos, sendo a implementação em *blockchain* até dez vezes mais eficiente para a transferência de dados, segundo os testes dos autores. Os autores determinaram por meio de que o tempo das consultas na rede de testes*blockchain* desenvolvida é 11,7 vezes mais rápido que em uma rede convencional, porém sua utilização é mais custosa devido à maior transferência de dados e o tempo de execução.

- *Combating health care fraud and abuse: Conceptualization and prototyping study of a blockchain antifraud framework*

  O trabalho de Mackey *et al.* (2020) descreve o desenvolvimento de uma rede *Ethereum* também tendo em vista a rede *Medicare*. O trabalho proposto não possui um sistema antifraude específico, sendo uma sugestão dos autores a implementação de detecção de fraude antes do processo, ao invés da análise das transações já realizadas. Os autores se apoiam na confiabilidade dos dados da rede *blockchain* e afirmam ser de possível integração com sistemas que estão sendo desenvolvidos para essa detecção.

- *A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry*

  Abbas *et al.* (2020) desenvolveram um sistema de gerenciamento da cadeia de fornecimento de medicamentos em uma rede *blockchain* permissionada (*Hyperledger*) associada a uma camada que utiliza aprendizagem profunda (*deep learning*) para ajudar na sugestão de medicamentos a partir de informações de *web sites* farmacêuticos bem conceituados, que apontam informações como os usos dos medicamentos e a avaliação dos mesmos. As técnicas utilizadas são o processamento de linguagem natural e o algoritmo *Lightbgm* para recomendas os medicamentos mais adequados aos pacientes. Os autores obtiveram resultados satisfatórios em seus testes, que demonstraram que o sistema se mantém estável com o aumento de usuários e eficiente na recomendação dos medicamentos.

- *Efficient privacy-preserving machine learning for blockchain network*

  Kim *et al.* (2019) utilizou um método de aprendizado de máquina distribuído, onde um nó computa os pesos globais para o algoritmo, obtidos a partir dos resultados dos outros nós e distribui os pesos para a rede. Esse método é utilizado para assegurar a privacidade de redes *blockchain* (*Hyperledger*) e prevenir ataques que busquem atrapalhar a acurácia do modelo. O modelo possui pesos globais para a agregação dos dados e mostrou ser eficiente nos casos de ataques adversariais, mesmo com um caso onde 30% dos nós eram maliciosos, o processamento e a classificação dos dados não foram prejudicados.

- *Descentralizing privacy: Using blockchain to protect personal data*

  O trabalho de Zyskind, Nathan e Pentland (2015) buscou combinar a utilização de uma rede *blockchain* integrada a um serviço exterior à rede, onde o serviço auxilia no acesso à rede e na consulta dos dados de acordo com permissões definidas pelo usuário, a fim de garantir a segurança e a privacidade dos mesmos, considerando a parte externa à rede, que não é descentralizada.

- *MedRec: Using blockchain for medical data access and permission management*

  Azaria *et al.* (2016) propuseram um sistema descentralizado para gerenciamento de registros médicos utilizando uma rede *blockchain*, onde os blocos indicam as informações

entre os pacientes e os provedores das informações, com as informações específicas das permissões de acesso, instruções para obtenção dos dados, que estão presentes em bancos de dados externos e um *hash* dos registros para que a integridade possa ser garantida. Os autores demonstraram assim que é possível utilizar uma rede *blockchain* para tratar dados em larga escala de maneira a permitir que os registros sejam auditáveis e seguros.

- *Supporting private data on hyperledger fabric with secure multiparty computation*

  Benhamouda, Halevi e Halevi (2018) propuseram a utilização de Computação Multiparte Segura para garantir a privacidade de dados dentro da arquitetura de uma rede blockchain implementada na plataforma *Hyperledger Fabric*. Os autores implementaram um protocolo de Computação Multiparte Segura para garantir que, ao mesmo tempo que a computação dos dados possa tratar todos os dados da rede (como no caso da detecção de anomalias), cada parte tem acesso aos dados que lhe são permitidos.

- *Connected Blockchain Federations for Sharing Electronic Health Records*

  Hashim, Shuaib e Sallabi (2022) propuseram uma arquitetura que busca integrar várias redes *blockchain* voltadas à área médica independentes e que podem utilizar lógica de negócio própria para cada uma, delegando as funções aos contratos inteligentes. A arquitetura é integrada a um armazenamento externo em um sistema de arquivos interplanetário, o que fornece um armazenamento externo à rede que ainda assim seja decentralizado. Os autores utilizaram a ferramenta *Hyperledger Fabric* para construir as redes de teste e, pelos seus resultados, puderam concluir que o sistema proposto tem um desempenho melhor comparado com um trabalho similar, diminuindo o tempo das transferências.

3.2.1.3 Trabalhos relacionados ao uso de aprendizagem federada e redes blockchain

- *A blockchain-based decentralized federated learning framework with committee consensus*

  Li *et al.* (2021) propuseram uma arquitetura de aprendizagem federada auxiliada por *blockchain* que consiste na utilização de um comitê, no qual são efetuadas as validações da rede por meio de um protocolo de consenso a fim de resolver os problemas de segurança que uma rede tradicional apresentaria. No fim do estudo, os autores puderam notar que a solução apresentada conseguiu prevenir ataques externos, como servidores ou pares que poderiam fornecer entradas maliciosas à rede de aprendizagem federada.

- *Crowdsfl: A secure crowd computing framework based on blockchain and federated learning*

  Li *et al.* (2020) propuseram uma arquitetura que se utiliza de *blockchain* e aprendizagem federada para resolver problemas de ataques a redes de *crowdsourcing* na área empresarial, onde partes do trabalho são terceirizadas ao público externo. Os autores utilizaram a aprendizagem e a rede descentralizada para classificar os trabalhadores em questão de sua

qualidade e também tendo em vista a distribuição de recompensas para os participantes, tendo como conclusão que essa distribuição pode ainda ser melhorada, embora os resultados encontrados tenham medidores de acurácia melhores que os trabalhos semelhantes.

- *Biscotti: A blockchain system for private and secure federated learning*

  Shayan *et al.* (2021) apresentou um sistema que integra *blockchain* e aprendizagem federada para resolver problemas na integração dessas tecnologias, como ataques maliciosos na rede, introdução de dados anômalos e vazamento de dados. Para isso, os autores integraram em uma arquitetura melhorias já estabelecidas, concluindo que a mesma foi a primeira a prover segurança no nível da rede sem precisar de uma parte centralizada ou parte específica para garantir a segurança da rede.

- *Blockchain technology and neural networks for the internet of medical things*

  Połap *et al.* (2020) propôs um sistema para compartilhamento de dados médicos utilizando aprendizagem federada e *blockchain*. Os autores utilizaram o aprendizado de máquina para analisar resultados de exames e ajudar os profissionais de saúde na análise de dados novos e na prevenção de possíveis futuras doenças baseadas nos dados atuais dos pacientes. Os autores se utilizaram de um servidor para agregar os resultados obtidos dos mais diversos dispositivos ligados à rede médica específica, e concluíram ao final de seu estudo que o modelo é benéfico no sentido de poder ter os modelos salvos independentemente de dados permanecerem armazenados nos dispositivos, porém ainda não puderam conduzir experimentos mais avançados com sua solução.

- *Baffle: Blockchain based aggregator free federated learning*

  Em Ramanan e Nakayama (2020), foi introduzido um modelo de aprendizagem federada assistido por *blockchain* que não possui uma unidade agregadora. Ao invés do modelo comum presente nas redes de aprendizagem federada, os autores utilizaram contratos inteligentes para realizar as operações de agregação, assim como a avaliação dos pesos dos modelos, a atualização do modelo global e a definição dos turnos. Os autores realizaram testes em um ambiente de renda de motoristas de taxi e puderam concluir que sua utilização foi efetiva para diminuir o custo das transações (uma vez que os dados são separados em partes) e uma estabilidade maior que outros modelos da literatura que foram utilizados como base para comparação.

- *Blockchain and federated learning-based distributed computing defence framework for sustainable society*

  Sharma, Park e Cho (2020) apresentaram uma arquitetura que utiliza aprendizagem federada e uma rede *blockchain* para a área de defesa computacional. Para garantir a segurança dos dados, os autores segregaram-lhes em organizações de acordo com o acesso e utilizaram uma camada de névoa (*fog layer*). Essa camada protege o acesso aos dados e

está otimizada a fim de selecionar algumas partes dos dados para seu treinamento com a finalidade de deixar o processamento de grande quantidade de dados menos custoso e mais performático. Os autores chegaram à conclusão de que o resultado obtido aponta que o modelo possui uma grande acurácia e que ele se mostrou privativo devido ao não compartilhamento dos dados de treinamento, porém apontaram uma limitação em relação aos incentivos fornecidos aos pares, uma vez que esse não foi um tópico de foco dos mesmos.

- *Blockchain-based federated learning for device failure detection in industrial IOT*

  Zhang *et al.* (2021) propuseram uma rede de aprendizagem federada assistida por *blockchain* para ambientes da Internet das Coisas, a fim de detectar falhas nos equipamentos. A arquitetura dividiu os pares das organizações entre os dispositivos presentes em nível de aplicação (clientes) e uma organização central que comanda o funcionamento da plataforma. Os autores concluíram que a solução foi eficaz na sua proposta.

- *Communication-efficient federated learning and permissioned blockchain for digital twin edge networks*

  Em Lu *et al.* (2021), os autores apresentaram uma rede *blockchain* permissionada aliada à utilização de aprendizagem federada para redes de gêmeos digitais. A aprendizagem é feita em nível dos clientes, onde os dados são coletados, validados e treinados. Após esse processo, os modelos são verificados e enviados para os servidores da rede para serem lá agregados. Os autores puderam concluir a partir dos resultados obtidos que a arquitetura teve uma eficiência maior do que o algoritmo base comparado e uma eficácia comparável, mantendo a segurança, sendo uma solução para as limitações de privacidade de dados da área.

- *Flchain: A blockchain for auditable federated learning with trust and incentive*

  Em Bao *et al.* (2019) propuseram a arquitetura *FLChain*, que utiliza uma rede *blockchain* para gerenciar redes de aprendizagem federada, assim buscando evitar o uso malicioso. A rede consegue garantir essa segurança por meio da seleção dos pares para treinarem na rede a partir de uma reputação de confiabilidade e intenção obtida a partir de avaliação realizada na rede. Os pares podem obter melhores avaliações na rede a partir de maior quantidade de recursos no treinamento também. Os gradientes obtidos no treinamento são mascarados localmente, o que garante a privacidade dos mesmos no processo. Os autores não fizeram uma comparação quantitativa com outras propostas, porém reafirmaram os benefícios positivos de sua arquitetura.

- *Vfchain: Enabling verifiable and auditable federated learning via blockchain systems*

  No trabalho de Peng *et al.* (2022), é apresentada uma arquitetura que utiliza de *blockchain* para auxiliar na auditoria de uma rede de aprendizagem federada. Um comitê de pares

confiáveis é responsável por coordenar a verificação e aprovação dos modelos submetidos pelos nós trabalhadores da rede ao modelo global. Todo esse processo é feito de forma descentralizada. Ao fim de seu trabalho, os autores concluíram que a arquitetura foi efetiva analisando os resultados obtidos em seus testes.

- *A blockchain-based federated learning method for smart healthcare*

  O trabalho de Chang, Fang e Sun (2021) propôs a utilização de aprendizagem federada assistida por uma rede *blockchain* visando as necessidades da Internet das Coisas voltada à área dos sistemas médicos. O sistema é dividido em camadas, tendo a camada dos usuários que contém os dispositivos que fornecem os dados e a camada dos nós da borda que realizam as tarefas dos servidores de aprendizagem federada, agregando os modelos. Além da contribuição relacionada à integração, os autores propuseram um algoritmo de privacidade diferencial onde é criado ruído de acordo com os processos das etapas do treinamento a fim de garantir maior privacidade dos dados. Os autores concluíram que a solução teve resultados de acurácia similares ao de uma rede de aprendizagem federada tradicional com um desempenho de tempo aceitável.

- *A framework for privacy-preservation of iot healthcare data using federated learning and blockchain technology*

  Em Singh *et al.* (2022) é proposta a utilização conjunta de aprendizagem federada e *blockchain* para redes inteligentes na área médica. Os sensores são responsáveis por fornecer os dados à rede, que por sua vez, executada na nuvem, processa os dados que são consumidos pelos dispositivos finais, sendo cada um desses grupos separados na arquitetura. A aprendizagem federada é utilizada para evitar que os dados sensíveis sejam enviados a um servidor para serem utilizados no treinamento, assim garantindo a segurança dos dados. Estes são armazenados externamente à rede, sendo seu acesso administrado pelo sistema proposto em relação a acesso, demarcação e concessão de permissões. Ao fim do estudo, os autores puderam concluir que cada uma das tecnologias utilizadas teve um benefício, como a rede *blockchain* auxiliado nas mudanças de escala da rede e na comunicação entre os pares, a utilização de um armazenamento na nuvem auxiliou na comunicação da rede, tendo como limitação seu desempenho depender do protocolo que une a rede *blockchain* à aprendizagem federada, porém apresentaram uma solução que seria fazer um modelo de confiabilidade na rede e um novo mecanismo de consenso.

- *Protecting personal healthcare record using blockchain federated learning technologies*

  Os autores em Aich *et al.* (2021) propuseram uma arquitetura que usa aprendizagem federada assistida por *blockchain* para proteger dados de sistemas médicos eletrônicos, sendo o acesso gerenciado pela plataforma e possuindo a aprendizagem federada um servidor para realizar a agregação dos dados. O trabalho, conforme apontado pelos autores,

não havia sido desenvolvido e testado no momento de sua escrita, assim não podendo ser comparável em relação a seu desempenho quantitativo.

- *Federated-learning based privacy preservation and fraud-enabled blockchain iomt system for healthcare*

  Lakhan *et al.* (2022) apresentaram uma arquitetura que une uma rede *blockchain* ao uso de aprendizagem federada para organizar e prover segurança contra fraudes a um cenário de Internet das Coisas em sistemas de registros médicos. Na rede, os dados são fornecidos pelos dispositivos e treinados por nós que compartilham os dados. O sistema, por sua vez realiza a ordenação e agenda os envios. A arquitetura realiza o treino nos nós utilizando diferentes modelos a fim de serem compartilhados com rede para serem computados, também aplicando o conceito de névoa a fim de garantir segurança e privacidade aos dados. Ao fim de seus experimentos, os autores concluíram que a arquitetura era mais eficiente em uso de energia e tempo do que as outras comparadas, porém durante o desenvolvimento não foram considerados algumas ameaças à rede, como ataques dinâmicos e em tempo de execução.

- *Federated blockchain system (FBS) for the healthcare industry*

  Em Eldin *et al.* (2023), os autores propuseram uma arquitetura que faz uso de redes *blockchain* para integrar organizações médicas de maneira permissionada com a finalidade de gerenciamento e compartilhamento de registros médicos eletrônicos. É utilizado armazenamento em nuvem devido ao tamanho dos recursos e o gerenciamento da identidade dos pacientes é único, sendo gerado em um cartão físico que possui sua chave privada. Os autores concluíram que o sistema proposto foi eficiente em seu desempenho, porém não apresentaram muitas informações sobre o ambiente em que os testes foram realizados e qual a sua solução para outros problemas presentes.

- *IoMT: A COVID-19 Healthcare System driven by Federated Learning and Blockchain*

  Samuel *et al.* (2022) propôs uma rede de aprendizagem federada assistida por uma rede *blockchain* de consórcio para a solução de problemas com redes que auxiliam no compartilhamento de informações sobre o vírus da COVID-19 e os dados dos pacientes. A rede de aprendizagem é executada de maneira paralela à rede *blockchain*, guardando os modelos. Após a coleta na superfície da rede, os dados são tratados para remoção dos incompletos ou errados e são salvos para o treinamento. O sistema separa as partes de acordo com sua confiabilidade. Os autores puderam concluir a partir de seus testes que seu projeto foi o primeiro a solucionar os problemas para que foi proposto.

- *FIDChain: Federated Intrusion Detection System for Blockchain-Enabled IoT Healthcare Applications*

  Ashraf *et al.* (2022) propuseram um mecanismo que integra redes *blockchain* a aprendizagem federada para prevenção de ataques de envenenamento de redes médicas voltadas à

área de Internet das Coisas. Eles utilizaram redes neurais artificiais e separaram a arquitetura em camadas, sendo uma para a obtenção dos dados pelos dispositivos, outra para os locais onde esses dados são verificados e normalizados, uma para o tráfego de dados, uma para a agregação dos pesos dos modelos e outras duas para a aplicação e para a lógica de negócio. Os autores puderam concluir que sua solução foi capaz de evitar ataques de envenenamento e teve desempenho melhor que os trabalhos comparados, tendo também uma maior velocidade para detecção de ataques devido a essa funcionalidade ficar presente numa camada mais local da rede.

- *A secure healthcare 5.0 system based on blockchain technology entangled with federated learning technique*

  Os autores em Rehman *et al.* (2022) propuseram a utilização da integração entre redes *blockchain* e aprendizagem federada para resolver ataques de invasão e para a predição de doenças em um contexto de Internet das Coisas. Os autores focaram nos dados de sensores, o que diferenciaram dos trabalhos correlatos. A rede agrega múltiplos dados a fim de extrair informações relevantes à sua aplicação, o que reduz problemas como repetição e falhas na informação. Os autores concluíram que o sistema foi eficiente, porém afirmam que o sistema é limitado pela quantidade de camadas invisíveis e que os fatores que impactam em seus resultados precisam de maior foco para trabalhos futuros.

- *Blockchain-based healthcare workflow for tele-medical laboratory in federated hospital IoT clouds*

  Os autores em Celesti *et al.* (2020) propuseram uma rede para integrar organizações médicas por meio do uso de *blockchain*. Os autores realizaram a integração da rede com armazenamento em nuvem para guardar os registros e outras informações clínicas. Por meio de um módulo de anonimização dos dados, os autores desassociam os dados de maneira a não permitir que os pacientes sejam discriminados. Dessa forma, os registros médicos possuem informações que apontam para os dados médicos, que por sua vez não possuem informações que individualizam o paciente. Os autores realizaram testes em uma rede pública e em uma rede híbrida *Ethereum* e concluíram que a híbrida teve melhor desempenho, embora não tenham se aprofundado muito na avaliação da rede, citando mais os benefícios que a área médica pode obter com seu estudo, como seu aumento promissor em robustez, desempenho e escalabilidade.

- *DITrust Chain: Towards Blockchain-Based Trust Models for Sustainable Healthcare IoT Systems*

  Em Abou-Nassar *et al.* (2020), os autores propuseram uma arquitetura que se utiliza de dois tipos de redes *blockchain* num contexto de redes médicas voltado à Internet das Coisas. A rede pública é responsável por ligar os objetos e a rede *ripple* é utilizada para limitar o acesso às informações de maneira permissionada. Os nós são divididos em zonas e nelas

são escolhidos pares confiáveis para realizar o consenso e a comunicação entre pares pode ser separada em associativa (entre pares da mesma zona) e agregativa. As requisições de agregação passam por validação a mais a fim de garantir a segurança dos dados. Ao fim de seu trabalho, os autores concluíram que seu trabalho apresentou melhor desempenho do que outros que foram comparados em escalabilidade, disponibilidade, interoperabilidade, autenticação, confiança, integridade dos dados, privacidade e confidencialidade.

- *When Collaborative Federated Learning Meets Blockchain to Preserve Privacy in Healthcare*

  Os autores de Houda *et al.* (2022) propuseram uma arquitetura chamada *HealthFed*, que integra uma rede de aprendizagem federada à *blockchain* para aplicações médicas. A arquitetura é separada em três planos, um para definir as regras da comunicação, um para colocá-las em prática e outro para a aplicação. São utilizados contratos inteligentes para adicionar e remover colaboradores à rede e a credibilidade dos pares é gerenciada nos registros por meio deles armazenados. Para evitar ataques de engenharia reversa dos dados, os autores utilizaram computação segura multiparte para garantir que os modelos locais possam ser divididos em várias partes, Os autores utilizaram o exemplo de dados de exames de câncer de mama para seus testes e concluíram por meio de testes quantitativos relacionados ao aprendizado de máquina que seu modelo possui melhores indicativos que outras soluções centralizadas.

- *A blockchain-orchestrated Federated Learning architecture for healthcare consortia*

  Passerat-Palmbach *et al.* (2019) propuseram uma arquitetura que utiliza de *blockchain* e aprendizagem federada para o tratamento de dados médicos. Para isso, apresentaram algumas soluções para problemas de segurança e privacidade, como a seleção aleatória de atualizações do modelo em cada rodada e a utilização de criptografia para trânsito dos dados e via *hardware*. Os autores também propuseram que a rede fosse auditável de maneira a preservar a privacidade dos dados. Apesar de suas propostas, os autores não apresentaram resultados experimentais para apoiar suas afirmações.

- *Trustworthy Privacy-preserving Hierarchical Ensemble and Federated Learning in Healthcare 4.0 with Blockchain*

  Stephanie *et al.* (2022) propuseram uma arquitetura que utiliza *blockchain* e aprendizagem federada a fim de lidar com dados heterogêneos de imagens em um contexto médico. Os autores dividiram a arquitetura entre organizações que têm redes *blockchain* particulares e que podem colaborar umas com as outras por meio de um protocolo seguro. Os autores testaram sua solução e compararam-na com soluções já existentes, provando ser melhor em desempenho do modelo e perceberam uma desvantagem na eficiência do modelo devido ao algoritmo escolhido, que requer maior computação. Além disso, os autores notaram

que consideraram a hetereogeneidade dos dados e transações apenas entre hospitais, e não entre os participantes de um hospital.

- *QFBN: Quorum Based Federated Blockchain Network for Healthcare System to Avoid Multiple Benefits and Data Breaches*

  Settipalli e Gangadharan (2022) propuseram uma arquitetura de rede feita tendo em vista a plataforma *Quorum* para integrar sistemas de planos de saúde a fim de evitar fraudes em que a mesma requisição é feita a diversas organizações diferentes, além de combater o vazamento de dados. Os autores integraram redes públicas com privadas. Os clientes possuem as privadas, onde os usuários podem realizar a aplicação. Por sua vez, o cliente é ligado a um nó distribuído por meio de um gerenciador de conexão, por meio do qual os dados são trafegados entre os clientes. As informações são armazenadas por meio das transações de maneira com que não possa haver duplicação das requisições, assim evitando fraudes.

- *Blockchain Management and Federated Learning Adaptation on Healthcare Management System*

  Turgay (2022) propôs uma arquitetura que une o uso de *blockchain* e aprendizagem federada para sistemas de gerenciamento médico. Os autores separaram a rede em camadas. A arquitetura integra as tecnologias de maneira flexível, sendo o livro-razão utilizado para armazenar os dados e a aprendizagem federada utilizada para assistir os sistemas médicos. Os nós mineradores são responsáveis por validar as atualizações do modelo. Os autores realizaram testes no desempenho da rede e na aprendizagem, indicando que o modelo é eficiente na proteção dos dados e de sua privacidade e confidencialidade, além de seu processamento e treinamento do modelo.

- *Federated Learning and Blockchain-Enabled Fog-IoT Platform for Wearables in Predictive Healthcare*

  Baucas, Spachos e Plataniotis (2023) propuseram uma solução que integra *blockchain* e aprendizagem federada utilizada para redes da Internet das Coisas com névoa em um contexto médico. O aprendizado é utilizado para realizar predições relacionadas ao monitoramento de informações de saúde dos pacientes por meio de dispositivos eletrônicos. Os autores utilizaram redes *blockchain* privadas para armazenar informações sobre os modelos e operações do sistema, utilizando o livro-razão para auditoria do sistema. Ao fim de seus experimentos, o trabalho propõe que seu modelo possui integridade e é efetivo mesmo em condições de dispositivos com menores capacidades computacionais, tratando com maior ênfase nos resultados qualitativos.

- *Detecting model misconducts in decentralized healthcare federated learning*

  Os autores em Kuo e Pham (2022) propuseram um modelo de aprendizagem federada descentralizado integrado com *blockchain* para detecção de usos indevidos em redes

médicas, realizando seus experimentos mais especificamente nas responsáveis por detectar problemas de saúde dos pacientes. O foco de seu trabalho foi a aprendizagem federada, sendo a rede *blockchain* utilizada para assistir no gerenciamento dos dados, implementando o algoritmo GloreChain (KUO; GABRIEL; OHNO-MACHADO, 2019 apud KUO; PHAM, 2022). Ao fim de seus trabalhos, os autores concluíram que o modelo apresentou um indicador de revocação alto, além de se mostrar pouco custoso e garantir a integridade e confiabilidade da aprendizagem.

3.2.2 Comparação com a pesquisa

A fim de facilitar a comparação entre os trabalhos apresentados e o presente estudo, foram separadas três categorias de trabalhos: os que se assemelham no uso de aprendizado de máquina na área médica, os que se assemelham no uso de redes *blockchain* com ou sem a associação ao aprendizado de máquina dando enfoque nos sistemas médicos, e, por fim, os trabalhos que utilizaram aprendizagem federada aliada ao conceito de *blockchain*, seja na área médica ou não, sendo essas duas últimas categorias escolhidas baseando-se, além da área explorada, na semelhança dos problemas tratados com os presentes problemas aqui apresentados a fim de facilitar a comparação.

3.2.2.1 Semelhança no aprendizado de máquina

Os trabalhos de Seo e Mendelevitch (2017), Bauder, Da Rosa e Khoshgoftaar (2018) e Yashraj *et al.* (2021) apresentam soluções para a situação das possíveis fraudes em serviços de registros médicos, tendo relação com a área de prescrições médicas, porém não apresentam uma solução para os problemas presentes nas redes tradicionais, como o ponto único de falha e a proveniente susceptibilidade a ataques e falhas. Além disso, apenas os trabalhos de Johnson e Khoshgoftaar (2019) e Yashraj *et al.* (2021) trataram sobre o desbalanceamento das classes, recorrente nesse setor onde as transações fraudulentas serão muito menos frequentes que as transações válidas.

A relação geral dos trabalhos apresentados nesta subseção com o presente estudo é apresentada no Quadro 1.

3.2.2.2 Semelhança nas redes blockchain

Em relação aos trabalhos que utilizaram redes *blockchain*, pode-se citar as semelhanças entre as propostas de Mackey *et al.* (2020), Abbas *et al.* (2020), Azaria *et al.* (2016), que apresentaram as redes descentralizadas como uma solução para o problema do ponto de falha único nas redes tradicionais. Além disso, pode-se citar a semelhança na utilização de algoritmos de aprendizado de máquina aliados à redes *blockchain* no trabalho de Kim *et al.* (2019), embora a sua proposta seja com um fim distinto, que é a preservação da privacidade e segurança dos dados, como citado através da prevenção de ataques adversariais. O trabalho de Hashim, Shuaib

Quadro 1 – Comparação entre os trabalhos relacionados a aprendizado de máquina

Trabalho	Semelhanças	Diferenças e limitações
Identifying frauds and anomalies in Medicare-B dataset	Identificação de fraudes e anomalias em registros médicos	Não apresenta uma solução para o problema de centralidade das redes tradicionais, não apresenta enfoque no tratamento do desbalanceamento de dados
Identifying medicare provider fraud with unsupervised machine learning	Identificação de fraudes em registros médicos com algoritmos de aprendizado de máquina não supervisionado	Não apresenta uma solução para o problema de centralidade das redes tradicionais
Medicare fraud detection using neural networks	Identificação de fraudes em registros médicos com algoritmos de aprendizado de máquina	Não apresenta uma solução para o problema de centralidade das redes tradicionais
Comparative Study of Using Various Machine Learning and Deep Learning-Based Fraud Detection Models For Universal Health Coverage Schemes	Identificação de fraudes em registros médicos com algoritmos de aprendizado de máquina, tratando do problema de desbalanceamento de classes	Não apresenta uma solução para o problema de centralidade das redes tradicionais
Medical fraud and abuse detection system based on machine learning	Identificação de fraudes em registros médicos com algoritmos de aprendizado de máquina	Não apresenta uma solução para o problema de centralidade das redes tradicionais

e Sallabi (2022) utilizou o conceito de redes *blockchain* federadas para compartilhamento de registros médicos entre várias organizações, o que se assemelha ao presente estudo porém difere em outros pontos, como a menor preocupação com privacidade e combate a fraudes e a não utilização de aprendizado de máquina.

A relação geral dos trabalhos apresentados nesta subseção com o presente estudo é apresentada no Quadro 2.

3.2.2.3 Semelhança na utilização de redes blockchain e aprendizagem federada

Os trabalhos de Li *et al.* (2021), Li *et al.* (2020), Shayan *et al.* (2021), Połap *et al.* (2020), Ramanan e Nakayama (2020), Bao *et al.* (2019), Sharma, Park e Cho (2020), Zhang *et al.* (2021), Lu *et al.* (2021), Peng *et al.* (2022) são similares na proposta de uma arquitetura que utiliza *blockchain* e aprendizagem federada. Todavia, elas não são focadas na área de registros médicos eletrônicos, sendo relacionadas a soluções de maneira mais genéricas ou em outra área. Dessa forma, pode-se comparar apenas sua implementação e estrutura, abstraindo-as das necessidades

Quadro 2 – Comparação entre os trabalhos relacionados a *blockchain*

Trabalho	**Semelhanças**	**Diferenças e limitações**
Blockchain paradigm for healthcare: Performance evaluation	Investigação dos benefícios de desempenho da utilização de redes *blockchain* na área médica	Não apresenta uma solução para os problemas de proteção de dados e detecção de fraudes
Combating health care fraud and abuse: Conceptualization and prototyping study of a blockchain antifraud framework	Utilização de rede *blockchain* para o compartilhamento de registros médicos eletrônicos	Não apresenta uma solução para os problemas de detecção de fraudes
A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry	Utilização de rede *blockchain* na área médica	Não trata sobre o compartilhamento de registros médicos eletrônicos
Efficient privacy-preserving machine learning for blockchain network	Utilização de rede *blockchain* e aprendizado de máquina para solucionar problemas de sistemas tradicionais	Enfoque na privacidade dos dados na utilização de aprendizado de máquina, não é específico da área da saúde
Descentralizing privacy: Using blockchain to protect personal data	Utilização de rede *blockchain* e solução para a proteção de dados sensíveis	Enfoque na privacidade dos dados, não é específico da área da saúde
MedRec: Using blockchain for medical data access and permission management	Utilização de rede *blockchain* para o compartilhamento de registros médicos eletrônicos, enfoque na segurança dos dados	Não apresenta uma solução para os problemas de detecção de fraudes
Supporting private data on hyperledger fabric with secure multiparty computation	Utilização de redes *blockchain* com enfoque na segurança dos dados pessoais	Enfoque na privacidade dos dados, não é específico da área da saúde
Connected Blockchain Federations for Sharing Electronic Health Records	Uso de blockchain para resolver problemas na área médica, busca por ser uma arquitetura que une várias organizações, é similar a uma das soluções por utilizar um banco de dados externo	Não utiliza aprendizado de máquina, não possui foco na privacidade e combate a fraudes, é voltada para a comunicação entre várias redes blockchain

específicas da área médica.

Já as pesquisas de Chang, Fang e Sun (2021), Singh *et al.* (2022), Aich *et al.* (2021), Lakhan *et al.* (2022), Eldin *et al.* (2023), Samuel *et al.* (2022), Ashraf *et al.* (2022), Rehman *et al.* (2022), Celesti *et al.* (2020), Abou-Nassar *et al.* (2020), Houda *et al.* (2022), Passerat-Palmbach *et al.* (2019), Stephanie *et al.* (2022), Settipalli e Gangadharan (2022), Turgay (2022), Baucas, Spachos e Plataniotis (2023), Kuo e Pham (2022) são mais específicas à área da saúde de maneira ampla, assistida por redes *blockchain* e aprendizagem federada. Apesar de sua semelhança na área médica e na utilização das tecnologias, a maioria delas não focou na detecção de fraudes. Dentre as que tiveram esse foco e as demais, nenhuma das propostas encontradas foi capaz de propor uma arquitetura segura contra ataques, que preservasse a privacidade, com desempenho considerável em relação às redes tradicionais e que seja agnóstico em relação às tecnologias utilizadas, interligando instituições médicas com a utilização de uma rede *blockchain*.

A comparação detalhada de cada trabalho é apresentada nos Quadros 3, 4, 5 e 6. Como pode-se perceber por meio das informações e comparações, muitos dos trabalhos utilizaram aprendizagem federada para auxiliar em processos médicos como a identificação de doenças e análise de informações obtidas por meio de monitoramento médico. Além disso, a busca pela utilização das tecnologias apresentadas com preocupações relacionadas a segurança e/ou privacidade não esteve presente em todos os trabalhos, porém está em parte considerável, ao menos parcialmente. É também notável que foram variadas as soluções para esses requerimentos, algumas baseando-se em métodos de segurança, outras na configuração da rede ou em como a arquitetura foi planejada, o que torna essas necessidades como algo que pode ser abordado de diversas maneiras e com diferentes preocupações a se observar, como custo computacional, impacto na usabilidade, impacto no desempenho e acessibilidade, por exemplo.

3.3 Definição da lacuna

Assim como mostrado na seção 3.2, nenhum dos trabalhos apresentou uma proposta para os problemas apresentados mais especificamente na área das prescrições médicas e que tivesse o propósito de ser independente de uma tecnologia, além de prover segurança e preservar a privacidade dos usuários.

Nesse contexto, a pesquisa aqui apresentada busca propor o uso de aprendizagem federada integrada de maneira flexível a uma rede *blockchain* a fim de impedir ameaças à integridade, privacidade, segurança e confiabilidade dos dados e da rede, o que não pôde ser identificado na literatura após a revisão bibliográfica de maneira integral.

Além disso, o presente estudo pretende dar exposição a tecnologias que são cada vez mais emergentes e que podem ser exploradas mais amplamente no futuro, apresentando uma perspectiva de melhoria contínua. Com o benefício de ser independente de tecnologia, também é beneficiada a aplicabilidade da arquitetura apresentada em diferentes ambientes com necessidades

distintas.

Quadro 3 – Comparação entre os trabalhos relacionados a blockchain e aprendizagem federada - parte 1

Trabalho	Semelhanças	Diferenças e limitações
A blockchain-based decentralized federated learning framework with committee consensus	Utilização de rede blockchain e aprendizagem federada para resolver problemas na segurança dos dados	Utiliza a separação de um comitê para a validação e não é focado na área médica
Crowdsfl: A secure crowd computing framework based on blockchain and federated learning	Utilização de rede blockchain e aprendizagem federada, preocupação com a segurança	Área diferente, classificação utilizada para outro fim, apresenta limitação no incentivo à participação na rede
Biscotti: A blockchain system for private and secure federated learning	Integração de blockchain e aprendizagem federada, preocupação com segurança e privacidade	Não tem uma área específica em seu desenvolvimento, assim não sendo focada na detecção de fraudes
Blockchain technology and neural networks for the internet of medical things	Uso de blockchain e aprendizagem federada na área médica, presença de servidor para agregar modelos	Focado na parte de Internet das Coisas, limitação nos experimentos e não é focado em segurança/privacidade
Baffle: Blockchain based aggregator free federated learning	Uso de blockchain e aprendizagem federada, preocupação com a descentralização, utiliza contratos inteligentes	Não tem uma área específica, não tem grande enfoque na segurança dos dados, a agregação é coordenada por meio dos contratos inteligentes
Blockchain and federated learning-based distributed computing defence framework for sustainable society	Uso de blockchain e aprendizagem federada, preocupação com a segurança e privacidade dos dados	É de diferente área, possui problemas quanto ao incentivo, diferentes necessidades em relação à proteção de dados devido à área de defesa
Blockchain-based federated learning for device failure detection in industrial iot	Uso de blockchain e aprendizagem federada, detecção de falhas	Difererente área, não envolve anomalias causadas de maneira maliciosa
Communication-efficient federated learning and permissioned blockchain for digital twin edge networks	Uso de blockchain e aprendizagem federada, presença de servidores para a agregação dos modelos e preocupação com a segurança	Diferente área, aprendizagem federada não é utilizada para detecção de anomalias, mas somente para garantir a privacidade dos dados
Flchain: A blockchain for auditable federated learning with trust and incentive	Uso de blockchain e aprendizagem federada, preocupação com a segurança e privacidade dos dados e com uso malicioso da rede	Não possui uma área específica, rede blockchain é utilizada de forma principal para auxiliar na aprendizagem federada e apresenta limitações na análise comparativa com a literatura

Quadro 4 – Comparação entre os trabalhos relacionados a blockchain e aprendizagem federada - parte 2

Trabalho	**Semelhanças**	**Diferenças e limitações**
Vfchain: Enabling verifiable and auditable federated learning via blockchain systems	Uso de blockchain e aprendizagem federada, preocupação com a segurança dos dados	Não possui uma área específica, rede blockchain é utilizada de forma principal para auxiliar na aprendizagem federada, agregação é feita de maneira diferente
A blockchain-based federated learning method for smart healthcare	Uso de blockchain e aprendizagem federada, área médica, preocupação com a privacidade dos dados	É voltada à área de Internet das Coisas, aprendizagem federada somente busca detectar falhas nos dispositivos
A framework for privacy-preservation of iot healthcare data using federated learning and blockchain technology	Uso de blockchain e aprendizagem federada, área médica, preocupação com a segurança e privacidade dos dados, presença de um armazenamento externo	É voltada à área de Internet das Coisas, aprendizagem federada é utilizada para outra finalidade que não identificação de anomalias, apresentou limitação no desempenho
Protecting personal healthcare record using blockchain federated learning technologies	Uso de blockchain e aprendizagem federada, área médica, preocupação com a segurança e privacidade dos dados	Não foi desenvolvida no período de sua implementação, pouco detalhamento e ausência de resultados quantitativos e qualitativos
Federated-learning based privacy preservation and fraud-enabled blockchain iomt system for healthcare	Uso de blockchain e aprendizagem federada, área médica, preocupação com a segurança e privacidade dos dados	É voltada à área de Internet das Coisas, não levou em conta algumas vulnerabilidades em sua implementação, busca encontrar fraudes em provedores na área médica
Federated blockchain system (FBS) for the healthcare industry	Uso de blockchain para integrar várias instituições na área médica, preocupação com a segurança dos dados e é semelhante a uma das soluções por utilizar um armazenamento externo	Não utiliza aprendizado de máquina, não possui foco no combate a fraudes, utiliza meio físico (cartão com chave privada) para garantir a segurança, não possui definição de como foram realizados os testes e soluções para outros problemas apresentados

Quadro 5 – Comparação entre os trabalhos relacionados a blockchain e aprendizagem federada - parte 3

Trabalho	Semelhanças	Diferenças e limitações
IoMT: A COVID-19 Healthcare System driven by Federated Learning and Blockchain	Uso de blockchain e aprendizagem federada na área médica, preocupação com a confiabilidade	É focado no compartilhamento de dados referentes a COVID-19, não possui foco no combate a fraudes em prescrições, é voltado à área de Internet das Coisas
FIDChain: Federated Intrusion Detection System for Blockchain -Enabled IoT Healthcare Applications	Uso de blockchain e apredizagem federada na área médica, preocupação com a segurança da rede	É focado mais à prevenção de ataques de envenenamento da rede, é voltado à área de Internet das Coisas
A secure healthcare 5.0 system based on blockchain technology entangled with federated learning technique	Uso de blockchain e aprendizagem federada na área médica, preocupação com a segurança da rede	É voltado para a área de Internet das Coisas, há limitações de desempenho apresentadas pelos autores
Blockchain-based healthcare workflow for tele-medical laboratory in federated hospital IoT clouds	Uso de blockchain e aprendizagem federada na área médica, integração de várias organizações médicas, preocupação com a privacidade dos dados	Falta de mais avaliações quantitativas da arquitetura, não é focado na identificação de fraudes, utiliza múltiplas redes blockchain
DITrust Chain: Towards Blockchain-Based Trust Models for Sustainable Healthcare IoT Systems	Uso de blockchain e aprendizagem federada na área médica, integração de várias partes na comunicação dos dados, preocupação com a segurança dos dados	Utiliza múltiplas redes blockchain, não é focado na identificação de fraudes, é voltado à área de Internet das Coisas
When Collaborative Federated Learning Meets Blockchain to Preserve Privacy in Healthcare	Uso de blockchain e aprendizagem federada na área médica, preocupação com a segurança e privacidade dos dados, uso de contratos inteligentes	É voltado à área de Internet das Coisas, não é focado na identificação de fraudes
A blockchain-orchestrated Federated Learning architecture for healthcare consortia	Uso de blockchain e aprendizagem federada na área médica, preocupação com a segurança dos dados	Necessita de hardware específico, ausência de experimentos para avaliação quantitativa da arquitetura
Trustworthy Privacy-preserving Hierarchical Ensemble and Federated Learning in Healthcare 4.0 with Blockchain	Uso de blockchain e aprendizagem federada na área médica, preocupação com a segurança e privacidade dos dados	Utiliza múltiplas redes blockchain, não é focado na identificação de fraudes, possui limitações em relação à eficiência do modelo utilizado

Quadro 6 – Comparação entre os trabalhos relacionados a blockchain e aprendizagem federada - parte 4

Trabalho	**Semelhanças**	**Diferenças e limitações**
QFBN: Quorum Based Federated Blockchain Network for Healthcare System to Avoid Multiple Benefits and Data Breaches	Uso de blockchain e aprendizagem federada na área médica, identificação de fraudes	É voltado à área dos seguros de vida, utiliza múltiplas redes blockchain
Blockchain Management and Federated Learning Adaptation on Healthcare Management System	Uso de blockchain e aprendizagem federada na área médica, preocupação com a segurança e privacidade dos dados	Não é focado na identificação de fraudes, é voltado à área de sistemas de gerenciamento médico
Federated Learning and Blockchain-Enabled Fog-IoT Platform for Wearables in Predictive Healthcare	Uso de blockchain e aprendizagem federada na área médica, uso do livro-razão para auditoria, preocupação com a segurança e privacidade dos dados	Não é focado na identificação de fraudes, é voltado à área de Internet das Coisas, aprendizado de máquina é utilizado para realizar predições sobre as informações de saúde dos pacientes
Detecting model misconducts in decentralized healthcare federated learning	Uso de blockchain e aprendizagem federada na área médica, detecção de mau uso da rede	Experimentos focados na área de detecção de mau uso nos sistemas de detecção de problemas de saúde dos pacientes, pouco foco na rede blockchain

CAPÍTULO

4

MODELO

Neste capítulo será apresentada a proposta de pesquisa. Na seção 4.1 é feita a descrição do desenvolvimento dos estudos e dos experimentos e são explicados os métodos empregados na pesquisa. Na seção 4.2 são apresentados e discutidos todos os resultados obtidos nos experimentos práticos.

4.1 Visão geral e metodologia

Nesta sessão é descrito o desenvolvimento e a metodologia do presente trabalho. Na Seção 4.1.1 é demonstrada como foi feita a primeira parte da pesquisa, com a descrição da arquitetura de rede desenvolvida na ferramenta *Hyperledeger Fabric* para transmissão de dados médicos. Na Seção 4.1.2 é apresentada a pesquisa por uma base de dados e a base escolhida, e, por fim, na Seção 4.1.3 é descrita a terceira e última parte da pesquisa, onde é apresentada a proposta de arquitetura final onde é integrada uma rede de aprendizagem federada a uma rede *blockchain* desenvolvida com a ferramenta *Hyperledger Besu*.

4.1.1 Primeiros experimentos com Hyperledger Fabric

Inicialmente foram realizados testes separados com uma rede *blockchain* configurada com a ferramenta *Hyperledger Fabric*. A rede elaborada é permissionada e dois contratos inteligentes foram propostos, um para transações entre entidades médicas e pacientes e outra para transações na área farmacêutica. O fluxograma da rede é representado pela Figura 2, onde podem-se observar as três partes envolvidas (paciente, profissional de saúde e setor farmacêutico) onde são feitas consultas e propostas de transação para a rede *blockchain*, ainda não possuindo a definição da rede de aprendizagem federada definida. Os testes foram realizados em uma rede com 1 nó ordenador e 5 pares e em outra com 8 nós ordenadores e 5 pares e os resultados foram obtidos com a ferramenta *Hyperledger Caliper*.

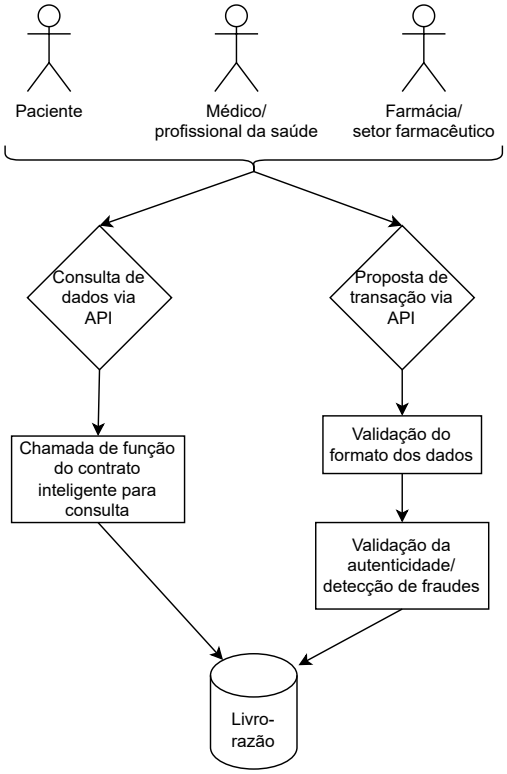

Figura 2 – Fluxograma da rede

4.1.2 Base de dados

Na primeira fase da pesquisa foram encontradas e selecionadas bases de dados que seriam de interesse para a pesquisa por meio de uma extensa pesquisa por bases de prescrições e registros médicos com a finalidade de identificação de prescrições anômalas. Para isso, utilizou-se de pesquisa em coleções de bases de dados na *internet* e contato com outros pesquisadores da área. Duas bases foram destacadas entre as demais:

- Base de dados de prescrições médicas de medicamentos do Brasil, do Hospital Nossa Senhora da Conceição (SC): <https://github.com/nlp-pucrs/prescription-outliers>

- Base de dados de prescrições médicas de medicamentos da Inglaterra, obtida por meio da junção de duas bases de dados de organizações inglesas: <https://opendata.nhsbsa.net/dataset/english-prescribing-data-epd>

A base de dados utilizada na pesquisa foi a primeira apresentada, correspondente às informações do Hospital Nossa Senhora da Conceição (SC), que possui 240.000 entradas de prescrições médicas datadas de Janeiro a Setembro de 2017 (SANTOS *et al.*, 2019). Os dados são descritos no Quadro 7. A escolha se deu devido à relevância dos dados para o estudo e a presença de rótulos, o que facilitaria a validação da classificação.

Quadro 7 – Descrição da base de dados utilizada nos testes

Coluna	Tipo de valor	Descrição
medication	Texto	O nome da medicação
frequency	Numérico	A frequência diária
dose	Numérico	A dose prescrita
target	Numérico (0 ou 1)	O rótulo da prescrição, sendo 0 em uma prescrição válida e 1 em uma com anomalias

4.1.3 Experimentos com Hyperledger Besu integrado a uma rede de aprendizagem federada

A estrutura e a solução do projeto foram posteriormente alteradas para utilizar aprendizagem federada supervisionada ao invés de aprendizado de máquina não supervisionado, o que se mostrou mais adequado devido às características da base de dados e do problema abordado, além de prover os benefícios da aprendizagem federada comparados à implementação anterior, que realizava o treinamento por parte dos responsáveis pelo processamento dos dados de maneira menos acoplada ao funcionamento geral da rede.

A implementação dos testes com o cliente *Hyperledger Besu* (FOUNDATION, 2023) deve-se à sua possibilidade de desenvolver redes públicas ou privadas além da maior adoção das redes *Ethereum* na literatura.

Foram propostas três arquiteturas similares porém com variações na escolha das informações armazenadas e no local de armazenamento a fim de comparar qual seria a abordagem ideal.

A rede é primeiramente implementada em um contexto maior (por exemplo, em um sistema público de saúde como o SUS). Organizações como hospitais, farmácias e laboratórios que produzem registros médicos podem se tornar clientes da rede juntando-se a ela e utilizando os contratos inteligentes referentes para executar transações na rede. A figura 3 apresenta a visão geral da proposta dentro de uma organização médica.

A partir daí, com os dados produzidos nesses estabelecimentos, são treinados os modelos locais para a identificação de fraudes e o modelo é enviado para o servidor agregador. A identificação de fraudes a partir do modelo construído pode ser implementada num conexto local e em um contexto global, a partir da necessidade.

Após definir as características comuns às implementações, segue-se a definição das particularidades de cada uma das soluções:

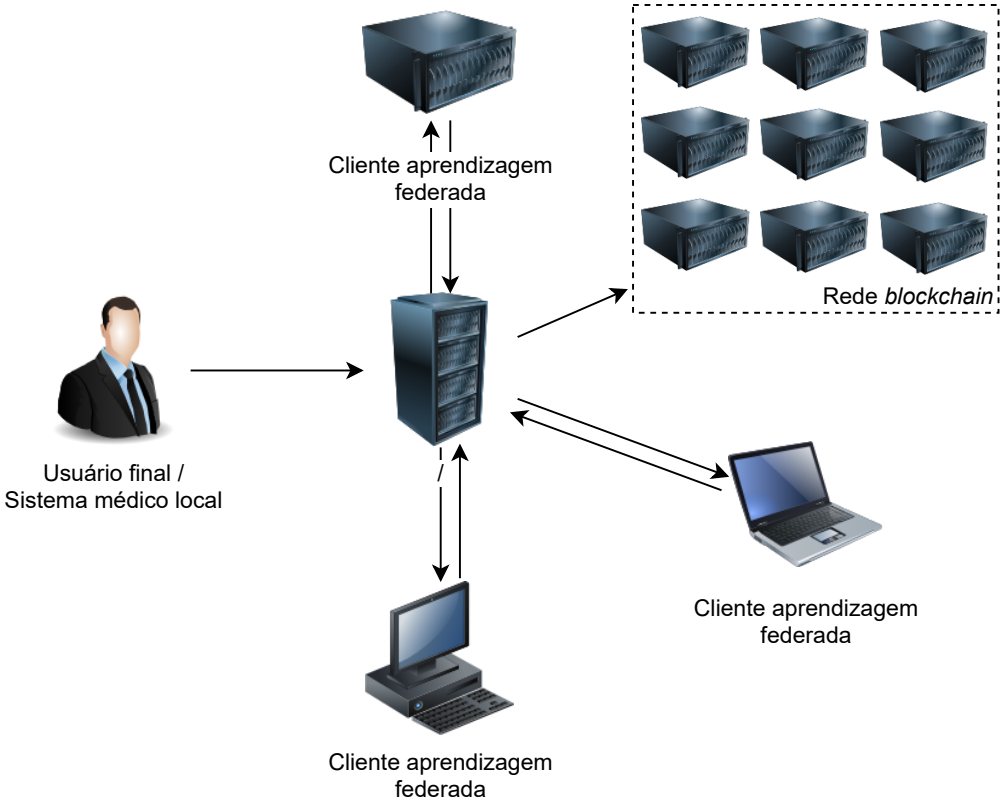

Figura 3 – Visão geral da proposta

Campo	Tipo	Descrição
round	Inteiro	Número da rodada
metrics	Texto	Métricas do treinamento (acurácia e perda) no formato JSON
model	Texto	Modelo serializado

Tabela 1 – Descrição dos campos salvos no livro-razão nas soluções A e B.

4.1.3.1 Solução A

A solução A consiste em armazenar todos os dados relacionados ao modelo e às suas métricas no livro-razão, para todos os clientes da rede de aprendizagem federada e também para o servidor. Com isso, é possibilitada a auditoria dos modelos de todos os nós que participam do treinamento em casos de suspeita de ataques.

O diagrama de sequência da solução A é apresentado na Figura 4. Como pode-se perceber, o fluxo se inicia na junção dos clientes ao servidor, que envia um sinal para iniciar o treinamento por rodada nos clientes, que treinam o modelo localmente com seus dados e após isso enviam as mudanças para o servidor, que faz a agregação dos modelos e salva as informações no livro-razão. Os dados salvos e sua descrição são presentes na Tabela 1. A esquematização do funcionamento da solução A é apresentada na Figura 5

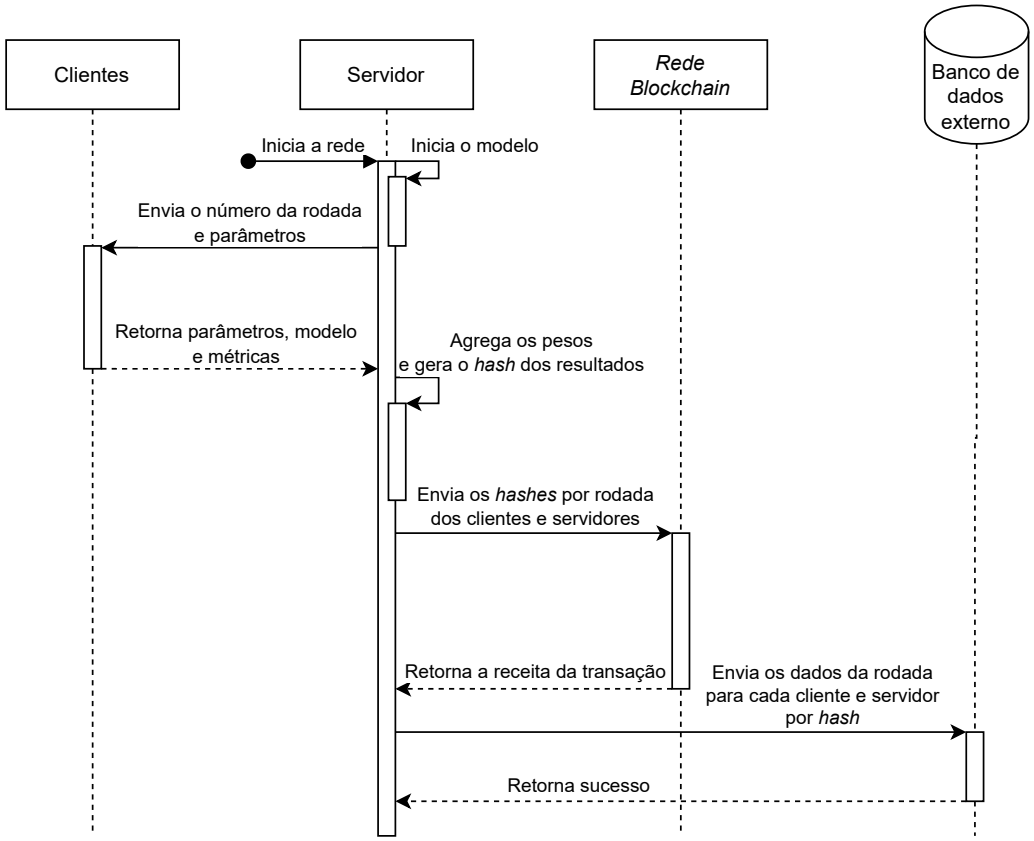

Figura 4 – Diagrama de sequência das soluções A e B

As configurações da rede, presentes mo bloco gênesis da rede, são apresentadas no Quadro 8.

Quadro 8 – Configuração da rede (bloco gênesis)

Configuração	**Valor**
Protocolo	QBFT
Duração da época (epoch length)	30000
Período de bloco (s)	5
Esgotamento da requisição (s)	10

4.1.3.2 Solução B

A solução B é similar à solução A, o fluxo de envio dos dados é o mesmo e o armazenamento em ambas é feito somente no livro-razão. Apenas o servidor envia o modelo na transação além das métricas, os modelos dos clientes são apenas enviados ao servidor para agregação ao modelo global, ou seja, apenas suas métricas são enviadas ao livro-razão.

O armazenamento do modelo após a agregação favorece parcialmente a auditoria podendo-se determinar se em alguma das rodadas houve algum tipo de ataque que prejudicasse o modelo global e a recuperação do estado anterior a ele.

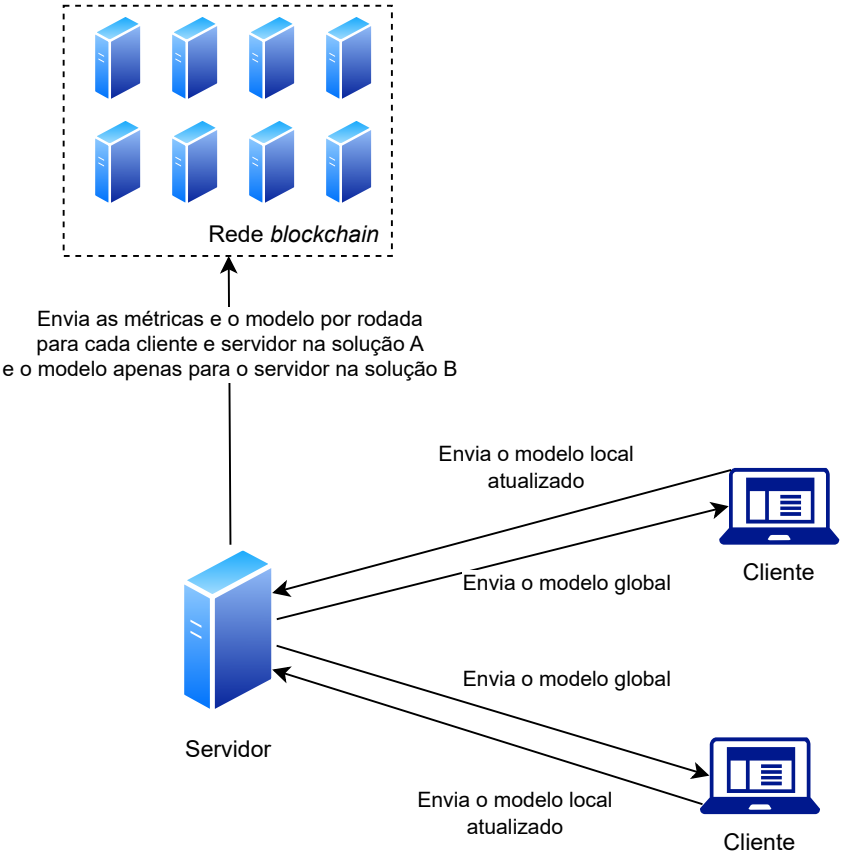

Figura 5 – Esquema do funcionamento das soluções A e B

O formato dos dados também é o mesmo representado na Tabela 1, embora o campo de modelo seja vazio para os clientes. A Figura 4 também apresenta o diagrama de sequência da solução B, sendo seu fluxo igual ao da solução anterior. A esquematização de seu funcionamento também é apresentada na Figura 5.

4.1.3.3 Solução C

A solução C é diferente das soluções anteriores em seu fluxo e nos dados armazenados. Os clientes enviam seus modelos para o servidor da aprendizagem federada, que por sua vez realiza a agregação e gera um *hash* das suas informações (modelos e métricas) e dos clientes e envia todas essas informações para um banco externo à rede, definindo a chave de cada entrada pelo seu *hash*, que por sua vez é enviado para o livro-razão. Assim, em caso de consulta das informações, deve-se consultar o livro-razão para a rodada e o nó esperado a fim de obter a chave para acessar os mesmos em um banco externo definido de acordo com a implementação da arquitetura. Seu diagrama de sequência é mostrado na Figura 6.

Esse modelo tem as mesmas vantagens de auditabilidade e recuperação após ataques das anteriores, porém se benificiando do menor armazenamento de dados no livro-razão.

O formato dos dados das entradas no livro-razão é definido na Tabela 2 e os dados

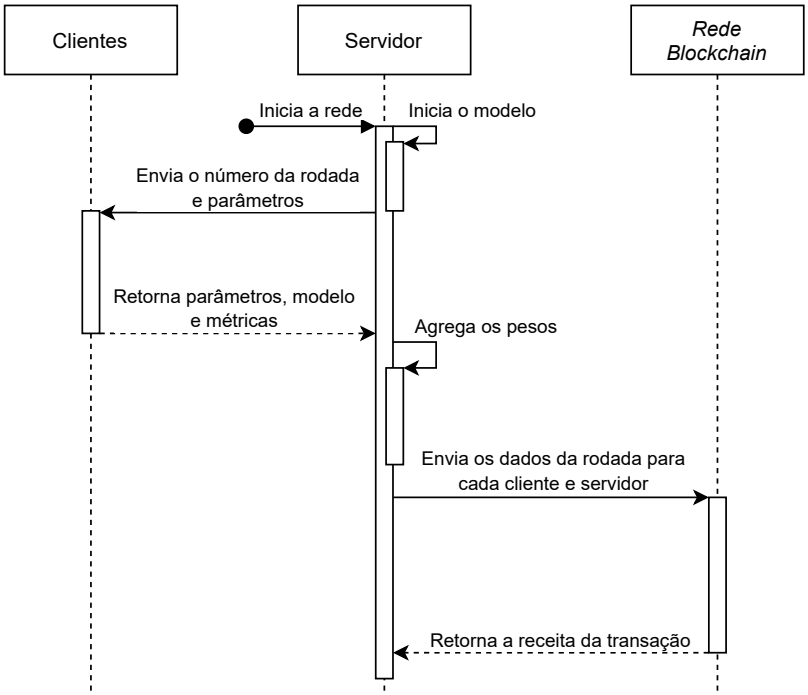

Figura 6 – Diagrama de sequência da solução C

Campo	Tipo	Descrição
round	Inteiro	Número da rodada
datahash	Texto	Hash dos dados

Tabela 2 – Descrição dos campos salvos no livro-razão na solução C.

Campo	Tipo	Descrição
_id	Texto	Hash da entrada
cid	Texto	Identificação do nó
round	Inteiro	Número da rodada
metrics	Texto	Métricas do treinamento (acurácia e perda) no formato JSON
model	Texto	Modelo serializado

Tabela 3 – Descrição dos campos salvos no banco de dados externo na solução C.

enviados ao banco externo são definidos na Tabela 3. A esquematização do funcionamento da solução C é apresentada na Figura 7.

4.1.4 Descrição dos contratos inteligentes

Como pode-se perceber pela estrutura dos dados, as soluções A e B compartilham do mesmo modelo das informações armazenadas no livro-razão. Sendo assim, um mesmo contrato inteligente foi proposto para as duas. Já a solução C possui uma estrutura diferente, requerindo

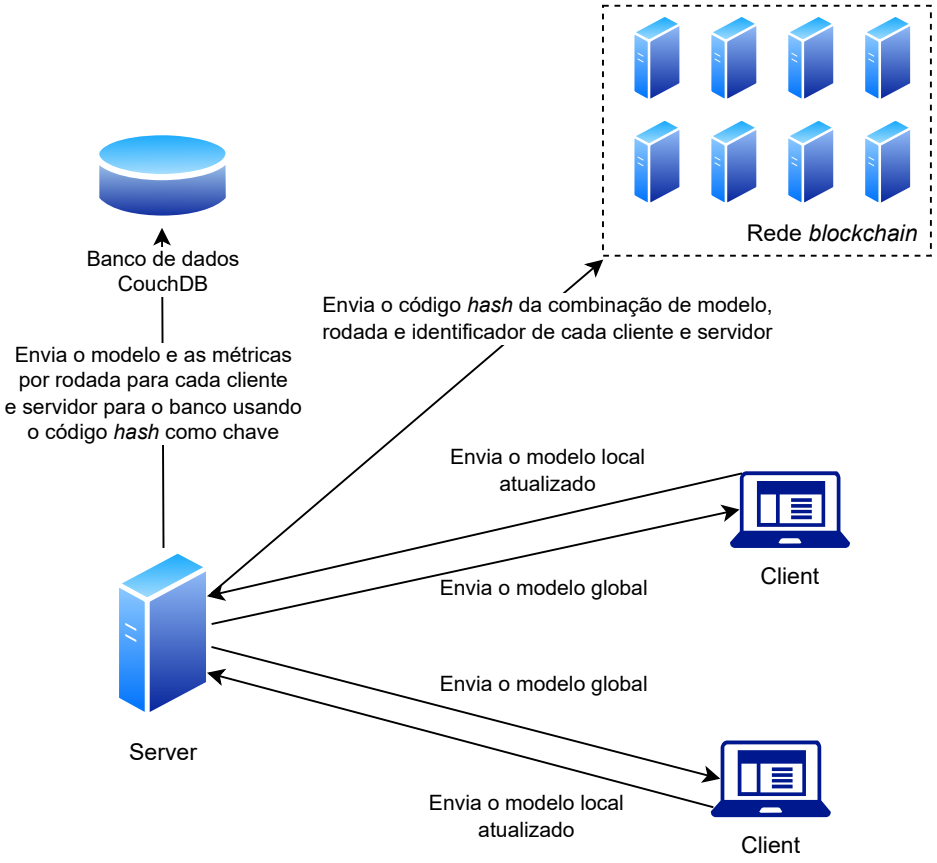

Figura 7 – Esquema do funcionamento da solução C

um contrato inteligente específico para ela. Os contratos inteligentes foram desenvolvidos na linguagem *solidity* e podem ser adaptados para todas as implementações de redes *blockchain* com contratos inteligentes que sejam Turing-completas.

4.1.4.1 *Contrato inteligente das soluções A e B*

As operações do contrato das soluções A e B é mostrado no Algoritmo 1 e é detalhado a seguir.

No contrato são definidas as estruturas dos dados e apenas duas funções para as operações principais:

A primeira é utilizada para enviar uma nova transação que recebe os dados advindos das rodadas da aprendizagem federada, validando e salvando dentro dos dados das rodadas do nó específico a nova, contendo as métricas e o modelo de acordo com a solução específica

A segunda função é apenas utilizada para fazer uma consulta no livro-razão, sem executar uma transação, para obter as rodadas de um par específico a partir de seu identificador.

Algoritmo 1 – Contrato inteligente das soluções A e B

```
1:  Contrato SoluçõesAeB
2:      Estrutura Rodada
3:          uint256 número_rodada
4:          string métricas
5:          string modelo
6:      Fim Estrutura Rodada
7:
8:      Estrutura Entrada
9:          string identificador_nó
10:         Rodada[] rodadas
11:     Fim Estrutura Entrada
12:
13:     address endereço_origem
14:     mapping(string ⇒ Entrada) public entradas
15:
16:     Construtor Flcontract()
17:         endereço_origem ← msg.remetente
18:     Fim Construtor
19:
20:     Função enviar(uint256 número_rodada, string métricas, string modelo, string identificador_nó)
21:         Requer número_rodada ≥ 0, "Número da rodada deve ser maior que 0"
22:         entrada ← entradas[identificador_nó]
23:         Se bytes(entrada.identificador_nó).length == 0
24:             entrada.identificador_nó ← identificador_nó
25:         Fim Se
26:         entrada.rodadas.push(Rodada(número_rodada, métricas, modelo))
27:         entradas[identificador_nó] ← entrada
28:     Fim Função
29:
30:     Função obterPeloNó(string identificador_nó) retorna Rodada[]
31:         Retorna entradas[identificador_nó].rodadas
32:     Fim Função
33: Fim Contrato
```

4.1.4.2 Contrato inteligente da solução C

As operações do contrato da solução C são similares às das outras, por isso o restante das funções foi omitida por ser ser igual. A implementação é demonstrada no Algoritmo 2.

As estruturas diferem das mostradas no Código Fonte 1 pois não possuem os campos de modelo e métricas, mas somente o *hash* para acessar as entradas no banco de dados externo.

Algoritmo 2 – Contrato inteligente da solução C

```
1:  Contrato SoluçãoC
2:      Estrutura Rodada
3:          uint256 número_rodada
4:          string hash
5:      Fim Estrutura Rodada
6:
7:      Estrutura Entrada
8:          string identificador_nó
9:          Rodada[] rodadas
10:     Fim Estrutura Entrada
11: Fim Contrato
```

4.2 Resultados e discussão

Na presente seção serão mostrados os resultados dos experimentos e a discussão sobre eles. Na Subseção 4.2.1 serão apresentados os resultados dos experimentos da primeira implementação da rede *blockchain*. Logo após, nas subseções 4.2.2 e 4.2.3 serão apresentados os resultados da implementação final para a detecção de anomalias e para a rede, respectivamente, comparando as três soluções propostas.

4.2.1 Resultados da primeira implementação - rede blockchain

Os testes de ambas implementaçõeos de redes *blockchain* tiveram como premissa a avaliação de indicadores e do desempenho geral das redes *blockchain* descritas nas Subseções 4.1.1 e 4.1.3.

Os resultados da avaliação de desempenho são descritos a seguir:

Com os testes inicias com a rotina de testes desenvolvida utilizando o kit de desenvolvimento para *Node.js* do *Hyperledger Fabric*, foram realizadas 100 transações em sequência em uma rede com 1 nó ordenador e 2 pares. Os tempos de cada transação são mostrados na Tabela 4, sendo a média 2,094 segundos, o desvio padrão de 0,09919 segundo, o valor máximo 3,056 segundos, o mínimo 2,072 segundos e a amplitude de 0,984 segundo.

Com os testes utilizando a ferramenta *Hyperledger Caliper*, são registradas a taxa de envio das transações (em transações por segundo); as latências máxima, mínima e média (em segundos) e a vazão (em transações por segundo) a partir do envio de transações por 60 segundos. Seus resultados são precisos mas podem não indicar o desempenho final das redes em um ambiente real, pois todos os processo do teste são controlados e sem variações de variáveis externas como carga e nós na rede.

Os resultados obtidos foram detalhados na Tabela 5 para o cenário de uma rede com 1 nó ordenador e 5 pares, sendo realizados 10 testes com a ferramenta e obtendo uma média de 174,52 transações por segundo, latências máxima, mínima e média de 2,767 segundos, 0,116

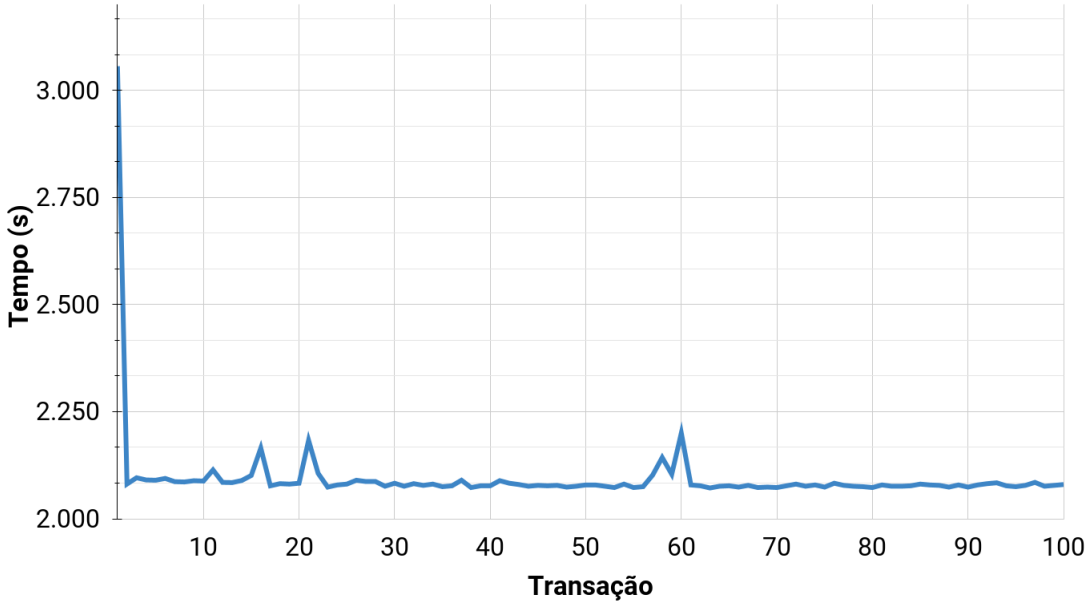

Figura 8 – Tempo das transações medidas com a rotina customizada de benchmark em uma rede com 1 nó ordenador e 2 pares

segundo e 1,005 segundo, respectivamente e uma vazão de 169,16 transações por segundo. O desvio padrão da taxa de envio foi de 5,165, enquanto das latências máxima, mínima e média foram de 0,306, 0,074 e 0,050, respectivamente, e da vazão 5,075.

Em um cenário de uma rede com 8 nós ordenadores e 5 pares, ainda utilizando a ferramenta *Hyperledger Caliper*, os resultados obtidos são os apresentados na Tabela 6, sendo as médias sendo 115,16 transações por segundo para a taxa de envio; 4,39, 0,119 e 1,569 para as latências máxima, mínima e média, respectivamente e 112,53 transações por segundo para a vazão. O desvio padrão da taxa de envio foi de 5,889, enquanto das latências máxima, mínima e média foram de 0,625, 0,052 e 0,094, respectivamente, e da vazão 6,300.

Os resultados indicam que houve um aumento menos que linear no tempo entre os dois testes, o que indica uma potencial escalabilidade da arquitetura. O tempo é maior com o aumento do número de pares devido à maior necessidade de tráfego de informações entre os pares e devido à política de validação e ordenação das transações, que nesse caso não depende somente de um nó ordenador.

4.2.2 Resultados da implementação final - detecção de anomalias

Foram coletadas 242 entradas no banco de dados da solução C dos testes executados. Os modelos modelos locais são treinados com uma partição aleatória da base de dados. A separação da base foi feita em uma proporção de 9:1 para treino e testes.

A acurácia média obtida pelo modelo foi de aproximadamente 98,37% e o valor médio

Tabela 4 – Métricas dos testes realizados a rotina de testes desenvolvida em uma rede com 1 nó ordenador e 2 pares

Número da transação	**Tempo (s)**
1	3,056
2	2,081
3	2,096
4	2,091
5	2,090
6	2,094
7	2,087
8	2,086
9	2,089
10	2,088
11	2,114
12	2,085
13	2,085
14	2,089
15	2,101
16	2,165
17	2,077
18	2,082
19	2,081
20	2,083
21	2,183
22	2,106
23	2,074
24	2,079
25	2,081
26	2,09
27	2,087
28	2,087
29	2,076
30	2,083
31	2,076
32	2,082
33	2,078
34	2,081
35	2,075
36	2,077
37	2,09
38	2,073
39	2,077
40	2,077

Número da transação	**Tempo (s)**
40	2,077
41	2,089
42	2,083
43	2,08
44	2,076
45	2,078
46	2,077
47	2,078
48	2,074
49	2,076
50	2,079
51	2,079
52	2,076
53	2,073
54	2,081
55	2,073
56	2,075
57	2,102
58	2,143
59	2,104
60	2,2
61	2,079
62	2,077
63	2,072
64	2,076
65	2,077
66	2,074
67	2,078
68	2,073
69	2,074
70	2,073
71	2,077
72	2,081
73	2,076
74	2,079
75	2,074
76	2,083
77	2,078
78	2,076
79	2,075

Número da transação	**Tempo (s)**
79	2,075
80	2,073
81	2,079
82	2,076
83	2,076
84	2,077
85	2,081
86	2,079
87	2,078
88	2,074
89	2,079
90	2,074
91	2,079
92	2,082
93	2,084
94	2,077
95	2,075
96	2,078
97	2,085
98	2,076
99	2,078
100	2,08
Média (s)	2,094
Desvio padrão (s)	0,09919
Valor máximo (s)	3,056
Valor mínimo (s)	2,072
Amplitude (s)	0,984

Tabela 5 – Métricas obtidas nos testes do *Hyperledger Caliper* de uma rede com 1 nó ordenador e 5 pares

Teste	**Taxa de envio (transações por segundo)**	**Latência máxima (s)**	**Latência mínima (s)**	**Latência média (s)**	**Vazão (transações por segundo)**
1	167,2	2,93	0,13	1,03	162,9
2	165,3	2,8	0,07	1,04	159
3	173,1	2,94	0,07	0,95	168,9
4	176,8	2,62	0,31	1,03	171,8
5	176,1	2,39	0,15	1,03	169,5
6	174,1	2,46	0,11	0,99	173,5
7	174,2	2,4	0,07	1,1	167,5
8	179,2	2,9	0,07	0,98	172,4
9	182,6	3,37	0,1	0,93	176,2
10	176,6	2,86	0,08	0,97	169,9
Média	174,52	2,767	0,116	1,005	169,16
Desvio padrão	5,165	0,306	0,074	0,050	5,075

Tabela 6 – Métricas obtidas nos testes do *Hyperledger Caliper* de uma rede com 8 nós ordenadores e 5 pares

Teste	**Taxa de envio (transações por segundo)**	**Latência máxima (s)**	**Latência mínima (s)**	**Latência média (s)**	**Vazão (transações por segundo)**
1	121,9	5,52	0,1	1,46	118,9
2	116,8	4,04	0,12	1,53	113
3	115,8	4,26	0,09	1,44	111,1
4	121,7	4,33	0,14	1,46	121,3
5	112,1	4,29	0,1	1,59	109,3
6	115,6	3,83	0,09	1,62	113,7
7	121,3	3,5	0,26	1,65	119,7
8	103,9	5,08	0,09	1,6	101,6
9	114,1	4,04	0,1	1,61	111,5
10	108,4	5,01	0,1	1,73	105,2
Média	115,16	4,39	0,119	1,569	112,53
Desvio padrão	5,889	0,625	0,052	0,094	6,300

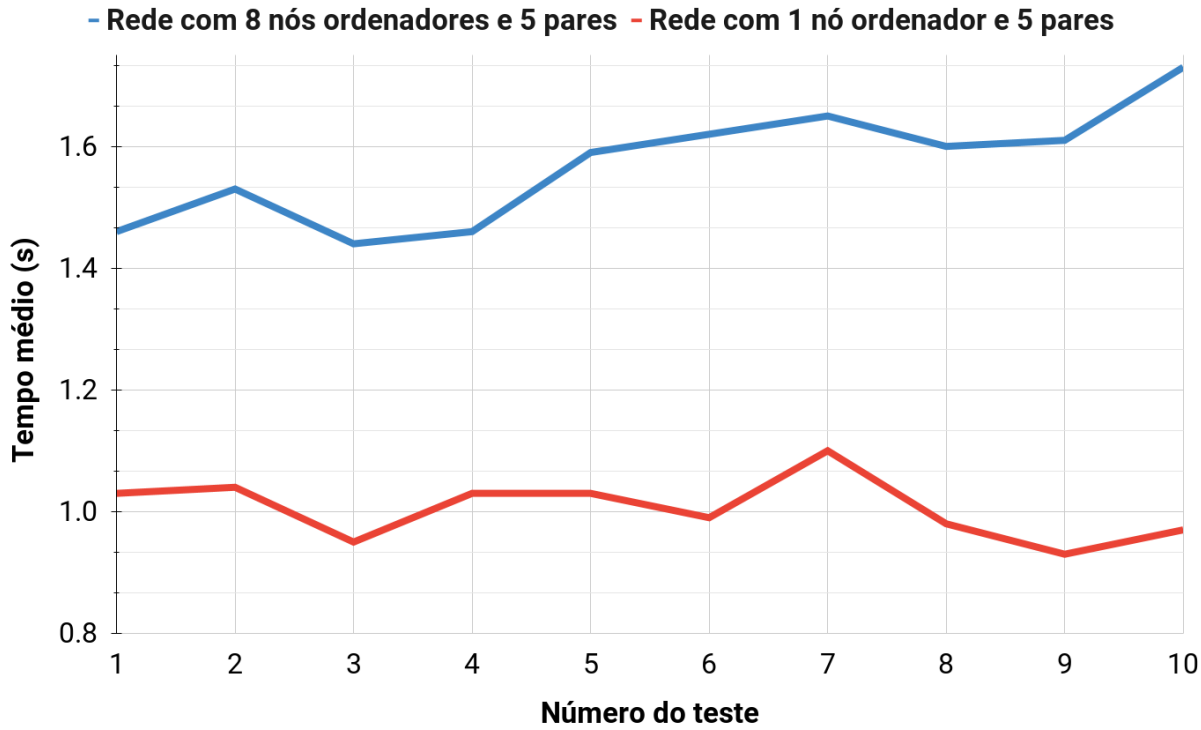

Figura 9 – Médias obtidas com os testes do *Hyperledger Caliper* em uma rede com 1 nó ordenador e 5 pares e em outra com 8 nós ordenadores e 5 pares

da perda foi de 67,82%. Os valores obtidos foram obtidos por meio da implementação do próprio modelo avaliando entradas escolhidas aleatoriamente para testes representando 10% da base de dados do cliente. Os resultados indicam que a classificação pode estar sendo afetada pela grande predominância de entradas da classe dominante, o que faz com que mais predições verdadeiras sejam feitas sem indicar com certeza a qualidade da classificação.

É relevante notar que não houve aproveitamento do que foi realizado nas implementações anteriores (apresentadas na Subseção 4.1.1) devido à inadequação do uso dos algoritmos escolhidos na implementação anterior para os dados escolhidos, uma vez que é um problema de classificação binária, com base de dado escolhida já sendo rotulada.

4.2.3 Resultados da implementação final - arquitetura completa

Os resultados obtidos pelos testes na nova arquitetura estão separados por qualidade analisada. Foram analisados o tempo das rodadas (para redes *blockchain* com 5 e 11 pares comparados com uma solução de aprendizagem federada sem *blockchain*), o uso de memória de acesso rápido (RAM), o uso de processamento (tempo de CPU) e o tamanho do bloco. Os resultados obtidos nos testes serão apresentados e posteriormente discutidos.

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Execução 1	9,247	9,931	9,960	10,111	10,000
Execução 2	8,707	10,021	10,084	9,986	9,913
Execução 3	9,673	10,143	9,941	10,244	9,878
Execução 4	6,540	9,982	10,048	9,994	10,037
Execução 5	10,929	10,013	10,209	9,891	10,050
Médias	9,019	10,018	10,048	10,045	9,976
Desvios-padrão	1,610	0,078	0,108	0,136	0,076

Tabela 7 – Tempos em segundos das rodadas de execuções dos testes da rede com 5 pares - solução A

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Execução 1	10,462	9,927	10,034	9,915	10,018
Execução 2	7,608	9,927	10,044	9,898	10,055
Execução 3	6,308	9,909	9,923	10,092	9,979
Execução 4	6,598	10,043	9,901	10,037	10,075
Execução 5	8,399	9,891	10,019	10,036	10,023
Médias	7,875	9,939	9,984	9,996	10,030
Desvios-padrão	1,668	0,060	0,067	0,085	0,037

Tabela 8 – Tempos em segundos das rodadas de execuções dos testes da rede com 5 pares - solução B

4.2.3.1 Medição do tempo das rodadas

Os testes para medição de tempo buscam avaliar a viabilidade da arquitetura proposta em questão de tempo para as transações. Foram feitos em execuções de cinco rodadas cada, sendo o tempo medido pela diferença entre o tempo de início da rodada e o tempo de fim, obtidos no servidor da aprendizagem federada. Os resultados das medidas de tempo para a rede *blockchain* com 5 pares integrada com uma rede de aprendizagem federada com um servidor e 2 clientes é apresentada nas Tabelas 7, 8 e 9 para as soluções A, B e C respectivamente. Para uma rede *blockchain* com 11 pares e a mesma quantidade de nós na rede de aprendizagem federada, os resultados obtidos são mostrados nas Tabelas 11 para a solução A, 12 para a solução B e 13 para a solução C. Por motivos de comparação, os tempos das rodadas para uma rede de aprendizagem federada igual às utilizadas nos testes anteriores porém sem a integração com a rede *blockchain* é apresentada na Tabela 10. Como pôde-se perceber, os resultados foram muito semelhantes entre as soluções, variando apenas em centésimos de segundos. A solução com a menor média foi a solução B.

O comparativo das médias dos tempos das rodadas de cada uma das soluções é apresentado na Tabela 14 e nos gráficos das Figuras 10 para 5 nós e 11 e 11 nós. Os resultados indicam que as diferenças entre no funcionamento das soluções não implica em mudanças significativas nos tempos de execução. Quanto às diferenças entre as diferentes configurações de rede indicam também que um aumento nos nós implica um aumento menos que linear nos tempos medidos, demonstrando um potencial de escalabilidade.

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Execução 1	10,963	9,799	10,074	9,979	9,930
Execução 2	7,572	10,052	10,230	9,858	9,983
Execução 3	10,471	10,031	9,893	10,085	9,927
Execução 4	10,037	9,978	10,080	9,938	10,000
Execução 5	10,642	10,052	9,957	10,010	10,036
Médias	9,937	9,982	10,047	9,974	9,975
Desvios-padrão	1,364	0,107	0,130	0,084	0,047

Tabela 9 – Tempos em segundos das rodadas de execuções dos testes da rede com 5 pares - solução C

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Execução 1	1,267	0,737	0,355	0,307	0,313
Execução 2	0,558	0,201	0,319	0,207	0,233
Execução 3	1,093	0,211	0,277	0,272	0,287
Execução 4	0,726	0,272	0,144	0,160	0,250
Execução 5	0,769	0,186	0,227	0,217	0,240
Médias	0,883	0,321	0,264	0,233	0,265
Desvios-padrão	0,289	0,235	0,083	0,058	0,034

Tabela 10 – Tempos em segundos das rodadas de execuções dos testes da rede sem *blockchain*

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Execução 1	12,121	14,980	14,904	15,091	15,007
Execução 2	13,211	15,008	14,952	15,003	14,971
Execução 3	14,811	15,061	14,926	15,089	14,931
Execução 4	15,327	15,018	15,048	14,979	15,064
Execução 5	11,190	15,090	14,928	14,967	15,009
Médias	13,332	15,031	14,952	15,026	14,996
Desvios-padrão	1,749	0,044	0,057	0,060	0,049

Tabela 11 – Tempos em segundos das rodadas de execuções dos testes da rede com 11 pares - solução A

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Execução 1	14,238	14,824	15,195	14,803	15,069
Execução 2	10,802	14,906	15,000	15,018	15,014
Execução 3	14,066	14,928	15,287	14,708	15,009
Execução 4	12,759	14,955	14,996	14,939	15,020
Execução 5	11,322	15,022	15,025	14,934	14,972
Médias	12,637	14,927	15,101	14,880	15,017
Desvios-padrão	1,559	0,072	0,133	0,123	0,035

Tabela 12 – Tempos em segundos das rodadas de execuções dos testes da rede com 11 pares - solução B

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Execução 1	12,292	14,947	15,005	14,980	15,036
Execução 2	12,168	15,131	14,942	14,969	14,944
Execução 3	12,988	15,109	14,945	14,957	14,952
Execução 4	14,621	14,946	15,059	14,982	14,915
Execução 5	15,433	15,000	15,020	14,977	15,076
Médias	13,500	15,027	14,994	14,973	14,985
Desvios-padrão	1,457	0,088	0,050	0,010	0,068

Tabela 13 – Tempos em segundos das rodadas de execuções dos testes da rede com 11 pares - solução C

	Solução A	Solução B	Solução C	Rede tradicional
Média 5 pares	9,821	9,565	9,983	-
Média 11 pares	9,959	9,741	9,651	-
Média sem rede blockchain	-	-	-	0,39312

Tabela 14 – Comparativo das médias de tempo das soluções em segundos

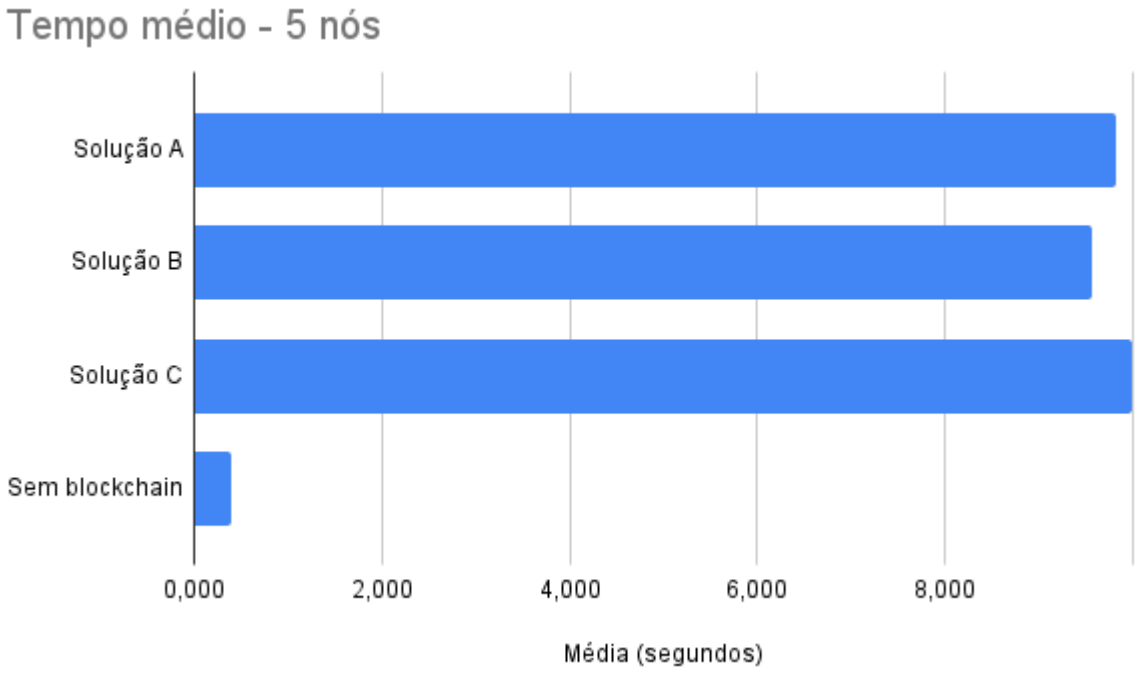

Figura 10 – Médias do tempo de rodada - 5 pares *blockchain*

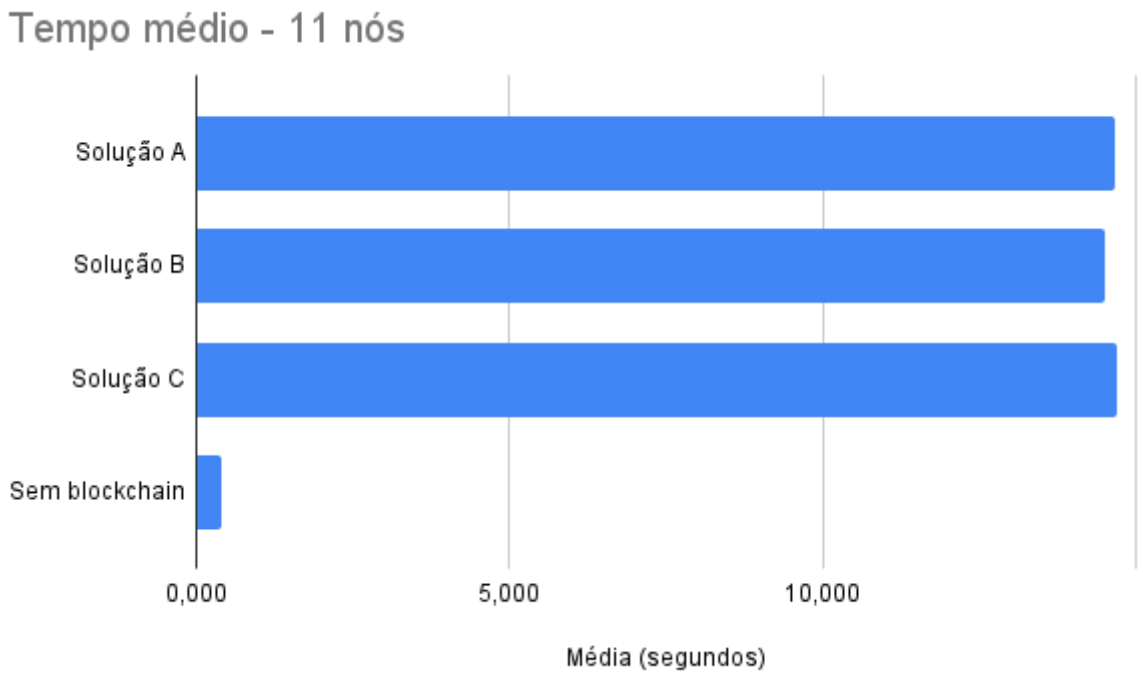

Figura 11 – Médias do tempo de rodada - 11 pares *blockchain*

4.2.3.2 Consumo de memória RAM por par

Como apresentado, foram investigadas as medidas de uso de memória RAM de cada um dos contêineres que executavam as rotinas dos pares. Foram capturadas as medidas em três pontos na execução, no início, no meio e no fim, para a rede *blockchain* com 11 pares conforme apresentado anteriormente. Os resultados são mostrados nas Tabelas 15, 16 e 17 para as soluções A, B e C respectivamente. A comparação final pode ser observada na Tabela 18 e no gráfico da Figura 12. Os resultados indicam uma pequena variação no uso de memória entre as soluções com o passar de tempo, que pode ser influenciado por outros fatores como tempo de execução dos contêineres dos pares. Sendo assim, conclui-se que não existem diferenças significativas entre elas.

4.2.3.3 Uso de CPU por par

Assim como a medição do uso de memória RAM, foi avaliado o tempo de CPU de cada par nos contêineres da rede. Também foram capturadas três medidas ao longo da execução das rodadas, uma no início, outra no meio e uma no término. As medidas foram feitas na rede *blockchain* com 11 pares. Os resultados são apresentados nas Tabelas 19, 20 e 21 para as soluções A, B e C respectivamente. A comparação entre as medidas é apresentada na tabela na Tabela 22 e no gráfico da Figura 13. Os resultados indicam uma pequena variação no uso de CPU entre as soluções com o passar do tempo, que também pode ser influenciado por outros fatores como tempo de execução dos contêineres dos pares. Sendo assim, também pode-se notar que não há

	00:00:00	**00:00:38**	**00:01:17**	**Média**
nó 1	431333,464	441942,704	469391,056	447555,7413
nó 2	219948,496	229553,52	256052,664	235184,8933
nó 3	426627,856	451937,776	461520,632	446695,4213
nó 4	550529,464	578004,96	587667,632	572067,352
nó 5	171955,248	196236,496	209066,52	192419,4213
nó 6 (RPC)	514109,368	523674,536	552239,984	530007,9627
nó 7	547068,984	589395,16	611533,664	582665,936
nó 8	507689,544	532972,784	566895,88	535852,736
nó 9	252345,4	278667,744	299874,6	276962,5813
nó 10	366478,424	391758,664	410923,808	389720,2987
nó 11	482101,84	512640,112	533892,096	509544,6827

Tabela 15 – Medições de uso da memória RAM em bytes dos nós da rede durante a execução dos testes - solução A

	00:00:00	**00:00:38**	**00:01:17**	**Média**
nó 1	603478,456	615057,744	642385,088	620307,096
nó 2	787342,432	794711,4	822007,816	801353,8827
nó 3	168538,032	193787,528	203256,272	188527,2773
nó 4	205481,776	214969,68	243332,384	221261,28
nó 5	622392,208	649735,624	658148,968	643425,6
nó 6 (RPC)	451551,56	463136,192	489412,08	468033,2773
nó 7	724817,392	742810,176	787150,272	751592,6133
nó 8	698669,16	730259,856	761759,6	730229,5387
nó 9	550484,968	575704,424	595658,184	573949,192
nó 10	738175,512	761311,208	782322,688	760603,136
nó 11	374849,544	406400,4	420059,808	400436,584

Tabela 16 – Medições de uso da memória RAM em bytes dos nós da rede durante a execução dos testes - solução B

	00:00:00	**00:00:38**	**00:01:17**	**Média**
nó 1	684551,256	694057,112	721477,736	700028,7013
nó 2	722539,12	732057,368	758440,656	737679,048
nó 3	800141,096	826424,408	834976,816	820514,1067
nó 4	276984,68	285481,376	313983,92	292149,992
nó 5	570069,672	595336,96	608085,248	591163,96
nó 6 (RPC)	450829,368	462406,68	488845,4	467360,4827
nó 7	775204,136	815330,2	185249,536	591927,9573
nó 8	793266	818663,768	852343,744	821424,504
nó 9	328451,04	353841,904	372870,112	351721,0187
nó 10	241715,048	266953,472	284027,648	264232,056
nó 11	574014,16	605612,416	623639,136	601088,5707

Tabela 17 – Medições de uso da memória RAM em bytes dos nós da rede durante a execução dos testes - solução C

	00:00	00:38	01:17
Solução A	406380,7353	429707,6778	450823,5033
Solução B	538707,3673	558898,5665	582317,56
Solução C	565251,416	586924,1513	549449,0865

Tabela 18 – Comparação de tamanho uso médio de RAM (KB)

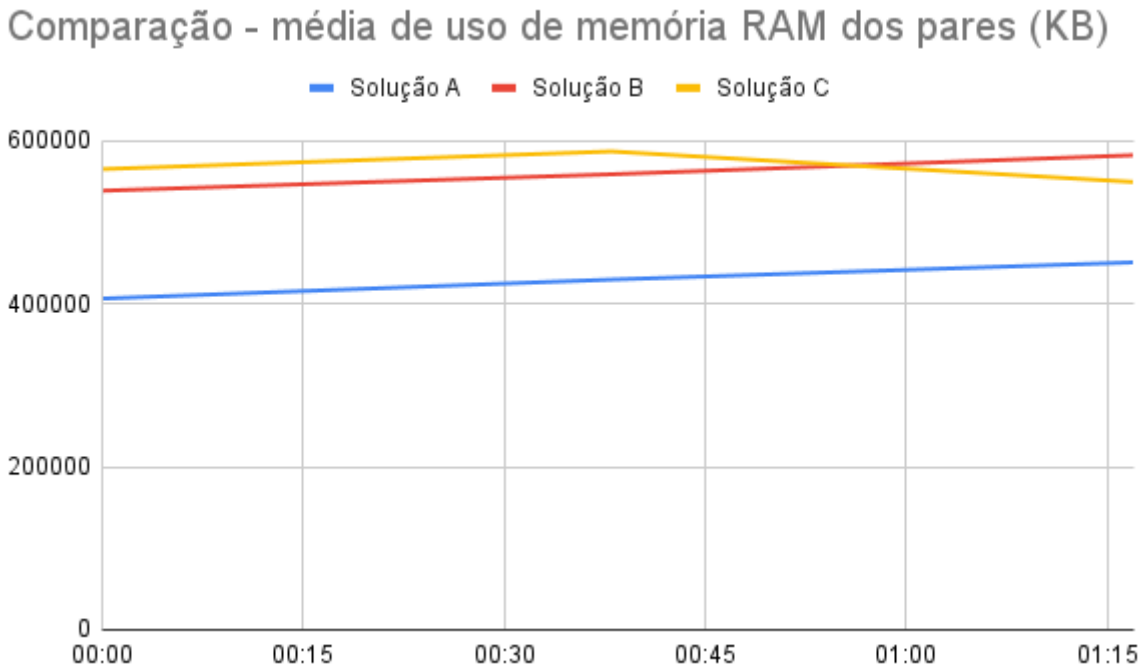

Figura 12 – Médias de uso de memória RAM (KB)

diferenças significantes a serem notadas.

4.2.3.4 Tamanho do bloco com as transações

Os números dos blocos das transações foram obtidos a partir de sua receita capturada nos testes, com base neles foi possível consultar o livro-razão para saber suas informações de tamanho. Foram feitas medidas dos blocos das transações depois de 5 rodadas da aprendizagem federada para cada uma das soluções. Os resultados das medições são apresentados nas Tabelas 23, 24 e 25 para as soluções A, B e C respectivamente. A comparação dos valores obtidos é feita na Tabela 26 e na Figura 14. Nota-se as diferenças esperadas entre as soluções, onde a solução A tem os maiores blocos tanto para os clientes quanto os servidores, a solução B é igual à A no tamanho dos blocos com a transação dos servidores porém tem tamanhos menores para os blocos com transações dos clientes. Já a solução C tem os menores tamanhos de blocos com as transações.

	00:00:00	**00:00:38**	**00:01:17**	**Média**
nó 1	54,88	55,09	55,6	55,19
nó 2	55,61	55,93	56,66	56,06666667
nó 3	53,9	54,24	54,45	54,19666667
nó 4	57,04	57,38	57,73	57,38333333
nó 5	53,39	53,97	54,39	53,91666667
nó 6 (RPC)	55	55,2	55,93	55,37666667
nó 7	64,03	65,56	66,42	65,33666667
nó 8	66,68	67,1	68,05	67,27666667
nó 9	63,34	63,66	64,02	63,67333333
nó 10	64,73	65,03	65,44	65,06666667
nó 11	62,81	63,23	63,65	63,23

Tabela 19 – Medições de uso de tempo de CPU dos nós da rede durante a execução dos testes - solução A

	00:00:00	**00:00:38**	**00:01:17**	**Média**
nó 1	75,87	76,09	76,47	76,14333333
nó 2	77,25	77,41	77,73	77,46333333
nó 3	76,52	76,77	76,96	76,75
nó 4	77,46	77,59	77,86	77,63666667
nó 5	75,34	75,6	75,81	75,58333333
nó 6 (RPC)	76,23	76,44	76,77	76,48
nó 7	86,53	87,45	88,95	87,64333333
nó 8	97,43	97,8	98,2	97,81
nó 9	92,58	92,81	93,17	92,85333333
nó 10	93,91	94,15	94,52	94,19333333
nó 11	92,93	93,26	93,54	93,24333333

Tabela 20 – Medições de uso de tempo de CPU dos nós da rede durante a execução dos testes - solução B

	00:00:00	**00:00:38**	**00:01:17**	**Média**
nó 1	84,48	84,69	85,04	84,73666667
nó 2	86,15	86,46	86,85	86,48666667
nó 3	84,67	84,93	85,12	84,90666667
nó 4	85,64	85,82	86,06	85,84
nó 5	82,93	83,24	83,52	83,23
nó 6 (RPC)	84,41	84,61	84,91	84,64333333
nó 7	96,73	97,45	98,1	97,42666667
nó 8	109,05	109,38	109,76	109,3966667
nó 9	103,43	103,72	104,13	103,76
nó 10	104,59	104,8	105,15	104,8466667
nó 11	103,03	103,37	103,73	103,3766667

Tabela 21 – Medições de uso de tempo de CPU dos nós da rede durante a execução dos testes - solução C

	00:00	00:38	01:17
Solução A	59,21909091	59,67181818	60,21272727
Solução B	83,82272727	84,12454545	84,54363636
Solução C	93,19181818	93,49727273	93,85181818

Tabela 22 – Comparação de tempo de CPU entre as soluções (s)

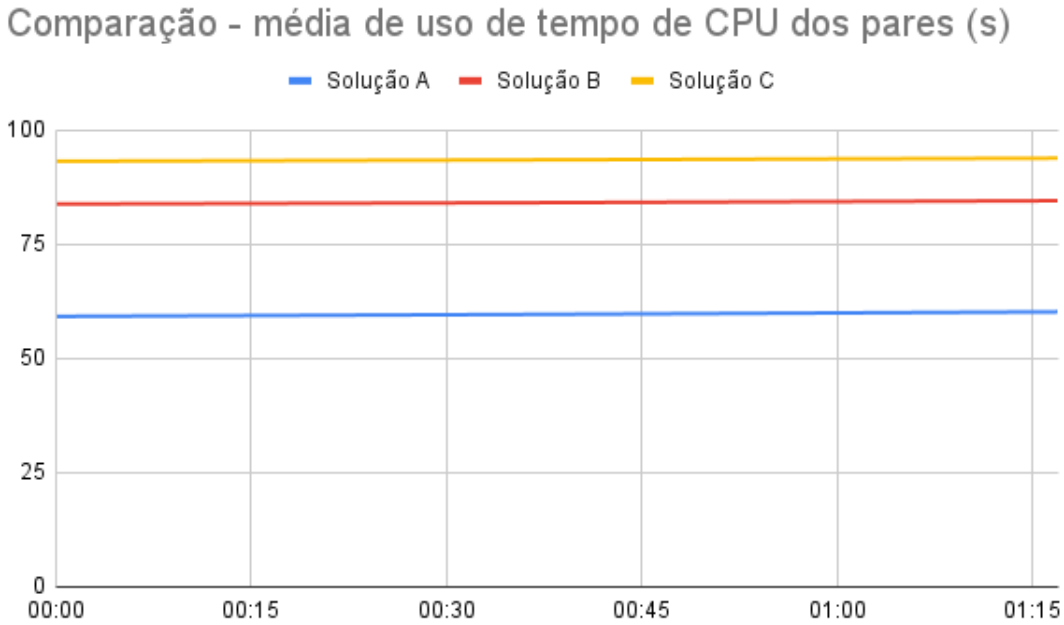

Figura 13 – Comparação de tempo de CPU entre as soluções (s)

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Servidor	2391	2391	2391	2391	2391
Cliente 1	2999	2999	2999	3000	3032
Cliente 2	2999	3031	2999	3032	3032

Tabela 23 – Tamanho do bloco em bytes por rodada - solução A

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Servidor	2391	2391	2391	2391	2391
Cliente 1	1397	1429	1397	1397	1429
Cliente 2	1397	1429	1397	1397	1429

Tabela 24 – Tamanho do bloco em bytes por rodada - solução B

	Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
Servidor	1205	1205	1205	1205	1205
Cliente 1	1205	1205	1205	1205	1205
Cliente 2	1205	1205	1204	1205	1205

Tabela 25 – Tamanho do bloco em bytes por rodada - solução C

	Tamanho do bloco com a transação do servidor (bytes)	Tamanho do bloco com a transação do cliente (bytes)
Solução A	2391	3012,2
Solução B	2391	1409,8
Solução C	1205	1204,9

Tabela 26 – Médias de tamanho do bloco com as transações por solução

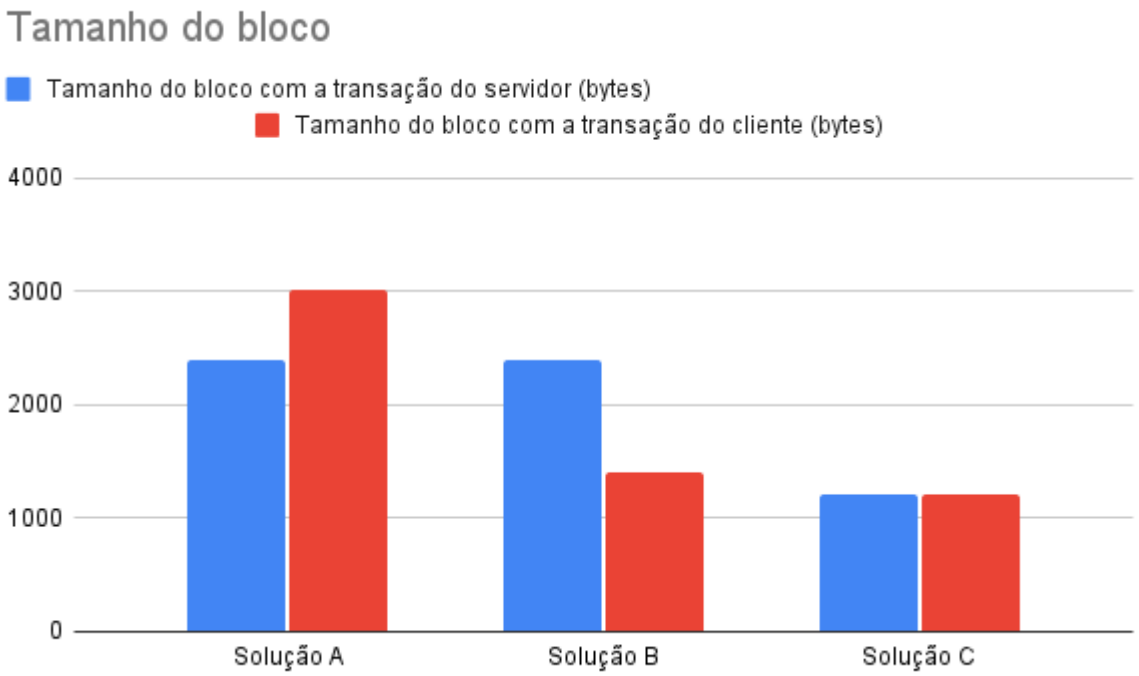

Figura 14 – Médias de tamanho do bloco com as transações por solução

4.3 Divulgação Científica

Publicação realizada na conferência *2021 IEEE 34th International Symposium on Computer-Based Medical Systems* (CBMS):

Garcia et al. (**2021**) Rodrigo Dutra Garcia et al. Towards a decentralized e-prescription system using smart contracts. *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*. (Jun.-2021), 556–561. doi: 10.1109/CBMS52027.2021.00037.

Repositório da primeira implementação (rede *blockchain*):

ZUTIÃO, Gabriel Augusto. **healthcare-blockchain**. 2020. Disponível em: https://github.com/gabrielz blockchain. Acesso em: 01 mai. 2023.

Repositório da implementação final:

ZUTIÃO, Gabriel Augusto. **flchain**. 2021. Disponível em: https://github.com/gabrielzut/flchain. Acesso em: 01 mai. 2023.

CAPÍTULO

5

CONSIDERAÇÕES FINAIS

Neste capítulo serão feitas observações e conclusões sobre o trabalho apresentado anteriormente e seus resultados. Na seção 5.1 serão discutidas suas principais contribuições e por fim é levantada a discussão final na seção 5.2.

5.1 Principais contribuições

Tendo em vista os problemas apresentados no capítulo 2 das redes tradicionais de compartilhamento de registros médicos eletrônicos, como fraudes, susceptibilidade a ataques e maiores necessidades em relação à privacidade dos dados, foi apresentado uma arquitetura que utiliza uma rede de aprendizagem federada assistida por *blockchain* para compartilhamento de informações que permitem a detecção de fraudes e a maior segurança dos dados.

A rede permite que dados de diversas organizações possam gerar um modelo global, que pode ser utilizado por uma entidade pública como o Sistema Único de Saúde no Brasil, por exemplo. Comprova a aplicabilidade do modelo aqui apresentado a utilização da base de dados do Hospital Nossa Senhora da Conceição (SC). Além disso, o algoritmo de aprendizado de máquina pode ser substituído ou configurado, além de outras configurações da rede.

Como apontado, devido à simplicidade da lógica do contrato inteligente, é seguro afirmar que a solução é independente de tecnologias, visto que qualquer tecnologia *blockchain* que permita contratos que são Turing-completos é capaz de ser integrada à rede.

De acordo com a classificação definida por Wang e Hu (2021) explicada na seção 2.6, a arquitetura aqui apresentada pode ser classificada como uma rede flexivelmente acoplada (*flexibly coupled*), uma vez que os pares da rede *blockchain* não necessariamente são clientes da rede de aprendizagem federada. Assim como apresentado na definição pelo autor, isso favorece com que as redes sejam mais independentes e diminui o uso de processamento e banda entre os pares.

Como uma rede flexivelmente acoplada, em um cenário inicialmente considerado, o servidor da aprendizagem federada seria uma instituição que estivesse acima na hierarquia, como uma organização federal. Cada uma das instituições abaixo, como hospitais, consultórios e farmácias, por exemplo, seriam clientes da aprendizagem federada e poderiam ter nós na rede *blockchain* separados. Assim, a rede de aprendizagem federada é flexivelmente acoplada à rede *blockchain*, uma vez que em caso de algum problema com a primeira, a segunda ainda poderia continuar operando.

Com a maior quantidade de dados para treinamento tendo em vista a agregação dos modelos, aumenta-se a viabilidade e a assertividade da detecção de anomalias, tornando todos os sistemas que podem ser integrados a ela mais robustos e capazes de resolver tais problemas. Além disso, por existir apenas a comunicação dos modelos nas atualizações entre os clientes e o servidor da aprendizagem federada, torna-se mais seguro e privado esse processo, uma vez que ataques que visam interceptar esses dados não poderão acessar os dados que treinaram os modelos, mas somente suas atualizações.

5.2 Discussão final

A seguir será apresentada a discussão final, iniciando com comentários acerca dos resultados obtidos na implementação anterior e na arquitetura final e posteriormente sobre o presente estudo e seu estado, suas limitações e sugestões de futuros trabalhos.

A partir dos resultados obtidos na implementação inicial, conclui-se que, em relação aos valores obtidos nos testes da rede *blockchain* das medidas de transações por segundo e latências máxima, mínima e médica e taxa de envio, os valores aumentaram de maneira menor que um crescimento linear relacionado ao aumento do número de pares de um teste para outro, mostrando ser plausivelmente escalável.

A partir dos resultados obtidos na arquitetura final, pode-se estabelecer conclusões acerca da detecção de anomalias (aprendizagem federada) e acerca do desempenho de toda a rede em si.

Em relação às medidas obtidas nos testes do modelo de classificação, pode-se perceber que a acurácia média entre as execuções teve um resultado muito alto (próximo a 1) e a perda média teve um resultado mediano, indicando, assim como foi notado sobre os resultados da implementação anterior, um possível desbalanceamento nos dados, fazendo com que o modelo acertasse com maior facilidade a classe majoritária em detrimento da classe minoritária, favorecendo uma acurácia alta mas que não reflete no desempenho final do algoritmo.

Sobre as medidas obtidas nos testes da arquitetura completa, observa-se que todas as soluções tiveram desempenho similar em relação ao tempo para cada rodada, sendo a solução B a mais performática em questão de centésimos de segundo. Além disso, assim como na implementação anterior, o aumento do número de nós provou que o aumento no tempo das

rodadas foi menos que linear, sendo assim, não apresenta um problema de escalabilidade. A solução B pode ter menor tempo médio devido ao menor tráfego de dados nas transações sem relação à solução A e na ausência do processo de armazenamento de dados e um banco de dados externo, como ocorre na solução C.

Em relação ao uso de memória RAM e de CPU, pode-se notar que na solução A os contêineres tiveram menores medidas em valores consideráveis em relação às outras soluções, embora não possa ser encontrado um motivo na arquitetura para isso. A partir desse ponto, pode-se sugerir que sejam levantados outros fatores que possam ter influenciado no resultado, como a ordem de execução dos testes em relação aos contêineres.

Em relação ao tamanho do bloco, o resultado esperado foi obtido, isto é: a solução A possui os maiores tamanhos de dados nos blocos com as transações, na solução B o tamanho dos blocos com as transições dos clientes é consideravelmente menor devido à ausência do modelo e na solução C os tamanhos são menores para ambos tipos de pares da rede (servidor e cliente). É relevante notar que, já nos testes executados, a solução C conseguiu uma redução considerável no tamanho dos blocos em relação às soluções A e B.

Tendo em vista os resultados obtidos, pode-se perceber que a integração de rede *blockchain* à aprendizagem federada teve um aumento considerável no tempo das rodadas, porém esse aumento é justificado pelas vantagens apresentadas em relação à integração.

Além disso, percebe-se que a solução foi capaz de realizar as operações propostas de maneira coordenada entre as redes, tendo particularidades em relação a cada uma das soluções.

Em relação à comparação entre as três soluções, nota-se que as soluções A e B têm vantagem em relação à C por não necessitarem de um banco de dados externo, podendo ser auditáveis de acordo com os modelos que são armazenados no livro-razão.

A solução C, por sua vez, provou ser mais vantajosa em relação ao tamanho dos blocos, o que indica um menor uso de banda nas comunicações entre os pares. Além disso, ela permite também uma auditabilidade em maior nível que a solução B, assim como ocorre na solução A.

É sugerida como melhoria uma comparação quantitativa mais extensiva do presente trabalho às redes tradicionais e a outras propostas alternativas que podem ter alguma relação ao trabalho.

Além disso, sugere-se para trabalhos futuros o tratamento do desbalanceamento da base de dados por meio de subamostragem e super-amostragem e a elaboração de testes com mais algoritmos, inclusive com algoritmos não supervisionados que poderiam ser úteis em implementações que não pudessem ser rotulados como foi o caso dos experimentos aqui realizados.

Por fim, no presente estudo foi apresentada uma arquitetura que utiliza uma rede de aprendizagem federada assistida por *blockchain* para solucionar problemas de segurança, privacidade e limitações gerais de sistemas de registros médicos eletrônicos.

A arquitetura provou ser funcional para os propósitos que foi elaborada em questão de funcionalidades. Além de ter resultados promissores em relação a seu desempenho, é muito adaptável em relação a tecnologias, pode ser utilizada em diversos âmbitos dos sistemas médicos e pode ser uma solução para a heterogeneidade dos registros médicos nos mais diversos sistemas.

REFERÊNCIAS

ABBAS, K.; AFAQ, M.; KHAN, T. A.; SONG, W. C. A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry. **Electronics (Switzerland)**, MDPI AG, v. 9, n. 5, may 2020. ISSN 20799292. Citado nas páginas 25, 42 e 51.

ABOU-NASSAR, E. M.; ILIYASU, A. M.; EL-KAFRAWY, P. M.; SONG, O. Y.; BASHIR, A. K.; EL-LATIF, A. A. Ditrust chain: Towards blockchain-based trust models for sustainable healthcare iot systems. **IEEE Access**, Institute of Electrical and Electronics Engineers Inc., v. 8, p. 111223–111238, 2020. ISSN 21693536. Citado nas páginas 48 e 54.

AICH, S.; SINAI, N. K.; KUMAR, S.; ALI, M.; CHOI, Y. R.; JOO, M. I.; KIM, H. C. Protecting personal healthcare record using blockchain federated learning technologies. In: . [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2021. v. 2021-February, p. 109–112. ISBN 9791188428069. ISSN 17389445. Citado nas páginas 46 e 54.

ANDROULAKI, E.; BARGER, A.; BORTNIKOV, V.; MURALIDHARAN, S.; CACHIN, C.; CHRISTIDIS, K.-n.; De Caro, A.; ENYEART, D.; MURTHY, C.; FERRIS, C.; LAVENTMAN, G.-n.; MANEVICH, Y.; NGUYEN, B.; SETHI, M.; SINGH, G.; SMITH, K.; SORNIOTTI, A.; STATHAKOPOULOU, C.; VUKOLIĆ, M.; COCCO, S. W.; YELLICK, J. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. In: **Proceedings of the 13th EuroSys Conference, EuroSys 2018**. ACM, 2018. v. 2018-Janua. ISBN 9781450355841. Disponível em: <https://doi.org/10.1145/3190508.3190538>. Citado na página 32.

ANGST, C. M.; AGARWAL, R. Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. **MIS Quarterly**, Management Information Systems Research Center, University of Minnesota, v. 33, n. 2, p. 339–370, 2009. ISSN 02767783. Disponível em: <http://www.jstor.org/stable/20650295>. Citado na página 25.

ASHRAF, E.; AREED, N. F.; SALEM, H.; ABDELHAY, E. H.; FAROUK, A. Fidchain: Federated intrusion detection system for blockchain-enabled iot healthcare applications. **Healthcare (Switzerland)**, MDPI, v. 10, 6 2022. ISSN 22279032. Citado nas páginas 26, 47 e 54.

AZARIA, A.; EKBLAW, A.; VIEIRA, T.; LIPPMAN, A. MedRec: Using blockchain for medical data access and permission management. In: **Proceedings - 2016 2nd International Conference on Open and Big Data, OBD 2016**. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2016. p. 25–30. ISBN 9781509040544. Citado nas páginas 25, 42 e 51.

BAO, X.; SU, C.; XIONG, Y.; HUANG, W.; HU, Y. Flchain: A blockchain for auditable federated learning with trust and incentive. In: . [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2019. p. 151–159. ISBN 9781728140247. Citado nas páginas 26, 45 e 52.

BAUCAS, M. J.; SPACHOS, P.; PLATANIOTIS, K. N. Federated learning and blockchain-enabled fog-iot platform for wearables in predictive healthcare. **IEEE Transactions on Computational Social Systems**, Institute of Electrical and Electronics Engineers (IEEE), p. 1–10, 1 2023. Citado nas páginas 50 e 54.

BAUDER, R. A.; Da Rosa, R. C.; KHOSHGOFTAAR, T. M. Identifying medicare provider fraud with unsupervised machine learning. In: **Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018**. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2018. p. 285–292. ISBN 9781538626597. Citado nas páginas 40 e 51.

BENHAMOUDA, F.; HALEVI, S.; HALEVI, T. Supporting private data on hyperledger fabric with secure multiparty computation. **Proceedings - 2018 IEEE International Conference on Cloud Engineering, IC2E 2018**, Institute of Electrical and Electronics Engineers Inc., p. 357–363, may 2018. Citado na página 43.

BEUTEL, D. J.; TOPAL, T.; MATHUR, A.; QIU, X.; FERNANDEZ-MARQUES, J.; GAO, Y.; SANI, L.; LI, K. H.; PARCOLLET, T.; PORTO, P.; GUSMãO, B. D.; LANE, N. D. Flower: A friendly federated learning research framework. 2022. Disponível em: <https://flower.dev>. Citado na página 27.

BOTTOU, L. Stochastic gradient descent tricks. In: ____. **Neural Networks: Tricks of the Trade: Second Edition**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 421–436. ISBN 978-3-642-35289-8. Disponível em: <https://doi.org/10.1007/978-3-642-35289-8_25>. Citado na página 36.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Diário Oficial [da] União**, Brasília, DF, 2018. ISSN 1677-7042. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm>. Citado na página 25.

BUTERIN, V. Ethereum: A next-generation smart contract and decentralized application platform. In: . [S.l.: s.n.], 2014. Citado nas páginas 26 e 33.

B.V., E. **Scopus**. 2023. <https://www.scopus.com/home.uri>, acessado em 27/02/2023. Citado na página 40.

CELESTI, A.; RUGGERI, A.; FAZIO, M.; GALLETTA, A.; VILLARI, M.; ROMANO, A. Blockchain-based healthcare workflow for tele-medical laboratory in federated hospital iot clouds. **Sensors (Switzerland)**, MDPI AG, v. 20, 5 2020. ISSN 14248220. Citado nas páginas 48 e 54.

CHAN, K. C.; ZHOU, X.; GURURAJAN, R.; ZHOU, X.; ALLY, M.; GARDINER, M. Integration of Blockchains with Management Information Systems. **Proceedings of the 2019 International Conference on Mechatronics, Robotics and Systems Engineering, MoRSE 2019**, Institute of Electrical and Electronics Engineers Inc., p. 157–162, dec 2019. Citado na página 32.

CHANG, Y.; FANG, C.; SUN, W. A blockchain-based federated learning method for smart healthcare. **Computational Intelligence and Neuroscience**, Hindawi Limited, v. 2021, 2021. ISSN 16875273. Citado nas páginas 26, 46 e 54.

CHO, Y. J.; WANG, J.; JOSHI, G. Client selection in federated learning: Convergence analysis and power-of-choice selection strategies. **arXiv preprint arXiv:2010.01243**, 2020. Citado na página 36.

CHRISTIDIS, K.; DEVETSIKIOTIS, M. **Blockchains and Smart Contracts for the Internet of Things**. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2016. 2292–2303 p. Citado na página 32.

CLIM, A.; ZOTA, R. D.; CONSTANTINESCU, R. Data exchanges based on blockchain in m-health applications. In: **Procedia Computer Science**. [S.l.]: Elsevier B.V., 2019. v. 160, p. 281–288. ISSN 18770509. Citado na página 39.

DIETTERICH, T. G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. **Neural Computation**, v. 10, n. 7, p. 1895–1923, 10 1998. ISSN 0899-7667. Disponível em: <https://doi.org/10.1162/089976698300017197>. Citado na página 34.

DOMINGOS, P. A few useful things to know about machine learning. **Communications of the ACM**, ACM, v. 55, n. 10, p. 78–87, 2012. Citado na página 33.

ELDIN, A. M.; HOSSNY, E.; WASSIF, K.; OMARA, F. A. Federated blockchain system (fbs) for the healthcare industry. **Scientific reports**, NLM (Medline), v. 13, p. 2569, 12 2023. ISSN 20452322. Citado nas páginas 47 e 54.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. d. L. F. d. **Inteligência artificial: uma abordagem de aprendizado de máquina**. [S.l.]: LTC, 2011. Citado nas páginas 33 e 34.

FOUNDATION, H. **Hyperledger Besu – Hyperledger Foundation**. 2023. <https://www.hyperledger.org/use/besu>, acessado em 11/01/2023. Citado nas páginas 26 e 63.

GOLDMAN, B. The news on the street: prescription drugs on the black market. **Cmaj**, Can Med Assoc, v. 159, n. 2, p. 149–150, 1998. Citado na página 27.

GOOGLE. **Google Acadêmico**. 2023. <https://scholar.google.com.br>, acessado em 11/01/2023. Citado na página 40.

HALEVY, A.; NORVIG, P.; PEREIRA, F. The unreasonable effectiveness of data. **IEEE Intelligent Systems**, IEEE, v. 24, n. 2, p. 8–12, 2009. Citado na página 33.

HASHIM, F.; SHUAIB, K.; SALLABI, F. Connected blockchain federations for sharing electronic health records. **Cryptography**, MDPI, v. 6, 9 2022. ISSN 2410387X. Citado nas páginas 43 e 52.

HOUDA, Z. A. E.; HAFID, A. S.; KHOUKHI, L.; BRIK, B. When collaborative federated learning meets blockchain to preserve privacy in healthcare. **IEEE Transactions on Network Science and Engineering**, IEEE Computer Society, 2022. ISSN 23274697. Citado nas páginas 49 e 54.

ISMAIL, L.; MATERWALA, H. Blockchain paradigm for healthcare: Performance evaluation. **Symmetry**, MDPI AG, v. 12, n. 8, aug 2020. ISSN 20738994. Citado nas páginas 31 e 41.

JAIN, A.; DUIN, R.; MAO, J. Statistical pattern recognition: a review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 1, p. 4–37, 2000. Citado na página 34.

JENSEN, P. B.; JENSEN, L. J.; BRUNAK, S. Mining electronic health records: Towards better research applications and clinical care. **Nature Reviews Genetics**, v. 13, n. 6, p. 395–405, 2012. Cited By :813. Disponível em: <www.scopus.com>. Citado na página 25.

JOHNSON, J. M.; KHOSHGOFTAAR, T. M. Medicare fraud detection using neural networks. v. 6, p. 63, 2019. Disponível em: <https://doi.org/10.1186/s40537-019-0225-0>. Citado nas páginas 41 e 51.

KEMMOE, V. Y.; STONE, W.; KIM, J.; KIM, D.; SON, J. Recent Advances in Smart Contracts: A Technical Overview and State of the Art. **IEEE Access**, v. 8, p. 117782–117801, 2020. ISSN 21693536. Citado nas páginas 33 e 39.

KIM, H.; KIM, S. H.; HWANG, J. Y.; SEO, C. Efficient privacy-preserving machine learning for blockchain network. **IEEE Access**, Institute of Electrical and Electronics Engineers Inc., v. 7, p. 136481–136495, 2019. ISSN 21693536. Citado nas páginas 42 e 51.

KLEINBAUM, D. G.; KLEIN, M. **Logistic Regression: A Self-Learning Text**. [S.l.]: Springer Science & Business Media, 2010. Citado nas páginas 34 e 35.

KONEčNý, J.; MCMAHAN, H. B.; YU, F. X.; RICHTáRIK, P.; SURESH, A. T.; BACON, D. **Federated Learning: Strategies for Improving Communication Efficiency**. arXiv, 2016. Disponível em: <https://arxiv.org/abs/1610.05492>. Citado na página 36.

KUO, T.-T.; GABRIEL, R. A.; OHNO-MACHADO, L. Fair compute loads enabled by blockchain: sharing models by alternating client and server roles. **Journal of the American Medical Informatics Association**, v. 26, p. 392–403, 5 2019. ISSN 1527-974X. Disponível em: <https://doi.org/10.1093/jamia/ocy180>. Citado na página 51.

KUO, T. T.; PHAM, A. Detecting model misconducts in decentralized healthcare federated learning. **International Journal of Medical Informatics**, Elsevier Ireland Ltd, v. 158, 2 2022. ISSN 18728243. Citado nas páginas 50, 51 e 54.

LAKHAN, A.; MOHAMMED, M. A.; NEDOMA, J.; MARTINEK, R.; TIWARI, P.; VIDYARTHI, A.; ALKHAYYAT, A.; WANG, W. Federated-learning based privacy preservation and fraud-enabled blockchain iomt system for healthcare. **IEEE Journal of Biomedical and Health Informatics**, Institute of Electrical and Electronics Engineers Inc., 2022. ISSN 21682208. Citado nas páginas 47 e 54.

LAPEYRE-MESTRE, M.; GONY, M.; CARVAJAL, A.; MACIAS, D.; CONFORTI, A.; D'INCAU, P.; HEERDINK, R.; STICHELE, R. Van der; BERGMAN, U.; GROUP, O. E. S. A european community pharmacy-based survey to investigate patterns of prescription fraud through identification of falsified prescriptions. **European addiction research**, S. Karger AG Basel, Switzerland, v. 20, n. 4, p. 174–182, 2014. Citado na página 39.

LI, X.; HUANG, K.; YANG, W.; WANG, S.; ZHANG, Z. On the convergence of fedavg on non-iid data. **arXiv preprint arXiv:1907.02189**, 2019. Citado na página 36.

LI, Y.; CHEN, C.; LIU, N.; HUANG, H.; ZHENG, Z.; YAN, Q. A blockchain-based decentralized federated learning framework with committee consensus. **IEEE Network**, Institute of Electrical and Electronics Engineers Inc., v. 35, p. 234–241, 3 2021. ISSN 1558156X. Citado nas páginas 26, 43 e 52.

LI, Z.; LIU, J.; HAO, J.; WANG, H.; XIAN, M. Crowdsfl: A secure crowd computing framework based on blockchain and federated learning. **Electronics (Switzerland)**, MDPI AG, v. 9, 5 2020. ISSN 20799292. Citado nas páginas 26, 43 e 52.

LU, Y.; HUANG, X.; ZHANG, K.; MAHARJAN, S.; ZHANG, Y. Communication-efficient federated learning and permissioned blockchain for digital twin edge networks. **IEEE Internet of Things Journal**, Institute of Electrical and Electronics Engineers Inc., v. 8, p. 2276–2288, 2 2021. ISSN 23274662. Citado nas páginas 45 e 52.

MACKEY, T. K.; MIYACHI, K.; FUNG, D.; QIAN, S.; SHORT, J. Combating health care fraud and abuse: Conceptualization and prototyping study of a blockchain antifraud framework. **Journal of Medical Internet Research**, JMIR Publications Inc., v. 22, n. 9, p. e18623, sep 2020. ISSN 14388871. Disponível em: <https://www.jmir.org/2020/9/e18623>. Citado nas páginas 25, 42 e 51.

MARKOFF, J. Entrepreneurs see a web guided by common sense. **The New York Times**, 2006. Disponível em: <https://www.nytimes.com/2006/11/12/business/12web.html>. Citado na página 25.

MCMAHAN, H. B.; MOORE, E.; RAMAGE, D.; HAMPSON, S.; ARCAS, B. A. y. Communication-efficient learning of deep networks from decentralized data. 2 2016. Disponível em: <http://arxiv.org/abs/1602.05629>. Citado nas páginas 26, 35 e 36.

MITCHELL, T. The discipline of machine learning. **Carnegie Mellon University**, v. 10, n. 1-2, 2006. Citado na página 33.

NAKAMOTO, S. Bitcoin: A peer-to-peer electronic cash system. **Decentralized Business Review**, p. 21260, 2008. Citado nas páginas 26 e 32.

NGUYEN, D. C.; DING, M.; PHAM, Q. V.; PATHIRANA, P. N.; LE, L. B.; SENEVIRATNE, A.; LI, J.; NIYATO, D.; POOR, H. V. Federated learning meets blockchain in edge computing: Opportunities and challenges. **IEEE Internet of Things Journal**, Institute of Electrical and Electronics Engineers Inc., v. 8, p. 12806–12825, 8 2021. ISSN 23274662. Citado na página 37.

PASSERAT-PALMBACH, J.; FARNAN, T.; MILLER, R.; GROSS, M. S.; FLANNERY, H. L.; GLEIM, B. A blockchain-orchestrated federated learning architecture for healthcare consortia. 10 2019. Disponível em: <http://arxiv.org/abs/1910.12603>. Citado nas páginas 49 e 54.

PENG, Z.; XU, J.; CHU, X.; GAO, S.; YAO, Y.; GU, R.; TANG, Y. Vfchain: Enabling verifiable and auditable federated learning via blockchain systems. **IEEE Transactions on Network Science and Engineering**, IEEE Computer Society, v. 9, p. 173–186, 2022. ISSN 23274697. Citado nas páginas 26, 45 e 52.

PHILLIPS, J. Prescription drug abuse: problem, policies, and implications. **Nursing outlook**, Elsevier, v. 61, n. 2, p. 78–84, 2013. Citado na página 27.

POŁAP, D.; SRIVASTAVA, G.; JOLFAEI, A.; PARIZI, R. M. Blockchain technology and neural networks for the internet of medical things; blockchain technology and neural networks for the internet of medical things. In: . [S.l.: s.n.], 2020. Citado nas páginas 44 e 52.

PRASAD, V.; JENA, A. B. Prespecified falsification end points: Can they validate true observational associations? **JAMA - Journal of the American Medical Association**, v. 309, n. 3, p. 241–242, 2013. Cited By :140. Disponível em: <www.scopus.com>. Citado na página 39.

RAMANAN, P.; NAKAYAMA, K. Baffle : Blockchain based aggregator free federated learning. In: . [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2020. p. 72–81. ISBN 9780738104959. Citado nas páginas 37, 44 e 52.

REHMAN, A.; ABBAS, S.; KHAN, M. A.; GHAZAL, T. M.; ADNAN, K. M.; MOSAVI, A. A secure healthcare 5.0 system based on blockchain technology entangled with federated learning technique. **Computers in Biology and Medicine**, Elsevier Ltd, v. 150, 11 2022. ISSN 18790534. Citado nas páginas 48 e 54.

SAMUEL, O.; OMOJO, A. B.; ONUJA, A. M.; SUNDAY, Y.; TIWARI, P.; GUPTA, D.; HAFEEZ, G.; YAHAYA, A. S.; FATOBA, O. J.; SHAMSHIRBAND, S. Iomt: A covid-19 healthcare system driven by federated learning and blockchain. **IEEE Journal of Biomedical and Health Informatics**, Institute of Electrical and Electronics Engineers Inc., 2 2022. ISSN 21682208. Citado nas páginas 47 e 54.

SANTOS, H. D. P. dos; ULBRICH, A. H. D. P. S.; WOLOSZYN, V.; VIEIRA, R. Ddc-outlier: Preventing medication errors using unsupervised learning. **IEEE Journal of Biomedical and Health Informatics**, v. 23, n. 2, p. 874–881, March 2019. ISSN 2168-2194. Citado nas páginas 27 e 62.

SEO, J.; MENDELEVITCH, O. Identifying frauds and anomalies in Medicare-B dataset. In: **Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS**. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2017. p. 3664–3667. ISBN 9781509028092. ISSN 1557170X. Citado nas páginas 40 e 51.

SETTIPALLI, L.; GANGADHARAN, G. R. Qfbn: Quorum based federated blockchain network for healthcare system to avoid multiple benefits and data breaches. **IEEE Consumer Electronics Magazine**, Institute of Electrical and Electronics Engineers Inc., 2022. ISSN 21622256. Citado nas páginas 50 e 54.

SHARMA, P. K.; PARK, J. H.; CHO, K. Blockchain and federated learning-based distributed computing defence framework for sustainable society. **Sustainable Cities and Society**, Elsevier Ltd, v. 59, 8 2020. ISSN 22106707. Citado nas páginas 44 e 52.

SHAYAN, M.; FUNG, C.; YOON, C. J.; BESCHASTNIKH, I. Biscotti: A blockchain system for private and secure federated learning. **IEEE Transactions on Parallel and Distributed Systems**, IEEE Computer Society, v. 32, p. 1513–1525, 7 2021. ISSN 15582183. Citado nas páginas 26, 44 e 52.

SINGH, S.; RATHORE, S.; ALFARRAJ, O.; TOLBA, A.; YOON, B. A framework for privacy-preservation of iot healthcare data using federated learning and blockchain technology. **Future Generation Computer Systems**, Elsevier B.V., v. 129, p. 380–388, 4 2022. ISSN 0167739X. Citado nas páginas 26, 46 e 54.

SOUTO, M. C. P.; LORENA, A. C.; DELBEM, A. C. B.; CARVALHO, A. C. P. d. L. F. Técnicas de aprendizado de máquina para problemas de biologia molecular. In: **Congresso da Sociedade Brasileira de Computação**. [S.l.]: SBC, 2003. Citado na página 34.

STEPHANIE, V.; KHALIL, I.; ATIQUZZAMAN, M.; YI, X. Trustworthy privacy-preserving hierarchical ensemble and federated learning in healthcare 4.0 with blockchain. **IEEE Transactions on Industrial Informatics**, IEEE Computer Society, 2022. ISSN 19410050. Citado nas páginas 49 e 54.

STRANG, J.; BABOR, T.; CAULKINS, J.; FISCHER, B.; FOXCROFT, D.; HUMPHREYS, K. Drug policy and the public good: Evidence for effective interventions. **The Lancet**, v. 379, n. 9810, p. 71–83, 2012. Citado na página 39.

TANWAR, S.; PAREKH, K.; EVANS, R. Blockchain-based electronic healthcare record system for healthcare 4.0 applications. **Journal of Information Security and Applications**, Elsevier Ltd, v. 50, feb 2020. ISSN 22142126. Citado na página 31.

TURGAY, S. Blockchain management and federated learning adaptation on healthcare management system. **International Journal of Intelligent Systems and Applications**, v. 14, p. 1–13, 10 2022. ISSN 2074904X. Disponível em: <https://www.mecs-press.org/ijisa/ijisa-v14-n5/v14n5-1.html>. Citado nas páginas 50 e 54.

UDDIN, M.; SALAH, K.; JAYARAMAN, R.; PESIC, S.; ELLAHHAM, S. Blockchain for drug traceability: Architectures and open challenges. **Health informatics journal**, SAGE Publications Sage UK: London, England, v. 27, n. 2, p. 14604582211011228, 2021. Citado na página 33.

WANG, Z.; HU, Q. Blockchain-based federated learning: A comprehensive survey. arXiv, 2021. Disponível em: <https://arxiv.org/abs/2110.02182>. Citado nas páginas 37 e 85.

WHITE, C.; READY, J.; KATZ, C. M. Examining how prescription drugs are illegally obtained: Social and ecological predictors. **Journal of Drug Issues**, v. 46, n. 1, p. 4–23, 2016. Disponível em: <https://doi.org/10.1177/0022042615608502>. Citado na página 25.

XU, J.; GLICKSBERG, B. S.; SU, C.; WALKER, P.; BIAN, J.; WANG, F. Federated learning for healthcare informatics. **Journal of Healthcare Informatics Research**, v. 5, p. 1–19, 2021. ISSN 2509-498X. Disponível em: <https://doi.org/10.1007/s41666-020-00082-4>. Citado na página 36.

YANG, Q.; LIU, Y.; CHEN, T.; TONG, Y. Federated machine learning: Concept and applications. 2 2019. Disponível em: <http://arxiv.org/abs/1902.04885>. Citado na página 36.

YASHRAJ, R.; #1, G.; SAI, S.; #2, M.; KUMAR, P.; #3, B. A Comparative Study of Using Various Machine Learning and Deep Learning-Based Fraud Detection Models For Universal Health Coverage Schemes. **International Journal of Engineering Trends and Technology**, v. 69, p. 96–102, 2021. Disponível em: <https://github.com/RohanYashraj/Healthcare-Fraud->. Citado nas páginas 41 e 51.

ZHANG, C.; XIAO, X.; WU, C. Medical fraud and abuse detection system based on machine learning. **International Journal of Environmental Research and Public Health**, v. 17, n. 19, p. 1–11, 2020. ISSN 16604601. Disponível em: <www.mdpi.com/journal/ijerph>. Citado na página 41.

ZHANG, W.; LU, Q.; YU, Q.; LI, Z.; LIU, Y.; LO, S. K.; CHEN, S.; XU, X.; ZHU, L. Blockchain-based federated learning for device failure detection in industrial iot. **IEEE Internet of Things Journal**, Institute of Electrical and Electronics Engineers Inc., v. 8, p. 5926–5937, 4 2021. ISSN 23274662. Citado nas páginas 26, 45 e 52.

ZHENG, Z.; XIE, S.; DAI, H.-N.; CHEN, W.; CHEN, X.; WENG, J.; IMRAN, M. An overview on smart contracts: Challenges, advances and platforms. **Future Generation Computer Systems**, v. 105, p. 475–491, 2020. Disponível em: <https://doi.org/10.1016/j.future.2019.12.019>. Citado na página 33.

ZYSKIND, G.; NATHAN, O.; PENTLAND, A. S. Decentralizing privacy: Using blockchain to protect personal data. In: **Proceedings - 2015 IEEE Security and Privacy Workshops, SPW 2015**. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2015. p. 180–184. ISBN 9781479999330. Citado nas páginas 25 e 42.


ICMC
USP
SÃO CARLOS