



UNIVERSIDADE ESTADUAL DA PARAÍBA
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA EM SAÚDE

HERÁCLIO ALMEIDA DA COSTA

**APRENDIZADO FEDERADO: ESTUDO DA INFLUÊNCIA DE ARTEFATOS SOBRE
O MODELO CENTRAL NA ANÁLISE DE MICROCALCIFICAÇÕES POR
MAMOGRAFIA**

Campina Grande - PB

2023

HERÁCLIO ALMEIDA DA COSTA

**APRENDIZADO FEDERADO: ESTUDO DA INFLUÊNCIA DE ARTEFATOS SOBRE
O MODELO CENTRAL NA ANÁLISE DE MICROCALCIFICAÇÕES POR
MAMOGRAFIA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia em Saúde da Universidade Estadual da Paraíba como requisito para obtenção do título de Mestre em Ciência e Tecnologia em Saúde.

Área de Concentração: Desenvolvimento de Produtos e Processos para Tecnologia em Saúde.

Tema: Inteligência Artificial aplicada ao diagnóstico médico.

Orientador: Prof. Dr. Edmar Candeia Gurjão

Campina Grande - PB
2023

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

C838a Costa, Heraclio Almeida da.
Aprendizado federado [manuscrito] : estudo da influência de artefatos sobre o modelo central na análise de microcalcificações por mamografia / Heraclio Almeida da Costa. - 2023.
48 p. : il. colorido.

Digitado.
Dissertação (Mestrado em Profissional em Ciência e Tecnologia em Saúde) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2023.
"Orientação : Prof. Dr. Edmar Candeia Gurjão, UFCG - Universidade Federal de Campina Grande."
1. Aprendizado federado. 2. Aprendizado de máquinas. 3. Mamografia. 4. Artefatos de imagem. I. Título
21. ed. CDD 610.28


HERÁCLIO ALMEIDA COSTA

**APRENDIZADO FEDERADO: ESTUDO DA INFLUÊNCIA DE ARTEFATOS
SOBRE O MODELO CENTRAL NA ANÁLISE DE MICROCALCIFICAÇÕES
POR MAMOGRAFIA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia em Saúde da Universidade Estadual da Paraíba como requisito para obtenção do título de Mestre em Ciência e Tecnologia em Saúde.

Dissertação aprovada em: 24/04/2023

BANCA EXAMINADORA:



Prof. Dr. Edmar Candeia Gurjão
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Frederico Moreira Bublitz
Universidade Estadual da Paraíba (UEPB)



Prof. Dra. Luciana Ribeiro Veloso
Universidade Federal de Campina Grande (UFCG)

RESUMO

Contexto: O crescente uso de técnicas de aprendizado de máquinas na área de diagnóstico médico levanta questionamentos sobre a segurança de dados sensíveis dos pacientes e sobre compartilhamento de dados para treinamento, fomentando o estudo de soluções inovadoras para tais problemas. A técnica de aprendizado federado permite o compartilhamento de informações para treinamento de algoritmos em colaboração multicêntrica, com agregação num servidor central destas informações sem a necessidade de compartilhamento dos bancos de dados, protegendo dados sensíveis e eliminando a necessidade de grande fluxo de dados para transmissão. Entretanto, existe desconhecimento por parte do servidor central a respeito da composição dos bancos de dados usados para treinamento. Quando do uso de redes convolucionais para diagnóstico por imagem, a falta de conhecimento do servidor diz respeito em especial à qualidade das imagens utilizadas e presença de artefatos. **Objetivos:** Avaliar a influência de artefatos de imagem no treinamento de um algoritmo de aprendizado de máquinas usando a técnica de aprendizado federado. **Métodos:** Utilizando-se imagens de mamografia, procedeu-se simulação de artefatos de movimento e de artefatos de poeira, avaliando-se diferentes cenários de treinamento. Uma rede neural convolucional de código aberto foi adaptada para classificar as imagens em 03 categorias: imagens sem microcalcificações; imagens com microcalcificações benignas; e imagens com microcalcificações suspeitas. Os treinamentos foram realizados antes e depois da adição dos artefatos simulados. **Resultados:** A adição de artefatos a imagens nos bancos de dados para treinamento resultou em influência no desempenho no algoritmo utilizado, variando em função do tamanho do banco de dados e do tipo de artefato, com potenciais efeitos benéficos. **Conclusão:** A utilização de imagens com artefatos na composição dos bancos de dados pode beneficiar o treinamento de algoritmos de aprendizado de máquina por técnica de aprendizado federado.

Palavras-chave: Aprendizado federado; Aprendizado de máquinas; Mamografia; Artefatos de imagem.

ABSTRACT

Background: The growing in use of machine learning techniques for medical diagnosis raises questions about the security of sensitive patient data and about sharing data for training, encouraging the study of innovative solutions to such problems. The federated learning technique allows the sharing of information for algorithm training in multicentric collaboration, with aggregation in a central server, without sharing databases, protecting sensitive data and eliminating the need for a large data stream for transmission. However, there is a lack of knowledge on the part of the central server regarding the composition of the databases used for training. When using convolutional networks for image diagnosis, the server's lack of knowledge concerns in particular the quality of the images used and the presence of artifacts.

Objectives: Evaluate the influence of image artifacts in the training of a machine learning algorithm using the federated learning technique. **Methods:** Using mammography images, motion artifacts and dust artifacts were simulated, evaluating different training scenarios. An open source convolutional neural network was adapted to classify the images into 03 categories: images without microcalcifications; images with benign microcalcifications; and images with suspicious microcalcifications. Training was performed before and after the addition of simulated artifacts. **Results:** The addition of artifacts to images in the training databases influenced the performance of the algorithm, varying according to the size of the database and the type of artifact, with potential beneficial effects. **Conclusion:** The use of images with artifacts in the composition of databases can benefit the training of machine learning algorithms using a federated learning technique.

Keywords: Federated learning; Machine learning; Mammography; Image artifacts.

SUMÁRIO

1. INTRODUÇÃO	7
2. HIPÓTESE	9
3. OBJETIVOS	10
3.1. Objetivo geral	10
3.2. Objetivos específicos	10
4. REFERENCIAL TEÓRICO	11
4.1 Aprendizado de máquina	11
4.2 Aprendizado federado	12
5. MÉTODO DA PESQUISA	16
5.1 Rede neural	17
5.2 Inserção de artefatos	18
5.3 Treinamento dos modelos	18
6. RESULTADOS	19
6.1 Treinamento do modelo sem federação	19
6.2 Treinamento por aprendizado federado antes da inserção de artefatos	19
6.3 Treinamento por aprendizado federado com inserção de artefatos (cenário 1)	24
6.4 Treinamento por aprendizado federado com inserção de artefatos (cenário 2)	27
6.5 Treinamento por aprendizado federado com inserção de artefatos (cenário 3)	31
6.6 Treinamento por aprendizado federado com inserção de artefatos (cenário 4)	34
6.7 Treinamento por aprendizado federado com exclusão das imagens com artefato do cliente A	38
6.8 Treinamento por aprendizado federado com exclusão das imagens com artefato do cliente C	39

7.	DISCUSSÃO	41
8.	CONCLUSÃO	44
9.	REFERÊNCIAS	45

1. INTRODUÇÃO

O uso de imagens de alta resolução do corpo humano permitiu um grande salto na capacidade de diagnóstico de doenças, utilizando-se do resultado da interação de energias eletromagnéticas ou mecânicas (no caso de ultrassonografia) com os tecidos biológicos (PANAYIDES et al, 2020). A evolução tecnológica vem sendo acelerada por agregação de sistemas computadorizados, causando maior potencial para o diagnóstico médico (SANTOS et al, 2019), com melhoria na acurácia dos exames, na consistência da interpretação de imagens médicas, na avaliação prognóstica e no suporte à decisão terapêutica (SANTOS et al, 2019).

O uso crescente de dados neste meio, incluindo os registros clínicos e imagens médicas diagnósticas, gerou um volume de informações sem precedentes, o que viabilizou novas aplicações, como o desenvolvimento de ferramentas baseadas em aprendizado de máquinas (AM) para fins diagnósticos e a distribuição de dados por tecnologia de computação na nuvem (SAHI & LI, 2016; LUNDERVOLD A & LUNDERVOLD A, 2018; PANAYIDES et al, 2019). Entretanto, apesar do enorme potencial das ferramentas computacionais baseadas em AM, ainda há limitações para seu uso na rotina clínica. Uma destas limitações se refere à heterogeneidade do registro dos dados médicos, o que limita seu uso em iniciativas de análise colaborativa entre diferentes centros, gerando a necessidade de ferramentas de padronização, como sistemas de informação eletrônicos específicos, a exemplo dos sistemas de armazenamento e comunicação de imagens (do inglês *picture archive and communication systems* - PACS) e dos sistemas de informação em radiologia (*radiology information systems* – RIS), bem como de formatos de arquivos digitais como o DICOM (*digital imaging and communications in medicine*) e o *Health Level 7* (HL7), facilitando a interoperabilidade dos dados (PANAYIDES et al, 2020).

Além da heterogeneidade dos dados, o uso e o desempenho de arquiteturas de AM também são limitados pelo tamanho e pela qualidade do dados utilizados nos treinamentos, usualmente sendo necessário um grande volume de dados e com diversidade de características capaz de resultar numa boa generalização do modelo treinado. Uma forma de minimizar tais limitações é o compartilhamento público de bancos de dados e o lançamento de desafios abertos à comunidade científica, que possibilitam troca de experiências com difusão de arquiteturas mais eficazes e de acesso a dados de populações distintas. Apesar de técnicas mais modernas

que buscam otimizar o treinamento de modelos com menor volume de dados, ainda há poucas iniciativas de compartilhamento de dados públicos, dificultando as tarefas do domínio médico (PANAYIDES et al, 2020).

Frente uma necessidade crescente de compartilhamento de dados, surge a preocupação quanto à preservação de informações sensíveis dos usuários dos sistemas eletrônicos de saúde. Uma série de questionamentos éticos e legais vem sendo levantada na busca de meios de proteção de dados e de formas de compartilhamento entre centros distintos, com o objetivo de manter segurança aliada a uma menor perda/transformação dos dados originais durante a transmissão, evitando que tenham sua utilidade reduzida (MENDES & VILELA, 2017). Diversas técnicas têm sido avaliadas para manutenção da privacidade de dados coletados, desde anonimização simples até recursos mais avançados de criptografia e baseados em redes neurais (SHAH & GULATI, 2016; SAHI & LI, 2016).

Uma proposta recente para resolver questões de privacidade relacionadas ao compartilhamento de informações entre múltiplos centros e a necessidade de transmissão de grande volume de dados para treinamento de algoritmos de AM em imagens médicas é o aprendizado federado - AF (WILLEMINK et al, 2020). Esta técnica permite a transmissão do resultado dos treinamentos locais e as informações embutidas neles sem o compartilhamento de dados sensíveis dos bancos utilizados no treinamento, gerando um modelo central que pode ser prontamente utilizado ou redistribuído para novas iterações locais (MACMAHAN et al., 2017). Nesta técnica, ocorre a transmissão do resultado de treinamentos locais sem compartilhamento do banco de dados, fazendo com que o responsável pelo agregamento central destas informações desconheça o conjunto de dados utilizados nos treinamentos (WILLEMINK et al, 2020). Especificamente no diagnóstico médico por imagem, tal desconhecimento gera potenciais suspeições sobre a qualidade das imagens utilizadas nos treinamentos, como no caso da presença de artefatos que configurem falhas e que comprometam a qualidade final da imagem, podendo resultar em informações perdidas ou mascaradas, com falhas de interpretação durante a leitura das mesmas. Diante da potencial influência que estes artefatos também possam gerar no processo de AM, surge a necessidade de avaliar o impacto do uso de imagens com artefatos no desempenho de redes neurais, especialmente nas treinadas por técnica de aprendizado federado.

2. HIPÓTESE

Baseando-se nos pressupostos de que a preparação de imagens médicas para AM requer confiabilidade dos dados de rotulação, qualidade das imagens, padronização de informações e uma boa representatividade populacional, minimizando eventuais vieses, tem-se por hipótese que o uso de imagens médicas com artefatos no AM no treinamento por técnica de aprendizado federado pode influenciar no desempenho dos modelos.

3. OBJETIVOS

3.1. Objetivo Geral

Avaliar a influência do uso de imagens de mamografia com artefatos no desempenho de rede neural treinada por técnica de aprendizado federado.

3.2. Objetivos Específicos

- Gerar um novo banco de dados derivado de banco de dados público, com recorte de áreas de interesse em imagens de mamografia;
- Selecionar um modelo de rede neural convolucional previamente concebido e de código aberto, adaptando-o aos resultados de interesse do estudo, procedendo-se treinamento do modelo com o banco de dados contendo todas as imagens disponíveis, com avaliação de seu desempenho;
- Segmentar o banco de dados e implementar o processo de AF, permitindo comparação do desempenho do modelo por esta técnica;
- Inserir artefatos de movimento e artefatos simulando microcalcificações em parte das imagens utilizadas nos treinamentos por AF, em cenários distintos, com avaliação do impacto no desempenho do modelo final.

4. REFERENCIAL TEÓRICO

4.1 Aprendizado de máquina

O conceito de aprendizado de máquina (AM) pode ser aplicado a todo algoritmo que, ao ser treinado por um banco de dados, consegue extrair informação e posteriormente realizar predições, com otimização de parâmetros para melhoria da performance de sua tarefa. Dentre as várias arquiteturas existentes, as redes neurais correspondem ao arquétipo do método, onde camadas de entrada, camadas ocultas e camadas de saída alinham-se com funções de ativação e pesos aplicados aos valores de entrada, lançando um resultado final que pode ser comparado com o desfecho inicialmente esperado. Esta comparação permite um ajuste dos pesos até que se obtenha a máxima eficiência possível do algoritmo (ERICKSON et al, 2017).

Os treinamentos normalmente utilizam de dois a três grupos de dados (aprendizado, validação e teste) e podem ser executados em técnicas distintas, como o treinamento supervisionado e o treinamento não supervisionado (ERICKSON et al, 2017). Tomando-se por exemplo lesões detectadas em exames de mamografia, no aprendizado supervisionado o banco de dados para treinamento possui rotulações com dados essenciais para classificação das lesões, permitindo que o algoritmo tenha todas as informações sobre a classificação que se propõe das imagens, por exemplo quanto a achados suspeitos para malignidade. No treinamento não supervisionado, os dados possuem diferenças passíveis de classificação, porém não são fornecidos com rotulações, resultando-se separação de resultados em grupos distintos pelo algoritmo, mas sem classificação do resultado.

Nas tarefas aplicadas às imagens médicas, as funções de segmentação e de classificação são as mais utilizadas, demonstrando bons resultados, em especial com o aumento do poder de processamento computacional, o que permitiu o desenvolvimento de modelos de redes neurais com um maior número de camadas, denominados modelos de aprendizado profundo (*deep learning*), sendo as redes neurais convolucionais modelos de destaque para análise de imagens médicas (PANAYIDES et al, 2020; ERICKSON et al, 2017; CHARTRAND et al, 2017). Nestas redes, filtros convolucionais são aplicados às imagens para extração de atributos, permitindo a progressão na rede de informações relevantes, bem como redimensionamento de matrizes, possibilitando melhor custo computacional sem perda de características essenciais das imagens.

No AM, deve-se atentar para alguns fatores limitantes de desempenho dos modelos, como o viés de seleção, que consiste em erro relacionado a características sistemáticas dos grupos populacionais utilizados num estudo clínico e que possam diferir significativamente dos grupos externos ao estudo, levando à formação de banco de dados utilizado com baixa representatividade da população geral. Caso ocorra, o viés de seleção pode tornar impraticável a validação externa do modelo de rede neural ou fazer com que o mesmo obtenha resultados substancialmente aquém do esperado quando utilizado em outros grupos populacionais (YU & JOHN, 2020). Outra preocupação sobre o treinamento de uma rede neural profunda é quanto ao dimensionamento dos bancos de dados, uma vez que essas redes usualmente precisam de grande volume de dados para atingirem uma boa performance, tornando-se necessário alto de tempo para a construção desses bancos (CHARTRAND et al, 2017). O compartilhamento massivo de dados poderia mitigar estes problemas, mas tal compartilhamento encontra diversas barreiras, incluindo fatores técnicos, motivacionais, econômicos, legais e éticos (VAN PANHUIS et al 2014).

4.2 Aprendizado Federado

O aprendizado federado é uma técnica de aprendizado de máquina que permite a união de informações e distribuição de modelos em diferentes dispositivos, como computadores, smartphones e dispositivos vestíveis, de forma colaborativa e remota, sem compartilhamento de dados sensíveis, tendo sido primeiramente proposta por Brendan MacMahan et al. (2017). A ideia consiste no compartilhamento de informações oriundas de treinamento local de um modelo de aprendizado de máquinas para um servidor central, que é responsável pela combinação destas informações com as informações resultantes do treinamento em outros dispositivos. Um modelo novo, com mesma estrutura, porém com pesos modificados, é gerado e redistribuído para os dispositivos, a fim de realizar treinamentos locais, num ciclo que pode ser pré-determinado ou contínuo (Figura 1). Esta técnica permite que os usuários obtenham o benefício de um modelo treinado com dados ricos sem a necessidade de um armazenamento central destes, poupando recursos de armazenamento e de transmissão massiva de dados, e mantendo a privacidade dos dados brutos entre os usuários em si e com o próprio servidor central (WILLEMINK et al, 2020).

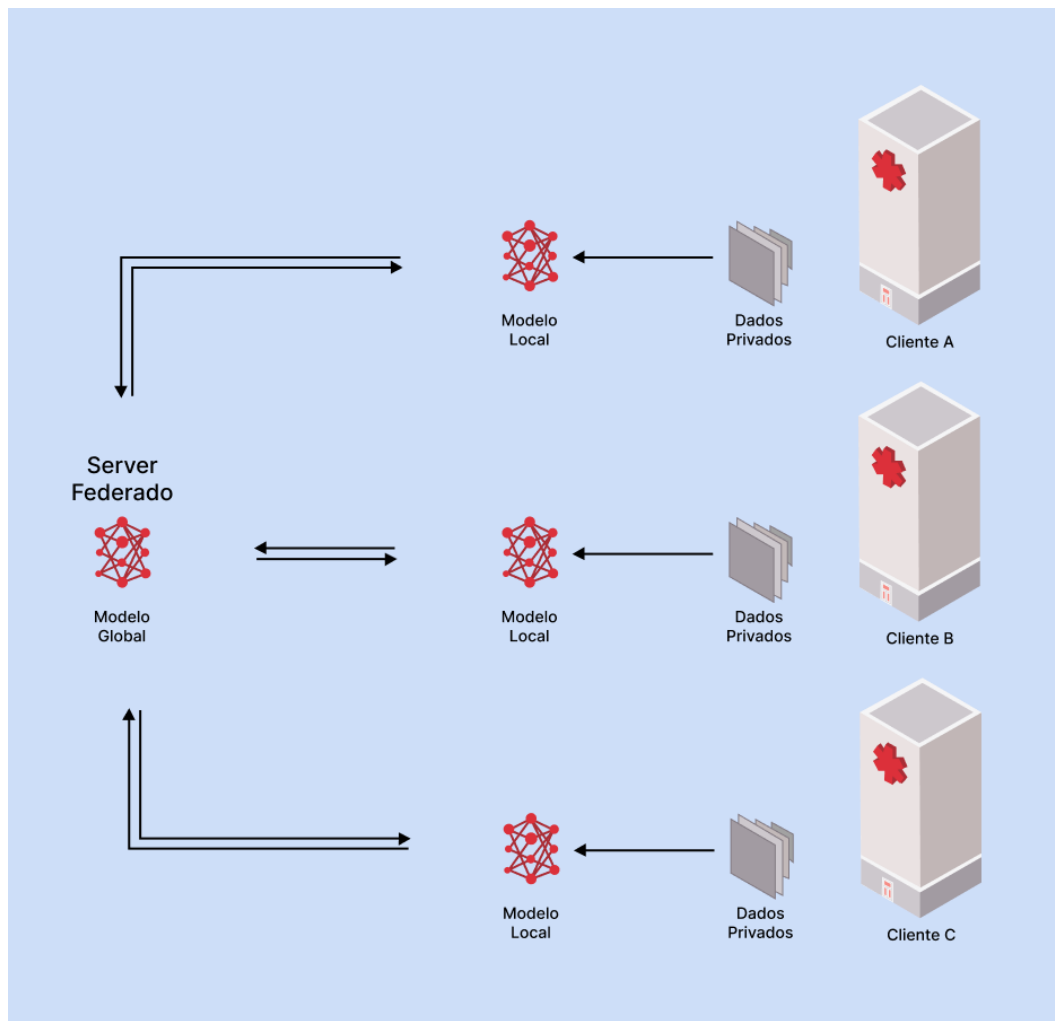


Figura 1 – Estrutura básica do aprendizado federado.

A técnica foi concebida com alguns pontos chaves para otimização, com destaque para a possibilidade de treinamento local com dados personalíssimos e distribuição massiva, além de desbalanceamento no volume de dados entre os clientes, assumindo-se inicialmente um modo de atualização sincrônico com rodadas de comunicação e redistribuição do modelo central atualizado para novos treinamentos locais. O algoritmo FederatedAveraging utilizado originalmente leva em consideração o volume de dados de cada cliente, procedendo uma média ponderada dos resultados dos gradientes:

$$\min_{w \in \mathbb{R}^d} f(w) \quad \text{onde} \quad f(w) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n f_i(w).$$

$$f(w) = \sum_{k=1}^K \frac{n_k}{n} F_k(w) \quad \text{onde} \quad F_k(w) = \frac{1}{n_k} \sum_{i \in \mathcal{P}_k} f_i(w).$$

Algoritmo FederatedAveraging. Os clientes K são indexados por k ; B é o tamanho do minilote local, E é o número de épocas locais e η é a taxa de aprendizado.

Server executes:
initialize w_0
for each round $t = 1, 2, \dots$ **do**
 $m \leftarrow \max(C \cdot K, 1)$
 $S_t \leftarrow$ (random set of m clients)
for each client $k \in S_t$ **in parallel do**
 $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
 $w_{t+1} \leftarrow \sum_{k=1}^K \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w): // Run on client k
 $\mathcal{B} \leftarrow$ (split \mathcal{P}_k into batches of size B)
for each local epoch i from 1 to E **do**
for batch $b \in \mathcal{B}$ **do**
 $w \leftarrow w - \eta \nabla \ell(w; b)$
return w to server

Entretanto, novos desenhos de fluxo de trabalho estão sendo desenvolvidos, incluído fluxos onde o processo de combinação de dados se dá também de forma descentralizada e dinâmica e sem necessidade de compartilhamento de todas as informações do modelo a cada rodada de treinamento, bem como variações da técnica para combinação mais eficiente das informações e em ambientes de grande heterogeneidade de dados (REIKE et al, 2020).

O aprendizado federado já possui amplo espectro de utilização, com aplicações nas áreas de saúde, mecanismos de recomendação, internet das coisas e *edge computing* (WEN et al, 2023). O uso desta técnica já está uso na saúde, com enfoque no diagnóstico por imagem, onde alguns desafios precisam ser superados (CHENG et al, 2021). Roth et al (2020) conduziram um estudo multicêntrico internacional para o desenvolvimento e aplicação prática de modelo de classificação para densidade mamária utilizando imagens de mamografia. Neste estudo, foi utilizado um modelo previamente concebido para classificação das imagens e foi implementado o algoritmo FederatedAveraging para a agregação dos resultados dos treinamentos locais, partindo-se de bancos de dados com características heterogêneas quanto a volume de dados, resoluções de imagem e técnicas de aquisição. Apesar de não ser atingida grande acurácia nos modelos, foi possível demonstrar que o aprendizado federado possibilitou melhoria da performance e uma melhor capacidade de generalização. Mais recentemente, frente a pandemia de COVID-19, pode-se observar um aumento do interesse promovendo aceleração

na implementação da técnica no diagnóstico por imagem (NAZ et al, 2021), porém outras áreas de conhecimento no meio médico tem se beneficiado da técnica de aprendizado federado, como os modelos de risco de mortalidade para diversas condições clínicas (OH & NADKARNI, 2023).

Apesar de bastante promissora, há algumas limitações a serem exploradas, como preocupações relacionadas a agentes maliciosos quanto ao *upload* de informações suspeitas, potencial grande número de clientes gerando alto custo de comunicação, a heterogeneidade de dados e a heterogeneidade de sistemas e dispositivos (ZHANG et al, 2022; DARZIDEHKALANI et al, 2022). Na utilização por diagnóstico por imagem, ao se proceder a preparação das imagens médicas dos bancos de dados, a confiabilidade dos dados de rotulação, a qualidade das imagens, a padronização de informações e a idoneidade do executante do treinamento são informações de grande relevância e são desconhecidas pelo servidor central e pelos demais participantes ao se aplicar o aprendizado federado (WILLEMINK et al, 2020). Dados locais com viés, intencional ou não, a utilização de dados com parâmetros fora do padrão proposto e ataques de segurança, como interceptação das transmissões, invasões aos bancos de dados locais e ataques de inversão do modelo, que permitem possível recriação de imagens, são problemas em potencial que abriam campo para estudo de ferramentas de mitigação de possíveis danos ao modelo central e de identificação de falta de equidade na composição dos bancos de dados (KAIROUZ et al 2021; PANDA et al, 2021, ZHANG K et al., 2022).

Tolpegin et al (2020) demonstraram que a simples inversão de rotulação nas imagens de treinamento de uma rede neural convolucional em um cliente foi capaz de impactar negativamente o modelo global, demonstrando-se a necessidade de métodos de proteção contra agentes maliciosos. O “envenenamento” intencional de um modelo não parece ser razoável ao se tratar de colaboração entre centros de saúde para o treinamento de rede para diagnóstico médico. Entretanto, preocupações com a qualidade das imagens utilizadas localmente mostraram-se pertinentes, especialmente no que se refere aos artefatos de imagens e falhas não intencionais de rotulações dos dados.

Outro desafio diz respeito ao armazenamento ao longo do tempo dos bancos de dados inicialmente utilizados para os treinamentos. No caso de imagens médicas, os custos de armazenamento podem se tornar grandes problemas frente ao volume dos arquivos, já havendo pesquisas sobre como otimizar o aprendizado federado ao longo do tempo sem que necessariamente haja o treinamento com os dados iniciais, podendo ser utilizados dados novos em substituição aos iniciais (Li B, 2022).

5 MÉTODO DA PESQUISA

O estudo foi realizado nas dependências do Departamento de Engenharia Elétrica da Universidade Federal de Campina Grande e foi direcionado para microcalcificações em exames de mamografia, achado de grande importância no diagnóstico de câncer de mama. Foi utilizado um banco de dados público de mamografias em formato DICOM, sem perda de qualidade em relação à aquisição e com anotações prévias a respeito de microcalcificações suspeitas, contendo originalmente 400 imagens (LOIZIDOU et al, 2021). Um médico radiologista com mais de 10 anos de experiência em análise de mamografias procedeu a seleção de áreas de interesse com recortes de dimensões padronizadas (228 x 228 pixels) nas imagens originais das mamografias, separando-as em três grupos: presença de microcalcificações suspeitas; presença de microcalcificações benignas; e ausência de microcalcificações (contendo apenas o tecido mamário). Tal medida permitiu a ampliação do número de imagens mantendo-se conformidade com as características básicas necessária para classificação no sistema BI-RADS (SICKLES et al, 2013). Na Figura 2, listamos exemplos de cada categoria.

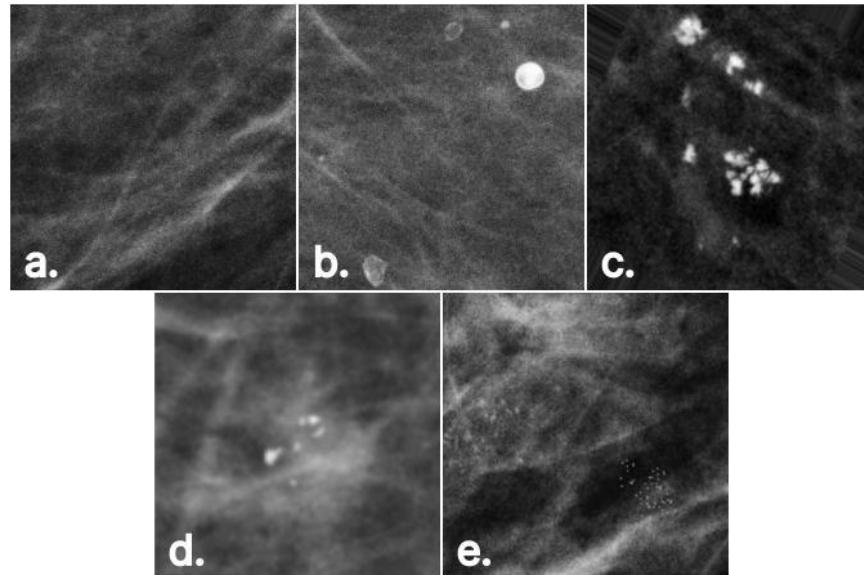


Figura 2 – Exemplos de imagens selecionadas de áreas de interesse obtidas a partir banco de dados original: (a) parênquima normal; (b) microcalcificações benignas; (c) Microcalcificações suspeitas de malignidade. Exemplo de imagens com aplicações de artefato: (d) imagem com artefato de movimento devido à aplicação do Blur; (e) Adição de artefato de poeira simulado microcalcificação (quadrante inferior direito da imagem).

Parte das imagens (10%) em cada categoria foi selecionada aleatoriamente e alocada para uso exclusivo na avaliação de desempenho dos modelos, sem uso para treinamento. As imagens restantes foram divididas em 03 clientes, simulando 03 centros médicos diferentes, em uma proporção de 3:2:1, denominados clientes A, B e C. Em cada banco, foram aplicadas técnicas de aumento de volume de dados (*data augmentation*), a exemplo de inversões horizontais, inversões verticais e rotações em 90°, elevando o número global de imagens sem que houvesse alteração nas características básicas das mesmas. O número final de imagens pode ser verificado na Tabela 1.

Classe	Cliente A	Cliente B	Cliente C
Normal	602	399	204
Benigno	594	397	204
Suspeito	248	174	84
Total	1444	970	492

Tabela 1 – Número de imagens nos bancos de dados para treinamento dos clientes A, B e C após técnicas de *data augmentation*.

5.1 Rede neural

Foi selecionada uma rede neural convolucional com arquitetura parcialmente pré-estabelecida e de código aberto, a Efficientnet (TAN M & LE QV, 2019), implementada em Python e utilizando as bibliotecas Keras e Tensor Flow. As camadas de saída do modelo foram adaptadas para três resultados possíveis: parênquima sem microcalcificações, presença de microcalcificações benignas e presença de microcalcificações suspeitas. O aprendizado federado foi iniciado com treinamento local do modelo com o banco de dados de cada cliente. A combinação central dos dados seguiu o algoritmo FederatedAveraging, utilizando a média ponderada dos pesos buscando manter a proporcionalidade da informação de acordo com o número de imagens em cada banco de dados. Para avaliação do modelo, utilizaram-se comandos em Python para a construção de curvas ROC e medição da área sob a curva (AUC) de cada categoria estudada. Também foram geradas matrizes de confusão, com os valores percentuais de cada célula, com posterior cálculo da acurácia global dos modelos.

5.2 Inserção de artefatos

A adição dos artefatos de imagem nas bases de dados foi feita através de simulação por ferramenta de desfoque (*blur*) para simular artefatos de movimento e a adição manual de diminutos focos com valores de cor próximo ao branco, simulando artefatos de poeira que podem levar a confusão com microcalcificações (Figura 2). A conformidade com os artefatos reais foi avaliada por médico radiologista com mais de 10 anos de experiência em avaliação de mamografias.

A inserção dos artefatos ocorreu em 04 cenários diferentes, formando novos bancos de dados: (1) adição de artefato de movimento em 10% das imagens do cliente A; (2) adição de artefato de movimento em 10% das imagens do cliente C; (3) adição de artefato de poeira a 10% das imagens do cliente A; (4) adição de artefato de poeira a 10% das imagens do cliente C. Optou-se por avaliar apenas os bancos de dados de dimensões extremas, permanecendo inalterado o intermediário.

5.3 Treinamento dos modelos.

Num primeiro momento, após a constituição do novo banco de dados global e retirada da amostra de cada categoria para formação de lotes de imagens para teste de desempenho, procedeu-se um treinamento sem a divisão dos dados em clientes, mantendo um cliente único que se beneficia por ter acesso a todas as imagens disponíveis para treinamento. Na sequência, após a divisão do banco de dados global, realizou-se o treinamento do modelo com a técnica de aprendizado federado entre 03 clientes, ainda sem a inserção de artefatos nas imagens. Na fase seguinte, para cada um dos quatro cenários já descritos anteriormente repetiu-se o processo de aprendizado federado, com geração de novas curvas ROC e avaliação da AUC e da acurácia dos modelos resultantes, permitindo uma comparação direta dos resultados para avaliação do impacto dos artefatos.

Por fim, frente os resultados negativos que foram obtidos com a adição de artefatos de poeira, procedeu-se a avaliação do modelo numa ação onde, ao invés da manutenção das imagens com artefatos, optou-se pela exclusão das mesmas antes do treinamento, levando a uma redução do banco de dados. Desta forma, novas duas situações foram geradas, uma onde procedeu-se a exclusão de 10% das imagens do banco de dados do cliente A e outra onde esta porcentagem foi aplicada ao banco de dados do cliente C.

6. RESULTADOS

6.1. Treinamento do modelo sem federação.

Antes da segmentação para o experimento de federação, foi realizado o treinamento do modelo utilizando-se todos os dados selecionados para este fim no banco de dados global, sendo realizadas 10 iterações. Foi registrado, então, o desempenho do modelo sem a inserção de artefatos e com o conhecimento dos dados totais, correspondendo a um cenário ideal onde seria possível ter acesso a todas as imagens.

Sob esta condição, o modelo obteve uma acurácia global de 84,9 %, com os seguintes valores de AUC:

	Normal	Benigno	Suspeito
AUC	0,97	0,94	0,92

6.2. Treinamento por aprendizado federado antes da inserção de artefatos.

Como primeira etapa do aprendizado federado, o modelo selecionado foi distribuído para cada um dos clientes, sendo procedido o treinamento local com 10 iterações e utilizando seus respectivos bancos de dados locais. O desempenho de cada modelo nos treinamentos locais pode ser verificado com as seguintes matrizes de confusão:

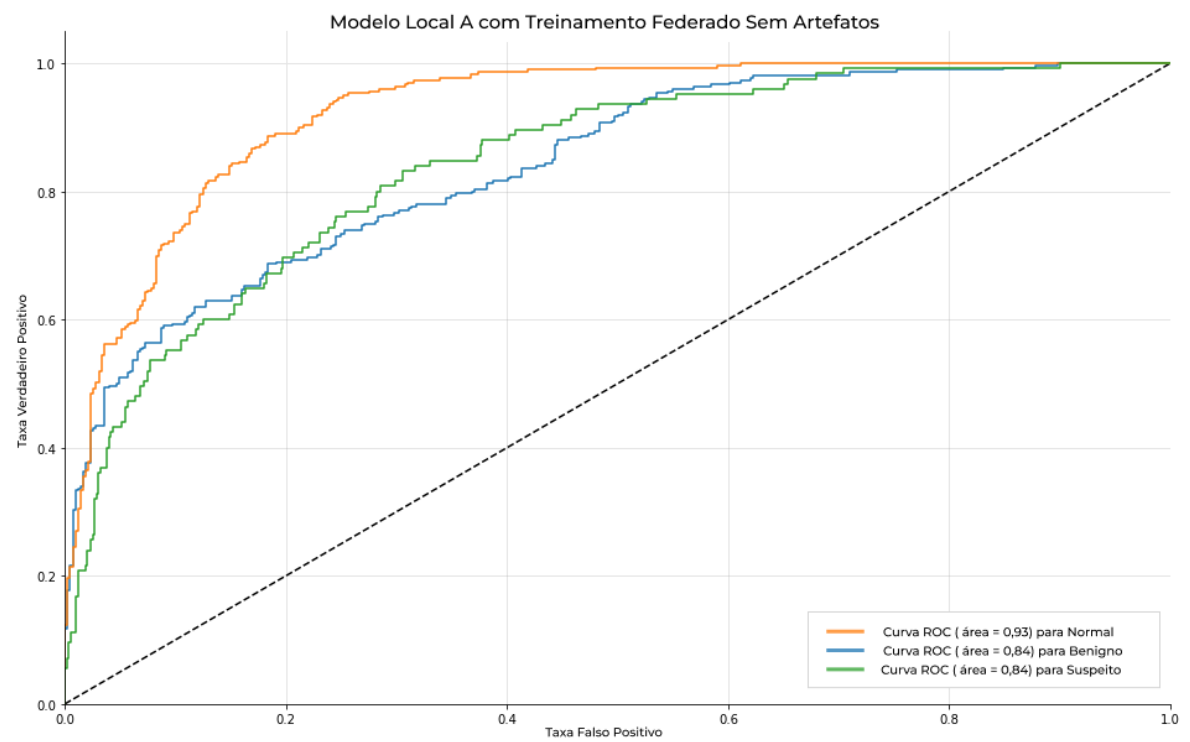
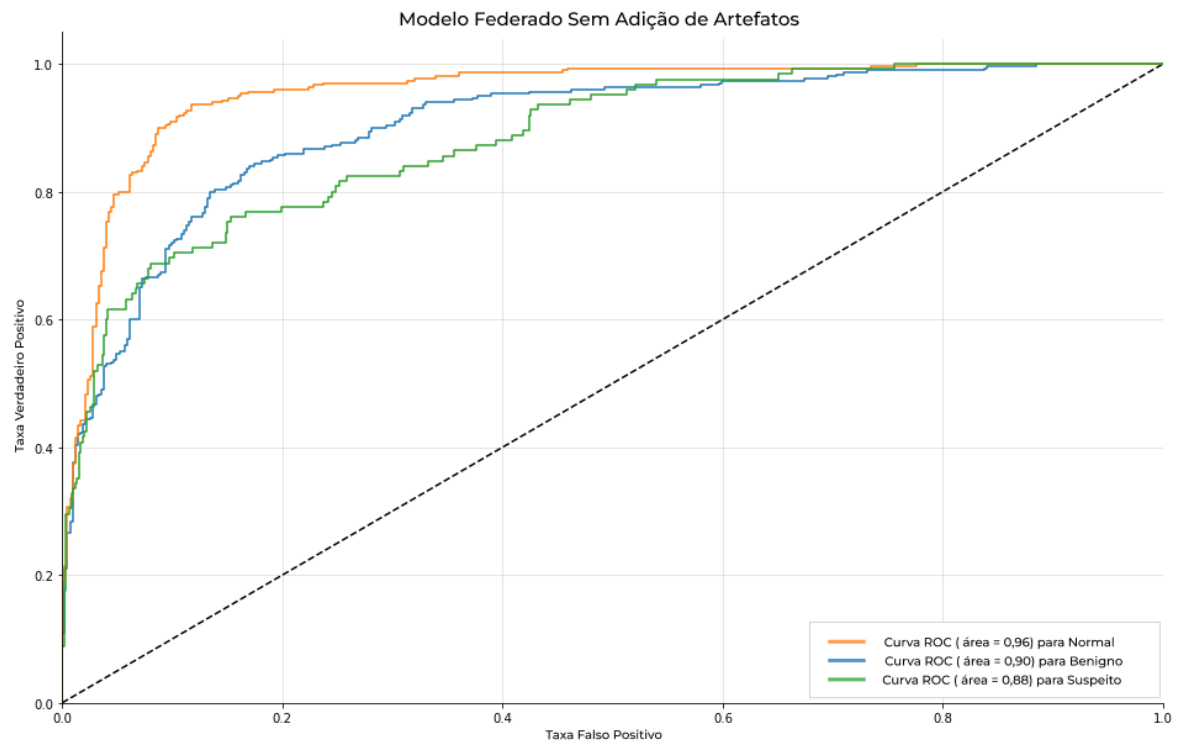
Modelo local A antes da federação (Acurácia de 74,17%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	51.66%	43.51%	79.01%
Verdadeiro Positivo	32.73%	32.46%	8.98%
Falso Negativo	8.56%	8.98%	8.29%
Falso Positivo	7.04%	15.06%	8.98%

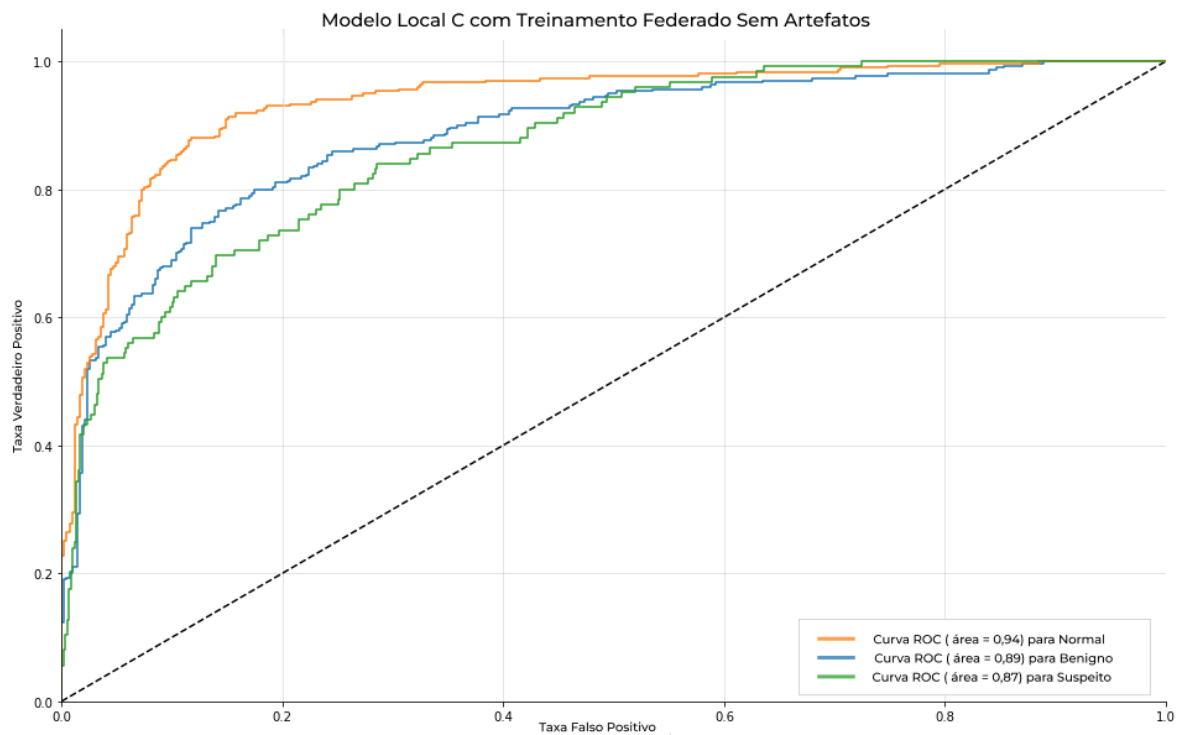
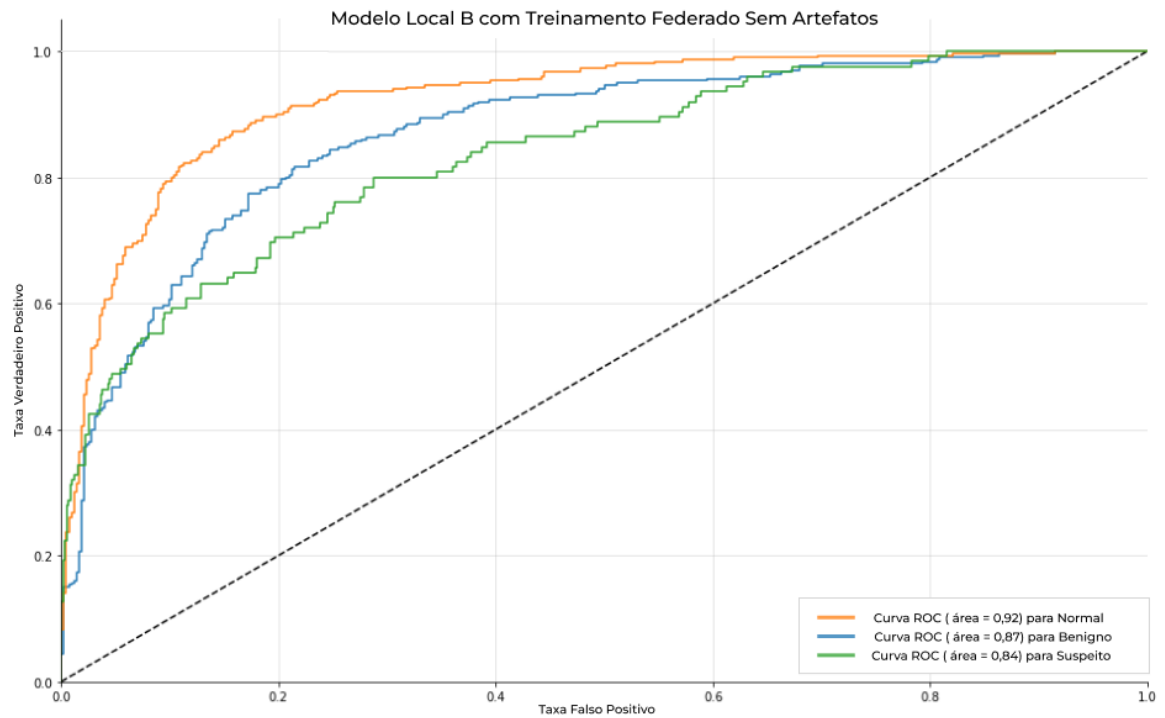
Modelo local B antes da federação (Acurácia de 75,41%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	50.97%	50.83%	73.62%
Verdadeiro Positivo	34.67%	29.01%	11.74%
Falso Negativo	6.63%	12.43%	5.52%
Falso Positivo	7.73%	7.73%	9.12%

Modelo local C antes da federação (Acurácia de 56,77%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	25%	49.59%	82.18%
Verdadeiro Positivo	40.06%	16.57%	0.14%
Falso Negativo	1.24%	24.86%	17.13%
Falso Positivo	33.70%	8.98%	0.55%

A seguir, executou-se a federação com o algoritmo FederatedAveraging, gerando um novo modelo, em cuja etapa de teste obteve uma AUC de 0,96 para o grupo de imagens normais, 0,90 para o grupo de microcalcificações benignas e 0,88 para o grupo de microcalcificações suspeitas, com uma acurácia geral de 78,45%. O modelo foi, então, redistribuído para os clientes locais a fim de uma avaliação de performance com seus pacotes de teste.

Os gráficos de curva ROC para o modelo federado e do desempenho deste nos clientes locais seguem abaixo:





A partir do aprendizado federado e da redistribuição do modelo para teste em cada cliente, foi possível a construção das seguintes matrizes de confusão:

Modelo Federado (Acurácia de 78,45%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	54.97%	43.92%	79.56%
Verdadeiro Positivo	33.29%	36.19%	8.98%
Falso Negativo	8.01%	5.25%	8.29%
Falso Positivo	3.73%	14.64%	3.18%

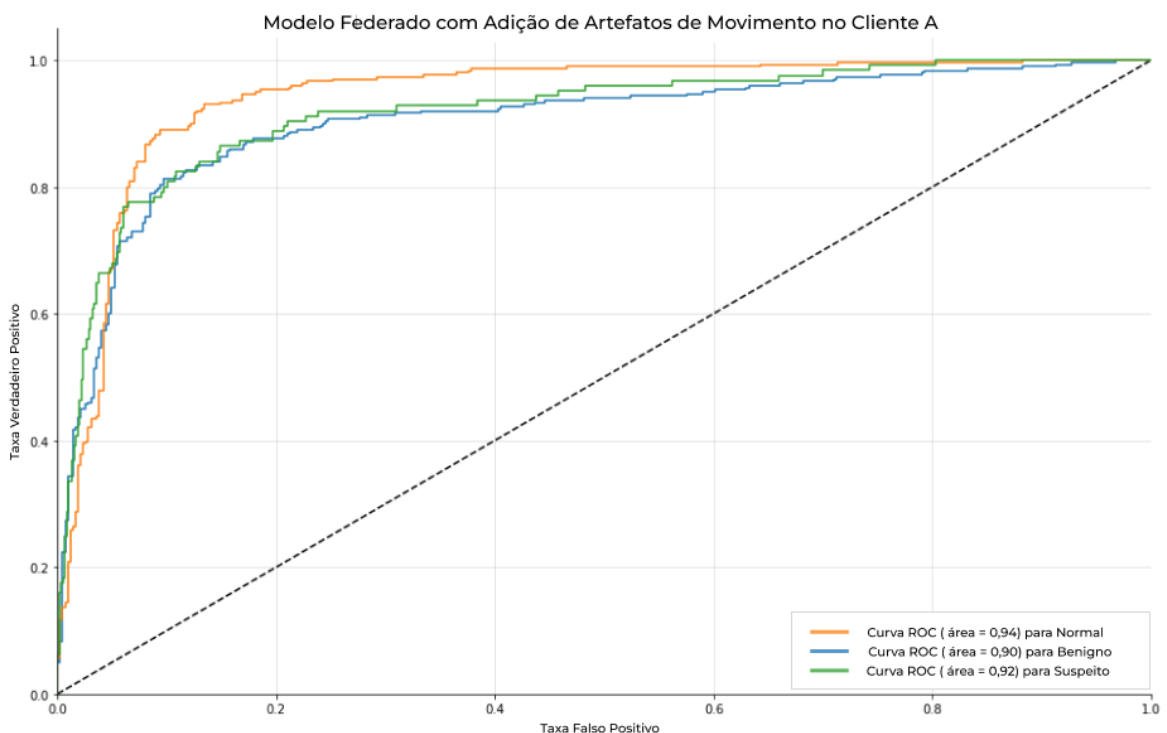
Modelo local A após federação (Acurácia de 69,96%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	52.90%	35.36%	79.70%
Verdadeiro Positivo	26.52%	34.67%	6.77%
Falso Negativo	14.78%	6.77%	10.50%
Falso Positivo	5.80%	23.20%	3.04%

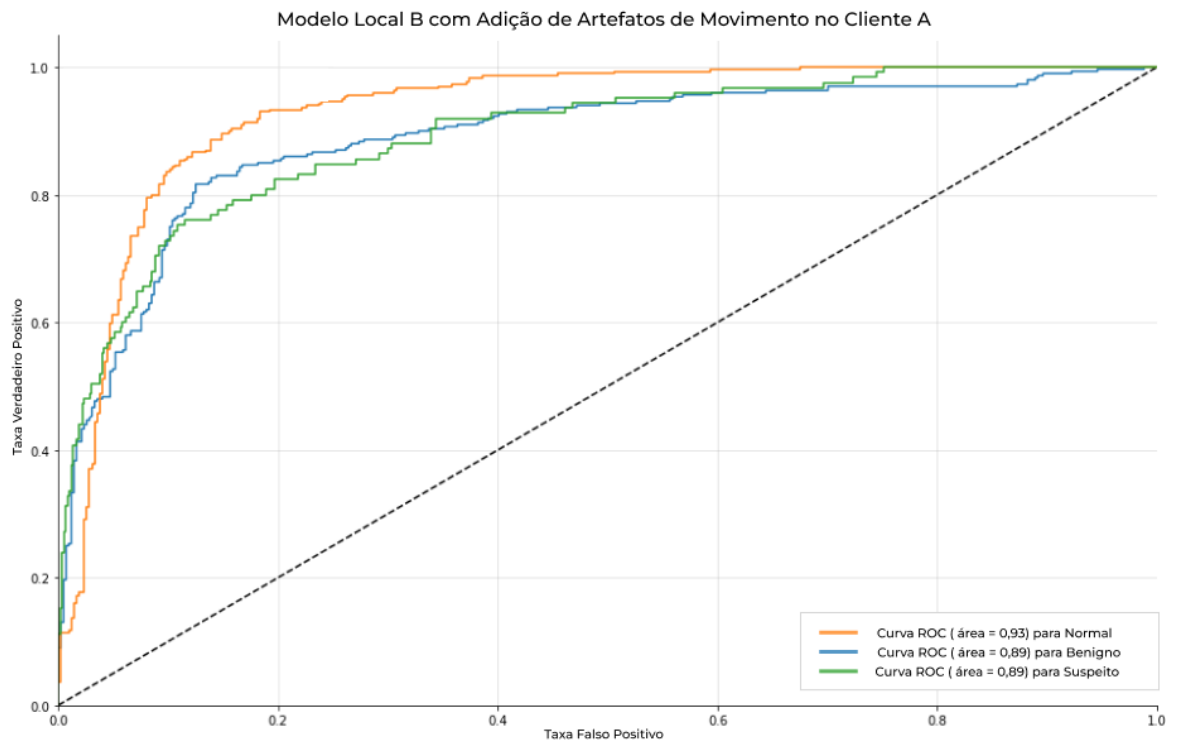
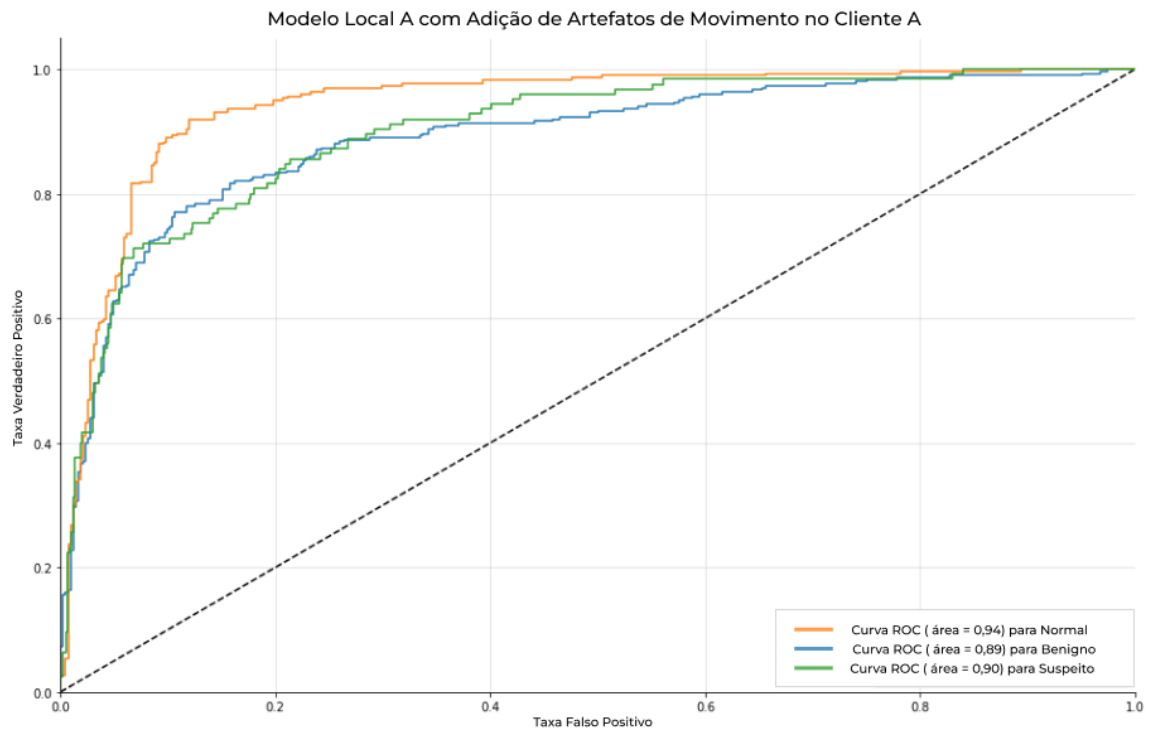
Modelo local B após federação (Acurácia de 75,83%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	52.35%	52.35%	78.31%
Verdadeiro Positivo	34,12%	34,12%	8.15%
Falso Negativo	7.18%	7.18%	9.12%
Falso Positivo	6.35%	6.35%	4.42%

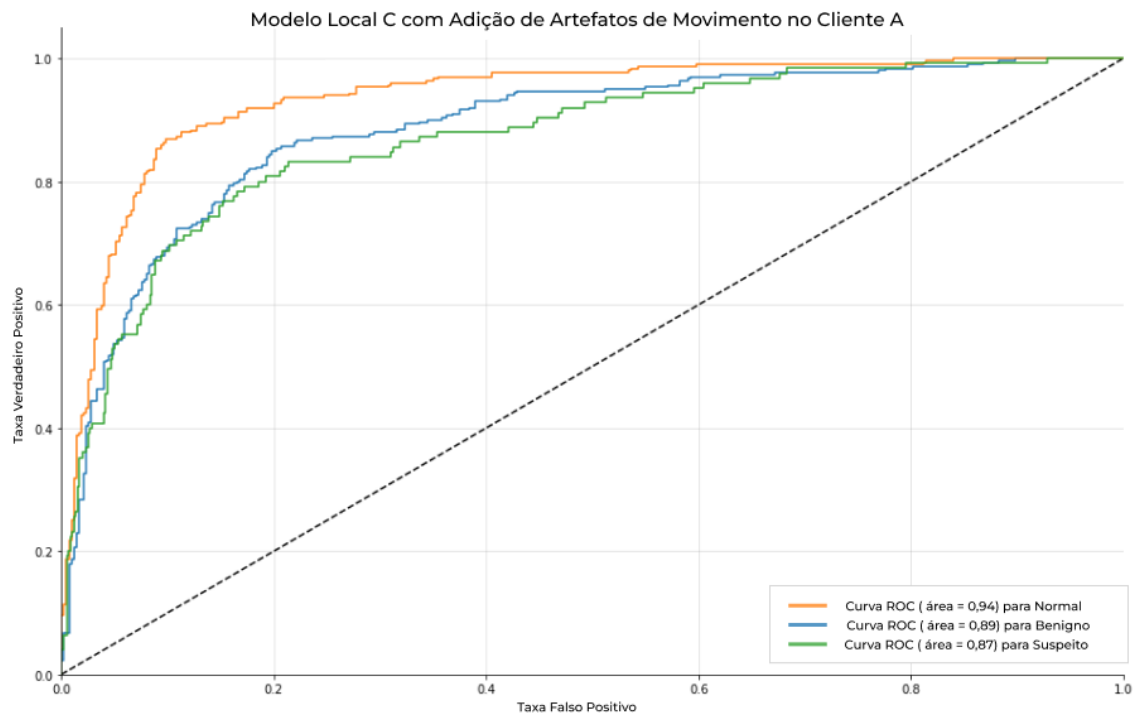
Modelo local C após federação (Acurácia de 70,44%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	57.18%	38.54%	74.72%
Verdadeiro Positivo	23.90%	36.60%	9.94%
Falso Negativo	17.40%	4.83%	7.32%
Falso Positivo	1.52%	20.03%	8.01%

6.3. Treinamento por aprendizado federado com inserção de artefatos (cenário 1).

Neste cenário, procedeu-se a aplicação de artefato de movimento em 10% das imagens do banco de dados do cliente A, havendo um aumento da acurácia do modelo federado para 81,63%, com queda da AUC em 2% para imagens normais, estabilidade para imagens benignas e aumento de 4,5% para imagens suspeitas. Seguem abaixo os respectivos gráficos ROC e matrizes de confusão para o modelo federado e seu desempenho local após redistribuição do mesmo:







Modelo Federado (Acurácia de 81,63%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	54,42%	50,14%	77,07%
Verdadeiro Positivo	34,53%	34,67%	12,43%
Falso Negativo	6,77%	6,77%	4,83%
Falso Positivo	4,28%	8,43%	5,66%

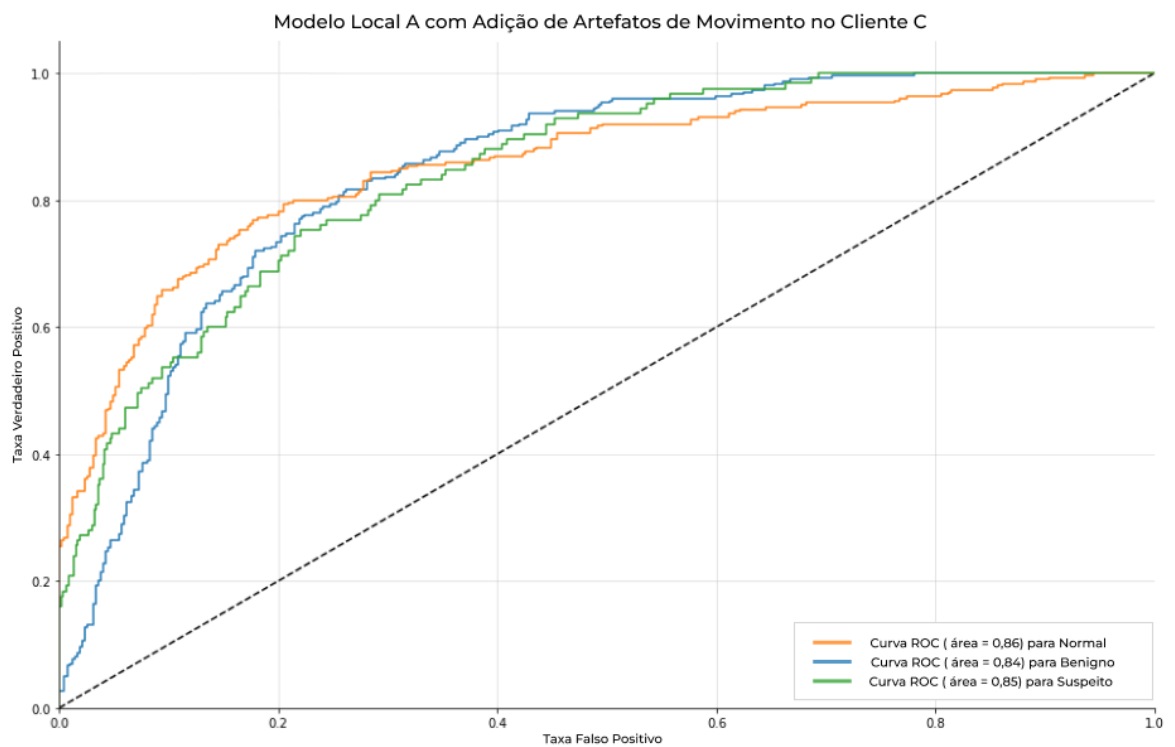
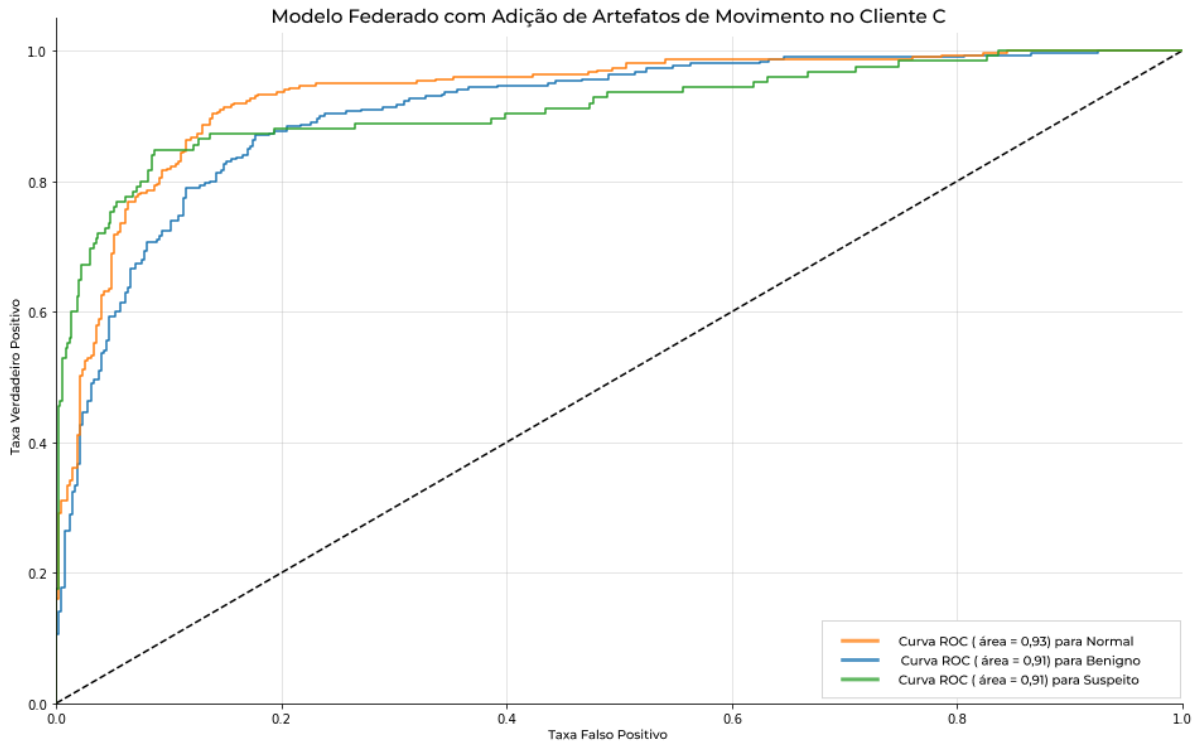
Modelo local A após federação (Acurácia de 80,39%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	49,31%	51,24%	79,83%
Verdadeiro Positivo	38,40%	32,73%	9,25%
Falso Negativo	2,90%	8,70%	8,01%
Falso Positivo	9,39%	7,32%	2,90%

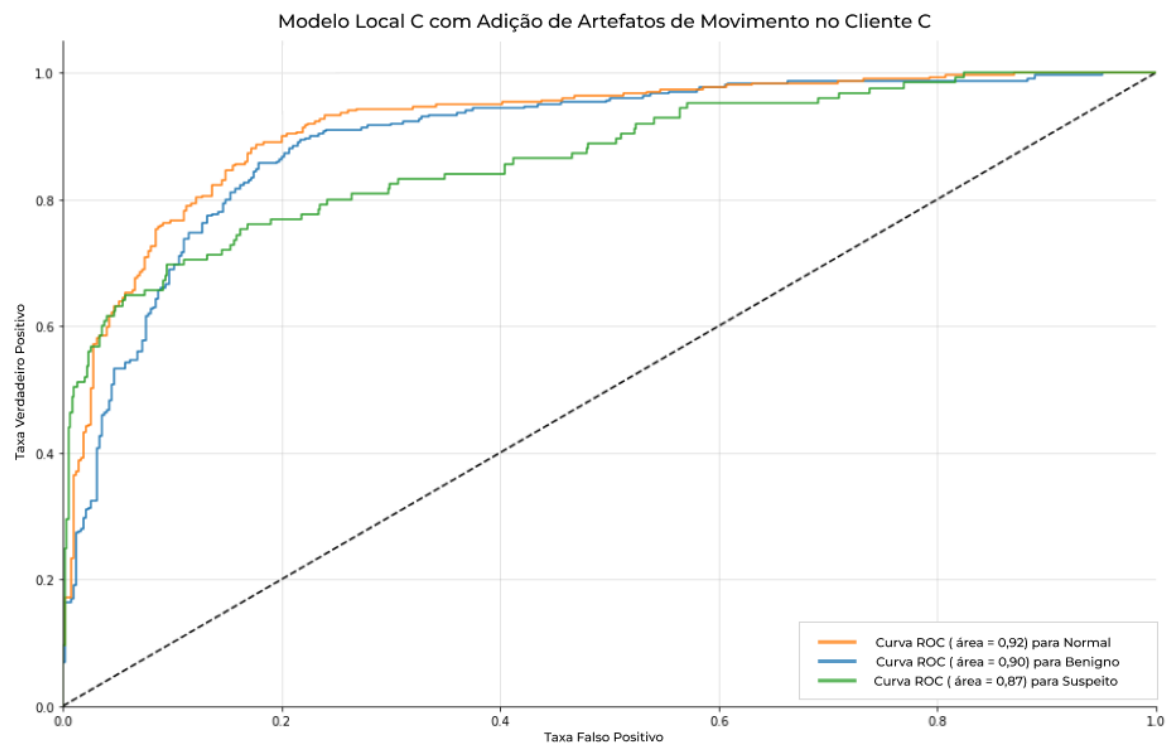
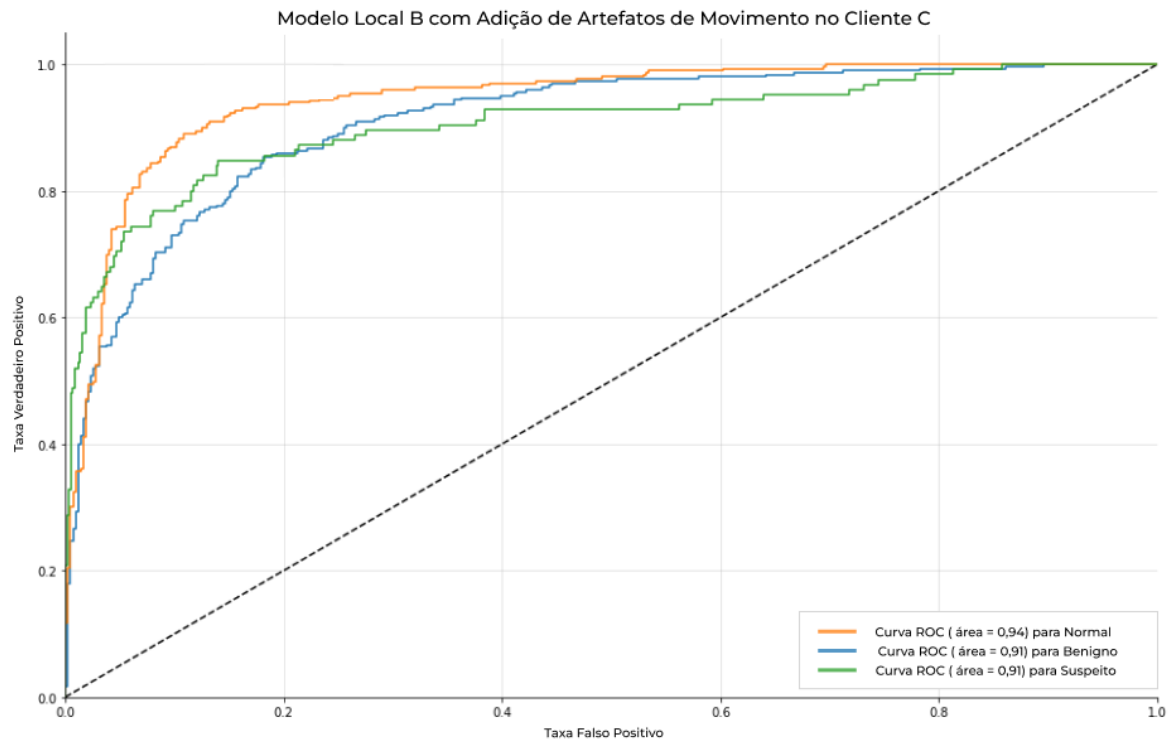
Modelo local B após federação (Acurácia de 79,01%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	53.31%	51.66%	74.03%
Verdadeiro Positivo	33.98%	32.32%	12.71%
Falso Negativo	7.32%	9.12%	4.56%
Falso Positivo	5.39%	6.91%	8.70%

Modelo local C após federação (Acurácia de 70,58%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	56.77%	38.95%	74.86%
Verdadeiro Positivo	23.34%	36.74%	10.50%
Falso Negativo	17.96%	4.70%	6.77%
Falso Positivo	1.93%	19.61%	7.87%

6.4. Treinamento por aprendizado federado com inserção de artefatos (cenário 2).

Neste cenário, procedeu-se a aplicação de artefato de movimento em 10% das imagens do banco de dados do cliente C, havendo uma estabilidade da acurácia do modelo federado em 78,45%, com queda da AUC em 3% para imagens normais, aumento em 1% para imagens benignas e aumento de 3% para imagens suspeitas. Seguem abaixo os respectivos gráficos ROC e matrizes de confusão para o modelo federado e seu desempenho local após redistribuição do mesmo:





Modelo Federado (Acurácia de 78,45%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	56,22%	44,61%	77,62%
Verdadeiro Positivo	28,59%	37,02%	12,85%
Falso Negativo	12,71%	4,42%	4,42%
Falso Positivo	2,49%	13,95%	5,11%

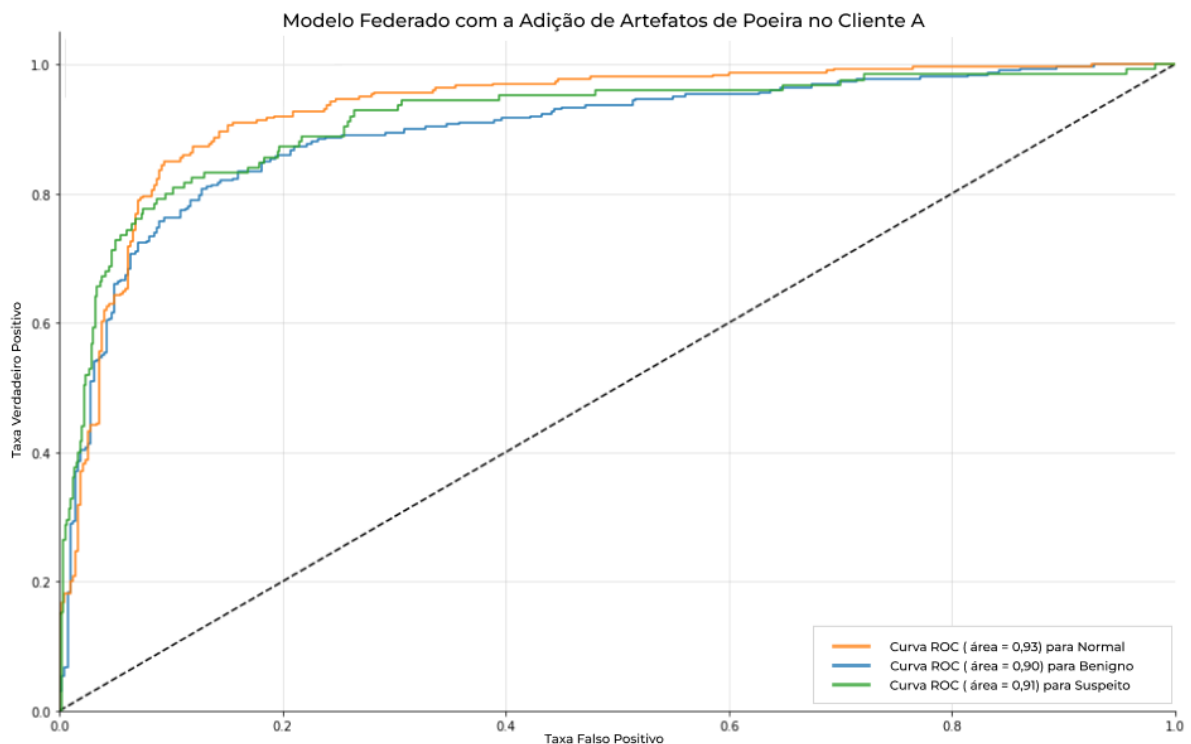
Modelo local A após federação (Acurácia de 64,23%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	56,22%	36,74%	71,27%
Verdadeiro Positivo	18,23%	36,46%	9,53%
Falso Negativo	23,07%	4,97%	7,73%
Falso Positivo	2,49%	21,82%	11,46%

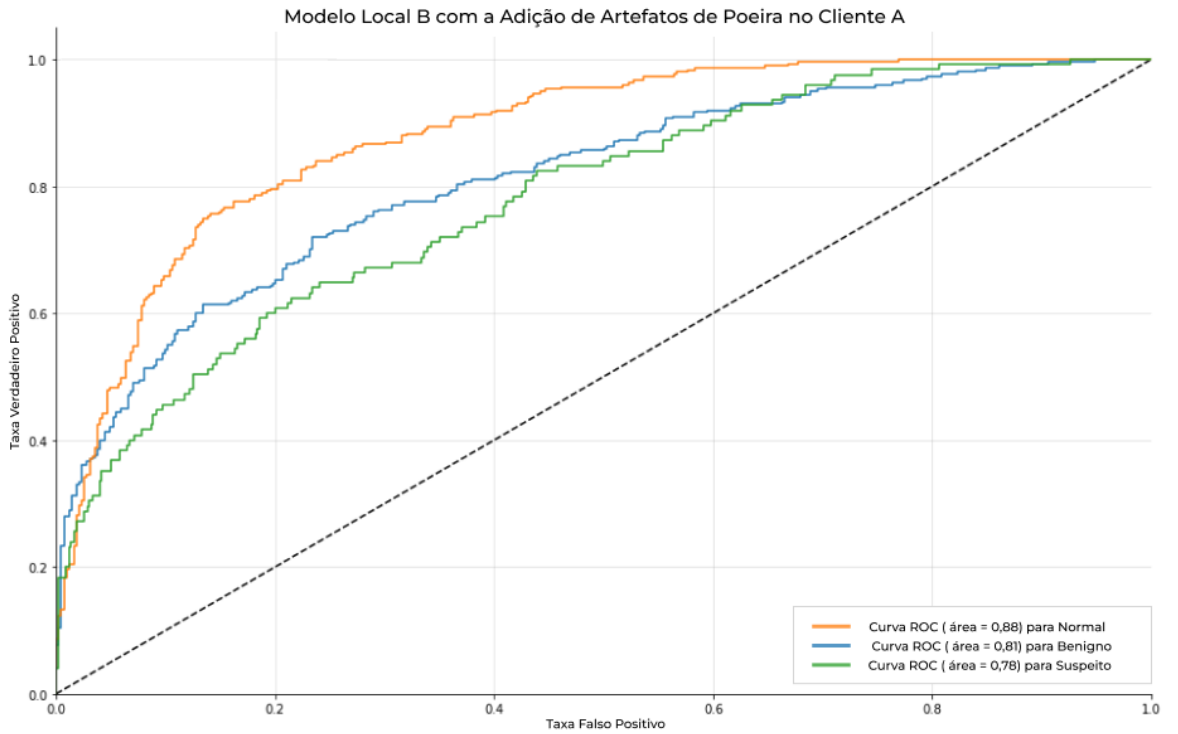
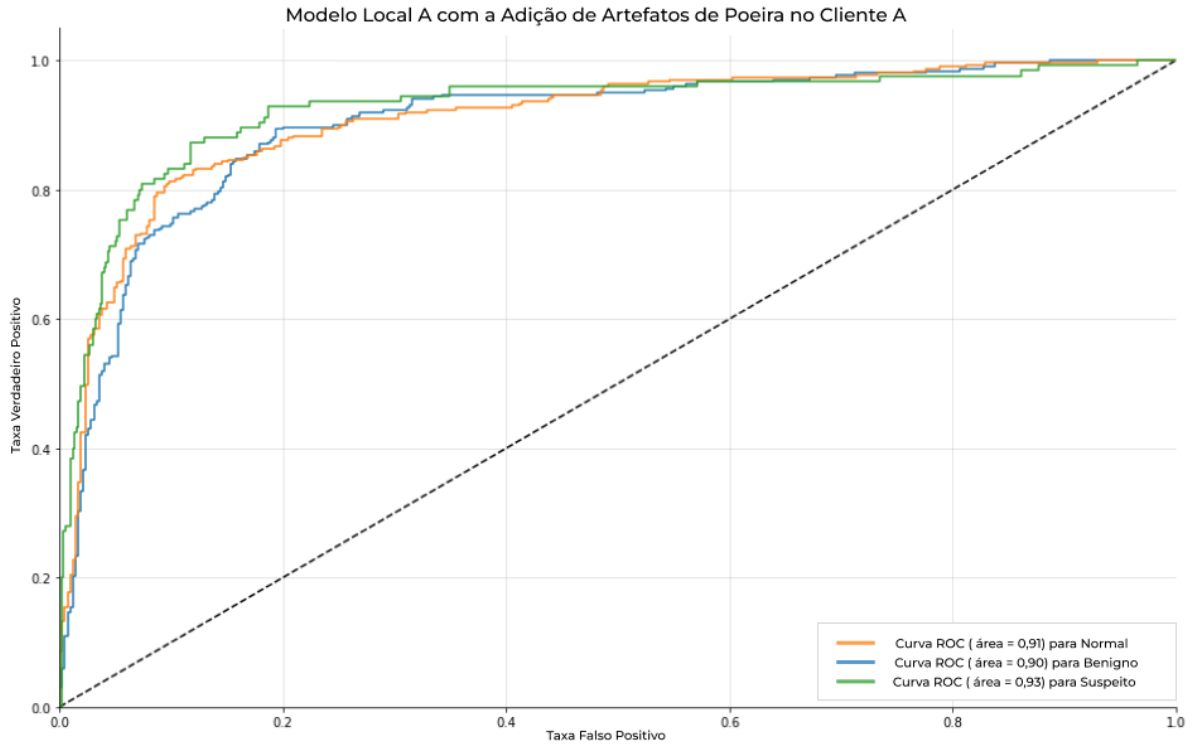
Modelo local B após federação (Acurácia de 80,11%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	55,66%	44,48%	79,97%
Verdadeiro Positivo	31,91%	37,02%	11,19%
Falso Negativo	9,39%	4,42%	6,08%
Falso Positivo	3,04%	14,09%	2,76%

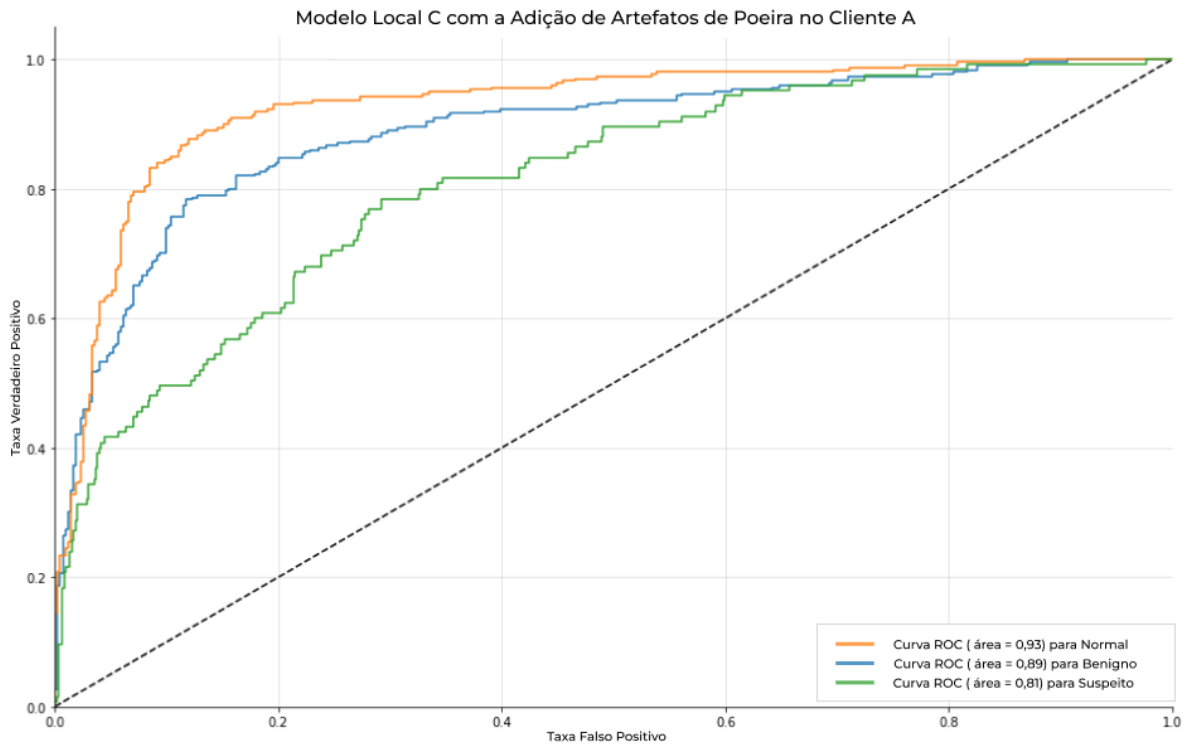
Modelo local C após federação (Acurácia de 76,24%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	53.87%	48.07%	74.31%
Verdadeiro Positivo	30.66%	34.25%	11.33%
Falso Negativo	10.64%	7.18%	5.94%
Falso Positivo	4.83%	10.50%	8.43%

6.5. Treinamento por aprendizado federado com inserção de artefatos (cenário 3).

Neste cenário, procedeu-se a aplicação de artefato de poeira simulando microcalcificações em 10% das imagens do banco de dados do cliente A, notando-se uma queda da acurácia do modelo federado para 73,76%, com queda da AUC em 3% para imagens normais, estabilidade para imagens benignas e aumento de 3% para imagens suspeitas. Seguem abaixo os respectivos gráficos ROC e matrizes de confusão para o modelo federado e seu desempenho local após redistribuição do mesmo:







Modelo Federado (Acurácia de 73,76%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	56,35%	40,47%	76,93%
Verdadeiro Positivo	24,17%	36,88%	12,71%
Falso Negativo	17,13%	4,56%	4,56%
Falso Positivo	2,35%	18,09%	5,80%

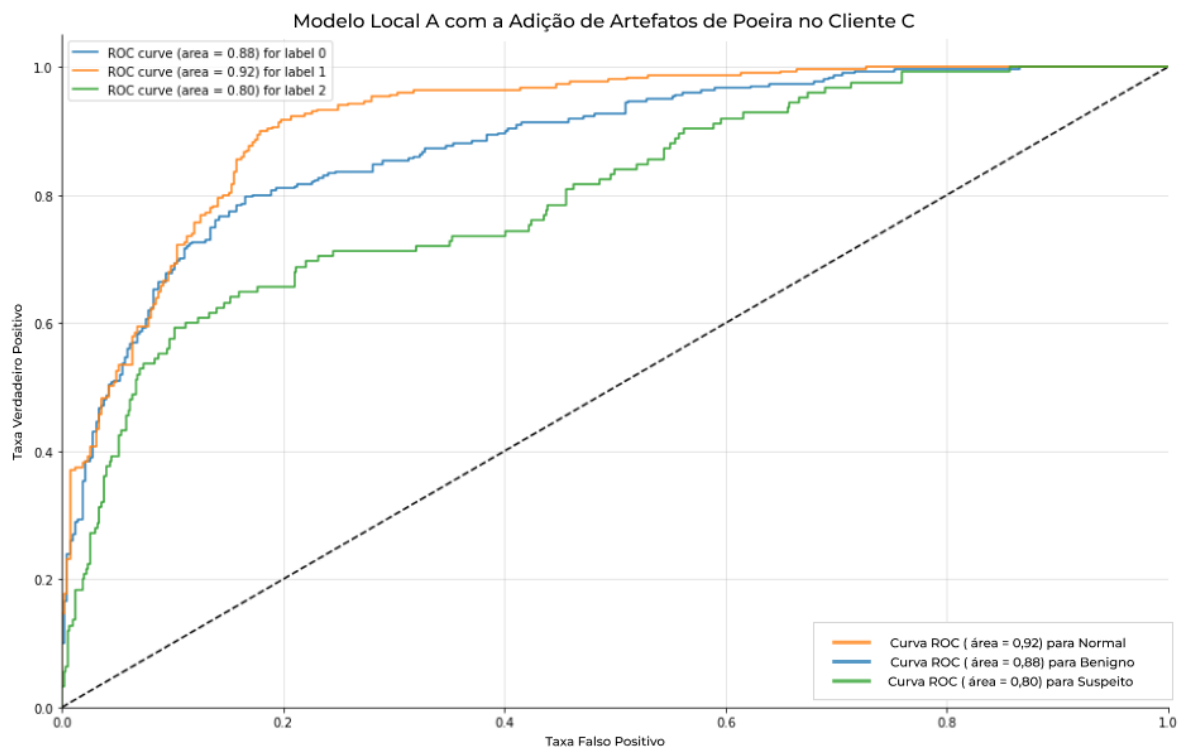
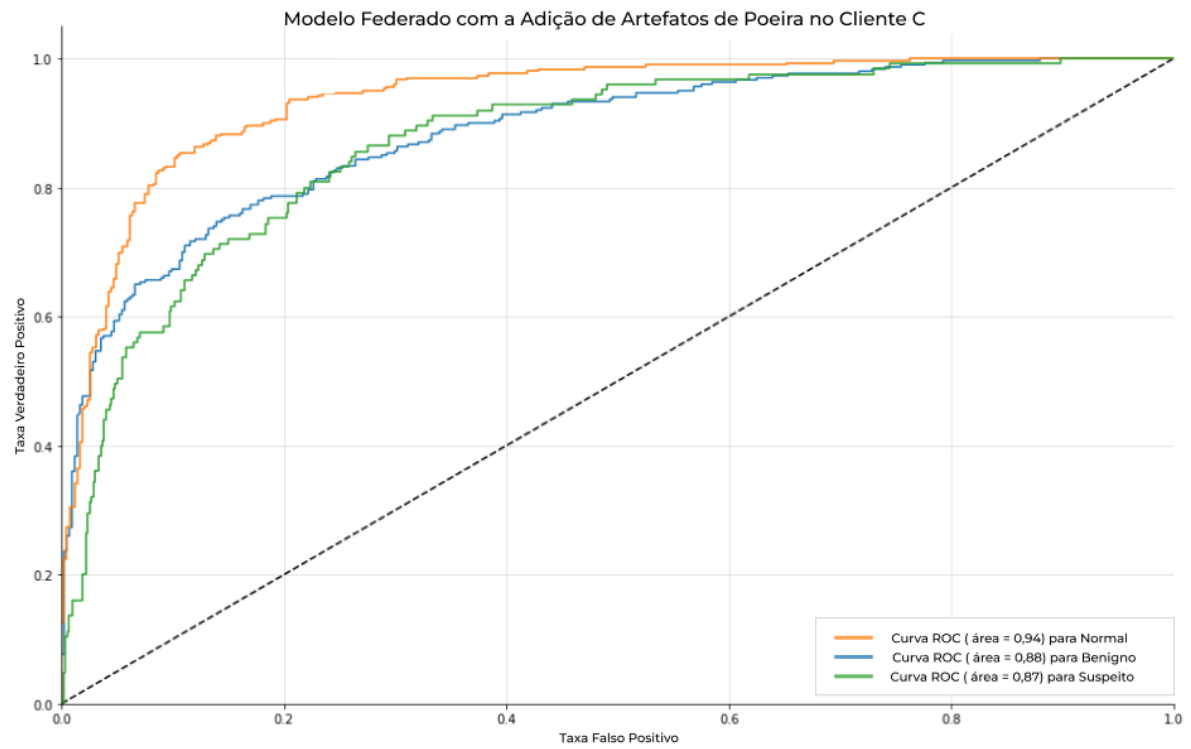
Modelo local A após federação (Acurácia de 79,14%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	54,97%	45,99%	78,18%
Verdadeiro Positivo	29,56%	37,02%	12,57%
Falso Negativo	11,74%	4,42%	4,70%
Falso Positivo	3,73%	12,57%	4,56%

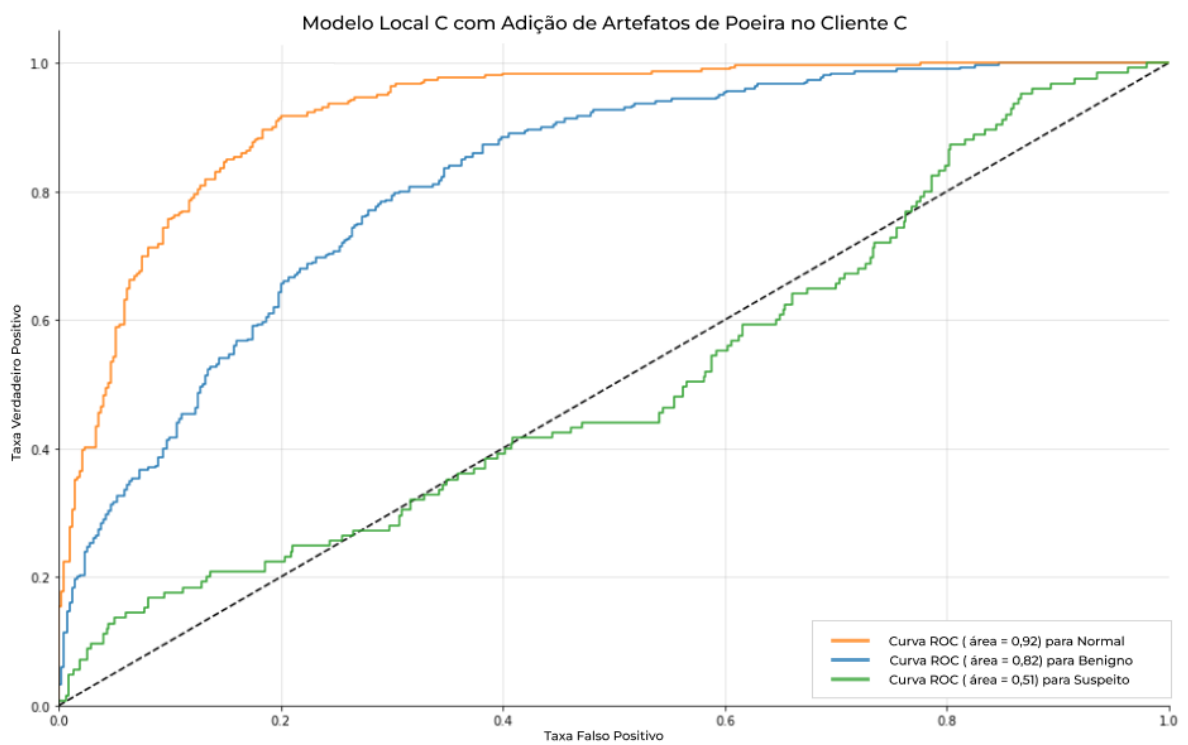
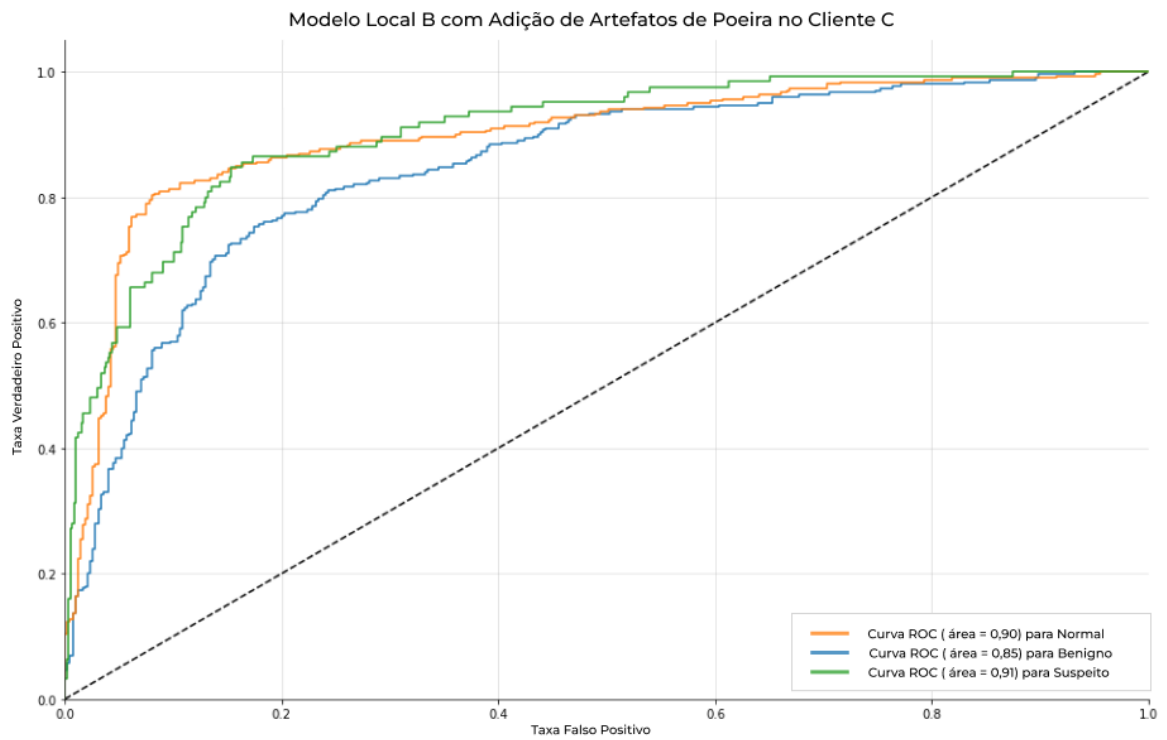
Modelo local B após federação (Acurácia de 59,12%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	56.35%	42.23%	59.53%
Verdadeiro Positivo	17.54%	30.11%	11.46%
Falso Negativo	23.76%	11.33%	5.80%
Falso Positivo	2.35%	15.33%	23.20%

Modelo local C após federação (Acurácia de 64,5%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	57.04%	35.22%	72.24%
Verdadeiro Positivo	18.23%	37.98%	8.29%
Falso Negativo	23.07%	3.45%	8.98%
Falso Positivo	1.66%	23.34%	10.50%

6.6. Treinamento por aprendizado federado com inserção de artefatos (cenário 4).

Neste cenário, procedeu-se a aplicação de artefato de poeira simulando microcalcificações em 10% das imagens do banco de dados do cliente C, ocorrendo uma queda da acurácia do modelo federado para 67,54%, com queda da AUC em 2% para imagens normais, queda em 2% para imagens benignas e queda de 1% para imagens suspeitas. Seguem abaixo os respectivos gráficos ROC e matrizes de confusão para o modelo federado e seu desempenho local após redistribuição do mesmo:





Modelo Federado (Acurácia de 67,54%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	56,22%	30,25%	81,08%
Verdadeiro Positivo	25,28%	38,67%	3,59%
Falso Negativo	16,02%	2,76%	13,67%
Falso Positivo	2,49%	28,31%	1,66%

Modelo local A após federação (Acurácia de 68,92%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	54.83%	38.40%	75.69%
Verdadeiro Positivo	24.17%	35.77%	8.98%
Falso Negativo	17.13%	5.66%	8.29%
Falso Positivo	3.87%	20.17%	7.04%

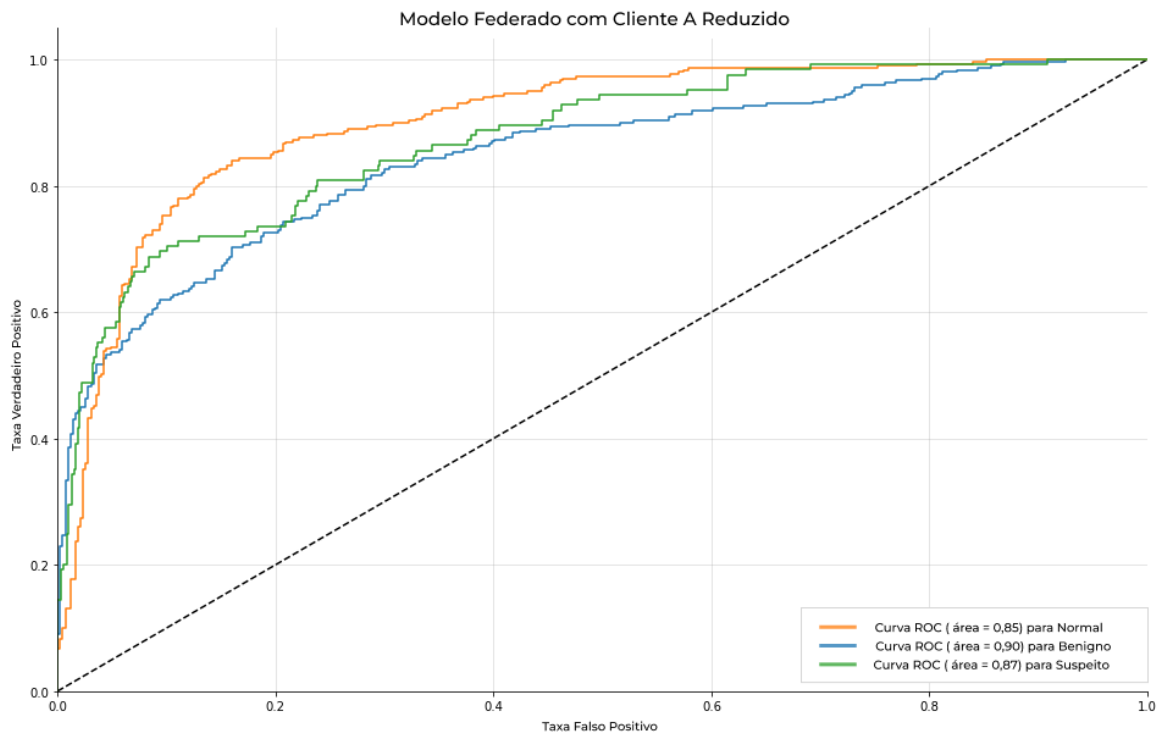
Modelo local B após federação (Acurácia de 65,33%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	56.63%	28.04%	80.66%
Verdadeiro Positivo	19.34%	38.40%	7.60%
Falso Negativo	21.96%	3.04%	9.67%
Falso Positivo	2.07%	30.52%	2.07%

Modelo local C após federação (Acurácia de 61,6%)			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	55.66%	23.76%	82.18%
Verdadeiro Positivo	21.82%	39.23%	0.55%
Falso Negativo	19.48%	2.21%	16.77%
Falso Positivo	3.04%	34.81%	0.55%

6.7. Treinamento por aprendizado federado com exclusão das imagens com artefato do cliente A.

Realizando o treinamento federado entre os clientes A, B e C sob estas condições, obteve-se um modelo federado com acurácia de 65,19%, com a seguinte matriz de confusão:

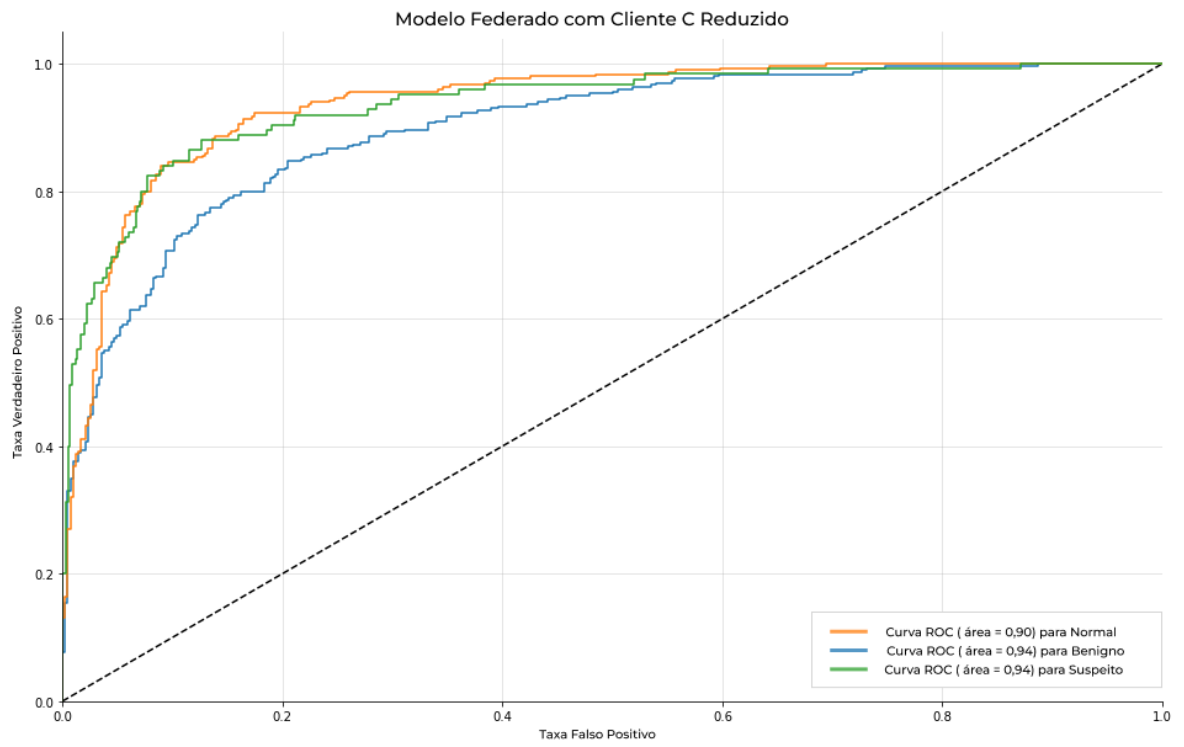
Modelo federado após exclusão de 10% das imagens do cliente A			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	56,91%	33,29%	75%
Verdadeiro Positivo	17,96%	36,19%	11,05%
Falso Negativo	23,34%	5,25%	6,22%
Falso Positivo	1,80%	25,28%	7,73%



6.8. Treinamento por aprendizado federado com exclusão das imagens com artefato do cliente C.

Realizando o treinamento federado entre os clientes A, B e C sob estas condições, obteve-se um modelo federado com acurácia de 76,10%, com a seguinte matriz de confusão:

Modelo federado após exclusão de 10% das imagens do cliente C			
	Normal	Benigno	Suspeito
Verdadeiro Negativo	55,52%	46,69%	73,90%
Verdadeiro Positivo	29,28%	32,32%	14,50%
Falso Negativo	12,02%	9,12%	2,76%
Falso Positivo	3,18%	11,88%	8,84%



7. DISCUSSÃO

Ao comparar-se o desempenho das técnicas sem federação e com federação, percebe-se uma queda da acurácia global dos modelos de 84,9% para 78,45 %, demonstrando que a transmissão do conhecimento dos clientes locais para combinação no servidor pode ser menos eficaz do que o treinamento com dos dados brutos dos bancos. Entretanto, as vantagens da técnica proposta por Brendan MacMahan et al (2017) se mantêm, destacando-se na área médica justamente por possibilitar trabalho conjunto multicêntrico com a preservação de privacidade dos dados e com redução da transmissão de grandes arquivos, ao tempo em que permite que os usuários se beneficiem de um modelo treinado com dados ricos. Por tais características, a técnica despertou grande interesse durante a pandemia da COVID-19, quando se gerou um ambiente com necessidade de esforços colaborativos mundiais para diagnóstico da doença, fomentando a pesquisa de AM com compartilhamento de informação entre múltiplos centros de diversos países (NAZ et al, 2022; SAMUEL et al, 2022; LIANG et al, 2022; ZHANG et al, 2021; DAYAN et al, 2021; BAI et al, 2021).

Frente o aumento do uso da técnica, a confiabilidade dos dados de rotulagem, a qualidade das imagens, a padronização das informações e a confiança no treinamento dos modelos locais tiveram aumento de sua importância, por figurarem informações relevantes e que são desconhecidas pelo servidor central e pelos demais participantes na aplicação do aprendizado federado (WILLEMINK et al, 2020). Dados locais enviados, intencionais ou não, bem como o uso de dados com parâmetros fora do padrão são problemas potenciais, abrindo campo para o estudo de ferramentas para mitigar possíveis danos ao modelo central e buscar identificar falta de equidade na composição das bases de dados (PANDA et al, 2021; KAIROUZ et al, 2021). Protocolos e recomendações dos tratamentos locais destes dados devem buscar uma padronização ampla, fazendo-se necessário entender as possíveis interferências que artefatos em imagens médicas possam causar no treinamento de modelos de AM, permitindo-se traçar estratégias de como lidar com tais imagens.

Neste estudo, o desempenho dos modelos após a adição dos artefatos de movimento demonstrou que a inclusão destas imagens no treinamento pode aumentar a acurácia final, no nosso caso ocorrendo uma melhora mais expressiva justamente quando a interferência ocorreu no cliente com maior número de imagens no banco de dados. Apesar de na prática médica tratar-se de artefato que pode dificultar a interpretação de imagens pelos radiologistas, a técnica de desfoque utilizada para simular o artefato é um dos recursos utilizados para aumento dos

bancos de dados (*data augmentation*) em aplicações de inteligência artificial (CHLAP et al, 2021).

Em um cenário distinto, Zech et al (2018) observaram que a validação externa de modelos de AM podem demonstrar piora de desempenho em função não só de substancial diferença na prevalência dos achados buscados entre o serviço onde o modelo foi treinado e o serviço onde foi testado, mas também pela influência de artefatos de dispositivos médicos, como tubos torácicos e marcadores de imagens. Estes dispositivos estavam presentes nas imagens do serviço onde ocorreu a validação do modelo, entretanto não estavam presentes nas imagens utilizadas para o treinamento, causando uma queda de acurácia do desempenho no serviço de validação. Nesse sentido, a falta destes artefatos pode configurar um viés de seleção na composição do banco de dados por excluir do treinamento a exposição a imagens artefatuais usualmente observadas na prática clínica, corroborando de forma indireta que a presença de artefatos nas imagens de treinamento pode ajudar na melhoria do modelo.

Ao se proceder a adição de artefatos de poeira, ocorreu degradação do modelo com redução de AUC, principalmente quando a interferência se deu no cliente com menor número de imagens, onde a acurácia do modelo federado passou de 78,45% para 67,54%, ocorrendo a maior queda de AUC na categoria de microcalcificações suspeitas, chegando a 0,51. Este resultado pode ter ocorrido devido à diminuição do número absoluto de imagens livres de interferência em um banco de dados local já pequeno, inviabilizando o treinamento adequado do modelo local e levando ao envenenamento do modelo central. Na avaliação do desempenho do treinamento local do cliente C antes da adição de artefatos, a AUC para a categoria de microcalcificações suspeitas era de 0,65. Este valor já era inicialmente baixo, demonstrando a possível inadequação de capacidade de treinamento do modelo por uma quantidade pequena de imagens.

Ao se optar pela redução do banco de dados em cada cliente, onde as imagens com artefatos de poeira simulando microcalcificações em tese seriam retiradas do treinamento do modelo, dois resultados distintos ocorreram. Na intervenção do banco de dados do cliente A, a acurácia global passou de 73,76% (treinamento com artefatos) para 65,19% (treinamento com redução do número de imagens), demonstrando-se que frente a necessidade de uma exclusão de dados em volume maior, a permanência das imagens com artefato nos bancos de treinamento pode ser menos deletéria do que o encolhimento do banco.

Na exclusão das imagens do cliente C, a acurácia global do modelo passou de 67,54% para 76,10%, um pouco abaixo da acurácia antes da adição de artefatos, que era de 78,45%. Em

relação à participação do cliente C no processo de federação, com o menor banco de dados frente os demais e com consequentes desempenhos sempre mais modestos, observa-se a influência negativa do mesmo sobre o modelo global. Com isso, levanta-se a necessidade de cogitar-se a existência de barreiras de entrada de desempenhos locais no processo de federação, como valores mínimos de desempenho do modelo no treinamento local ou a criação de um pacote de teste junto ao servidor central para que possa ser avaliado o desempenho dos modelos antes da agregação. O uso de técnicas e novos algoritmos que ajudem a eliminar ou reduzir a influência negativa sobre o modelo global já está sendo proposto para aplicações práticas (Shi et al, 2022; Varno et al, 2022; Yu et al, 2022).

8. CONCLUSÃO

Os resultados do estudo demonstram que, diante de um banco de dados de clientes locais com dimensões minimamente adequadas para o treinamento do modelo de AM, a presença de artefatos de imagem pode beneficiar o desempenho do modelo. Também se observa que, em determinadas situações, a presença de imagens com artefatos pode levar a uma piora no desempenho do modelo, porém em intensidade menor do que a exclusão das imagens contaminadas do banco de dados para treinamento. Esses achados sugerem que em trabalhos conjuntos entre instituições médicas, a adoção de um protocolo padronizado para manutenção de imagens com artefatos na composição de bancos de dados deve ser incentivada quando se utiliza a técnica de aprendizagem federada. Quanto a modelos que eventualmente levem a um prejuízo do modelo central na federação, o envio de informes sobre os desempenhos locais junto aos metadados deve servir de suporte para medidas de mitigação do prejuízo.

No entanto, este estudo teve limitações por não fazer uso de um modelo de AM desenhado especificamente para imagens médicas, por utilizar artefatos por meio de simulação e por fazer uso de um número original limitado de imagens, uma vez que o objetivo não era a criação de um algoritmo para uso clínico, mas tão somente avaliar o comportamento de um modelo sob influência de imagens com artefatos. Salienta-se que diferentes técnicas de treinamento de máquinas e a utilização de outros tipos de imagens médicas, como aquisições volumétricas de tomografia computadorizada ou imagens multiparamétricas de ressonância magnética, podem gerar resultados distintos quanto ao à manutenção de imagens com artefatos na composição de bancos de dados, havendo necessidade de novos estudos para abordar tais situações.

REFERÊNCIAS

- Ayyala, R. S.; Chorlton, M.; Behrman, R. H.; Kornguth, P. J.; Slanetz, P. J. Digital mammographic artifacts on full-field systems: What are they and how do I fix them? **Radiographics**. 2008;28(7):1999-2008.
- Bai, X. et al. Advancing COVID-19 diagnosis with privacy-preserving collaboration in artificial intelligence. **Nat Mach Intell**. 2021;3(12):1081-1089.
- Barrett, J.; Keat, N. Artifacts in CT: Recognition and Avoidance. **Radiographics**. 2004;24(6):1679-1691+1796.
- Brendan McMahan, H. et al. Communication-efficient learning of deep networks from decentralized data. **Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017**.
- Chartrand, G. et al. Deep Learning: A Primer for Radiologists. **Radiographics**. 2017;37(7): 2113-2131.
- Cheng, P. M. et al. Deep learning: An update for radiologists. **Radiographics**. 2021;41(5):1427-1445.
- Chlap, P. et al. A review of medical image data augmentation techniques for deep learning applications. **J Med Imaging Radiat Oncol**. 2021 Aug;65(5):545-563.
- Darzidehkalani, E.; Ghasemi-Rad, M.; van Ooijen, P. M. A. Federated Learning in Medical Imaging: Part II: Methods, Challenges, and Considerations. **J Am Coll Radiol**. 2022 Aug;19(8):975-982.
- Dayan, I. et al. Federated learning for predicting clinical outcomes in patients with COVID-19. **Nat Med**. 2021;27(10):1735-1743.
- Erickson, B. J. Et al. Machine Learning for Medical Imaging. **Radiographics**. 2017 Mar-Apr;37(2):505-515.
- Kairouz, P. et al. Advances and open problems in federated learning. **Found Trends Mach Learn**. 2021;14(1-2):1-210.

Liang, H. et al. Artificial intelligence for stepwise diagnosis and monitoring of COVID-19. **Eur Radiol**. 2022;32(4):2235-2245.

Loizidou, K.; Skouroumouni, G.; Pitris, C.; Nikolaou, C. Digital subtraction of temporally sequential mammograms for improved detection and classification of microcalcifications. **Eur Radiol Exp**. 2021;5(1).

Lundervold, A. S.; Lundervold A. An overview of deep learning in medical imaging focusing on MRI. **Z Med Phys**. 2019;29(2):102-127.

Mendes, R.; Vilela, J. Privacy-Preserving Data Mining: Methods, Metrics and Applications. **IEEE Access**. 2017;vol 5, pp 10562-10582.

Naz, S.; Phan, K. T.; Chen, Y-P. A comprehensive review of federated learning for COVID-19 detection. **Int J Intell Syst**. 2022;37(3):2371-2392.

Oh, W.; Nadkarni, G. N. Federated Learning in Health care Using Structured Medical Data. **Adv Kidney Dis Health**. 2023 Jan;30(1):4-16.

Panayides, A. S. et al. Radiogenomics for Precision Medicine With a Big Data Analytics Perspective. **IEEE J Biomed Health Inform**. 2019;23(5):2063-2079.

Panayides, A.S. et al. AI in Medical Imaging Informatics: Current Challenges and Future Directions. **IEEE J Biomed Health Inform**. 2020 Jul;24(7):1837-1857.

Panda, A. et al. SparseFed: Mitigating Model Poisoning Attacks in Federated Learning with Sparsification. **ArXiv**. December 2021. Disponível em: <https://arxiv.org/abs/2112.06274>. Acesso em: 07 de jun. 2022.

Rieke, N. et al. The future of digital health with federated learning. **npj Digit Med**. 2020;3(1).

Roth, H. R. et al. Federated Learning for Breast Density Classification: A Real-World Implementation. **Lect Notes Comput Sci**. 2020;12444 LNCS:181-191.

Sahi, A.; Lai, D.; Li, Y. Security and privacy preserving approaches in the eHealth clouds with disaster recovery plan. **Comput Biol Med**. 2016;78:1-8.

Samuel, O. et al. IoMT: A COVID-19 Healthcare System driven by Federated Learning and Blockchain. **IEEE J Biomed Health Inform.** 2022.

Santos, M. K. et al. Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine. **Radiol Bras.** 2019;52(6):387-396

Shah, A.; Gulati, R. Privacy Preserving Data Mining: Techniques, Classification and Implications - A Survey. **International Journal of Computer Applications.** 2016; 137:40-46.

Shi, Y. et al. Optimization Strategies for Client Drift in Federated Learning: A review. **International Journal of Computer Applications.** 2022; 214:1168-1173.

Sickles, E. A.; D'Orsi, C. J.; Bassett, L. W. et al. ACR BI-RADS® Mammography. In: **ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System.** Reston, VA, American College of Radiology; 2013.

Tan, M.; Le, Q.V. Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. **ArXiv.** 2019. Disponível em: <https://arxiv.org/abs/1905.11946>. Acesso em: 2 de jun de 2022.

Tolpegin, V.; Truex, S.; Gursoy, M. E.; Liu, L. Data poisoning attacks against federated learning systems. **Lect Notes Comput Sci.** 2020;12308 LNCS:480-501.

van Panhuis, W. G. et al. A systematic review of barriers to data sharing in public health. **BMC Public Health.** 2014;14(1).

Varno, F. et al. AdaBest: Minimizing Client Drift in Federated Learning via Adaptive Bias Estimation. **ArXiv.** 2022. Disponível em: <https://arxiv.org/abs/2204.13170>. Acesso em: 3 de abril de 2023.

Walz-Flannigan, A. I.; Brossoit, K. J.; Magnuson, D. J.; Schueler, B. A. Pictorial review of digital radiography artifacts. **Radiographics.** 2018;38(3):833-846.

Wen, J. et al. A survey on federated learning: challenges and applications. **Int J Mach Learn Cybern.** 2023;14(2):513-535..

Willeminck, M. J. et al. Preparing Medical Imaging Data for Machine Learning. **Radiology**. 2020 Apr;295(1):4-15.

Yu, A. C.; Eng, J. One Algorithm May Not Fit All: How Selection Bias Affects Machine Learning Performance. **Radiographics**. 2020;40(7):1932-1937.

Yu X, Li L, He X, Chen S, Jiang L. Federated Learning Optimization Algorithm for Automatic Weight Optimal. **Comput Intell Neurosci**. 2022;Nov(9):8342638.

Zech, J. R. et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. **PLoS Med**. 2018 Nov 6;15(11):e1002683.

Zhang, K. et al. Challenges and future directions of secure federated learning: a survey. **Front Comput Sci**. 2022;16(5):165817.

Zhang, W. et al. Dynamic-Fusion-Based Federated Learning for COVID-19 Detection. **IEEE Internet Things J**. 2021;8(21):15884-15891.