FUCAPE PESQUISA E ENSINO S/A – FUCAPE ES

LUIZ FELIPE VILELA PINTO

IA CENTRADA NO SER HUMANO: Uma abordagem integrada em busca de um melhor modelo de detecção de fraudes na saúde.

VITÓRIA

2025

LUIZ FELIPE VILELA PINTO

IA CENTRADA NO SER HUMANO: Uma abordagem integrada em busca de um melhor modelo de detecção de fraudes na Saúde.

Tese apresentada ao Programa de Pós-Graduação em Ciências Contábeis e Administração, da Fucape Pesquisa e Ensino S/A, como requisito parcial para obtenção do título de Doutor em Ciências Contábeis e Administração – Nível Profissionalizante.

Orientador: Prof° Dr.: Danilo Soares Monte Mor

VITÓRIA

2025

LUIZ FELIPE VILELA PINTO

IA CENTRADA NO SER HUMANO: Uma abordagem integrada em busca de um melhor modelo de detecção de fraudes na Saúde.

Tese apresentada ao Programa de Pós-Graduação em Ciências Contábeis e Administração da Fucape Pesquisa e Ensino S/A, como requisito parcial para obtenção do título de Doutor em Ciências Contábeis e Administração.

Aprovada em 31 de outubro de 2025.

BANCA EXAMINADORA

Prof° Dr.: Danilo Soares Monte Mor
Fucape Pesquisa e Ensino S/A

Prof° Dr.: José Mario Bispo Sant'anna
Fucape Pesquisa e Ensino S/A

Prof° Dr.: Poliano Bastos da Cruz
Fucape Pesquisa e Ensino S/A

Prof° Dr.: Everlan Elias Montibeler
UFES Universidade Federal do Espírito Santo

Prof° Dr.: Octavio Locatelli
Fucape Pesquisa e Ensino S/A

AGRADECIMENTOS

O caminho do conhecimento é uma jornada que vai muito além das conquistas individuais. É construído por pequenos passos que, somados, contribuem para a formação de uma sociedade mais justa e com oportunidades ampliadas. Quando esses avanços são aplicados na prática, eles transformam realidades, promovendo dignidade e gerando melhorias concretas na vida das pessoas.

Concluir esta tese de doutorado representa uma trajetória marcada por desafios, aprendizados e superações. Cada passo só foi possível graças à presença constante de Deus, que me concedeu força, sabedoria e resiliência para continuar, mesmo diante das adversidades. A Ele dedico minha mais profunda gratidão.

Agradeço também ao meu orientador, Prof. Dr. Danilo Soares Monte Mor, por sua orientação firme, pelas valiosas contribuições acadêmicas e pelo incentivo constante, fundamentais para o desenvolvimento deste trabalho.

Sou igualmente grato à minha família, aos amigos, aos professores e a todos que, de alguma forma, caminharam ao meu lado. O apoio recebido, os gestos de incentivo e as lições transmitidas tiveram papel fundamental na conclusão desta jornada, etapa por etapa.

A todos que contribuíram para tornar este sonho realidade, registro minha sincera e profunda gratidão.

RESUMO

A relevância do setor de saúde, que movimenta grandes volumes financeiros e impacta diretamente a vida de milhões de pessoas, torna-o alvo constante de fraudes, gerando riscos significativos à sustentabilidade e à qualidade do atendimento. Diante desse cenário, o conjunto dos trabalhos desenvolvidos nesta tese buscou contribuir para soluções mais eficazes de prevenção e detecção de fraudes, com foco específico em cooperativas médicas, estrutura que representa parcela expressiva da saúde suplementar no Brasil e que, por sua forma de organização e governança, apresenta desafios particulares para controle de eventos de fraudes. O primeiro estudo científico validou que a abordagem de IA supervisionada, integrada ao conhecimento humano, apresenta maior assertividade na detecção de eventos suspeitos de fraude em comparação a métodos puramente autônomos, destacando a importância do fator humano para aumentar a precisão e reduzir falsos alarmes. O artigo tecnológico complementou essa análise, mapeando as áreas mais sensíveis dentro das cooperativas médicas e identificando os tipos de fraudes mais recorrentes, oferecendo uma visão aprofundada sobre vulnerabilidades operacionais e comportamentais. A partir dessa radiografia, delinearam-se diretrizes para o fortalecimento de ações de controle e mitigação de riscos. Por fim, o produto tecnológico consolidou essas descobertas no desenvolvimento da plataforma antifraude denominada GloomTrace®, estruturada para integrar aprendizado de máquina com a experiência de especialistas. De forma integrada, os resultados desta tese reforçam que o caminho mais seguro para combater fraudes em cooperativas médicas combina tecnologia de ponta, IA supervisionada e participação ativa de profissionais, ampliando a confiabilidade dos processos e promovendo uma gestão mais transparente, ágil e eficiente.

Palavras-chave: Inteligência Artificial; Detecção de Fraudes; Saúde; Cooperativas Médicas.

ABSTRACT

The relevance of the healthcare sector, which handles large financial volumes and directly impacts the lives of millions of people, makes it a constant target for fraud, generating significant risks to sustainability and the quality of care. In this context, the set of studies developed in this dissertation aims to contribute to more effective solutions for fraud prevention and detection, with a specific focus on medical cooperatives, a structure that represents a significant share of the supplementary healthcare market in Brazil and, due to its organizational and governance model, presents particular challenges for fraud control. The first scientific study validated that a supervised AI approach, integrated with human expertise, achieves greater accuracy in detecting suspected fraud events compared to purely autonomous methods, highlighting the importance of the human factor in increasing precision and reducing false positives. The technological paper complemented this analysis by mapping the most sensitive areas within medical cooperatives and identifying the most recurrent types of fraud, providing an in-depth view of operational and behavioral vulnerabilities. Based on this overview, guidelines were outlined to strengthen control measures and mitigate risks. Finally, the technological product consolidated these findings in the development of the antifraud platform named GloomTrace®, designed to integrate machine learning algorithms with the expertise of specialists. Taken together, the results of this dissertation reinforce that the safest way to combat fraud in medical cooperatives combines advanced technology, supervised AI, and the active participation of professionals, increasing process reliability and promoting more transparent, agile, and efficient management.

Keywords: Artificial Intelligence; Fraud Detection; Healthcare; Medical Cooperatives.

SUMÁRIO

INTRODUÇÃO GERAL

O setor de saúde ocupa posição central na economia, movimentando recursos expressivos e, por isso, tornando-se especialmente vulnerável a práticas fraudulentas (OECD, 2024; IBGE, 2024; NHCAA, n.d.; IESS, 2017). Esse contexto reforça a necessidade urgente de desenvolver métodos mais eficientes para prevenção e detecção de irregularidades.

Nos últimos anos, pesquisadores de diferentes países têm explorado o uso de tecnologias emergentes, como algoritmos de aprendizado de máquina, como estratégia para mitigar fraudes (Shahana et al., 2023; Askary et al., 2018; Fedyk et al., 2022). No entanto, estudos também evidenciam limitações em modelos puramente autônomos, seja pela complexidade dos dados, seja pela resistência de profissionais em adotar soluções que não oferecem transparência ou participação humana (Commerford et al., 2022; Nakao et al., 2023).

Nesse contexto, surge o conceito de Inteligência Artificial Centrada no Ser Humano (Human-Centered AI - HCAI), que propõe combinar a capacidade de processamento dos algoritmos com a experiência de especialistas, tornando os sistemas mais precisos, confiáveis e alinhados a princípios de governança responsável (Shneiderman, 2020; Xu et al., 2023; Jiang et al., 2023).

Apesar do avanço conceitual em torno do uso de algoritmos de aprendizado de máquina para auditoria e controle de fraudes, estudos apontam que a aplicação prática de modelos supervisionados de IA integrados à expertise humana ainda é incipiente em contextos reais, principalmente em organizações de saúde suplementar que operam com estruturas cooperativas (Massi et al., 2020; Sun et al., 2018;

Shahana et al., 2023). Pesquisas como a de Massi et al. (2020) demonstram que métodos autônomos são dominantes na literatura, mas carecem de contextualização operacional e ajustes por especialistas, fator crucial para reduzir falsos positivos e alinhar os resultados aos protocolos clínicos reais.

Além disso, cooperativas médicas, que representam uma fatia relevante da saúde suplementar brasileira (IESS, 2017), operam sob o princípio da autogestão e da corresponsabilidade entre os cooperados, o que cria particularidades na governança, na estrutura de auditoria e nos fluxos de decisão que diferem de modelos empresariais convencionais. Essas especificidades aumentam a complexidade na aplicação de sistemas automatizados e justificam a necessidade de metodologias híbridas, que combinem o potencial de detecção algorítmica com ajustes e validações feitas por profissionais humanos (Nakao et al., 2023; Xu et al., 2024).

Apesar de alguns avanços, o uso prático de plataformas inteligentes voltadas especificamente para este segmento ainda é restrito. Muitos estudos permanecem no nível exploratório ou experimental, sem evoluir para soluções escaláveis que contemplem as variáveis regulatórias, de confidencialidade e de adaptação aos fluxos internos de cooperativas (Commerford et al., 2022; Jiang et al., 2023).

Diante desse cenário, esta tese está estruturada em três capítulos complementares e integrados: (i) um artigo científico que investiga, com base em dados reais, a eficácia de uma abordagem de Inteligência Artificial centrada no ser humano para detecção de eventos suspeitos de fraude; (ii) um artigo tecnológico que mapeia vulnerabilidades e propõe uma arquitetura antifraude para cooperativas médicas; e (iii) o memorial descritivo do Produto Técnico Tecnológico (PTT), que apresenta a concepção, a arquitetura e as principais funcionalidades da plataforma

digital GloomTrace, evidenciando seu design, módulos operacionais e aplicação prática. As três peças se articulam de forma a fornecer fundamentação teórica, validação prática e apresentação estruturada do produto final, abrangendo sua concepção, arquitetura e principais funcionalidades.

Esta pesquisa se justifica teoricamente ao contribuir com a literatura recente sobre aplicações práticas de IA supervisionada no combate a fraudes em saúde, além de avançar no debate sobre abordagens centradas no ser humano (Shneiderman, 2020; Nakao et al., 2023; Xu et al., 2024). Do ponto de vista prático, oferece subsídios para gestores aprimorarem seus controles, reduzirem perdas financeiras e assegurarem a sustentabilidade das cooperativas, alinhando-se, ainda, a diretrizes de governança corporativa, responsabilidade ética e uso transparente de dados, em conformidade com normas como a LGPD.

Portanto, esta tese visa não apenas preencher lacunas acadêmicas, mas também contribuir para a melhoria efetiva da gestão antifraude, articulando três trabalhos complementares, a validação comparativa de métodos, o estudo exploratório e o desenvolvimento de um produto tecnológico, em um ciclo coerente que gera valor para o sistema de saúde suplementar e para a sociedade como um todo.

ARTIGO CIENTÍFICO: IA CENTRADA NO SER HUMANO: UMA ABORDAGEM INTEGRADA EM BUSCA DE UM MELHOR MODELO DE DETECÇÃO DE FRAUDES EM COOPERATIVAS MÉDICAS.

RESUMO

A relevância do setor de saúde, que envolve grande parte da população e volumes financeiros expressivos, chama a atenção de indivíduos e organizações que buscam fraudar o sistema, gerando prejuízos e riscos para sua sustentabilidade. Nos últimos anos, houve um aumento significativo nos estudos sobre o tema, com foco no desenvolvimento de tecnologias capazes de mitigar esses eventos. A inteligência artificial (IA) tem se mostrado uma importante aliada na detecção de padrões atípicos relacionados a possíveis irregularidades. Nesse contexto, o presente estudo tem como objetivo avaliar se uma abordagem de IA integrada e centrada no ser humano pode melhorar a precisão na identificação de eventos suspeitos de fraude, em comparação aos modelos autônomos de aprendizado de máquina. O estudo contribui para o avanço do conhecimento aplicado ao comparar dois métodos voltados à identificação desses eventos: um baseado exclusivamente em um modelo autônomo de aprendizado de máquina, e outro que integra o conhecimento de especialistas humanos ao processo computacional. Além disso, os resultados da pesquisa podem orientar gestores na construção de plataformas antifraude, contribuindo diretamente para a redução dos riscos na gestão, otimização de recursos e melhor qualidade no atendimento ao paciente. Esses resultados também serviram de subsídio para o artigo tecnológico e para o desenvolvimento do Produto Técnico Tecnológico descrito nesta tese.

Palavras-chave: Fraude; Saúde Suplementar; Cooperativa Médica; Inteligência Artificial.

1 INTRODUÇÃO

O setor de saúde ocupa um lugar de destaque na economia, envolvendo uma parcela expressiva do PIB mundial. Países membros da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) destinaram, em média, 9,7% do PIB para despesas com saúde em 2021, com destaque para os Estados Unidos, que alocaram 18,6%, totalizando aproximadamente US$ 4 trilhões (OECD.Stat, base de dados da OCDE; Instituto Brasileiro de Geografia e Estatística (IBGE), 2024). No Brasil, esse percentual foi o mesmo (9,7%), com a maior parte das despesas sendo custeada por famílias e instituições filantrópicas (5,7%) e o restante pelo governo (4%). Ainda assim, o gasto per capita brasileiro representa cerca de um terço da média dos países da OCDE, indicando diferenças substanciais de capacidade de investimento no setor (IBGE, 2024).

Esses volumes de recursos tornam o setor suscetível a ações fraudulentas. Nos Estados Unidos, estimativas da National Health Care Anti-Fraud Association (NHCAA) apontam que entre 3% e 10% dos gastos com saúde podem ser desviados por meio de fraudes (National Health Care Anti-Fraud Association, n.d.). No Brasil, segundo levantamento do Instituto de Estudos de Saúde Suplementar (IESS, 2017), cerca de 19,1% das despesas da saúde suplementar foram atribuídas a fraudes ou desperdícios. Tais números justificam a intensificação de esforços voltados à prevenção e detecção de irregularidades.

Estudos têm destacado o potencial da inteligência artificial (IA) como ferramenta para mitigar práticas fraudulentas, por meio da detecção de padrões atípicos em grandes volumes de dados clínicos e administrativos (Askary et al., 2018; Rahahleh et al., 2021; Fedyk et al., 2022). Modelos de aprendizado de máquina,

especialmente os não supervisionados, têm sido aplicados em estudos que buscam agrupar atendimentos similares e identificar casos que destoam desses padrões (Massi et al., 2020; Matloob et al., 2020; Settipalli et al., 2022).

Por outro lado, diversos autores chamam atenção para as limitações da adoção irrestrita de modelos autônomos. A ausência de transparência, a dificuldade de interpretação dos modelos e a baixa sensibilidade ao contexto específico podem comprometer a aceitação dos resultados por profissionais de saúde e auditores (Commerford et al., 2022; Nakao et al., 2023; Shneiderman, 2020).

Diante disso, cresce o interesse por abordagens que combinem capacidades técnicas da IA com o conhecimento especializado dos profissionais, conceito conhecido como IA centrada no ser humano (HCAI - Human-Centered Artificial Intelligence). Essa abordagem busca desenvolver sistemas que mantenham o ser humano no centro das decisões, permitindo maior controle, adaptação ao contexto e potencial para aceitação mais ampla (Shneiderman, 2020; Xu et al., 2024).

Embora o tema da HCAI tenha avançado em diversos campos, há escassez de aplicações práticas documentadas na área da saúde suplementar, especialmente na detecção de anomalias associadas ao pagamento de prestadores de serviços de saúde. Essa lacuna fundamenta o presente estudo, que tem como objetivo comparar dois métodos para detecção de fraudes: o primeiro utiliza exclusivamente o algoritmo DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), proposto por Ester et al. (1996), voltado à identificação de agrupamentos com base na densidade dos dados; o segundo combina a análise de especialistas clínicos à calibragem de um modelo supervisionado, implementado por meio do algoritmo Random Forest, desenvolvido por Breiman (2001).

As guias identificadas como suspeitas por ambos os métodos serão avaliadas por critérios definidos por especialistas, que não terão conhecimento prévio sobre qual abordagem gerou os alertas. O objetivo é comparar o desempenho dos métodos na identificação de eventos suspeitos de fraude, fornecendo evidências empíricas sobre a efetividade da integração entre inteligência artificial (IA) e conhecimento de especialistas humanos.

Importante destacar que este estudo não tem como objetivo a apuração efetiva das fraudes, mas sim a identificação de eventos suspeitos de fraude com base em padrões incomuns de comportamento assistencial e financeiro.

Diante desse cenário, a questão que orienta este estudo é:

Uma abordagem de inteligência artificial centrada no ser humano pode melhorar a precisão na detecção de eventos suspeitos de fraude em comparação aos modelos autônomos?

A hipótese central deste estudo é que a abordagem integrada e centrada no ser humano apresenta melhor desempenho na geração de alertas relacionados a eventos suspeitos de fraude, em comparação aos modelos autônomos. Os resultados obtidos têm potencial para orientar a implementação de sistemas antifraude mais sensíveis às particularidades institucionais, contribuindo para o aperfeiçoamento da gestão e para o uso racional dos recursos no setor de saúde suplementar.

Além disso, este estudo contribui para o avanço do conhecimento aplicado sobre o uso da inteligência artificial no setor de saúde suplementar, ao comparar metodologicamente abordagens autônomas e integradas. Os resultados podem servir como base para o desenvolvimento de sistemas antifraude mais sensíveis ao contexto

institucional, oferecendo suporte a gestores na adoção de estratégias mais eficazes de controle, otimização de recursos e qualificação dos processos de controles internos e regulação.

2 REFERENCIAL TEÓRICO

2.1 A IA COMO ALIADA NO COMBATE À FRAUDE

Nos últimos anos, houve um aumento significativo nos estudos sobre o tema fraude. O interesse pelo tema tem crescido globalmente, com destaque para os Estados Unidos, que concentram o maior volume de publicações e citações. Países como China, Taiwan e Grécia também fizeram contribuições significativas sobre o tema (Shahana et al., 2023). Vários estudos destacam a importância do uso da inteligência artificial na identificação de eventos suspeitos e, consequentemente, na mitigação de riscos, tornando o uso de algoritmos uma ferramenta relevante para os gestores da saúde (Sun et al., 2018; Matloob et al., 2020; Settipalli et al., 2022; Massi et al., 2020).

Segundo Sun et al. (2018), padrões suspeitos podem ser identificados por meio da análise de divergência entre clusters de pacientes. Os autores realizaram um estudo com dados de uma seguradora de saúde da China, envolvendo mais de 40 milhões de registros de internações de 10 mil pacientes durante 5 anos. O método utilizado calculava, por meio de algoritmos, a similaridade entre internações hospitalares, criando agrupamentos com comportamentos semelhantes. Após a coleta dos dados e criação dos agrupamentos, médicos especialistas participaram do processo ajustando imperfeições nos clusters, como a padronização de nomes de diagnósticos, medicamentos e tratamentos. O resultado dessas correções gerou uma base consolidada com cerca de 68.000 internações, 700.000 prescrições médicas e 3.200 pacientes. A partir disso, foi possível extrair o significado semântico típico de cada cluster e detectar divergências camufladas que indicavam comportamentos

atípicos. Segundo os autores, o método alcançou melhora superior a 15% na precisão em relação às abordagens tradicionais utilizadas como parâmetro.

Matloob et al. (2020) propuseram uma metodologia baseada na detecção de sequências incomuns de serviços prestados ou prescrições realizadas para os pacientes. Utilizando o algoritmo Prefixspan (para mineração de sequência frequente), uma árvore de predição compacta e o Teorema de Bayes, os autores analisaram dados de um hospital privado na cidade de Islamabad, no Paquistão. A metodologia parte da premissa de que fraudadores procuram se camuflar no comportamento dos pacientes regulares, dificultando a detecção por abordagens convencionais. As regras iniciais de sequências frequentes e raras por especialidade foram geradas com base em séries temporais extraídas de registros transacionais de atendimentos ao longo de cinco anos. Com essas sequências, os algoritmos puderam identificar comportamentos que não correspondiam a padrões típicos da especialidade. Várias reuniões com especialistas da área médica foram conduzidas para validar e ajustar o modelo proposto. O sistema atingiu uma precisão média de 85% na identificação de eventos suspeitos.

Settipalli et al. (2022) reforçam a dificuldade de validar manualmente grandes volumes de sinistros utilizando auditorias convencionais, considerando a necessidade de envolver especialistas médicos na análise. Como alternativa, os autores propõem um modelo baseado na decomposição de séries temporais em três componentes: tendência, sazonalidade e irregularidade. O objetivo é separar os eventos que seguem padrões dos que representam desvios, chamados de desvios conceituais ou inesperados, que podem estar associados a práticas fraudulentas. Eles criticam abordagens tradicionais baseadas apenas em regras fixas ou clustering com

distâncias e bandas estáticas, argumentando que esses métodos não consideram variações temporais nem realizam análise prévia dos desvios. O modelo proposto busca ser adaptativo às mudanças nos dados de sinistros e reduzir alarmes falsos. Após a eliminação dos padrões sazonais e de tendência, os resíduos são analisados por meio de agrupamento topológico, o que permite destacar variações irregulares que se distanciam dos comportamentos esperados. A etapa seguinte será o desenvolvimento de um mecanismo de similaridade adaptativo para diferenciar ruído e desvio real, com base em aprendizado de máquina semi-supervisionado.

Massi et al. (2020) propuseram um modelo não supervisionado voltado à detecção de práticas de upcoding associadas ao sistema de Grupos de Diagnósticos Relacionados (DRG). O estudo foi conduzido na região da Lombardia, Itália, entre 2013 e 2015, com dados de cerca de 400 mil internações hospitalares, 130 mil pacientes e 183 hospitais, com foco em casos de insuficiência cardíaca. O algoritmo K-means foi aplicado ao banco de registros de alta hospitalar para agrupar os hospitais por semelhança de comportamento. Hospitais localizados nas extremidades dos clusters foram considerados discrepantes. Esses casos foram destacados visualmente para que auditores humanos realizassem uma análise mais profunda. A proposta metodológica dos autores visa oferecer uma ferramenta de apoio à decisão, sem exigir classificação prévia dos hospitais. O modelo se alinha à ideia de análise exploratória baseada em padrões emergentes de dados.

Embora esses estudos representem avanços significativos na identificação de anomalias em dados de saúde, observa-se que grande parte dos modelos revisados não contempla a participação ativa de especialistas humanos no processo analítico. Essa lacuna reforça a relevância da proposta deste estudo, que busca integrar o

conhecimento de especialistas ao uso de modelos computacionais na identificação de eventos suspeitos no sistema de saúde suplementar.

2.2 PROBLEMAS E DESAFIOS DA IA

No mundo contemporâneo, a resistência de profissionais em confiar em conselhos gerados automaticamente por tecnologias emergentes ainda é uma barreira à adoção da inteligência artificial (IA) em atividades de auditoria e controle interno (Commerford et al., 2022). Segundo Nakao et al. (2023), a aceitação plena de modelos autônomos de IA está longe de se consolidar, em parte devido à falta de transparência e explicabilidade de muitas soluções adotadas no mercado. Esse cenário gera desconforto em processos de tomada de decisão que excluem a participação humana.

Commerford et al. (2022) mostram que a credibilidade da fonte de informação influencia fortemente o julgamento dos auditores. Em contextos que envolvem tanto especialistas humanos quanto modelos baseados em IA, a origem do conselho impacta diretamente as recomendações formuladas. A disposição dos profissionais em aceitar as sugestões fornecidas pelos algoritmos é, portanto, determinante para a eficácia dessas soluções.

De acordo com Yeomans et al. (2019), existe uma tendência a confiar mais em informações oriundas de fontes humanas do que em fontes automatizadas. Para Castelo et al. (2019), esse comportamento está associado à crença de que algoritmos não são capazes de lidar com subjetividades do mundo real, o que os tornaria menos eficazes em tarefas que exigem intuição. Griffith (2018) reforça essa ideia ao afirmar

que indivíduos tendem a valorizar mais as fontes que percebem como experientes e confiáveis.

Além da questão da confiança, há ainda o fenômeno da chamada "aversão a algoritmos". Estudos como os de Yeomans et al. (2019) e Commerford et al. (2022) indicam que profissionais demonstram resistência quando decisões ou evidências são geradas por sistemas automatizados, mesmo quando esses sistemas são tecnicamente mais precisos. Essa aversão pode levar a escolhas decisões menos eficazes por parte dos auditores e impactar a adoção prática dessas ferramentas.

Fukas et al. (2022) apontam que a falta de transparência é um dos principais obstáculos na implementação de soluções baseadas em IA para o controle e tomada de decisão. A falta de clareza influencia não apenas a compreensão dos resultados, mas também a confiança nos sistemas e nas empresas que os desenvolvem (Avin et al., 2021).

Do ponto de vista técnico, Settipalli et al. (2022) alertam que o uso inadequado de modelos de aprendizado de máquina, sem a devida análise de desvios por especialistas humanos, pode resultar em inconsistências nos resultados. Shi et al. (2022) complementam, destacando os desafios na correção de bases de dados imperfeitas e no ajuste de modelos aplicados de maneira genérica.

A complexidade dos modelos também levanta preocupações sobre sua compreensão e governança. Zakaria (2021) argumenta que, à medida que os modelos de IA se tornam mais sofisticados, aumenta a necessidade de disseminar uma base teórica clara, com definições, categorias e fundamentos que aproximem os usuários da lógica desses sistemas. Essa clareza é fundamental para sua aceitação no contexto da auditoria.

Do ponto de vista ético e regulatório, Christ et al. (2021) e Emett et al. (2023) indicam que o uso de sistemas de IA totalmente autônomos ainda não é plenamente aceito por empresas, em razão de incertezas sobre responsabilidade legal, confiabilidade e controle. Nesse sentido, Costanza-Chock et al. (2022) sugerem que padrões éticos e orientações de órgãos reguladores são essenciais para mitigar riscos e fortalecer a confiança nos sistemas.

A literatura também evidencia que os princípios criados até o momento para promover o uso responsável da IA ainda deixam lacunas, especialmente quanto à confiabilidade no desenvolvimento dessas soluções (Lehner et al., 2021). Nesse contexto, o próprio autor ressalta que permanece o desafio de alinhar avanços tecnológicos às expectativas legais e sociais.

Como alternativa à abordagem autônoma, Jiang et al. (2023) defendem a adoção de modelos de IA centrados no ser humano. Essa filosofia rompe com o paradigma de substituição e propõe uma cooperação mais próxima entre pessoas e tecnologia. Xu et al. (2024) reforçam que os sistemas sociotécnicos devem considerar a percepção dos usuários e integrar as capacidades humanas ao projeto dos modelos.

Shneiderman (2020) sustenta que a combinação de controle humano e automação pode resultar em melhores soluções. Para isso, é necessário desenvolver sistemas nos quais as decisões críticas sejam sustentadas tanto por algoritmos quanto por julgamento especializado.

Mesmo com as resistências e os desafios éticos, diversos autores reconhecem o potencial da IA para transformar os sistemas de controle. Askary et al. (2018) afirmam que a tecnologia pode melhorar a qualidade das informações geradas. Fedyk et al. (2022) e Rahahleh et al. (2021) defendem que, com uso adequado, a IA pode

reduzir falhas em auditorias e controles internos, mesmo que os efeitos sobre o trabalho humano se materializem gradualmente.

Por fim, Abdulameer et al. (2022) e Liu et al. (2021) apontam que o papel dos auditores está se transformando. A tendência é de redução da força de trabalho dedicada a tarefas repetitivas e de aumento da integração entre profissionais e sistemas inteligentes. Essa transição representa, ao mesmo tempo, um desafio e uma oportunidade para o futuro da auditoria e da governança organizacional.

2.3 IA CENTRADA NO SER HUMANO

A integração entre mecanismos de inteligência artificial (IA) e especialistas humanos, conhecida na literatura como IA centrada no ser humano (*Human-Centered AI – HCAI*), tem como premissa ampliar, e não substituir, as capacidades humanas. Essa abordagem é defendida por diversos estudos recentes que ressaltam a importância da sinergia entre avanços tecnológicos e habilidades humanas (Xu et al., 2023; Nakao et al., 2023; Shneiderman, 2020; Jiang et al., 2023; Xu et al., 2024; Nasarian et al., 2024; Dong et al., 2024).

Shneiderman (2020) destaca que as tecnologias baseadas em IA oferecem vantagens como algoritmos sofisticados, sensores avançados e grandes volumes de dados. No entanto, o verdadeiro ganho ocorre quando essas tecnologias são combinadas com estruturas organizacionais que valorizam as capacidades humanas. A estrutura HCAI propõe que os usuários mantenham controle adequado sobre os sistemas, mesmo diante de altos níveis de automação, o que contribui para um desempenho superior, mais ético e transparente.

Xu et al. (2023) observam que sistemas de IA estão se tornando mais próximos do comportamento humano, apresentando capacidades cognitivas e adaptativas. Essa evolução aumenta os riscos de decisões automatizadas não serem compreendidas ou aceitas, principalmente em contextos sensíveis. Por isso, autores como Nakao et al. (2023) defendem o desenvolvimento de modelos que incorporem justiça, imparcialidade e sensibilidade ao contexto, reforçando a importância de abordagens mais responsáveis e ajustadas à realidade dos usuários.

Além dos cientistas de dados, o desenvolvimento de soluções de IA deve envolver outras partes interessadas, como usuários finais e especialistas no domínio. Nakao et al. (2023) argumentam que, para garantir justiça e transparência, é necessário entender como diferentes grupos percebem esses valores e, a partir disso, construir interfaces mais compreensíveis e acessíveis.

Problemas relacionados à justiça algorítmica, como vieses estatísticos e sociais, são amplamente discutidos na literatura. Quando decisões de alto impacto são tomadas exclusivamente por sistemas automatizados, sem envolvimento humano, há um risco real de perda de legitimidade e confiança (Nakao et al., 2023). Nesse sentido, Shneiderman (2020) propõe uma distinção clara entre automação computacional e controle humano, defendendo o equilíbrio entre ambos como ideal para o desenvolvimento de soluções seguras e eficazes.

Dong et al. (2024) reforçam a necessidade de compreender como os indivíduos reagem ao uso da IA em funções que envolvem julgamento e autoridade. Em seu estudo, observaram uma forte rejeição à substituição de gestores humanos por algoritmos, especialmente quando as decisões envolviam aspectos emocionais ou relacionais. Curiosamente, o desempenho ou motivação dos participantes não variou

significativamente em função do tipo de gestor, humano ou algorítmico, mas as percepções de justiça e intenção de permanência foram sensíveis à forma como a decisão era percebida, e não ao agente decisor em si. Os próprios autores alertam, contudo, para as limitações metodológicas dos estudos sobre reações à gestão algorítmica, que podem gerar resultados inconsistentes. Isso reforça a importância de interpretar esses achados com cautela e de ampliar o escopo das pesquisas em contextos reais.

Outro exemplo citado por Nakao et al. (2023) envolve o uso de IA por agentes de crédito. Nesse cenário, os profissionais enxergavam os sistemas automatizados como ferramentas auxiliares, e não como substitutos da experiência humana. Havia consciência de que os modelos de IA não captam plenamente o contexto decisório, sobretudo quando lidam com informações subjetivas ou ausentes dos dados estruturados. Assim, cientistas de dados optaram por um processo híbrido, em que as recomendações geradas pelos algoritmos eram verificadas por especialistas antes da adoção final.

Shneiderman (2020) reforça esse entendimento ao discutir quando a IA deve servir como ferramenta de automação e quando deve funcionar como ampliação das capacidades humanas. Segundo o autor, os usuários demonstram maior aceitação quando sentem que mantêm o controle, podem antecipar ações do sistema e têm possibilidade de intervenção. Esse comportamento é observável no uso cotidiano de tecnologias como elevadores, dispositivos médicos e câmeras, onde a confiança na automação se dá sem abdicar do controle humano.

Jiang et al. (2023) observam que o foco exclusivo em automação, comum em modelos tradicionais de desenvolvimento de IA, gerou desconfiança e resistência. Em

contrapartida, a abordagem HCAI representa uma mudança de paradigma ao valorizar a colaboração entre pessoas e máquinas. Os autores defendem que soluções eficazes exigem mais do que boa tecnologia: dependem também de estruturas sociais e organizacionais adequadas.

Xu et al. (2024) vão além e afirmam que a HCAI deve ser entendida como uma filosofia de desenvolvimento. Ela promove a valorização das habilidades humanas desde o projeto inicial, com foco na mitigação de riscos e maximização dos benefícios da IA. O autor argumenta que erros do passado, como o excesso de centralização tecnológica, precisam ser superados com modelos que coloquem o ser humano no centro do processo.

Essa mudança de abordagem exige liderança por parte dos desenvolvedores, que devem promover a colaboração interdisciplinar e estimular a integração entre áreas técnicas, sociais e organizacionais. Segundo Xu et al. (2023), isso fortalece uma cultura voltada à construção de sistemas mais confiáveis, úteis e funcionais.

Contudo, ainda persistem tensões. Jiang et al. (2023) relatam que parte da comunidade técnica continua presa ao debate sobre substituição versus protagonismo humano. Para os autores, a estrutura HCAI resolve esse impasse ao oferecer um caminho em que a IA funciona como aliada e não como concorrente.

Shneiderman (2020) alerta para um risco adicional: a “arrogância algorítmica”, termo utilizado para descrever a postura de desenvolvedores que superestimam a capacidade dos sistemas autônomos. Essa atitude pode comprometer a segurança e usabilidade dos sistemas. Em contrapartida, projetos bem-sucedidos priorizam o entendimento dos usuários sobre o funcionamento do sistema e permitem intervenção humana quando necessário.

Além da governança, outro desafio enfrentado pelos modelos modernos de IA é a sua complexidade crescente. Nasarian et al. (2024) observam que o aumento de camadas, hiperparâmetros e interações internas torna os modelos preditivos menos compreensíveis. Por isso, defendem que a interpretabilidade deve ser priorizada, mesmo que isso implique alguma perda em precisão ou desempenho.

Esse debate impulsionou o surgimento de áreas como o Aprendizado de Máquina Interpretável (IML) e a Inteligência Artificial Explicável (XAI). Essas abordagens visam tornar os modelos mais transparentes e auditáveis, principalmente em setores sensíveis como saúde, finanças e segurança (Nakao et al., 2023).

A literatura também utiliza os conceitos de *Black-Box*, *Gray-Box* e *White-Box* para classificar o grau de transparência dos modelos. Os modelos *Black-Box* têm alta precisão, mas baixa interpretabilidade, dificultando a compreensão de seus resultados. Os *White-Box* são projetados para serem compreensíveis, ainda que em geral menos precisos. Já os *Gray-Box* representam uma solução intermediária: buscam equilíbrio entre precisão e explicabilidade (Nasarian et al., 2024).

Xu et al. (2024) propõem um quadro metodológico abrangente para o desenvolvimento de soluções HCAI sustentáveis e escaláveis. No nível técnico, recomendam integrar funções humanas à IA, evitando projetos exclusivamente automatizados. No nível do usuário, sugerem considerar os valores da HCAI desde o início. E no nível ético, defendem a construção de sistemas que conquistem a confiança dos usuários e sejam socialmente responsáveis.

Para viabilizar essa agenda, os autores sugerem medidas concretas:

(1) formação de talentos interdisciplinares em HCAI;

(2) criação de fundos governamentais para pesquisa aplicada;

(3) desenvolvimento de projetos colaborativos intersetoriais;

(4) definição de padrões da Organização Internacional de Normalização (ISO) e do Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE) que formalizem metodologias centradas no ser humano.

2.4 ALGORITMOS PARA DETECÇÃO DE FRAUDES

Diversos estudos aplicados à área da saúde têm utilizado algoritmos de aprendizado de máquina, supervisionados ou não supervisionados, para detectar eventos suspeitos de fraude (Matloob et al., 2020; Settipalli et al., 2022; Massi et al., 2020; Cheng et al., 2024; Hong et al., 2024; Wang et al., 2020). A literatura indica que, na maioria dos casos, recorre-se a métodos não supervisionados devido à ausência de rótulos nos conjuntos de dados, buscando identificar padrões de comportamento e detectar anomalias (Matloob et al., 2020). Já os algoritmos supervisionados exigem dados rotulados, com entradas e respectivas classificações, para aprender a mapear corretamente essas associações e realizar previsões para novos dados (Settipalli et al., 2022).

No entanto, métodos supervisionados podem apresentar baixa sensibilidade a valores discrepantes e dificuldades para capturar a complexidade de bases de dados extensas e heterogêneas, comuns na área da saúde (Hong et al., 2024). Além disso, quando surgem novos padrões de comportamento, há risco de perda de acurácia, pois esses modelos tendem a reproduzir padrões antigos aprendidos no treinamento. Por essas razões, a literatura frequentemente prioriza abordagens não supervisionadas, mais adaptáveis a dados não rotulados e variáveis. Há também uma tendência

crescente de integrar as duas estratégias, combinando a capacidade adaptativa da IA com o conhecimento especializado humano (Massi et al., 2020; Cheng et al., 2024).

Dois dos algoritmos mais utilizados na literatura para detecção de padrões e anomalias são o DBSCAN e o Random Forest (Fernández-Delgado et al., 2014; Schubert et al., 2017). O DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), proposto por Ester et al. (1996), é um método não supervisionado de agrupamento por densidade, eficaz na identificação de clusters com formatos variados, robusto a ruídos e outliers, e capaz de operar sem conhecimento prévio sobre a quantidade de agrupamentos, característica especialmente útil em bases médicas ruidosas e despadronizadas (Wang et al., 2020; Schubert et al., 2017).

Por sua vez, o Random Forest, desenvolvido por Breiman (2001) e consolidado em implementações como a de Liaw e Wiener (2002), é um algoritmo supervisionado baseado em múltiplas árvores de decisão, que operam de forma paralela para gerar previsões mais robustas e estáveis. É reconhecido por seu bom desempenho em tarefas de classificação e regressão, sendo considerado um dos classificadores mais eficientes em análises comparativas (Fernández-Delgado et al., 2014), inclusive em aplicações na detecção de padrões irregulares na saúde (Schonlau et al., 2020).

No contexto da saúde suplementar, a combinação desses algoritmos com validação por especialistas representa um caminho promissor para aprimorar as estratégias de detecção de eventos suspeitos de fraude, conciliando a eficiência computacional com a interpretação clínica qualificada.

2.5 HIPÓTESE

A ocorrência de eventos suspeitos de fraude no setor de saúde tem gerado crescente preocupação entre executivos e gestores da área. A adoção de tecnologias emergentes, como a inteligência artificial, vem sendo considerada uma aliada estratégica na tentativa de mitigar tais ocorrências. No entanto, a aceitação irrestrita de modelos autônomos ainda enfrenta barreiras, especialmente pela ausência de transparência, precisão e sensibilidade ao contexto, o que pode gerar desconfiança e resistência à adoção.

Como alternativa, a literatura tem destacado a abordagem conhecida como IA centrada no ser humano (HCAI), que propõe a integração entre os sistemas inteligentes e a expertise de profissionais. Em vez de substituir o julgamento humano, essa abordagem visa complementar suas capacidades, promovendo soluções mais ajustadas ao contexto institucional e às particularidades da operação em saúde suplementar.

Com base nesse entendimento, formula-se a seguinte hipótese a ser testada no presente estudo:

H1: Uma abordagem de IA centrada no ser humano melhora a precisão da identificação de eventos suspeitos de fraude, em comparação a modelos autônomos.

3. METODOLOGIA

3.1 BASE DE DADOS

Para alcançar o objetivo do presente estudo, foram utilizados dados de uma cooperativa médica de grande porte, localizada na região sudeste do país, com mais de 2.500 médicos cooperados e aproximadamente 400.000 beneficiários. De acordo com a classificação da Agência Nacional de Saúde Suplementar (ANS), cooperativas com mais de 100.000 beneficiários são consideradas de grande porte. A cooperativa envolvida no estudo manifestou grande preocupação com eventos suspeitos de fraude relacionados aos atendimentos vinculados ao método ABA (Applied Behavior Analysis), em razão do volume expressivo de sessões autorizadas e do impacto financeiro associado a esse tipo de serviço. Tal cenário motivou a definição deste escopo como foco central da presente pesquisa.

O método ABA, ou Análise do Comportamento Aplicada, é uma abordagem terapêutica baseada em evidências, amplamente utilizada no tratamento de indivíduos com transtorno do espectro autista (TEA). Fundamenta-se na aplicação de princípios da análise do comportamento para promover o desenvolvimento de habilidades funcionais e reduzir comportamentos disfuncionais (Cooper et. al., 2020).

A amostra analisada corresponde aos pagamentos de serviços de terceiros realizados no ano de 2024, especificamente relacionados ao método ABA. Foram extraídas 194.942 guias de pagamento, referentes ao atendimento de 3.277 beneficiários, contemplando principalmente sessões de fonoaudiologia, psicoterapia e terapia ocupacional.

3.2 MODELOS DE APRENDIZADO DE MÁQUINA UTILIZADOS

Foram empregados dois tipos de modelos de aprendizado de máquina para dar suporte à detecção de eventos suspeitos de fraude: modelo não supervisionado e modelo supervisionado.

O modelo não supervisionado foi aplicado a um conjunto de dados sem rótulos prévios, com o objetivo de agrupar guias de atendimento com características semelhantes e identificar anomalias que pudessem sugerir comportamentos atípicos. Nesse contexto, foi utilizado o algoritmo DBSCAN, conforme recomendado por Wang et al. (2020), em função de sua capacidade de identificar agrupamentos com formatos arbitrários, sua robustez a ruídos e outliers e o fato de não exigir a definição prévia do número de clusters uma vantagem importante para bases heterogêneas como as da saúde suplementar.

O modelo supervisionado, por sua vez, foi treinado com um conjunto de dados rotulado, definido com base em parâmetros de referência estabelecidos por especialistas da cooperativa. Foi utilizado o algoritmo Random Forest, reconhecido por sua precisão e estabilidade em tarefas de classificação (Schonlau et al., 2020). Esse modelo constrói múltiplas árvores de decisão durante o treinamento e realiza a classificação final por meio do voto majoritário entre as árvores, o que contribui para maior robustez e capacidade de generalização.

Ambos os algoritmos foram implementados com a biblioteca scikit-learn, desenvolvida em Python. Segundo Pedregosa et al. (2011), essa biblioteca reúne diversos algoritmos de aprendizado de máquina supervisionado e não supervisionado,

sendo amplamente utilizada pela comunidade científica por sua simplicidade e eficiência.

3.3 MÉTODOS DE EXECUÇÃO E ANÁLISE

Para atender ao objetivo deste estudo, foram definidos dois métodos complementares para a identificação de eventos suspeitos de fraude: um método autônomo e outro integrado à experiência humana.

O método autônomo baseou-se na aplicação do algoritmo DBSCAN a um conjunto de dados sem rótulos. Essa abordagem não supervisionada permitiu formar agrupamentos por densidade e identificar automaticamente registros considerados atípicos. Diversas combinações de parâmetros foram testadas durante o processo de calibragem, com atenção especial aos valores de min_cluster_size e min_samples. A configuração final selecionada (min_cluster_size = 12.200 e min_samples = 1) resultou na formação de 8 clusters distintos e na identificação de 12.807 outliers, o que corresponde a 6,57% da base analisada.

A qualidade da segmentação foi avaliada por meio do Silhouette Score, que atingiu 0,673, um valor considerado excelente em contextos com variabilidade e ruído. Os agrupamentos revelaram perfis homogêneos de atendimentos, com variações controladas na quantidade de sessões e nos valores unitários por atendimento. Já os registros classificados como outliers apresentaram características extremas ou incomuns, como valores unitários muito baixos (R$ 8,00) ou elevados (acima de R$ 120,00), além de quantidade de sessões superior a mil em uma única guia de serviço. A concentração desses atendimentos em poucos prestadores reforçou sua classificação como eventos suspeitos de fraude, por se distanciarem

significativamente dos padrões observados na maioria da base. Esta abordagem mostrou-se eficaz na detecção automática de padrões anômalos, sendo utilizada como base para construção do modelo supervisionado.

O método integrado, por sua vez, incorporou os princípios da inteligência artificial centrada no ser humano, promovendo a combinação entre algoritmos e o conhecimento de especialistas na gestão da cooperativa. Os especialistas realizaram uma análise do comportamento dos prestadores e definiram aqueles que melhor representavam o perfil ideal de prestação de serviços segundo a visão da operadora, considerando parâmetros gerenciais, assistenciais e de qualidade, além de tratar os eventos classificados como exceções. Com base nessas definições, estruturou-se um conjunto de dados rotulado, no qual os atendimentos foram classificados conforme sua aderência aos padrões estabelecidos.

Esse conjunto rotulado, composto por 19.794 registros, foi utilizado para treinar um modelo supervisionado com o algoritmo Random Forest, cuja escolha se deu por sua robustez, estabilidade e capacidade de generalização. A técnica de validação cruzada foi empregada para ajuste de hiperparâmetros e aprimoramento da acurácia preditiva.

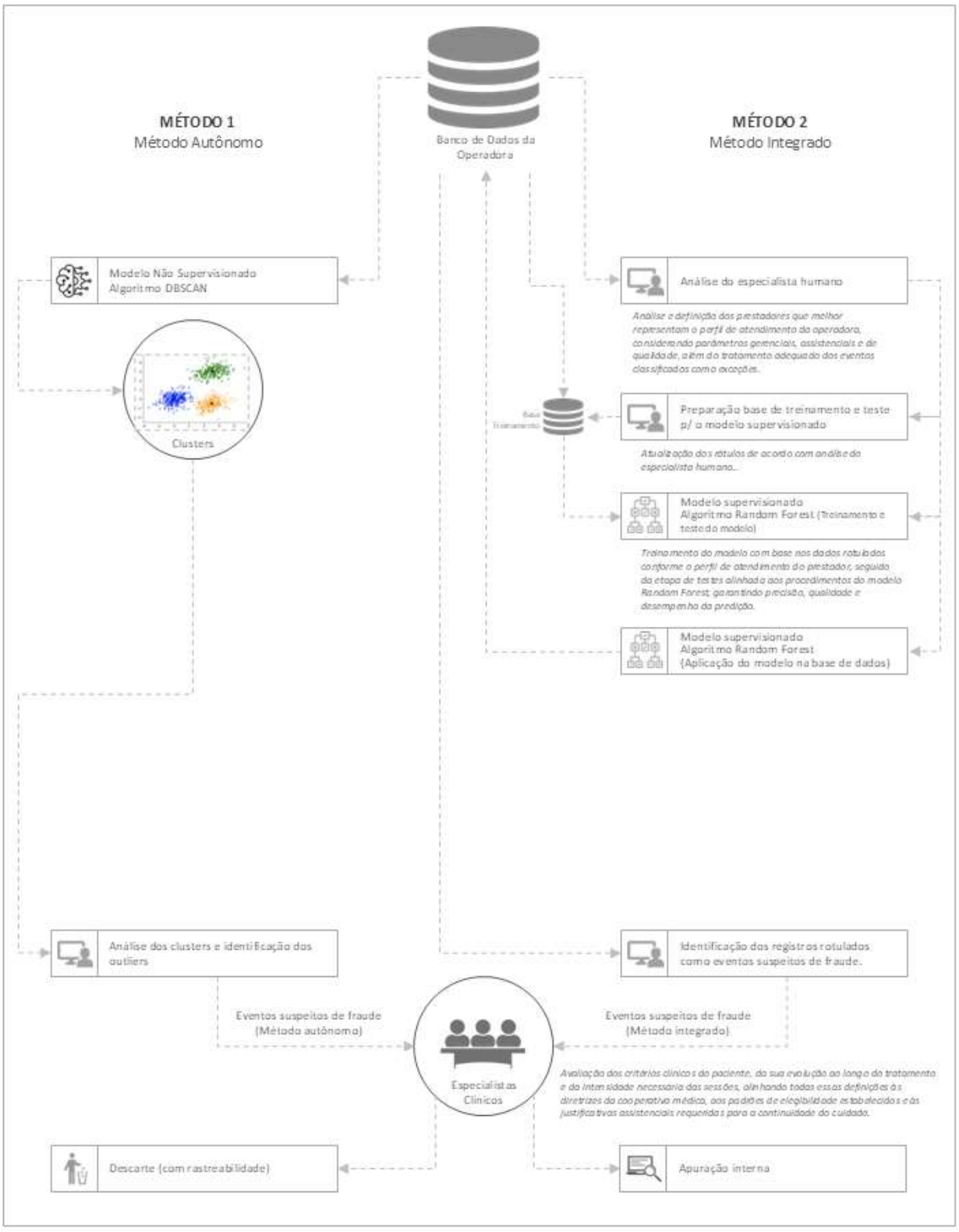
Após o treinamento e teste, o modelo foi aplicado aos demais registros da base, classificando automaticamente os atendimentos de acordo com os perfis aprendidos. Como resultado, foram classificadas 8.882 guias com alta probabilidade de representarem eventos suspeitos de fraude (4,56% do total), conforme os padrões aprendidos pelo modelo. Esse volume representa a proporção da base que mais se afastou dos perfis definidos como aderentes pelos especialistas, indicando maior risco

potencial. Cada previsão foi acompanhada de uma pontuação de confiança, permitindo distinguir entre previsões sólidas e casos incertos.

Os registros com baixa confiança preditiva foram tratados como potenciais eventos suspeitos de fraude e encaminhados à avaliação humana. No entanto, o principal diferencial da abordagem centrada no ser humano não está apenas na validação final, mas sim na etapa anterior de construção do modelo supervisionado, em que os especialistas da cooperativa definiram os parâmetros e os critérios de conduta esperados com base na realidade prática, nos objetivos assistenciais e na estratégia institucional. Essa participação ativa dos especialistas garantiu que o modelo refletisse as diretrizes e particularidades da organização, promovendo um monitoramento orientado por inteligência artificial, mas ancorado em supervisão qualificada.

A Figura 1 apresenta a arquitetura geral dos métodos empregados, evidenciando a atuação isolada do método autônomo e a integração sequencial entre DBSCAN, especialistas humanos e Random Forest no método integrado.

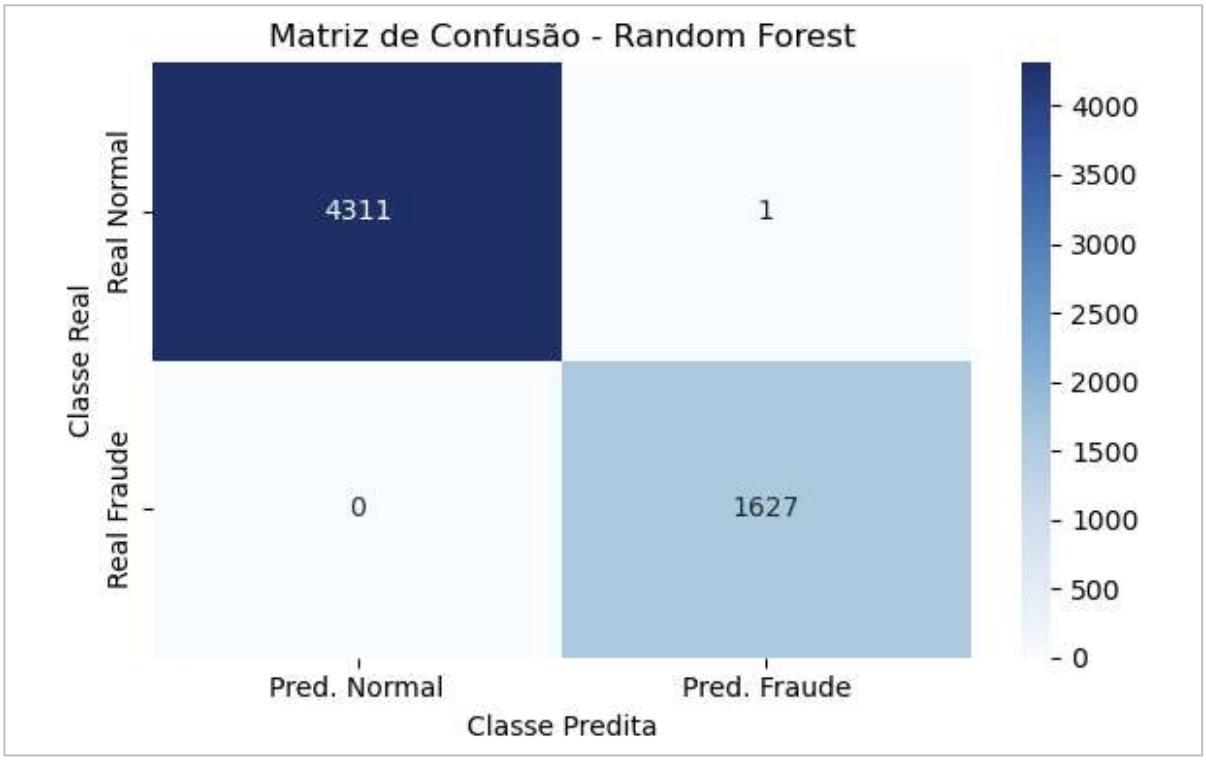
Figura 1 - Visão Geral – Métodos Autônomo e Integrado.



Elaborado pelo autor (2025).

Após o treinamento do modelo supervisionado, foram obtidos resultados altamente satisfatórios em termos de desempenho preditivo. A Figura 2 apresenta a matriz de confusão gerada, que evidencia a capacidade do modelo em separar corretamente os registros classificados como normais e os eventos suspeitos de fraude.

Figura 2 - Matriz de Confusão – Random Forest.



Elaborado pelo autor (2025).

A análise da matriz revela que houve apenas dois falsos negativos, ou seja, a grande maioria dos eventos suspeitos rotulados foram corretamente identificados, e apenas dois falsos positivos foram registrados.

No caso da classe “Evento Suspeito”, o modelo classificou 1.627 registros como positivos, todos eles verdadeiros positivos, sem ocorrência de falsos positivos, resultando em uma precisão de 100%.

A Tabela 1 apresenta métricas clássicas utilizadas na análise do desempenho obtido por modelos voltados à classificação binária, incluindo precisão, sensibilidade, F1-score e acurácia. Essas métricas permitem quantificar, de forma objetiva, o comportamento do modelo em relação à sua capacidade de identificar corretamente as classes “Normal” e “Evento Suspeito”.

Tabela 1 - Métricas de desempenho – Random Forest.

	Precision	Recall	f1-score	Support
Normal	1.0000	0.9998	0.9999	4312
Evento Suspeito	0.9994	1.0000	0.9997	1627
Accuracy			0.9998	**5939**
Macro avg	0.9997	0.9999	0.9998	5939
Weighted avg	0.9998	0.9998	0.9998	5939

Elaborado pelo autor (2025).

A precisão (*precision*) representa a proporção de registros classificados como positivos que, de fato, pertencem à classe positiva. É definida por:

$$Precisão = \frac{VP}{VP + FP}$$

em que:

VP (verdadeiros positivos) é o número de eventos suspeitos corretamente classificados como tal;

FP (falsos positivos) refere-se aos atendimentos normais incorretamente classificados como suspeitos.

A sensibilidade (recall) mede a proporção de instâncias da classe positiva que foram corretamente identificadas pelo modelo. Sua fórmula é:

$$Sensibilidade = \frac{VP}{VP + FN}$$

em que:

FN (*falsos negativos*) corresponde ao número de eventos suspeitos classificados incorretamente como normais.

O F1-score corresponde à média harmônica entre precisão e sensibilidade, sendo útil para sintetizar essas duas métricas em um único valor, especialmente em contextos onde há desequilíbrio entre as classes. É calculado por:

$$F1 = 2 * \frac{Precisão * Sensibilidade}{Precisão + Sensibilidade}$$

Neste caso, como a precisão e a sensibilidade da classe “Evento Suspeito” são próximas e elevadas, o F1-score também atinge valor aproximado de 0,99.

Por fim, a acurácia representa a proporção total de classificações corretas, tanto positivas quanto negativas, em relação ao número total de instâncias. É definida por:

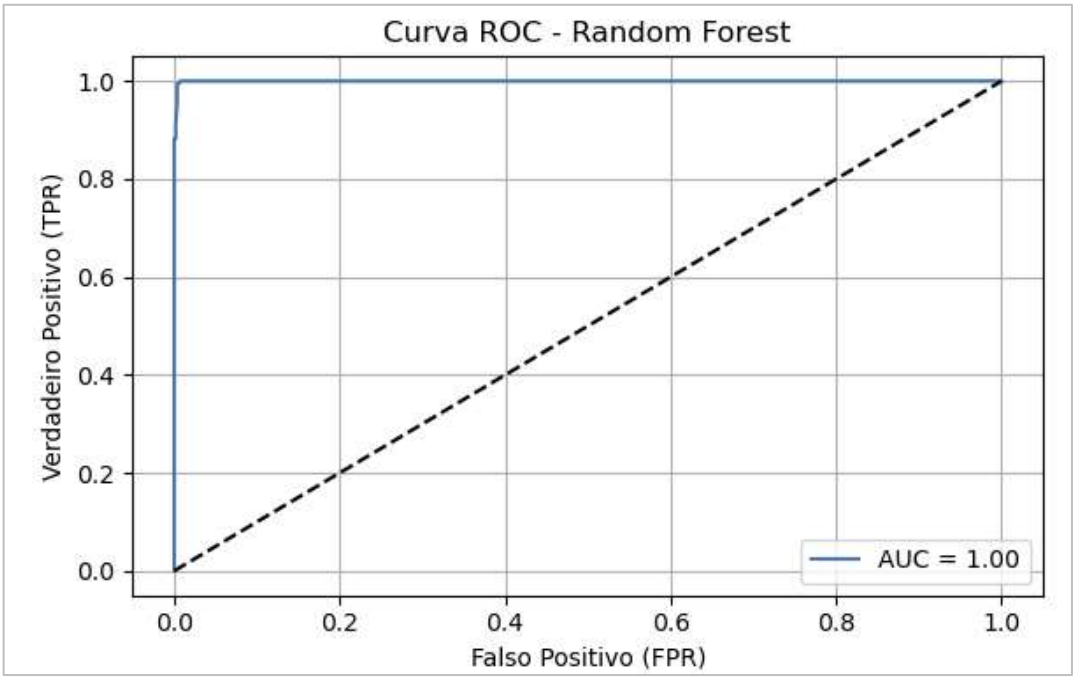
$$Acurácia = \frac{VP + VN}{VP + FP + FN + VN}$$

Onde, VN (*verdadeiros negativos*) indica os atendimentos normais corretamente classificados.

Com 4.311 registros corretamente classificados como normais (VN) e 1.627 corretamente classificados como eventos suspeitos (VP), sobre um total de 5.939 atendimentos, o modelo obteve uma acurácia de aproximadamente 99,98%.

A análise conjunta dessas métricas fornece uma visão detalhada do desempenho do modelo supervisionado, permitindo avaliar sua eficácia na distinção entre as classes, sem recorrer a interpretações subjetivas.

Figura 3 - Curva ROC – Random Forest.



Elaborado pelo autor (2025).

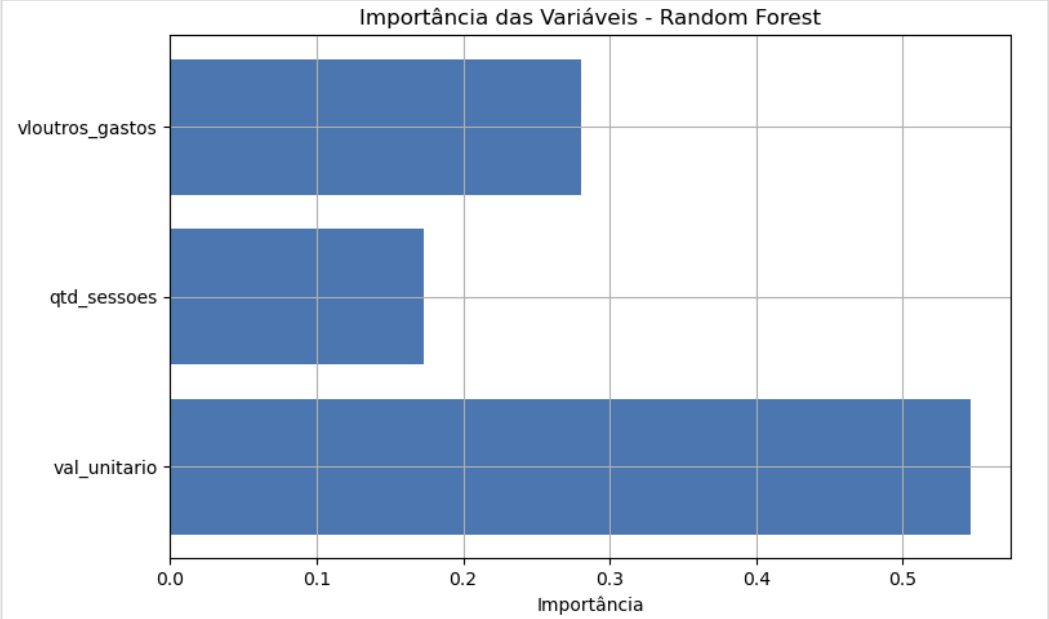
A Figura 3 apresenta a curva ROC (Receiver Operating Characteristic) gerada para o modelo Random Forest. A área sob a curva (AUC) foi de 1,00, indicando que, dentro do conjunto avaliado, o modelo foi capaz de distinguir perfeitamente entre as classes “Normal” e “Evento Suspeito”. Esse resultado sugere alta capacidade discriminativa do classificador.

Entretanto, é importante destacar que valores máximos de AUC devem ser interpretados com cautela, pois podem indicar possível sobreajuste (overfitting). O sobreajuste ocorre quando o modelo aprende não apenas os padrões gerais presentes nos dados de treinamento, mas também ruídos, flutuações e características específicas desse conjunto, o que pode comprometer sua capacidade de

generalização para novos dados. Por isso, apesar da AUC perfeita observada, torna-se essencial validar o modelo em diferentes conjuntos e avaliar seu desempenho em dados reais ou prospectivos.

Além das métricas apresentadas, foi avaliada a importância relativa das variáveis utilizadas pelo modelo. Conforme mostrado na Figura 4, o valor unitário da sessão foi o atributo mais relevante, seguido pelo valor de outros gastos e pela quantidade de sessões. Essa hierarquia sugere que guias com valores elevados por sessão e valor de outros gastos tendem a se alinhar aos perfis considerados suspeitos.

Figura 4 – Importância das variáveis no modelo Random Forest.



Elaborado pelo autor (2025).

Esses resultados demonstram não apenas a eficácia técnica do modelo, mas também sua aderência aos critérios definidos pelos especialistas na fase de rotulagem. Para os registros classificados com baixa confiança preditiva, o modelo recomendou o encaminhamento à avaliação humana, fortalecendo a proposta de integração entre inteligência artificial e supervisão especializada. Essa abordagem

representa um avanço importante na construção de sistemas de monitoramento baseados em inteligência artificial centrada no ser humano.

3.4 MÉTODO DE ANÁLISE DOS RESULTADOS

A etapa de análise dos resultados teve como objetivo avaliar a efetividade dos dois métodos empregados na identificação de eventos suspeitos de fraude. Conforme descrito anteriormente, os métodos aplicados, um autônomo, baseado exclusivamente em aprendizado não supervisionado, e outro integrado à experiência de especialistas, geraram agrupamentos distintos. Enquanto o primeiro operou de forma totalmente independente, o segundo foi construído com base em parâmetros definidos por profissionais da cooperativa.

Para comparar os resultados dos modelos, foi desenvolvido um algoritmo em Python capaz de validar os registros identificados como suspeitos por ambas as abordagens: o método autônomo (DBSCAN) e o método supervisionado (Random Forest). A validação foi realizada com base em regras clínicas e operacionais previamente definidas por especialistas clínicos, permitindo identificar, de forma objetiva, os eventos classificados como suspeitos que não apresentavam desvios relevantes. Esses casos foram tratados como falsos positivos, evidenciando as limitações de cada método.

Cabe destacar que o critério utilizado pelo especialista humano na fase de treinamento do modelo supervisionado, baseado essencialmente na análise do comportamento dos prestadores e na identificação daquele que melhor representa o perfil de atendimento desejado pela operadora, não corresponde aos parâmetros

empregados pelos especialistas clínicos, que avaliam diretamente a coerência e a justificativa clínica do atendimento.

No caso do método autônomo (DBSCAN), dos 12.807 eventos sinalizados como suspeitos, 11.641 foram validados pelas regras dos especialistas, enquanto 1.166 foram identificados como falsos positivos, o que representa um percentual de acerto de 90,90% e uma taxa de falsos positivos de 9,10%.

Já o método supervisionado (Random Forest) apresentou desempenho superior. Dos 8.882 eventos classificados como suspeitos, 8.386 foram confirmados como válidos, com apenas 496 registros considerados falsos positivos, resultando em um percentual de acerto de 94,42% e uma taxa de falsos positivos de 5,58%.

A partir dessa análise, foi possível obter subsídios para a comparação dos métodos adotados, considerando tanto a abrangência quanto a precisão dos sinais gerados. Os resultados dessa avaliação são apresentados no capítulo seguinte.

4 RESULTADOS

A análise dos resultados permitiu comparar o desempenho de dois métodos aplicados à identificação de eventos suspeitos de fraude: um modelo autônomo, baseado em aprendizado não supervisionado (DBSCAN), e um modelo supervisionado (Random Forest), treinado com base em parâmetros e critérios definidos por especialistas da cooperativa. A avaliação consistiu na análise técnica e clínica das listagens geradas por cada abordagem, conforme descrito no capítulo anterior.

O modelo supervisionado (Random Forest) apresentou desempenho superior tanto em termos de acurácia quanto de aderência aos critérios institucionais. Das 8.882 guias classificadas como suspeitas por esse modelo, 8.386 (94,42%) foram validadas como condutas passíveis de investigação, enquanto 496 guias (5,58%) foram descartadas como falsos positivos após revisão especializada.

O modelo não supervisionado (DBSCAN), por sua vez, identificou um total de 12.807 eventos suspeitos, dos quais 11.641 (90,90%) foram validados conforme os critérios clínicos e operacionais adotados, e 1.166 (9,10%) foram considerados falsos positivos. Apesar de sua capacidade de sinalizar um maior número absoluto de ocorrências, o método apresentou menor precisão, evidenciada pela maior taxa de alarmes injustificados.

Esses achados evidenciam que o modelo supervisionado foi mais eficaz na discriminação entre variações legítimas e padrões anômalos relevantes, refletindo maior aderência às diretrizes institucionais. Essa superioridade pode ser atribuída à utilização de rótulos construídos com base em conhecimento especializado,

permitindo ao modelo aprender padrões representativos de práticas consideradas suspeitas no contexto da cooperativa.

Adicionalmente, o modelo supervisionado demonstrou elevada consistência e desempenho estatístico, conforme demonstrado na etapa de validação cruzada (vide capítulo 3.3), apresentando altos índices de sensibilidade e especificidade. Já o modelo DBSCAN se mostrou valioso como ferramenta exploratória, especialmente na identificação de outliers extremos, mas sua maior suscetibilidade à geração de falsos positivos limita seu uso como solução autônoma para apoio à processos de verificação.

Em síntese, os resultados obtidos conferem maior robustez e confiabilidade ao modelo supervisionado (Random Forest), reforçando sua aplicabilidade como ferramenta estratégica de apoio à auditoria médica em ambientes complexos e sensíveis como o da saúde suplementar.

5 CONCLUSÃO

Este estudo teve como objetivo avaliar a efetividade de duas abordagens baseadas em inteligência artificial na identificação de eventos suspeitos de fraude em serviços de saúde, com foco específico nas sessões do método ABA. Foram comparados dois modelos: um autônomo, fundamentado em aprendizado não supervisionado (DBSCAN), e outro supervisionado, baseado no algoritmo Random Forest, calibrado com o apoio de especialistas da cooperativa.

A comparação entre as abordagens permitiu observar diferenças significativas na acurácia e na confiabilidade dos resultados gerados. O modelo supervisionado apresentou maior precisão e menor incidência de falsos positivos, enquanto o modelo não supervisionado, embora útil na detecção exploratória de padrões extremos, produziu um volume mais elevado de sinalizações não confirmadas. A avaliação técnica e clínica das guias classificadas como suspeitas evidenciou a superioridade do método supervisionado na triagem de eventos com indícios mais consistentes de anomalia.

Dessa forma, confirmou-se a hipótese central deste trabalho: a integração entre inteligência artificial e conhecimento humano especializado, característica da abordagem centrada no ser humano (HCAI), resultou em maior efetividade na detecção de eventos atípicos e redução de alertas imprecisos. A sinergia entre o poder computacional dos algoritmos e o julgamento contextualizado dos especialistas mostrou-se mais eficiente do que a adoção de modelo autônomo isolado, sobretudo em um domínio tão complexo e sensível quanto o da saúde suplementar.

Entre as principais contribuições deste estudo, destaca-se a aplicação prática de técnicas de aprendizado de máquina em um ambiente real de monitoramento e detecção de irregularidades na área da saúde, aliada à construção de um processo comparativo entre métodos supervisionados e não supervisionados. Ademais, a validação empírica da relevância da participação humana na parametrização, interpretação e refinamento de modelos preditivos reforça a importância de abordagens híbridas no desenvolvimento de soluções baseadas em IA.

Como limitação, ressalta-se que o estudo não envolveu a verificação formal da procedência dos eventos classificados como suspeitos, uma vez que o foco metodológico esteve voltado à triagem inicial com base em indícios. Além disso, os dados analisados foram oriundos de uma única cooperativa médica, o que pode restringir a generalização dos achados para outras operadoras de planos de saúde com perfis distintos.

Diante dos resultados obtidos, recomenda-se, para estudos futuros, a ampliação da base de dados com a inclusão de variáveis clínicas mais robustas e indicadores históricos de reincidência, bem como a replicação da abordagem em diferentes instituições do setor. Sugere-se, ainda, o aprofundamento em técnicas de explicabilidade algorítmica, com o objetivo de ampliar a transparência e a aceitabilidade institucional dos modelos preditivos utilizados em detecção e prevenção de fraudes.

Em síntese, os achados deste trabalho reforçam o potencial da inteligência artificial centrada no ser humano como estratégia eficaz, ética e sustentável para apoiar o processo de auditoria e controle de fraudes na saúde suplementar,

promovendo maior eficiência na alocação de recursos e na identificação de práticas que merecem investigação especializada.

REFERÊNCIAS

Abdulameer, M., Mansoor, M. M., Alchuban, M., Rashed, A., Al-Showaikh, F., & Hamdan, A. (2022). The impact of artificial intelligence (AI) on the development of accounting and auditing profession. In *Technologies, Artificial Intelligence and the Future of Learning Post-COVID-19: The Crucial Role of International Accreditation* (pp. 201–213). Springer International Publishing. https://doi.org/10.1007/978-3-030-93921-2_12

Askary, S., Abu-Ghazaleh, N., & Tahat, Y. A. (2018). Artificial intelligence and reliability of accounting information. In *Challenges and Opportunities in the Digital Era: 17th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2018* (pp. 315–324). Springer International Publishing. https://doi.org/10.1007/978-3-030-02131-3_28

Avin, S., Belfield, H., Brundage, M., Krueger, G., Wang, J., Weller, A., Anderljung,M., Krawczuk, I., Krueger, D., Lebensold, J., Maharaj, T., & Zilberman, N. (2021). Filling gaps in trustworthy development of AI. *Science*, *374*(6573), 1327–1329. https://doi.org/10.1126/science.abi7176

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bushra, A. A., & Yi, G. (2021). Comparative analysis review of pioneering DBSCAN and successive density-based clustering algorithms. *IEEE Access*, *9*, 87918–87935. https://doi.org/10.1109/access.2021.3089036

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Cheng, F., Yan, C., Liu, W., & Lin, X. (2024). Research on medical insurance anti-gang fraud model based on the knowledge graph. *Engineering Applications of Artificial Intelligence*, *134*, 108627. https://doi.org/10.1016/j.engappai.2024.108627

Christ, M. H., Emett, S. A., Summers, S. L., & Wood, D. A. (2021). Prepare for takeoff: Improving asset measurement and audit quality with drone-enabled inventory audit procedures. *Review of Accounting Studies, 26*(2), 665–685. https://doi.org/10.1007/s11142-020-09574-5

Commerford, B. P., Dennis, S. A., Joe, J. R., & Ulla, J. W. (2022). Man versus machine: Complex estimates and auditor reliance on artificial intelligence. *Journal of Accounting Research, 60*(1), 171–201. https://doi.org/10.1111/1475-679x.12407

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson.

Costanza-Chock, S., Raji, I. D., & Buolamwini, J. (2022, June). Who audits the auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In *Proceedings of the 2022 ACM Conference on Fairness*, Accountability, and Transparency (pp. 1571–1583). https://doi.org/10.1145/3531146.3533213

Deloitte. (2023). Global health care outlook 2023. https://www.deloitte.com/content/dam/assets-shared/docs/industries/life-sciences-health-care/2023/2023-health-care-outlook-final.pdf

Dong, M., Bonnefon, J. F., & Rahwan, I. (2024). Toward human-centered AI management: Methodological challenges and future directions. *Technovation*, *131*. https://doi.org/10.1016/j.technovation.2024.102953

Emett, S. A., Kaplan, S. E., Mauldin, E. G., & Pickerd, J. S. (2023). Auditing with data and analytics: External reviewers' judgments of audit quality and effort. *Contemporary Accounting Research*, *40*(4), 2314–2339. https://doi.org/10.1111/1911-3846.12894

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 226–231).

Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies, 27*(3), 938–985. https://doi.org/10.1007/s11142-022-09697-x

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real-world classification problems? *Journal of Machine Learning Research*, *15*, 3133–3181. http://jmlr.org/papers/v15/delgado14a.html

Fukas, P., Rebstadt, J., Menzel, L., & Thomas, O. (2022). Towards explainable artificial intelligence in financial fraud detection: Using Shapley additive explanations to explore feature importance. In X. Franch, G. Poels, F. Gailly, & M. Snoeck (Eds.), *Advanced Information Systems Engineering (Lecture Notes in Computer Science,* Vol. 13295, pp. 109–126). Springer. https://doi.org/10.1007/978-3-031-07472-1_7

Fraudes na Saúde Suplementar. (2021). *Fraudes na saúde suplementar*. JurisHealth https://media.jurishealth.com.br/site/pages-ADZ8ZmW6J7IsuiA.pdf

Griffith, E. E. (2018). When do auditors use specialists' work to improve problem representations of and judgments about complex estimates? *The Accounting Review, 93*(4), 177–202. https://doi.org/10.2308/accr-51926

Hong, B., Lu, P., Xu, H., Lu, J., Lin, K., & Yang, F. (2024). Health insurance fraud detection based on multi-channel heterogeneous graph structure learning. *Heliyon, 10*(9). https://doi.org/10.1016/j.heliyon.2024.e30045

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics, 178*(4), 977–1007. https://doi.org/10.1007/s10551-022-05049-6

Instituto Brasileiro de Geografia e Estatística. (2024). *Conta-satélite de saúde: Brasil: 2010–2021*. https://biblioteca.ibge.gov.br/visualizacao/livros/liv102075_informativo.pdf

Instituto de Estudos de Saúde Suplementar. (2017). *Impacto das fraudes e dos desperdícios sobre os gastos da saúde suplementar: Estimativa para 2017*. IESS. https://www2.iess.org.br/cms/rep/analise_especial_fraudes_estimativa_2017.pdf

Jiang, J., Karran, A. J., Coursaris, C. K., Léger, P. M., & Beringer, J. (2023). A situation awareness perspective on human-AI interaction: Tensions and opportunities. *International Journal of Human–Computer Interaction, 39*(9), 1789–1806. https://doi.org/10.1080/10447318.2022.2093863

Lehner, O. M., Ittonen, K., Silvola, H., Ström, E., & Wührleitner, A. (2022). Artificial intelligence based decision-making in accounting and auditing: Ethical challenges and normative thinking. *Accounting, Auditing & Accountability Journal, 35*(9), 109–135. https://doi.org/10.1108/aaaj-09-2020-4934

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22. https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest

Liu, G., Guo, J., Zuo, Y., Wu, J., & Guo, R. Y. (2020). Fraud detection via behavioral sequence embedding. *Knowledge and Information Systems, 62*, 2685–2708. https://doi.org/10.1007/s10115-019-01433-3

Liu, X. (2021). Transformation from financial accounting to management accounting in the age of artificial intelligence. In *2021 International Conference on Big Data Analytics for Cyber-Physical System in Smart City* (pp. 1185–1195). Springer Singapore. https://doi.org/10.1007/978-981-16-7466-2_131

Massi, M. C., Ieva, F., & Lettieri, E. (2020). Data mining application to healthcare fraud detection: A two-step unsupervised clustering method for outlier detection with administrative databases. *BMC Medical Informatics and Decision Making*, *20*, 1–11. https://doi.org/10.1186/s12911-020-01143-9

Matloob, I., Khan, S. A., & Rahman, H. U. (2020). Sequence mining and prediction-based healthcare fraud detection methodology. *IEEE Access, 8*, 143256–143273. https://doi.org/10.1109/access.2020.3013962

Nakao, Y., Strappelli, L., Stumpf, S., Naseer, A., Regoli, D., & Gamba, G. D. (2023). Towards responsible AI: A design space exploration of human-centered artificial intelligence user interfaces to investigate fairness. *International Journal of Human–Computer Interaction, 39*(9), 1762–1788. https://doi.org/10.1080/10447318.2022.2067936

Nasarian, E., Alizadehsani, R., Acharya, U. R., & Tsui, K. L. (2024). Designing interpretable ML system to enhance trust in healthcare: A systematic review. *Information Fusion*,*108*, 102412. https://doi.org/10.1016/j.inffus.2024.102412

National Health Care Anti-Fraud Association. (n.d.). *The challenge of health care fraud*. https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud/

Organization for Economic Co-operation and Development. (2024). *Health expenditure and financing (SHA) [Dataset*]. OECD. https://stats.oecd.org/index.aspx?DataSetCode=SHA

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R.., Dubourg, V., V., Vanderplas, Cournapeau, D., Brucher, M., Perrot J., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830. https://www.jmlr.org/papers/v12/pedregosa11a.html?utm_source=chatgpt.com

Rahahleh, M. H., Hamzah, A. H. B., & Rashid, N. (2021). The artificial intelligence in the audit on reliability of accounting information and earnings manipulation detection. In *The Big Data-Driven Digital Economy* (pp. 315–323). Springer. https://doi.org/10.1007/978-3-030-73057-4_24

Settipalli, L., Gangadharan, G. R., & Fiore, U. (2022). Predictive and adaptive drift analysis on decomposed healthcare claims using ART-based topological clustering. *Information Processing & Management*, *59*(3), 102887. https://doi.org/10.1016/j.ipm.2022.102887

Settipalli, L., & Gangadharan, G. R. (2023). WMTDBC: An unsupervised multivariate analysis model for fraud detection in health insurance claims*. Expert Systems with Applications, 215*, 119259. https://doi.org/10.1016/j.eswa.2022.119259

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), 3–29. https://doi.org/10.1177/1536867X20909688

Schubert, E., Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems, 42*(3), 1–21. https://doi.org/10.1145/3068335

Shahana, T., Lavanya, V., & Bhat, A. R. (2023). State of the art in financial statement fraud detection: A systematic review. *Technological Forecasting and Social Change*, *192*, 122527. https://doi.org/10.1016/j.techfore.2023.122527

Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: A systemic review. *Neural Computing and Applications*, *34*(17), 14327–14339. https://doi.org/10.1007/s00521-022-07472-2

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Sun, C., Li, Q., Li, H., Shi, Y., Zhang, S., & Guo, W. (2018). Patient cluster divergence based healthcare insurance fraudster detection. *IEEE Access*, *7*, 14162–14170. https://doi.org/10.1109/access.2018.2886680

Wang, Y., Gu, Y., & Shun, J. (2020, June). Theoretically-efficient and practical parallel DBSCAN. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data* (pp. 2555–2571). https://doi.org/10.1145/3318464.3380582

Wen, Y., & Holweg, M. (2023). A phenomenological perspective on AI ethical failures: The case of facial recognition technology. *AI & Society,* 1–18. https://doi.org/10.1007/s00146-023-01648-7

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to human interaction with AI systems: New challenges and opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human–Computer Interaction*, *39*(3), 494–518. https://doi.org/10.1080/10447318.2022.2041900

Xu, W., & Gao, Z. (2024, May). Enabling human-centered AI: A methodological perspective. In *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICHMS59971.2024.10555771

Yeomans, M., Shah, A. K., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, *32,* 403–414. https://doi.org/10.1002/bdm.2118

Zakaria, H. (2021). The use of artificial intelligence in e-accounting audit. In *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success* (pp. 341–356). Springer. https://doi.org/10.1007/978-3-030-62796-6_20

ARTIGO TECNOLÓGICO: FRAUDES EM COOPERATIVAS MÉDICAS: O GRANDE DESAFIO PARA UMA GESTÃO CONTEMPORÂNEA.

RESUMO

A crescente preocupação com práticas irregulares em cooperativas médicas tem impulsionado a busca por soluções mais eficientes e tecnicamente viáveis para o controle de fraudes. Este artigo apresenta os resultados de um levantamento prático realizado junto a gestores de uma cooperativa de grande porte, com foco na identificação das áreas mais vulneráveis e dos principais tipos de fraudes operacionais. A partir desse diagnóstico, são propostas diretrizes para o desenvolvimento de uma arquitetura tecnológica especializada, capaz de monitorar condutas atípicas de forma contínua. O modelo sugere o uso de inteligência artificial para identificar padrões de desvio com base em dados históricos e perfis comportamentais, combinando automação com supervisão humana especializada. A proposta inclui requisitos funcionais, módulos operacionais e fluxos técnicos, priorizando a integração com sistemas já utilizados pela instituição e a flexibilidade na definição de parâmetros. O objetivo é oferecer uma base prática e adaptável para a construção de plataformas antifraude mais eficazes, voltadas à realidade da saúde suplementar brasileira, servindo também como fundamento para o desenvolvimento do Produto Técnico Tecnológico descrito nesta tese.

Palavras-chave: Prevenção de Fraudes; Saúde Suplementar; Cooperativas Médicas; Tecnologia da Informação.

1 O SETOR DE SAÚDE E SEUS DESAFIOS NO COMBATE À FRAUDE.

O setor de saúde consolidou-se, nas últimas décadas, como um dos pilares relevantes da economia global, absorvendo uma parcela expressiva de recursos públicos e privados. Segundo dados da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), em 2021, os países membros destinaram, em média, 9,7% de seu Produto Interno Bruto para despesas com saúde (OECD, 2024; IBGE, 2024).

Nos Estados Unidos, essa proporção foi ainda mais significativa, atingindo 18,6% do PIB no mesmo período, com investimentos que ultrapassaram a marca de 4 trilhões de dólares em ações de prevenção, assistência e gestão de cuidados. No Brasil, a estrutura de financiamento segue a média da OCDE, mas revela forte participação direta de famílias e instituições privadas. Em 2021, foram aplicados 9,7% do PIB em gastos com saúde, sendo 5,7% provenientes de despesas particulares e 4% oriundos do setor público (IBGE, 2024).

A elevada movimentação de recursos, entretanto, traz consigo riscos relevantes relacionados à má utilização de verbas e práticas fraudulentas. Dados da National Health Care Anti-Fraud Association (NHCAA) indicam que, nos Estados Unidos, estimativas apontam que fraudes podem responder por valores entre 3% e 10% do total investido em saúde (NHCAA, n.d.). No contexto nacional, um levantamento do Instituto de Estudos de Saúde Suplementar (IESS) revelou que, em 2017, aproximadamente 19% das despesas da saúde suplementar foram impactadas por desperdícios ou irregularidades. Apesar de não ser um dado recente, o número ainda serve como referência importante diante da escassez de estudos atualizados sobre o tema no Brasil (IESS, 2017).

Esse cenário crítico tem levado operadoras e cooperativas médicas a buscar métodos mais modernos e eficazes de auditoria, controle e prevenção. Além dos modelos tradicionais de análise, gestores passaram a considerar o uso de tecnologias emergentes, como algoritmos de aprendizado de máquina e soluções baseadas em inteligência artificial, como caminhos viáveis para aprimorar a detecção de padrões anômalos e acelerar processos de triagem (Askary et al., 2018; Fedyk et al., 2022; Rahahleh et al., 2021; Shahana et al., 2023).

Contudo, estudos recentes indicam que a simples adoção de sistemas autônomos, sem considerar as características institucionais e culturais das organizações, não garante resultados efetivos. Em muitos casos, a ausência de transparência nos critérios utilizados por modelos automatizados gera desconfiança, resistência e baixa adesão entre os profissionais das áreas envolvidas (Commerford et al., 2022; Nakao et al., 2023; Jiang et al., 2023).

Nesse contexto, o conceito de Inteligência Artificial Centrada no Ser Humano (Human-Centered AI - HCAI) vem ganhando relevância por propor a combinação entre ferramentas automatizadas e o conhecimento técnico de especialistas, com o objetivo de garantir que os sistemas apoiem, e não substituam, a tomada de decisão (Shneiderman, 2020; Xu et al., 2024). A proposta valoriza a transparência, a rastreabilidade e a participação ativa dos profissionais, aspectos fundamentais em um setor onde decisões impactam diretamente vidas e recursos financeiros expressivos.

No caso das cooperativas médicas, o desafio é ainda maior. Essas instituições operam com governança compartilhada, múltiplos fluxos assistenciais e uma estrutura decisória que exige equilíbrio entre aspectos clínicos, operacionais e financeiros. Compreender quais setores estão mais expostos, como ocorrem as fraudes e quais

lacunas existem nos controles internos é essencial para direcionar o desenvolvimento de soluções tecnológicas que respeitem as particularidades desse modelo.

Este artigo parte justamente desse entendimento prático e institucional, buscando propor uma arquitetura antifraude que se baseia em evidências reais, sem desconsiderar o contexto operacional da saúde suplementar. A seguir, são discutidos casos e iniciativas que demonstram como a combinação entre tecnologia e conhecimento especializado tem se mostrado eficaz na mitigação de riscos e no fortalecimento dos controles.

2 ALIADOS NO COMBATE À FRAUDE

Os avanços tecnológicos têm sido cada vez mais discutidos como ferramentas fundamentais para reforçar a segurança e a sustentabilidade do setor de saúde, especialmente em ambientes que operam com grandes volumes de dados e múltiplos fluxos de atendimento. Tecnologias baseadas em inteligência artificial, mineração de dados e aprendizado de máquina vêm sendo aplicadas em diferentes países como aliadas estratégicas no enfrentamento de fraudes e desperdícios que afetam operadoras, hospitais e sistemas de saúde suplementar (Askary et al., 2018; Fedyk et al., 2022).

Sun et al. (2018) demonstraram, em estudo conduzido em uma seguradora chinesa, que a aplicação de métodos de clusterização combinada à intervenção de médicos especialistas pode aprimorar a identificação de condutas suspeitas, como cobranças indevidas e prolongamentos não justificados de internações. A validação e padronização feitas por especialistas aumentaram significativamente a eficiência do

modelo, reforçando a relevância de abordagens híbridas que integrem inteligência artificial e conhecimento técnico humano.

Outro estudo relevante, conduzido por Matloob et al. (2020), explorou o uso de algoritmos para mapear sequências de atendimento em um hospital no Paquistão. A combinação entre o algoritmo Prefixspan e árvores de decisão probabilísticas permitiu gerar padrões típicos de cada especialidade. Quando confrontadas com os dados reais dos pacientes, essas sequências revelaram anomalias antes imperceptíveis. A etapa de revisão com profissionais da saúde foi essencial para validar os padrões e reduzir falsos positivos.

A importância da dimensão temporal também é evidenciada por Settipalli et al. (2022), que destacam a necessidade de adaptar os modelos de detecção às variações sazonais e tendências. Ao utilizar técnicas de decomposição de séries temporais e agrupamentos topológicos, os autores demonstraram como diferenciar desvios naturais de comportamentos potencialmente fraudulentos, proporcionando uma análise mais precisa e contínua.

O estudo de Massi et al. (2020) traz à tona o fenômeno do upcoding, a reclassificação indevida de tratamentos para categorias com maior remuneração. A partir de modelos não supervisionados, os pesquisadores agruparam internações com base em diagnósticos e custos, sinalizando hospitais com desvios relevantes em relação à média regional. As visualizações dos clusters foram utilizadas como apoio à decisão dos auditores, demonstrando o papel da inteligência visual como facilitadora da análise técnica.

Mais recentemente, autores como Xu et al. (2024) e Nakao et al. (2023) reforçam a importância de se adotar uma abordagem centrada no ser humano (HCAI)

ao aplicar sistemas inteligentes em ambientes sensíveis como a saúde. Para Xu et al. (2024), o engajamento dos profissionais na configuração, validação e interpretação dos alertas é essencial para mitigar vieses algorítmicos e aumentar a aceitação institucional. Nakao et al. (2023), por sua vez, apontam que a falta de clareza nas interfaces e a ausência de explicações compreensíveis limitam a adoção prática dessas soluções.

Essa necessidade de transparência e confiança também é discutida por Commerford et al. (2022), ao analisar o uso de IA em auditorias contábeis. O estudo mostra que, mesmo quando os algoritmos apresentam alta precisão, a credibilidade dos resultados depende diretamente da percepção de que há uma lógica clara por trás dos alertas. Essa lógica é especialmente aplicável ao contexto da saúde suplementar, onde erros de classificação podem gerar desconfiança e impactos financeiros relevantes.

No Brasil, ainda que a adoção de soluções automatizadas esteja em estágio inicial, o debate vem ganhando corpo. Relatórios como o do Instituto de Estudos de Saúde Suplementar (IESS, 2017) estimam perdas de quase 20% nas despesas do setor devido a fraudes e desperdícios. Apesar da gravidade do cenário, muitas operadoras ainda baseiam suas estratégias em auditorias retrospectivas e manuais, que, isoladamente, não são suficientes para lidar com a complexidade e o volume de dados gerados diariamente.

Diante disso, o aprendizado extraído dos estudos mencionados aponta para um denominador comum: os melhores resultados são alcançados quando há integração entre inteligência artificial e análise humana especializada. Em todos os casos relatados, o fator decisivo foi o envolvimento ativo de profissionais capazes de

interpretar os dados com base na realidade assistencial, jurídica e institucional do setor.

Esse entendimento fundamenta a proposta do presente estudo, que busca ir além das aplicações genéricas de IA para sugerir uma arquitetura tecnológica orientada à realidade das cooperativas médicas brasileiras. O modelo a ser apresentado valoriza o alinhamento entre tecnologia, cultura organizacional e processos institucionais, considerando que soluções híbridas, que unem automação com julgamento humano, representam o caminho mais seguro para fortalecer os controles internos e a sustentabilidade do setor.

3 ENTREVISTAS – METODOLOGIA E RESULTADOS

Durante os meses de janeiro e fevereiro de 2024, foram realizadas entrevistas com gestores das principais áreas de controle de uma cooperativa médica de grande porte. De acordo com a classificação da Agência Nacional de Saúde Suplementar (ANS), cooperativas com mais de 100 mil beneficiários se enquadram nesta categoria.

As entrevistas seguiram um roteiro estruturado com perguntas pré-definidas, aplicadas de forma presencial. O objetivo foi obter informações diretas dos gestores sobre riscos de fraudes no ambiente institucional, com base em fatos observados, ocorrências registradas e experiências práticas acumuladas ao longo de sua atuação.

As entrevistas foram conduzidas de maneira informal, sem gravação ou exigência de compromisso formal por parte dos participantes. O intuito foi reunir informações práticas sobre ocorrências de fraudes, setores com maior exposição a riscos e falhas percebidas nos mecanismos de controle interno, considerando a experiência operacional dos gestores em suas áreas de atuação.

Ao todo, foram realizados 13 encontros, envolvendo 22 gestores distribuídos em 20 áreas distintas da cooperativa. A Tabela 1 apresenta a codificação dos encontros (E1 a E13) e os respectivos registros. Os entrevistados foram estimulados a apontar as áreas mais suscetíveis a fraudes, os tipos de irregularidades mais recorrentes, as consequências observadas e as estratégias de mitigação já implementadas.

Áreas envolvidas: Comercial (Vendas e Pós-vendas), Suprimentos, Contratos, Regulação, Produção Médica, Contas Médicas, Cadastro, Financeiro, Controladoria, Almoxarifado, Tecnologia da Informação, Jurídico, Conselho Técnico, Intercâmbio, Loja, Assistencial (Hospitalar, Ambulatorial e Atenção à Saúde), Marketing, Ouvidoria, Ciência de Dados, Riscos, Governança e Compliance.

Como resultado, observou-se um forte destaque para a área Comercial, apontada por 77% dos entrevistados como a mais sensível à ocorrência de fraudes, especialmente nos processos de vendas e pós-vendas. Outras áreas com maior frequência de citação (acima de 30%) incluem Suprimentos, Contratos, Produção Médica e Cadastro.

Tabela 1 - Resumo das entrevistas e áreas mais citadas como vulneráveis.

Área	%	Total	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
Comercial	77%	10	1	1				1	1	1	1	1	1	1	1
Suprimentos / Compras	54%	7			1	1		1	1		1	1		1	
Contratos	38%	5	1			1	1		1		1				
Regulação / Produção Médica	38%	5		1		1	1						1	1	
Cadastro	31%	4		1					1					1	1
Financeiro	23%	3				1				1				1	
Controladoria	15%	2								1			1		
Almoxarifado	15%	2				1	1								
Tecnologia	8%	1							1						
Jurídico	8%	1		1											
Conselho Técnico	8%	1	1												
Intercâmbio	8%	1				1									
Loja	8%	1						1							
Assistencial	8%	1						1							
Marketing	8%	1											1		
Ouvidoria	0%	0													
Ciência de Dados	0%	0													
Riscos	0%	0													
Governança e Compliance	0%	0													

Elaborado pelo autor (2025).

Além da identificação de áreas críticas, os entrevistados também descreveram tipos específicos de fraudes já observadas no cotidiano da cooperativa. Os relatos foram organizados na Tabela 2, agrupados em seis categorias principais conforme a parte envolvida no evento fraudulento: beneficiários, agentes comerciais, cooperados, funcionários, terceiros/prestadores e sistemas.

A categoria “Cooperados” concentrou o maior número de ocorrências relatadas, totalizando 27% dos registros, o que evidencia a necessidade de mecanismos de monitoramento voltados à conduta médica. Também se destacaram as fraudes relacionadas à elegibilidade de beneficiários e solicitações indevidas sessões de tratamento, exames, materiais e prescrições.

Algumas áreas assistenciais, como Hospitalar, Ambulatorial e Atenção à Saúde, foram mencionadas de forma preliminar, mas exigem aprofundamento específico em entrevistas futuras com gestores diretamente vinculados às unidades de atendimento.

Tabela 2 - Tipos de fraudes citadas pelos entrevistados.

Partes Envolvidas	Tipos de Fraudes	Total	%
BENEFICIÁRIOS	Reembolso sem desembolso	4	16%
	Utilização de carteirinha por outra pessoa	4	
	Recibos falsos	2	
COMERCIAL	Elegibilidade (falsificação de documentos/informações)	6	17%
	Adesões de associações	2	
	Vendas não efetivadas (simulação de vendas)	2	
	Contratos não reajustados (ex: 2017) ou com reajustes abaixo do indicado pelo cálculo atuário	1	
FUNCIONÁRIOS	Contratações direcionadas em TI	3	16%
	Desvios de medicamentos	2	
	Benefícios para colaboradores concedidos por fornecedores/prestadores	2	
	Funcionário fazendo solicitação para ele mesmo	1	
	Uso de fornecedores inabilitados	1	
	Compras de urgência	1	
COOPERADO	Registro indevido de materiais (ex: OPME - prescrições a maior)	6	27%
	Registro indevidos de exames	4	
	Médicos que recebem por cirurgias, mas não participam	2	
	Turismo médico	1	
	Retornos após 30 dias (consultas indevidas)	1	
	Pedido de receita gerando registro de exame não realizado	1	
	Registro de ponto indevido por médicos	1	
	Autorizações indevidas	1	
TERCEIROS	Reembolso método ABA	4	17%
	Conflito de interesse - Indicações conectadas entre médicos e prestadores	4	
	Fornecedores/prestadores com vínculo com médicos/funcionários	3	
SISTEMAS	Perfil de acesso - uso indevido dos sistemas	3	6%
	Acesso indevido ao prontuário do paciente	1	

Elaborado pelo autor (2025).

4 CONTRIBUIÇÃO PARA UMA GESTÃO MAIS SEGURA

Fraudes em cooperativas médicas representam riscos significativos tanto do ponto de vista financeiro quanto institucional. Diante desse cenário, este estudo busca oferecer uma proposta concreta de solução tecnológica voltada à identificação e prevenção de condutas atípicas, com foco específico na realidade das cooperativas.

A proposta consiste na sugestão de construção de uma plataforma antifraude, baseada em tecnologias emergentes como inteligência artificial e aprendizado de máquina. Essa plataforma visa analisar dados de forma contínua, com múltiplas camadas de monitoramento e validação, a partir de três componentes principais:

Perspectiva de monitoramento, cada módulo da plataforma se baseia em uma perspectiva específica (ex: produção médica, sessões de tratamento, elegibilidade, solicitações de materiais), conforme ilustrado na Tabela 3.

Tabela 3 - Perspectivas de Monitoramento.

Perspectiva de Monitoramento	Elemento Investigado
Produção médica	Médico
Compra de OPME	Evento Cirúrgico
Atuação Comercial	Contrato
Uso do Plano	Beneficiário
Prestação de Serviços	Terceiro/Fornecedor

Elaborado pelo autor (2025).

Método de análise em três fases, organizado em 8 etapas, utilizado para implantar o monitoramento em cada perspectiva:

Figura 1 – Plataforma Antifraude.

Elaborado pelo autor (2025).

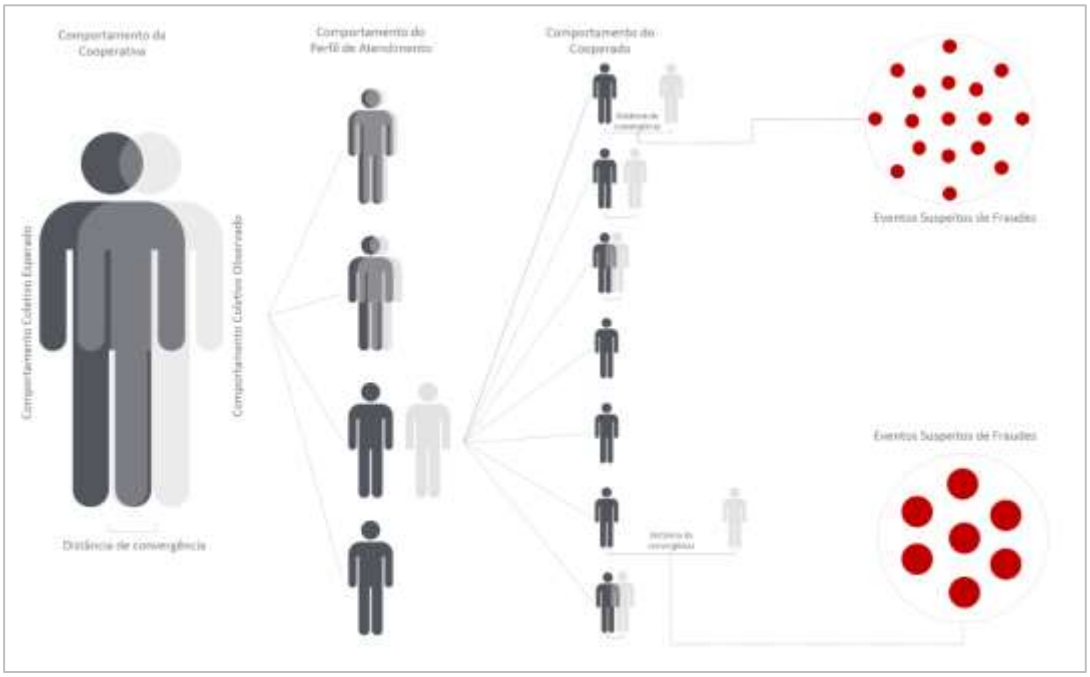
Fase 1 – Identificação de padrões esperados: Utiliza-se aprendizado de máquina para agrupar comportamentos semelhantes. Com base nesses agrupamentos, especialistas definem os “perfis esperados”, levando em conta aspectos como diretrizes institucionais, estratégias assistenciais e qualidade de atendimento.

Fase 2 – Classificação e clusterização de dados reais: A plataforma compara os dados reais com os perfis esperados, tanto em nível individual quanto coletivo. Essa análise permite identificar o grau de distanciamento do comportamento atual em relação ao padrão estabelecido.

Fase 3 – Geração de alertas e tratamento: Quando desvios significativos são detectados, a plataforma emite alertas para análise. O sistema registra as atividades de investigação, conclusões, justificativas e encaminhamentos.

Para exemplificar o funcionamento, a perspectiva de produção médica foi utilizada. Nessa abordagem, a plataforma analisa os atendimentos realizados pelos médicos cooperados, considerando variáveis como tipo de procedimento, materiais utilizados, exames solicitados, perfil dos pacientes e valores envolvidos. A partir disso, define-se um “Perfil de Atendimento”, que representa o comportamento esperado para grupos de profissionais com características semelhantes.

Figura 2 – Exemplo utilizando a perspectiva de produção médica.



Elaborado pelo autor (2025).

Essa classificação é mais precisa do que outras já existentes, como o código de especialidade ou área de atuação, que muitas vezes estão desatualizados. No caso da cooperativa analisada, apenas cerca de 10% dos médicos possuíam o campo “área de atuação” preenchido corretamente.

Com o perfil definido, a produção real de cada médico é comparada com o padrão esperado e com o comportamento coletivo de profissionais com perfil

semelhante. Esse processo permite identificar desvios individuais e desvios por grupo, sinalizando possíveis eventos suspeitos de fraude.

Os alertas gerados são encaminhados para análise pelos responsáveis técnicos, que podem registrar evidências, justificativas ou decisões de encaminhamento. O sistema permite análises por dashboards, filtros e histórico temporal, garantindo transparência, rastreabilidade e impessoalidade no processo de apuração.

Além da perspectiva de produção médica, outras perspectivas podem ser monitoradas pela mesma plataforma, como: guias de pagamento, contratos, autorizações, estoque, faturamento e relacionamento com prestadores. A estrutura foi desenhada para se integrar aos sistemas já utilizados pela cooperativa, com a criação de um barramento de dados capaz de consolidar e organizar informações provenientes de diferentes fontes.

A proposta visa oferecer um ambiente seguro, orientado por dados, e com ferramentas capazes de sustentar decisões mais precisas e técnicas. Com a análise automatizada e a participação ativa de especialistas, a plataforma reduz subjetividades, evita desgastes pessoais no processo de auditoria e fortalece a governança.

No futuro, a consolidação dos dados permitirá comparações entre cooperativas e análises longitudinais sobre a evolução dos comportamentos. Com isso, espera-se contribuir não apenas com a prevenção de fraudes, mas também com a construção de uma cultura institucional mais robusta, transparente e sustentável.

5 CONCLUSÕES DO ESTUDO

Este estudo tecnológico buscou oferecer uma contribuição concreta para o enfrentamento das fraudes em cooperativas médicas, por meio de um diagnóstico prático e da proposta de uma solução baseada em tecnologias emergentes. A partir de entrevistas com gestores de uma cooperativa de grande porte, foi possível identificar as áreas mais sensíveis a irregularidades e os tipos de fraudes mais frequentemente observados no contexto institucional.

Com base nesse mapeamento, estruturou-se uma proposta de plataforma antifraude orientada à realidade das cooperativas médicas brasileiras, combinando algoritmos de inteligência artificial com a participação ativa de especialistas. A solução apresentada vai além da simples automação de auditorias, ela propõe um modelo de monitoramento contínuo, parametrizável, transparente e integrado aos sistemas operacionais já existentes.

A principal inovação reside na adoção de um método de análise em múltiplas fases, capaz de identificar comportamentos atípicos com base em padrões esperados definidos pelos próprios especialistas da instituição. Isso garante que os alertas gerados estejam alinhados às diretrizes estratégicas, práticas assistenciais e especificidades culturais de cada cooperativa.

Além disso, a plataforma propõe uma arquitetura que favorece a rastreabilidade das decisões, o armazenamento estruturado de dados e a padronização do processo de tratamento dos alertas, contribuindo para uma apuração mais técnica, impessoal e sustentável.

Do ponto de vista prático, o estudo permite que gestores de outras cooperativas tenham acesso a uma leitura atualizada sobre os riscos de fraude no setor, com base em evidências concretas. A metodologia utilizada e os achados obtidos também podem servir de referência para o desenho de novas estratégias de controle, monitoramento e governança.

Em resumo, a proposta aqui apresentada busca apoiar uma gestão mais segura, proativa e orientada por dados, alinhada às necessidades do setor de saúde suplementar e às exigências crescentes de integridade, eficiência e transparência.

REFERÊNCIAS

Askary, S., Abu-Ghazaleh, N., & Tahat, Y. A. (2018). Artificial intelligence and reliability of accounting information. In Challenges and Opportunities in the Digital Era: 17th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2018 (pp. 315–324). Springer International Publishing. https://doi.org/10.1007/978-3-030-02131-3_28

Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? Review of Accounting Studies, 27(3), 938–985. https://doi.org/10.1007/s11142-022-09697-x

Fraudes na Saúde Suplementar. (2021). *Fraudes na saúde suplementar*. JurisHealth https://media.jurishealth.com.br/site/pages-ADZ8ZmW6J7IsuiA.pdf

Deloitte. (2023). *Global health care outlook: 2023.* https://www.deloitte.com/content/dam/assets-shared/docs/industries/life-sciences-health-care/2023/2023-health-care-outlook-final.pdf

Instituto Brasileiro de Geografia e Estatística. (2024). *Conta-satélite de saúde: Brasil: 2010–2021*. IBGE, Coordenação de Contas Nacionais. https://biblioteca.ibge.gov.br/visualizacao/livros/liv102075_informativo.pdf

Massi, M. C., Ieva, F., & Lettieri, E. (2020). Data mining application to healthcare fraud detection: A two-step unsupervised clustering method for outlier detection with administrative databases. *BMC Medical Informatics and Decision Making*, *20*(160), 1–11. https://link.springer.com/article/10.1186/s12911-020-01143-9

Matloob, I., Khan, S. A., & Rahman, H. U. (2020). Sequence mining and prediction-based healthcare fraud detection methodology. IEEE Access, 8, 143256–143273. https://doi.org/10.1109/access.2020.3013962

National Health Care Anti-Fraud Association. (n.d.). *The challenge of health care fraud*. https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud/

OECD. (2024). Health expenditure and financing (SHA) [Dataset]. OECD Health Statistics 2024. https://stats.oecd.org/index.aspx?DataSetCode=SHA

Rahahleh, M. H., Hamzah, A. H. B., & Rashid, N. (2021). The artificial intelligence in the audit on reliability of accounting information and earnings manipulation detection. In The Big Data-Driven Digital Economy (pp. 315–323). Springer. https://doi.org/10.1007/978-3-030-73057-4_24

Settipalli, L., Gangadharan, G. R., & Fiore, U. (2022). Predictive and adaptive drift analysis on decomposed healthcare claims using ART-based topological clustering. *Information Processing & Management*, *59*(3), 102887. https://doi.org/10.1016/j.ipm.2022.102887

Shahana, T., Lavanya, V., & Bhat, A. R. (2023). State of the art in financial statement fraud detection: A systematic review. *Technological Forecasting and Social Change*, *192*, 122527. https://doi.org/10.1016/j.techfore.2023.122527

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Sun, C., Li, Q., Li, H., Shi, Y., Zhang, S., & Guo, W. (2018). Patient cluster divergence based healthcare insurance fraudster detection. IEEE Access, 7, 14162–14170. https://doi.org/10.1109/access.2018.2886680

PRODUTO TECNOLÓGICO: GLOOMTRACE®: UMA PLATAFORMA DIGITAL ESTRUTURADA PARA DETECÇÃO E TRATAMENTO DE FRAUDES

RESUMO

O presente memorial descritivo apresenta o desenvolvimento de uma plataforma antifraude voltada a cooperativas médicas, concebida como um Produto Técnico Tecnológico (PTT). O produto foi idealizado para responder aos desafios enfrentados por operadoras de saúde suplementar no combate a práticas irregulares, especialmente diante das limitações dos métodos tradicionais de verificação retroativa e manual de processos. A proposta tecnológica consiste em uma solução híbrida, que combina modelos de inteligência artificial (IA) com validação humana especializada, estruturada a partir dos princípios da IA centrada no ser humano. A plataforma realiza o monitoramento contínuo de diferentes perspectivas de risco, com base na comparação entre padrões esperados e comportamentos reais. Quando identificadas divergências relevantes, o sistema gera alertas parametrizáveis e auditáveis, os quais podem ser analisados por especialistas por meio de uma interface transparente e rastreável. Do ponto de vista técnico, a plataforma incorpora algoritmos de aprendizado de máquina, estruturas de dados integradas e dashboards interativos que possibilitam análises segmentadas e acompanhamento por diferentes critérios. Seu design modular e flexível permite a integração com sistemas legados e a replicação do modelo em diferentes operadoras. O produto encontra-se atualmente em fase de teste e ajustes finais, com aplicação prevista em ambiente real de uma cooperativa médica de grande porte. Este PTT busca contribuir de forma direta para o fortalecimento dos mecanismos de controle, a mitigação de riscos financeiros e o reforço das práticas de conformidade e transparência institucional. Está diretamente alinhado aos achados apresentados no artigo científico e à proposta arquitetural discutida no artigo tecnológico. Além disso, promove avanços acadêmicos e técnicos ao propor uma abordagem inovadora, aderente à realidade da saúde suplementar brasileira e às exigências contemporâneas de governança, transparência e eficiência.

Palavras-chave: Fraude; Saúde Suplementar; Cooperativa Médica; Inteligência Artificial.

1. INTRODUÇÃO

O setor de saúde suplementar brasileiro movimenta montantes expressivos de recursos e cumpre papel estratégico na cobertura assistencial à população. Cooperativas médicas, em particular, exercem uma função relevante nesse ecossistema, reunindo milhares de profissionais e representando uma parcela significativa dos gastos nacionais em saúde (IESS, 2017). Estima-se que aproximadamente 19% das despesas da saúde suplementar estejam associadas a irregularidades, evidenciando a magnitude do problema e a necessidade de mecanismos de controle mais eficazes (IESS, 2017).

As práticas irregulares nesse segmento assumem diferentes formatos e podem ocorrer em várias etapas da cadeia de prestação de serviços, desde a produção assistencial até o faturamento (NHCAA, 2020). A detecção desses eventos é frequentemente dificultada por fatores como a dispersão de dados em múltiplos sistemas, a ausência de padrões unificados e a necessidade de analisar grandes volumes de informações em prazos curtos (IESS, 2017; OECD, 2023).

Historicamente, muitas organizações têm recorrido a métodos de verificação retroativa e manual de processos, que, embora importantes, apresentam limitações evidentes. Essas abordagens tendem a ser lentas, demandar elevado esforço humano e oferecer capacidade reduzida de detecção em tempo oportuno, especialmente diante da crescente complexidade das operações e do aumento do volume de dados (IESS, 2017).

Embora soluções baseadas em inteligência artificial (IA) venham sendo testadas no setor, a ausência de transparência nas decisões algorítmicas e a

dificuldade de interpretação de resultados comprometem a aceitação institucional dessas tecnologias (Commerford et al., 2022; Nakao et al., 2023). Nesse sentido, abordagens que integrem IA e análise humana especializada, como a inteligência artificial centrada no ser humano (Human-Centered AI – HCAI), mostram-se mais promissoras, por aliarem automação à avaliação técnica e contextual de especialistas (Shneiderman, 2020; Xu et al., 2023).

O presente Produto Técnico Tecnológico (PTT) descreve o desenvolvimento da plataforma GloomTrace®, concebida para integrar modelos de aprendizado de máquina a mecanismos de validação humana especializada, estruturada sob os princípios da IA centrada no ser humano. A proposta combina monitoramento contínuo de diferentes perspectivas de risco com análise baseada na comparação entre comportamentos esperados e ocorrências reais.

O desenvolvimento da plataforma contou com a colaboração de gestores e profissionais das áreas assistencial, jurídica e administrativa de uma cooperativa de grande porte, que contribuíram para o mapeamento de riscos, a definição de parâmetros de controle e a validação das funcionalidades operacionais. A arquitetura modular da solução possibilita integração com sistemas legados, adaptabilidade a diferentes contextos e replicação em outras operadoras. Além de gerar alertas parametrizáveis e rastreáveis, a plataforma oferece dashboards interativos que permitem análises por múltiplos critérios, como perfil de prestador, período e reincidência.

Com este PTT, busca-se contribuir não apenas para o aprimoramento tecnológico no setor de saúde suplementar, mas também para o reforço das práticas

de conformidade e transparência institucional, promovendo um ambiente mais seguro, eficiente e alinhado às demandas contemporâneas de governança.

2 DISCUSSÃO TÉCNICA

A crescente complexidade dos serviços de saúde, associada à digitalização de dados assistenciais, criou um ambiente desafiador para o fortalecimento dos mecanismos de controle em operadoras e cooperativas médicas. Os processos manuais de verificação, ainda amplamente utilizados, apresentam limitações quanto à abrangência, frequência e profundidade das análises realizadas (IESS, 2017). Essas restrições operacionais favorecem a persistência de condutas atípicas e práticas irregulares, que muitas vezes passam despercebidas até que os prejuízos se tornem significativos.

O Produto Técnico Tecnológico aqui descrito foi concebido para romper com esse modelo predominantemente reativo, por meio da construção de uma solução tecnológica que amplia a capacidade de monitoramento, promove rastreabilidade das decisões e melhora a efetividade da triagem de eventos suspeitos. A proposta parte da integração entre recursos de inteligência artificial e análise humana especializada, combinando automação com julgamento técnico, alinhada ao paradigma da Inteligência Artificial Centrada no Ser Humano (HCAI).

A plataforma utiliza como base um repositório integrado de dados, alimentado por diversas fontes internas da cooperativa, incluindo movimentações assistenciais, registros de autorizações, históricos de pagamento, dados de prestadores, elegibilidade e contratos. A arquitetura de dados foi desenhada para permitir o cruzamento entre diferentes dimensões do atendimento, possibilitando a construção

de perfis comportamentais e a detecção de variações anômalas. Esse repositório foi tratado com técnicas de uniformização, qualificação e unificação, garantindo confiabilidade e consistência analítica.

Do ponto de vista técnico, a solução combina métodos supervisionados e não supervisionados de aprendizado de máquina, calibrados com técnicas de validação cruzada para garantir equilíbrio entre precisão, sensibilidade e confiabilidade. O desenvolvimento envolveu o uso de Python com a biblioteca *scikit-learn*, Anaconda com Jupyter Notebook para exploração e validação de dados, Java (Spring) para implementação do back-end, React para construção do front-end e PostgreSQL como banco de dados principal. Essa combinação tecnológica garante robustez, escalabilidade e flexibilidade para integração com diferentes sistemas legados.

Um diferencial relevante da plataforma é a substituição das classificações formais por agrupamentos baseados em comportamento real, construídos a partir da prática observada. Essa estratégia permite captar nuances que passariam despercebidas por modelos tradicionais, aumentando a sensibilidade do sistema sem comprometer sua precisão. Além disso, a plataforma oferece um painel de validação para uso pelos especialistas da cooperativa, com filtros por tipo de evento, período, perfil de prestador, valor financeiro e reincidência.

O processo de validação é totalmente rastreável. Cada alerta gerado pelo sistema pode ser analisado individualmente por um especialista humano, que registra sua decisão (descartar ou encaminhar para apuração), a justificativa técnica e o histórico de ações relacionadas. Esse registro possibilita verificações futuras e confere legitimidade ao processo, reduzindo subjetividades e conflitos internos. Essa lógica está diretamente alinhada aos princípios da HCAI, conforme discutido por

Shneiderman (2020) e Xu et al. (2024), que destacam a importância da sinergia entre automação e controle humano em contextos sensíveis como o da saúde.

O desenvolvimento do produto exigiu o envolvimento de múltiplos conhecimentos técnicos, como ciência de dados, estatística aplicada, engenharia de software, modelagem de processos, aspectos jurídicos do ambiente digital e compliance. Também foi essencial a participação de gestores da cooperativa na definição dos parâmetros, validação de premissas e avaliação dos resultados esperados. Essa articulação interdisciplinar aumentou a robustez do produto e garantiu sua aderência à realidade institucional.

Desde a concepção, a plataforma foi projetada com enfoque na usabilidade e na integração harmônica às rotinas das equipes responsáveis pelo tratamento dos eventos. Isso envolveu não apenas a conformidade com a LGPD e demais requisitos legais, mas também o cuidado com a clareza das interfaces, a simplicidade na navegação e a adaptação do fluxo de validação às práticas já consolidadas da cooperativa. Tal abordagem reduz a curva de aprendizado dos usuários, aumenta a adesão ao sistema e garante que a operação diária não seja prejudicada pela introdução da nova tecnologia.

Por fim, o sistema foi estruturado para ser modular e replicável. Sua lógica pode ser facilmente adaptada a outras operadoras de saúde, desde que observadas as especificidades locais e os parâmetros regulatórios de cada instituição. O produto encontra-se atualmente em fase de testes e ajustes finais, com previsão de aplicação prática em uma cooperativa de grande porte.

Assim, a discussão técnica evidencia que o produto vai além de um sistema automatizado de alerta. Trata-se de uma infraestrutura de apoio à decisão, construída

com base em evidências, adaptável ao contexto institucional e orientada à melhoria dos controles internos. Ao promover um modelo híbrido de detecção de eventos suspeitos, associando algoritmos a critérios técnicos humanos, a plataforma oferece uma abordagem eficaz, transparente e sustentável para o enfrentamento de práticas irregulares em cooperativas médicas.

3 DESIGN

O desenvolvimento da plataforma antifraude GloomTrace® seguiu uma abordagem incremental, organizada em quatro etapas principais: (i) mapeamento dos processos internos da cooperativa; (ii) organização e tratamento das bases de dados; (iii) construção e testes dos modelos de inteligência artificial; e (iv) desenvolvimento da interface de validação e painel gerencial.

A primeira etapa concentrou-se no mapeamento dos fluxos operacionais críticos e na identificação dos pontos de controle mais relevantes para monitoramento automatizado. Essa atividade envolveu gestores e especialistas de áreas estratégicas, apoiando-se em reuniões, documentos institucionais e análise de dados históricos.

Em seguida, foi estruturado o repositório integrado de dados, reunindo informações de múltiplas fontes internas, como sistemas de autorização, faturamento, cadastro de prestadores e contratos. As bases passaram por processos de limpeza, padronização e enriquecimento, resultando em variáveis consolidadas que ampliaram a capacidade analítica da plataforma.

Na terceira etapa, foram implementadas e calibradas técnicas supervisionadas e não supervisionadas de aprendizado de máquina, utilizando validação cruzada para assegurar equilíbrio entre precisão, sensibilidade e confiabilidade. Essa camada de

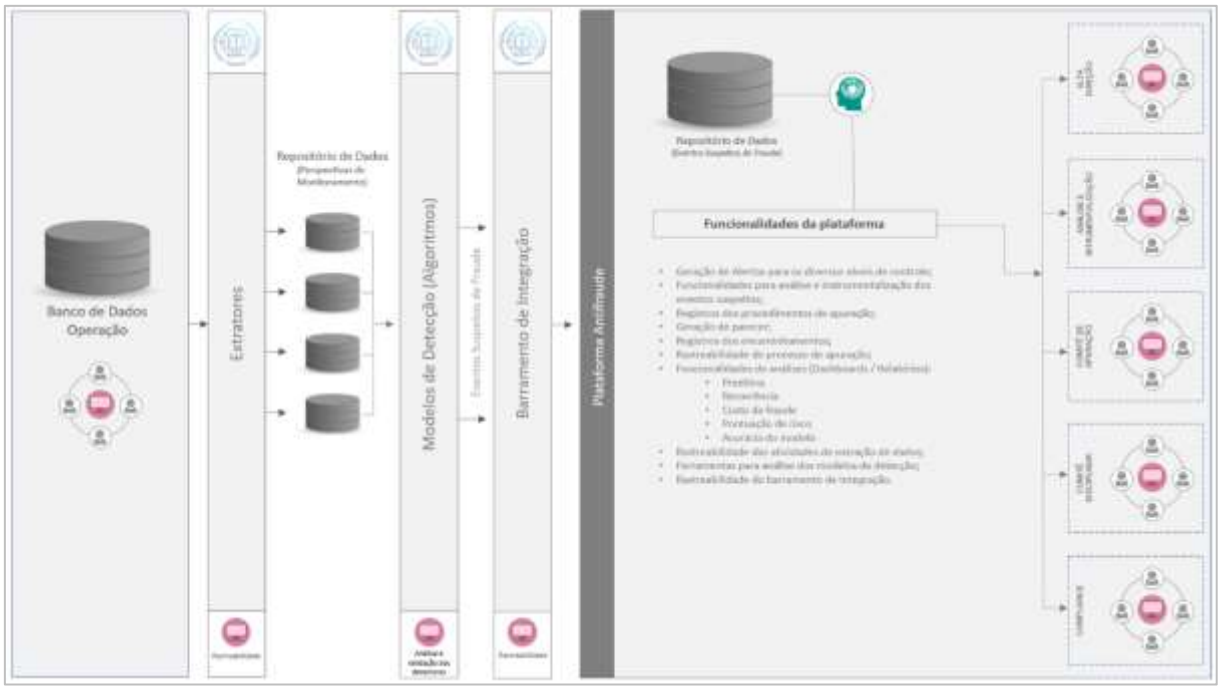
detecção foi projetada para operar de forma integrada ao fluxo institucional, gerando resultados compatíveis com as funcionalidades de visualização e análise da plataforma.

Por fim, a interface foi construída com foco em usabilidade, rastreabilidade e transparência. O sistema permite filtrar eventos por diferentes critérios e registrar as decisões tomadas, vinculando-as ao usuário autenticado e mantendo histórico de ações para auditoria e governança interna.

O projeto contou com a atuação integrada de cientistas de dados, desenvolvedores de software, auditores médicos e assistenciais, gestores operacionais e equipes de compliance, garantindo que os requisitos técnicos, clínicos e regulatórios fossem atendidos. A arquitetura modular, parametrizável e compatível com sistemas legados assegura flexibilidade para adaptações e replicabilidade do modelo em outras instituições.

A Figura 1 apresenta a arquitetura funcional da plataforma, contemplando as camadas de importação de dados, motores de detecção, mecanismos de rastreabilidade e painéis de validação humana, evidenciando o fluxo de informações e a integração com sistemas institucionais.

Figura 1 – Arquitetura funcional da plataforma GloomTrace.



Elaborado pelo autor (2025).

4 DETALHAMENTO DO PRODUTO (MVP)

A plataforma antifraude GloomTrace® encontra-se em estágio de MVP (Minimum Viable Product) e foi desenvolvida para atender cooperativas médicas que atuam no regime de saúde suplementar. Seu propósito é oferecer um conjunto integrado de funcionalidades voltadas ao monitoramento contínuo de eventos, identificação de situações potencialmente irregulares, validação técnica e suporte à tomada de decisão, adotando como fundamento o conceito de Inteligência Artificial Centrada no Ser Humano (HCAI).

A estrutura geral do produto compreende três módulos principais. O primeiro é o módulo de integração e coleta de dados, responsável por importar automaticamente informações assistenciais e financeiras provenientes dos sistemas operacionais da cooperativa, como faturamento, autorizações, contratos e cadastro de prestadores. As

conexões podem ser realizadas por arquivos estruturados (CSV, Excel) ou via APIs, quando disponíveis. Os dados recebidos passam por tratamento automático de consistência, unificação de variáveis e criação de atributos derivados, como frequência de atendimentos por prestador, valor médio por sessão e padrões de sazonalidade.

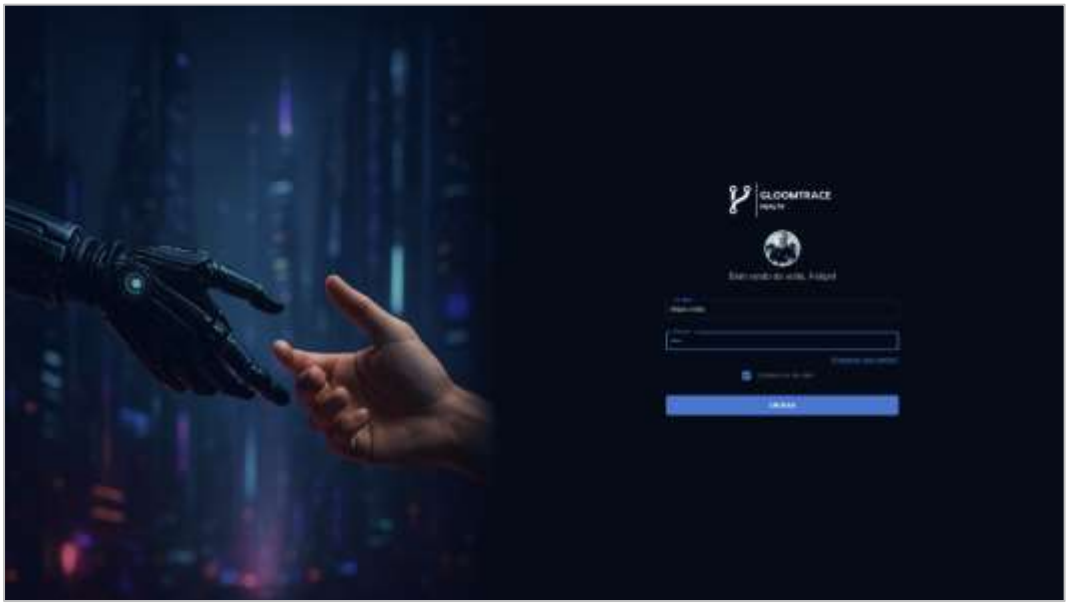
O segundo módulo é o de processamento e análise inteligente, que reúne os algoritmos de aprendizado de máquina calibrados para identificar condutas atípicas a partir de perfis históricos reais. São utilizadas abordagens supervisionadas e não supervisionadas, com técnicas de validação cruzada para cálculo de probabilidades e priorização por grau de risco. Os alertas são gerados acompanhados das variáveis críticas que influenciaram sua classificação, assegurando transparência na lógica de funcionamento.

O terceiro módulo é o de validação humana e acompanhamento, destinado aos especialistas da cooperativa. Nele, os alertas são apresentados com dados contextualizados, incluindo histórico do prestador, valores financeiros, reincidência e perfil de produção. A interface permite a aplicação de filtros, acesso a documentos associados e registro de decisões, seja para descartar o evento ou encaminhá-lo para apuração. Todo o processo é rastreável, com registro vinculado ao usuário autenticado, garantindo integridade institucional e conformidade com a LGPD.

O acesso à plataforma GloomTrace® é controlado por um sistema de autenticação individualizada, que vincula cada usuário a um perfil previamente definido pela instituição. Esses perfis determinam permissões específicas para visualização, análise e intervenção nos eventos suspeitos, assegurando que informações sensíveis sejam acessadas apenas por pessoas autorizadas. Essa

configuração garante conformidade com a LGPD, reforça a rastreabilidade das ações executadas e permite identificar de forma precisa o responsável por cada operação realizada no sistema (Figura 2 - Controle de acesso ao sistema).

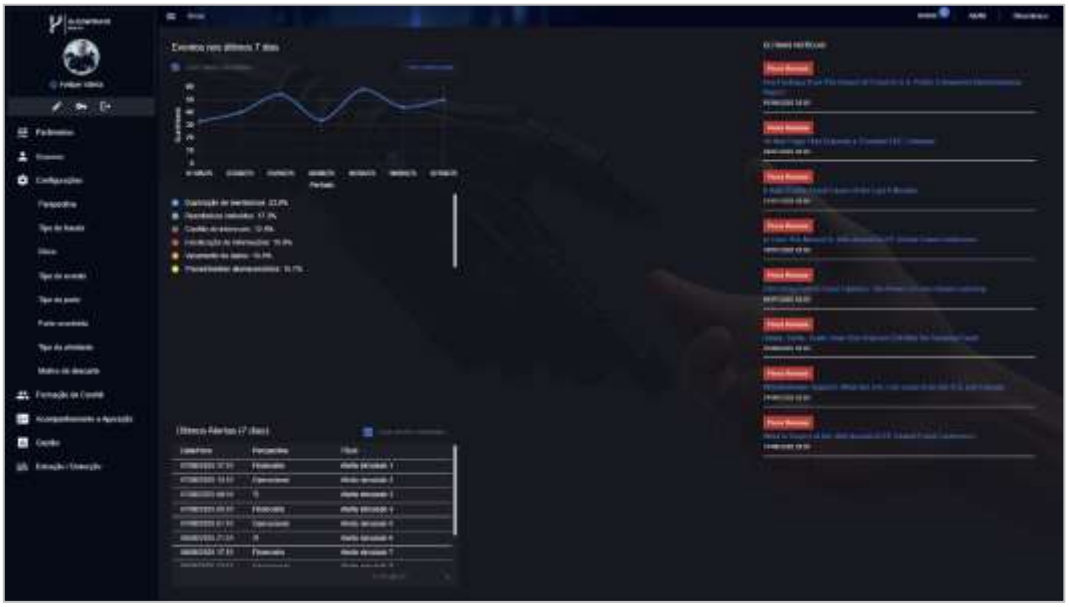
Figura 2 – Controle de acesso ao sistema.



Elaborado pelo autor (2025).

A plataforma GloomTrace oferece ampla parametrização de acordo com as necessidades institucionais, permitindo definir diferentes perspectivas de monitoramento, como produção médica, cooperados, prestadores e beneficiários, entre outras. Também é possível configurar os tipos de fraudes a serem acompanhados, atribuindo níveis de risco e categorizando detalhadamente os eventos que caracterizam cada ocorrência. O sistema possibilita ainda registrar as partes envolvidas, definir as atividades a serem executadas no processo de apuração e especificar os motivos para o descarte de eventos inicialmente classificados como suspeitos. Essa flexibilidade assegura que a plataforma se adapte a diferentes contextos operacionais, fortalecendo a precisão na detecção e a eficácia das ações de controle (Figura 3 – Configurações do sistema).

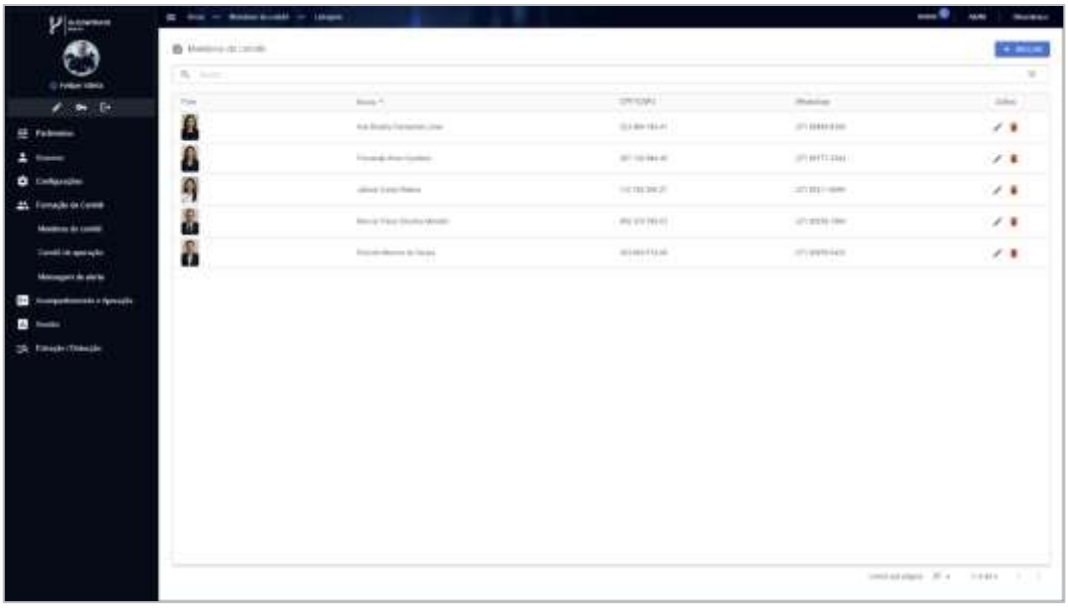
Figura 3 – Parametrizações do sistema.



Elaborado pelo autor (2025).

A funcionalidade de criação de comitês de apuração possibilita configurar grupos responsáveis pela análise e condução de investigações internas, registrando os membros que os compõem e definindo, para cada participante, o papel que desempenhará no processo. Essa estrutura garante clareza nas responsabilidades, alinhamento com os protocolos institucionais e rastreabilidade das ações, além de contribuir para a transparência e a eficácia das decisões tomadas (Figura 4 - Criação de comitês de apuração).

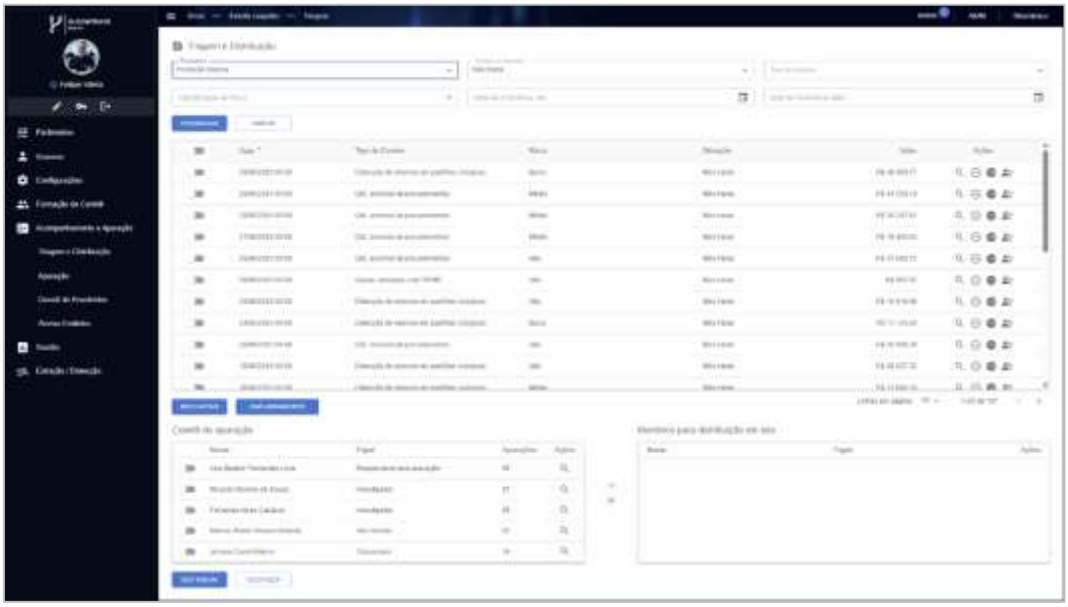
Figura 4 – Comitê de apuração.



Elaborado pelo autor (2025).

A etapa de triagem e distribuição de eventos suspeitos constitui um recurso essencial no processo de apuração da plataforma GloomTrace. Nessa fase, a área de controle realiza uma análise preliminar para verificar a pertinência do evento detectado, podendo descartá-lo quando não houver indícios suficientes ou se situações semelhantes já tiverem sido investigadas anteriormente. Nos casos de descarte, o registro permanece armazenado na base de rastreabilidade do sistema, possibilitando consultas ou fiscalizações futuras. Quando o evento é considerado procedente, ele é direcionado automaticamente ao comitê de apuração responsável, que conduzirá a análise detalhada e definirá as ações cabíveis (Figura 5 - Triagem e apuração).

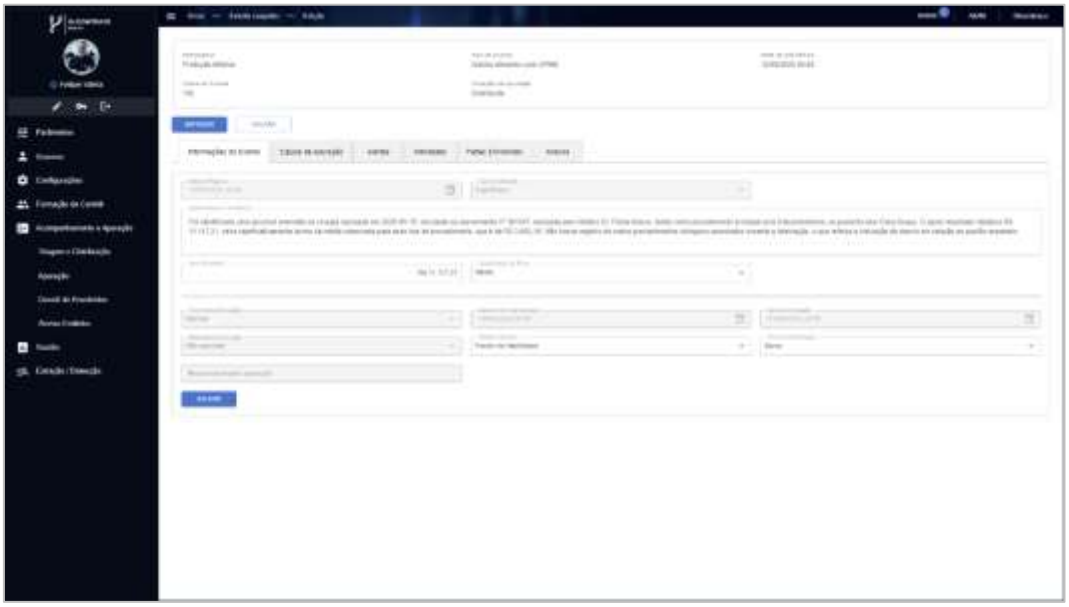
Figura 5 – Triagem e distribuição.



Elaborado pelo autor (2025).

No acompanhamento de um evento em investigação, a plataforma GloomTrace permite registrar um conjunto abrangente de informações, muitas delas geradas automaticamente pelos algoritmos na fase de detecção. Entre os dados registrados estão a equipe responsável pela condução da apuração, os alertas emitidos pelo sistema, as atividades executadas, as partes envolvidas e os documentos anexos relacionados às ações de apuração. Essa funcionalidade garante que todo o histórico do caso seja documentado de forma estruturada e rastreável, assegurando transparência e confiabilidade ao processo investigativo (Figura 6 - Acompanhamento e apuração).

Figura 6 – Acompanhamento e apuração.



Elaborado pelo autor (2025).

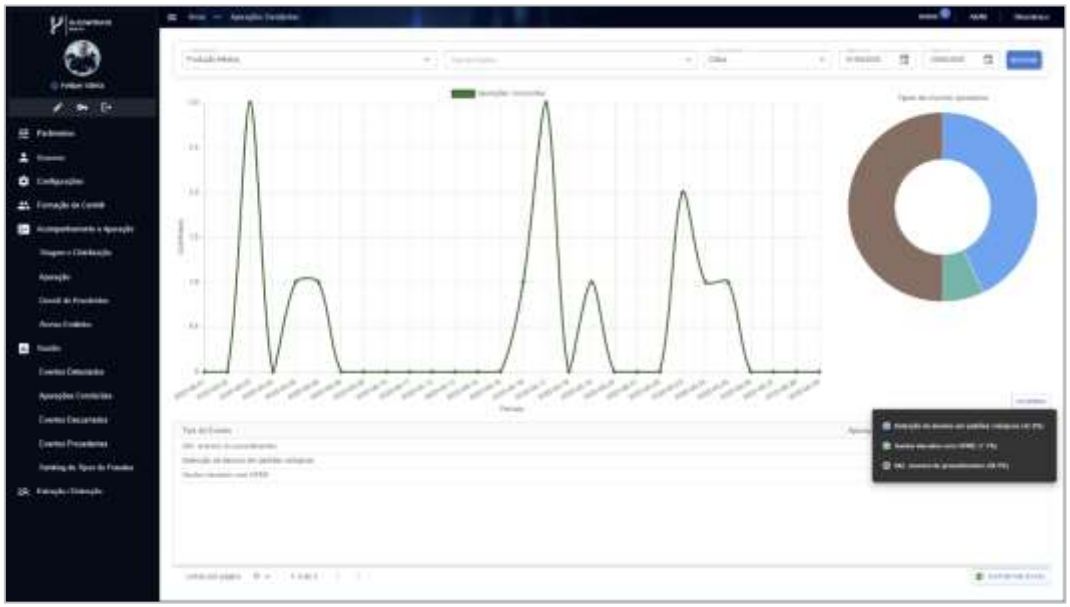
O dossiê de envolvidos consolida informações sobre todos os participantes relacionados a um evento suspeito, reunindo dados cadastrais, histórico de interações, reincidência, registros financeiros e vínculos institucionais. Essa organização favorece uma análise contextual mais precisa, permitindo ao especialista compreender o cenário de forma abrangente antes de tomar decisões no processo de apuração (Figura 7 – Dossiê de envolvidos).

Figura 7 – Dossiê de envolvidos.

Fonte: Autor (2025).

Os painéis de gestão da plataforma GloomTrace reúnem visualizações interativas que consolidam indicadores essenciais para o acompanhamento das atividades de controle. A interface possibilita filtrar informações por perspectiva de monitoramento, tipo de evento e período, apresentando gráficos sobre eventos detectados, apurações concluídas, eventos descartados, eventos procedentes e o ranking dos tipos de fraudes identificadas. Esses recursos permitem à alta gestão acompanhar a evolução das ocorrências, identificar padrões e subsidiar decisões estratégicas com base em informações atualizadas e rastreáveis (Figura 8 – Painéis de gestão).

Figura 8 – Painéis de gestão.



Elaborado pelo autor (2025).

A plataforma foi concebida para operar em ambiente de servidor em nuvem ou infraestrutura local com suporte a Docker. Sua arquitetura tecnológica combina Java (Spring Framework) no back-end, responsável pela lógica de negócios, APIs e integração com os módulos de processamento; React no front-end, oferecendo interface web interativa e responsiva; Python, com bibliotecas como Pandas e Scikit-Learn, para análise e modelagem dos dados, executado em ambiente Anaconda com JupyterLab; e PostgreSQL como banco de dados central para armazenamento dos dados tratados e resultados analíticos. Também há integração com ferramentas de visualização e exportação em formatos como Excel e PDF, sem dependência obrigatória de sistemas externos, mas com possibilidade de integração a sistemas legados via APIs RESTful.

O MVP foi concebido para permitir aplicação prática imediata, com investimento inicial reduzido e preservando a infraestrutura tecnológica já em uso. Sua flexibilidade

permite personalização de parâmetros, pesos e critérios de risco, adequando-se aos processos internos de cada instituição. Entre os benefícios operacionais e institucionais proporcionados estão o fortalecimento dos mecanismos de controle com adição de uma camada extra de segurança, a detecção independente e desvinculada dos fluxos internos, o sigilo sobre a lógica dos algoritmos para reduzir riscos de manipulação, a identificação de eventos com menor viés humano priorizando evidências comportamentais, a detecção automatizada e contínua em tempo quase real, a visão unificada para análise e apuração de alertas e a possibilidade de implantação discreta e evolutiva de novos mecanismos de controle.

A versão atual está em fase de testes com dados históricos reais de uma cooperativa médica de grande porte, com previsão de implantação em ambiente operacional nos próximos meses. O ciclo de feedback institucional servirá como base para ajustes e evolução contínua da plataforma, ampliando gradualmente seu escopo de monitoramento e sua capacidade preditiva. Essas características consolidam a GloomTrace como uma ferramenta estratégica de apoio à decisão, capaz de transformar o modelo de controle vigente sem causar rupturas nos sistemas legados e garantindo aderência às exigências regulatórias e de governança do setor.

5 CONCLUSÃO

O desenvolvimento da plataforma antifraude voltada a cooperativas médicas surgiu como resposta prática a um problema crônico do setor de saúde suplementar: a dificuldade de identificar, com agilidade e precisão, condutas assistenciais atípicas que possam configurar fraude, desperdício ou má alocação de recursos. A partir da construção de um produto baseado em inteligência artificial e controle técnico

humano, foi possível criar uma infraestrutura funcional e inovadora para o monitoramento contínuo de riscos operacionais.

A aderência do produto à área de Ciências Contábeis e Administração é evidente, pois ele atua diretamente sobre os sistemas de controle interno, auditoria, prevenção de perdas e suporte à decisão gerencial, todos eixos fundamentais da boa governança corporativa. Ao fortalecer a capacidade de detecção precoce de desvios, o produto contribui para a eficiência dos processos, a sustentabilidade econômica das instituições e a integridade da gestão.

No campo da inovação, o produto se diferencia por abandonar modelos normativos engessados e incorporar um arcabouço híbrido, baseado na combinação entre algoritmos de aprendizado de máquina e análise crítica de especialistas humanos. Ao utilizar perfis comportamentais reais em substituição a classificações estáticas, a solução introduz uma lógica mais realista e ajustada à prática institucional. A interface de validação rastreável e impessoal também representa um avanço em relação aos sistemas convencionais de controle.

A complexidade técnica e institucional do produto foi significativa, exigindo o envolvimento de múltiplos conhecimentos e a colaboração direta com profissionais da própria cooperativa médica. A construção do repositório de dados, o tratamento de inconsistências, a definição dos parâmetros clínicos e a calibragem dos modelos de IA demandaram conhecimento especializado em ciência de dados, auditoria, engenharia de software, conformidade legal e proteção de dados (LGPD) e gestão de riscos.

A aplicabilidade da plataforma foi uma prioridade desde a concepção do MVP. O sistema pode ser facilmente integrado a ambientes operacionais já existentes, sem

a necessidade de substituição de sistemas legados. Sua arquitetura modular e parametrizável facilita a replicação em outras operadoras, com adaptação aos respectivos perfis assistenciais e regulamentações locais. Essa flexibilidade técnica amplia o escopo de impacto e favorece a escalabilidade da solução.

Por fim, o impacto esperado vai além da redução de perdas financeiras. A plataforma promove maior rastreabilidade, impessoalidade e técnica nas decisões, contribuindo para um ambiente de maior governança, transparência e confiança institucional. Ao transformar dados brutos em alertas qualificados, validados por especialistas e organizados de forma inteligente, a solução oferece às cooperativas uma nova abordagem de controle: mais proativa, mais confiável e mais sustentável.

O produto encontra-se em fase de testes e ajustes finais, com previsão de aplicação prática em ambiente real nos próximos meses. Os resultados esperados incluem não apenas a melhoria dos indicadores operacionais e financeiros, mas também a validação institucional de um novo modelo de combate à fraude, mais alinhado à realidade do setor e às exigências contemporâneas de responsabilidade institucional e uso ético da tecnologia.

REFERÊNCIAS

Abdulameer, M., Mansoor, M. M., Alchuban, M., Rashed, A., Al-Showaikh, F., & Hamdan, A. (2022). The impact of artificial intelligence (AI) on the development of accounting and auditing profession. In *Technologies, Artificial Intelligence and the Future of Learning Post-COVID-19: The Crucial Role of International Accreditation* (pp. 201-213). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-93921-2_12

Commerford, B. P., Dennis, S. A., Joe, J. R., & Ulla, J. W. (2022). Man versus machine: complex estimates and auditor reliance on artificial intelligence. *Journal of Accounting Research*, *60*(1), 171-201. https://doi.org/10.1111/1475-679X.12407

Instituto de Estudos de Saúde Suplementar. (2017). Impacto das fraudes e dos desperdícios sobre os gastos da saúde suplementar: Estimativa para 2017. Instituto de Estudos de Saúde Suplementar. https://www2.iess.org.br/cms/rep/analise_especial_fraudes_estimativa_2017.pdf

Nakao, Y., Strappelli, L., Stumpf, S., Naseer, A., Regoli, D., & Gamba, G. D. (2023). Towards responsible AI: A design space exploration of human-centered artificial intelligence user interfaces to investigate fairness. *International Journal of Human–Computer Interaction*, *39*(9), 1762–1788. https://doi.org/10.1080/10447318.2022.2067936

National Health Care Anti-Fraud Association. (n.d.). The challenge of health care fraud. https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud/

Organisation for Economic Co-operation and Development (OECD). (2023). Health at a glance 2023: OECD indicators. OECD Publishing. https://doi.org/10.1787/7a7afb35-en

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Xu, W., & Gao, Z. (2024, May). Enabling human-centered AI: A methodological perspective. In *2024 IEEE 4th International Conference on Human-Machine Systems* (ICHMS) (pp. 1–6). IEEE. https://doi.org/10.1109/ICHMS59971.2024.10555771

CONCLUSÃO GERAL

Esta tese apresentou uma abordagem integrada para o aprimoramento da detecção de fraudes em cooperativas médicas, estruturada em três componentes complementares: o artigo científico, o artigo tecnológico e o Produto Técnico Tecnológico (PTT). Em conjunto, esses trabalhos demonstram que modelos de Inteligência Artificial alinhados aos princípios da Inteligência Artificial Centrada no Ser Humano representam uma solução tecnicamente superior e institucionalmente mais adequada para o setor de saúde suplementar.

O artigo científico forneceu evidências empíricas ao comparar um método não supervisionado com um modelo supervisionado calibrado por especialistas. Os resultados revelaram que o modelo supervisionado alcançou maior precisão e menor ocorrência de falsos positivos, confirmando que a participação humana no processo de definição de padrões melhora substancialmente o desempenho preditivo dos algoritmos. A hipótese de que abordagens híbridas superam métodos autônomos foi confirmada.

O artigo tecnológico ampliou a compreensão do fenômeno ao mapear vulnerabilidades internas das cooperativas médicas e ao identificar fragilidades operacionais que favorecem ocorrências fraudulentas. O estudo evidenciou que a mitigação de riscos exige integração entre governança, processos formais e tecnologias inteligentes, permitindo construir diretrizes aplicáveis às particularidades do modelo cooperativista.

O Produto Técnico Tecnológico, representado pela plataforma GloomTrace®, incorporou de forma prática os achados dos dois artigos iniciais. A plataforma

apresenta módulos de triagem, investigação, gerenciamento e visualização estratégica de eventos suspeitos. Sua estrutura foi concebida segundo os princípios da IA centrada no ser humano, garantindo supervisão, rastreabilidade, explicabilidade e participação ativa de especialistas em todas as etapas do processo de análise. O produto demonstra a viabilidade de uma solução concreta e escalável para o controle de fraudes no setor.

A integração dos três trabalhos permite afirmar que a detecção eficaz de fraudes na saúde suplementar depende de modelos híbridos. Os algoritmos oferecem velocidade e sensibilidade para identificar padrões atípicos, enquanto especialistas fornecem interpretação contextual e capacidade de julgamento, elementos essenciais para garantir precisão e legitimidade institucional. A IA isolada não é suficiente para compreender nuances assistenciais e comportamentais, por isso abordagens centradas no ser humano apresentam resultados superiores em termos de eficácia e aceitação organizacional.

Entre as contribuições da tese destacam-se a demonstração empírica da superioridade de modelos supervisionados calibrados por especialistas, o diagnóstico estruturado das vulnerabilidades gerenciais que favorecem fraudes e o desenvolvimento de um produto tecnológico que materializa os princípios estudados em um sistema aplicável e alinhado às necessidades reais do setor.

Como limitações, ressalta-se que o estudo se baseou em dados de uma única cooperativa médica, o que pode limitar a generalização dos achados. Além disso, a pesquisa concentrou-se na triagem de eventos suspeitos, não na apuração conclusiva das fraudes. Recomenda-se que estudos futuros ampliem o escopo institucional,

incorporem dados clínicos mais detalhados e investiguem técnicas de explicabilidade que aumentem a transparência e a confiança no uso de IA.

Em síntese, esta tese evidencia que a Inteligência Artificial Centrada no Ser Humano oferece uma estratégia ética, eficiente e sustentável para fortalecer os mecanismos de auditoria e controle na saúde suplementar. A aplicação integrada de tecnologia, governança e supervisão humana contribui para sistemas de saúde mais confiáveis, transparentes e resilientes.