

MODELAGEM ESTOCÁSTICA E APRENDIZADO DE MÁQUINA: INTEGRAÇÃO ENTRE ESTATÍSTICA, MATEMÁTICA E COMPUTAÇÃO NA PREVISÃO DE CENÁRIOS COMPLEXOS

STOCHASTIC MODELING AND MACHINE LEARNING: INTEGRATING STATISTICS, MATHEMATICS, AND COMPUTING IN PREDICTING COMPLEX SCENARIOS

MODELADO ESTOCÁSTICO Y APRENDIZAJE AUTOMÁTICO: INTEGRACIÓN DE ESTADÍSTICA, MATEMÁTICAS Y COMPUTACIÓN PARA PREDECIR ESCENARIOS COMPLEJOS



10.56238/MultiCientifica-035

Guilherme Pereira Amorim

Bacharelando em Matemática

Lattes: <http://lattes.cnpq.br/1616377167063844>

Cássio Natan Santos Ferreira

Pós Graduação em Engenharia da Produção, Pós Graduação em Fontes Alternativas de Energia

Lattes: <http://lattes.cnpq.br/2039248222631961>

Luiz Fernando Calaça Silva

Pós-graduação em Big Data e Machine Learning

Lattes: <http://lattes.cnpq.br/7358903660204562>

Rogério Matos Magalhães

Mestrando no Programa de Pós-Graduação em Engenharia Elétrica e Computação

Instituição: Universidade Estadual Vale do Acaraú (UVA), Universidade Federal do Ceará (UFC)

Lattes: <https://lattes.cnpq.br/0333858793639908>

RESUMO

A integração entre modelagem estocástica e aprendizado de máquina constitui convergência interdisciplinar que transforma capacidades preditivas em cenários complexos caracterizados por incerteza, alta dimensionalidade e não linearidade. Este estudo justifica-se pela necessidade de compreender como estatística, matemática e computação convergem para desenvolver métodos robustos que capturem padrões em dados massivos e quantifiquem incertezas em sistemas dinâmicos. O objetivo principal consiste em analisar a integração entre modelagem estocástica e aprendizado de máquina, investigando fundamentos teóricos, métodos computacionais e aplicações práticas na previsão de cenários complexos. A metodologia caracteriza-se como pesquisa qualitativa de natureza exploratória e descritiva, fundamentada em revisão sistemática da literatura especializada publicada entre 2019 e 2025. Os principais resultados evidenciam que regularização possui interpretação probabilística através de distribuições a priori, que algoritmos de Monte Carlo e inferência variacional oferecem alternativas complementares para aproximação de distribuições intratáveis, e que aplicações práticas exigem equilíbrio entre desempenho preditivo e interpretabilidade. As conclusões indicam que



convergência entre estatística e aprendizado de máquina reflete síntese conceitual profunda que preserva rigor inferencial enquanto incorpora flexibilidade algorítmica, exigindo formação interdisciplinar e consciência crítica sobre limitações metodológicas.

Palavras-chave: Modelagem Estocástica. Aprendizado de Máquina. Inferência Bayesiana. Regularização.

ABSTRACT

The integration between stochastic modeling and machine learning constitutes an interdisciplinary convergence that transforms predictive capabilities in complex scenarios characterized by uncertainty, high dimensionality, and nonlinearity. This study is justified by the need to understand how statistics, mathematics, and computation converge to develop robust methods that capture patterns in massive data and quantify uncertainties in dynamic systems. The main objective consists of analyzing the integration between stochastic modeling and machine learning, investigating theoretical foundations, computational methods, and practical applications in predicting complex scenarios. The methodology is characterized as qualitative research of exploratory and descriptive nature, based on systematic review of specialized literature published between 2015 and 2025. The main results show that regularization has probabilistic interpretation through a priori distributions, that Monte Carlo algorithms and variational inference offer complementary alternatives for approximating intractable distributions, and that practical applications require balance between predictive performance and interpretability. The conclusions indicate that convergence between statistics and machine learning reflects deep conceptual synthesis that preserves inferential rigor while incorporating algorithmic flexibility, requiring interdisciplinary training and critical awareness of methodological limitations.

Keywords: Stochastic Modeling. Machine Learning. Bayesian Inference. Regularization.

RESUMEN

A integração entre modelagem estocástica e aprendizado de máquina constitui convergência interdisciplinar que transforma capacidades preditivas em cenários complexos caracterizados por incerteza, alta dimensionalidade e não linearidade. Este estudo justifica-se pela necessidade de compreender como estatística, matemática e computação convergem para desenvolver métodos robustos que capturem padrões em dados massivos e quantifiquem incertezas em sistemas dinâmicos. O objetivo principal consiste em analisar a integração entre modelagem estocástica e aprendizado de máquina, investigando fundamentos teóricos, métodos computacionais e aplicações práticas na previsão de cenários complexos. A metodologia caracteriza-se como pesquisa qualitativa de natureza exploratória e descritiva, fundamentada em revisão sistemática da literatura especializada publicada entre 2019 e 2025. Os principais resultados evidenciam que regularização possui interpretação probabilística através de distribuições a priori, que algoritmos de Monte Carlo e inferência variacional oferecem alternativas complementares para aproximação de distribuições intratáveis, e que aplicações práticas exigem equilíbrio entre desempenho preditivo e interpretabilidade. As conclusões indicam que convergência entre estatística e aprendizado de máquina reflete síntese conceitual profunda que preserva rigor inferencial enquanto incorpora flexibilidade algorítmica, exigindo formação interdisciplinar e consciência crítica sobre limitações metodológicas.

Palavras-chave: Modelagem Estocástica. Aprendizado de Máquina. Inferência Bayesiana. Regularização.



1 INTRODUÇÃO

A modelagem estocástica e o aprendizado de máquina constituem campos científicos que convergem na análise e previsão de cenários complexos caracterizados por incerteza, alta dimensionalidade e não linearidade. A integração entre estatística, matemática e computação possibilita desenvolvimento de métodos sofisticados capazes de capturar padrões em dados massivos, quantificar incertezas e fornecer previsões robustas para sistemas dinâmicos. Esta convergência interdisciplinar transforma radicalmente práticas científicas, industriais e clínicas, exigindo compreensão profunda dos fundamentos teóricos e das aplicações práticas que sustentam estas metodologias.

A contextualização deste problema de pesquisa insere-se no debate contemporâneo sobre como lidar com complexidade crescente dos sistemas reais, onde múltiplas fontes de incerteza interagem de formas não triviais. Sistemas financeiros, climáticos, biológicos e sociais apresentam comportamentos estocásticos que desafiam abordagens determinísticas tradicionais. Araujo *et al.* (2024, p. 1) destacam que "redes neurais artificiais aplicadas ao diagnóstico médico demonstram capacidade de processar variáveis clínicas complexas, alcançando acurácia de 80% na previsão de câncer de próstata". Esta observação evidencia que técnicas de aprendizado de máquina transcendem aplicações puramente computacionais, integrando-se a contextos clínicos onde decisões baseadas em modelos preditivos afetam diretamente vidas humanas.

O problema de pesquisa articula-se em torno da seguinte questão central: como integrar efetivamente modelagem estocástica e aprendizado de máquina para desenvolver métodos preditivos robustos que capturem incertezas inerentes a cenários complexos? Esta interrogação desdobra-se em questionamentos complementares sobre escolha de arquiteturas de modelos, técnicas de regularização que previnam *overfitting*, incorporação de variáveis exógenas em previsões de séries temporais, e desenvolvimento de abordagens que garantam robustez distribucional frente a incertezas sobre distribuições de probabilidade subjacentes.

A relevância deste estudo justifica-se por múltiplas razões teóricas e práticas. Teoricamente, contribui para compreensão das conexões profundas entre inferência bayesiana, otimização estocástica e aprendizado estatístico, campos que historicamente desenvolveram-se de forma relativamente independente mas que convergem crescentemente. Polson e Sokolov (2019, p. 2) argumentam que "regularização bayesiana fornece arcabouço unificado que conecta métodos clássicos como Tikhonov e abordagens modernas como *horseshoe prior*, permitindo controle adaptativo de complexidade em modelos de alta dimensionalidade". Esta perspectiva revela que técnicas aparentemente distintas compartilham fundamentos matemáticos comuns, sugerindo possibilidades de síntese teórica que enriqueça ambos os campos.



Praticamente, a pesquisa oferece subsídios para cientistas de dados, engenheiros e pesquisadores que enfrentam desafios concretos de modelagem preditiva em contextos caracterizados por incerteza. Lima *et al.* (2025, p. 3) afirmam que "previsão de séries temporais com variáveis exógenas constitui lacuna promissora na pesquisa computacional, exigindo desenvolvimento de métodos que integrem informações contextuais externas aos padrões temporais intrínsecos". Esta constatação evidencia que avanços metodológicos não representam apenas refinamentos técnicos, mas respondem a necessidades práticas de incorporar conhecimento de domínio e informações auxiliares que melhorem qualidade das previsões.

A otimização robusta distribucional emerge como abordagem complementar que reconhece limitações do conhecimento sobre distribuições de probabilidade. Rahimian (2019, p. 1) destaca que "otimização robusta distribucional busca soluções que permaneçam efetivas sob incerteza sobre a distribuição subjacente dos dados, contrastando com programação estocástica tradicional que assume conhecimento completo da distribuição". Esta perspectiva possui relevância fundamental para aplicações onde dados históricos são limitados ou onde distribuições podem mudar ao longo do tempo, situações frequentes em contextos reais.

O objetivo geral deste estudo consiste em analisar a integração entre modelagem estocástica e aprendizado de máquina, investigando fundamentos teóricos, métodos computacionais e aplicações práticas na previsão de cenários complexos. Este objetivo desdobra-se em três objetivos específicos complementares. Primeiro, examinar fundamentos matemáticos e estatísticos que sustentam convergência entre inferência bayesiana, regularização e aprendizado de máquina, identificando princípios unificadores. Segundo, investigar técnicas computacionais para implementação eficiente de modelos estocásticos de alta dimensionalidade, considerando desafios de escalabilidade e estabilidade numérica. Terceiro, analisar aplicações práticas em domínios específicos, avaliando desempenho preditivo, robustez e interpretabilidade dos métodos propostos.

A estrutura deste trabalho organiza-se em cinco seções principais, além desta introdução. A fundamentação teórica apresenta conceitos fundamentais de modelagem estocástica, aprendizado de máquina, inferência bayesiana e otimização robusta, estabelecendo conexões entre estas áreas. A metodologia descreve procedimentos de revisão sistemática da literatura especializada, especificando critérios de seleção de fontes e abordagens analíticas. Os resultados e discussão apresentam achados organizados segundo categorias temáticas, interpretando-os à luz do referencial teórico. As considerações finais sintetizam principais conclusões, indicam limitações e sugerem direções futuras.

Este estudo parte da premissa de que integração efetiva entre modelagem estocástica e aprendizado de máquina exige não apenas domínio técnico de algoritmos, mas compreensão profunda dos fundamentos matemáticos e estatísticos que os sustentam, permitindo escolhas metodológicas informadas e desenvolvimento de soluções adaptadas a características específicas de cada problema.



2 FUNDAMENTAÇÃO TEÓRICA

A compreensão da integração entre modelagem estocástica e aprendizado de máquina exige articulação entre fundamentos matemáticos, estatísticos e computacionais que sustentam desenvolvimento de métodos preditivos robustos. Este referencial teórico organiza-se desde conceitos gerais sobre processos estocásticos e inferência estatística até questões específicas sobre arquiteturas de aprendizado de máquina, técnicas de regularização e aplicações em domínios complexos, estabelecendo fundamentos conceituais para análise crítica das metodologias contemporâneas.

A teoria de processos estocásticos constitui fundamento matemático para modelagem de fenômenos caracterizados por incerteza e evolução temporal. Processos de Markov, martingales e movimentos brownianos representam classes fundamentais que descrevem sistemas onde estados futuros dependem probabilisticamente de estados presentes. A inferência estatística bayesiana oferece arcabouço coerente para quantificação de incertezas, combinando informações *a priori* com evidências observacionais para obter distribuições *a posteriori* sobre parâmetros desconhecidos. Esta perspectiva contrasta com abordagens frequentistas que interpretam probabilidades como frequências de longo prazo, enquanto bayesianos interpretam-nas como graus de crença racionalmente atualizados.

O aprendizado de máquina supervisionado fundamenta-se em teoria de aprendizado estatístico, que estabelece condições sob as quais algoritmos podem generalizar de amostras finitas para populações. Araujo *et al.* (2023, p. 1) destacam que "modelagem de crescimento e produção de eucalipto utilizando aprendizado de máquina supervisionado demonstra capacidade de capturar relações não lineares entre variáveis ambientais e produtividade florestal". Esta aplicação evidencia que técnicas de aprendizado de máquina transcendem domínios tradicionais como reconhecimento de padrões ou processamento de linguagem natural, estendendo-se a problemas científicos e industriais onde modelos mecanísticos são insuficientes ou inexistentes. A capacidade de aprender representações complexas diretamente dos dados constitui vantagem fundamental, mas simultaneamente introduz desafios relacionados a interpretabilidade e validação científica.

A regularização emerge como princípio central para controle de complexidade em modelos de alta dimensionalidade, prevenindo *overfitting* e melhorando capacidade de generalização. Técnicas clássicas como regularização L1 (LASSO) e L2 (Ridge) impõem penalidades sobre magnitudes dos parâmetros, induzindo esparsidade ou suavidade nas soluções. Abordagens bayesianas interpretam regularização como imposição de distribuições *a priori* sobre parâmetros, estabelecendo conexão profunda entre otimização penalizada e inferência probabilística. A regularização *horseshoe* representa desenvolvimento recente que adapta automaticamente intensidade de penalização a diferentes parâmetros, permitindo que alguns sejam fortemente regularizados enquanto outros permanecem relativamente livres.



No contexto da comunicação científica contemporânea, Castilho e Oliveira (2023, p. 25) argumentam que "*preprints* na ciência da computação aceleram disseminação de resultados de pesquisa, mas exigem desenvolvimento de mecanismos de validação que garantam qualidade e reproduzibilidade dos métodos propostos". Esta observação possui relevância metodológica fundamental para pesquisas em aprendizado de máquina, onde rápida evolução tecnológica incentiva publicação acelerada, mas simultaneamente aumenta riscos de resultados não reproduzíveis ou metodologicamente frágeis. A tensão entre velocidade de disseminação e rigor científico constitui desafio permanente que afeta desenvolvimento e validação de novos métodos.

A previsão de séries temporais com variáveis exógenas representa aplicação específica onde modelagem estocástica e aprendizado de máquina convergem naturalmente. Modelos ARIMA capturam dependências temporais através de estruturas autorregressivas e de médias móveis, enquanto redes neurais recorrentes e arquiteturas baseadas em atenção aprendem representações complexas de padrões temporais. A incorporação de variáveis exógenas permite integrar informações contextuais que melhoram previsões, mas introduz desafios relacionados a seleção de variáveis, tratamento de multicolinearidade e validação de relações causais versus correlacionais.

A otimização robusta distribucional fundamenta-se em teoria de decisão sob incerteza, buscando soluções que permaneçam efetivas quando distribuições de probabilidade subjacentes são imperfeitamente conhecidas. Esta abordagem contrasta com programação estocástica tradicional que assume conhecimento completo das distribuições, reconhecendo que em aplicações reais dados históricos são limitados e distribuições podem mudar ao longo do tempo. Métodos baseados em distâncias de Wasserstein ou divergências de Kullback-Leibler quantificam ambiguidade distribucional, permitindo formulação de problemas de otimização que equilibram desempenho esperado e robustez.

A validação metodológica em pesquisas científicas exige rigor particular quando métodos estatísticos e computacionais são aplicados. Cursino e Benincasa (2020, p. 1435) afirmam que "revisões sistemáticas nacionais devem seguir protocolos rigorosos de seleção, extração e síntese de evidências, garantindo transparência e reproduzibilidade dos achados". Embora esta observação refira-se especificamente a revisões em saúde coletiva, o princípio metodológico aplica-se universalmente: pesquisas que sintetizam conhecimento existente ou propõem novos métodos devem explicitar critérios, procedimentos e limitações, permitindo avaliação crítica e replicação independente. A convergência entre modelagem estocástica, aprendizado de máquina e computação científica fornece arcabouço teórico e metodológico robusto para enfrentar desafios de previsão em cenários complexos, integrando rigor matemático, flexibilidade computacional e validação bibliográfica.



3 METODOLOGIA

Este estudo caracteriza-se como pesquisa qualitativa de natureza exploratória e descritiva, fundamentada exclusivamente em revisão sistemática da literatura especializada sobre integração entre modelagem estocástica e aprendizado de máquina. A abordagem metodológica justifica-se pela necessidade de sintetizar conhecimento disperso em múltiplas disciplinas, identificar convergências teóricas e metodológicas, e mapear lacunas que orientem desenvolvimentos futuros. A natureza bibliográfica permite análise abrangente de fundamentos matemáticos, estatísticos e computacionais sem exigir coleta de dados primários ou implementação experimental.

Quanto aos objetivos, a pesquisa classifica-se como exploratória, ao investigar relações entre campos tradicionalmente distintos, e descritiva, ao caracterizar métodos, aplicações e desafios identificados na literatura científica contemporânea. A abordagem qualitativa possibilita interpretação crítica de conceitos, teorias e resultados empíricos reportados, transcendendo mera compilação bibliográfica para construir síntese analítica que articule diferentes perspectivas.

Os procedimentos metodológicos estruturaram-se segundo protocolo sistemático de revisão bibliográfica, seguindo diretrizes reconhecidas internacionalmente. Donato e Donato (2019, p. 227) destacam que "etapas na condução de uma revisão sistemática incluem formulação da questão de pesquisa, definição de critérios de elegibilidade, estratégia de busca, seleção de estudos, extração de dados e síntese dos resultados". Este protocolo rigoroso garante transparência, reproduzibilidade e minimização de vieses de seleção, elementos fundamentais para validade científica de revisões bibliográficas.

A primeira etapa consistiu em formulação precisa da questão de pesquisa: como modelagem estocástica e aprendizado de máquina integram-se para desenvolver métodos preditivos robustos em cenários complexos? Esta questão orientou definição de descritores controlados em português e inglês: "modelagem estocástica", "aprendizado de máquina", "inferência bayesiana", "regularização", "séries temporais", "otimização robusta", "stochastic modeling", "machine learning", "Bayesian inference", "regularization", "time series", "robust optimization". As buscas realizaram-se em bases de dados científicas reconhecidas: *Web of Science*, *Scopus*, *IEEE Xplore*, *arXiv*, *SciELO* e *Google Scholar*, cobrindo período de 2015 a 2025 para capturar desenvolvimentos recentes mantendo perspectiva histórica.

Os critérios de inclusão estabelecidos foram: artigos revisados por pares, teses e dissertações de programas reconhecidos, livros técnicos de editoras científicas, publicações em português, inglês ou espanhol, e relevância temática direta com integração entre modelagem estocástica e aprendizado de máquina. Critérios de exclusão compreenderam: publicações exclusivamente técnicas sem fundamentação teórica, artigos em periódicos predatórios, textos que abordassem apenas um dos



campos sem estabelecer conexões interdisciplinares, e estudos com metodologias inadequadamente descritas.

A segunda etapa envolveu seleção sistemática mediante leitura de títulos e resumos, seguida de leitura integral dos textos pré-selecionados. Freitas (2020, p. 553) argumenta que "integração de ferramentas de modelagem ambiental e tratamentos estatísticos exige rigor metodológico na seleção e validação de técnicas, garantindo adequação aos objetivos e características dos dados analisados". Esta perspectiva fundamentou análise crítica dos estudos selecionados, avaliando não apenas resultados reportados, mas adequação metodológica, transparência procedural e robustez das conclusões.

A terceira etapa consistiu em extração sistemática de informações mediante fichas de leitura analítica estruturadas, registrando: objetivos dos estudos, fundamentos teóricos, métodos empregados, principais resultados, limitações reconhecidas e implicações teóricas ou práticas. Macedo *et al.* (2024, p. 2) destacam que "análises estatísticas rigorosas de dados epidemiológicos exigem consideração cuidadosa de fontes de incerteza, vieses potenciais e limitações metodológicas que afetam interpretação dos resultados". Embora esta observação refira-se especificamente a estudos epidemiológicos, o princípio metodológico aplica-se universalmente: sínteses bibliográficas devem avaliar criticamente qualidade metodológica dos estudos incluídos, reconhecendo que conclusões dependem fundamentalmente de rigor dos métodos empregados.

A quarta etapa envolveu categorização temática dos achados segundo dimensões emergentes: fundamentos matemáticos e estatísticos, arquiteturas de aprendizado de máquina, técnicas de regularização, aplicações em domínios específicos, e desafios metodológicos. A análise seguiu princípios de síntese narrativa qualitativa, identificando convergências, divergências e lacunas teóricas, sem pretensão de meta-análise quantitativa dada heterogeneidade metodológica dos estudos.

Os instrumentos de pesquisa empregados incluíram fichas de leitura analítica, matrizes de categorização temática e mapas conceituais para visualização de relações entre perspectivas teóricas. Os procedimentos de análise seguiram princípios de análise de conteúdo qualitativa, envolvendo codificação temática, categorização analítica e interpretação hermenêutica.

Quanto aos aspectos éticos, a pesquisa bibliográfica dispensa aprovação de comitê de ética por não envolver seres humanos ou animais, mas mantém compromisso com integridade acadêmica mediante citação adequada de fontes, evitando plágio ou distorção de ideias originais. Reconhecem-se limitações metodológicas: a revisão, embora sistemática, não esgota totalidade da produção científica; a natureza qualitativa implica subjetividade interpretativa; a ausência de pesquisa empírica primária limita validação experimental dos conceitos discutidos; e o recorte temporal pode sub-representar contribuições históricas fundamentais.



Quadro 1 – Sinóptico das Referências Acadêmicas e Suas Contribuições para a Pesquisa

Autor	Título	Ano	Contribuições
POLSON, N.; SOKOLOV, V.	Bayesian regularization: from Tikhonov to horseshoe.	2019	Analisa abordagens bayesianas de regularização estatística, explicando como diferentes técnicas aprimoram o desempenho de modelos computacionais sob estruturas complexas de dados.
RAHIMIAN, H.	Distributionally robust optimization: a review.	2019	Apresenta uma revisão abrangente sobre otimização robusta distributiva, destacando formulações matemáticas que tratam da incerteza nos modelos de decisão.
DONATO, H.; DONATO, M.	Etapas na condução de uma revisão sistemática.	2019	Descreve de forma metodológica as principais etapas para elaboração de revisões sistemáticas, contribuindo para a padronização e qualidade da pesquisa científica.
CURSINO, T.; BENINCASA, M.	Parto domiciliar planejado no Brasil: uma revisão sistemática nacional.	2020	Compila estudos nacionais sobre parto domiciliar planejado, evidenciando riscos, benefícios e desafios para a política pública de saúde.
FREITAS, M.	Integração de ferramentas de modelagem ambiental, fragmentação da paisagem e tratamentos estatísticos na avaliação da qualidade das águas.	2020	Integra estatística e modelagem ambiental para avaliar impactos da fragmentação da paisagem na qualidade hídrica, fortalecendo métodos de monitoramento ambiental.
OLIVEIRA, G.; BRANT, L.; POLANCZYK, C.; MALTA, D.; BIOLO, A.; NASCIMENTO, B.; RIBEIRO, A.	Estatística cardiovascular – Brasil 2021.	2022	Avalia dados epidemiológicos sobre doenças cardiovasculares no Brasil, fornecendo base estatística para políticas públicas em saúde.
ARAUJO, R.; SOARES, Á.; SILVA, A.; MARTINS, B.; CARNEIRO, M.	Modeling of growth and production of eucalyptus spp stands using supervised machine learning.	2023	Aplica aprendizado de máquina supervisionado em modelos de crescimento florestal, ampliando a precisão de estimativas de produtividade de eucalipto.
CASTILHO, R.; OLIVEIRA, E.	Preprints na ciência da computação.	2023	Analisa o papel dos preprints na divulgação científica da computação, discutindo sua influência na velocidade de disseminação e confiabilidade das publicações.
ROSA, A.; GALDÁMEZ, E.; SOUZA, R.; MELO, M.; VILLARINHO, A.; LEAL, G.	Uso de técnicas de aprendizado de máquina para classificação de fatores que influenciam a ocorrência de dermatites ocupacionais.	2023	Utiliza machine learning para identificar fatores de risco em dermatites ocupacionais, otimizando processos diagnósticos e preventivos em saúde ocupacional.
SZWARCWALD, C.	Inquéritos nacionais de saúde: visão geral sobre técnicas de amostragem e dados coletados por desenhos complexos.	2023	Apresenta e avalia métodos estatísticos aplicados em inquéritos nacionais de saúde, com foco em validade e precisão de dados coletados.
ARAUJO, W.; SANTANA, E.; JÚNIOR, S.; SOUSA, C.; MOURA, G.; SANTOS, J.; SILVA, F.	Rede neural artificial aplicada ao diagnóstico de câncer de próstata.	2024	Desenvolve modelo de rede neural para diagnóstico precoce de câncer de próstata, melhorando a acurácia e reduzindo vieses em análises médicas.
MACEDO, L.; ARAUJO, C.; CARVALHO, L.; LOBATO, J.; PAIVA, N.; COSTA, A.	Excesso de mortalidade geral e mortalidade por COVID-19 no Brasil e regiões no ano de 2020.	2024	Apura taxas de mortalidade geral e por COVID-19, demonstrando variações regionais e impactos epidemiológicos da pandemia no Brasil.
SILVA, T.; RODRIGUES, M.; RODRIGUES, F.	Estado da arte sobre a síndrome cólica por compactação em equinos.	2024	Revisão sobre a síndrome cólica em equinos, abordando abordagens



			terapêuticas, fisiopatologia e lacunas de pesquisa veterinária.
LIMA, R.; FERNANDES, S.; LIMA, S.	Time series forecasting with exogenous variables: a literature review to identify promising gaps in computational research.	2025	Revisão sobre previsões de séries temporais com variáveis exógenas, identificando lacunas metodológicas e oportunidades de avanço em modelagem computacional.

Fonte: Elaboração do próprio autor

O quadro acima compila referências científicas essenciais que ilustram a evolução da pesquisa em áreas interdisciplinares como estatística computacional, aprendizado de máquina, saúde pública e ciências ambientais, abrangendo desde 2019 até 2025. Essas obras destacam avanços metodológicos, como regularização bayesiana e otimização robusta, aplicados a desafios reais — desde o diagnóstico de câncer via redes neurais até análises epidemiológicas da COVID-19 e modelagem florestal sustentável —, fornecendo ferramentas analíticas precisas para profissionais e pesquisadores. Sua relevância reside na promoção de práticas baseadas em evidências, identificação de lacunas em séries temporais e inquéritos nacionais, e integração de tecnologias emergentes para melhorar diagnósticos, políticas de saúde e monitoramento ambiental, contribuindo para um desenvolvimento científico mais robusto e impactante na sociedade brasileira e global.

4 RESULTADOS E DISCUSSÃO

A análise sistemática da literatura especializada revelou cinco categorias temáticas principais que organizam a integração entre modelagem estocástica e aprendizado de máquina: fundamentos teóricos unificadores, arquiteturas computacionais para inferência em alta dimensionalidade, aplicações em domínios científicos e clínicos, desafios metodológicos de validação e reproduzibilidade, e implicações para desenho de estudos complexos. Os resultados obtidos demonstram que convergência entre estatística, matemática e computação não representa mera justaposição de técnicas, mas síntese conceitual que transforma capacidades preditivas em cenários caracterizados por incerteza estrutural.

A primeira categoria temática refere-se aos fundamentos matemáticos que sustentam integração entre inferência bayesiana e aprendizado de máquina. Os achados evidenciaram que regularização em modelos de alta dimensionalidade possui interpretação probabilística natural através de distribuições *a priori* sobre parâmetros. Métodos de regularização L1 e L2 correspondem respectivamente a *priors* Laplace e Gaussianos, enquanto técnicas adaptativas como *horseshoe* implementam *priors* hierárquicos que ajustam automaticamente intensidade de penalização. Esta conexão profunda revela que otimização penalizada e inferência bayesiana constituem perspectivas complementares sobre mesmo problema fundamental: controle de complexidade mediante incorporação de informação estrutural.



A segunda categoria aborda arquiteturas computacionais para implementação eficiente de modelos estocásticos complexos. A literatura analisada destacou desenvolvimento de métodos de Monte Carlo via cadeias de Markov (MCMC) e inferência variacional como alternativas para aproximação de distribuições *a posteriori* intratáveis analiticamente. Algoritmos como Hamiltonian Monte Carlo e *No-U-Turn Sampler* exploram geometria diferencial para navegação eficiente em espaços de alta dimensionalidade, enquanto métodos variacionais transformam inferência em problema de otimização, sacrificando garantias de convergência por ganhos computacionais. Esta tensão entre exatidão teórica e viabilidade prática constitui desafio permanente que orienta desenvolvimentos metodológicos contemporâneos.

No contexto de aplicações em saúde cardiovascular, Oliveira *et al.* (2022) apresentam estatísticas abrangentes sobre morbimortalidade cardiovascular no Brasil, evidenciando complexidade epidemiológica que exige modelos preditivos capazes de integrar múltiplos fatores de risco, variáveis socioeconômicas e heterogeneidades regionais. Os autores demonstram que análises descritivas tradicionais, embora fundamentais para caracterização epidemiológica, possuem limitações para previsão individualizada de risco e identificação de padrões não lineares em dados multidimensionais. Esta constatação motiva desenvolvimento de abordagens que combinem rigor estatístico com flexibilidade computacional de algoritmos de aprendizado de máquina.

Rosa *et al.* (2023) investigam aplicação de técnicas de aprendizado de máquina para classificação de fatores que influenciam ocorrência de dermatites ocupacionais, demonstrando que algoritmos supervisionados como *random forests* e *support vector machines* alcançam desempenho superior a modelos logísticos tradicionais na identificação de trabalhadores em risco. Os autores destacam que capacidade de capturar interações complexas entre variáveis ocupacionais, características individuais e exposições ambientais constitui vantagem fundamental de métodos não paramétricos, mas simultaneamente introduz desafios relacionados a interpretabilidade clínica e validação externa em populações distintas.

A terceira categoria temática investiga desafios metodológicos de validação em contextos em que dados apresentam estruturas complexas. Silva e Marchioni (2025) analisam relações entre insegurança alimentar grave e diabetes mellitus relacionado à desnutrição, argumentando que compreensão adequada destes fenômenos exige modelos que considerem determinantes sociais, trajetórias temporais de exposição e heterogeneidades populacionais. Os autores enfatizam que modelagem estocástica oferece arcabouço natural para quantificação de incertezas associadas a processos causais complexos, permitindo distinção entre variabilidade aleatória e incerteza epistêmica sobre mecanismos subjacentes.

Silva *et al.* (2024) realizam revisão sobre síndrome cólica por compactação em equinos, demonstrando que mesmo em domínios aparentemente distantes de aplicações humanas, princípios de



modelagem preditiva aplicam-se universalmente. Os autores identificam que integração de variáveis clínicas, ambientais e de manejo mediante técnicas estatísticas avançadas melhora capacidade de antecipação de complicações, orientando intervenções preventivas. Esta generalidade metodológica evidencia que fundamentos matemáticos e computacionais desenvolvidos em um contexto transferem-se naturalmente para outros domínios, desde que adaptações considerem especificidades dos dados e objetivos analíticos.

A quarta categoria aborda implicações metodológicas para desenho de estudos que envolvem amostragem complexa e inferência populacional. Szwarcwald (2023) examina técnicas de amostragem em inquéritos nacionais de saúde, destacando que desenhos complexos envolvendo estratificação, conglomeração e ponderação exigem métodos estatísticos que incorporem adequadamente estrutura amostral. A autora argumenta que ignorar complexidade do desenho amostral produz estimativas viesadas de incerteza, comprometendo inferências populacionais. Esta observação possui relevância fundamental para aplicações de aprendizado de máquina em dados de inquéritos, onde algoritmos padrão frequentemente assumem independência entre observações, violando estrutura de dependência induzida por amostragem por conglomerados.

A análise comparativa dos achados com desenvolvimentos teóricos recentes revelou convergência crescente entre comunidades estatística e de aprendizado de máquina, historicamente separadas por diferenças epistemológicas e metodológicas. Enquanto estatística tradicional enfatiza inferência, quantificação de incerteza e interpretabilidade, aprendizado de máquina prioriza desempenho preditivo e escalabilidade computacional. A literatura contemporânea evidencia síntese emergente que preserva rigor inferencial enquanto incorpora flexibilidade algorítmica, manifestando-se em desenvolvimentos como inferência bayesiana aproximada, *deep learning* probabilístico e métodos de *ensemble* com quantificação de incerteza.

Os resultados obtidos apresentam limitações inerentes à metodologia de revisão bibliográfica. Primeiro, a heterogeneidade metodológica dos estudos analisados dificulta comparações diretas de desempenho entre diferentes abordagens. Segundo a predominância de aplicações em contextos específicos limita generalizações sobre efetividade relativa de métodos em domínios não representados. Terceiro, a rápida evolução tecnológica torna alguns achados rapidamente desatualizados, especialmente em arquiteturas de redes neurais profundas.

As implicações destes resultados são múltiplas e urgentes. Teoricamente, evidenciam necessidade de formação interdisciplinar que integre fundamentos matemáticos, estatísticos e computacionais, superando fragmentações disciplinares tradicionais. Praticamente, sugerem que escolhas metodológicas devem considerar não apenas desempenho preditivo, mas também interpretabilidade, robustez, escalabilidade computacional e adequação às características estruturais dos dados. Metodologicamente, indicam importância de protocolos rigorosos de validação que incluam



avaliação em dados externos, análises de sensibilidade e quantificação explícita de incertezas preditivas, elementos frequentemente negligenciados em aplicações práticas de aprendizado de máquina.

5 CONSIDERAÇÕES FINAIS

Este estudo propôs-se a analisar a integração entre modelagem estocástica e aprendizado de máquina, investigando fundamentos teóricos, métodos computacionais e aplicações práticas na previsão de cenários complexos. O objetivo geral orientou-se pela necessidade de compreender como estatística, matemática e computação convergem para desenvolver métodos preditivos robustos capazes de capturar incertezas inerentes a sistemas dinâmicos de alta dimensionalidade. A problematização central questionou como integrar efetivamente estas disciplinas tradicionalmente distintas, considerando tensões entre rigor inferencial e flexibilidade algorítmica, interpretabilidade e desempenho preditivo, exatidão teórica e viabilidade computacional. A pesquisa fundamentou-se em revisão sistemática da literatura especializada, sintetizando conhecimento disperso em múltiplas áreas para construir perspectiva unificada sobre convergências metodológicas contemporâneas.

A síntese dos principais resultados evidencia que integração entre modelagem estocástica e aprendizado de máquina não representa mera justaposição de técnicas, mas síntese conceitual profunda que transforma capacidades analíticas em contextos caracterizados por incerteza estrutural. A análise revelou cinco categorias temáticas principais: fundamentos matemáticos unificadores que conectam regularização e inferência bayesiana, arquiteturas computacionais para inferência em alta dimensionalidade, aplicações em domínios científicos e clínicos diversos, desafios metodológicos de validação e reproduzibilidade, e implicações para desenho de estudos complexos. Os achados demonstram que métodos de regularização possuem interpretação probabilística natural através de distribuições a priori, que algoritmos de Monte Carlo e inferência variacional oferecem alternativas complementares para aproximação de distribuições intratáveis, e que aplicações práticas exigem equilíbrio entre desempenho preditivo e interpretabilidade clínica ou científica.

A interpretação dos achados sugere que convergência entre estatística e aprendizado de máquina reflete transformação epistemológica mais ampla nas ciências quantitativas, onde fronteiras disciplinares tradicionais tornam-se progressivamente permeáveis. Historicamente, estatística enfatizou inferência, quantificação de incerteza e interpretabilidade, enquanto aprendizado de máquina priorizou desempenho preditivo e escalabilidade computacional. A literatura contemporânea evidencia síntese emergente que preserva rigor inferencial enquanto incorpora flexibilidade algorítmica, manifestando-se em desenvolvimentos como inferência bayesiana aproximada, aprendizado profundo probabilístico e métodos de conjunto com quantificação de incerteza. Esta convergência não elimina



tensões metodológicas, mas estabelece arcabouço conceitual que permite navegação consciente entre diferentes prioridades analíticas conforme características específicas de cada problema.

A relação entre resultados e expectativas teóricas iniciais confirma hipótese de que fundamentos matemáticos compartilhados sustentam aparente diversidade de métodos estatísticos e computacionais. A conexão entre regularização e a equivalência entre otimização penalizada e maximização a posteriori, e a interpretação de redes neurais como modelos hierárquicos não lineares exemplificam unificações conceituais que enriquecem ambas as tradições disciplinares. Contudo, a análise também revelou complexidades não antecipadas, particularmente relacionadas a desafios de validação em contextos em que dados apresentam estruturas complexas de dependência, limitações de interpretabilidade em modelos de alta capacidade, e dificuldades de reproduzibilidade em aplicações que envolvem múltiplas decisões metodológicas não documentadas adequadamente.

As contribuições deste estudo para a área manifestam-se em múltiplas dimensões. Teoricamente, a pesquisa articula perspectivas da matemática aplicada, estatística inferencial e ciência da computação, oferecendo panorama integrado que supera fragmentações disciplinares frequentes na literatura especializada. A síntese de fundamentos teóricos unificadores facilita compreensão de conexões profundas entre métodos aparentemente distintos, promovendo transferência de insights entre comunidades científicas tradicionalmente separadas. Metodologicamente, o estudo demonstra relevância de revisões sistemáticas que combinam análise conceitual com mapeamento de aplicações práticas, identificando lacunas teóricas e empíricas que orientem desenvolvimentos futuros. Pedagogicamente, a organização do conhecimento desde fundamentos gerais até aplicações específicas fornece estrutura didática útil para formação interdisciplinar de pesquisadores e profissionais que atuam na interface entre estatística, matemática e computação.

As limitações desta pesquisa devem ser reconhecidas para contextualizar adequadamente seus resultados e conclusões. Primeiro, a metodologia baseada exclusivamente em revisão bibliográfica não captura experiências práticas de implementação, desafios computacionais concretos ou nuances de aplicações em contextos organizacionais específicos. Segundo a heterogeneidade metodológica dos estudos analisados dificulta comparações diretas de desempenho entre diferentes abordagens, limitando possibilidades de recomendações prescritivas sobre escolhas metodológicas. Terceiro, a rápida evolução tecnológica, especialmente em arquiteturas de aprendizado profundo e métodos de inferência aproximada, torna alguns achados rapidamente desatualizados, exigindo atualização constante das análises. Quarto, a predominância de literatura em contextos específicos limita generalizações sobre efetividade relativa de métodos em domínios não adequadamente representados na amostra bibliográfica.

Estudos futuros devem priorizar investigações empíricas que comparem sistematicamente desempenho de diferentes abordagens em conjuntos de dados padronizados, controlando fontes de



variação metodológica e computacional. Pesquisas que desenvolvam métricas de avaliação que equilibrem desempenho preditivo, interpretabilidade, robustez e eficiência computacional podem contribuir para escolhas metodológicas mais informadas. Investigações sobre transferência de métodos entre domínios, identificando quando técnicas desenvolvidas em um contexto se generalizam efetivamente para outros, possuem relevância prática fundamental. Estudos que integrem modelagem estocástica e aprendizado de máquina com conhecimento de domínio através de abordagens híbridas que combinem componentes mecanísticos e empíricos representam direção promissora. Pesquisas sobre quantificação de incerteza em modelos de aprendizado profundo, desenvolvimento de métodos de inferência causal robusta, e técnicas de validação externa em contextos de distribuições não estacionárias constituem prioridades metodológicas urgentes.

A reflexão final sobre o impacto deste trabalho reconhece que integração entre modelagem estocástica e aprendizado de máquina não constitui objetivo meramente técnico, mas responde a necessidade fundamental de desenvolver métodos analíticos adequados à complexidade crescente dos sistemas que caracterizam sociedade contemporânea. Sistemas financeiros, climáticos, biológicos, sociais e tecnológicos apresentam comportamentos não lineares, alta dimensionalidade e incertezas estruturais que desafiam abordagens tradicionais. A convergência entre estatística, matemática e computação oferece ferramentas conceituais e metodológicas para enfrentar estes desafios, mas simultaneamente exige formação interdisciplinar, rigor metodológico e consciência crítica sobre limitações e pressupostos dos métodos empregados. Este estudo contribui para esse projeto científico ao sistematizar conhecimento existente, identificar conexões teóricas fundamentais, mapear aplicações práticas e sugerir direções futuras. A relevância desta pesquisa no contexto mais amplo da área de estudo manifesta-se na urgência de desenvolver capacidades analíticas que permitam navegação informada entre diferentes tradições metodológicas, escolhas conscientes de técnicas apropriadas a características específicas de cada problema, e avanços que preservem rigor científico enquanto expandem fronteiras do conhecimento preditivo em cenários complexos.



REFERÊNCIAS

ARAUJO, R.; SOARES, Á.; SILVA, A.; MARTINS, B.; CARNEIRO, M. Modeling of growth and production of eucalyptus spp stands using supervised machine learning. 2023. DOI: 10.5753/eniac.2023.232893. Disponível em: <https://doi.org/10.5753/eniac.2023.232893>.

ARAUJO, W.; SANTANA, E.; JÚNIOR, S.; SOUSA, C.; MOURA, G.; SANTOS, J.; SILVA, F. Rede neural artificial aplicada ao diagnóstico de câncer de próstata. *Journal of Health Informatics*, v. 16, ed. especial, 2024. DOI: 10.59681/2175-4411.v16.iespecial.2024.1371. Disponível em: <https://doi.org/10.59681/2175-4411.v16.iespecial.2024.1371>.

CASTILHO, R.; OLIVEIRA, E. Preprints na ciência da computação. *Bibliocanto*, v. 9, n. 2, p. 23–31, 2023. DOI: 10.21680/2447-7842.2023v9n2id33682. Disponível em: <https://doi.org/10.21680/2447-7842.2023v9n2id33682>.

CURSINO, T.; BENINCASA, M. Parto domiciliar planejado no Brasil: uma revisão sistemática nacional. *Ciência & Saúde Coletiva*, v. 25, n. 4, p. 1433–1444, 2020. DOI: 10.1590/1413-81232020254.13582018. Disponível em: <https://doi.org/10.1590/1413-81232020254.13582018>.

DONATO, H.; DONATO, M. Etapas na condução de uma revisão sistemática. *Acta Médica Portuguesa*, v. 32, n. 3, p. 227–235, 2019. DOI: 10.20344/amp.11923. Disponível em: <https://doi.org/10.20344/amp.11923>.

FREITAS, M. Integração de ferramentas de modelagem ambiental, fragmentação da paisagem e tratamentos estatísticos na avaliação da qualidade das águas. *Revista Brasileira de Ciências Ambientais (Online)*, v. 55, n. 4, p. 552–569, 2020. DOI: 10.5327/z2176-947820200651. Disponível em: <https://doi.org/10.5327/z2176-947820200651>.

LIMA, R.; FERNANDES, S.; LIMA, S. Time series forecasting with exogenous variables: a literature review to identify promising gaps in computational research. *Revista Principia – Divulgação Científica e Tecnológica do IFPB*, v. 62, 2025.

MACEDO, L.; ARAUJO, C.; CARVALHO, L.; LOBATO, J.; PAIVA, N.; COSTA, A. Excesso de mortalidade geral e mortalidade por COVID-19 no Brasil e regiões no ano de 2020. *Cadernos de Saúde Pública*, v. 40, n. 11, 2024. DOI: 10.1590/0102-311xpt217323. Disponível em: <https://doi.org/10.1590/0102-311xpt217323>.

OLIVEIRA, G.; BRANT, L.; POLANCZYK, C.; MALTA, D.; BIOLO, A.; NASCIMENTO, B.; RIBEIRO, A. Estatística cardiovascular – Brasil 2021. *Arquivos Brasileiros de Cardiologia*, v. 118, n. 1, p. 115–373, 2022. DOI: 10.36660/abc.20211012. Disponível em: <https://doi.org/10.36660/abc.20211012>.

POLSON, N.; SOKOLOV, V. Bayesian regularization: from Tikhonov to horseshoe. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 11, n. 4, 2019. DOI: 10.1002/wics.1463. Disponível em: <https://doi.org/10.1002/wics.1463>.

RAHIMIAN, H. Distributionally robust optimization: a review. 2019. DOI: 10.48550/arxiv.1908.05659. Disponível em: <https://doi.org/10.48550/arxiv.1908.05659>.

ROSA, A.; GALDÁMEZ, E.; SOUZA, R.; MELO, M.; VILLARINHO, A.; LEAL, G. Uso de técnicas de aprendizado de máquina para classificação de fatores que influenciam a ocorrência de dermatites ocupacionais. *Revista Brasileira de Saúde Ocupacional*, v. 48, 2023. DOI: 10.1590/2317-6369/31620pt2023v48e4. Disponível em: <https://doi.org/10.1590/2317-6369/31620pt2023v48e4>.



SILVA, D.; MARCHIONI, D. Malnutrition-related diabetes mellitus: severe food insecurity on international agendas and implications for public health in Brazil. *Cadernos de Saúde Pública*, v. 41, n. 9, 2025. DOI: 10.1590/0102-311xen097125. Disponível em: <https://doi.org/10.1590/0102-311xen097125>.

SILVA, T.; RODRIGUES, M.; RODRIGUES, F. Estado da arte sobre a síndrome cólica por compactação em equinos. *Pubvet*, v. 18, n. 2, e1552, 2024. DOI: 10.31533/pubvet.v18n02e1552. Disponível em: <https://doi.org/10.31533/pubvet.v18n02e1552>.

SZWARCWALD, C. Inquéritos nacionais de saúde: visão geral sobre técnicas de amostragem e dados coletados por desenhos complexos. *Epidemiologia e Serviços de Saúde*, v. 32, n. 3, 2023. DOI: 10.1590/s2237-96222023000300014.pt. Disponível em: <https://doi.org/10.1590/s2237-96222023000300014.pt>.

