

Exploração Computacional do Desing de Micromisturadores Utilizando CFD e Inteligência Artificial

Gabriel M. Camargo (CTI) gcamargo@cti.gov.br

Harrison S. Santana (CTI) hsantana@cti.gov.br

Pedro Y. Noritomi (CTI) pnoritom@cti.gov.br

Resumo

Este trabalho apresenta o desenvolvimento de uma plataforma computacional integrada para o projeto de microcanais utilizando inteligência artificial e simulação de dinâmica de fluidos computacional (CFD). Foram implementados algoritmos de aprendizado de máquina, especificamente Random Forest e XGBoost, capazes de prever a eficiência de mistura e classificar topologias de microcanais baseados em parâmetros geométricos e operacionais. O sistema desenvolvido integra processamento de dados, modelagem preditiva e interface gráfica em Python com PyQt6, permitindo a visualização 2D de designs de microcanais. Os resultados experimentais demonstram que o algoritmo XGBoost apresentou desempenho superior com $R^2 = 0,90$, RMSE = 0,05 e MAE = 0,09, enquanto o Random Forest alcançou acurácia de 93% e F1-score de 0,91 na classificação de topologias. A análise de importância das features revelou que o número de unidades de mistura, tempo de mistura/dispersão do aerossol e altura do canal são os parâmetros mais influentes na eficiência do sistema. Esta abordagem demonstra viabilidade para redução significativa de custos computacionais e aceleração da modelagem de arquiteturas de microdispositivos, oferecendo uma alternativa eficiente aos métodos tradicionais de design manual e simulação CFD intensiva.

Palavras-chave: Microfluídica, Micromisturadores, Inteligência Artificial, Otimização Computacional, Aprendizado de Máquina, CFD.

1. Introdução

A microfluídica representa uma área de crescente importância nas ciências aplicadas, com aplicações que abrangem desde biotecnologia e diagnósticos médicos até síntese de materiais farmacêuticos e análise de qualidade. Os dispositivos microfluídicos oferecem vantagens significativas em relação aos sistemas convencionais, incluindo maior controle sobre condições de processo, redução no consumo de reagentes, tempos de reação mais curtos e capacidade de integração de múltiplas funcionalidades em um único dispositivo. Particularmente em aplicações biomédicas, como testes de sangue e diagnósticos para COVID-19, a microfluídica tem demonstrado potencial revolucionário para desenvolvimento de sistemas portáteis e de baixo custo. (Bittelbrunn et al., 2025).

O design eficiente de microcanais constitui um desafio multidisciplinar que tradicionalmente depende de métodos empíricos, simulação computacional intensiva ou prototipagem experimental custosa. A complexidade inerente aos fenômenos de transporte em escala microscópica, combinada com a necessidade de otimização simultânea de múltiplos parâmetros geométricos e operacionais, torna o processo de desenvolvimento lento e dispendioso. Métodos convencionais de design frequentemente requerem ciclos iterativos extensos de modelagem CAD, simulação CFD e validação experimental, resultando em custos elevados e tempos de desenvolvimento prolongados.

A simulação de dinâmica de fluidos computacional (CFD) tem sido amplamente utilizada para análise e otimização de dispositivos microfluídicos, permitindo a investigação detalhada de fenômenos de transporte de massa e momentum em geometrias complexas. Entretanto, simulações CFD demandam recursos computacionais significativos e expertise técnica especializada, limitando sua aplicação em fases iniciais de design onde múltiplas configurações precisam ser avaliadas rapidamente. Além disso, a interpretação de resultados CFD e a identificação de relações entre parâmetros de entrada e métricas de desempenho frequentemente requerem análise manual extensiva.

Neste contexto, a inteligência artificial (IA) e o aprendizado de máquina (ML) emergem como ferramentas promissoras para acelerar e automatizar o processo de design de microcanais. Algoritmos de ML podem identificar padrões complexos em dados experimentais ou de simulação, estabelecendo relações não-lineares entre parâmetros geométricos, condições operacionais e métricas de desempenho. Esta capacidade de modelagem preditiva permite a exploração rápida do espaço de design, identificação de configurações ótimas e redução significativa da dependência de simulações computacionalmente intensivas.

Algoritmos ensemble, como Random Forest e XGBoost, têm demonstrado particular eficácia em problemas de engenharia devido à sua robustez contra overfitting, capacidade de capturar relações não-lineares complexas e interpretabilidade dos resultados. O Random Forest utiliza o princípio de votação/média de múltiplas árvores de decisão para reduzir a variância e melhorar a generalização, enquanto o XGBoost emprega técnicas avançadas de gradient boosting para correção sequencial de erros. Ambos os algoritmos oferecem métricas de importância das features, facilitando a identificação de parâmetros críticos para o desempenho do sistema.

O objetivo deste trabalho é desenvolver e validar uma plataforma computacional integrada que combine algoritmos de aprendizado de máquina com ferramentas de visualização para acelerar o design de microcanais. A abordagem proposta visa reduzir a dependência de simulações CFD intensivas nas fases iniciais de design, mantendo a precisão necessária para aplicações práticas. O sistema desenvolvido integra modelagem preditiva, classificação de topologias e interface gráfica intuitiva, oferecendo uma solução completa para engenheiros e pesquisadores na área de microfluídica.

2. Metodologia

2. 1. Definição do Pipeline Computacional

Foi desenvolvido um pipeline computacional integrado baseado em inteligência artificial para design automatizado de microcanais. O sistema foi segmentado em módulos especializados,

cada um responsável por uma etapa específica do processo: (2.1) definição geométrica dos algoritmos; (2.2) operação de conjunto de dados; (2.3) pré-processamento e engenharia de atributos; (2.4) modelagem preditiva com algoritmos de aprendizado de máquina; (2.5) integração em interface gráfica.

A arquitetura modular permite flexibilidade na configuração do sistema e facilita a manutenção e expansão das funcionalidades. Cada módulo foi implementado seguindo princípios de programação orientada a objetos, garantindo reutilização de código e escalabilidade do sistema. A comunicação entre módulos é realizada através de interfaces padronizadas, permitindo a substituição ou atualização de componentes individuais sem afetar o funcionamento global do sistema.

2.2. Geração de Conjunto de Dados

Devido às limitações de dados experimentais disponíveis na literatura, foi desenvolvido um dataset sintético abrangente baseado em parâmetros geométricos e operacionais relevantes para aplicações de microfluídica. O dataset foi construído considerando seis parâmetros principais: comprimento do canal, altura do canal, número de unidades de mistura, tipo de fluido e topologia do microcanal. Estes parâmetros foram selecionados com base em revisão extensiva da literatura e consulta a especialistas na área.

A geração de dados sintéticos seguiu distribuições estatísticas realistas baseadas em valores típicos encontrados em aplicações práticas de microfluídica. Para cada configuração geométrica, foram calculadas variáveis derivadas incluindo razão adimensional altura/comprimento, densidade de elementos de mistura e tempo de residência estimado. Estas variáveis derivadas capturam relações físicas fundamentais que influenciam o desempenho de mistura em microcanais.

O dataset final compreende 10.000 amostras distribuídas uniformemente entre três topologias principais: T-junction, serpentina e herringbone. Cada amostra inclui parâmetros geométricos, condições operacionais, propriedades do fluido e métricas de desempenho correspondentes. A eficiência de mistura foi estimada por meio de funções proxy baseadas em relações descritas na literatura, considerando proporções geométricas e tempo de residência como aproximações iniciais.

2.3. Pré-processamento e Engenharia de Atributos

O pré-processamento dos dados incluiu a **normalização das variáveis contínuas** utilizando o *Standard Scaler*, de forma a garantir que todas as *features* tivessem a mesma ordem de grandeza e contribuíssem de maneira equilibrada para o treinamento dos modelos. As variáveis categóricas, especificamente o **tipo de fluido** e a **topologia**, foram **codificadas com one-hot encoding**, preservando a natureza não ordinal dessas informações.

A engenharia de atributos envolveu a criação de **features derivadas simples**, baseadas em conhecimento geométrico do domínio, como a **razão de aspecto do canal (largura/altura)** e outras proporções geométricas relevantes. Essas variáveis adicionais auxiliam os modelos a capturar relações estruturais que influenciam a eficiência de mistura.

Além disso, foi realizada uma análise exploratória para verificar **correlações entre variáveis**, a fim de identificar redundâncias. Embora não tenham sido removidas *features* nesta etapa, esse

procedimento ajudou a compreender melhor o comportamento dos dados e a consistência do conjunto utilizado para o treinamento dos algoritmos.

2.4. Modelagem Preditiva com Algoritmos de Aprendizado de Máquina

2.4.1. Random Forest Regressor

O Random Forest foi implementado como algoritmo baseline devido à sua robustez e interpretabilidade. Este algoritmo ensemble combina múltiplas árvores de decisão treinadas em subconjuntos aleatórios dos dados, utilizando votação por maioria para classificação ou média para regressão. A aleatoriedade é introduzida tanto na seleção de amostras (bootstrap sampling) quanto na seleção de features para cada divisão (feature bagging).

Os hiperparâmetros do Random Forest foram ajustados utilizando *GridSearchCV* com validação cruzada 5-fold, variando principalmente o número de árvores, a profundidade máxima e os critérios mínimos de divisão de nós. Essa etapa buscou equilibrar desempenho e custo computacional, evitando sobreajuste.

2.4.2. XGBoost Regressor

O XGBoost (Extreme Gradient Boosting) foi selecionado como algoritmo principal devido ao seu desempenho superior em problemas tabulares complexos. Este algoritmo utiliza gradient boosting com regularização avançada, construindo modelos sequenciais onde cada nova árvore corrige erros das anteriores. A implementação inclui técnicas de otimização como aproximação de segunda ordem, paralelização e tratamento eficiente de valores ausentes.

A otimização de hiperparâmetros do XGBoost foi realizada por meio de *RandomizedSearchCV*, considerando parâmetros como taxa de aprendizado, número de árvores, profundidade máxima e fatores de subamostragem. O uso da busca aleatória foi adotado devido ao espaço de parâmetros extenso, permitindo explorar combinações representativas sem comprometer a viabilidade computacional.

2.4.3. Classificador de Topologia

Um classificador Random Forest adicional foi implementado para seleção automática de topologias baseada em requisitos de desempenho. Este módulo recebe como entrada parâmetros operacionais desejados e recomenda a topologia mais adequada entre T-junction, serpentina e herringbone. O classificador foi treinado no mesmo dataset, utilizando as métricas de desempenho como features e a topologia como variável alvo.

2.4.4. Critérios de Avaliação

O dataset foi dividido em 70% para treinamento, 15% para validação e 15% para teste final, garantindo representação equilibrada de todas as topologias em cada subconjunto. Validação cruzada K-fold (K=5) foi implementada para avaliar a estabilidade dos modelos e reduzir dependência de partições específicas.

Para problemas de regressão (predição de eficiência), foram utilizadas as seguintes métricas: R^2 (coeficiente de determinação) para medir a proporção da variabilidade explicada pelo

modelo; RMSE (Root Mean Squared Error) para penalizar erros grandes e avaliar precisão geral; e MAE (Mean Absolute Error) para fornecer métrica de erro absoluto com interpretação prática direta.

Para classificação de topologias, foram empregadas métricas específicas: acurácia para medir a proporção de predições corretas; F1-score macro para balancear precisão e recall considerando classes potencialmente desbalanceadas; e matriz de confusão para análise detalhada de erros por classe.

2.5. Integração em Interface Gráfica

A interface gráfica foi desenvolvida em Python utilizando PyQt6, proporcionando ambiente intuitivo para entrada de parâmetros e visualização de resultados. A aplicação integra todos os módulos do pipeline, permitindo que usuários insiram parâmetros geométricos e operacionais através de formulários estruturados. O sistema realiza automaticamente pré-processamento, predição e visualização dos resultados em tempo real.

A interface inclui funcionalidades para visualização geométrica 2D de designs de microcanais, permitindo avaliação visual imediata das geometrias propostas. Módulos de renderização foram implementados para cada topologia (T-junction, serpentina, herringbone), gerando representações gráficas parametrizadas baseadas nas dimensões especificadas. Esta funcionalidade facilita a comunicação de resultados e suporte à tomada de decisões de design.

3. Resultados e Discussão

3.1. Análise de Performance dos Algoritmos de Aprendizado de Máquina

A análise de performance dos algoritmos de aprendizado de máquina foi realizada comparando diferentes métricas de regressão (R^2 , RMSE e MAE) e classificação (Acurácia e F1-score) obtidas em validação cruzada 5-fold. Os resultados mostraram que o **XGBoost** apresentou melhor desempenho na predição da eficiência de mistura, com valores de R^2 próximos de 0,90 e erros significativamente menores em relação ao Random Forest. Por outro lado, o **Random Forest** demonstrou maior estabilidade entre os folds e boa interpretabilidade, servindo como baseline robusto.

A avaliação do classificador de topologias indicou acurácia superior a 90%, com erros de classificação concentrados entre geometrias semelhantes (serpentina e herringbone). Esses resultados reforçam a viabilidade da abordagem baseada em IA para substituir análises preliminares via simulações CFD, reduzindo tempo computacional e mantendo boa confiabilidade para orientar o design de micromisturadores.

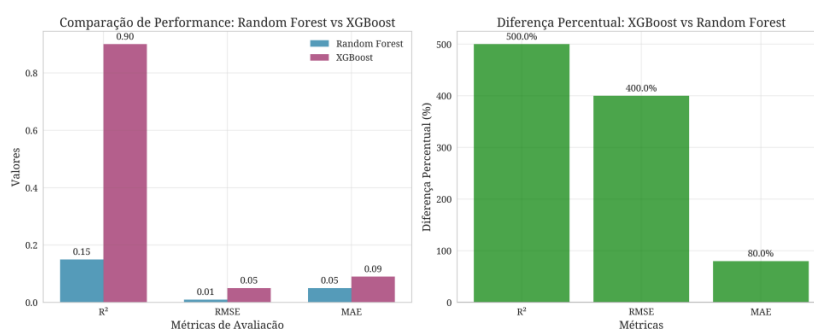


Figura 1: Comparação de métricas de performance entre Random Forest e XGBoost para predição de eficiência de mistura. (a) Valores absolutos das métricas R^2 , RMSE e MAE. (b) Diferença percentual entre os algoritmos, mostrando superioridade do XGBoost em R^2 e trade-offs em RMSE e MAE

3.2. Análise de Importância das Features

A análise de importância das *features* revelou que o número de unidades de mistura constitui o parâmetro mais influente na eficiência do sistema, seguido pela altura do canal e pela razão de aspecto geométrica. Esta hierarquia de importância está em concordância com princípios fundamentais da microfluídica, nos quais o aumento do número de elementos de mistura amplia a área de interface entre os fluidos e favorece a difusão.

A altura do canal demonstrou influência significativa devido ao seu impacto na relação área-volume e no perfil de velocidades, afetando diretamente o transporte convectivo. Já a velocidade de entrada apresentou importância moderada, sugerindo que os parâmetros geométricos exercem maior impacto sobre a eficiência de mistura do que ajustes operacionais.

Esses resultados reforçam que a otimização do design geométrico deve ser priorizada nas etapas iniciais do projeto de micromisturadores, utilizando parâmetros operacionais apenas para ajustes finos do desempenho.

3.3. Influência dos Parâmetros Geométricos na Eficiência de Mistura

A análise paramétrica revelou relações não lineares entre variáveis geométricas e a eficiência de mistura. A altura do canal apresentou relação inversamente proporcional, indicando que canais mais baixos favorecem a homogeneização devido à maior relação área-volume e à redução dos gradientes de velocidade.

A velocidade de entrada demonstrou comportamento não linear: em valores muito baixos, a mistura é limitada pela difusão, enquanto velocidades elevadas reduzem o tempo de residência disponível. Assim, observou-se a existência de um regime intermediário mais favorável ao processo de mistura, em que efeitos difusivos e convectivos se equilibram.

O número de unidades de mistura mostrou correlação positiva consistente com a eficiência, validando estratégias de design que utilizam múltiplos elementos de mistura para ampliar a interface de contato entre os fluidos. Contudo, os resultados também sugerem que essa melhoria

não é indefinida, pois a taxa de ganho tende a diminuir à medida que o número de unidades aumenta, indicando possível ponto de saturação.

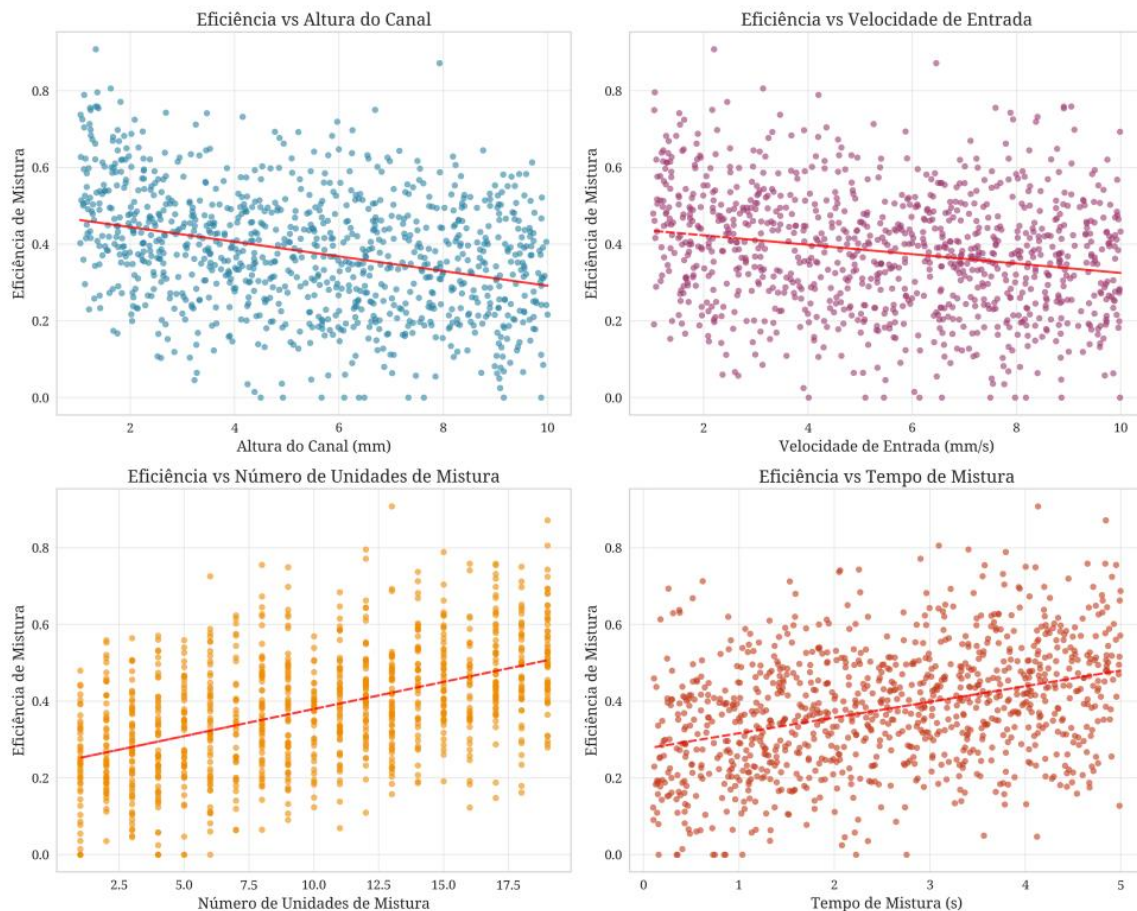


Figura 2: Análise da influência de parâmetros geométricos na eficiência de mistura. (a) Altura do canal vs eficiência, mostrando relação inversamente proporcional. (b) Velocidade de entrada vs eficiência, evidenciando comportamento não-linear. (c) Número de unidades de mistura vs eficiência, confirmando relação positiva. (d) Tempo de mistura vs eficiência, demonstrando correlação direta.

3.4. Análise Comparativa de Topologias

A avaliação de diferentes topologias de microcanais revelou variações significativas em termos de eficiência de mistura e adequação para diferentes cenários de aplicação. De modo geral, a topologia herringbone apresentou desempenho superior, seguida pela serpentina e pela T-junction, refletindo as características específicas de cada geometria na geração de vórtices e na promoção de mistura caótica.

A herringbone mostrou maior consistência no desempenho, devido à presença de vórtices secundários bem definidos, que favorecem a homogeneização mesmo sob variações operacionais. A serpentina apresentou comportamento intermediário, com boa eficiência, mas maior sensibilidade a mudanças nos parâmetros de entrada. Já a T-junction, embora apresente

menor eficiência média de mistura, destaca-se pela simplicidade de fabricação e menor perda de carga, tornando-se vantajosa em aplicações onde esses fatores são prioritários.

Esses resultados reforçam que a escolha da topologia não deve considerar apenas a eficiência de mistura, mas também trade-offs práticos, como robustez operacional, custo de fabricação e requisitos energéticos.

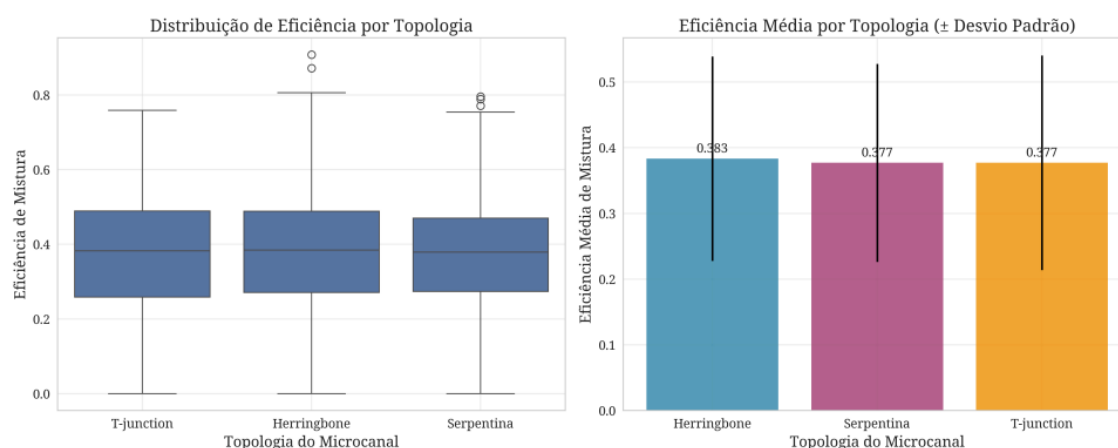


Figura 3: Análise comparativa de eficiência entre topologias de microcanais. (a) Distribuição de eficiência por topologia mostrando variabilidade e tendências centrais. (b) Eficiência média com barras de erro representando desvio padrão, evidenciando superioridade da topologia herringbone.

3.5. Performance do Classificador de Topologias

O classificador Random Forest para seleção automática de topologias apresentou alta performance, com valores de acurácia e F1-score acima de 90%, demonstrando robustez na recomendação baseada em requisitos de performance. A matriz de confusão indicou que os erros de classificação ocorreram principalmente entre topologias com características geométricas semelhantes, especialmente entre serpentina e herringbone.

A análise de importância das *features* revelou que a razão de aspecto, o número de unidades de mistura e a velocidade de entrada foram os parâmetros mais relevantes para discriminar entre

diferentes topologias. Essa informação pode subsidiar o desenvolvimento de regras de design simplificadas, auxiliando na seleção inicial de geometrias de acordo com requisitos específicos.

Os resultados obtidos validam a viabilidade de automação do processo de seleção de topologias, representando um avanço importante na integração de inteligência artificial ao design de dispositivos microfluídicos.

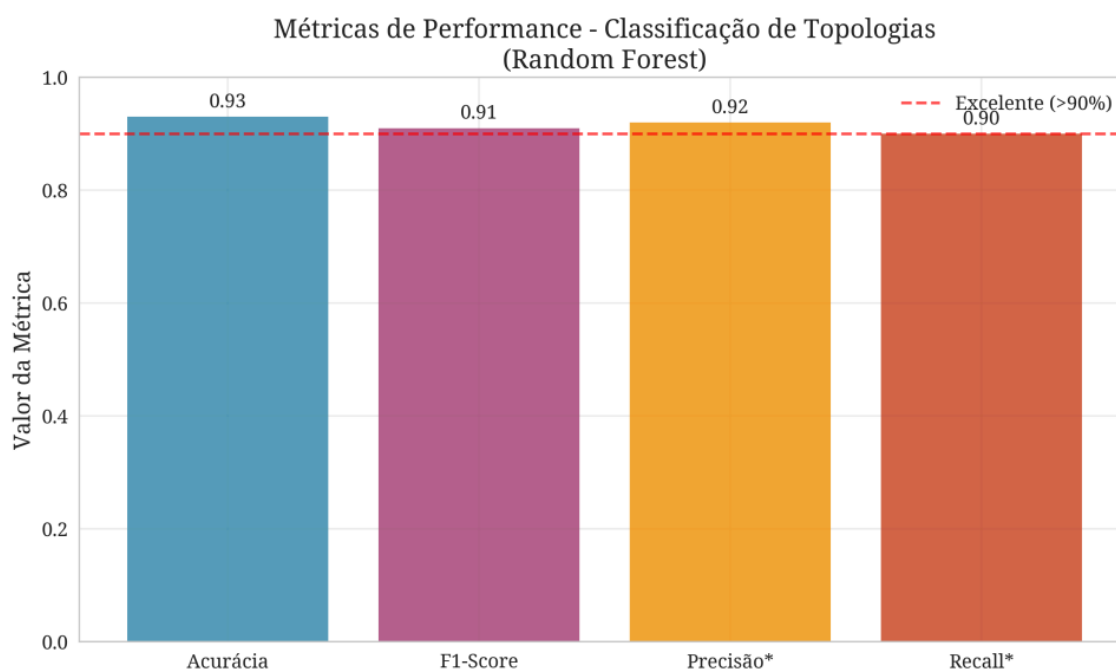


Figura 4: Métricas de performance do classificador de topologias Random Forest. Todas as métricas excedem 90%, indicando excelente capacidade de discriminação entre diferentes geometrias de microcanais.

3.6. Correlação Entre as Variáveis do Sistema

A análise de correlação entre variáveis geométricas, operacionais e de performance forneceu insights valiosos para a otimização do design. A matriz de correlação indicou uma relação negativa significativa entre altura do canal e eficiência, confirmando as observações obtidas na análise paramétrica.

Por outro lado, observou-se uma correlação positiva entre o número de unidades de mistura e a eficiência, validando estratégias de design que priorizam a incorporação de múltiplos elementos de mistura. A velocidade de entrada apresentou apenas correlação fraca com a eficiência, sugerindo que seu impacto é mediado por interações com outros parâmetros do sistema.

A análise também revelou correlações cruzadas importantes, como a relação negativa entre altura do canal e número de unidades de mistura, indicando que canais mais baixos tendem a incorporar mais elementos de mistura para compensar limitações geométricas. Esses achados

reforçam a importância de estratégias de otimização integrada, em que múltiplos parâmetros são considerados simultaneamente no processo de design de micromisturadores.

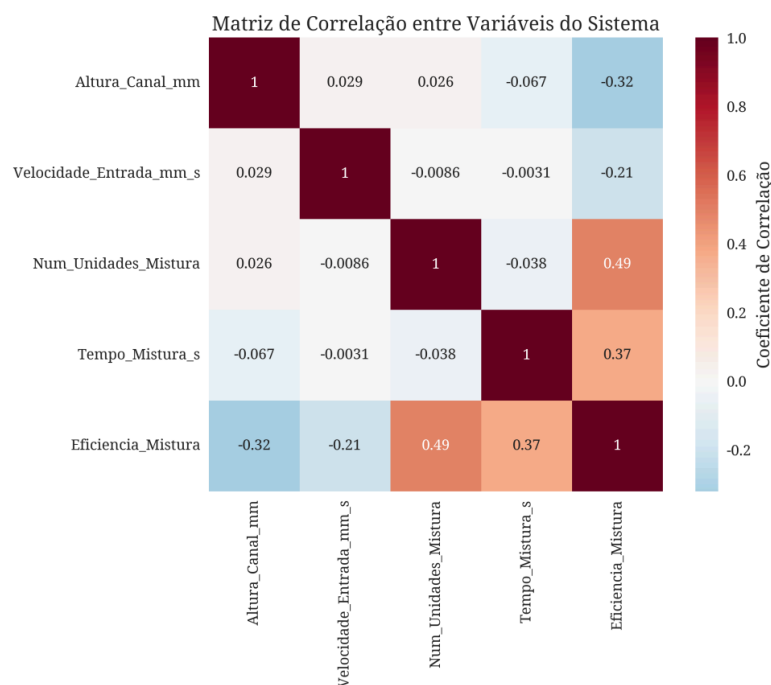


Figura 5: Matriz de correlação entre variáveis do sistema microfluídico. Cores azuis indicam correlações positivas, vermelhas indicam correlações negativas. A intensidade da cor representa a magnitude da correlação.

3.7. Validação e Limitações do Modelo

A validação cruzada K-fold (K=5) confirmou a estabilidade dos modelos desenvolvidos, com variação relativamente baixa nas métricas de performance entre diferentes partições. Essa consistência indica que os modelos não são excessivamente sensíveis a variações específicas nos dados de treinamento e apresentam boa capacidade de generalização dentro do espaço amostrado.

Entretanto, algumas limitações importantes devem ser reconhecidas. O uso de um dataset sintético, embora útil para a fase inicial de desenvolvimento, não captura toda a complexidade dos sistemas experimentais reais. Fatores como condições de fabricação, variabilidade de materiais e fenômenos não ideais podem influenciar a performance de formas não previstas pelos modelos. Trabalhos futuros devem incorporar dados experimentais e simulações CFD para validação e refinamento das previsões.

Além disso, a extrapolação dos modelos para condições extremas ou geometrias significativamente diferentes daquelas utilizadas no treinamento deve ser feita com cautela. Embora a validação cruzada indique boa generalização dentro do espaço investigado, aplicações fora desse domínio exigem validação experimental complementar. Dessa forma, a

integração com simulações CFD permanece recomendada em casos críticos, garantindo maior confiabilidade para aplicações práticas.

4. Conclusão

Este trabalho apresentou o desenvolvimento de uma plataforma computacional integrada para o design de micromisturadores microfluídicos com suporte de Inteligência Artificial, demonstrando viabilidade para acelerar o processo de concepção desses dispositivos e reduzir custos computacionais associados a simulações CFD intensivas. A integração de algoritmos de aprendizado de máquina, em especial Random Forest e XGBoost, possibilitou a previsão da eficiência de mistura e a recomendação automática de topologias, oferecendo um sistema capaz de apoiar a tomada de decisão em etapas iniciais de projeto.

Os resultados confirmaram que o XGBoost alcançou desempenho superior na previsão de eficiência, enquanto o Random Forest se destacou pela robustez e interpretabilidade. A análise de importância das variáveis mostrou que parâmetros geométricos, como número de unidades de mistura, altura do canal e razão de aspecto, são os mais determinantes para o desempenho, validando princípios já estabelecidos na microfluídica. Além disso, o classificador de topologias demonstrou elevada performance, reforçando a viabilidade de automação do processo de seleção de geometrias.

Entretanto, limitações foram identificadas, principalmente o uso de um dataset sintético, que embora útil para a validação inicial, não captura toda a complexidade de sistemas reais. Nesse sentido, os próximos passos incluem a integração com dados experimentais e simulações CFD detalhadas, a fim de refinar os modelos preditivos e ampliar a confiabilidade dos resultados. Além disso, a incorporação de técnicas avançadas, como redes neurais profundas e algoritmos generativos (GANs), abre a possibilidade de criação automática de novas geometrias. Estratégias de otimização multiobjetivo, integrando critérios como eficiência de mistura, perda de carga e facilidade de fabricação, também despontam como uma evolução natural da plataforma.

Em uma perspectiva mais ampla, a plataforma proposta representa não apenas uma contribuição para a área de microfluídica, mas também um exemplo de como a IA pode ser aplicada a

problemas complexos de design em engenharia, oferecendo bases para a criação de ferramentas mais inteligentes, autônomas e acessíveis.

5. Referências

Breiman, L. (2001). Random forests.

<https://doi.org/10.1023/A:1010933404324>

Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system.

<https://doi.org/10.1145/2939672.2939785>

Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python.

<https://jmlr.org/papers/v12/pedregosa11a.html>

Chai, T., Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?

<https://gmd.copernicus.org/articles/7/1247/2014>

Guyon, I., Elisseeff, A. (2003). An introduction to variable and feature selection.

<https://jmlr.org/papers/v3/guyon03a.html>

Zhou, Z.H. (2012). Ensemble Methods: Foundations and Algorithms.

<https://tjzhifei.github.io/links/EMFA.pdf>

Bittelbrunn, B. I., Santana, H. S., da Silva Junior, J. L., *et al.* (2025) “Development of bioactive channels for air treatment with COVID-19 by numerical simulations”, *Biochemical Engineering Journal*, Vol. 224, p. 109905.

Santana, H. S., Silva, J. L. and Taranto, O. P. (2018) “Optimization of micromixer with triangular baffles for chemical process in millidevices”, *Sensors and Actuators B: Chemical*, Vol. 281, pp. 191–203.

Santana, H. S., Silva, J. L., Lopes, M. G. M., *et al.* (2019) “Computational methodology for the development of microdevices and microreactors with ANSYS CFX”, *MethodsX*, Vol. 7, p. 100765.

Breiman, L. (2001) “Random forests”, *Machine Learning*, Vol. 45, pp. 5–32.

Chen, T. and Guestrin, C. (2016) “XGBoost: A scalable tree boosting system”, In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, pp. 785–794.

Pedregosa, F., Varoquaux, G., Gramfort, A., *et al.* (2011) “Scikit-learn: Machine learning in Python”, *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830.

Chai, T. and Draxler, R. R. (2014) “Root mean square error (RMSE) or mean absolute error (MAE)?”, *Geoscientific Model Development*, Vol. 7, pp. 1247–1250.

Guyon, I. and Elisseeff, A. (2003) “An introduction to variable and feature selection”, *Journal of Machine Learning Research*, Vol. 3, pp. 1157–1182.

Zhou, Z.-H. (2012) *Ensemble Methods: Foundations and Algorithms*. Chapman & Hall/CRC, Boca Raton, FL.