

Estimativa de Emissões em Veículos a Combustão baseado em Redes Neurais LSTM e Análise de Incertezas via Monte Carlo Dropout

Ludmila Gomes, Bárbara Cavalcante, Mateus de Sousa, José Hélio Bento da Silva, Iguatemi E. Fonseca e Juan M. Villanueva

Resumo— Este artigo propõe um softsensor baseado em redes neurais LSTM para estimar as emissões de CO₂ em veículos a combustão, utilizando dados obtidos via OBD-II. A abordagem permite modelar sequências temporais complexas em condições reais de condução. Para quantificar a incerteza das previsões, foi empregada a técnica de Monte Carlo Dropout, permitindo a geração de intervalos de confiança em tempo real. Os resultados indicam alta precisão na estimativa das emissões, com incertezas bem caracterizadas, contribuindo para aplicações em monitoramento ambiental e otimização veicular.

Palavras-Chave— Emissões veiculares, Softsensor, Redes neurais LSTM, OBD-II, Monte Carlo Dropout, Incerteza, Estimativa de CO₂.

Abstract— This paper proposes a softsensor based on LSTM neural networks to estimate CO₂ emissions in combustion vehicles using OBD-II data. The approach models complex temporal sequences under real driving conditions. To quantify prediction uncertainty, the Monte Carlo Dropout technique is applied, enabling real-time confidence interval estimation. Results show high accuracy in emission estimates with well-characterized uncertainty, supporting applications in environmental monitoring and vehicle performance optimization.

Keywords— Vehicle emissions, Softsensor, LSTM neural networks, OBD-II, Monte Carlo Dropout, Uncertainty, CO₂ estimation.

I. INTRODUÇÃO

As emissões de dióxido de carbono (CO₂) provenientes de veículos representam uma parcela relevante dos gases de efeito estufa, contribuindo para a mudança climática e a poluição do ar urbano. No Brasil, o setor de transportes responde por cerca de 47

Sensores físicos diretos são, em geral, caros e restritos a ambientes controlados. Como alternativa, destacam-se os soft-sensors — modelos computacionais que estimam variáveis não diretamente mensuráveis com base em dados observáveis. Em veículos a combustão, a interface OBD-II fornece informações como rotação, carga do motor e temperatura, viabilizando estimativas indiretas de emissões com alta frequência [2].

Ludmila Gomes, Universidade Federal da Paraíba, João Pessoa-Paraíba, e-mail: lvgg@academico.ufpb.br. Bárbara Cavalcante, Universidade Federal da Paraíba, João Pessoa-Paraíba, e-mail: barbara.cavalcante@academico.ufpb.br. Mateus de Sousa, Universidade Federal da Paraíba, João Pessoa-Paraíba, e-mail: mateus.sousa@estudante.cear.ufpb.br. José Hélio Bento da Silva, Universidade Federal da Paraíba, João Pessoa-Paraíba, e-mail: jose.dasilva@estudante.cear.ufpb.br. Iguatemi E. Fonseca, Universidade Federal da Paraíba, João Pessoa-Paraíba, e-mail: iguatemi@ci.ufpb.br. Juan M. Villanueva, Universidade Federal da Paraíba, João Pessoa-Paraíba, e-mail: jmauricio@cear.ufpb.br

Redes neurais artificiais, especialmente as do tipo LSTM, são adequadas para capturar dependências temporais complexas desses dados [3]. Contudo, para aplicações sensíveis como o monitoramento ambiental, é igualmente importante quantificar a incerteza associada às previsões. O método Monte Carlo Dropout se destaca nesse contexto por permitir a obtenção de distribuições probabilísticas das saídas do modelo, fornecendo intervalos de confiança para as estimativas [4].

Este artigo propõe um softsensor baseado em redes LSTM para estimativa de emissões veiculares a partir de dados OBD-II, incorporando análise de incertezas via Monte Carlo Dropout. Essa abordagem permite avaliar a confiabilidade das previsões e contribui com soluções de baixo custo e alta robustez, com aplicações potenciais em estudos ambientais, gestão de frotas e políticas públicas. As principais contribuições incluem: a) estimativa de emissões com dados OBD-II sem sensores físicos; b) uso de redes LSTM para modelagem temporal dos dados; e c) incorporação de técnicas de quantificação de incerteza, ainda pouco exploradas nesse domínio.

II. ESTADO DA ARTE

O uso de dados obtidos via interface OBD-II tem ganhado destaque como alternativa viável para estimar parâmetros de desempenho e impacto ambiental de veículos em tempo real. Diversos estudos têm explorado softsensors baseados em aprendizado de máquina para inferir variáveis difíceis de medir diretamente, como o consumo de combustível e as emissões de poluentes.

Torregrossa et al. (2020) desenvolveram um sistema de inferência de emissões de CO₂ e NOx a partir de dados OBD-II utilizando redes feedforward, com resultados promissores em ciclos de condução simulados. No entanto, a abordagem não contemplava a estrutura sequencial dos dados, limitando sua capacidade preditiva em situações dinâmicas [2].

Chen et al. (2021) aplicaram redes neurais do tipo LSTM para prever o consumo de combustível e emissões instantâneas com base em séries temporais obtidas de veículos híbridos. A rede demonstrou bom desempenho em trajetos urbanos, mas não abordou aspectos de confiabilidade ou quantificação de incertezas nas previsões, fator crucial para aplicações críticas em ambiente real [5].

Zhou et al. (2022) combinaram LSTM com Attention Mechanisms para estimar emissões de CO₂ em frotas comerciais

utilizando dados GPS e OBD-II. Embora o modelo apresentasse alta acurácia, o estudo focou apenas na previsão pontual, sem indicar margens de erro ou confiança nas estimativas, o que limita sua aplicabilidade em diagnósticos automatizados e decisões regulatórias [6].

Em relação à quantificação de incertezas, Gal e Ghahramani (2016) introduziram o uso do Monte Carlo Dropout como técnica de aproximação bayesiana em redes neurais profundas, demonstrando que esse método permite representar incertezas epistemológicas com baixo custo computacional. Apesar disso, seu uso em aplicações automotivas, particularmente na estimativa de emissões, ainda é incipiente [4].

Mais recentemente, Liu et al. (2023) investigaram a aplicação de técnicas bayesianas em redes convolucionais para estimativas de emissões veiculares, mas o estudo não incluiu modelos recorrentes como LSTM, nem explorou dados OBD-II como fonte primária de entrada [7].

Apesar dos avanços recentes, observa-se uma lacuna relevante na integração de modelos recorrentes como LSTM com mecanismos de quantificação de incertezas voltados à estimativa de emissões em veículos convencionais. A maioria dos trabalhos foca em previsões pontuais e omitem a robustez das previsões em condições operacionais diversas, como variações de carga, temperatura ambiente ou estilo de condução. Além disso, poucos estudos exploram plenamente o potencial de dados embarcados via OBD-II para estimativas em tempo real com caracterização probabilística da incerteza. Diante desse panorama, este trabalho busca preencher essa lacuna ao propor o desenvolvimento de um softsensor baseado nas redes neurais artificiais do tipo em LSTM com análise de incertezas via Monte Carlo Dropout, oferecendo uma solução robusta, explicável e viável para aplicações reais de monitoramento ambiental veicular.

III. FUNDAMENTAÇÃO TEÓRICA

Nesta seção serão apresentados os fundamentos preliminares sobre softsensors, interfaces OBD-II para aquisição de dados de veículos, introdução a redes neurais artificiais e análise de incertezas em modelos baseados no método Monte Carlo Dropout.

A. Softsensors e Estimativa de Emissões

Softsensors, ou sensores virtuais, são modelos matemáticos ou de aprendizado de máquina que estimam variáveis de difícil medição direta a partir de variáveis acessíveis, muitas vezes com auxílio de sensores físicos embarcados. No contexto automotivo, as variáveis fornecidas pela interface OBD-II (OnBoard Diagnostics), como rotação do motor, carga do acelerador, massa de ar admitida e velocidade do veículo, oferecem uma base rica para inferir grandezas ambientais relevantes, como consumo de combustível e emissões de CO₂ [8].

A utilização de softsensors permite a realização de estimativas em tempo real, com menor custo e complexidade quando comparado a sensores de emissão direta, que frequentemente requerem sistemas laboratoriais ou instrumentação invasiva nos gases de escape. Assim, a modelagem indireta baseada

em dados de sensores embarcados tem se mostrado uma estratégia promissora, especialmente quando associada a técnicas modernas de aprendizado de máquina.

B. Interface OBD-II como Fonte de Dados

O protocolo OBD-II é um padrão amplamente adotado em veículos modernos, fornecendo acesso a parâmetros internos do motor e do sistema de controle do veículo. A leitura desses dados pode ser feita com baixo custo e permite coleta contínua em condições reais de operação. Entre as variáveis comumente disponíveis estão: RPM, temperatura do motor, posição do acelerador, pressão absoluta do coletor, entre outras, que influenciam diretamente na geração de emissões.

A vantagem do OBD-II reside em sua padronização e disponibilidade em praticamente toda a frota a combustão fabricada a partir da década de 2000, o que o torna ideal para aplicações em larga escala e estudos populacionais de impacto ambiental veicular [9].

C. Redes Neurais Artificiais

As redes neurais recorrentes (RNNs) são arquiteturas projetadas para lidar com dados sequenciais, pois permitem que informações de entradas anteriores influenciem a saída atual. No entanto, RNNs tradicionais enfrentam dificuldades em capturar dependências de longo prazo, devido ao problema do desaparecimento ou explosão do gradiente durante o treinamento [10].

Para superar essa limitação, foram propostas as redes LSTM, uma variante de RNNs, capaz de aprender dependências temporais de longo prazo. A LSTM utiliza um mecanismo de portas (gate system) que controla o fluxo de informação, permitindo ao modelo aprender quando "lembrar" ou "esquecer" informações passadas. Essa estrutura torna a LSTM particularmente eficaz para modelar séries temporais complexas e ruidosas, como os dados obtidos de sensores veiculares sob diferentes condições de operação [3].

No contexto deste trabalho, a escolha da LSTM foi motivada por sua capacidade superior de capturar padrões dinâmicos ao longo do tempo, fundamentais para estimar variáveis como emissões de CO₂, que são fortemente dependentes do histórico de operação do motor. Modelos estáticos ou redes neurais tradicionais tendem a apresentar desempenho inferior ao não considerar adequadamente as transições temporais rápidas (ex: aceleração repentina, mudança de marcha, desaceleração), que afetam diretamente os níveis de emissão. Além disso, a LSTM permite construir modelos robustos mesmo em cenários com ruído ou dados incompletos, frequentes em sistemas embarcados como o OBD-II.

D. Monte Carlo Dropout e Incertezas em Redes Neurais Artificiais

A quantificação de incertezas em modelos de aprendizado de máquina é essencial quando se deseja confiabilidade em contextos críticos, como decisões ambientais, controle veicular ou diagnóstico automático. O método Monte Carlo Dropout propõe uma forma eficiente de estimar incertezas epistemológicas (relacionadas ao modelo) em redes neurais profundas.

Introduzido por Gal e Ghahramani (2016), o método consiste em manter ativa a técnica de dropout (aleatorização de unidades neurais) não apenas durante o treinamento, mas também na inferência. Ao realizar várias inferências estocásticas com o mesmo dado de entrada, é possível obter uma distribuição das previsões, permitindo a construção de intervalos de confiança e o cálculo da variância associada à saída do modelo [4].

Essa abordagem é especialmente relevante quando o modelo opera em cenários fora do padrão de treinamento, como condições extremas de condução ou uso de combustível fora da faixa média, oferecendo uma camada adicional de segurança e interpretabilidade ao sistema.

IV. METODOLOGIA PROPOSTA

A metodologia seguida neste trabalho se apoia no modelo metodológico CRISP-DM (Cross-Industry Standard Process for Data Mining) [11], amplamente utilizado em projetos de ciência de dados. Assim, as etapas principais desse modelo, a citar: (i) compreensão do negócio, (ii) compreensão dos dados, (iii) preparação dos dados, (iv) modelagem, (v) avaliação, (vi) implantação, são seguidas a fim de estimar em tempo real as emissões de CO₂ de veículos a combustão utilizando dados provenientes da interface OBD-II, combinando redes neurais LSTM com análise de incertezas via Monte Carlo Dropout. A Figura 1 ilustra o diagrama de fluxo da metodologia proposta neste trabalho.

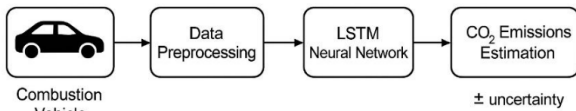


Fig. 1. Estimativa de Emissões de CO₂ em Veículos a Combustão com Softsensors e Análise de Incertezas via Monte Carlo Dropout.

A. Coleta de Dados via OBD-II e Pré-processamento dos Dados

A coleta de dados foi realizada por meio da interface OBD-II instalada no veículo, que fornece variáveis operacionais críticas em tempo real, como rotação do motor (RPM), carga do motor (engine load), temperatura do líquido de arrefecimento, posição do acelerador (throttle position), velocidade do veículo, entre outras. O aplicativo Torque Pro foi utilizado para visualização das informações coletadas e, finalmente, a aplicação Conect2AI, cujo desenvolvimento é demonstrado por Medeiros et al. [14] foi usada para o acesso completo aos dados em arquivos CSV.

Nesta etapa, os dados brutos são tratados para garantir a consistência e a qualidade do conjunto de dados. Posteriormente, os processos realizados serão discutidos mais detalhadamente, englobando limpeza, normalização, segmentação dos dados, entre outras etapas.

B. Modelagem com Rede Neural LSTM

O modelo LSTM é treinado para mapear sequências de entrada (variáveis do OBD-II) em saídas correspondentes às emissões estimadas de CO₂. A arquitetura foi ajustada considerando:

- Duas camadas LSTM empilhadas, seguidas por camadas densas;
- Função de perda do tipo Mean Squared Error (MSE);
- Otimizador Adam.

Para avaliar o desempenho dos modelos preditivos, foram utilizadas as métricas de Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) e R-squared (R²). E, resumidamente, a métrica MAE calcula a média dos erros absolutos entre os valores previstos e reais, MAPE expressa esse erro em termos percentuais e, finalmente, o R² indica quão bem os dados se ajustam ao modelo preditivo, variando de 0 a 1 [12].

C. Estimativa de Incertezas com Monte Carlo Dropout

Durante a fase de inferência, o modelo LSTM aplica dropout de forma estocástica (ativado mesmo em modo de teste) para simular uma distribuição de modelos. São realizadas N inferências com a mesma entrada, e a saída final é obtida por média das previsões. A variância entre as previsões permite calcular intervalos de confiança (ex.: 95%) para cada estimativa de CO₂, oferecendo robustez e interpretabilidade.

V. RESULTADOS

Esta seção apresenta os resultados obtidos a partir da aplicação do modelo proposto. O processo foi dividido em cenários progressivos, começando com uma configuração básica e evoluindo para cenários mais complexos e realistas, de forma a validar a robustez do softsensor e a eficácia da análise de incertezas.

A. Aquisição e Pré-processamento dos Dados

Os dados foram coletados via interface OBD-II de um veículo a combustão durante trajetos urbanos e rodoviários de 30 a 60 minutos, com amostragem de 1 Hz e mais de 100 variáveis monitoradas (como RPM, velocidade, temperatura do motor e carga). No pré-processamento, valores faltantes foram preenchidos e instâncias com velocidade nula removidas. Os dados foram divididos em conjuntos de treinamento (70%), validação (15%) e teste (15%), normalizados com MinMax e estruturados em sequências temporais de 64 passos para uso em modelos LSTM.

B. Seleção de Variáveis

Inicialmente, foi realizada a seleção de atributos para predição de valores de gramas de CO₂ emitidos. Destaca-se a necessidade da seleção, tendo em vista que o data-set originalmente apresenta 104 atributos, devido à inviabilidade de uso do conjunto completo, por questões de eficiência computacional. Para tanto, foi realizada uma análise dos dados, que resultou na remoção manual de atributos irrelevantes. Basicamente, foram

removidos atributos que não se relacionavam com os fenômenos mecânicos e físicos relativos ao processo de combustão e geração de movimento no veículo, informações do módulo OBD-II coletadas e dados dos GPS. Ao fim do processo de seleção manual, foram obtidos 31 atributos restantes.

Com as 31 colunas acima, foi realizado o cálculo da matriz de correlação, usando o método de Pearson. Com os resultados, foram retirados os atributos de correlação baixa (removendo atributos que se apresentassem abaixo de 0.6) ou de correlação equivalente (como os atributos 'Accelerator PedalPosition D(%)' e 'Accelerator PedalPosition E(%)', e o atributo de maior correlação foi mantido). Finalmente, os atributos restantes são verificados abaixo: Engine RPM (rpm), Engine Load (%), Absolute Throttle Position B(%), Accelerator Pedal Position (%), Fuel Rail Pressure (psi), Intake Manifold Pressure (psi), co2_grams_per_sec, Speed (km/h).

C. Modelo LSTM

Com base no trabalho realizado por Singh et al. [13], a arquitetura do modelo LSTM desenvolvido apresenta duas camadas LSTM com 128 neurônios e ativação tangente hiperbólica, com uso de dropout após cada camada LSTM. Em seguida, é adicionada uma camada densa de 240 neurônios e ativação ReLU, novamente seguida por uma camada dropout. Finalmente, a última camada densa com neurônio único e ativação linear, cujo resultado é a predição de CO 2. O modelo foi treinado com a função de perda MSE e otimizador Adam. Totalizando 119.761 parâmetros. Os seguintes critérios de parada foram utilizados: Early Stopping com paciência de 20 épocas; e Redução da taxa de aprendizado com base na perda de validação. Na Figura 2 se ilustram as curvas de perda, a média do erro absoluto (MAE) e MAPE para o treinamento e validação deste modelo.

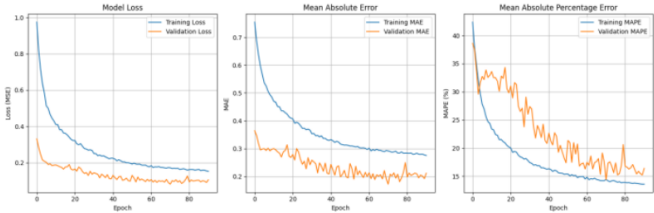


Fig. 2. Gráfico representativo das curvas de MAE no treinamento e validação do primeiro modelo desenvolvido.

Na Figura 3 acima, se ilustra o processo de inferência deste modelo, incluindo a avaliação da incerteza por meio do Monte Carlo Dropout. Como pode ser observado, o valor da incerteza é bastante grande, o que motivou a sintonizar de forma mais adequada os hiperparâmetros do Monte Carlo Dropout para limitar a incerteza a um determinado valor.

Para resolver este problema, foi utilizada a técnica de Grid Search sobre os parâmetros do dropout, regularização l2 e quantidade de camadas LSTM utilizadas, a fim de obter o menor valor possível de desvio padrão no processo da aplicação do Monte Carlo Dropout e, conseqüentemente, obter o modelo que apresenta menor grau de incerteza possível sobre as predições realizadas. Os resultados do procedimento de Grid

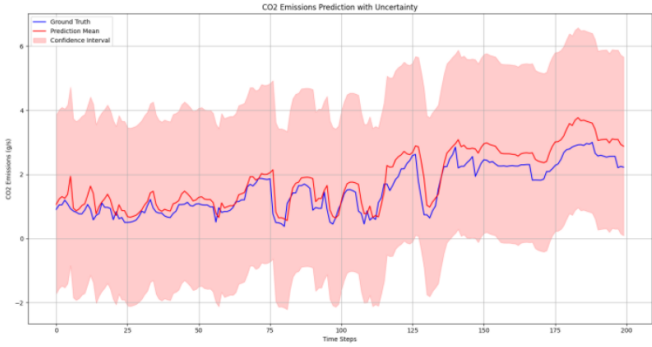


Fig. 3. Resultados sobre conjunto de teste, com valores reais, valores preditos e intervalo de confiança sobre as predições do primeiro modelo.

Search demonstraram que a configuração ótima apresentava os seguintes valores:

- Dropout = 0.5;
- Regularização = 0,0001;
- 2 camadas LSTM.

A partir da sintonização dos hiperparâmetros, pode-se comparar os resultados obtidos dos modelos desenvolvidos, considerando que o primeiro deles apresentou as melhores métricas, mas alto desvio padrão, enquanto que o segundo apresentou métricas também boas, mas com o menor desvio padrão possível.

As Figuras 5 e 6 dizem respeito aos resultados após a sintonização adequada dos hiperparâmetros com base no processo de Grid Search. Dessa forma, é possível encontrar configurações específicas para a rede LTSM e configuração adequada do Monte Carlo Dropout para obter uma incerteza menor possível.

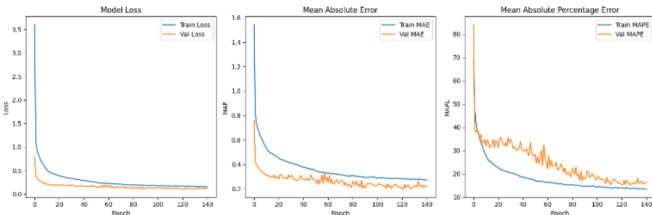


Fig. 4. Curvas de perda, MAE e MAPE, respectivamente, do treinamento e validação do segundo modelo desenvolvido (pós Grid Search).

D. Métricas de Desempenho

Considerando os dois modelos desenvolvidos, a Tabela 1 apresenta os resultados médios obtidos nos diferentes cenários apresentados anteriormente.

TABELA I
RESULTADOS MÉDIOS OBTIDOS NOS DIFERENTES CENÁRIOS.

	Melhor Modelo LSTM	Modelo LSTM pós-Grid Search
MAE	0,246	0,252
MAPE	15,8%	16,4%
R2	0,839	0,789
σ médio	1,392	0,25

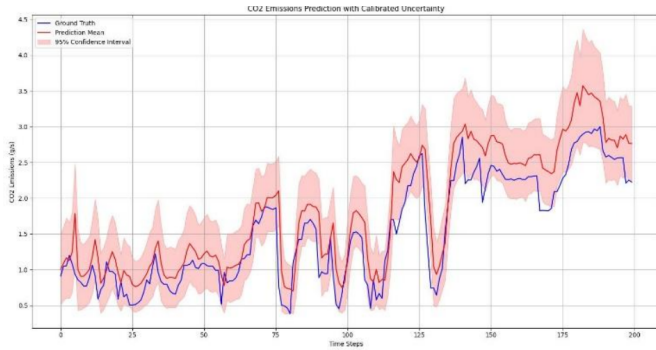


Fig. 5. Resultados sobre conjunto de teste, com valores reais, valores preditos e intervalo de confiança sobre as previsões (pós Grid Search).

Visualizando a tabela comparativa, percebe-se que, mediante uma pequena perda em relação às métricas sobre as previsões, foi possível diminuir consideravelmente o desvio padrão sobre os resultados obtidos. Assim, houve um decrescimento mínimo sobre as previsões realizadas pelo novo modelo, mas pode-se considerar que a confiança sobre ele aumentou.

E. Avaliação de Incerteza com Monte Carlo Dropout

Durante a inferência, foi aplicada a técnica de Monte Carlo Dropout com 1000 amostras por instância, mantendo ativo o dropout (taxa = 0,5) na rede LSTM. Com isso, foram calculados o desvio padrão e os intervalos de confiança para cada previsão de CO₂.

A Fig. 6 ilustra as previsões médias e os respectivos intervalos de confiança de 95% (IC95%) para um trecho urbano. Observa-se que os maiores intervalos de incerteza ocorrem em momentos de variação abrupta da carga ou aceleração rápida — cenários em que o modelo tem menor representatividade no conjunto de treinamento.

Em média, a largura do intervalo de confiança (IC) com 95% foi de:

- Modelo sem grid search: $\pm 2,784$ gCO₂/s;
- Modelo pós Grid Search: $\pm 0,7$ gCO₂/s.

Os resultados indicam que o softsensor baseado em LSTM apresenta boa acurácia mesmo em trajetos com variações intensas de condução. A análise de incertezas permite identificar momentos de menor confiança do modelo, sendo uma ferramenta valiosa para aplicações em sistemas críticos. Além disso, o uso de dados OBD-II demonstrou ser uma abordagem prática e replicável, com potencial para aplicações em frotas urbanas e sistemas embarcados de monitoramento ambiental.

VI. CONCLUSÕES

Os resultados indicam que o softsensor baseado em LSTM apresenta boa acurácia mesmo em trajetos com variações intensas de condução. A análise de incertezas permite identificar momentos de menor confiança do modelo, sendo uma ferramenta valiosa para aplicações em sistemas críticos. Além disso, o uso de dados OBD-II demonstrou ser uma abordagem prática e replicável, com potencial para aplicações em frotas urbanas e sistemas embarcados de monitoramento ambiental.

Os resultados demonstraram que o modelo proposto atinge elevados níveis de acurácia. A aplicação do Monte Carlo Dropout permitiu a geração de intervalos de confiança em tempo real, evidenciando que a incerteza do modelo aumenta em cenários com maior variabilidade dinâmica — como acelerações bruscas e variações de carga — o que corrobora a importância da análise probabilística para aplicações em diagnósticos automotivos e sistemas embarcados. Como limitações, reconhece-se que o modelo depende da disponibilidade e qualidade das variáveis medidas via OBD-II, que podem variar conforme fabricante e protocolo do veículo. Além disso, a generalização para diferentes tipos de motorização (diesel, híbrido, flex) requer expansão da base de dados e eventual reparametrização do modelo. Como trabalhos futuros, propõe-se a aplicação do modelo em frotas comerciais com conectividade em nuvem, a integração com dados geoespaciais (GPS e relevo) para melhor contextualização das emissões, e a investigação de arquiteturas híbridas LSTM-Attention para maior precisão e uso da técnica de Monte Carlo Dropout para maior interpretabilidade da incerteza nas estimativas.

REFERÊNCIAS

- [1] Ministério da Ciência, Tecnologia e Inovações (MCTI), "Sistema de Estimativa de Emissões de Gases de Efeito Estufa (SEEG)," 2023.
- [2] D. Torregrossa, et al., "Soft Sensors for Emission Estimation in Automotive Engines Using OBD-II Data," *Sensors*, v. 20, n. 22, 2020.
- [3] K. Greff, et al., "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, v. 28, n. 10, pp. 2222–2232, 2017.
- [4] Y. Gal e Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 2016.
- [5] Y. Chen, et al., "Fuel Consumption and Emission Estimation of Hybrid Vehicles Using LSTM Networks," *Applied Energy*, v. 298, 2021.
- [6] X. Zhou, et al., "Deep Learning for CO₂ Emission Prediction Based on OBD and GPS Data," *Transportation Research Part D: Transport and Environment*, v. 105, 2022.
- [7] H. Liu, et al., "Bayesian Convolutional Networks for Vehicle Emission Estimation under Uncertainty," *Neural Computing and Applications*, v. 35, 2023.
- [8] P. Kadlec, et al., "Soft Sensors for Process Monitoring: A Review," *Computers & Chemical Engineering*, v. 33, pp. 795–814, 2009.
- [9] T. Wenzel, et al., "OBD-II Vehicle Diagnostics: A Comprehensive Overview," *Transportation Research Board*, 2018.
- [10] Y. Bengio, et al., "Learning Long-Term Dependencies with Gradient Descent is Difficult," *IEEE Transactions on Neural Networks*, v. 5, n. 2, 1994.
- [11] P. Chapman, J. Clinton, T. Khabaza, T. Reinartz, C. Shearer, e R. Wirth, "CRISP-DM 1.0: Step-by-step data mining guide," *CRISP-DM Consortium*, 2000. Disponível em: <https://www.the-modeling-agency.com/crisp-dm.pdf>.
- [12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. Disponível em: <https://jmlr.org/papers/v12/pedregosalla.html>.
- [13] M. Singh, R. K. Dubey, et al., "Deep Learning Model Based CO₂ Emissions Prediction Using Vehicle Telematics Sensors Data" *IEEE Transactions on Intelligent Vehicles*, vol. 8, 2023.
- [14] M. Medeiros, M. Azevedo, M. Amaral, C. M. D. Viegas, M. Silva, I. Silva, D. G. Costa, "A Framework for Efficient Communication in IoT Applications in the Automotive Industry with OBD-II Edge" *2023 Symposium on Internet of Things (SIoT)*, 2023.