

Metodologia para Rotulação de Estilos de Condução via Dados de Telemetria Veicular: Uma Abordagem Híbrida com Clusterização e Métodos Classificatórios

Lorenzo B. S. Lopes¹, Gabriel G. Nobre², Vinicius A. O. Martins³, Gabriela N. Lopes¹,
Roberto G. Ribeiro⁴, Heitor S. R. Filho², Antônio A. F. Loureiro²

¹Dep. de Engenharia Elétrica, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil

²Dep. de Ciência da Computação (DCC), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil

³Prog. de Pós Graduação em Eng. Elétrica da Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil

⁴Dep. de Computação e Sistemas (DECSI), Universidade Federal de Ouro Preto (UFOP), João Monlevade, Brasil

lorenzobicalho@hotmail.com, gabrielgnobre7@gmail.com, viniciusomartins@gmail.com,

nuneslopesgabriela@gmail.com, roberto.gomes@ufop.edu.br, ramossh@dcc.ufmg.br, loureiro@dcc.ufmg.br

Resumo—A análise das práticas de condução veicular tem ganhado relevância crescente devido à sua ampla aplicabilidade em sistemas inteligentes de transporte, especialmente no que se refere à eficiência energética no setor automotivo. Nesse cenário, o desenvolvimento de métodos eficazes para rotular e classificar estilos de condução é essencial. Este estudo propõe uma metodologia para atribuição de rótulos a estilos de condução veicular, utilizando dados de telemetria automotiva coletados via barramento Controller Area Network (CAN). A abordagem integra o agrupamento não supervisionado K-Means com múltiplas estratégias de classificação, incluindo um modelo de lógica fuzzy, um processo determinístico e um método baseado em indicadores estatísticos. Os resultados demonstram que a combinação dessas técnicas proporciona uma interpretação mais precisa e robusta dos estilos de condução. Essa validação cruzada entre as técnicas reforça o potencial de abordagens não supervisionadas na identificação de perfis de condutores, com aplicações relevantes para o setor automotivo, especialmente aquelas voltadas à sustentabilidade e ao uso mais racional de recursos.

Palavras-chave—Classificação de estilos de condução; Agrupamento K-Means; Telemetria automotiva; Eficiência energética.

Abstract—The analysis of driving behavior has gained increasing importance due to its wide applicability in intelligent transportation systems, particularly in enhancing energy efficiency in the automotive sector. In this context, the development of effective methods for labeling and classifying driving styles is essential. This study proposes a methodology for labeling vehicle driving styles using automotive telemetry data collected via the Controller Area Network (CAN) bus. The approach integrates unsupervised K-Means clustering with multiple classification strategies, including a fuzzy logic model, a deterministic method, and a statistical indicators-based method. The results demonstrate that integrating these techniques yields a more precise and robust interpretation of driving styles. This cross-validation highlights the strong potential of unsupervised methods for identifying driver profiles, with significant implications for the automotive industry, particularly in advancing sustainability and enhancing resource efficiency.

Index Terms—Driving style classification; K-Means clustering; Automotive telemetry; Energy efficiency.

I. INTRODUÇÃO

Questões relacionadas à energia e às mudanças climáticas têm ganhado destaque global, especialmente após a definição, em setembro de 2015, dos 17 Objetivos de Desenvolvimento Sustentável (ODS) pela Organização das Nações Unidas (ONU). Para atingir tais metas, é necessário o comprometimento de diferentes setores, com destaque para o setor energético, uma vez que a energia é elemento fundamental para todas as atividades [1]. Ainda em 2015, a adoção do Acordo de Paris durante a Conferência das Nações Unidas sobre Mudanças Climáticas reafirmou a urgência de ações coordenadas no enfrentamento do aquecimento global [2].

No contexto nacional, o Brasil tem buscado alinhar sua indústria automotiva às metas globais de sustentabilidade por meio do Programa de Mobilidade Verde e Inovação (Mover), sancionado em junho de 2024 como substituto do Rota 2030. A iniciativa amplia as exigências ambientais para o setor automotivo, promovendo a adoção de tecnologias mais limpas e eficientes [3].

Diante das limitações das tecnologias tradicionais de economia de combustível, dos desafios para mitigar congestionamentos urbanos e da ainda baixa adesão de veículos movidos a energias alternativas, o eco-driving surge como uma alternativa viável e de rápida implementação para promover a eficiência energética e a redução de emissões no setor de transportes no curto prazo [4]. O termo eco-driving refere-se a um conjunto de práticas que visam otimizar o consumo de energia durante a condução de veículos. Segundo [5], trata-se de um processo de tomada de decisão que envolve, entre outros aspectos, o comportamento do condutor ao volante.

Nesse cenário, o estudo e a aplicação de perfis de condução vêm ganhando relevância por sua ampla aplicabilidade em

sistemas inteligentes de transporte. As informações sobre o estilo de direção dos motoristas podem ser utilizadas em iniciativas como gestão de frotas, seguros personalizados, programas de eco-driving e intervenções para segurança viária. Além disso, essas aplicações favorecem mudanças comportamentais por meio de feedbacks, avaliações de infraestrutura viária, treinamentos corporativos e estratégias de decisão em diferentes níveis — do estratégico ao operacional [6].

Considerando o impacto significativo do comportamento de condução no consumo de energia, nas emissões veiculares e na segurança viária, torna-se crucial desenvolver métodos eficazes para classificar e analisar os estilos de condução. Um dos principais desafios nessa área é a ausência de dados rotulados, especialmente quando se trabalha com conjuntos de dados de telemetria coletados sem rotulagem prévia. A rotulagem manual, embora precisa, é altamente impraticável para grandes volumes de dados devido ao tempo e esforço extensivos necessários. Conforme destacado em [7], essa limitação impulsiona a necessidade de métodos automáticos de rotulagem que sejam capazes de categorizar o comportamento de condução de forma eficiente e consistente.

A rotulagem automática é geralmente realizada por meio de técnicas de aprendizado de máquina, métodos baseados em heurísticas ou pela combinação de ambos. Dentre essas abordagens, os algoritmos de agrupamento — em especial o K-Means [8] — são frequentemente adotados devido à sua simplicidade e eficácia na identificação de padrões sem a necessidade de rótulos prévios. No entanto, estudos anteriores utilizam um conjunto limitado de variáveis ou focam em aspectos muito específicos do comportamento de condução, o que restringe a generalização e a abrangência da análise.

Neste contexto, o principal objetivo deste trabalho é desenvolver e avaliar uma metodologia para a rotulagem de comportamentos de condução utilizando o algoritmo de agrupamento K-Means aplicado a dados de telemetria automotiva coletados da rede CAN (Controller Area Network). A metodologia combina múltiplas abordagens visando alcançar uma caracterização abrangente e robusta dos estilos de condução. Diferentemente da maioria dos trabalhos existentes — que frequentemente dependem de um conjunto restrito de sinais ou aplicam técnicas isoladas —, este estudo integra o agrupamento não supervisionado com três métodos consolidados na literatura: (i) modelos de lógica fuzzy baseados nas acelerações lateral e longitudinal; (ii) análise estatística utilizando jerk (derivada da aceleração) para capturar variações dinâmicas no comportamento de condução; e (iii) regras determinísticas baseadas em limiares empíricos para variáveis-chave como rotação do motor (RPM), posição do acelerador e ângulo do volante. Esta abordagem integrada permite a identificação de padrões de condução sob múltiplas perspectivas, suprimindo lacunas da literatura e viabilizando uma rotulagem ampla e confiável do comportamento dos condutores.

Ao analisar como essas metodologias se alinham, se complementam ou divergem, o estudo busca não apenas comparar seus resultados, mas também utilizá-los como ferramenta para caracterizar e validar os agrupamentos gerados pelo K-Means.

Esse processo, em última instância, viabiliza uma rotulagem robusta dos estilos de condução — calmo, normal e agressivo —, proporcionando uma compreensão ampla dos padrões de comportamento dos motoristas e reforçando o potencial do aprendizado não supervisionado para essa aplicação.

O restante deste artigo está organizado da seguinte forma: na Seção II apresenta-se a revisão da literatura, contextualizando os métodos existentes para classificação de estilos de condução. A Seção III detalha o conjunto de dados utilizado, as variáveis extraídas, os métodos de processamento e a aplicação dos algoritmos de classificação. Na Seção IV são apresentados os resultados obtidos. Por fim, na Seção V são discutidos os resultados, suas implicações e limitações, seguidos das conclusões na Seção VI.

II. REVISÃO DA LITERATURA

A análise do comportamento de motoristas e da eficiência operacional dos veículos depende diretamente da qualidade e diversidade dos dados coletados. Diversos dispositivos e métodos têm sido utilizados na literatura para essa finalidade, variando em custo, confiabilidade e aplicabilidade.

Além dos sensores, diversos métodos de coleta e avaliação de comportamento de condutores foram discutidos por [9], incluindo questionários autodeclarados, simuladores de direção, avaliação por especialistas, registradores embarcados e sensores de smartphones. Cada um desses métodos possui vantagens e desvantagens específicas. Por exemplo, questionários são de baixo custo, mas altamente sujeitos à memória e percepção do motorista; simuladores permitem a simulação de cenários críticos, mas não garantem realismo; enquanto sensores embarcados fornecem dados em tempo real, porém com custo elevado e possibilidade de manipulação.

Estudos recentes demonstram o potencial dos dados provenientes via CAN para análise do comportamento veicular e classificação de estilos de direção. A rede CAN é um protocolo de comunicação serial amplamente utilizado na indústria automotiva, permitindo a troca eficiente e segura de dados entre diferentes unidades de controle eletrônico, sensores e atuadores do veículo [10]. Com taxas de transmissão de até 1 Mbit/s, a rede veicular CAN possibilita o controle distribuído em tempo real, sendo integrada tanto em sistemas críticos como o controle do motor e freios ABS quanto em funções auxiliares, como acionamento de janelas e luzes.

Dentre as principais vantagens do uso de dados da rede veicular CAN estão a alta frequência de amostragem, a padronização parcial dos sinais, a não intrusividade na condução do motorista e a riqueza de informações sobre o funcionamento do veículo. No entanto, as limitações incluem a variação entre fabricantes quanto à codificação de sinais, a necessidade de hardware específico (como interfaces OBD-II e microcontroladores programáveis) e o desafio de mapear corretamente os identificadores da CAN para sinais interpretáveis.

Fugiglando et al. [11] realizaram uma das primeiras análises em larga escala utilizando dados da CAN, coletando 2418 sinais distintos durante 55 dias com 64 motoristas em condições reais e não controladas. O estudo evidenciou a

viabilidade de capturar padrões de direção a partir dos dados da rede CAN em ambientes urbanos.

A classificação de estilos de direção é uma etapa fundamental para a análise comportamental de motoristas, especialmente em contextos que envolvem segurança viária e eficiência energética. A literatura propõe diversas formas de categorizar os estilos de condução, variando em número de categorias, critérios utilizados e contexto de aplicação (como ambientes urbanos, rodoviários ou simulados).

De forma geral, a maioria dos estudos adota uma classificação em três grupos principais: agressivo, normal e cauteloso [12]–[14]. Essa abordagem busca representar um espectro de comportamentos que varia desde manobras arriscadas e impulsivas até condução defensiva e previsível.

A identificação de eventos de direção agressiva é essencial para avaliar o comportamento do condutor e seu impacto na segurança e na eficiência energética. Estudos indicam que variáveis como velocidade, aceleração, frenagem, localização e horários de condução estão entre os principais parâmetros para análise de risco de acidentes [6]. Entre os eventos monitorados para detectar comportamentos agressivos estão frenagens bruscas, acelerações rápidas e curvas fechadas [15]–[17], mudanças de faixa sem sinalização [17], [18], velocidade excessiva e variações rápidas de velocidade [18], [19]. Também são utilizados indicadores como aceleração lateral e longitudinal [20] e *jerk* (derivada da aceleração) [21], além de dados extraídos da rede CAN, como posição dos pedais, rotação do motor e ângulo do volante [11]. A construção de índices compostos inclui ainda variáveis como taxa de inércia e gradiente do pedal do acelerador [19]. A combinação desses elementos permite classificar automaticamente estilos de direção, auxiliando modelos preditivos, sistemas de pontuação e feedback em tempo real.

Entre os métodos de classificação de comportamento de motoristas explorados na literatura, destacam-se abordagens como a análise de clusters para categorização de condutores e a análise do perfil de *jerk* — variações abruptas na aceleração — como proposta por [6], que combina informações estatísticas do *jerk* com o tipo de via e o nível de congestionamento para classificar o estilo de condução. No campo da lógica fuzzy, [21] propuseram um sistema de inferência que utiliza como entradas o valor normativo da aceleração longitudinal e lateral, além da velocidade, permitindo a classificação instantânea dos estilos de direção com baixo custo computacional.

III. METODOLOGIA

Este trabalho propõe uma metodologia para rotulagem e análise de estilos de condução com base em dados veiculares. Inicialmente, foi realizada a seleção de variáveis relevantes de uma base pública de dados obtidos via barramento CAN, seguida da extração de variáveis derivadas, como *jerk*, aceleração normalizada e deflexão do volante, que enriquecem a análise do comportamento de direção. Em seguida, aplicou-se o algoritmo K-Means para agrupar os dados em padrões de condução sem a necessidade de rótulos prévios.

Para interpretar os clusters formados, os dados também foram classificados por três abordagens supervisionadas inspiradas na literatura: uma baseada em regras determinísticas, outra utilizando lógica fuzzy e uma terceira fundamentada na análise estatística do *jerk*. Foi então proposto um método para consolidar essas classificações, agrupando os segmentos com base na concordância de pelo menos duas das três abordagens. A análise das médias das variáveis permitiu compreender as características de cada cluster e validar a coerência dos grupos encontrados, possibilitando a atribuição de perfis de condução como calmo, normal e agressivo.

A. Coleta de Dados

O conjunto de dados utilizado neste trabalho foi originalmente coletado por [22], a partir de um ambiente real de condução, com o objetivo de capturar variáveis que não poderiam ser completamente representadas em simulações. Para isso, foi utilizada uma combinação entre o sistema OBD-II (On Board Diagnostics) e uma ferramenta de captura para extração de características, permitindo extrair dados diretamente da rede CAN dos veículos.

A coleta compreendeu cerca de 23 horas de condução, totalizando aproximadamente 46 km de percurso (ida e volta) entre a Universidade da Coreia e o Estádio da Copa do Mundo em Sangam. Participaram da coleta 10 motoristas diferentes, rotulados de A a J, sendo que os dados foram registrados em intervalos de um segundo ao longo das viagens. O veículo utilizado pertence a um modelo recente da KIA Motors Corporation, e, ao todo, foram extraídas 51 variáveis distintas associadas ao comportamento de condução.

Entre os diversos sinais transmitidos pela rede CAN do veículo, foram selecionadas apenas as variáveis mais relevantes para as análises deste trabalho, com o objetivo de reduzir a dimensionalidade e o ruído da base de dados. As colunas escolhidas refletem diretamente o comportamento de condução, sendo elas: consumo de combustível, valor instantâneo relacionado ao uso de combustível; posição do pedal do acelerador, obtida por meio do sensor de posição da borboleta; rotação do motor, em rotações por minuto (RPM); velocidade do veículo, medida instantaneamente; aceleração longitudinal, ao longo do eixo principal do veículo; aceleração lateral, associada à realização de curvas; e ângulo do volante, correspondente ao grau de esterçamento.

Sinais como pedais e volante refletem diretamente as ações do condutor, enquanto variáveis como velocidade, rotação do motor e acelerações expressam sensações perceptíveis durante a condução. Esses parâmetros podem indicar padrões de direção específicos, como tendência a acelerar bruscamente, conduzir em alta velocidade ou realizar curvas agressivas.

Uma coluna de identificação única (ID) foi adicionada para cada amostra do conjunto. Além disso, a coluna de tempo foi utilizada para segmentar o conjunto de dados em sessões independentes — denominadas *runs* —, correspondentes a cada viagem registrada. Essa separação foi baseada em reinicializações do índice temporal. A extração de atributos foi realizada iterativamente por *run*.

B. Extração de características (Features)

Foram extraídas variáveis adicionais relevantes para a análise do comportamento do condutor.

1) *Aceleração normalizada (acc_norm)*: A aceleração normalizada é uma métrica derivada da composição vetorial das acelerações longitudinal e lateral do veículo, representando a força g total experimentada em cada instante. De acordo com [20], sensores acelerômetros de dois eixos podem ser utilizados para capturar essa grandeza, permitindo identificar padrões de direção ao comparar diferentes estilos de condução, como partidas e frenagens bruscas. O estudo mostra que a aceleração combinada é um bom indicador para classificar o desempenho do condutor.

A norma da aceleração em um instante n é calculada conforme (1).

$$acc_norm(n) = \sqrt{Accel_{long}^2(n) + Accel_{lat}^2(n)} \quad (1)$$

Essa métrica sintetiza os efeitos dinâmicos associados à movimentação do veículo em curvas, acelerações e frenagens, e é particularmente útil para detectar comportamentos agressivos na condução.

2) *Deflexão do volante*: O estudo de [23] realizou testes práticos com motoristas para analisar o tempo necessário de variação no ângulo de direção em diferentes contextos de direção. Os resultados mostram que a deflexão do volante é uma métrica significativa para distinguir estilos de condução, sendo utilizada como parâmetro na classificação do comportamento do condutor.

A deflexão do volante foi definida como a variação absoluta entre valores consecutivos do ângulo de direção (2), em que Δa representa a diferença no ângulo de direção e Δt o intervalo de tempo entre duas medições sucessivas.

$$K = \frac{\Delta a}{\Delta t} = \frac{|a_2 - a_1|}{t_2 - t_1} \quad (2)$$

3) *Jerk*: Jerk é definido como a taxa de variação da aceleração, ou seja, a derivada de segunda ordem da velocidade em relação ao tempo. Essa métrica é particularmente relevante para a análise do conforto e da segurança na condução, pois mudanças bruscas na aceleração (valores elevados de jerk) podem causar desconforto aos passageiros e aumentar o risco de acidentes [21], [24].

Neste trabalho, o jerk foi estimado por meio da segunda diferença da velocidade do veículo (convertida para m/s), conforme (3).

$$J(t) = \frac{d^2 v(t)}{dt^2} \quad (3)$$

Inspirados na abordagem de [21], foi utilizado o coeficiente de variação γ como métrica, definido como a razão entre o desvio padrão do jerk (SD_J), calculado dentro da janela móvel, e o valor médio de referência (J_{ref}), que corresponde ao jerk do estilo de condução normal no tipo de via em que o motorista se encontra.

Para fins de classificação, o estilo de condução do motorista no instante atual t_e é classificado com base nas seguintes características do jerk, extraídas dentro do intervalo de tempo $t \in [(t_e - \omega), t_e]$, onde t_e representa o instante atual e ω é o tamanho da janela de observação. Para cada instante de tempo t_e , o desvio padrão foi calculado dentro de uma janela móvel de 10 segundos.

$$\gamma = \frac{SD_J}{J_{ref}} \quad (4)$$

Como os dados foram coletados em ambientes reais e não controlados, não foi possível associar cada segmento de trajeto a um tipo específico de via. Assim, foi adotado o valor médio global de jerk estimado por [24], igual a $0,2732 \text{ m/s}^3$, como referência para normalização.

C. K-means

Após a obtenção das features adicionais, procedeu-se com a análise de agrupamento utilizando o algoritmo K-Means, com o objetivo de identificar padrões emergentes nos dados de direção. Além disso, foram removidas todas as entradas em que a velocidade apresentava valor igual a zero, por não representarem efetivamente trechos de condução.

Para a análise de clusters, foram selecionadas as seguintes variáveis numéricas: consumo de combustível, valor instantâneo relacionado ao consumo; posição relativa do pedal do acelerador, obtida por meio do sensor de posição da borboleta; rotação do motor em RPM (rotações por minuto); velocidade instantânea do veículo; aceleração composta, que corresponde à composição vetorial das acelerações longitudinal e lateral do veículo; e ângulo de deflexão, definido como a variação absoluta entre valores consecutivos do ângulo de direção. Estas variáveis foram escolhidas por representarem aspectos distintos, porém complementares, do comportamento de condução.

Os dados foram normalizados utilizando Z-score (subtração da média e divisão pelo desvio padrão), garantindo que todas as variáveis contribuíssem igualmente para o cálculo das distâncias no espaço multidimensional do K-Means.

O algoritmo K-Means foi aplicado com $k = 3$ clusters e 10 replicações, buscando garantir maior robustez e minimizar o risco de convergência a mínimos locais. A escolha de três clusters foi fundamentada em análises exploratórias prévias, estudos da literatura [11] e na coerência com o número de classes utilizado nas outras abordagens de classificação.

D. Análise Cruzada com Classificações Independentes

Com os dados previamente agrupados, tornou-se necessário definir rótulos representativos para cada cluster. Para isso, foram empregados três sistemas independentes de classificação, possibilitando uma análise cruzada entre os agrupamentos obtidos por K-Means e as classes atribuídas por métodos supervisionados. Essa abordagem permitiu avaliar o grau de alinhamento ou divergência entre técnicas supervisionadas e não supervisionadas, além de evidenciar características predominantes em cada grupo.

1) *Classificação Fuzzy*: Seguindo a proposta de [20], foi implementado um Sistema de Inferência Fuzzy (FIS) com duas variáveis de entrada: aceleração combinada (ou normalizada) e velocidade do veículo. O sistema classifica as amostras em três categorias: Calmo, Normal e Agressivo. Ambas as variáveis refletem o comportamento de velocidade, tanto em trechos retilíneos quanto em curvas, indicando tendências como excesso de velocidade ou acelerações bruscas.

A Fig. 1 apresenta a série temporal da velocidade e da aceleração normalizadas para uma viagem do motorista A. Já a Fig. 2 mostra a relação entre velocidade, aceleração e o grau de pertinência calculado pelo sistema fuzzy, em que o azul escuro representa uma condução calma e o vermelho escuro, uma condução agressiva.

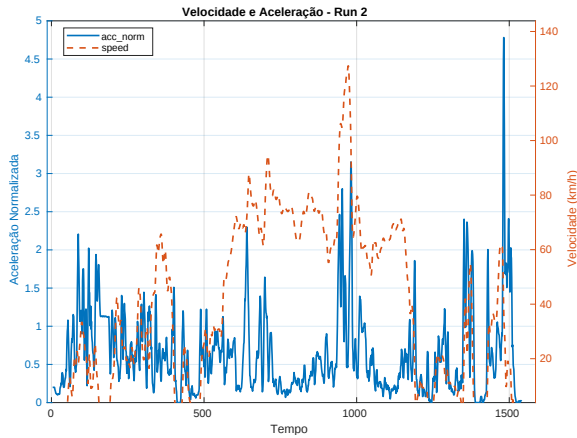


Fig. 1. Série temporal: vel. x acel. em viagem do motorista A.

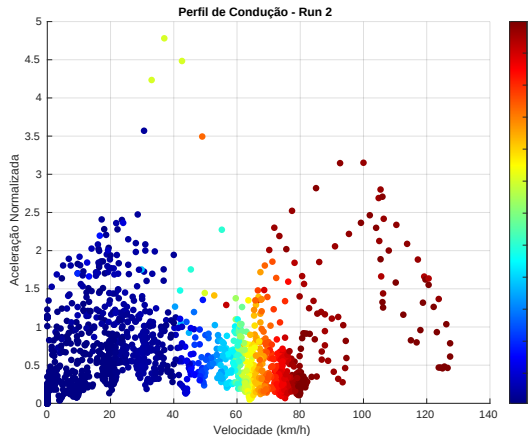


Fig. 2. Relação entre velocidade, aceleração e grau de pertinência fuzzy, onde azul escuro indica condução calma e vermelho escuro, condução agressiva.

2) *Classificação Determinística*: Inspirado na abordagem de [23], foi realizada uma classificação binária baseada em regras simples definidas por limiares em três variáveis: posição do pedal do acelerador — normal (≤ 38), agressivo (> 38); rotação do motor em RPM (rotações por minuto) — normal (≤ 3300), agressivo (> 3300); e deflexão do volante — normal (< 355), agressivo (≥ 355).

Esses parâmetros refletem situações de uso intensivo do motor, pressionamento acentuado do pedal e movimentos bruscos do volante, frequentemente associados a ultrapassagens, mudanças de faixa repentinas ou curvas fechadas.

A Fig. 3 apresenta a relação entre as variáveis RPM, posição do pedal do acelerador, deflexão do volante, aceleração e o score de agressividade calculado pelo classificador determinístico. Esse score foi obtido por meio de uma média ponderada das variáveis normalizadas de posição do acelerador, rotação do motor e deflexão do volante, sendo atribuídos pesos iguais de 0,33 para cada uma. As normalizações foram realizadas considerando os valores máximos definidos para cada variável — 100 para a posição do pedal, 6000 para RPM e 500 para deflexão.

Cada variável contribui para o score apenas quando ultrapassa um limiar, definido com base em critérios de domínio: 38 para a posição do pedal, 3300 para RPM e 355 para deflexão. Assim, o score assume valor zero quando nenhuma dessas variáveis ultrapassa os respectivos limiares, classificando a amostra como “Normal”. Caso contrário, qualquer contribuição diferente de zero resulta na classificação como “Agressivo”.

A visualização permite observar como a combinação dessas variáveis influencia diretamente na definição do perfil de condução no modelo determinístico, destacando as regiões do espaço de variáveis que mais concentram os comportamentos classificados como agressivos.

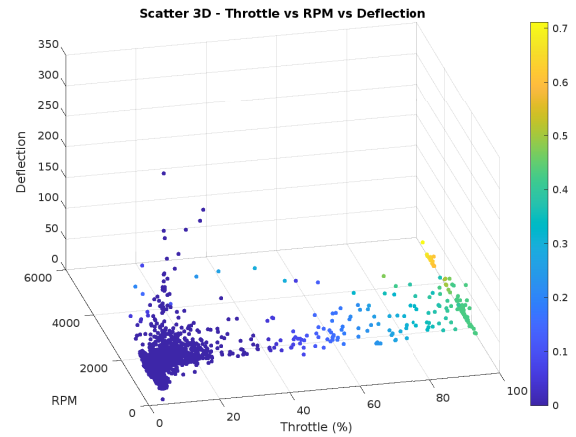


Fig. 3. Relação entre RPM, pedal do acelerador, deflexão do volante e aceleração no cálculo do score de agressividade do classificador determinístico.

3) *Classificação por Jerk*: Com base na métrica γ , definida pela razão entre o desvio padrão do jerk (SD_J) e um valor de referência (J_{ref}), o estilo de condução foi classificado conforme os intervalos sugeridos por [21]: calmo ($\gamma < 0.5$), normal ($0.5 \leq \gamma < 1$) e agressivo ($\gamma \geq 1$).

A Fig. 4 apresenta uma série temporal da velocidade e dos resultados de classificação utilizando o método baseado em jerk, aplicado a uma janela deslizante de aproximadamente 10 segundos sobre um trecho de cerca de 200 segundos da condução de um motorista. A cada instante t , a classificação é

exibida apenas quando há uma mudança em relação à classe anterior. Na figura, é possível observar como variações bruscas de aceleração, representadas pelo jerk, estão diretamente associadas às mudanças nos perfis de condução ao longo do trajeto.

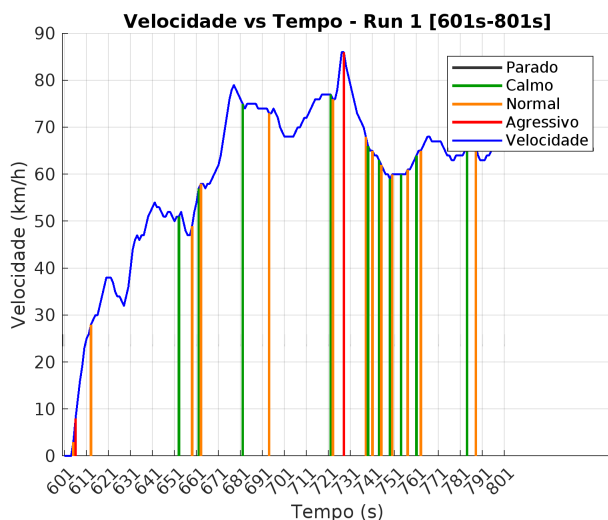


Fig. 4. Série temporal da velocidade e da classificação do estilo de condução utilizando o método baseado em jerk, com janelas de aproximadamente 10 segundos, ao longo de 200 segundos de uma viagem.

4) *Agrupamento por Concordância de Classificações*: Com base nas classificações obtidas pelos três métodos, foi realizada uma categorização qualitativa das amostras segundo a concordância entre os classificadores. A cada amostra foi atribuído um identificador de grupo (1, 2 ou 3), com base nas seguintes combinações: Grupo 1 (Calmo) — Determinístico = Normal, Fuzzy = Calmo, Jerk = Calmo; Grupo 2 (Normal) — todos os três classificadores indicam "Normal"; Grupo 3 (Agressivo) — todos os três classificadores indicam "Agressivo".

O agrupamento foi considerado válido quando ao menos duas das três classificações apresentaram concordância. Nos casos sem nenhuma convergência entre os métodos, a amostra foi atribuída ao **Grupo 0**. Para essas instâncias, a classificação final foi inferida por validação cruzada com os clusters obtidos anteriormente.

IV. RESULTADOS

TABELA I
RESUMO ESTATÍSTICO POR GRUPO COM MÉDIAS DAS VARIÁVEIS SELECIONADAS

Grupo	Cont.	%	Combustível	Velocidade
0	1560	2.13	2663.2	44.574
1	18409	25.18	743.51	25.285
2	51301	70.16	795.1	39.768
3	1846	2.52	3163.8	63.085

Grupo	Acc_Norm	Pedal	RPM	Deflexão
0	0.90409	76.297	2280.6	7.3435
1	0.64109	11.193	1475.6	10.739
2	0.54096	12.484	1634.4	6.3667
3	1.0899	74.344	2723.1	8.8732

Nesta seção, são apresentados os principais resultados obtidos a partir da análise dos dados de condução, considerando tanto os métodos de classificação utilizados quanto os agrupamentos gerados.

A Tabela I apresenta um resumo estatístico das principais variáveis analisadas, segmentadas por grupos definidos a partir da concordância entre os três classificadores independentes (fuzzy, determinístico e jerk). Essa tabela permite visualizar padrões médios de comportamento associados aos perfis de condução identificados.

TABELA II
COINCIDÊNCIA ENTRE CLASSIFICADORES

Tipo de coincidência	Contagem	Percentual
Coincidência total (3/3)	17 234	23.57
Coincidência parcial (2/3)	54 322	74.30
Nenhuma coincidência (0/3)	1 560	2.13

TABELA III
DISTRIBUIÇÃO DE GRUPOS POR CLUSTER (ANÁLISE CRUZADA)

Cluster	Grupo	Contagem	Percentual
1	0	18	0.06
	1	11 678	36.00
	2	20 701	63.82
	3	39	0.12
2	0	136	0.38
	1	6 462	17.88
	2	29 326	81.17
	3	207	0.57
3	0	1 406	30.91
	1	269	5.91
	2	1 274	28.01
	3	1 600	35.17

A coluna Contagem mostra o número de registros atribuídos a cada grupo, e Percentual representa a proporção desses registros em relação ao total. As colunas Combustível, Velocidade, Acc_Norm, Pedal, RPM e Deflexão apresentam, respectivamente, as médias dos principais indicadores de condução para cada grupo: consumo de combustível, velocidade média, aceleração normalizada, nível de acionamento do pedal, rotação do motor e deflexão do volante. Esses valores médios ajudam a caracterizar o estilo de condução predominante em cada grupo, permitindo distinguir perfis, por exemplo, mais econômicos, mais agressivos ou mais conservadores.

A Tabela II detalha a coincidência entre os classificadores, evidenciando a taxa de concordância total, parcial ou nenhuma entre os métodos empregados. Além disso, são exploradas separadamente as coincidências por tipo de classificação (Calmo, Normal e Agressivo), o que permite entender a robustez e a sobreposição entre os critérios adotados.

A Tabela III realiza uma análise cruzada entre os clusters gerados pelo algoritmo K-Means e os grupos definidos com base nas classificações independentes. Essa correspondência oferece uma visão da coerência entre métodos supervisionados e não supervisionados, além de auxiliar na interpretação dos agrupamentos automáticos.

TABELA IV
MÉDIAS DAS VARIÁVEIS POR CLUSTER

Cluster	Contagem	Percentual	Combustível	Velocidade
1	32 436	44.36	461.01	16.88
2	36 131	49.41	985.90	52.36
3	4 549	6.22	3054.9	55.42
Parâmetros adicionais por Cluster				
Cluster	Acc_Norm	Pedal	RPM	Deflexão
1	0.701	8.25	1197.7	13.52
2	0.442	13.33	1915.4	2.48
3	0.939	77.72	2538.2	5.28

Por fim, a Tabela IV apresenta as médias das mesmas variáveis da Tabela I, agora calculadas para cada cluster identificado. Essa apresentação permite uma análise comparativa dos perfis típicos de cada grupo, de acordo com os padrões aprendidos pelo algoritmo de agrupamento.

Além da análise de coincidência geral, foi realizada uma avaliação individual para cada estilo de condução — Agressivo, Normal e Calmo — com base na concordância entre os três classificadores independentes.

Para o estilo Agressivo, observou-se que em apenas 85 casos (0,12%) houve total concordância entre os classificadores. Em 1.761 observações (2,41%), dois classificadores apontaram o estilo como agressivo, enquanto em 17.463 registros (23,88%) apenas um dos sistemas indicou esse padrão. A maioria dos dados (53.807 observações, correspondendo a 73,59%) não foi rotulada como agressiva por nenhum dos classificadores.

No caso do estilo Normal, houve concordância total em 10.583 observações (14,47%). Em 40.718 casos (55,69%), dois classificadores identificaram esse comportamento. Apenas um classificador apontou estilo normal em 20.748 observações (28,38%), enquanto em 1.067 registros (1,46%) nenhum dos métodos classificou a condução como normal.

Para o estilo Calmo, os três classificadores concordaram em 6.566 amostras (8,98%). Já em 48.501 observações (66,33%) houve concordância entre dois classificadores. Em 15.935 casos (21,79%) apenas um método apontou esse padrão, e em 2.114 observações (2,89%) não houve indicação de condução calma por nenhum dos classificadores.

V. DISCUSSÃO

Como a base de dados utilizada foi originalmente concebida para o desafio de identificação de motoristas, e não especificamente para classificação de estilo de condução, é esperado — e justificado — que a maior parte dos dados seja classificada como pertencente aos estilos "normal" ou "calmo", conforme pode ser observado na Tabela I. Esse padrão se mantém também na Tabela III, onde os grupos identificados como normal ou calmo predominam.

Na Tabela I, observa-se que o Grupo 0, correspondente aos casos de não concordância entre classificadores, representa cerca de 2% dos dados. Em termos de médias das variáveis, esse grupo se assemelha tanto ao Grupo 2 quanto ao Grupo 3, revelando sua natureza ambígua. Essa indefinição afeta especialmente o perfil agressivo: enquanto 23,88% dos dados

tiveram ao menos um classificador apontando para esse estilo, apenas 2,53% tiveram pelo menos dois classificadores em concordância com essa classificação. Já para os estilos normal e calmo, as taxas de concordância (duas ou mais classificações coincidentes) são significativamente mais altas: 70,16% e 75,31%, respectivamente.

Ainda na Tabela I, percebe-se que os casos em que dois classificadores indicam "calmo" somam 66,33%, e em que todos indicam "calmo", 8,98%. No entanto, o Grupo 1 (classificado como calmo) representa apenas cerca de 25% do total. Isso se deve ao fato de que o classificador determinístico só possui dois rótulos: normal e agressivo. Assim, em situações em que os classificadores fuzzy e jerk apontam "calmo" e "normal", ou vice-versa, o sistema de decisão priorizou a classificação como "normal". Isso explica por que o Grupo 2 (normal) acabou sendo significativamente maior do que o Grupo 1.

Esse comportamento também se reflete na Tabela III, na qual tanto o Cluster 1 quanto o Cluster 2 apresentam maior proporção de amostras do Grupo 2. Entretanto, há uma distinção: o Cluster 1 é mais equilibrado entre os Grupos 1 e 2, indicando que as observações do Grupo 2 nesse cluster apresentam características mais próximas do perfil calmo.

Embora o agrupamento por compatibilidade entre classificadores apresente resultados coerentes, ele possui algumas limitações importantes. Primeiramente, os itens classificados como Grupo 0 não apresentam um critério claro de decisão, o que compromete sua interpretabilidade. Além disso, a ausência da classe "calmo" no classificador determinístico dificulta a separação efetiva entre os estilos calmo e normal. O modelo também trata os classificadores com igual peso, ignorando que uma classificação isolada pode ser mais representativa do que outras duas mais incertas. A lógica de decisão utilizada é arbitrária, pois optar por critérios mais conservadores — em que uma única classificação como "agressivo" seja suficiente — ou mais restritivos — que exijam três concordâncias para definir a classe — pode levar a resultados extremos, como aumento de falsos positivos ou excesso de exclusões. Além disso o modelo é limitado às regras previamente estabelecidas, não sendo capaz de identificar novos padrões ou comportamentos não previstos pelos classificadores.

A utilização de agrupamento não supervisionado por *clusters* atenua essas limitações. Como pode ser observado nas Tabelas III e IV, o uso do K-Means permitiu redirecionar observações inicialmente classificadas como Grupo 0, além de refinar a separação entre os estilos. A presença de todos os grupos em todos os clusters evidencia que o modelo foi capaz de ajustar a sensibilidade dos classificadores iniciais e reconhecer novos padrões não contemplados anteriormente. A separação entre os estilos normal e calmo também foi melhor distribuída, enquanto a baixa quantidade de dados agressivos se manteve, o que reforça a justificativa inicial baseada na natureza da base de dados.

A análise das médias das variáveis por cluster mostra coerência com os rótulos atribuídos: o Cluster 3, associado ao estilo agressivo, apresenta os maiores valores médios em todas as variáveis — exceto na deflexão do volante. Este

último ponto é esperado, pois maiores deflexões costumam ocorrer em baixas velocidades, típicas de manobras calmas, como estacionamento ou conversões em baixa velocidade.

Destaca-se ainda a diferença no consumo de combustível: o estilo agressivo apresenta uma média de consumo aproximadamente três vezes maior que o estilo normal, e cerca de 6,6 vezes maior que o estilo calmo. Esse dado evidencia o impacto direto do estilo de condução sobre a eficiência energética, reforçando a relevância da análise e indicando um caminho promissor para futuras investigações e iniciativas de incentivo à condução econômica e segura.

VI. CONCLUSÃO

Este trabalho buscou explorar diferentes estratégias para a classificação de estilos de condução a partir de dados de telemetria veicular. Foram consideradas abordagens consolidadas na literatura, como modelos fuzzy aplicados às dinâmicas lateral e longitudinal, análise estatística baseada no jerk (variação da aceleração) e regras determinísticas fundamentadas em limites operacionais de parâmetros do veículo, tais como rotação do motor, posição do pedal do acelerador e deflexão do volante. A partir desse panorama, desenvolveu-se uma metodologia de agrupamento utilizando o algoritmo K-Means, com o objetivo de investigar sua capacidade de identificar perfis de condução de forma não supervisionada e avaliar como seus resultados se relacionam com as classificações obtidas pelas demais abordagens.

O trabalho demonstra a compatibilidade parcial entre os métodos clássicos presentes na literatura e destaca como podem ser utilizados como ferramenta auxiliar para a interpretação dos agrupamentos gerados. A análise cruzada permitiu validar os agrupamentos obtidos, além de oferecer uma visão mais robusta dos diferentes perfis de condução. Essa abordagem contribui para o avanço na compreensão do comportamento dos condutores, reforçando o potencial de técnicas não supervisionadas na detecção de padrões de condução, com impactos diretos em aplicações de segurança, eficiência energética e manutenção preditiva no setor automotivo.

A principal contribuição deste trabalho consiste em um conjunto estruturado de métodos e técnicas que amplia as possibilidades de classificação de estilos de direção. Com o uso da metodologia desenvolvida, métodos supervisionados de aprendizado de máquina podem ser treinados e aplicados em sistemas veiculares para essa finalidade. Como trabalho futuro, propõe-se o desenvolvimento de um sistema abrangente que integre a metodologia apresentada a modelos de treinamento e classificação em tempo real.

REFERÊNCIAS

- [1] Y. Simsek, W. G. Santika, M. Anisuzzaman, T. Urmee, P. A. Bahri, and R. Escobar, "An analysis of additional energy requirement to meet the sustainable development goals," *Journal of Cleaner Production*, vol. 272, p. 122646, 2020.
- [2] X. Pan, W. Chen, S. Zhou, L. Wang, J. Dai, Q. Zhang, X. Zheng, and H. Wang, "Implications of near-term mitigation on china's long-term energy transitions for aligning with the paris goals," *Energy economics*, vol. 90, p. 104865, 2020.
- [3] FUNDEP/UFGM, "Mover — mobilidade verde e inovação," <https://mover.fundep.ufmg.br/mover/>, 2025, acesso em: 6 maio 2025.
- [4] N. Xu, X. Li, Q. Liu, and D. Zhao, "An overview of eco-driving theory, capability evaluation, and training applications," *Sensors*, vol. 21, no. 19, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/19/6547>
- [5] M. S. Alam and A. McNabola, "A critical review and assessment of eco-driving policy technology: Benefits limitations," *Transport Policy*, vol. 35, pp. 42–49, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0967070X14001152>
- [6] H. Singh and A. Kathuria, "Profiling drivers to assess safe and eco-driving behavior – a systematic review of naturalistic driving studies," *Accident Analysis Prevention*, vol. 161, p. 106349, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0001457521003808>
- [7] S. Bouhsissin, N. Sael, and F. Benabbou, "Driver behavior classification: A systematic literature review," *IEEE Access*, vol. PP, 02 2023.
- [8] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.
- [9] M. Ghatee and H. Eftekhari, "Hybrid of discrete wavelet transform and adaptive neuro fuzzy inference system for overall driving behavior recognition," *Transportation Research Part F Traffic Psychology and Behaviour*, vol. 58, pp. 782–796, 07 2018.
- [10] Robert Bosch GmbH, *CAN Specification*, version 2.0 ed., Robert Bosch GmbH, Postfach 50, D-7000 Stuttgart 1, 1991, also known as the CAN 2.0 specification.
- [11] U. Fugiglando, E. Massaro, P. Santi, S. Milardo, K. Abida, R. Stahlmann, F. Netter, and C. Ratti, "Driving behavior analysis through can bus data in an uncontrolled environment," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, 10 2017.
- [12] D. Dörr, D. Grabengieser, and F. Gauterin, "Online driving style recognition using fuzzy logic," in *17th international IEEE conference on intelligent transportation systems (ITSC)*. IEEE, 2014, pp. 1021–1026.
- [13] B. Higgs and M. Abbas, "A two-step segmentation algorithm for behavioral clustering of naturalistic driving styles," in *16th international ieee conference on intelligent transportation systems (itsc 2013)*. IEEE, 2013, pp. 857–862.
- [14] L. Xu, J. Hu, H. Jiang, and W. Meng, "Establishing style-oriented driver models by imitating human driving behaviors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2522–2530, 2015.
- [15] M. Fazeen, B. Gozick, R. Dantu, M. Bhukhiya, and M. C. González, "Safe driving using mobile phones," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1462–1468, 2012.
- [16] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan, "Mobile phone based drunk driving detection," in *2010 4th international conference on pervasive computing technologies for healthcare*. IEEE, 2010, pp. 1–8.
- [17] R. Vaiana, T. Iuele, V. Astarita, M. V. Caruso, A. Tassitani, C. Zaffino, and V. P. Giofrè, "Driving behavior and traffic safety: an acceleration-based safety evaluation procedure for smartphones," *Modern Applied Science*, vol. 8, no. 1, p. 88, 2014.
- [18] L. James, *Road rage and aggressive driving: Steering clear of highway warfare*. Prometheus Books, 2009.
- [19] K. Kim, J. Park, and J. Lee, "Fuel economy improvement of urban buses with development of an eco-drive scoring algorithm using machine learning," *Energies*, vol. 14, p. 4471, 07 2021.
- [20] A. Aljaafreh, N. Alshabat, and M. Najim Al-Din, "Driving style recognition using fuzzy logic," in *2012 IEEE International Conference on Vehicular Electronics and Safety, ICVES 2012*, 07 2012, pp. 460–463.
- [21] Y. Murphey, R. Milton, and L. Kiliaris, "Driver's style classification using jerk analysis," in *2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems*, 05 2009, pp. 23 – 28.
- [22] B. I. Kwak, J. Woo, and H. K. Kim, "Know your master: Driver profiling-based anti-theft method," in *PST 2016*, 2016.
- [23] M. Karrouchi, I. Nasri, R. Mohammed, I. Atmane, K. Hirech, A. Mes-saoudi, M. Mustapha, and K. Kassmi, "Driving behavior assessment: A practical study and technique for detecting a driver's condition and driving style," *Transportation Engineering*, vol. 14, p. 100217, 11 2023.
- [24] L. G. de Almeida, "Driver's behavior classification in vehicular communication networks for commercial vehicles." Master's thesis, UNIVERSIDADE FEDERAL DE ITAJUBÁ - UNIFEI, Itajubá, 2023.