

Identificação do Comportamento de Motoristas: Uma Abordagem Baseada em Teoria da Informação

Micael S. Santos¹, Gean S. Santos¹, Andre L. L. Aquino¹

¹Instituto de Computação – Universidade Federal de Alagoas – Campus A. C. Simões
Cidade Universitária – 57072-970 – Maceió – AL – Brasil

{michael.santos, gean.santos, alla}@laccan.ufal.br

Resumo. Neste trabalho, propomos a identificação do comportamento do motorista com o uso do algoritmo Random Forest e Long Short-Term Memory (LSTM), baseado em medidas de teoria da informação, como Entropia de Shannon, Complexidade Estatística e Informação de Fisher. Os modelos LSTM e Random Forest foram aplicados em dados provenientes dos sensores acelerômetro e giroscópio em veículos. Tais dados foram rotulados como: *slow*, *normal*, e *aggressive*. Comparamos a metodologia padrão da literatura com a nossa proposta por meio das medidas de acurácia, área sob a curva ROC (AUC), e precisão. Seguindo a literatura obtivemos: com Random Forest 60 % de acurácia, 58 % de AUC, e 61 % de precisão; com LSTM 56 à 58 % de acurácia, 52 % de AUC, e 68 à 73 % de precisão. Seguindo nossa proposta obtivemos: com Random Forest 52 à 92 % de acurácia, 54 à 94 % de AUC, e 61 à 100 % de precisão; com LSTM 61 à 80 % de acurácia, 58 à 78 % de AUC, e 56 à 85 % de precisão.

Abstract. In this work, we propose the identification of driver behavior using the Random Forest and Long Short-Term Memory (LSTM) algorithms, based on information theory measures such as Shannon Entropy, Statistical Complexity, and Fisher Information. The LSTM and Random Forest models were applied to data from accelerometer and gyroscope sensors in vehicles. These data were labeled as: *slow*, *normal*, and *aggressive*. We compared the standard methodology in the literature with our proposed approach using accuracy, area under the ROC curve (AUC), and precision.

Following the literature, we obtained: with Random Forest, 60% accuracy, 58% AUC, and 61% precision; with LSTM, 56 to 58% accuracy, 52% AUC, and 68 to 73% precision. Following our proposed approach, we obtained: with Random Forest, 52 to 92% accuracy, 54 to 94% AUC, and 61 to 100% precision; with LSTM, 61 to 80% accuracy, 58 to 78% AUC, and 56 to 85% precision.

1. Introdução

Neste trabalho, propomos o uso do algoritmo Random Forest e LSTM baseado em medidas de teoria da informação, Entropia de Shannon [Nascimento and Prudente 2016], Informação de Fisher [Mateos et al. 2021], e Complexidade Estatística [López-Ruiz et al. 1995], para reconhecer o comportamento do motorista a partir de dados de direção veicular.

Trabalhos recentes tem utilizado novas medidas para melhorar o desempenho de modelos de classificação, especialmente as medidas de Teoria da

Informação [Aquino et al. 2015]. Dados de direção podem ser obtidos via simulador [Miyajima et al. 2007], OBD-II [Girma et al. 2019, Park and Kim 2019], sensores de smartphone (acelerômetro, GPS, etc.), e sensores fisiológicos [Schneegass et al. 2013], além de câmeras e microfones nos veículos [Wu and Ye 2009]. Optamos por utilizar dados do OBD-II, que fornecem informações já disponíveis no carro, sem a necessidade de adicionar sensores extras.

De acordo com [Manderna and Kumar 2022], a combinação de dados de telemática veicular com algoritmos de aprendizado de máquina permite reconhecer estilos de direção distintos, identificar comportamentos e padrões de direção, e detectar comportamentos perigosos. Aplicações como identificação de padrões, caracterização [Martinelli et al. 2020], previsão [Singh and Dubey 2023], e identificação do tipo de trajeto [Bernardi et al. 2018] também são possíveis. Estudos recentes agruparam dados para identificar perfis de motoristas e condições de estrada [Vaiti et al. 2022], utilizando tanto modelos clássicos de aprendizado de máquina [Uvarov and Ponomarev 2021, Xun et al. 2019] quanto modelos de aprendizado profundo [Girma et al. 2019, Park and Kim 2019].

Diferentemente de outras abordagens, nossa proposta se destaca pelo uso eficaz de métricas da Teoria da Informação em conjunto com o algoritmo Random Forest. Isso nos permite capturar de forma mais completa o comportamento intrínseco do sistema dinâmico representado pelos dados sensoriados, resultando em inferências de maior qualidade feitas pelo algoritmo. Nossa comparação entre Random Forest e LSTM mostrou que, enquanto o modelo LSTM é mais eficaz com uma janela de tamanho 180, o Random Forest obteve melhores resultados com uma janela de tamanho 480, reforçando a importância da escolha adequada dos parâmetros de janela para cada abordagem.

Contudo, é crucial reconhecer as limitações inerentes a este estudo. Um dos principais desafios reside na disponibilidade e qualidade dos dados obtidos. Caso os dados não sejam suficientemente representativos ou contenham ruído, a precisão do nosso modelo pode ser comprometida. Além disso, a questão do viés de seleção é pertinente, já que os dados coletados podem não ser aleatórios, potencialmente distorcendo os resultados.

Logo, este trabalho tem como objetivo ampliar os avanços recentes no reconhecimento do comportamento do motorista, enfatizando a importância das técnicas de processamento de dados e aprendizado de máquina. Além disso, propõe uma abordagem inovadora que combina algoritmos de aprendizado de máquina com medidas da Teoria da Informação para aprimorar a classificação e previsão de comportamentos de direção.

2. Identificação do Comportamento de Motoristas com Teoria da Informação

O conjunto de dados utilizado, possui uma sequência de dados da classe *Aggressive*, depois uma sequência da classe *Normal*, e então uma sequência da classe *Slow*. Primeiro, cada recurso (*feature* — coluna) foi separado nas três classes existentes para então serem divididos em subconjuntos de tamanho fixo (janelas). Cada janela é uma série temporal, e cada uma é deslocada uma amostra da sua seguinte, tendo assim a maior sobreposição possível para se obter o máximo de subconjuntos.

Para cada janela foi gerado o padrão de Ordinais de Bandt-Pompe [Bandt and Pompe 2002] para transformar os dados brutos em um histograma. Especificamente, o método de simbolização BP atribui distribuições de probabilidade a partir da série temporal em consideração, ou seja, a causalidade temporal do processo.

Nesse sentido, dada uma série temporal $\mathbf{X}(t) = \{x_t : t = 1, \dots, N\}$, uma dimensão de incorporação $D \geq 2 (D \in \mathbb{N})$, e um atraso de incorporação $\tau \in \mathbb{N}$, calculamos os padrões ordinais de ordem D (comprimento do padrão) gerados por $(s) \mapsto (x_{s-(D-1)\tau}, x_{s-(D-2)\tau}, \dots, x_{s-\tau}, x_s)$. Em seguida, atribuímos a cada ponto no tempo s um vetor D -dimensional resultante da avaliação da sequência no tempo $s - (D - 1)\tau, \dots, s - \tau, s$. Incorporamos mais informações sobre o passado no vetor ao considerar um valor de D mais alto. De acordo com o padrão, o significado da ordem D em relação ao tempo s é a permutação $\pi = \{r_0, r_1, \dots, r_{D-1}\}$ de $\{0, 1, \dots, D - 1\}$ é definido como $x_{5-r_{D-1}\tau} \leq x_{5-r_{D-2}\tau} \leq \dots \leq x_{5-r_1\tau} \leq x_{5-r_0\tau}$.

Portanto, convertemos os dados produzidos para o símbolo único π . Para obter resultados não ambíguos, definimos $r_i < r_{i-1}$ se $\chi_{s-r_i} = \chi_s r_{i-1}$. Se $X(t)$ seguir uma distribuição ligeiramente contínua, a probabilidade de valores iguais é zero. Portanto, calculamos as frequências relativas associadas para todos $D!$ e as possíveis permutações π de ordem D , nas quais esta sequência ordenada particular foi encontrada na série temporal, dividida pelo número total de sequências. Assim, o histograma $P \equiv \{p(\pi)\}$ é definido como

$$p(\pi) = \frac{\#\{s \text{ do tipo } \pi : s \leq N - (D - 1)\tau\}}{N - (D - 1)\tau}, \quad (1)$$

onde $\#$ é a cardinalidade do conjunto.

Com base no histograma $p(\pi)$ calculamos as medidas de teoria da informação. A primeira delas é a Entropia de Shannon, uma medida global de autoinformação. Seja $\mathcal{X} = \{x_j : j = 1, \dots, M\}$ uma variável aleatória discreta de comprimento $M < \infty$ cujas características de distribuição são descritas pela função de probabilidade $P = \{p_i : i = 1, \dots, M\}$. p_i representa a probabilidade do estado i , e $\sum_{i=1}^M p_i = 1$, e M é o número de estados possíveis do sistema verificado. A conhecida entropia de Shannon é

$$S[P] = - \sum_{i=1}^M p_i \ln p_i, \quad (2)$$

Entre eles, $p_i \ln p_i = 0$ se $p_i = 0$, está relacionado ao processo físico descrito por P . Também é útil definir a chamada entropia de Shannon normalizada para avaliar a autoinformação de forma normalizada, denotada por

$$\mathcal{H}[P] = \frac{S[P]}{S_{max}} = \frac{S[P]}{S[P_e]} = \frac{S[P]}{\ln M}. \quad (3)$$

A segunda medida utilizada é a Complexidade Estatística (CE) $\mathcal{C}_{JS}[P]$ [Lamberti et al. 2004], que pode identificar detalhes dinâmicos críticos na série temporal. López-Ruiz *et al.* [López-Ruiz et al. 1995] propuseram esta medida de complexidade com base no produto de funções, $\mathcal{C}_{JS}[P] = \mathcal{H}[P] \mathcal{Q}_{JS}[P, P_e]$, onde $\mathcal{H}[P] \in [0, 1]$ é a Entropia de Shannon normalizada, e \mathcal{Q}_{JS} é o desequilíbrio com base na divergência de Jensen-Shannon (JS). Neste sentido, \mathcal{Q}_{JS} é expresso por

$$\begin{aligned} \mathcal{Q}_{JS} &= Q_0 \mathcal{J}_S[P, P_e] \\ &= Q_0 \left\{ S \left[\frac{P + P_e}{2} \right] - \left[\frac{S[P] + S[P_e]}{2} \right] \right\}, \end{aligned} \quad (4)$$

Onde Q_0 é uma constante de normalização, enquanto \mathcal{J}_S é a divergência de JS para quantificar a diferença entre distribuições de probabilidade. A presença de estrutura de correlação é quantificada na CE [Olivares et al. 2019], que mede a complexidade da série temporal. A Função de Distribuição de Probabilidade (FDP) desempenha um papel crucial nesse contexto, representando a probabilidade de ocorrência de diferentes eventos na série temporal. A propriedade global da CE é que seu valor não muda com diferentes layouts de FDP. Assim, $\mathcal{C}_{JS}[P]$ quantifica a desordem, mas também o grau de estrutura correlacionada.

Por fim, consideramos a medida de Informação de Fisher (FI) para analisar aspectos locais das mudanças no conteúdo de informação fornecido por uma série temporal. Ela possui diferentes interpretações e cálculos; entre outras coisas, a quantidade de informação extraída de um processo mede a capacidade de estimar parâmetros ou o estado desordenado de um sistema ou fenômeno [Olivares et al. 2019]. Definimos como

$$FI[P] = F_0 \sum_{i=1}^{N-1} (\sqrt[3]{p_{i+1}} - \sqrt[3]{p_i})^2, \quad (5)$$

onde F_0 é uma constante de normalização definida por

$$F_0 = \begin{cases} 1 & \text{se } p_{i^*} = 1 \text{ para } i^* = 1 \text{ ou } i^* = N \\ & \text{e } p_i = 0 \forall i \neq i^* \\ 1/2 & \text{caso contrário.} \end{cases} \quad (6)$$

De acordo com Olivares *et al.* [Olivares et al. 2019], a sensibilidade local de FI às FDPs discretas requer a ordem de i dos valores discretos em $P = \{p_i : i = 1, \dots, N\}$ ao somar a partir da Eq. 5. É a distância entre duas probabilidades relacionadas. Portanto, diferentes ordens resultarão em diferentes valores de FI e, assim, em sua natureza local. Vale salientar que para a extração das medidas de teoria da informação, foi utilizada a biblioteca *Ordpy* [Pessa and Ribeiro 2021].

De posse das medidas calculadas, foi utilizado o *Random Forest* para realizar a tarefa de classificação. Durante a configuração do modelo, especificamos os seguintes parâmetros: número de estimadores 50; número mínimo de amostras para dividir um nó 6; e profundidade máxima da árvore 10. Após a configuração dos parâmetros, o modelo foi treinado com os dados de treinamento preparados, utilizando-os para aprender padrões e relações nos dados que permitissem a classificação posterior dos comportamentos de direção. Inicialmente, os dados brutos de aceleração e giroscópio foram utilizados para treinar e testar o modelo. Em seguida, foram aplicadas medidas de teoria para gerar características informativas. Estas características foram calculadas com diferentes janelas de tempo e deslocamentos, resultando em múltiplos conjuntos de dados informativos.

Para fins de comparação, foi utilizado um modelo de LSTM para realizar a tarefa de classificação em nosso estudo. A escolha por esse modelo se deu devido à sua capacidade de capturar padrões temporais complexos em sequências de dados, tornando-o adequado para a análise de séries temporais, como os dados de movimento que estávamos lidando. Durante a configuração do modelo, especificamos os seguintes parâmetros: unidades LSTM consideramos o intervalo de 32 a 512, com um incremento de 32 unidades; taxa de dropout variando de 0.0 a 0.5 com um incremento de 0.1; e taxa de aprendizagem

fixa em 0.001. Após isso, foi realizada a otimização dos parâmetros com o intuito de maximizar a eficácia do modelo na classificação dos padrões de condução. Utilizamos o Keras Tuner para realizar uma busca aleatória no espaço de hiperparâmetros definido, executando várias tentativas e avaliando o desempenho do modelo em cada configuração com base na acurácia de validação. Este processo nos permitiu identificar a configuração ótima que resultou na melhor performance do modelo.

Em seguida, com a configuração ideal em mãos, procedemos com o treinamento do modelo LSTM usando os conjuntos de dados preparados. Em vista disso, avaliamos o desempenho do modelo utilizando o conjunto de dados de teste, que não foi utilizado durante o treinamento ou a validação. Essa avaliação nos permitiu verificar como o modelo generalizou para dados não vistos, fornecendo uma estimativa do desempenho do modelo em situações do mundo real.

3. Metodologia de Avaliação

Para nossas avaliações foi utilizado o conjunto de dados "Driving Behavior", disponível na plataforma *Kaggle*, criado por [Popescu and Ion 2022]. Este conjunto de dados consiste em dois arquivos CSV: *train_motion_data.csv* e *test_motion_data.csv*, que possui dados utilizados para treino e teste de modelos respectivamente. Cada registro do conjunto de dados está associado a uma etiqueta que classifica o comportamento do motorista. Maiores detalhes podem ser encontrados em [Popescu and Ion 2022]

O comportamento *Aggressive* é caracterizado por movimentos bruscos, incluindo viradas repentinas para a esquerda ou direita, além de acelerações e frenagens intensas, *Normal* indica direção moderada, sem movimentos bruscos, e *Slow* corresponde a velocidade abaixo da média, com acelerações e frenagens suaves [Cojocar and Popescu 2022].

Por outro lado, o conjunto de dados de teste consiste em 3084 registros, distribuídos em 1273 para *Slow*, 997 para *Normal* e 814 para *Aggressive*. Essa distribuição reflete as proporções do conjunto de dados de treinamento, embora com valores ligeiramente menores. A rota escolhida para coleta de dados de teste foi ligeiramente mais curta, resultando em um menor número total de instâncias. No entanto, a consistência nas proporções entre as classes é mantida, garantindo a validade e a representatividade dos nossos resultados de teste.

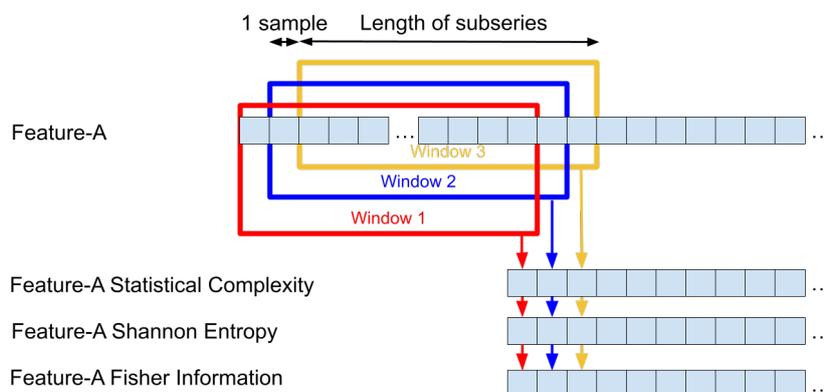


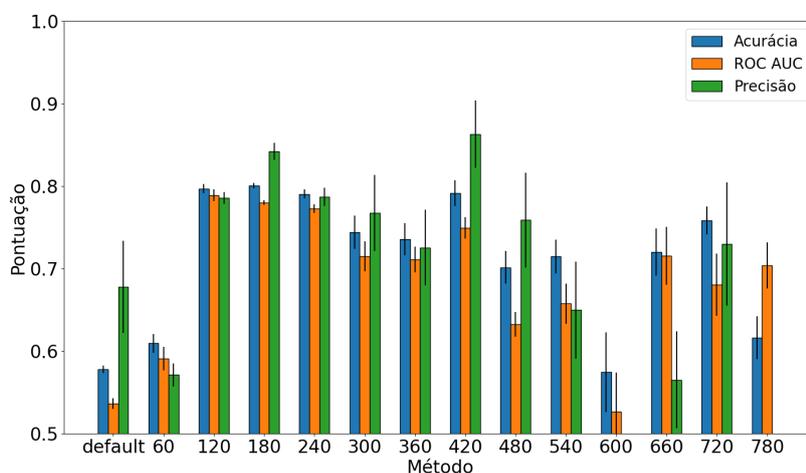
Figura 1. Cálculo de medidas de Teoria da Informação [Santos 2024].

Avaliamos o desempenho dos modelos em diferentes tamanhos de janela de 60

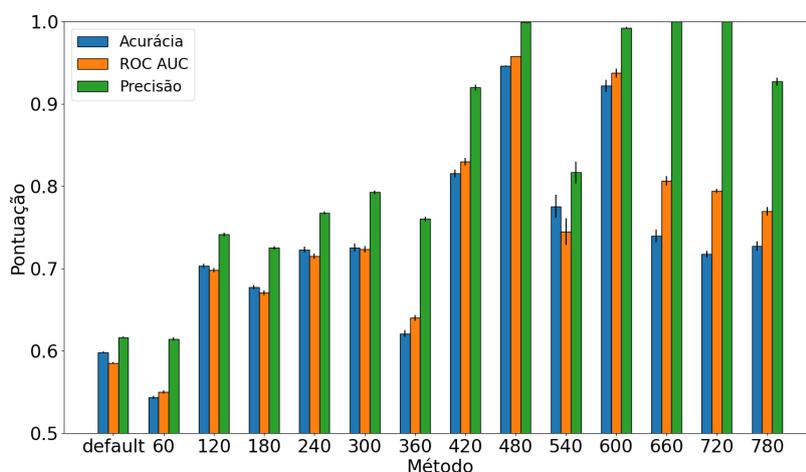
diferentes conjuntos de dados.

4. Resultados

Nesta seção, apresentamos os resultados da aplicação dos modelos de aprendizagem de máquina LSTM e Random Forest para a classificação do comportamento do motorista, com base em dados provenientes de sensores de acelerômetro e giroscópio veicular. A escolha desses modelos específicos foi motivada por trabalhos anteriores na literatura. Em particular, os autores do artigo que criaram o conjunto de dados que utilizamos, [Popescu and Ion 2022] empregaram esses dois modelos para classificar o comportamento do motorista. Portanto, optamos por utilizar os mesmos modelos para permitir uma comparação direta e verificar se a aplicação de medidas de teoria da informação poderia melhorar significativamente a performance dos modelos de classificação.



(a) LSTM



(b) Random Forest

Figura 2. Avaliação dos modelos.

Ao analisar a Figura 2a, vemos que a implementação de medidas de teoria da

informação no conjunto de dados original melhora consideravelmente as métricas de desempenho de um modelo LSTM destinado à classificação do comportamento do motorista. Nota-se que o uso de uma janela de tamanho 180 é particularmente eficaz, resultando nos valores mais altos de acurácia e curva ROC, e o segundo maior valor de precisão. Logo, podemos deduzir que nessa janela o modelo consegue capturar padrões de comportamento do motorista de forma mais precisa, sem perder informações cruciais presentes nos dados. Por outro lado, ao expandir a janela para tamanhos maiores que 180 há uma tendência de decréscimo nas pontuações, isso ocorre devido que ao aumentar o tamanho da janela de extração de medidas de Teoria da Informação diminui-se a quantidade total de dados extraídos e os algoritmos de Deep Learning requerem uma maior informação temporal.

Agora ao analisar o gráfico do Random Forest, Figura 2b, percebemos que a aplicação de medidas de teoria da informação, especialmente com uma janela de tamanho 480, no qual proporciona um aumento significativo nas métricas de desempenho do modelo. Este método alcançou valores próximos aos máximo das métricas abordadas, sugerindo que uma maior informação temporal (tamanho maior de janela para cálculo das medidas de Teoria da Informação) permite ao modelo capturar padrões complexos e incorporando mais informação nos dados para o treinamento do classificador. Contudo, também há uma tendência de decréscimo após esse tamanho de janela devido a menor informação temporal.

Isto posto, a aplicação da teoria da informação revelou-se uma boa estratégia para o aprimoramento dos modelos de classificação do comportamento do motorista. A análise comparativa entre os modelos LSTM e Random Forest demonstrou que a inserção de medidas de entropia de Shannon, complexidade estatística e informação de Fisher pode significativamente influenciar o desempenho preditivo. Porém, observamos que é necessário identificar a melhor escolha do tamanho da janela em relação a quantidade de dados disponível e o modelo utilizado.

O modelo Random Forest obteve melhor desempenho de uma janela de tamanho 480, alcançando métricas acima de 92 %. Esse resultado evidencia que o pré-processamento realizado com medidas de teoria da informação, aplicadas aos dados de sensores utilizados, teve um impacto positivo nos resultados de classificação do modelo Random Forest, aprimorando sua capacidade de classificação. Portanto, a seleção do tamanho da janela e a aplicação de medidas de teoria da informação devem ser consideradas em conjunto com as características específicas de cada modelo de aprendizagem de máquina.

5. Conclusão

Nesta trabalho, apresentamos a aplicação dos modelos de aprendizagem de máquina LSTM e Random Forest para a classificação do comportamento do motorista e exploramos o impacto da inserção de medidas de teoria da informação no desempenho de ambos os modelos.

Ao analisar os resultados, observamos que a implementação de medidas de teoria da informação no conjunto de dados original melhora consideravelmente as métricas de desempenho de um modelo LSTM, com uma janela de amostragem de tamanho 180, para a classificação do comportamento do motorista. Por outro lado, o modelo Random Forest obteve melhor desempenho com uma janela de tamanho 480, alcançando métricas próximas à máxima. Isso evidencia que o pré-processamento realizado com medidas de

teoria da informação teve um impacto positivo nos resultados de classificação.

Em conclusão, nossos resultados demonstram que a teoria da informação pode melhorar significativamente a performance dos modelos de classificação do comportamento do motorista. No entanto, é crucial considerar as particularidades de cada algoritmo ao otimizar os hiperparâmetros e a abordagem de pré-processamento dos dados. O estudo abre oportunidades para futuras investigações sobre a otimização dos parâmetros e a aplicação de medidas de teoria da informação em outros contextos de aprendizado de máquina e análise de dados veiculares.

Como trabalho futuro, sugere-se a adoção de uma *baseline* diferente para comparação. Por exemplo, pode-se investigar a aplicação de Redes Neurais Convolucionais (CNNs) em conjunto com Long Short-Term Memory (LSTM) para explorar a eficácia na captura de padrões espaciais e temporais dos dados sensoriais. Além disso, se faz necessário um maior estudo para otimizar a escolha da janela de tempo de cálculo das medidas de informação concomitantemente com a escolha dos hiperparâmetros. A exploração de diferentes fontes de dados, como dados de sensores adicionais (câmeras, microfones) também representa uma direção promissora para expandir e enriquecer a análise do comportamento do motorista.

Referências

- Aquino, A. L., Cavalcante, T. S., Almeida, E. S., Frery, A. C., and Rosso, O. A. (2015). Characterization of vehicle behavior with information theory. *The European Physical Journal B*, 88(10):257.
- Bandt, C. and Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 88:174102.
- Bernardi, M. L., Cimitile, M., Martinelli, F., and Mercaldo, F. (2018). Driver and path detection through time-series classification. *Journal of Advanced Transportation*, 2018:1758731.
- Cojocar, I. and Popescu, P.-S. (2022). Building a driving behaviour dataset. In *Romanian Conference on Human-Computer Interaction*.
- Girma, A., Yan, X., and Homaifar, A. (2019). Driver identification based on vehicle telematics data using lstm-recurrent neural network. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 894–902.
- Lamberti, P. W., Martin, M., Plastino, A., and Rosso, O. A. (2004). Intensive entropic non-triviality measure. *Physica A: Statistical Mechanics and its Applications*, 334(1-2):119–131.
- López-Ruiz, R., Mancini, H. L., and Calbet, X. (1995). A statistical measure of complexity. *Physics Letters A*, 209(5-6):321–326.
- Manderna, A. and Kumar, S. (2022). Effective long short-term memory based-driver identification in its. In *2022 International Conference on Inventive Computation Technologies (ICICT)*.
- Martinelli, F., Mercaldo, F., Orlando, A., Nardone, V., Santone, A., and Kumar, A. (2020). Human behavior characterization for driving style recognition in vehicle system. *Computers & Electrical Engineering*, 83.

- Mateos, D. M., Gómez-Ramírez, J., and Rosso, O. A. (2021). Using time causal quantifiers to characterize sleep stages. *Chaos, Solitons & Fractals*, 146:110798.
- Miyajima, C., Nishiwaki, Y., Ozawa, K., Wakita, T., Itou, K., Takeda, K., and Itakura, F. (2007). Driver modeling based on driving behavior and its evaluation in driver identification. *Proceedings of the IEEE*, 95(2):427–437.
- Nascimento, W. S. and Prudente, F. V. (2016). Study of shannon entropy in the context of quantum mechanics: An application to free and confined harmonic oscillator. In *Quim. Nova*, pages 757–764.
- Olivares, F., Souza, L., Legnani, W., and Rosso, O. A. (2019). Informational time causal planes: A tool for chaotic map dynamic visualization. In Legnani, W. and Moschandreou, T. E., editors, *Nonlinear Systems-Theoretical Aspects and Recent Applications*, chapter 5. IntechOpen, Rijeka.
- Park, K. H. and Kim, H. K. (2019). This car is mine!: Automobile theft countermeasure leveraging driver identification with generative adversarial networks. *CoRR*, abs/1911.09870.
- Pessa, A. A. B. and Ribeiro, H. V. (2021). ordpy: A python package for data analysis with permutation entropy and ordinal network methods. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(6):063110.
- Popescu, P.-S. and Ion, C. (2022). Driving behavior.
- Santos, G. S. (2024). Use of information theory measures extracted from obd-ii interface data for driver identification. Mestrado em informática, Universidade Federal de Alagoas, Instituto de Computação, Maceió.
- Schneegass, S., Pfleging, B., Broy, N., Heinrich, F., and Schmidt, A. (2013). A data set of real world driving to assess driver workload. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '13, page 150–157, New York, NY, USA. Association for Computing Machinery.
- Singh, M. and Dubey, R. K. (2023). Deep learning model based co2 emissions prediction using vehicle telematics sensors data. *IEEE Transactions on Intelligent Vehicles*, 8(1):768–777.
- Uvarov, K. and Ponomarev, A. (2021). Driver identification with obd-ii public data. In *2021 28th Conference of Open Innovations Association (FRUCT)*, pages 495–501.
- Vaiti, T., Tisljaric, L., Erdelic, T., and Caric, T. (2022). Traffic emissions clustering using obd-ii dataset based on machine learning algorithms. *Transportation Research Procedia*, 64:364–371. International Scientific Conference “The Science and Development of Transport - Znanost i razvitak prometa”.
- Wu, J.-D. and Ye, S.-H. (2009). Driver identification using finger-vein patterns with radon transform and neural network. *Expert Syst. Appl.*, 36:5793–5799.
- Xun, Y., Sun, Y., and Liu, J. (2019). An experimental study towards driver identification for intelligent and connected vehicles. In *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*.