



WALE-A: SISTEMA PARA CLASSIFICAÇÃO DE EVENTOS CRÍTICOS DE DIREÇÃO

Maximiliano Martins de Faria

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientador: Geraldo Bonorino Xexéo

Rio de Janeiro
Setembro de 2023

WALE-A: SISTEMA PARA CLASSIFICAÇÃO DE EVENTOS CRÍTICOS DE
DIREÇÃO

Maximiliano Martins de Faria

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientador: Geraldo Bonorino Xexéo

Aprovada por: Prof. Geraldo Bonorino Xexéo

Prof. Jano Moreira de Souza

Prof. Geraldo Zimbrão da Silva

Prof. Carlos Eduardo Ribeiro de Mello

Prof. Eduardo Soares Ogasawara

RIO DE JANEIRO, RJ – BRASIL

SETEMBRO DE 2023

Martins de Faria, Maximiliano

WALE-a: Sistema para classificação de eventos críticos de direção/Maximiliano Martins de Faria. – Rio de Janeiro: UFRJ/COPPE, 2023.

XIII, 106 p.: il.; 29, 7cm.

Orientador: Geraldo Bonorino Xexéo

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2023.

Referências Bibliográficas: p. 92 – 106.

1. Driving Behavior. 2. crash prediction system. 3. collision avoidance system. 4. Driving maneuvers. 5. accident prediction. 6. safety-critical events. 7. wavelet. 8. time series classification. 9. time series transformation.
I. Bonorino Xexéo, Geraldo. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

*Para Sabrina, uma mulher
maravilhosa que me acompanha
desde então.*

Agradecimentos

A conclusão de uma Tese de Doutorado, sem dúvida, é o resultado de anos dedicados ao estudo e à perseverança. Esta jornada é frequentemente resumida como 10% de inspiração e 90% de transpiração, marcada por momentos de felicidade, angústia e apreensão. Descrever o esforço investido em um trabalho desse porte é uma tarefa desafiadora, mas, para além disso, essa trajetória influencia a vida de todos ao nosso redor. Assim, gostaria de expressar meus sinceros agradecimentos a algumas pessoas fundamentais que me acompanharam nessa jornada.

Primeiramente, desejo expressar minha gratidão à minha família. Ela constitui a base de tudo, e tenho a fortuna de contar com uma sólida estrutura familiar, composta por meu pai, Nelson, e minha mãe, Terezinha, que sempre incentivaram os estudos como uma ferramenta de empoderamento. Os filhos são a razão de nossas vidas, e a conclusão deste trabalho é uma maneira de transmitir a relevância da educação a todos. Muitas vezes, a motivação provém do fato de que, ao término deste percurso, estou proporcionando um exemplo valioso aos meus filhos. Além da família nuclear, ao longo desta trajetória, tive a felicidade de encontrar Sabrina, minha parceira de vida. Desejo expressar meu profundo agradecimento a ela por seu constante apoio e pela crença em minha capacidade de chegar a este ponto com sucesso. Também não posso deixar de lembrar do meu orientador, Prof. Geraldo Xexeo, cuja orientação foi como uma luz nos momentos de incertezas onde sempre surgia com ideias geniais iluminando as dúvidas. Finalmente, gostaria de homenagear uma pessoa que não pode estar presente para testemunhar esta conquista: meu avô, o saudoso Sr. José Faria. Ele desempenhou um papel crucial em minha formação e é uma fonte inesgotável de inspiração.

Além da família e do amor de minha companheira, a espiritualidade é um pilar essencial na vida de um indivíduo. Hoje, a fé desempenha um papel significativo na minha vida e minha evolução como ser humano. Agradeço a Deus e a mãe maior, Nossa Senhora da Aparecida, por estarem presentes me oferecendo a compreensão de que existe algo além que transcende a nossa própria existência. Reconhecer essa dimensão é gratificante para mim. Agradeço, de todo coração, ao meu Deus e mãe eternos por minha vida e por ser quem eu sou.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

WALE-A: SISTEMA PARA CLASSIFICAÇÃO DE EVENTOS CRÍTICOS DE DIREÇÃO

Maximiliano Martins de Faria

Setembro/2023

Orientador: Geraldo Bonorino Xexéo

Programa: Engenharia de Sistemas e Computação

Para melhorar a segurança na direção, os Sistemas de Previsão de Acidentes (CPS) são usados para identificar e classificar eventos críticos de direção (SCE), ajudando os motoristas a antecipar possíveis acidentes. Este estudo introduz um novo algoritmo destinado a CPS, fundindo técnicas avançadas de inteligência artificial e técnicas de pré-processamento para estabelecer um sistema mais preciso. Nossa proposta denominada de **WALE-a**, examina dados de séries temporais coletados por sensores veiculares, empregando uma abordagem de decomposição de *wavelets* em conjunto com um classificador linear. Isso permite a categorização de eventos, como eventos como *crash*, *near-crash* e manobras abruptas. Através da análise abrangente do desempenho de diversos algoritmos de aprendizado de máquina, o nosso algoritmo se sobressai.

Os resultados obtidos revelaram a notável eficácia do algoritmo **WALE-a** em diversos conjuntos de dados, em especial nos cenários com os *datasets accel_x* e *accel_y*, originários de acelerômetros. As acurácias de nossa proposta foram superiores a 88%, alcançando mais de 90% no conjunto de dados *accel_y*. Além disso, demonstrou uma notável capacidade em distinguir os eventos críticos e não críticos de forma precisa e consistente, superando outros classificadores. A força do algoritmo foi reiterada ao assegurar seis das dez primeiras posições em nosso *ranking* de classificação. Foi comprovado, também, que o algoritmo **WALE-a** se destaca ao equilibrar acurácia e tempo de processamento, alcançando consistentemente uma acurácia acima de 80% e mantendo tempos de processamento inferiores a 20 milissegundos em todas as três versões do algoritmo. Por fim, nossa pesquisa evidenciou o papel crucial da inteligência artificial e das técnicas de pré processamento na construção dos bons sistemas de predição de acidentes.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

WALE-A: A SYSTEM FOR CLASSIFICATION OF SAFETY-CRITICAL EVENTS

Maximiliano Martins de Faria

September/2023

Advisor: Geraldo Bonorino Xexéo

Department: Systems Engineering and Computer Science

In order to improve driving safety, Crash Prediction Systems (CPS) are used to identify and classify critical driving events (SCE), helping drivers anticipate potential accidents. Our team has introduced a new algorithm for CPS that utilizes artificial intelligence techniques to create a safer driving experience. Our algorithm analyzes sensor time series data using wavelet decomposition and a linear classifier, classifying events such as crashes, near-crashes, and abrupt maneuvers. We trained, tested, and evaluated our algorithm against various machine learning algorithms, including our algorithm, WALE-a, achieving accuracy scores greater than 90% for various datasets, particularly with features of accelerometer data. Our algorithm has shown remarkable efficiency, achieving an average accuracy rate of more than 80% while maintaining processing times below 20 milliseconds. These results indicate that the WALE-a algorithm can reduce road accidents and promote safer driving conditions.

Sumário

Agradecimentos	v
Lista de Figuras	xi
Lista de Tabelas	xiii
1 Introdução	1
1.1 Contextualização	1
1.2 Motivação	3
1.3 Contribuições	4
1.4 Objetivo do trabalho	5
1.5 Organização do trabalho	5
2 Revisão da Literatura	7
2.1 Planejamento da revisão	9
2.1.1 Objetivos da revisão	9
2.1.2 Questões da pesquisa	9
2.1.3 Os termos de busca	10
2.1.4 Repositórios de artigos científicos	11
2.1.5 Restrições da pesquisa	12
2.1.6 Critérios para seleção dos artigos	12
2.2 Execução da revisão	14
2.2.1 Identificação e seleção dos artigos	14
2.2.2 Avaliação de qualidade dos artigos retornados	14
2.2.3 Extração dos dados	15
2.3 Relatório da revisão	16
2.3.1 Resultados Encontrados	16
2.4 Consolidação do conhecimento	31
2.5 Oportunidade de pesquisa	32
3 <i>Driving Behavior</i> conceitos e aplicações	36
3.1 <i>Driving Behavior</i> e os termos relacionados	37

3.1.1	<i>Driving Style</i> : Definição e Importância	37
3.1.2	<i>Driving Maneuvers</i>	38
3.1.3	<i>Driving Behavior</i>	40
3.1.4	<i>Driving Event</i>	42
3.2	Identificação e classificação de <i>Driving Event</i> (DE)	45
3.3	Aplicações para <i>Driving Behavior</i> e <i>Driving Event</i>	47
3.3.1	<i>Driver Assistance System</i>	48
3.3.2	<i>Eco Driving</i>	48
3.3.3	<i>Road Condition Monitoring</i>	49
3.3.4	<i>Fleet Management</i>	49
3.3.5	<i>Insurance Applications</i>	50
3.3.6	<i>Crash Prediction Systems</i>	50
4	Sistemas de predição de acidentes (CPS)	52
4.1	Abordagens para CPS	53
4.2	Conjuntos de dados	55
4.3	Técnicas utilizadas	57
5	Metodologia e dados	59
5.1	Coleta de Dados	60
5.2	Preparação dos Dados	60
5.3	Treinamento dos Classificadores	61
5.4	Avaliação de Desempenho	62
6	WALE-a: Camadas Adaptativas Baseada em Wavelets	64
6.1	Classificação de Séries Temporais	65
6.2	Transformação de Séries Temporais	65
6.3	Estrutura do Algoritmo WALE-a	67
6.4	As Camadas do WALE-a	69
6.4.1	Camada de Decomposição	69
6.4.2	Camada de Extração	70
6.4.3	Camada de <i>Pooling</i>	71
6.4.4	Camada de Classificação	72
7	Experimentos	74
7.1	Configuração do Experimento	74
7.2	Coleta de Dados	74
7.3	Preparação dos Dados	75
7.3.1	Identificação de Eventos	77
7.4	Amostragem	80

7.5	Ajuste de Algoritmos	81
7.6	Aprendizado	81
8	Análise dos Resultados	83
9	Conclusão	90
	Referências Bibliográficas	92

Lista de Figuras

1.1	WALE-a overview	5
2.1	Exemplo de ficha catalográfica a ser utilizada neste trabalho	15
2.2	Arquitetura proposta por MESEGUER <i>et al.</i> (2013)	17
2.3	Arquitetura proposta por CASTIGNANI <i>et al.</i> (2015)	18
2.4	Arquitetura proposta por AL-DOORI <i>et al.</i> (2014)	19
2.5	Processos do KDD segundo HAN <i>et al.</i> (2011)	20
2.6	Processos de Mineração de Dados utilizada neste trabalho	21
2.7	Metodologia proposta por BANERJEE <i>et al.</i> (2014) com alguns processos de Mineração aparentes	21
2.8	Classificação de <i>Driving Behavior</i> proposta por KALSOOM e HALIM (2013)	22
2.9	Algoritmo proposto por SAIPRASERT <i>et al.</i> (2017) com a visão de processos de Mineração aparente	23
2.10	Modelo de análise de <i>Driving Behavior</i> encontrado neste trabalho	27
2.11	Tabela com o <i>features set</i> usado em WU <i>et al.</i> (2016)	30
3.1	Termos relacionados com <i>Driving Behavior</i> baseado no trabalho de MARTINEZ <i>et al.</i> (2018)	37
3.2	Sistema de coordenada do acelerômetro relativa ao veículo FAZEEN <i>et al.</i> (2012)	40
3.3	Sinal da aceleração do eixo <i>x</i> do acelerômetro relativo a manobra “Mudar de faixa para direita/esquerda” FAZEEN <i>et al.</i> (2012)	41
3.4	Modelo conceitual para identificação de <i>Driving Maneuvers</i>	42
3.5	Modelo de relações de <i>Driving Behavior</i> e <i>Driving Event</i>	43
3.6	Algoritmo de identificação de evento de acordo com o trabalho de SAIPRASERT <i>et al.</i> (2017)	44
3.7	Uso do DTW na comparação de sinais para detecção de eventos críticos baseado no trabalho de SAIPRASERT <i>et al.</i> (2017)	45
3.8	Classificação de evento com modelo de aprendizado de máquina baseado no trabalho de YADAV <i>et al.</i> (2019)	46

3.9	Modelo geral utilizado nas análises de DB e DE para este trabalho . . .	46
3.10	Modelo para identificação e classificação de evento crítico	47
4.1	Modelo para predição de acidentes por localidade e horário	54
4.2	Modelo para sistema completo de predição de acidentes	55
4.3	Construção do conjunto de dados do trabalho WANG <i>et al.</i> (2010) . .	56
5.1	Metodologia do processo	59
5.2	Processo de validação da metodologia	62
6.1	Aplicação de filtros passa-alta e passa-baixa ao sinal	66
6.2	Encadeamento dos bancos de filtro na decomposição do sinal	67
6.3	Decomposição de séries temporais	68
6.4	Visão geral das camadas do WALE-a	68
6.5	Visão geral da Camada de Decomposição do WALE-a	70
6.6	Visão da Camada de Extração do WALE-a	71
6.7	Visão da Camada de Pooling do WALE-a	73
7.1	Sistema de execução dos experimentos <i>Classification Constructor</i> . .	75
7.2	Visualização de registros nulos e a quantidade removida	75
7.3	Número de eventos selecionados e os tamanhos das janelas	76
7.4	Funcionamento da técnica de janela deslizante	78
7.5	Cálculo das três métricas nas janelas	78
7.6	Cálculo dos limites para cada métrica	79
7.7	Seleção dos dados relevantes de cada série das métricas	79
7.8	Visão geral do algoritmo “Seleção de Janelas”	80
8.1	Desempenho da acurácia entre classificadores	85
8.2	Desempenho dos classificadores entre conjuntos de dados	86
8.3	Relação entre a <i>recall</i> e especificidade dos classificadores	88
8.4	Relação entre a acurácia e o tempo de processamento dos classificadores	89

Lista de Tabelas

2.1	Principais características da Revisão Sistemática	8
2.2	Etapas da RS	9
2.3	Questões da pesquisa	10
2.4	Termos de busca	10
2.5	Expressão de busca	11
2.6	Repositórios de artigos científicos	12
2.7	Questões para avaliação de qualidade dos artigos	13
2.8	Respostas e pontuação	13
2.9	Resultados da 1ª seleção	14
2.10	Resultados da 2ª seleção - Avaliação de qualidade	15
2.11	Técnicas de pré-processamento mais usadas	24
2.12	Técnicas de processamento mais usadas	34
2.13	Sensores utilizados para análises de <i>Driving Behavior</i>	35
3.1	<i>Driving Maneuvers</i> mais relevantes baseado nos estudos de GERDES (2006), HÜLNHAGEN <i>et al.</i> (2010) e SATHYANARAYANA <i>et al.</i> (2012)	39
3.2	Dados/sinais citados em diversos estudos FAZEEN <i>et al.</i> (2012), SAI-PRASERT <i>et al.</i> (2017) e FUGIGLANDO <i>et al.</i> (2018)	40
3.3	Eventos críticos	44
6.1	Lista das métricas usadas	72
7.1	WALE-a versions	80
7.2	Algorithms list.	82
8.1	The top ten specific classifiers	87

Capítulo 1

Introdução

1.1 Contextualização

Nas últimas décadas, o crescimento global tem impulsionado uma expansão significativa dos bens de consumo, como os veículos automotores WU *et al.* (2016). No entanto, esse crescimento tem sido acompanhado por um aumento no congestionamento urbano e nos níveis de acidentes de trânsito FUGIGLANDO *et al.* (2018). É notório que os problemas de tráfego exercem o maior impacto nos acidentes automobilísticos JÚNIOR *et al.* (2017) FUGIGLANDO *et al.* (2018). Essa tendência compromete, consideravelmente, a segurança viária e o bem-estar das pessoas, uma vez que a interação entre indivíduos e veículos é um dos principais desencadeadores de acidentes rodoviários SAIPRASERT *et al.* (2017). De acordo com a Organização Mundial da Saúde (OMS) ORGANIZATION (2018), os acidentes de trânsito são um problema global responsável por mais 1,3 milhões de mortes em 2018.

Os acidentes de trânsito resultam de uma interação complexa de diversos fatores, que abrangem desde erros humanos, até as más condições veiculares e problemas na infraestrutura rodoviária SAIPRASERT *et al.* (2017). Dados provenientes de pesquisas realizadas nos Estados Unidos, evidenciam que em torno de 50% dos acidentes rodoviários estão diretamente ligados a comportamentos inseguros por parte dos condutores ENGELBRECHT *et al.* (2015). Esses comportamentos inseguros se manifestam de variadas maneiras como, por exemplo, a falta de atenção por parte do motorista durante a condução. Ainda que a atividade de dirigir possa ser considerada fácil, mesmo um breve momento de desatenção pode resultar em acidente. Os achados do estudo conduzido por Sathyannarayana *et al.* SATHYANARAYANA *et al.* (2012) revelaram que mais de 75% das colisões e 65% das quase colisões, decorreram da falta de atenção do motorista. Outros fatores que exercem impacto nos acidentes incluem atitudes imprudentes, fadiga, direção sob condições de sono, uso de substâncias como drogas e álcool. Todos esses aspectos se manifestam por meio

de comportamentos de condução anormais ou agressivas, que podem ser observadas em manobras como frenagens, acelerações bruscas, curvas acentuadas ou mudanças abruptas de faixa MOHAMAD *et al.* (2011); SAIPRASERT *et al.* (2014).

Esses desafios são enfrentados a partir do desenvolvimento de ferramentas voltadas para a promoção da segurança nas vias urbanas e rodovias das grandes cidades. A literatura científica de alguns anos para cá tem intensificado estudos no monitoramento do comportamento de direção, como forma de melhorar a segurança nesses locais SAIPRASERT *et al.* (2017). Esses estudos são conhecidos como análises de *Driving Behavior* e buscam entender todo ambiente de direção em volta do motorista, desde suas ações, tomadas de decisão, inclusive os eventos que são aplicados no veículo. *Driving Behavior* é um processo cíclico onde o motorista reconhece o ambiente a sua volta, decide uma ação (acelerar ou frear), e a executa de forma prudente com segurança, ou algumas vezes de forma imprudente e/ou perigosa WAKITA *et al.* (2005).

As análises de *Driving Behavior* focam na forma de como o motorista conduz seu veículo, principalmente, naquelas atitudes consideradas irresponsáveis e agressivas. A identificação e classificação são as tarefas principais nas análises de *Driving Behavior* CASTIGNANI *et al.* (2015) ENGELBRECHT *et al.* (2015). Diversos estudos apontam que a identificação e classificação de *Driving Behavior* contribuem para melhorar a segurança dos motoristas e das rodovias WU *et al.* (2016) DÖRR *et al.* (2014). As análises de *Driving Behavior* procuram identificar as manobras de um motorista e classificá-la como “segura” ou “perigosa”.

Aliado a isso, o avanço da indústria automotiva desenvolveu veículos com sensores capazes de captar determinados sinais durante os eventos de direção FUGIGLANDO *et al.* (2018). Estes sensores contribuem para a análise de *Driving Behavior*, uma vez que coletam os dados de direção de um motorista, armazenando-os para serem avaliados posteriormente SAIPRASERT *et al.* (2017). Esses dados são utilizados em diversas aplicações pela indústria desde os “sistema de transporte inteligente”, “sistema de assistência ao motorista”, “gerenciamento de frotas”, “mercado de seguro” e dos “sistemas de predição de acidentes” (CPS).

De acordo com o exposto acima, fica evidente a importância das análises de *Driving behavior* aliados aos sensores instalados nos veículos. Esta união pode ser utilizada para reconhecer padrões de direção e ainda auxiliar nas tomadas de decisões automáticas, como emissão de alertas sobre possíveis perigos eminentes WAHAB *et al.* (2007); JOHNSON e TRIVEDI (2011). Neste cenário destacam-se os sistemas de predição de acidentes (CPS). Estes sistemas, nos últimos anos, tornaram-se ferramentas fundamentais na prevenção de acidentes de trânsito XIONG *et al.* (2019); HALIM *et al.* (2016a). O CPS ajuda os motoristas a se prevenirem de potenciais problemas aumentando sua segurança ao volante, identificando e classificando os

eventos críticos de direção. No entanto, CPS precisa equilibrar precisão e rapidez para evitar alarmes falsos e feedback atrasados, como demonstrado por estudos anteriores WANG *et al.* (2010); XIONG *et al.* (2019); BA *et al.* (2017); ELAMRANI ABOU ELASSAD *et al.* (2020a); QU *et al.* (2017).

1.2 Motivação

Com base nas informações apresentadas, surge um desafio real e indispensável: desenvolver ferramentas capazes de aprimorar a segurança na direção. A busca por sistemas que possam alertar os motoristas situações de perigo, com base em seus próprios dados de condução, está determinada a se tornar uma tecnologia obrigatória nos próximos anos. A utilização dos dados dos veículos para identificar padrões de direção específicos, classificando-os de acordo com os níveis de risco, representa um avanço significativo para a segurança dos motoristas e das rodovias em geral.

A detecção de potenciais riscos na direção é uma questão recorrente, tanto na literatura científica quanto na aplicação prática no setor automobilístico. O cerne desse tema reside na identificação, classificação e previsão de eventos críticos de direção, com o objetivo de emitir alertas ou tomar medidas preventivas. Em contrapartida, os indicadores de segurança utilizados para a detecção de riscos de acidentes frequentemente se baseiam em suposições inconsistentes, como a consideração de velocidade ou aceleração constantes, o que pode comprometer a precisão da avaliação de uma ocorrência XIONG *et al.* (2019). Nesse contexto, a abordagem orientada a dados e o emprego de técnicas de inteligência artificial, impulsionadas pelo crescimento notável nos últimos anos, mostram-se mais eficazes. Ao aproveitar os dados dos sensores veiculares, essas abordagens proporcionam maior assertividade HALIM *et al.* (2016a). As aplicações mais proeminentes no campo do *Driving behavior* são os **sistemas de previsão de acidentes (CPS)**.

Desenvolver um CPS é uma tarefa intrincada e desafiadora, dado o escopo diversificado das etapas envolvidas. O processo pode ser dividido em três categorias principais: **identificação**, **classificação** e **previsão** dos eventos críticos. Neste estudo, concentraremos nossos esforços nas etapas de **identificação** e **classificação**. Os dados utilizados em nossa pesquisa foram coletados a partir de vários sensores veiculares. Empregamos técnicas de janela deslizante para extrair eventos críticos, que posteriormente serão categorizados mediante diversas abordagens, incluindo algoritmos tradicionais de aprendizado de máquina, tais como SVM, árvores de decisão e redes neurais HALIM *et al.* (2016a); GUTIERREZ-OSORIO e PEDRAZA (2020); CASTRO e KIM (2016).

O aprendizado de máquina baseado nos dados de direção representa o ápice da tecnologia para o desenvolvimento de CPS. Para viabilizar a operação desses siste-

mas, torna-se crucial identificar os eventos críticos de segurança a fim de antever potenciais acidentes. Assim, diante de volumes consideráveis de dados, uma abordagem de aprendizado de máquina se mostra viável para a marcação desses eventos. Inicialmente, os eventos críticos são identificados para, posteriormente, serem classificados em três níveis de gravidade: acidentes, quase-acidentes ou manobras abruptas KLUGER *et al.* (2016).

1.3 Contribuições

O presente trabalho introduz uma abordagem inovadora para a **identificação** de eventos críticos de segurança (SCE) e, posteriormente, a **classificação** desses eventos como acidentes, quase-acidentes ou outros comportamentos perigosos de direção. A metodologia adotada aproveita dados reais provenientes dos sensores veiculares do projeto SHRP2 *Naturalistic Direction Study* (NDS) fornecidos pelo Instituto da Virgínia, EUA DAVOODI *et al.* (2020).

No que se refere à **identificação** dos eventos, desenvolvemos um processo coeso e claro, que abrange a coleta, transformação e extração dos eventos considerados críticos. Esse processo foi estruturado em etapas distintas, cada uma cumprindo uma função específica, tornando-o de fácil compreensão e passível de reprodução. No âmbito da **classificação**, além de construirmos um processo específico para a tarefa, criamos um algoritmo denominado **WALE-a**, especialmente concebido para classificar eventos de segurança. Esse algoritmo, fundamentado em Redes Neurais de Convolução (CNN), emprega a decomposição *wavelet* e um classificador linear para categorizar as séries temporais dos sensores veiculares. A Figura 1.1 oferece uma visão panorâmica de nossa proposta e ilustra como as tarefas de **identificação** e **classificação** se interrelacionam.

Nosso trabalho representa um avanço na construção dos sistemas de previsão de acidentes (CPS), com contribuições relevantes tanto para o âmbito acadêmico, quanto para a aplicação prática no mercado. Tais contribuições se materializam na formulação de um novo algoritmo de análise de séries temporais, capaz de classificar eventos críticos de segurança de maneira eficaz. Além disso, foi uma prioridade nossa garantir que os processos de identificação e classificação fossem escaláveis, de rápida execução e facilmente interpretáveis e reproduzíveis. O algoritmo de classificação que desenvolvemos foi concebido para ser flexível e pode ser utilizado com diversas funções *wavelet* simultaneamente, o que resulta em um aprimoramento no reconhecimento de padrões. Por fim, demonstramos que a **identificação** e **classificação** de eventos críticos de segurança podem ser realizadas com êxito, utilizando um número reduzido de sensores de baixo custo.

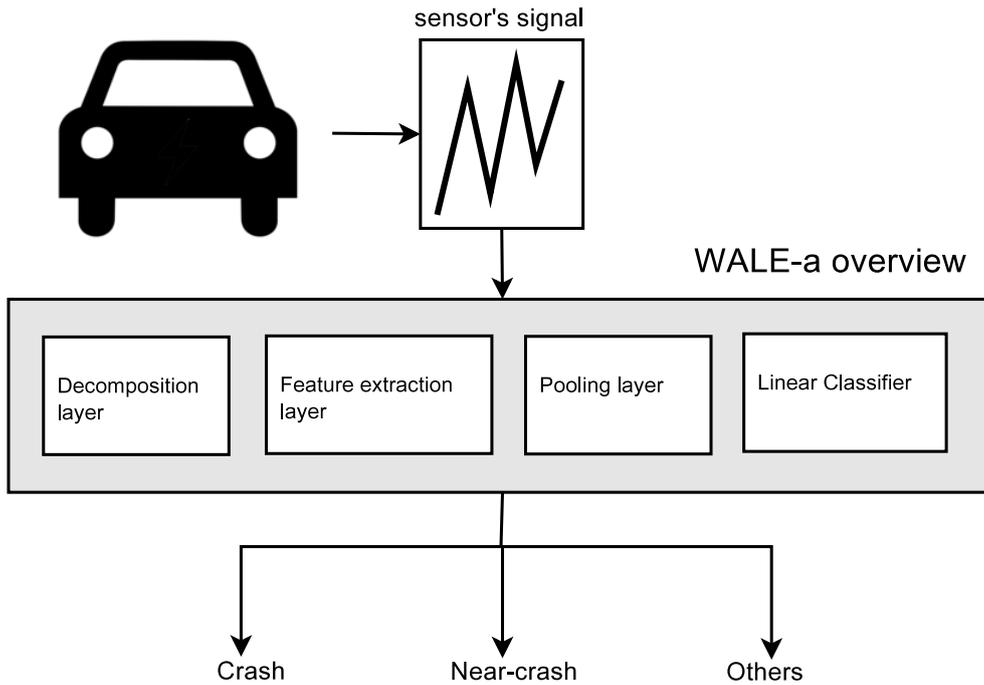


Figura 1.1: WALE-a overview

1.4 Objetivo do trabalho

O objetivo central da nossa pesquisa é aprimorar os sistemas de previsão de acidentes, por meio da introdução de novas abordagens na **identificação** e **classificação** de séries temporais obtidas a partir dos sensores veiculares. Impulsionados pelo conhecimento adquirido em *Driving Behavior* e pela motivação de aperfeiçoar a segurança no trânsito, buscamos enfrentar o desafio apresentado pela detecção de eventos críticos de segurança, visando uma classificação mais assertiva desses eventos como acidentes, quase-acidentes ou outros comportamentos perigosos de direção. Nossas contribuições se refletem na proposição de um processo de identificação claro e reproduzível, além do desenvolvimento de um algoritmo inovador denominado **WALE-a**, fundado em Redes Neurais de Convolução (CNN), enriquecido pela decomposição *wavelet* e um classificador linear. Por meio dessas metas, almejamos fortalecer a construção de Sistemas de Previsão de Acidentes (CPS), aportando novos métodos que otimizem a **identificação** e **classificação** de eventos críticos, ao mesmo tempo que utilizam eficientemente os dados dos sensores de veículos.

1.5 Organização do trabalho

A estrutura desta tese foi organizada de forma coesa para facilitar a compreensão progressiva do seu conteúdo. A sequência de capítulos inicia com o Capítulo 2, onde conduzimos uma revisão sistemática da literatura e compartilhamos as conclusões e

insights obtidos dessa análise. Avançando, o Capítulo 3 explora os conceitos essenciais relacionados ao *Driving Behavior*, proporcionando o embasamento necessário para compreender o contexto subjacente ao estudo. Em seguida no Capítulo 4, introduzimos os fundamentos dos Sistemas de Predição de Colisões (*Crash Prediction Systems*), estabelecendo as bases para a investigação das abordagens propostas neste trabalho. A metodologia empregada na condução dos experimentos é descrita no Capítulo 5, oferecendo uma visão das etapas planejadas. A proposta do novo algoritmo de classificação, chamado **WALE-a**, é detalhada no Capítulo 6, onde explicamos sua formulação e funcionamento em profundidade. A execução prática dos experimentos é abordada no Capítulo 7, onde compartilhamos como as estratégias que foram implementadas. No Capítulo 8, analisamos os resultados dos experimentos, destacando as conclusões e ideias resultantes. Finalmente, no Capítulo 9, consolidamos as conclusões deste estudo, enfatizando suas contribuições e as lições aprendidas ao longo do processo.

Capítulo 2

Revisão da Literatura

Nesta seção, abordaremos a metodologia empregada em nossa revisão da literatura. Nossa investigação teve início em 2018, com o objetivo primordial de compreender o panorama abrangente do *Driving Behavior*, englobando seus conceitos, tecnologias, aplicações e sua utilização na análise e aprimoramento da segurança rodoviária. Partimos do pressuposto de que a partir dessas bases sólidas, poderíamos identificar lacunas potenciais que se transformassem em oportunidades de pesquisa. Para concretizar esse propósito, nossa primeira etapa consistiu em planejar e executar uma revisão sistemática da literatura, empregando nossas questões de pesquisa como ferramenta para mapear áreas ainda não exploradas.

Revisão Sistemática (RS) é um tipo de revisão de literatura com foco nas questões de uma pesquisa específica. É uma forma de identificar, avaliar e sintetizar todas as informações pertinentes às questões da pesquisa KITCHENHAM (2004). Realizar uma RS é uma decisão importante para qualquer projeto de pesquisa, haja visto, que a mesma se utiliza de métodos claros e bem definidos que podem ser criticados e reutilizados por outros pesquisadores DYBA *et al.* (2007).

Grande parte das pesquisas existentes se iniciam com uma revisão da literatura. A RS é um tipo de revisão da literatura que sintetiza toda informação existente de forma estruturada e clara. Na RS são realizadas estratégias de buscas e seleção com objetivo de identificar os estudos mais relevantes possíveis para, em seguida, serem analisados conforme os critérios de inclusão e exclusão relativos às questões da pesquisa. Este processo demanda tempo e mais esforço do que as revisões tradicionais da literatura. Uma das grandes vantagens em se realizar RS é o fato desta prover informações sobre um determinado fenômeno de forma robusta e transferível KITCHENHAM (2004).

Pesquisadores e tomadores de decisão são encorajados a usar as pesquisas mais relevantes e recentes possíveis. A RS vai ao encontro desta meta, pois preenche lacunas não exploradas e cria guias para essas pesquisas, possibilitando que as mesmas sejam analisadas no futuro CENTRE FOR REVIEWS AND DISSEMINATION

(2009). A RS destaca-se por algumas características próprias, estas são apresentadas na tabela 2.1.

Tabela 2.1: Principais características da Revisão Sistemática

Características	
✓	A RS se inicia com a criação de um protocolo que aborda todo planejamento, condução e relatório da pesquisa
✓	A RS é baseada em uma estratégia de busca, pré definida, que tem como objetivo encontrar todo estudo relevante à pesquisa
✓	A RS inclui uma estratégia de seleção com todos os estudos incluídos e os excluídos com suas devidas justificativas
✓	A RS é um tipo de estudo com etapas de execução pré definidas, documentadas, de conteúdo robusto e valor científico

A RS é largamente usada em áreas como a medicina, mas não limitada a esta. É comum que outras ciências utilizem esta metodologia para identificar, analisar e publicar estudos sobre determinadas questões em aberto ou pouco exploradas. Outros campos incluem psicologia, pesquisas educacionais, sociologia, gestão de negócios, entre outros. Os pesquisadores dessas áreas estão usando RS para identificar e analisar novos assuntos, abrindo caminho para outras iniciativas de pesquisa e desenvolvimento. CENTRE FOR REVIEWS AND DISSEMINATION (2009).

Este trabalho seguiu as ideias e publicações de Barbara Kitchenham e do Centro de Revisão e Disseminação (CRD) da Universidade York (KITCHENHAM, 2004; CENTRE FOR REVIEWS AND DISSEMINATION, 2009; KITCHENHAM *et al.*, 2009a,b). A tabela 2.2 descreve os passos que devem ser adotados para a realização de uma revisão sistemática da literatura.

Na *etapa de planejamento*, estabelecemos os objetivos da Revisão Sistemática (RS), selecionamos os repositórios de artigos científicos pertinentes e criamos os termos de busca, bem como os critérios para inclusão e exclusão de trabalhos. Na *etapa de execução*, aplicamos os termos de busca nos repositórios de artigos identificados previamente. Aqui, analisamos os resultados à luz de critérios de qualidade para a seleção de trabalhos relevantes. Na *etapa de relatório*, organizamos, resumimos e apresentamos os resultados obtidos nas etapas anteriores. Embora essas etapas da revisão, conforme ilustrado na Tabela 2.2, sejam apresentadas em sequência, é importante destacar que, em muitos casos, elas se desdobram de maneira interativa. Por exemplo, um planejamento inicial pode ser sucedido por uma fase de execução experimental, permitindo ajustes no planejamento original, o que, por sua vez, influencia tanto na execução subsequente quanto na geração dos relatórios.

Tabela 2.2: Etapas da RS

Etapa	Atividade
1	Planejamento: <ul style="list-style-type: none"> ✓ Identificação dos recursos necessários para revisão sistemática ✓ Desenvolvimento do protocolo de revisão, incluindo o planejamento necessário para todas as etapas da revisão
2	Execução: <ul style="list-style-type: none"> ✓ Execução da estratégia de pesquisa definida na fase de planejamento ✓ Seleção de estudos relevantes ✓ Avaliação da relevância dos estudos selecionados ✓ Extração e síntese de informações relevantes
3	Relatórios: <ul style="list-style-type: none"> ✓ Geração de indicadores quantitativos e qualitativos ✓ Relatar os resultados da revisão na forma de relatórios técnicos e artigos científicos

2.1 Planejamento da revisão

Como já exposto acima, a revisão inicia com o desenvolvimento do protocolo de revisão, que descreve os objetivos e as atividades a serem desenvolvidas até o relatório final.

2.1.1 Objetivos da revisão

Os objetivos da revisão são:

- (a) Explorar as análises de *Driving Behavior* e suas aplicações;
- (b) Entender como técnicas de Mineração de Dados podem ajudar no entendimento de *Driving Behavior*;
- (c) Identificar possíveis lacunas de pesquisas nas análises de *Driving Behavior*;

2.1.2 Questões da pesquisa

A segunda etapa do protocolo de revisão é especificar as questões de pesquisa a serem respondidas. Estas questões são apresentadas na Tabela 2.3.

Deve-se notar que as questões da pesquisa estão intimamente alinhadas aos objetivos da pesquisa já descritos. Por exemplo, as questões Q₁, Q₅ e Q₆ estão alinhadas ao objetivo (a), assim como, as questões Q₂, Q₃ e Q₄ estão alinhadas ao objetivo (b). Por fim sem uma ligação direta com alguma questão, o objetivo (c) foi reservado para as oportunidades futuras de pesquisa.

Tabela 2.3: Questões da pesquisa

Id	Questão
Q ₁	Qual arquitetura usada nos projetos de análise de <i>Driving Behavior</i> ?
Q ₂	Como as áreas de Mineração de Dados e <i>Driving Behavior</i> se relacionam?
Q ₃	Quais as técnicas de Mineração de Dados mais usadas em análises de <i>Driving Behavior</i> ?
Q ₄	Como são aplicadas as técnicas de Mineração de dados nas análises de <i>Driving Behavior</i> ?
Q ₅	Quais os dispositivos usados nas coletas de dados de <i>Driving Behavior</i> ?
Q ₆	Como são montados os conjuntos de dados nas análises de <i>Driving Behavior</i> ?

2.1.3 Os termos de busca

A partir das questões da pesquisa Q₁, ..., Q₆, são obtidos os termos de buscas que, por sua vez, são aplicados nos repositórios dos artigos científicos. O termo da busca é uma expressão composta de palavras chaves extraídas das questões da pesquisa, e concatenadas através de booleanos *AND* e *OR*. Na tabela 2.4 são apresentados os termos de buscas usados na revisão.

Tabela 2.4: Termos de busca

Dimensão	Termo
Population	"driving condition", "driving behavior", "driving recognition", "driving event", "driving maneuver", "driving safe", "driving style"
Intervention	"data mining", "machine learning", "pattern recognition", "artificial intelligence"
Comparison	"data analytics", "technique, prediction", "insights", "process", "architecture", "task"
Outcome	"improve safety", "risk", "accident", "accident detection", "anomaly detection", "abnormal driving"

Com base nas ideias de Barbara Kitchenham sobre RS KITCHENHAM *et al.* (2009a), uma estrutura composta de cinco dimensões (ou seja, *Population*, *Intervention*, *Comparison* e *Outcome*) foi usada para organizar os termos de busca na tabela 2.4. Esta é usada para facilitar a leitura e o entendimento do que se deseja pesquisar. Esta metodologia é chamada de P.I.C.O.¹ propostos em KITCHENHAM (2004) e CENTRE FOR REVIEWS AND DISSEMINATION (2009).

A metodologia P.I.C.O., *Population* refere-se à área de conhecimento a ser estudada na pesquisa. A *Intervention* descreve os paradigmas, métodos, técnicas e

¹Palavras concatenadas através do booleano *AND* que formam a expressão de busca.

conceitos aplicados nas populações. A *Comparison* descreve os termos que se deseja comparar. Estes podem ser técnicas, processos, atividades entre outros. Finalmente, o *Outcome* descreve os resultados relacionados a fatores de importância para a pesquisa, como por exemplo, maior confiabilidade, redução de acidentes, aumento da segurança, entre outros.

A expressão de busca usada na pesquisa, está apresentada na tabela 2.5. Esta expressão é resultado da concatenação dos termos de busca já apresentados na tabela 2.4. Pode-se notar que os termos de cada dimensão foram agrupados com booleano *OR* e as dimensões foram agrupadas pelo booleano *AND*.

Tabela 2.5: Expressão de busca

Concatenação dos termos de busca
(“driving condition”OR “driving behavior”OR “driving event”OR “driving maneuver”OR “driving pattern”OR “driving safe”OR “driving style”) AND
(“data mining”OR “machine learning”OR “pattern recognition”OR “artificial intelligence”) AND
(“data analytics”OR “technique, prediction”OR “insights”OR “process”, “architecture”OR “task”) AND
(“improve safety”OR “risk”OR “accident”OR “accident detection”, “anomaly detection”OR “abnormal driving”)

A expressão de busca sofreu diversas alterações ao longo da pesquisa. As primeiras buscas iniciaram-se em Janeiro de 2018 e as últimas em Agosto de 2019. Este processo é iterativo, ou seja, na medida que são retornados os artigos científicos, a expressão vai sofrendo pequenos ajustes. É importante ressaltar que a expressão apresentada sofre variações de acordo com o repositório que era aplicada, pois cada repositório tem suas restrições quanto ao tamanho da expressão e a quantidade de booleanos.

2.1.4 Repositórios de artigos científicos

O próximo passo na etapa de planejamento é definir os repositórios de artigos científicos a serem utilizados na execução da revisão. Importante ressaltar que os repositórios de pesquisas devem ser eletrônicos, virtuais e estarem disponíveis (acessadas via *Internet*), com mecanismos próprios de busca e permissão para uso de operadores lógicos (*OR* e/ou *AND*). Os repositórios escolhidos são bases largamente conhecidas e usadas em pesquisas científicas TRAINA e TRAINA JR (2009). A tabela 2.6 apresenta os repositórios selecionados.

Tabela 2.6: Repositórios de artigos científicos

Repositório	Endereço eletrônico
Springer Link	link.springer.com
Periódicos CAPES	www.capes.gov.br
ISI Web of Science	www.isiknowledge.com
Scopus	www.scopus.com
Science@Direct	www.sciencedirect.com
IEEE Explorer	ieeexplore.ieee.org

2.1.5 Restrições da pesquisa

Como o inglês é a língua padrão para publicações científicas, a revisão relatada neste trabalho considera apenas os trabalhos escritos nesta língua. Entre eles, serão selecionados somente artigos publicados em revistas, conferências, congressos e *workshops*. Além disso, esses documentos devem ter sido publicados entre os meses de Janeiro de 2008 e Abril de 2019.

A revisão deste trabalho abrange cerca de onze anos de pesquisa acadêmica e científica e, sempre que possível, foram aplicados filtros nos repositórios de artigos para restringir algumas áreas de interesse como: ciência da computação, engenharia automotiva, ciência da decisão, negócios e administração. Além disso, os artigos selecionados dos repositórios necessitam satisfazer determinados critérios de inclusão e exclusão, discutidos a seguir.

2.1.6 Critérios para seleção dos artigos

Esta fase é muito importante para a revisão. Aqui, todos os estudos retornados das buscas aplicadas nos repositórios passam por duas etapas de seleção. A primeira etapa é mais simples, e visa descartar artigos sem relação alguma com a pesquisa. Já na segunda etapa os artigos passam por filtros mais refinados, adotando-se alguns critérios que garantam a qualidade da pesquisa. A seguir serão detalhados os dois critérios utilizados neste trabalho.

1^a Seleção de artigos

Nesta etapa tem início a primeira seleção dos artigos retornados da expressão de busca. O critério nesta fase é relativamente simples, verifica-se se os mesmos estão relacionados com a pesquisa, a partir da leitura do título, resumo e conclusão dos artigos retornados. Esta relação deve ser coerente com as questões da pesquisa, ou seja, os estudos devem abordar, pelos menos: *Driving behavior* e Mineração de dados. Os estudos que não tiverem relação alguma com a pesquisa serão descartados.

2ª Seleção de artigos

Na segunda fase de seleção, serão incluídos os artigos escolhidos na etapa anterior (seção 2.1.6). Nessa etapa, a escolha dos artigos é embasada em um processo de avaliação por pontuação. Os artigos selecionados passarão por uma análise mais aprofundada, além de considerar o título, o resumo e a conclusão. O objetivo é determinar se cada artigo responde integralmente às questões de pesquisa, permitindo categorizar as respostas como “SIM”, “PARCIAL” ou “NÃO”. Essa abordagem visa simplificar a elaboração das fichas catalográficas para o relatório final. Vale mencionar que as perguntas foram levemente adaptadas para facilitar a compreensão, sem comprometer o significado original. A Tabela 2.7 apresenta as perguntas a serem respondidas nesta etapa.

Tabela 2.7: Questões para avaliação de qualidade dos artigos

Id	Questão para Qualidade
QQ ₁	O artigo cita que tipo de arquitetura é utilizada nos projetos de análise de <i>Driving Behavior</i> ?
QQ ₂	O artigo apresenta como áreas de Mineração de Dados e <i>Driving Behavior</i> se relacionam?
QQ ₃	O artigo apresenta quais técnicas de Mineração de Dados e/ou Aprendizado de Máquina são usados em análises de <i>Driving Behavior</i> ?
QQ ₄	O artigo cita como são aplicadas técnicas de Mineração de dados e/ou Aprendizado de Máquina nas análises de <i>Driving Behavior</i> ?
QQ ₅	O artigo apresenta quais são dispositivos usados nas coletas de dados de <i>Driving Behavior</i> ?
QQ ₆	O artigo cita como são montados os conjuntos de dados para análises de <i>Driving Behavior</i> ?

A medida que os artigos vão sendo lidos cada pergunta da tabela 2.1.5 deve ser respondida. As respostas são relacionadas com uma pontuação que varia de 0 a 2. Um artigo pode receber no mínimo zero e no máximo doze pontos. Na tabela 2.8 apresenta-se as respostas e a pontuação ligada à mesma.

Tabela 2.8: Respostas e pontuação

Respostas	Pontuação
SIM	2
PARCIAL	1
NÃO	0

A pontuação final é resultado da soma de cada resposta relativa às perguntas da tabela 2.7. Nesta etapa os artigos que não alcançarem o mínimo de dois pontos serão descartados. Os artigos selecionados seguem para a fase final da pesquisa.

Estes farão parte da ficha catalográfica do relatório final, que são as perguntas da tabela 2.7 preenchidas com as informações retiradas de cada artigo.

2.2 Execução da revisão

Aqui, inicia-se, de fato, a execução de tudo que foi planejado na seção 2.1. Esta seção descreve as etapas de execução da revisão, como a aplicação da expressão de busca (tabela 2.5) nos repositórios apresentados na tabela 2.6 e a execução dos critérios de seleção dos artigos e análises dos estudos retornados.

2.2.1 Identificação e seleção dos artigos

Durante os meses de Abril de 2018 a Junho de 2018 a expressão de busca (tabela 2.5) foi aplicada em cada repositório (tabela 2.6). Conforme já mencionado, por se tratar de um processo interativo foram executadas diversas rodadas de busca que provocaram ajustes, tanto nas etapas de planejamento como nas etapas de execução da pesquisa.

Nesta etapa foram aplicadas as restrições da pesquisa (seção 2.1.5) e a 1ª seleção de artigos (seção 2.1.6). Aqui foram executadas as primeiras buscas, seleções e coleta de estudos, assim como, foram aplicados ajustes nos mecanismos de busca de cada repositório, filtros por anos, tipos de estudos e o idioma.

No final desta etapa, as buscas retornaram 291 artigos científicos, cujos resultados da 1ª seleção, são apresentados na tabela 2.9, abaixo:

Tabela 2.9: Resultados da 1ª seleção

Estudos	Quantidade
Aceitos	107
Rejeitados	166
Duplicados	18

2.2.2 Avaliação de qualidade dos artigos retornados

Esta etapa, uma das mais importantes da revisão, ocorreu entre os meses de Agosto e Janeiro de 2019. Aqui são aplicadas estratégias para avaliação da qualidade dos estudos retornados da fase anterior (seção 2.2.1). Esta estratégia foi definida no protocolo de revisão e está descrita na seção 2.1.6. Neste estágio, ocorrem leituras mais detalhadas de cada artigo, onde informações de cada estudo, tais como: objetivo do artigo, metodologia aplicada, resultados encontrados, limitações da pesquisa e conclusão, são importantes, pois ajudarão nas respostas do questionário apresentado na tabela 2.7.

Importante ressaltar que não existe um consenso entre os pesquisadores do que seja qualidade, o que existem são sugestões. Neste trabalho, a avaliação de qualidade foi aplicada nos estudos que retornaram da seção anterior, onde cada um passou por um questionário com perguntas relacionadas às questões da pesquisa e, as respostas, ligadas a um sistema de pontuação. Os artigos selecionados foram os que alcançaram acima de 1 (um) ponto. Na tabela 2.10 são apresentados os resultados.

Tabela 2.10: Resultados da 2^a seleção - Avaliação de qualidade

Estudos	Quantidade
Aceitos	96
Rejeitados	11

Em janeiro de 2019, concluímos a inclusão de 96 (noventa e seis) artigos que formaram a base integral da nossa revisão sistemática, e forneceram os elementos para a elaboração das fichas catalográficas. Por meio desses estudos, fomos capazes de abordar as questões centrais da nossa pesquisa, como apresentado na Tabela 2.3.

2.2.3 Extração dos dados

Nesta seção são preenchidas as fichas catalográficas, com os dados dos estudos retornados da seção 2.2.2 relativos à avaliação de qualidade. A ficha é uma planilha simples contendo todos os estudos retornados com o nome, um pequeno resumo sobre o estudo, e as respostas referentes a pesquisa da tabela 2.7. Na figura 2.1 é apresentado um exemplo de ficha catalográfica que será utilizada neste trabalho.

Figura 2.1: Exemplo de ficha catalográfica a ser utilizada neste trabalho

2.3 Relatório da revisão

Nesta seção, consolidaremos e sintetizaremos todo o conhecimento adquirido por meio da revisão. O propósito do relatório é estabelecer a base para o Capítulo 2 da tese, auxiliar na identificação de oportunidades de pesquisa e também servir como alicerce teórico para futuros artigos científicos. Conforme delineado por KITCHENHAM (2004), um relatório de revisão segue uma estrutura que engloba informações essenciais da pesquisa, como título, autores, resumo, questões abordadas, métodos empregados, resultados obtidos e outros elementos relevantes. Para o escopo deste trabalho, o relatório servirá como uma síntese do que foi descoberto ao longo da pesquisa.

2.3.1 Resultados Encontrados

Nesta seção, serão apresentados os resultados obtidos até fevereiro de 2019, consolidando as respostas extraídas dos 96 (noventa e seis) artigos aprovados e aceitos. Todas as informações levantadas são pertinentes às questões de pesquisa, conforme detalhado na Tabela 2.3.

Questão Q₁: Qual arquitetura usada nos projetos de análise de *Driving Behavior*?

Esta seção tem por objetivo investigar as abordagens arquiteturais adotadas nos projetos de análise de *Driving Behavior*. Compreender a estrutura desses projetos é fundamental para discernir suas propostas. A compreensão da arquitetura de um sistema é de grande relevância, conforme afirmado por JAAKKOLA e THALHEIM (2011). A arquitetura de um sistema consiste em um modelo conceitual que delinea a estrutura, o contexto e suas perspectivas. Além disso, a arquitetura desempenha um papel crucial na análise, no planejamento e no desenvolvimento de sistemas.

No contexto deste estudo, a arquitetura é examinada no sentido de como os autores, dos artigos selecionados, projetaram seus experimentos e seus sistemas. Para alcançar essa compreensão, adotamos uma abordagem resumida da arquitetura, focando em dois tipos principais: Arquitetura de *Software* e Arquitetura de *Hardware*. Conforme delineado por RECHTIN (1992), estes tipos abordam diferentes perspectivas:

- (1) Arquitetura de *Software* - Refere-se a uma visão de alto nível da estrutura do sistema, englobando elementos de *software*, suas inter-relações e características.
- (2) Arquitetura de *Hardware* - Envolve a visão física do sistema, englobando seus componentes e as relações entre eles.

A grande maioria dos trabalho revisados citam a arquitetura dos seus projetos. Alguns apontam a arquitetura do sistema, outros referem-se à arquitetura dos algoritmos utilizados, ou focam na arquitetura física montada. Alguns não mencionam nenhum deles. Pode-se destacar o estudo de MESEGUER *et al.* (2013), onde o autor propõe um sistema para classificação de *Driving Behavior* e a via que o veículo circula, utilizando uma rede neural. Na figura 2.2 é apresentada a arquitetura proposta no estudo. Neste estudo o autor separa a arquitetura em *software* (itens 2,3 e 4 da figura 2.2), onde estão compostas as aplicações que coletam os dados de direção do motorista (aceleração, velocidade, R.P.M.) e para, depois, enviá-las a um servidor que executará suas análises referentes ao comportamento da direção. Já na arquitetura de *hardware* (item 1 da figura 2.2) o autor usa um dispositivo chamado ODB-II² instalado no veículo. Este, por sua vez, transmite os sinais coletados para um celular onde está instalado a aplicação de coleta dos sinais.

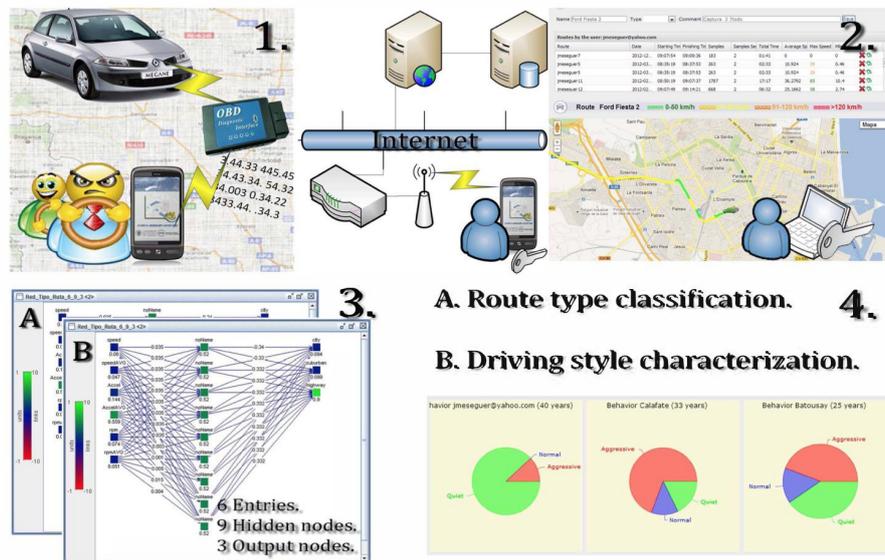


Figura 2.2: Arquitetura proposta por MESEGUER *et al.* (2013)

No artigo CASTIGNANI *et al.* (2015) autor também menciona a arquitetura utilizada. Nele, é proposto um sistema para detecção de eventos considerados “perigosos”. No artigo, a arquitetura mais detalhada é a de *software*, em que o autor apresenta como irá funcionar o sistema proposto. A figura 2.3 mostra a arquitetura do sistema, nota-se que o sistema proposto coleta os sinais de direção através dos sensores instalados no veículo, depois utiliza um sistema *Fuzzy* para inferir que tipo de manobra arriscada foi executada pelo motorista. Apesar de não mencionar diretamente a arquitetura de *hardware*, fica claro que foram usados sensores como acelerômetro, giroscópio, magnetômetro e GPS, todos pertencentes a um celular

² *On-board diagnostics* (ODB), interface usada para coleta de diversos sinais de um veículo que tenha o padrão SAE J1962 implantado.

instalado no veículo.

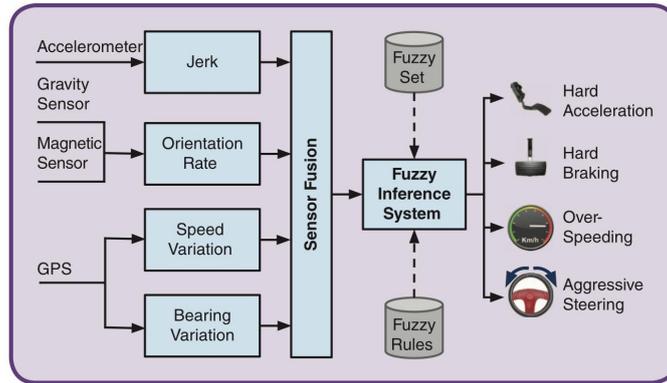


Figura 2.3: Arquitetura proposta por CASTIGNANI *et al.* (2015)

Em AL-DOORI *et al.* (2014) o autor propõe um modelo de inteligência artificial, baseada em uma rede neural e lógica *fuzzy*, para classificar *Driving Behavior* com objetivo de otimizar o uso da bateria de carros elétricos. No artigo é citado a arquitetura do sistema para representar o modelo, isto está claro na figura 2.4. Neste estudo está aparente a visão da arquitetura de *software* (figura 2.4). Já na arquitetura de *hardware*, apesar de não ser detalhada, o autor cita que os sinais usados pelo sistema foram coletados da *CAN bus*³ do veículo.

Diversos outros estudos mencionam arquitetura de *software* e *hardware*. A primeira é sempre a mais detalhada, pode-se citar os estudos de: CASTIGNANI *et al.* (2015), EREN *et al.* (2012), ARAÚJO *et al.* (2012), VASCONCELOS *et al.* (2017), CARMONA *et al.* (2015), WU *et al.* (2016), YLIZALITURRI-SALCEDO *et al.* (2015), FUGIGLANDO *et al.* (2018) e JÚNIOR *et al.* (2017). Foi constatado a partir das análises das arquiteturas (*software* e *hardware*), que grande parte dos estudos adotaram um estilo de arquitetura conhecido como “Cliente/Servidor” (segundo RECHTIN (1992)). Neste cenário, o cliente é representado pelo veículo onde fica instalado os sensores, e estes, por sua vez, enviam os sinais coletados para algum servidor responsável pelas análises *Driving Behavior*.

Questão Q₂: Como as áreas de Mineração de Dados e *Driving Behavior* se relacionam?

Um dos objetivos deste trabalho é entender como a área da Mineração de Dados pode auxiliar no entendimento de *Driving Behavior* (seção 2.1.1). Nesse sentido, é imprescindível verificar como as duas disciplinas se relacionam, ou seja, entender se nas análises empregadas de *Driving Behavior* dos artigos selecionados, seguem, de

³ *Controller Area Network (CAN bus)* é o padrão que especifica o barramento interno dentro dos veículos modernos (a partir 1994). Este barramento interliga diversos microcontroladores que capturam sinais de vários dispositivos dos carros (freios, aceleração, volante entre outros).

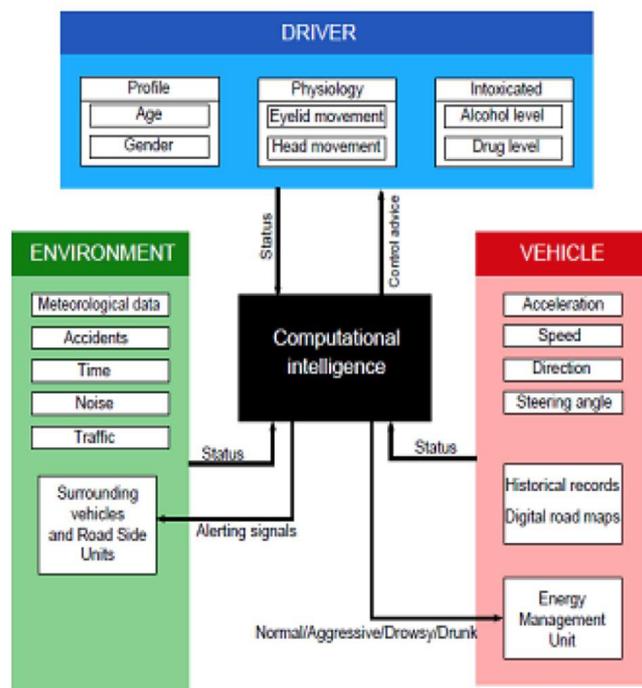


Figura 2.4: Arquitetura proposta por AL-DOORI *et al.* (2014)

alguma forma, os conceitos, processos e arquitetura da Mineração de dados. Para isto, a compreensão da mineração de dados para este trabalho deve estar esclarecida.

Segundo MADNI *et al.* (2017) Mineração de dados é usada para transformar dados brutos em informação útil ou em conhecimento. O termo é largamente usado para se referir a um processo de coleta, processamento e aprendizado a partir de bases de dados TAYLOR *et al.* (2016). Em HAN *et al.* (2011) o autor cita que Mineração de Dados pode ser definida de várias maneiras. Muitos empregam o termo como sinônimo de *knowledge discovery from data* (KDD), já outros acreditam que seja apenas mais um passo dentro do processo de KDD. Essas duas visões não são inteiramente antagônicas e sim apenas formalidades conceituais. A figura 2.5 mostra os processos do KDD.

Para este trabalho optou-se simplificar a definição de Mineração de dados e então associá-la a disciplina de KDD como um sinônimo. Desta forma, os processos de KDD apresentados na figura 2.5 estarão relacionados com a Mineração de dados. Este relacionamento irá contribuir para um melhor entendimento dos artigos analisados, uma vez que facilita a extração de conhecimento oriundo da aplicação das técnicas de limpezas, das técnicas de transformação e das técnicas de mineração. A figura 2.6 apresenta a visão da Mineração de dados utilizada neste trabalho.

A partir das premissas expostas, todos os artigos selecionados foram analisados quanto aos processos apresentados na figura 2.6. Estes são detalhados a seguir:

- (1) **Integração dos dados** - Esta etapa representa a coleta de dado bruto a partir

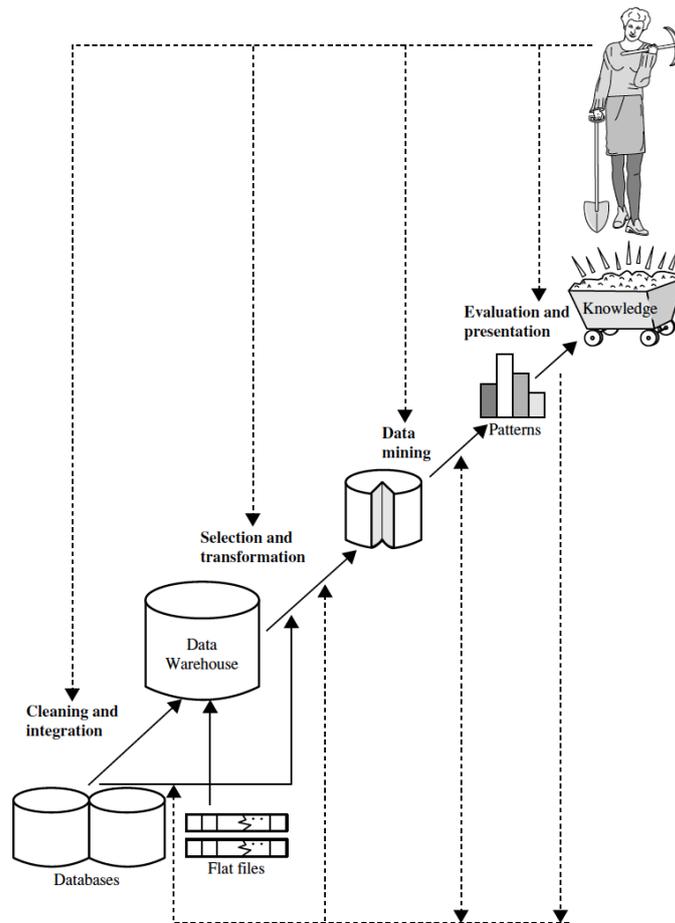


Figura 2.5: Processos do KDD segundo HAN *et al.* (2011)

de uma ou mais fontes. Em caso de mais de uma fonte de dados estas devem ser consolidadas em uma única;

- (2) **Pré processamento** - Nesta etapa ocorrem várias sub-etapas como: Limpeza dos dados com a eliminação de inconsistências, transformações dos dados para formatos e visões mais apropriadas, redução dos dados para um melhor processamento e análises, amostragens de dados entre outras;
- (3) **Processamento** - Aqui, os dados são analisados e acontece, de fato, a aplicação das técnicas de Mineração. O objetivo é identificar padrões para que possamos, de alguma forma, classificá-los. Nesta etapa as informações coletadas são transformadas em conhecimento;
- (4) **Pós processamento** - Etapa onde ocorre a apresentação do que foi descoberto e o conhecimento adquirido pela mineração.

A grande maioria dos artigos que mencionam alguma técnica da Mineração de dados em seu conteúdo, não citam diretamente seus processos. Outros já são mais detalhistas quanto aos processos, pode-se destacar o trabalho de BANERJEE *et al.*

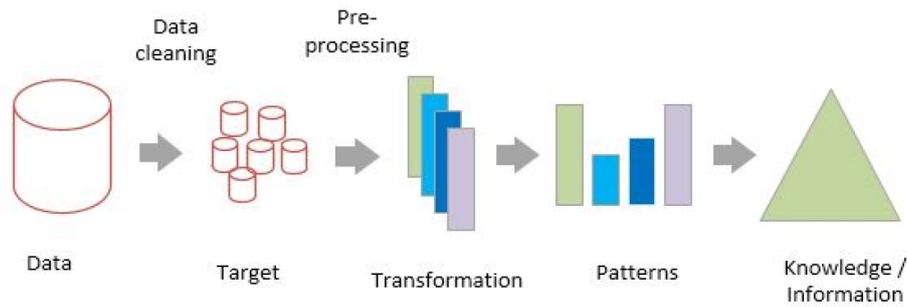


Figura 2.6: Processos de Mineração de Dados utilizada neste trabalho

(2014) que propõe um novo método para identificar e classificar *Driving Behavior* de acordo com o segmento da via onde se trafega em um determinado momento. Neste trabalho, os autores citam os processos de Mineração de dados como etapas dentro da metodologia proposta, deixando claro as etapas de integração de dados, pré processamento (limpeza e seleção de dados) e processamento (classificação de *Driving Behavior*). A figura 2.7 mostra a metodologia proposta por eles.

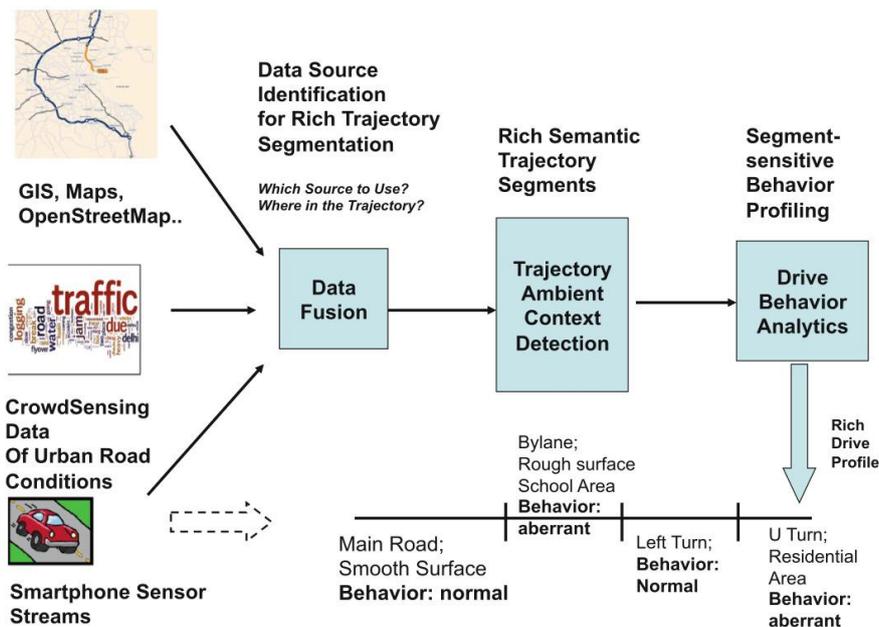


Figura 2.7: Metodologia proposta por BANERJEE *et al.* (2014) com alguns processos de Mineração aparentes

Em MESEGUER *et al.* (2013) o autor propõe um sistema para classificação de *Driving Behavior* e vias, a partir de uma rede neural. Da mesma forma, neste trabalho, fica clara a relação com Mineração de Dados, já que o estudo cita os processos de coleta e integração de dados, além das etapas de pré processamento como limpeza e normalização dos dados; as etapas de processamento, com a implementação da rede neural, e por fim, as etapas de pós processamento, com a apresentação dos resultados das classificações de *Driving Behavior* propostas. A figura 2.2 apresenta

uma visão dos processos utilizados.

Em KALSOOM e HALIM (2013) os autores propõe um sistema de classificação de *Driving Behavior* por meio de *clustering* de determinados sinais de direção para então agrupá-los por tipos, no caso “Normal” ou “Perigoso”. Neste estudo são apenas citados os processos de integração de dados e processamento. A figura 2.8 mostra a visão do sistema de classificação, pode-se notar os processos integração e coleta de dados da Mineração de Dados.

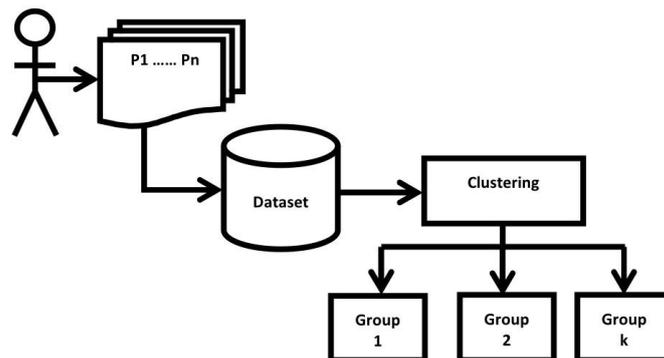


Figura 2.8: Classificação de *Driving Behavior* proposta por KALSOOM e HALIM (2013)

Em SAIPRASERT *et al.* (2017) são propostos três algoritmos para detecção de *Driving Event* críticos. Neste trabalho a visão de processos da Mineração de dados é utilizada, principalmente, na descrição da coleta e integração de dados, nas técnicas de pré processamento e nas técnicas de processamento. Na figura 2.9 é apresentado um dos algoritmos propostos no estudo. Aqui fica clara a preocupação do autor com os processos da Mineração de dados.

Em TAYLOR *et al.* (2016) o autor propõe o uso de técnicas de Mineração de dados, usando os sinais da *CAN bus* de um veículo para classificar o tipo de via que está trafegando. Neste artigo é citado o uso de todas as etapas da Mineração de dados, apesar do trabalho focar apenas nas técnicas de pré processamento como: seleção e extração de dados. Diversos outros estudos como em: VASCONCELOS *et al.* (2017), WAHAB *et al.* (2007), MA e ANDREASSON (2007), CHONG *et al.* (2013), CASTIGNANI *et al.* (2015), LIN *et al.* (2014), ENGELBRECHT *et al.* (2015), CARPATOREA *et al.* (2016), AL-DOORI *et al.* (2014), ALVAREZ *et al.* (2014), WU *et al.* (2016) e FUGIGLANDO *et al.* (2018), mencionam o uso de técnicas de Mineração de Dados, porém com foco na etapa de processamento de dados, ou seja, todos discutem a aplicação das técnicas de processamento em seus algoritmos. Em resumo, a interconexão entre Mineração de Dados e *Driving Behavior* é representada por um processo unificado, englobando a coleta de dados, a aplicação de técnicas de pré-processamento, a análise baseada nos *insights* gerados pelas téc-

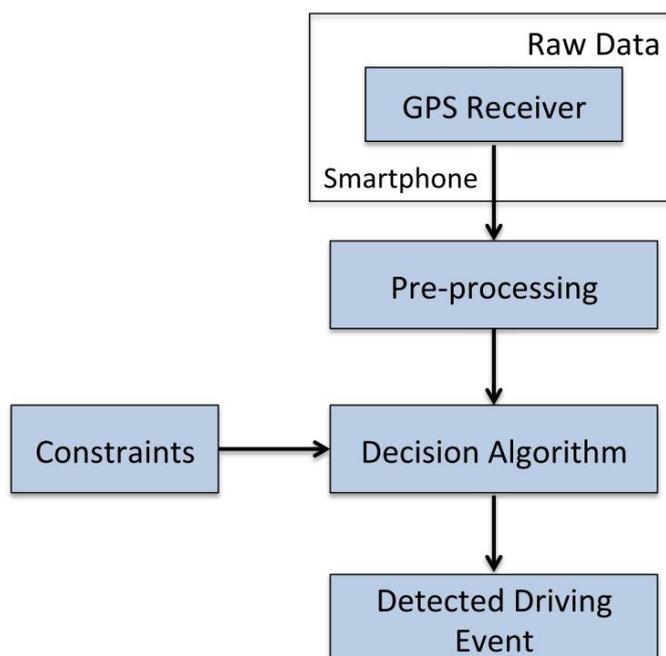


Fig. 2 Rule-based algorithm

Figura 2.9: Algoritmo proposto por SAIPRASERT *et al.* (2017) com a visão de processos de Mineração aparente

nicas de processamento e, por fim, a apresentação do conhecimento adquirido. O foco primordial reside na utilização das ferramentas da Mineração de Dados, para transformar dados brutos em entendimentos valiosos, aprofundando a compreensão do *Driving Behavior* e ampliando sua aplicabilidade em diversos cenários.

Questão Q₃: Quais as técnicas de Mineração de Dados mais usadas em análises de *Driving Behavior*?

Para abordar a questão Q₃ sobre as técnicas de Mineração de Dados mais utilizadas nas análises de *Driving Behavior*, é importante esclarecer que as técnicas a serem discutidas estão limitadas às etapas de “Pré processamento” e “Processamento”. Nesses estágios, uma variedade de abordagens são empregadas, e as tabelas 2.11 e 2.12 fornecem um panorama das técnicas predominantes relacionadas a cada estudo. Nas tabelas 2.11 e 2.12, podemos identificar 52 (cinquenta e dois) artigos que diretamente mencionam o emprego de técnicas de mineração e/ou aprendizado de máquina. Importante destacar que a ausência de certos artigos nas tabelas não implicam que não estejam relacionados à pesquisa, uma vez que muitos deles abordam o tema, mas não discutem explicitamente técnicas de Mineração de Dados. Alguns artigos podem empregar técnicas estatísticas ou métricas para delimitar condições específicas. Essas abordagens também são válidas para inclusão, uma vez que este estudo

Tabela 2.11: Técnicas de pré-processamento mais usadas

Processo	Técnica	Artigos
Pré processamento	Filtro de Kalman	CASTIGNANI <i>et al.</i> (2015); WU <i>et al.</i> (2016); MA e ANDREASSON (2007); MEIRING e MYBURGH (2015); CASTIGNANI <i>et al.</i> (2013)
	Filtro <i>Savitzky-Golay</i>	BAGDADI (2013)
	Normalização de dados	MESEGUER <i>et al.</i> (2013); CARPATOREA <i>et al.</i> (2016); FUGIGLANDO <i>et al.</i> (2018); ARAÚJO <i>et al.</i> (2012); ALVAREZ <i>et al.</i> (2014); YADAV <i>et al.</i> (2019); CASTRO e KIM (2016)
	<i>Moving Average</i>	SAIPRASERT <i>et al.</i> (2017); ENGELBRECHT <i>et al.</i> (2015); JOHNSON e TRIVEDI (2011); EREN <i>et al.</i> (2012); MA e ANDREASSON (2007); WU <i>et al.</i> (2016); PHOLPRASIT <i>et al.</i> (2015)
	<i>Linear discriminant analysis (LDA)</i>	SATHYANARAYANA <i>et al.</i> (2012)
	<i>Principal components analysis (PCA)</i>	CASTIGNANI <i>et al.</i> (2015); TAYLOR <i>et al.</i> (2016); FUGIGLANDO <i>et al.</i> (2018); BA <i>et al.</i> (2017); HALIM <i>et al.</i> (2016b)
	Interpolação	SAIPRASERT <i>et al.</i> (2017); FAZEEN <i>et al.</i> (2012); KLUGER <i>et al.</i> (2016)
	Filtro <i>Low-pass</i>	JOHNSON e TRIVEDI (2011), WU <i>et al.</i> (2016)
	Filtro <i>High-pass</i>	FAZEEN <i>et al.</i> (2012); EREN <i>et al.</i> (2012)
	Detecção de <i>Outlier</i>	FUGIGLANDO <i>et al.</i> (2018); VASCONCELOS <i>et al.</i> (2017)

busca compreender as análises de *Driving Behavior*, porém, na questão Q₃, estamos particularmente interessados em técnicas de Mineração e/ou aprendizado de máquina.

No contexto das análises de *Driving Behavior*, observamos que técnicas como a normalização e suavização dos dados (por meio do filtro de *Kalman* e *moving average*), são amplamente mencionadas. Isso está relacionado à natureza frequentemente bruta dos dados coletados (séries temporais), exigindo um tratamento preliminar. A tabela 2.11 sugere que, no mínimo, as etapas de normalização, transformação e seleção de dados são aplicadas. Uma análise semelhante pode ser aplicada às técnicas de processamento, onde se observa que algoritmos de aprendizado de máquina, tais como *Support Vector Machine (SVM)*, *Artificial Neural Network (ANN)*, *K-means* e *Random Forest*, são os mais referenciados. A maioria dos estudos trata da classificação de situações perigosas de direção com base no *Driving Behavior*, frequentemente empregando abordagens de aprendizado supervisionado. Essa tendência justifica o uso disseminado dessas técnicas nesse contexto.

Destaca-se ainda que as análises de *Driving Behavior* frequentemente lidam com dados brutos provenientes de sensores, especialmente séries temporais. Esses dados demandam um processamento inicial para permitir a classificação com base em padrões específicos. Assim, podemos concluir que, em última instância, as análises de *Driving Behavior* estão intrinsecamente ligadas a um problema de classificação de séries temporais.

Questão Q₄: Como são aplicadas as técnicas de Mineração de dados na classificação de *Driving Behavior*?

A análise da aplicação das técnicas de Mineração de Dados, na classificação de *Driving Behavior*, estende-se diretamente da questão anterior. Nas tabelas 2.11 e 2.12 encontram-se as técnicas frequentemente empregadas nos estudos selecionados. A maneira como essas técnicas são aplicadas varia entre os artigos, dependendo dos procedimentos conduzidos pelos autores e de seus objetivos específicos. Após examinar os cinquenta e dois estudos, surge um padrão evidente na forma como essas técnicas são empregadas, levando em consideração os requisitos e objetivos individuais de cada um.

Para uma visão ampla, a figura 2.10 ilustra um modelo geral do *Driving Behavior*, resultado das análises dos artigos selecionados. É importante observar que a maioria dos artigos destaca as tarefas de **identificação** e **classificação** dos comportamentos de direção dos motoristas em seus contextos de pesquisa.

No que diz respeito à identificação, trata-se da seleção dos dados que contêm eventos potenciais, sejam eles perigosos ou não. O dado bruto geralmente é uma série temporal, frequentemente dividida em janelas de tamanhos variados contendo, tanto manobras, e eventos críticos ou não. Assim, é fundamental que o sistema de *Driving Behavior* possa isolar a porção relevante contendo os eventos, para posterior divisão em pequenas séries temporais. Enquanto alguns estudos mencionam abertamente suas abordagens de identificação, outros não o fazem, embora a identificação seja uma etapa obrigatória. No geral, a identificação baseia-se em regras que empregam diversas métricas (como desvio padrão, média e *jerk*), onde o dado extraído do sinal (que gerou um determinado evento), é comparado às regras para determinar o tamanho do evento, seu início e fim. Por exemplo, os estudos SAIPRASERT *et al.* (2017); CARMONA *et al.* (2015); YLIZALITURRI-SALCEDO *et al.* (2015); PAEFGEN *et al.* (2012), não citam a identificação diretamente, mas mencionam que o dado bruto foi tratado e dividido, antes da classificação. Já em JOHNSON e TRIVEDI (2011); PHOLPRASIT *et al.* (2015), a identificação é bem descrita, aqui o dado bruto passa por uma etapa que utiliza algumas métricas (*simple moving average* (SMA), desvio padrão e média), para verificar que pedaços ultrapassam um determinado limite. Os dados que ultrapassaram são separados e considerados importantes para serem classificados. Esta técnica é uma forma eficiente dos algoritmos se preocuparem apenas com partes de dados que detêm informações relevantes.

A etapa de classificação, por sua vez, lida com a avaliação da criticidade das manobras ou eventos. Nossos estudos indicam que essa tarefa é conduzida por meio de várias abordagens, incluindo o uso de bancos de exemplos, regras/limites e modelos treinados de aprendizado de máquina. As abordagens com limites são mais

comuns, onde a série temporal proveniente da etapa de identificação, é comparada a um limite específico. Se o limite for excedido, o evento é classificado como crítico. Por exemplo, SAIPRASERT *et al.* (2017); MOHAMAD *et al.* (2011); JOHNSON e TRIVEDI (2011); CARMONA *et al.* (2015); MA e ANDREASSON (2007); PA-EFGEN *et al.* (2012), usam sinais de GPS e acelerômetro para determinar se um sinal ultrapassa um limite específico e, caso afirmativo, classificam o evento como perigoso. Abordagens baseadas em regras também são usadas, como CASTIGNANI *et al.* (2015, 2013, 2014), onde regras de lógica *Fuzzy* são empregadas para atribuir pontuações aos eventos e determinar sua criticidade.

As abordagens que utilizam bancos de exemplos são mais elaboradas e precisas. Nesse caso, os sinais identificados são comparados a exemplos de manobras e eventos, onde algoritmos de comparação de séries temporais determinam a classificação dos mesmos. SAIPRASERT *et al.* (2017); JOHNSON e TRIVEDI (2011); YLIZALITURRI-SALCEDO *et al.* (2015); PHOLPRASIT *et al.* (2015), usam a técnica *Dynamic Time Warping* (DTW) para essa comparação.

Finalmente, surge a última abordagem, que representa tanto o cerne moderno quanto o foco principal do nosso estudo. Nessa abordagem, a classificação do sinal ocorre mediante a utilização de um modelo pré-treinado de inteligência artificial, empregado para categorizar os dados MESEGUER *et al.* (2013); AL-DOORI *et al.* (2014); QU *et al.* (2017); CHONG *et al.* (2013); MEIRING e MYBURGH (2015); GUOYING (2016). Essa estratégia se revela a mais complexa, variando consideravelmente entre os diferentes estudos. A maioria desses estudos utiliza técnicas de aprendizado de máquina para adquirir um entendimento dos eventos críticos, o que lhes permite criar seus próprios modelos personalizados.

Dentro dessa categoria, revisões como aquelas realizadas por HALIM *et al.* (2016a) e MEIRING e MYBURGH (2015), exploram extensivamente a aplicação de técnicas de mineração de dados e aprendizado de máquina para identificar padrões de comportamento de direção arriscada. Tais revisões destacam notavelmente a relevância de abordagens com uso de *artificial neural network* (ANN) e *support vector machine* (SVM). Adicionalmente, SATHYANARAYANA *et al.* (2012) propõem a utilização da rede de área do veículo (*controller area network* - CAN), em conjunto com diversas técnicas de aprendizado de máquina, como SVM, K-nn e *linear discriminant analysis* (LDA), para classificar uma ampla gama de manobras (curvas à direita e esquerda, mudanças de faixa, pilotagem em linha reta, paradas, entre outras). Além disso, em FUGIGLANDO *et al.* (2018), os autores empregam os dados da CAN dos veículos, combinando-os com técnicas de clusterização, a fim de agrupar diferentes tipos de motoristas com base na gravidade de suas condutas. Por sua vez, MESEGUER *et al.* (2013) e WANG *et al.* (2017) utilizam técnicas de ANN para classificar o tipo de *Driving Behavior*, avaliando se ele é agressivo ou não.

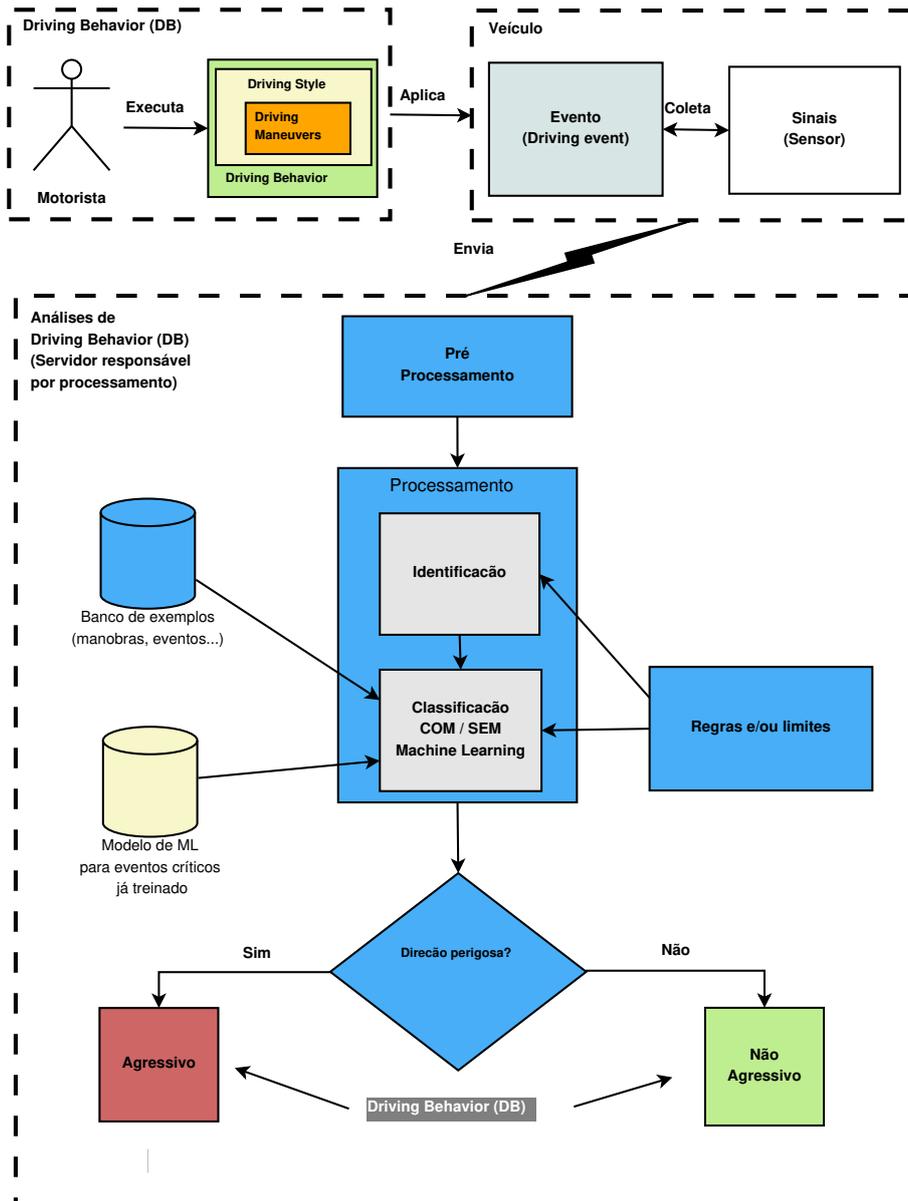


Figura 2.10: Modelo de análise de *Driving Behavior* encontrado neste trabalho

A representação visual do modelo de *Driving Behavior* na figura 2.10, fornece uma visão integrada das estratégias comuns adotadas nos estudos de análises de *Driving Behavior*. O modelo se divide em três elementos-chave: **motorista**, **veículo** e **central de análise** de *Driving Behavior*. No nível do **motorista**, ocorrem as manobras de direção (*Driving Maneuvers*) e o estilo de direção (*Driving Style*), ambos contribuindo para o *Driving Behavior*. Essas ações desencadeiam eventos de direção (*Driving Event*) no **veículo**, onde sensores capturam os sinais relevantes, como aceleração, frenagem, *jerk* e ângulo da direção. Os sinais coletados são transmitidos à **central de análise**, onde ocorrem as etapas de Mineração de Dados: pré processamento e processamento. Essas etapas são delineadas nas tabelas 2.11 e 2.12. Alguns estudos, como CASTIGNANI *et al.* (2015) e TOLEDO *et al.* (2008),

discutem as características dessa **central**, que normalmente inclui servidores, um gerenciador de banco de dados e algoritmos de classificação de *Driving Behavior*. O objetivo final é identificar e classificar o *Driving Behavior* de um motorista após uma viagem.

A relação entre *Driving Behavior*, *Driving Style*, *Driving Event* e *Driving Maneuvers*, será explorada na seção subsequente. Esses elementos são essenciais nas análises de *Driving Behavior*, e compreender essa relação ajuda engenheiros e pesquisadores a desenvolverem suas arquiteturas e sistemas. Para isso, é crucial entender os participantes desse processo (conforme a figura 2.10), e como cada um contribui para a análise de *Driving Behavior*. Como exemplo, CASTIGNANI *et al.* (2015) utiliza manobras (*Driving Maneuvers*) e eventos de direção (*Driving Event*), para criar um sistema de identificação de manobras e classificação de eventos como seguros ou arriscados. Em ENGELBRECHT *et al.* (2015), os autores buscam identificar comportamentos de direção perigosa para melhorar a segurança rodoviária, propondo um sistema de classificação de *Driving Behavior* que identifica *Driving Events* em janelas de tempo e, a partir disso, classifica as manobras correspondentes. Além disso, JOHNSON e TRIVEDI (2011) desenvolve um sistema que alerta o motorista sobre manobras perigosas. Nesses exemplos, todas as abordagens empregam técnicas de Mineração de Dados (especificamente *Machine Learning*), para a classificação de comportamentos de direção.

Outro elemento fundamental na classificação do *Driving Behavior* são os sensores, responsáveis por capturar os sinais dos veículos. Os sensores desempenham um papel essencial na identificação e classificação, fornecendo sinais fundamentais. Consequentemente, uma boa avaliação de *Driving Behavior* envolve a seleção dos sinais mais relevantes para que a classificação se torne mais confiável e robusta. Embora não haja um consenso sobre quais sinais são ideais, a literatura sugere que a escolha depende das metas do estudo. Em TOLEDO *et al.* (2008), por exemplo, os sinais de aceleração e velocidade são utilizados para detectar comportamentos perigosos e reduzir acidentes. Em WAHAB *et al.* (2007), os autores focam na segurança rodoviária, usando sinais como a posição do acelerador, do freio e a velocidade. Em ARAÚJO *et al.* (2012) e MESEGUER *et al.* (2013), o consumo de combustível é o alvo, com sinais como velocidade, aceleração, consumo e RPM sendo cruciais. Alguns estudos podem não mencionar explicitamente os sinais, mas abordam os sensores e os eventos, permitindo inferir os sinais subjacentes. Os sinais mais comuns identificados neste estudo são:

- (1) *Steering wheel angle*
- (2) *Yaw rate*
- (3) *Braking (negative acceleration)*

- (4) *Lateral acceleration*
- (5) *Longitudinal acceleration*
- (6) *Jerk*
- (7) *Vehicle speed*
- (8) *Vehicle RPM*

Questão Q₅: Quais os dispositivos usados nas coletas de dados de *Driving Behavior*?

Nesta questão o objetivo é descobrir quais os sensores são mais usados nos veículos para a coleta dos sinais e conseqüentemente, na classificação de *Driving Behavior*. Em todos os estudos dependendo do tipo, os sensores são instalados dentro ou fora dos veículos. É por meio dos sensores que os sinais são coletados e a partir destes que os padrões das manobras e eventos podem ser identificadas. Existem diversos tipos de sensores, por exemplo: IMU (*Inertial Measurement Unit*), GPS, Acelerômetro, Giroscópio dentre outros. O IMU ENGELBRECHT *et al.* (2015) coleta sinais como: velocidade, posição do acelerador, aceleração, etc. Além desse, outros sensores dos *smartphones* ARAÚJO *et al.* (2012), podem coletar sinais como: aceleração, mudança de direção, frenagem, etc. Também existem os sensores usados independentes, a exemplo dos: GPS, acelerômetro, magnetômetro, entre outros. Na tabela 2.13 é apresentado os sensores encontrados nas análises dos artigos.

Obs.: Importante frisar que os estudos que utilizam a interface *ODB-II* estão coletando os sinais da *CAN-Bus* do veículo, mesmo que não mencionem este fato explicitamente. O contrário já não é verdadeiro, pois já, existem outros tipos de dispositivos que também podem coletar informações da *CAN-Bus*.

Questão Q₆: Como são montados os conjuntos de dados nas análises de *Driving Behavior*?

Nesta questão, o estudo busca compreender a elaboração dos conjuntos de dados (*datasets*) derivados dos sinais capturados pelos sensores. Conforme ilustrado na figura 2.10, as análises de *Driving Behavior* geralmente ocorrem em um centro de dados central, onde os sinais dos veículos são recebidos e passam por tratamentos para identificação e classificação. A montagem de um conjunto de dados está intrinsecamente relacionada com os objetivos da pesquisa que o utiliza.

Originalmente, um conjunto de dados é composto por séries temporais brutas, apresentando o eixo do tempo e o eixo dos sinais capturados, também conhecido como características (*features*) do conjuntos de dados. As *features* podem ser sinais

como aceleração lateral, aceleração longitudinal, velocidade, entre outros. Esses conjuntos de dados são empregados tanto na identificação de eventos quanto na classificação dos mesmos. A maioria dos artigos, embora não mencionem explicitamente a identificação, utiliza conjuntos de dados para ambas as etapas. Na identificação, o conjunto frequentemente permanece na forma bruta (séries temporais), enquanto na classificação, ele pode permanecer bruto ou ser enriquecido por meio de transformações com métricas como média, desvio padrão, variância, entre outras. A qualidade da seleção de características (*feature set*) desempenha um papel crucial na classificação de *Driving Behavior*. Contudo, não há um padrão rígido sobre como extrair ou selecionar essas *features*, pois varia dependendo da abordagem do estudo de classificação WU *et al.* (2016). No entanto, existem práticas recomendadas, muitas delas provenientes das técnicas consagradas de Mineração de Dados.

Alguns estudos destacam a extração e a construção de conjuntos de características como uma preocupação central. Em ARAÚJO *et al.* (2012), por exemplo, a partir de sinais brutos, os autores criaram um *feature set* chamado conjunto F , contendo dados como velocidade média, velocidade mínima, velocidade máxima, aceleração mínima, aceleração máxima, tempo de parada do veículo e rotações por minuto (RPM). Em LIN *et al.* (2014), os autores mencionam dados presentes em seu *feature set*, tais como média, desvio padrão, variância e valor absoluto. Já em FUGIGLANDO *et al.* (2018), houve um esforço para detalhar mais o *feature set*, que foi extraído dos sinais do *CAN-bus* do veículo. Os dados do conjunto incluem valor absoluto do sinal (aceleração, velocidade etc.), diferença entre dois intervalos de tempo do sinal, valor máximo do sinal, média, mediana e desvio padrão. Em WU *et al.* (2016), os autores também detalham seu *feature set*, como evidenciado na figura 2.11 onde a tabela extraída de seu estudo, apresenta o conjunto de características usado.

Table 4. The features extracted from the valid data.

Feature	Description
max	Maximum value of the valid data
min	Minimum value of the valid data
avg	Mean value of the valid data
std	Standard deviation of the valid data
peak	The quantity of peak value
pvam	The rate of data which exceeds the mean value
mad	Mean absolute deviation of the valid data
eng	Energy of the valid data
cov	Covariance matrix of the valid data
k	Slope between the maximum and minimum
t	The duration time of the valid data

Figura 2.11: Tabela com o *features set* usado em WU *et al.* (2016)

Como observado, há diversas maneiras de montar um *dataset* com um *feature set*. Não existe um padrão fixo para construir esses conjuntos de características, e as métricas e dados utilizados variam significativamente. No entanto, algumas conclusões foram alcançadas. Por exemplo, os estudos aproveitaram práticas da Mineração de Dados para realizar a extração e a composição de seus conjuntos de características. Além disso, nos estudos que detalharam os conjuntos de características utilizados, puderam ser identificados alguns dados e métricas frequentemente empregados, tais como:

- Valor absoluto do sinal
- Mediana
- Média
- Valor mínimo do sinal
- Variância
- Desvio padrão

2.4 Consolidação do conhecimento

Após um ano de pesquisa, nossa revisão sistemática revelou fatos essenciais sobre as análises de *Driving Behavior* e o papel das técnicas de mineração de dados e aprendizado de máquina, em fortalecer essa disciplina para aprimorar a segurança rodoviária. Essa pesquisa não apenas atingiu nossos primeiros dois objetivos, conforme delineados na seção de objetivos (2.1.1), mas também os desdobrou de maneira aprofundada através das seis questões investigadas (2.3). A jornada metodológica, seguindo os passos de uma revisão sistemática, resultou na análise minuciosa de mais de 90 estudos selecionados. Estes estudos destacaram a complexidade e amplitude do campo do *Driving Behavior*, onde um grande conjunto de abordagens foi explorada. Descobrimos que o conceito de *Driving Behavior* engloba uma diversidade de comportamentos, desde manobras comuns até manobras críticas, e sua análise ganha um papel primordial na melhoria da segurança nas estradas. Além disso, a importância dos sinais coletados por meio de sensores nos veículos, incluindo acelerômetros, IMUs e sensores de *smartphones*, emergiu como um elemento central para a obtenção de dados necessários para a análise detalhada dos comportamentos dos motoristas.

Ao mergulhar nas respostas às questões, nos deparamos com uma rica diversidade de técnicas de Mineração de Dados empregadas para abordar a complexidade do *Driving Behavior*. Desde algoritmos de aprendizado de máquina até métodos

baseados em regras e comparações de séries temporais, todas foram exploradas para identificar e classificar os eventos de direção. Nossas análises permitiram esclarecer o cenário ao apresentar um modelo holístico do *Driving Behavior*, representado pela figura 2.10. Esse modelo destaca as etapas de identificação e classificação, que variam desde técnicas simples até abordagens avançadas, como Redes Neurais Artificiais e *Support vector machine* (SVM). Em adição, investigamos como os conjuntos de dados são construídos. Revelamos uma profusão de técnicas e abordagens na seleção e transformação das características (*features*) utilizadas nos conjuntos de dados.

Em síntese, a análise de *Driving Behavior* emerge como uma disciplina multidisciplinar, respaldada por um vasto conjunto de técnicas de Mineração de Dados. Ao coletar e processar os sinais dos sensores instalados nos veículos, essa análise se traduz em uma contribuição indispensável para a segurança rodoviária, permitindo a detecção de eventos de alto risco e viabilizando o desenvolvimento de sistemas de alerta e prevenção. Nossa pesquisa deixou clara a importância de utilizar técnicas de mineração de dados e/ou aprendizado de máquina na classificação de padrões críticos de direção. A complexidade deste processo fica evidente pela variedade de abordagens adotadas e pela importância atribuída à seleção de características nos conjuntos de dados. À medida que avanços interdisciplinares e tecnológicos continuam a moldar a área, a análise de *Driving Behavior* permanece como um agente-chave na promoção de estradas mais seguras e conscientes.

2.5 Oportunidade de pesquisa

Nossa revisão sistemática resultou na identificação de oportunidades para pesquisas futuras no campo da análise de *Driving Behavior*. Destacamos especialmente a importância da aplicação de técnicas de inteligência artificial (IA) para aprimorar a segurança viária, em especial no contexto do *Crash Prediction System* (CPS), aplicação conhecida no ambiente de *Driving Behavior*. Ao aperfeiçoarmos algoritmos de aprendizado de máquina capazes de classificar eventos críticos, elevamos o nível de segurança do modelo como um todo. Em consonância, emergem perspectivas promissoras ao explorar novas técnicas de aprendizado de máquina, em conjunto com sensores para identificar e classificar eventos de direção agressiva, preenchendo uma demanda crucial por aprimoramentos nesse campo.

Outra área repleta de oportunidades é a aplicação de técnicas de aprendizado de máquina em dados provenientes de experimentos reais de direção. A diversidade de tipos de dados coletados por diferentes estudos aponta para a necessidade de testes mais amplos com novas abordagens, especialmente empregando dados reais de coletas, a fim de economizar recursos e maximizar o reaproveitamento dos dados. Além disso, nossa pesquisa destaca a relevância da otimização dos conjuntos de

dados, que constituem a base dos algoritmos. A seleção, combinação e transformação das características (*features*) são elementos cruciais na construção de conjuntos de dados robustos, ressaltando a necessidade de novos métodos que explorem esses aspectos.

Dados brutos oriundos de veículos frequentemente assumem a forma de séries temporais, sujeitas a influências de tendências, sazonalidades e propriedades não estacionárias. Nesse contexto, a decomposição de séries temporais surge como uma solução promissora para abordar essas complexidades e pode ser empregada como entrada para técnicas de aprendizado de máquina. As oportunidades que emergem dessa análise apontam para uma jornada contínua na direção de análises de *Driving Behavior* mais eficazes, contribuindo para a promoção da segurança nas estradas e o aprimoramento das estratégias de prevenção de acidentes.

Resumindo, a partir de agora nossa pesquisa visa explorar novos algoritmos, aplicados a dados brutos, para aprimorar a acurácia na identificação e classificação de eventos críticos de direção. Nosso foco se concentrará na utilização de técnicas de aprendizado de máquina em conjunto com tratamentos mais sofisticados dos dados, visando ressaltar características específicas que facilitem o reconhecimento de padrões pelos algoritmos. Nossa abordagem também manterá premissas importantes como a facilidade, escalabilidade e compreensão das transformações de dados, com ênfase na velocidade da aplicação e eficiência de custos nos experimentos.

Tabela 2.12: Técnicas de processamento mais usadas

Processo	Técnica	Artigos
Processamento	<i>Artificial Neural Network (ANN)</i>	MESEGUER <i>et al.</i> (2013); AL-DOORI <i>et al.</i> (2014); CHONG <i>et al.</i> (2013); JÚNIOR <i>et al.</i> (2017); ALVAREZ <i>et al.</i> (2014); GUOYING (2016)
	<i>Bayesian Network</i>	JÚNIOR <i>et al.</i> (2017); CASTRO e KIM (2016); EREN <i>et al.</i> (2012)
	<i>C4.5 Decision Tree</i>	WU <i>et al.</i> (2016); CASTRO e KIM (2016); TAYLOR <i>et al.</i> (2016)
	<i>Cross Validation</i>	FUGIGLANDO <i>et al.</i> (2018); TAYLOR <i>et al.</i> (2016); CHONG <i>et al.</i> (2013); JÚNIOR <i>et al.</i> (2017); CASTRO e KIM (2016); ZHU <i>et al.</i> (2018); PAEFGEN <i>et al.</i> (2013); BAGNALL <i>et al.</i> (2017)
	<i>Dynamic Time warping (DTW)</i>	BANERJEE <i>et al.</i> (2014); ENGELBRECHT <i>et al.</i> (2015); SAIPRASERT <i>et al.</i> (2017); JOHNSON e TRIVEDI (2011); JÚNIOR <i>et al.</i> (2017); EREN <i>et al.</i> (2012); WAHLSTRÖM <i>et al.</i> (2017); XIONG <i>et al.</i> (2019); MEIRING e MYBURGH (2015); PHOLPRASIT <i>et al.</i> (2015); ISMAIL FAWAZ <i>et al.</i> (2019);
	<i>Fuzzy C-means</i>	MA e ANDREASSON (2007); HALIM <i>et al.</i> (2016b)
	<i>Fuzzy Logic</i>	CASTIGNANI <i>et al.</i> (2015); ARAÚJO <i>et al.</i> (2012); CHONG <i>et al.</i> (2013); AL-DOORI <i>et al.</i> (2014); LIN <i>et al.</i> (2014); MEIRING e MYBURGH (2015); MA e ANDREASSON (2007); CASTIGNANI <i>et al.</i> (2013); LU <i>et al.</i> (2009); CASTIGNANI <i>et al.</i> (2014)
	<i>Fuzzy Neural Network</i>	WAHAB <i>et al.</i> (2007); CHONG <i>et al.</i> (2013)
	<i>Gaussian Mixture Model (GMM)</i>	WAKITA <i>et al.</i> (2005); WAHAB <i>et al.</i> (2007); GUOYING (2016)
	<i>Hidden Markov Model (HMM)</i>	MEIRING e MYBURGH (2015); MITROVIC (2005); WANG <i>et al.</i> (2010); LÄNGKVIST <i>et al.</i> (2014)
	<i>Hierarchical Clustering</i>	KALSOOM e HALIM (2013)
	<i>K-means Clustering</i>	KALSOOM e HALIM (2013); GUOYING (2016); FUGIGLANDO <i>et al.</i> (2018); XIONG <i>et al.</i> (2019); HALIM <i>et al.</i> (2016b); GUTIERREZ-OSORIO e PEDRAZA (2020); KLUGER <i>et al.</i> (2016); YANG ZHENG <i>et al.</i> (2014)
	<i>K-NN Classification/Clustering</i>	BANERJEE <i>et al.</i> (2014); WU <i>et al.</i> (2016); SATHYANARAYANA <i>et al.</i> (2012); BAGNALL <i>et al.</i> (2017); ISMAIL FAWAZ <i>et al.</i> (2019); HU <i>et al.</i> (2013)
	<i>Maximum Likelihood (ML)</i>	ENGELBRECHT <i>et al.</i> (2015); JUN <i>et al.</i> (2007)
	<i>Naive Bayes</i>	WU <i>et al.</i> (2016); TAYLOR <i>et al.</i> (2016)
	<i>Random Forest</i>	JÚNIOR <i>et al.</i> (2017); EREN <i>et al.</i> (2012); TAYLOR <i>et al.</i> (2016); ZHU <i>et al.</i> (2018); WANG <i>et al.</i> (2017); LONGHI e NANNI (2019)
	<i>Reinforcement Learning</i>	CHONG <i>et al.</i> (2013)
<i>Support Vector Machine (SVM)</i>	QU <i>et al.</i> (2017); MEIRING e MYBURGH (2015); GUOYING (2016); SATHYANARAYANA <i>et al.</i> (2012); HALIM <i>et al.</i> (2016b); WANG <i>et al.</i> (2017); WU <i>et al.</i> (2016), JÚNIOR <i>et al.</i> (2017)	

Tabela 2.13: Sensores utilizados para análises de *Driving Behavior*

Sensores	Artigos
Acelerômetro	TOLEDO <i>et al.</i> (2008), BANERJEE <i>et al.</i> (2014), JOHNSON e TRIVEDI (2011) e MITROVIC (2005)
<i>Inertial Measurement Unit (IMU)</i>	CARMONA <i>et al.</i> (2015), VASCONCELOS <i>et al.</i> (2017), PAEFGEN <i>et al.</i> (2012) e ENGELBRECHT <i>et al.</i> (2015)
<i>Smartphones sensors</i>	BANERJEE <i>et al.</i> (2014), MESEGUER <i>et al.</i> (2013), CASTIGNANI <i>et al.</i> (2015), ENGELBRECHT <i>et al.</i> (2015), HAQUE <i>et al.</i> (2016), SAIPRASERT <i>et al.</i> (2017), ALVAREZ <i>et al.</i> (2014), CARMONA <i>et al.</i> (2015), EREN <i>et al.</i> (2012), VASCONCELOS <i>et al.</i> (2017), YLIZALITURRI-SALCEDO <i>et al.</i> (2015), JÚNIOR <i>et al.</i> (2017), FAZEEN <i>et al.</i> (2012), PAEFGEN <i>et al.</i> (2012) e ARAÚJO <i>et al.</i> (2012)
GPS	TOLEDO <i>et al.</i> (2008), MA e ANDREASSON (2007), BANERJEE <i>et al.</i> (2014), CASTIGNANI <i>et al.</i> (2015), ENGELBRECHT <i>et al.</i> (2015), JOHNSON e TRIVEDI (2011), CARMONA <i>et al.</i> (2015) e MITROVIC (2005)
<i>CAN-Bus</i> do veículo	CARMONA <i>et al.</i> (2015), TAYLOR <i>et al.</i> (2016), JOHNSON e TRIVEDI (2011), AL-DOORI <i>et al.</i> (2014) e ALVAREZ <i>et al.</i> (2014)
Interface <i>ODB-II</i> do veículo	ARAÚJO <i>et al.</i> (2012), MESEGUER <i>et al.</i> (2013), CARMONA <i>et al.</i> (2015), VASCONCELOS <i>et al.</i> (2017), YLIZALITURRI-SALCEDO <i>et al.</i> (2015), JÚNIOR <i>et al.</i> (2017) e PAEFGEN <i>et al.</i> (2012)

Capítulo 3

Driving Behavior conceitos e aplicações

Driving Behavior é a disciplina que investiga o comportamento de um motorista dentro de um veículo. O termo é usado para representar diferentes conceitos relacionados às ações de condução de um motorista como: acelerar, frear, virar à esquerda ou à direita, entre outras MEIRING e MYBURGH (2015). Segundo WAKITA *et al.* (2005) *Driving Behavior* é um processo cíclico onde o motorista reconhece o ambiente a sua volta, decide uma ação (acelerar ou frear) e a executa de forma prudente com segurança ou algumas vezes de forma imprudente e perigosa. Em ENGELBRECHT *et al.* (2015) *Driving Behavior* é representado como um conjunto de ações de direção que ao serem reconhecidas auxiliam na identificação do padrão de direção que está sendo executado, por exemplo: um motorista liga a seta, vira o veículo para a esquerda, segue em frente e ultrapassa o carro à frente.

Driving Behavior é um conceito complexo, pois, descreve como um motorista opera os controles de um veículo e como este interage com o meio em sua volta, isto torna a identificação e classificação deste ambiente essencial para o tema. Identificar e classificar *Driving Behavior* irá ajudar não apenas na prevenção de acidentes, mas também, no desenvolvimento de modelos de condução que possam auxiliar na execução de manobras com maior confiabilidade trazendo conforto e segurança aos motoristas MARTINEZ *et al.* (2018); FUGIGLANDO *et al.* (2018). Existem diversos conceitos, tipos de caracterização e aplicações relacionados à *Driving Behavior*. Neste capítulo será feito um resumo dos assuntos mais relevantes, o que irá facilitar o entendimento do leitor sobre nosso trabalho.

3.1 *Driving Behavior* e os termos relacionados

Driving Behavior é um conceito muito amplo influenciado por diversos fatores o que naturalmente dificulta sua descrição. Ao longo dos anos surgiram vários termos, muitas vezes usados de forma equivocada, o que dificultou ainda mais o seu uso. Na literatura científica são encontrados termos como: *Driving Behavior*, *Driving Style*, *Driving Event*, *Driving Pattern*, *Driving Maneuvers* entre outros. De acordo com MARTINEZ *et al.* (2018) é necessária uma descrição mais concisa dos termos para se evitar confusão aos leitores. Na figura 3.1 é apresentada a relação entre os termos mais usados.

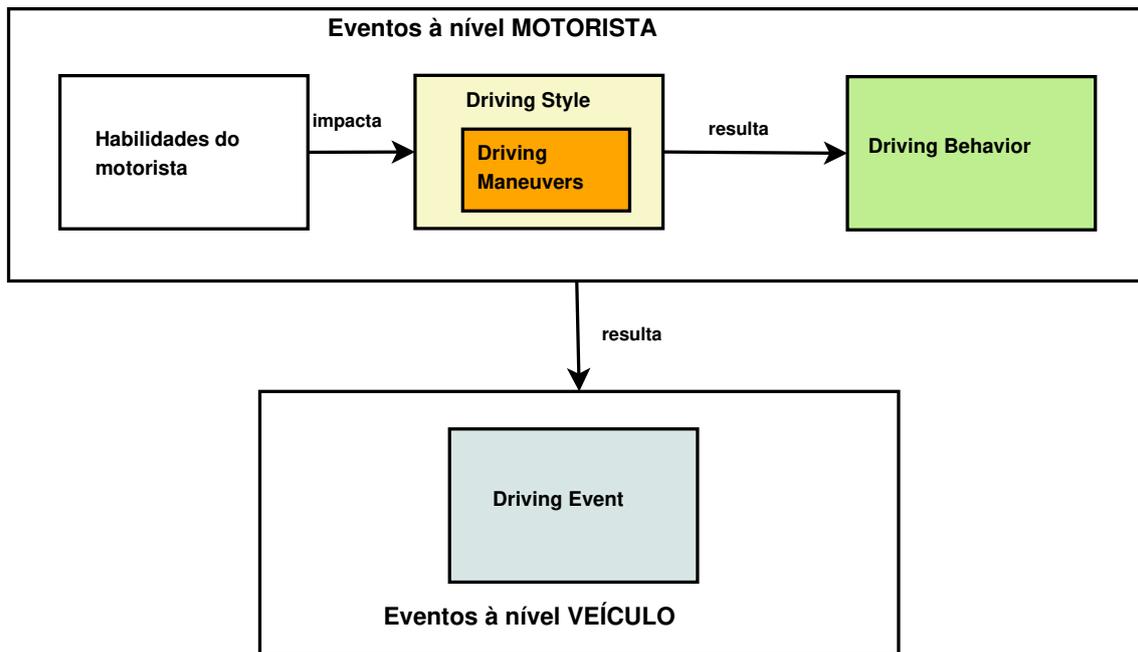


Figura 3.1: Termos relacionados com *Driving Behavior* baseado no trabalho de MARTINEZ *et al.* (2018)

3.1.1 *Driving Style*: Definição e Importância

O termo *Driving Style* refere-se à maneira como um motorista conduz um veículo, englobando aspectos como a pressão aplicada no acelerador, a técnica utilizada para virar o volante em curvas e a abordagem ao frear, entre outras ações MARTINEZ *et al.* (2018) MEIRING e MYBURGH (2015). Essas ações podem ser identificadas, quantificadas e categorizadas. A compreensão de *Driving Style* está intrinsecamente ligada aos conceitos de *Driving Maneuvers* e *Driving Event*, conforme apresentado no estudo de JOHNSON e TRIVEDI (2011). No contexto deste estudo, *Driving Style* é considerado como uma série de *Driving Maneuvers* interdependentes (ver Figura 3.1).

O *Driving Style* de um motorista possui impactos diretos não apenas no comportamento de direção (*Driving Behavior*), mas também na segurança do condutor, no desgaste mecânico do veículo e até mesmo no consumo de combustível ARAÚJO *et al.* (2012) MESEGUER *et al.* (2013). Diversos estudos evidenciam que os padrões de *Driving Style* praticados pelos motoristas têm o potencial de melhorar ou comprometer a segurança automotiva AL-DOORI *et al.* (2014) MARTINEZ *et al.* (2018).

Nesse contexto, a identificação do *Driving Style* torna-se crucial à medida que crescem as exigências por segurança, eficiência econômica e práticas ecologicamente sustentáveis. A literatura apresenta uma variedade de estudos que buscam identificar e classificar o *Driving Style*. Estes estudos classificam o estilo de direção de acordo com níveis de criticidade e apesar dos objetivos serem os mesmos, os níveis empregados diferem conforme o foco de cada estudo. Por exemplo, o trabalho de LIN *et al.* (2014) categoriza o *Driving Style* como “Cauteloso”, “Mediano”, “Especialista” e “Imprudente”. Já MEIRING e MYBURGH (2015) utilizam as categorias “Normal”, “Agressivo”, “Desatencioso” e “Bêbado”. Por sua vez, ENGELBRECHT *et al.* (2015) simplifica a classificação em “Seguro” e “Não seguro”.

Na maioria desses estudos, os algoritmos de identificação e classificação de *Driving Style* são frequentemente abordados de forma conjunta, alternando entre abordagens separadas e combinadas. No presente trabalho, optamos por abordar essas duas tarefas de maneira distinta. A identificação concentra-se em determinar o início e o fim de cada manobra (*Driving Maneuvers*). Em contrapartida, a classificação visa atribuir um grau de criticidade à manobra identificada, indicando se ela é "Segura" ou "Não segura".

3.1.2 *Driving Maneuvers*

Complementando e detalhando os desafios apresentados na identificação de *Driving Style*, os conceitos para *Driving Maneuvers* podem ser entendidos como: “A ação executada pelo motorista que resulta em uma manobra aplicada no veículo”. O termo é também empregado em várias situações que muitas vezes se confundem. Para este trabalho, *Driving Maneuvers* (manobras do motorista) são consideradas o menor pedaço dentro de todo o sistema de *Driving Behavior*. Este conceito é apoiado por estudos como GERDES (2006) que identifica *Driving Maneuvers* como 1 (uma) tarefa de direção que pode ser classificada e, como em HÜLNHAGEN *et al.* (2010) que considera *Driving Maneuvers* uma sequência de elementos básicos de direção. Inspirados nestes estudos, a tabela 3.1 apresenta os tipos *Driving Maneuvers* mais relevantes.

Com as definições de *Driving Maneuvers* apresentada, torna-se evidente a impor-

Tabela 3.1: *Driving Maneuvers* mais relevantes baseado nos estudos de GERDES (2006), HÜLNHAGEN *et al.* (2010) e SATHYANARAYANA *et al.* (2012)

Manobras mais comuns
Parar
Seguir em frente
Virar à direita
Virar à esquerda
Mudar de faixa para direita
Mudar de faixa para esquerda
Seguir carro a frente
Ultrapassar

tância em identificá-los, dado que são peças fundamentais na classificação de *Driving Style* e, por sua vez, na classificação de *Driving Behavior*. Para identificação de *Driving Maneuvers* são utilizados os dados, também chamados de sinais, provenientes de sensores instalados nos veículos. A partir desses sensores são extraídas as informações que ajudam a reconhecer uma manobra. Geralmente estes dados/sinais são agrupados em intervalos de tempo para então serem identificados. Diversos estudos com o em SAIPRASERT *et al.* (2017); FAZEEN *et al.* (2012); EREN *et al.* (2012), relacionam a identificação de *Driving Maneuvers* com os sinais originados de sensores como: Acelerômetro, Giroscópio, *Global Positioning System* (GPS), Magnetômetro entre outros. Esses estudos utilizam o sistema de coordenadas dos sensores (eixos x , y e z) para auxiliar na identificação dos *Driving Maneuvers*. Na figura 3.2 é apresentado como o sistema de coordenadas se posiciona em relação a um veículo.

Através dos eixos de coordenadas dos sensores, as manobras podem ser reconhecidas. Por exemplo, na figura 3.3 a partir do sinal da aceleração, eixo x do acelerômetro, é possível reconhecer uma manobra “Mudar de faixa para direita/esquerda”. Existem diversos sinais utilizados para reconhecer manobras, na tabela 3.2 são demonstrados alguns destes. A relação entre **manobra**, **sinal** e **sensor** é uma característica importante no reconhecimento das manobras e muito explorado na literatura científica. O resultado desta relação é: “Uma manobra executada por um motorista é aplicada ao veículo, que através de seus sensores captura seus sinais e armazena seus valores”. A figura 3.4 detalha um modelo conceitual usado neste trabalho para demonstrar a identificação de *Driving Maneuvers*.

No que tange a classificação de *Driving Maneuvers* vale o que foi apresentado em *Driving Style*, ou seja, os termos usados para classificar geralmente são: manobra “Normal” ou manobra “Agressiva”, podendo variar de acordo com o objetivo do estudo. Esses termos são muito usados em estudos que focam na classificação de manobras e na segurança das rodovias LIN *et al.* (2014)ENGELBRECHT *et al.* (2015)SAIPRASERT *et al.* (2017). Neste ponto deve-se mencionar que muitos estu-

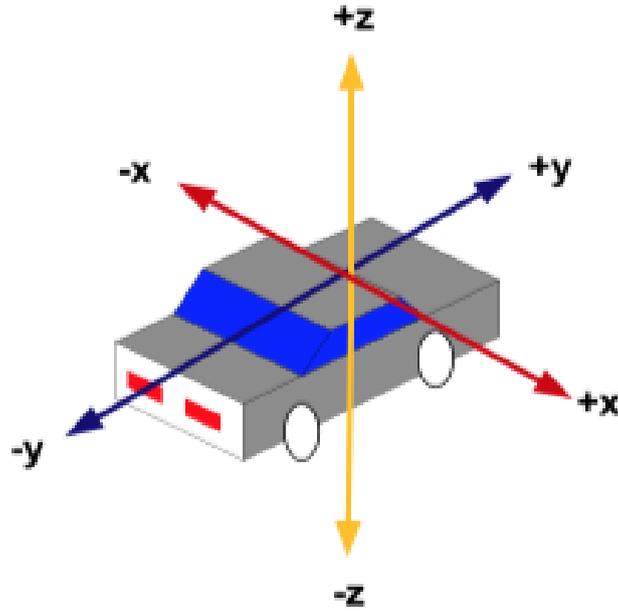


Figura 3.2: Sistema de coordenada do acelerômetro relativa ao veículo FAZEEN *et al.* (2012)

Tabela 3.2: Dados/sinais citados em diversos estudos FAZEEN *et al.* (2012), SAI-PRASERT *et al.* (2017) e FUGIGLANDO *et al.* (2018)

Sinais mais comuns
Aceleração longitudinal (eixo y do sensor acelerômetro)
Aceleração lateral (eixo x do sensor acelerômetro)
Revolução por Minuto (R.P.M)
Taxa de virada (eixo z do giroscópio)
Posição do veículo (latitude e longitude)
Velocidade

dos se preocupam em apenas identificar e classificar que tipo de manobras (tabela 3.1) se tratam, deixando a classificação da criticidade (“Normal” e “Agressivo”) para as análises de *Driving Style* ou *Driving Behavior*.

3.1.3 *Driving Behavior*

O termo *Driving Behavior* refere-se ao conjunto de estilos e manobras de direção executados por um motorista. Durante uma viagem, um motorista pode alternar entre diferentes estilos de direção, variando de ações moderadas a ações mais bruscas, dependendo do trecho da estrada ou das condições climáticas por exemplo. O objetivo do *Driving Behavior* é compreender o modo geral de condução ao longo de uma viagem ou em um intervalo de tempo predefinido. Esse comportamento pode ser descrito com termos como “Normal” ou “Agressivo”, semelhante ao adotado

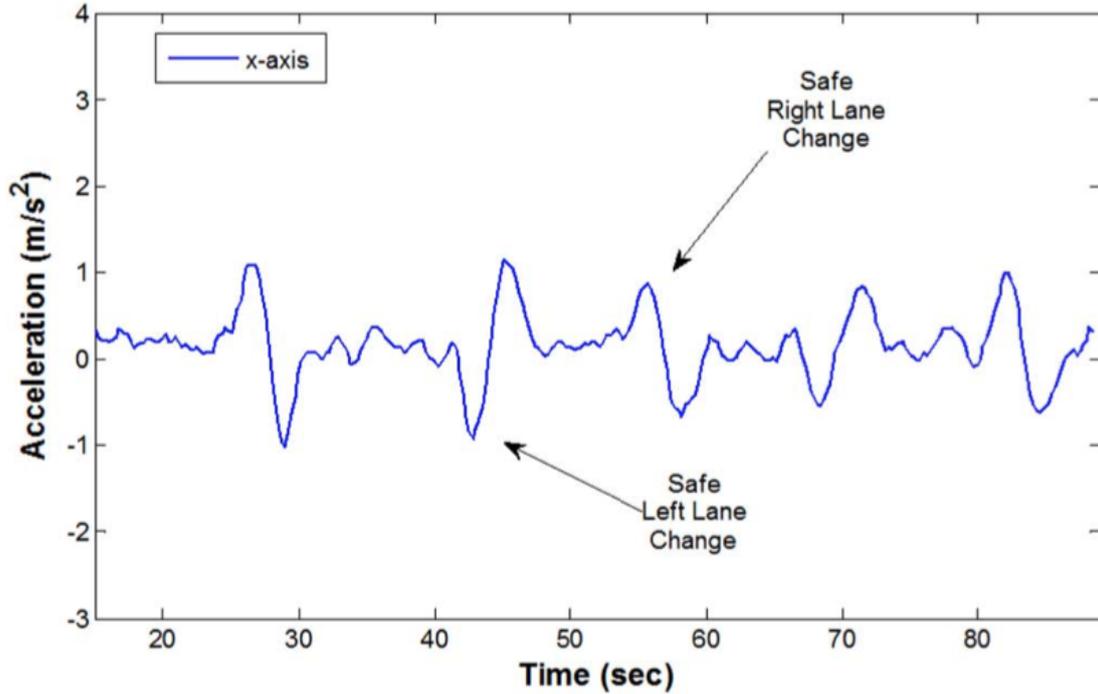


Figura 3.3: Sinal da aceleração do eixo x do acelerômetro relativo a manobra “Mudar de faixa para direita/esquerda”FAZEEN *et al.* (2012)

em *Driving Style* e *Driving Maneuvers*. A representação do *Driving Behavior* em relação aos estilos de direção executados é demonstrada na equação 3.1, baseada no trabalho de AL-DOORI *et al.* (2014):

$$DB = \{DS_{t=1}, DS_{t=2}, DS_{t=3}, \dots, DS_{t=n}\} \quad (3.1)$$

Onde DB representa o ambiente de direção, o DS o estilo de direção operados em intervalos (t) de tempo.

A figura 3.1 e a equação 3.1 ilustram a abrangência de termos e conceito do *Driving Behavior*. É importante notar que este conceito pode se tornar ainda mais complexo pois, além dos estilos de direção do motorista, fatores externos como as condições da estrada e do clima também podem influenciar CASTIGNANI *et al.* (2015); BANERJEE *et al.* (2014). A identificação e classificação de *Driving Behavior* são desafios relevantes, amplamente explorados na literatura científica. As classificações do *Driving Behavior* têm foco na melhoria da segurança rodoviária, conforme citados em WU *et al.* (2016); JÚNIOR *et al.* (2017); EREN *et al.* (2012); VAIANA *et al.* (2014), na eficiência de combustível, observado em MESEGUER *et al.* (2013); CASTIGNANI *et al.* (2015); AL-DOORI *et al.* (2014); ARAÚJO *et al.* (2012), na gestão de frotas logísticas, de acordo com CASTIGNANI *et al.* (2015); ENGELBRECHT *et al.* (2015); MEIRING e MYBURGH (2015) e, por fim, no mercado de seguros automotivos, demonstrados em CASTIGNANI *et al.* (2015);

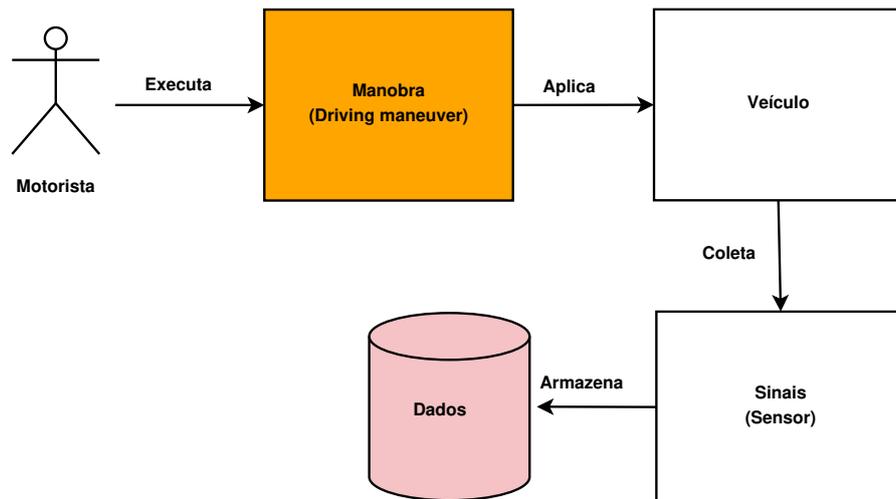


Figura 3.4: Modelo conceitual para identificação de *Driving Maneuvers*

JÚNIOR *et al.* (2017); ENGELBRECHT *et al.* (2015); MEIRING e MYBURGH (2015); TOLEDO *et al.* (2008). Na seção 2.3.1, apresentamos os detalhes da identificação e classificação do *Driving Behavior*. Em resumo, o *Driving Behavior* analisa como um todo o modo de condução de um motorista, abrangendo os estilos de direção mais frequentes durante uma viagem e as manobras mais comuns executadas por ele.

3.1.4 *Driving Event*

Driving Event é outro componente essencial nos estudos de Comportamento de Direção (*Driving Behavior*), sendo este o resultado das ações executadas pelo motorista enquanto conduz o veículo MITROVIC (2005). Os elementos de *Driving Event* assemelham-se aos de *Driving Maneuvers* (descritos na seção 3.1.2) e a principal diferença é que os eventos estão relacionados ao veículo, enquanto as manobras estão ligadas ao motorista MARTINEZ *et al.* (2018); JOHNSON e TRIVEDI (2011). O termo refere-se às mudanças no estado do veículo, como velocidade, aceleração, viradas para a esquerda ou direita, frenagem, entre outras JÚNIOR *et al.* (2017); MITROVIC (2005).

A identificação e classificação de *Driving Event* são cruciais para a segurança nas estradas JOHNSON e TRIVEDI (2011); MITROVIC (2005), permitindo alertar o motorista sobre comportamentos inadequados e, assim, reduzir o risco de acidentes SAIPRASERT *et al.* (2017); JÚNIOR *et al.* (2017). A melhoria da segurança rodoviária também utiliza a detecção de *Driving Event*, com o objetivo de identificar e classificar os eventos como “Seguros” ou “Perigosos” WAHLSTRÖM *et al.* (2017); MITROVIC (2005). A maioria dos sistemas focados em aumentar a segurança, melhorar a eficiência de combustível e otimizar o gerenciamento de frotas, se

concentram em identificar e classificar eventos “Perigosos”, “Agressivos” ou “Críticos”.

A detecção de eventos agressivos é realizada pela interpretação dos dados (sinais) coletados dos veículos, de maneira semelhante ao processo usado em *Driving Maneuvers*. Como mencionado, os eventos estão relacionados ao veículo, ou seja, uma manobra executada pelo motorista leva a uma mudança de estado do veículo, gerando um evento. Essa relação entre *Driving Event* e *Driving Maneuvers* é discutida em alguns estudos como ENGELBRECHT *et al.* (2015); JOHNSON e TRIVEDI (2011), embora nem sempre fique clara a diferença entre eles, e às vezes sejam usados de forma análoga. Neste trabalho, a distinção entre os conceitos é importante para uma compreensão precisa do que está sendo desenvolvido. A figura 3.5 apresenta um modelo conceitual que relaciona os principais conceitos ora utilizados. A figura destaca os principais elementos nos quais os algoritmos de detecção podem ser aplicados.

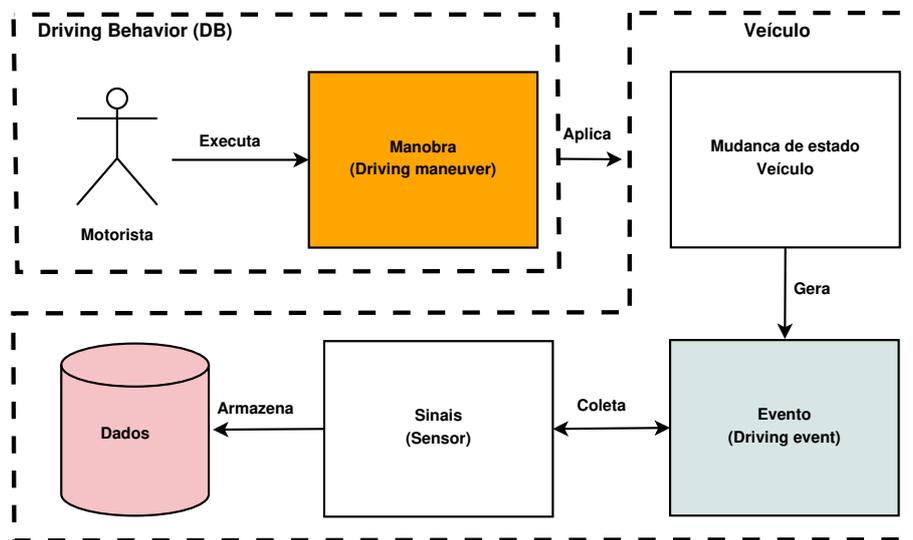


Figura 3.5: Modelo de relações de *Driving Behavior* e *Driving Event*

Quanto aos dados (sinais) gerados pelas manobras dos motoristas, eles são os elementos-chave para a detecção de eventos. A literatura conta com diversos estudos que utilizam a combinação de sensores e sinais, todos com foco na identificação de eventos agressivos. Por exemplo, em YLIZALITURRI-SALCEDO *et al.* (2015), os autores usam os sinais do acelerômetro para identificar e classificar eventos agressivos. Em MARTINEZ *et al.* (2018), os autores enfatizam a seleção de sinais como fundamental para a detecção de eventos, citando “frenagem”, “aceleração” e “mudanças de faixa”, como sinais promissores para classificação de eventos agressivos. Em JÚNIOR *et al.* (2017), os autores testam vários algoritmos de Aprendizado de Máquina para reconhecimento de eventos e sua subsequente classificação como agressivos ou não. A tabela 3.3 apresenta alguns eventos agressivos amplamente explorados na literatura.

Tabela 3.3: Eventos críticos

Eventos críticos mais comuns
Aceleração agressiva
Frenagem agressiva
Virada para esquerda agressiva
Virada para direita agressiva
Mudança de faixa para esquerda agressiva
Mudança de faixa para direita agressiva

Três abordagens predominantes são usadas para a detecção de eventos críticos, as duas primeiras estão relacionadas a tarefa de identificação e a última à tarefa de classificação. A primeira abordagem baseia-se em limites e/ou regras aplicados aos dados (sinais). A segunda compara dados, geralmente séries temporais, em um intervalo de tempo, onde uma série A (um evento de direção) é comparada com outra série B (evento modelo que representa um evento crítico) WAHLSTRÖM *et al.* (2017). A terceira abordagem classifica o sinal utilizando como um modelo pré-treinado de inteligência artificial, que categoriza os dados MESEGUER *et al.* (2013); AL-DOORI *et al.* (2014); QU *et al.* (2017); CHONG *et al.* (2013); MEIRING e MYBURGH (2015); GUOYING (2016). A primeira abordagem é apresentada na figura 3.6 e ilustra um algoritmo que identifica um evento. A segunda é exemplificada pela figura 3.7, que utiliza o *Dynamic Time Warping* (DTW) para comparar duas séries temporais e assim identificar o evento. Por último, a figura 3.8 apresenta a terceira abordagem na classificação de um evento, usando um modelo de Aprendizado de Máquina.

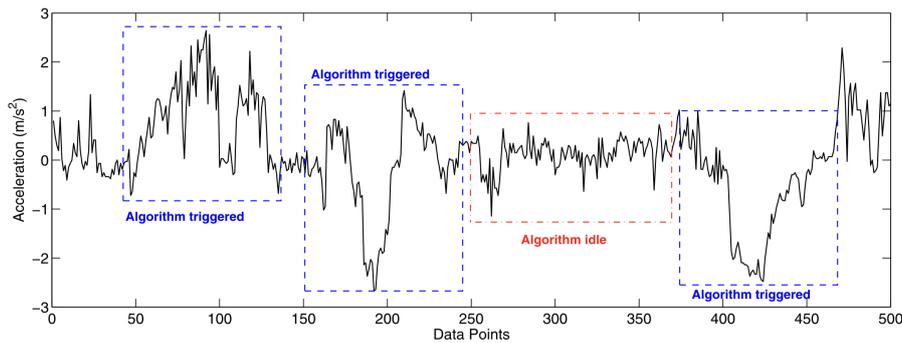


Figura 3.6: Algoritmo de identificação de evento de acordo com o trabalho de SAI-PRASERT *et al.* (2017)

De acordo com as análises anteriores sobre *Driving Behavior* e suas relações com os conceitos de *Driving Style*, *Driving Maneuvers* e *Driving Event*, é possível observar muitas semelhanças nas técnicas e abordagens de identificação e classificação. Embora as diferenças sejam sutis, é importante apresentá-las e discuti-las. É claro que é possível analisar o *Driving Behavior* apenas com o *Driving Style* ou apenas

com *Driving Event*, ou até mesmo considerar tudo como *Driving Behavior*. Isso não está incorreto, mas compreender os limites de cada conceito é fundamental para garantir que a pesquisa alcance claramente os resultados definidos em seus objetivos. *Driving Behavior* trata da totalidade relacionada ao motorista, enquanto *Driving Event* diz respeito ao veículo como um todo. No entanto, isso não significa que ao classificar apenas *Driving Event* não estejamos analisando *Driving Behavior*, já que esses conceitos estão interligados por uma sequência de agentes, ações e resultados, conforme ilustrado nas figuras 3.1 e 3.5.

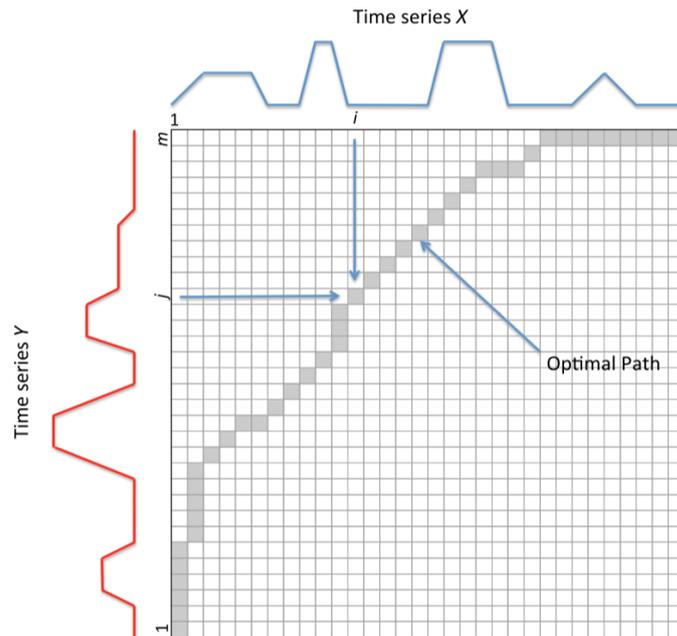


Figura 3.7: Uso do DTW na comparação de sinais para detecção de eventos críticos baseado no trabalho de SAIPRASERT *et al.* (2017)

3.2 Identificação e classificação de *Driving Event* (DE)

Nesta seção, consolidaremos o conhecimento adquirido nas análises de *Driving Behavior* (DB) e seus conceitos em relação às tarefas de identificação e classificação. A partir de agora, nossa pesquisa se concentrará exclusivamente nos conceitos que utilizaremos para testar nossos algoritmos. É relevante destacar que os conceitos de *Driving Style* serão incorporados no termo DB, assim como os conceitos de *Driving Maneuvers*, embora continuemos a nos referir a este último como “manobra”. O cerne do nosso trabalho permanece nas análises de DB, porém com uma abordagem mais específica. Nosso foco agora está na “**identificação e classificação de *Driving Event* por meio da aplicação de Aprendizado de Máquina a sinais**

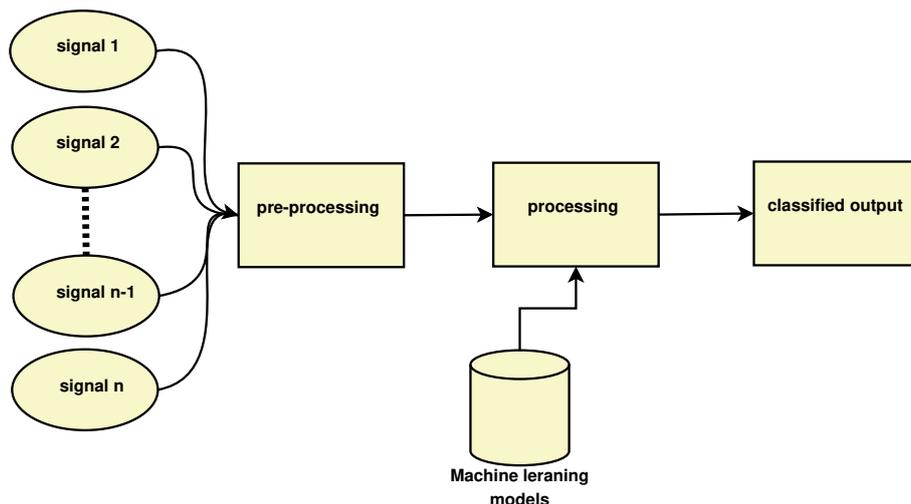


Figura 3.8: Classificação de evento com modelo de aprendizado de máquina baseado no trabalho de YADAV *et al.* (2019)

provenientes de sensores instalados em veículos, com o objetivo principal de detectar eventos críticos de direção (SCE)”. O modelo geral em que nossas abordagens serão aplicadas é ilustrado na figura 3.9.

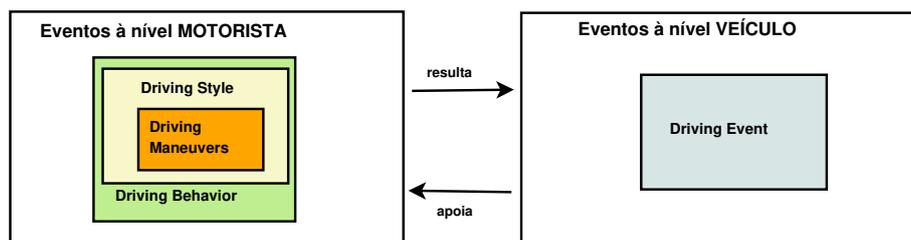


Figura 3.9: Modelo geral utilizado nas análises de DB e DE para este trabalho

A literatura sobre identificação e classificação de DE é extensa e abrange uma variedade de sensores e algoritmos WAHLSTRÖM *et al.* (2017). É fundamental reforçar que a identificação ocorre por meio de sensores instalados nos veículos que geram dados (sinais) que, quando organizados e interpretados, resultam em sua classificação. Por exemplo, um motorista em um veículo realiza várias manobras, que impactam no estado do veículo, gerando eventos de direção, como aceleração agressiva acima de um determinado limite. Os sensores, como acelerômetros, coletam os sinais de aceleração do veículo, enviando-os a uma central de dados, conforme discutido na seção 2.3.1 e representado na figura 2.10. Os dados recebidos pela central são identificados como eventos de direção e, em seguida, classificados. Ao final, é possível constatar todos os eventos de direção executados pelo motorista, permitindo uma análise detalhada do seu estilo de direção.

Para a tarefa de identificação, nosso trabalho empregará uma abordagem baseada em regras (figura 3.6). Ou seja, os dados brutos (séries temporais), após

pré-processamento, serão submetidos a um algoritmo de identificação que selecionará apenas as janelas de tempo que contenham eventos. Isso permitirá que nossos algoritmos de Aprendizado de Máquina avaliem se o evento é crítico ou não. Posteriormente, os eventos identificados serão confrontados com modelos de inteligência artificial para a sua classificação. Em nosso trabalho, os eventos críticos que desejamos classificar são **acidentes**, **quase-acidentes** ou **manobras abruptas**. A figura 3.10 apresenta o modelo utilizado neste trabalho para a identificação e classificação de eventos críticos. Nas próximas seções, detalharemos mais profundamente nossa abordagem.

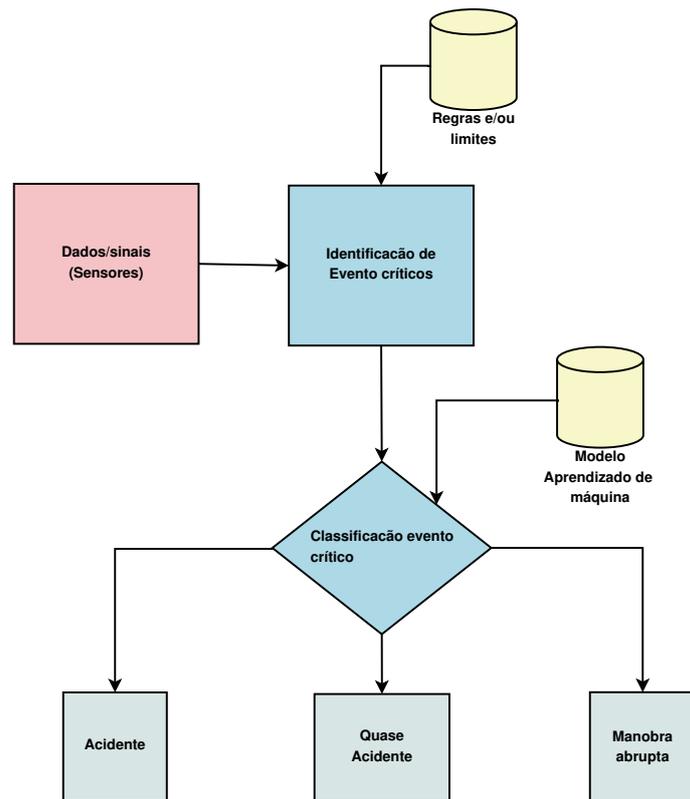


Figura 3.10: Modelo para identificação e classificação de evento crítico

3.3 Aplicações para *Driving Behavior* e *Driving Event*

Atualmente, uma variedade de aplicações surge a partir das análises de *Driving Behavior*, as quais são agrupadas com base em seus propósitos. É fundamental destacar que, com frequência, essas aplicações abordam a classificação de *Driving Style* e *Driving Event*. No entanto, conforme apresentado pela equação 3.1, os conceitos de *Driving Style* estão incorporados ao *Driving Behavior*. Para nosso trabalho, o foco recai na análise do ponto de vista do veículo, ou seja, em *Driving*

Event. Isso não descarta os estudos que exploram aplicações a partir das análises de *Driving Style* ou unicamente *Driving Behavior*, uma vez que como já apresentado nas figuras 3.9 e 3.5, existe uma relação próxima entre os conceitos, sendo frequentemente usados de forma intercambiável na literatura. A seguir, destacamos as aplicações mais mencionadas.

3.3.1 *Driver Assistance System*

Driver Assistance System (DAS) ganhou força nos últimos anos com o avanço da tecnologia nos veículos DÖRR *et al.* (2014). São diversas as razões para aumento na visibilidade desses sistemas, mas principalmente no que tange segurança das rodovias e dos motoristas JOHNSON e TRIVEDI (2011). DAS pode contribuir para aumentar a atenção do motorista, mitigando situações perigosas com fornecimento de informações sobre o tráfego e sobre o seu entorno. O DAS é desenvolvido especificamente para aprimorar, automatizar e adaptar os sistemas do veículo para reforçar a segurança, enriquecer a experiência de condução e o conforto da viagem MEIRING e MYBURGH (2015).

DAS foi introduzido na indústria automobilística com objetivo de aumentar a segurança e conforto do motorista. Em busca deste objetivo a indústria implantou diversas inovações que hoje estão presentes em veículos mais modernos, tais como: sistema automático de luzes, controles de cruzeiro adaptativo, sistema de frenagem inteligente, assistente de navegação, detecção de tráfego com recálculo de rotas, sistemas de alertas para o motorista, sistema de notificação para condições meteorológicas, visão traseira, entre outros MEIRING e MYBURGH (2015).

DAS pode se adaptar de acordo com o comportamento de direção (*Driving Behavior*), em tempo real, provendo informações e alertas ao motorista baseado em seu estilo de direção BERGASA *et al.* (2014), por exemplo, se o motorista tem um comportamento agressivo na direção, o DAS pode emitir alertas avisando sobre possíveis riscos como aceleração excessiva, sugerindo correção daquele comportamento FAZEEN *et al.* (2012).

3.3.2 *Eco Driving*

O conceito de *Eco Driving* é utilizado para representar comportamento de direção que resulte no uso eficiente de combustível. Em tempos onde a preocupação com meio ambiente se tornou fundamental, o uso consciente de combustível é uma meta a ser perseguida, principalmente devido às emissões de poluentes geradas. Estudos apontam que a diferença entre um comportamento de direção (*Driving Behavior*) normal e agressiva, pode resultar em mais 40% de emissão de gases (como CO_2) na atmosfera a favor da direção agressiva VAIANA *et al.* (2014).

O meio urbano tem sido grande responsável pela emissão de gases poluentes na atmosfera, além do uso excessivo de combustíveis fósseis, conforme verificado por MEIRING e MYBURGH (2015). Nas grandes cidade tem sido investigado sinais como: aceleração, velocidade, mudança de marcha e frenagem, que podem influenciar no uso ineficiente de combustível e conseqüentemente no aumento de emissão de poluentes. Muitos estudos apontam para que se alcance uma direção *Eco-friendly*, sendo necessária a diminuição de acelerações e frenagens bruscas WAHLSTRÖM *et al.* (2017).

3.3.3 *Road Condition Monitoring*

O monitoramento das vias e ruas, urbanas ou não, é uma aplicação importante, pois, a partir deste monitoramento é possível aumentar a qualidade da direção, visto que uma via em condições ruins, pode acarretar no aumento do custo de manutenção e do uso de combustível, além de reduzir o conforto da viagem e aumentar o risco de acidentes WAHLSTRÖM *et al.* (2017) MEIRING e MYBURGH (2015).

O monitoramento das condições das vias é um aspecto importante na segurança. O sistema pode manter motoristas informados sobre rotas, condições de tráfego, etc. O reconhecimento destas condições podem ser alcançadas através do uso de acelerômetros e GPS. Estudos como FAZEEN *et al.* (2012) usam os sinais do acelerômetro para detectar anomalias nas vias como: buracos e asfaltos irregulares. Uma rodovia em condições precárias, além de prejudicial à segurança dos motoristas, pode contribuir para os congestionamentos SAIPRASERT *et al.* (2017).

3.3.4 *Fleet Management*

Os sistemas de monitoramento de frotas logísticas cresceram muito nos últimos anos, a partir da inclusão de uma gama de pequenas aplicações como: controle de manutenção veicular, rastreamento, controle de velocidade, economia de combustível, entre outros JOHNSON e TRIVEDI (2011). O gerenciamento das frotas logísticas auxiliam as empresas minimizarem riscos que veículos e motoristas estão sujeitos, ampliando a eficiência e reduzindo custos. Alguns estudos na área de *Fleet Management* são de grande contribuição, a exemplo do sistema MIROAD (*Mobile-Sensor-Platform for Intelligent Recognition Of Agressive Driving*), sugerido por JOHNSON e TRIVEDI (2011), para o monitoramento do comportamento de direção da frota, a partir de alertas sonoros para os motoristas em uma direção agressiva.

As aplicações de *Fleet Management* buscam, principalmente, a segurança da frota. Muitos desses sistemas oferecem o rastreamento dos veículos e o monitoramento da velocidade. O controle de velocidade é essencial. É a maneira que a empresa tem de gerenciar o tipo de dirigibilidade de seus motoristas, para mitigar

riscos de acidentes. Os motoristas desses sistemas de *Fleet Management* são responsáveis por seus veículos e peças fundamentais do sistema. Seu acompanhamento é uma condição “*sine qua non*” para as empresas que oferecem esse tipo de serviços. O monitoramento do motorista, muitas vezes, passa por algoritmos de pontuação como em CASTIGNANI *et al.* (2015) que propôs um *software* chamado de *SenseFleet*. A partir dos sinais do acelerômetro é detectado os eventos agressivos e utilizado um sistema de pontuação para refletir um fator de risco para cada motorista da frota.

3.3.5 *Insurance Applications*

A indústria de seguros, especialmente o setor de seguros automotivos, geralmente calcula seus prêmios com base em dados estatísticos que possam influenciar o custo dos sinistros. Esses dados incluem, por exemplo, características do veículo (tipo, valor e etc), e o perfil do motorista (idade, sexo, estado civil, experiência, dentre outros). Esses dados nem sempre representam a realidade de cada motorista e seu risco. Entretanto, a análise do comportamento de condução pode ajudar a fornecer uma representação mais precisa de um indivíduo. Estudos como de CASTIGNANI *et al.* (2015), TOLEDO *et al.* (2008) e PAEFGEN *et al.* (2012) propõem algum tipo de sistema individual de pontuação mediante a detecção de eventos agressivos. Os resultados indicaram que os algoritmos detectaram essas ocorrências com precisão. Esse tipo de sistema pode ser usado em *Insurance Applications* com intuito de beneficiar bons motoristas, alertando-os sobre sua direção, diminuindo o risco de acidentes e, ainda, os beneficiando com premiações e, portanto, reduzindo o valor do seu seguro.

3.3.6 *Crash Prediction Systems*

Nos últimos anos, os Sistemas de Predição de Acidentes (*Crash Prediction Systems*) emergiram como uma tecnologia inovadora e essencial para a segurança dos motoristas. Esses sistemas têm como objetivo emitir alertas, identificar colisões iminentes ou até mesmo prever acidentes. Entre as aplicações voltadas à detecção de colisões, estão sistemas que realizam chamadas para serviços de emergência. Ou seja, ao detectar sinais abruptos (como um acidente), o sistema do veículo dispara alertas para serviços de emergência, fornecendo informações sobre o acidente, sua gravidade e localização ZALDIVAR *et al.* (2011). Essa abordagem é notavelmente útil, pois os serviços de emergência podem receber informações detalhadas para agir de forma eficiente.

Também merecem destaque as aplicações voltadas à prevenção de acidentes, que buscam antecipar possíveis colisões antes mesmo de ocorrerem. Esse tipo de aplicação geralmente atribui uma pontuação ao motorista, avaliando o risco associado à

sua forma de condução. Isso permite acompanhar a evolução da habilidade do motorista e emitir alertas sobre possíveis riscos. Um exemplo é o estudo de BAGDADI (2013), no qual o autor propõe um sistema de detecção de possíveis acidentes com base no comportamento de direção do motorista e nos eventos resultantes. O autor utiliza a avaliação do *jerk*¹, medida que permite identificar eventos de frenagem crítica e reconhecer acidentes devido à alta correlação entre eles.

Os sistemas de predição de acidentes (CPS) são o foco central do nosso trabalho. Nas próximas seções, detalharemos o funcionamento desses sistemas, abordaremos os estudos mais recentes sobre o tema e demonstraremos como nossa proposta contribui para a melhoria desses sistemas. Nossa pesquisa irá discutir e apresentar um modelo para a identificação de eventos críticos de direção e sua classificação quanto à gravidade. Em nosso caso, buscamos identificar acidentes, quase-acidentes e manobras abruptas a partir de grandes volumes de dados. Nosso modelo desempenha um papel fundamental na estrutura de um CPS e, por conseguinte, acreditamos que nossa proposta esteja alinhada ao objetivo principal desses sistemas: oferecer maior segurança aos motoristas e às rodovias em geral.

¹Em física, *jerk* é a taxa de variação da aceleração de um objeto.

Capítulo 4

Sistemas de predição de acidentes (CPS)

Esta seção está alinhada às descobertas da revisão sistemática da literatura apresentada no capítulo 2. Conforme discutido por Alvesson e Sandberg em ALVESSON e SANDBERG (2011), os pesquisadores identificam as novas oportunidades por meio de revisões e exploração de questões em aberto, as quais desempenham um papel crucial na identificação de novas descobertas. Nossa revisão mapeou oportunidades de pesquisa no campo da análise de comportamento de direção, com foco especial na aplicação em sistema de predição de acidentes (CPS), com a utilização de técnicas de aprendizado de máquina para aprimorar a segurança viária. A exploração de técnicas de aprendizado de máquina em conjunto com sensores emergem como iniciativas promissoras, para identificar eventos de direção e classificá-los quanto ao grau de criticidade. Além disso, identificamos lacunas na otimização dos conjuntos de dados, enfatizando a importância dos métodos de seleção e transformação. Também observamos oportunidades na aplicação de abordagens baseadas na decomposição de séries temporais, estas com potencial de aprimorar o reconhecimento de padrões pelos algoritmos de aprendizado de máquina. Neste capítulo apresentaremos os estudos relacionados a CPS e sua utilização, em conjunto com técnicas de inteligência artificial.

A predição de acidentes é uma área com vasto potencial de pesquisa, particularmente com a evolução das técnicas de *Machine Learning*. Esse cenário abriu novas perspectivas e desafios para o campo. Em estudos recentes, como mencionados por BAGDADI (2013); ELLISON e DANIELS (2012), diversos métodos têm sido propostos para classificar possíveis acidentes com base no *Driving Behavior* do motorista. Em estudos adicionais, como os realizados por HALIM *et al.* (2016b); WANG *et al.* (2017), os pesquisadores apresentaram suas abordagens para a identificação do *Driving Behavior*, acrescentando a utilização de técnicas de aprendizado de máquina. Essa abordagem é importante para o ambiente de direção, bem como

a aplicação de técnicas de aprendizado de máquina, que emergem como estratégias promissoras nos Sistemas de Predição de Acidentes (CPS).

Com a crescente quantidade de dados gerados por sensores nos veículos, as técnicas de aprendizado de máquina assumem um papel importante nos *Crash Prediction Systems* (CPS), pois a acurácia de seus modelos é essencial para um desempenho eficiente desses sistemas. Vale ressaltar que o objetivo central dos CPS é avaliar a probabilidade de ocorrência de um acidente antes mesmo que ele se materialize. Isso enfatiza a necessidade de abordagens preditivas sólidas e eficazes, que contribuem para a prevenção de acidentes em ambientes de direção. Vários estudos de análise de segurança rodoviária empregaram uma variedade de técnicas para compreender padrões de colisões e identificar riscos potenciais em cenários de direção, conforme destacado por XIONG *et al.* (2019); KLUGER *et al.* (2016); WANG *et al.* (2017); NAKAOKA *et al.* (2008).

Os CPS têm a capacidade de antecipar acidentes iminentes, o que os torna sistemas essenciais e amplamente explorados no âmbito da segurança viária, conforme evidenciado por WANG *et al.* (2010); XIONG *et al.* (2019); BA *et al.* (2017); ELAMRANI ABOU ELASSAD *et al.* (2020a); QU *et al.* (2017). Esses sistemas podem auxiliar os motoristas em situações de risco, emitindo alertas ou mesmo tomando decisões autônomas que visam evitar possíveis perigos, como observado por BA *et al.* (2017); ELAMRANI ABOU ELASSAD *et al.* (2020a); QU *et al.* (2017). No entanto, antes de prever um potencial acidente, é imperativo que o sistema seja capaz de identificar e classificar padrões críticos de segurança. Adicionalmente, esses sistemas também precisam equilibrar precisão, tempo de previsão e custo computacional na detecção de eventos críticos, conforme pontuado por BA *et al.* (2017).

A seguir, serão apresentadas as abordagens mais relevantes na literatura, juntamente com os conjuntos de dados empregados e as técnicas mais comumente utilizadas.

4.1 Abordagens para CPS

Dentro da literatura científica, diversos trabalhos se dedicam à criação de *Crash Prediction Systems* (CPS), explorando variadas abordagens, conjuntos de dados e técnicas para aprimorar a segurança no ambiente de direção. Tal amplitude de pesquisas pode ser notada em estudos como XIONG *et al.* (2019); HALIM *et al.* (2016a); BA *et al.* (2017); ELAMRANI ABOU ELASSAD *et al.* (2020a); QU *et al.* (2017); CASTRO e KIM (2016); KLUGER *et al.* (2016); ZHU *et al.* (2018); LONGHI e NANNI (2019); AKIN e AKBA (2010); AIT-MLOUK e AGOUTI (2019); REN *et al.* (2018).

Dentro do campo dos CPS, podem ser identificadas três principais tarefas: a

identificação de eventos de direção, a **classificação** de eventos críticos e a **previsão** de acidentes em potencial. Essas tarefas, por vezes, podem ser apresentadas isoladamente ou interconectadas. Por exemplo, em abordagens como CASTRO e KIM (2016); ZHU *et al.* (2018); PAEFGEN *et al.* (2013); AKIN e AKBA (2010), modelos de aprendizado de máquina são aplicados para identificar padrões indicativos de possíveis causas de acidentes, o que pode indiretamente contribuir para a classificação e previsão de futuros incidentes dentro dos CPS.

Outros trabalhos focam especificamente na classificação e identificação. KLUGER *et al.* (2016), por exemplo, empregou a transformação de *Fourier* combinado com clusterização para detectar eventos críticos de segurança, a partir das acelerações longitudinais dos veículos, caracterizando tais eventos como colisões, quase colisões e manobras abruptas. WANG *et al.* (2017) utilizaram variações de *jerk* para classificar padrões e, assim, identificar acidentes. Além disso, XIONG *et al.* (2019); NAKAOKA *et al.* (2008) buscaram identificar padrões de direção insegura para, em seguida, classificá-los em níveis de criticidade através de um índice de risco.

Outros trabalhos enfatizaram a previsão de acidentes, construindo modelos com base em dados históricos para prever incidentes em janelas de tempo distintas. LONGHI e NANNI (2019) classificou motoristas e previu a ocorrência de acidentes futuro usando um modelo de aprendizado de máquina, enquanto QU *et al.* (2017) empregou dados de radares para prever possíveis choques traseiros em diferentes horários. REN *et al.* (2018) desenvolveu um modelo de aprendizado profundo para prever riscos de acidentes com base em correlações espaciais e temporais entre localizações. Na figura 4.1 é apresentada a arquitetura do artigo. AIT-MLOUK e AGOUTI (2019) explorou regras de associação para prever acidentes em uma janela de tempo predefinida.

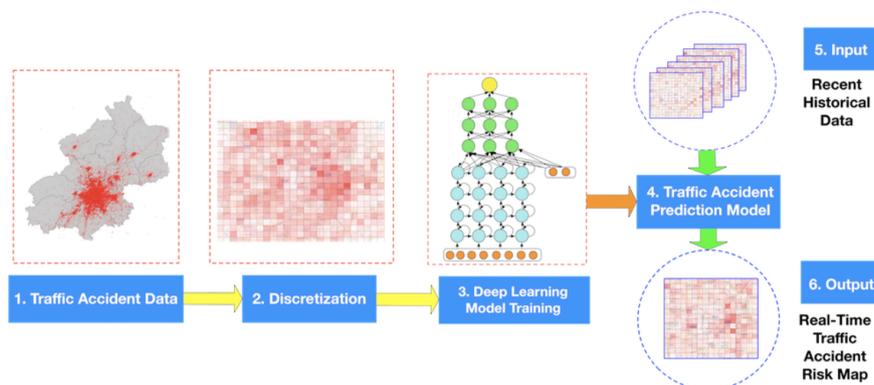


Figura 4.1: Modelo para previsão de acidentes por localidade e horário

Por fim, existem trabalhos mais abrangentes que englobam identificação, classificação e previsão. Trabalhos como WANG *et al.* (2010); BA *et al.* (2017); ELAMRANI ABOU ELASSAD *et al.* (2020a), construíram seus modelos com base nos

dados de sensores de veículos, acrescidos de dados psicológicos e tempo. Os trabalhos analisam pedaços anteriores aos acidentes para ensinar o modelo, para então, usá-lo na predição das janelas subsequentes de tempo. Na figura 4.2 é apresentada a arquitetura de BA *et al.* (2017), mas que representa muito bem os demais trabalhos.

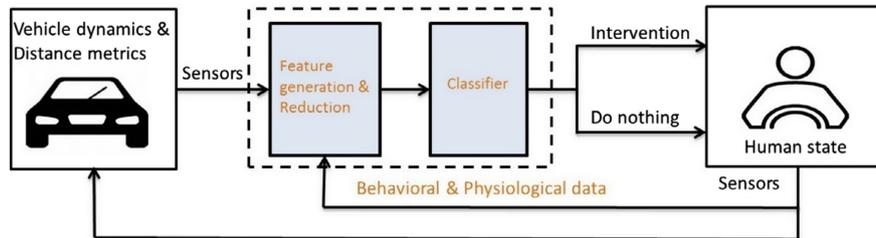


Figura 4.2: Modelo para sistema completo de predição de acidentes

4.2 Conjuntos de dados

Além das abordagens mencionadas sobre sistemas de predição de acidentes (CPS), é fundamental destacar a construção das fontes de dados, que representa um componente de extrema relevância nas análises de acidentes HALIM *et al.* (2016a); QU *et al.* (2017); GUTIERREZ-OSORIO e PEDRAZA (2020). Os dados empregados nessas análises são obtidos a partir de sensores instalados em veículos, e baseado nesses dados é possível discernir padrões de manobras e eventos. Tais conjuntos de dados consistem em séries temporais primárias, com uma dimensão temporal e um eixo representativo das características capturadas. O cerne dos CPS repousa na utilização desses dados para resolver uma questão de classificação de série temporal.

Antes que os conjuntos de dados sejam aplicados nos CPS, eles passam por um processo de pré-processamento, englobando etapas como limpeza, seleção e transformação. Isso é conduzido para resolver questões específicas, como já explorado no capítulo anterior. Cada pesquisa adota técnicas de seleção e transformação de acordo com seus objetivos. Deste modo, surgem técnicas adequadas para tratar conjuntos de dados brutos em formato de séries temporais, bem como métodos apropriados para lidar com conjuntos de dados previamente transformados WU *et al.* (2016); FUGIGLANDO *et al.* (2018); ARAÚJO *et al.* (2012); LIN *et al.* (2014). A estrutura usual dos conjuntos de dados nos CPS assume a forma de uma matriz, em que as linhas denotam eventos e as colunas representam suas características (*features*). Tais *features* podem ser representadas pelo tempo ou por um conjunto de métricas derivadas dos sinais coletados, conforme ilustrado na Figura 2.11, no estudo de WU *et al.* (2016).

No que tange às transformações dos dados, uma ampla variedade de abordagens é empregada. Por exemplo, em KLUGER *et al.* (2016), os autores aplicaram a

Transformada de *Fourier* discreta para para ajustar o conjunto de dados, a fim de identificar e classificar situações de direção segura ou insegura. De modo similar, em WANG *et al.* (2010), os dados provenientes de diversos sensores foram usados para construir um conjunto de dados transformado, mediante derivação dos dados originais por meio de uma janela deslizante (Figura 4.3). Por sua vez, AKIN e AKBA (2010) adotaram transformações mais simples, envolvendo médias, mínimos, máximos e desvio padrão. Há também casos nos quais os pesquisadores utilizaram diretamente os dados brutos das séries temporais dos sensores em seus conjuntos de dados, como ilustrado nos trabalhos de BA *et al.* (2017); ELAMRANI ABOU ELAS-SAD *et al.* (2020b).

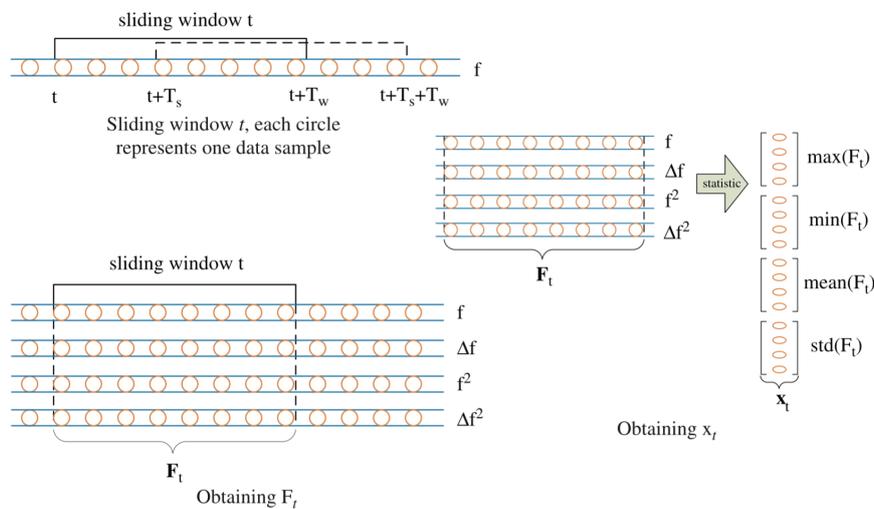


Figura 4.3: Construção do conjunto de dados do trabalho WANG *et al.* (2010)

Frequentemente, os estudos obtêm os conjuntos de dados prontos para uso, visto que nem todos os projetos realizam coleta direta dos dados provenientes dos sensores. Em todos os casos, a origem dos dados é claramente indicada. Tais conjuntos podem ser oriundos de fontes públicas, serem adquiridos por meio de órgãos de trânsito ou coletados especificamente para a pesquisa em questão HALIM *et al.* (2016a); GUTIERREZ-OSORIO e PEDRAZA (2020). Os conjuntos de dados providos por departamentos de trânsito são bastante comuns, geralmente devido à acessibilidade dos mesmos, que em alguns casos são de domínio público.

Outra faceta relevante na construção dos conjuntos de dados envolve os tipos de dados utilizados e sua representatividade. Nos artigos sobre CPS os conjuntos de dados são categorizados de acordo com suas representações. Os tipos de representação mais comuns em análises de CPS incluem:

1. **Informações espaço-tempo** - Esses conjuntos englobam dados sobre a localização e o momento do acidente, incluindo informações como horário, local e tipo de estrada;

2. **Informações ambientais** - Esses conjuntos contêm dados acerca do clima e das condições ambientais, além de dados da superfície da estrada e iluminação;
3. **Informações demográficas do motorista** - Esses conjuntos apresentam informações sobre o motorista, como gênero, idade, escolaridade, entre outras;
4. **Informações fisiológicas do motorista** - Esses conjuntos englobam informações sobre aspectos fisiológicos, tais como frequência cardíaca, frequência respiratória e dilatação dos olhos;
5. **Dados dos sensores do veículo** - Essa é a categoria mais recorrente, compreendendo informações oriundas de sensores como acelerômetros, giroscópios e GPS.

Os dados provenientes de sensores, por sua ampla utilização, possuem um respaldo robusto na literatura. Autores como XIONG *et al.* (2019); BA *et al.* (2017); ELAMRANI ABOU ELASSAD *et al.* (2020a); QU *et al.* (2017); KLUGER *et al.* (2016); PAEFGEN *et al.* (2013); WANG *et al.* (2017) empregaram dados de sensores veiculares (acelerômetros, giroscópios e GPS) para capturar séries temporais, as quais foram posteriormente transformadas em novos conjuntos de dados mediante diversas técnicas, desde janelas deslizantes até métricas mais simples, como média e desvio padrão. Outras duas categorias notáveis são as de **Informações espaço-tempo** e **Informações ambientais**. Exemplificam-se aqui os trabalhos de CASTRO e KIM (2016); ZHU *et al.* (2018); LONGHI e NANNI (2019); AKIN e AKBA (2010); AIT-MLOUK e AGOUTI (2019); REN *et al.* (2018), que empregaram esses tipos de dados para compreensão das principais características dos eventos críticos, fundamentando-se em informações espaciais-temporais e ambientais para a construção de seus modelos.

4.3 Técnicas utilizadas

As técnicas de aprendizado de máquina desempenham um papel central na exploração dos melhores enfoques para os CPS. Uma investigação criteriosa das técnicas empregadas é essencial para a compreensão desses sistemas.

Dentre as técnicas frequentemente utilizadas, o aprendizado supervisionado emerge como o paradigma predominante na análise de padrões de acidentes, uma vez que os CPS regularmente operam sob um contexto de classificação. Nesse sentido, as Redes Neurais Artificiais despontam como uma escolha proeminente para técnicas de classificação, particularmente quando se trata de identificar comportamentos de direção insegura. Essa abordagem pode ser observada em muitas investigações, como nos estudos de HALIM *et al.* (2016a); ELAMRANI ABOU ELASSAD *et al.*

(2020a); QU *et al.* (2017); CASTRO e KIM (2016); ZHU *et al.* (2018); PAEFGEN *et al.* (2013); WANG *et al.* (2017); LONGHI e NANNI (2019); AKIN e AKBA (2010). Além de detectar direção insegura, essas investigações também abordam a identificação das condições precursoras de acidentes e a proposição de modelos para estimar a gravidade dos incidentes GUTIERREZ-OSORIO e PEDRAZA (2020).

Outra abordagem de destaque é a *Support Vector Machine* (SVM). Largamente empregada na classificação de eventos como acidentes ou situações normais, essa técnica figura em estudos como ELAMRANI ABOU ELASSAD *et al.* (2020a); QU *et al.* (2017); GUTIERREZ-OSORIO e PEDRAZA (2020); WANG *et al.* (2017); LONGHI e NANNI (2019).

Por fim, as técnicas de Árvore de Decisão, em especial o método de *Random Forest*, merecem destaque por sua recorrência e relevância nos resultados obtidos. As árvores de decisão constituem ferramentas valiosas em aplicações de classificação de padrões, oferecendo flexibilidade e capacidade de lidar com conjuntos de dados complexos e incompletos. Além disso, sua natureza visualmente interpretável contribui para uma compreensão mais intuitiva dos resultados. Essas técnicas de Árvore de Decisão estão presentes em diversos trabalhos, tais como ELAMRANI ABOU ELASSAD *et al.* (2020a); CASTRO e KIM (2016); ZHU *et al.* (2018); PAEFGEN *et al.* (2013); WANG *et al.* (2017); LONGHI e NANNI (2019).

Capítulo 5

Metodologia e dados

Neste capítulo apresentaremos o processo metodológico desenvolvido para alcançar os objetivos de nossa pesquisa. Nosso propósito é **identificar** e **classificar** eventos críticos rodoviários, usando para isso dados de acelerômetros, giroscópios e GPS. Neste sentido, desenvolvemos um novo algoritmo que pode examinar e classificar um evento crítico, como acidente, quase-acidente ou manobras abruptas, usando um processo de decomposição de *wavelet* multinível. Neste capítulo, detalharemos o processo usado para aquisição dos dados, pré-processamento de dados, processamento com a incorporação de técnicas de aprendizado de máquina e avaliação de resultados.

O processo consiste em cinco subprocessos e começa com a coleta de dados do SHRP2 *Naturalistic Driving Study* DAVOODI *et al.* (2020). Em seguida, o pré-processamento envolve etapas como limpeza, transformação de dados, **identificação** de eventos, extração dos dados relevantes e a divisão dos dados em conjuntos de treinamento e teste. Por fim, treinamos os classificadores e avaliamos os resultados para a geração do modelo. A figura 5.1 mostra todo nosso processo metodológico. A seguir detalharemos os subprocessos e suas atividades.

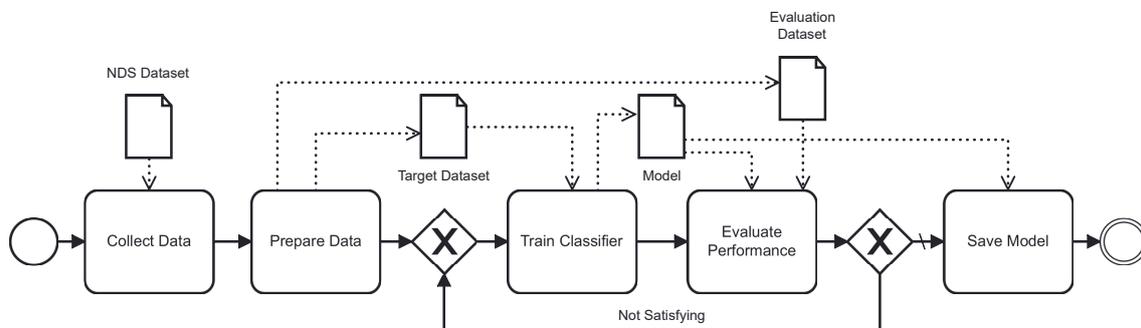


Figura 5.1: Metodologia do processo

5.1 Coleta de Dados

A fase de coleta de dados aborda a aquisição das informações utilizadas em nossa pesquisa. É relevante ressaltar que nosso trabalho não incluiu a criação de um processo de coleta próprio. Ao invés disso, utilizamos dados já previamente capturados por sensores veiculares. Para nosso estudo, empregamos os conjuntos de dados provenientes do projeto SHRP2 *Naturalistic Driving Study* (NDS) DAVOODI *et al.* (2020). Esses dados foram disponibilizados pelos engenheiros do projeto e obtidos através de serviços de armazenamento de acesso público. O projeto SHRP2 tem como propósito a pesquisa no contexto de ambientes de direção e segurança viária. Esses dados são oferecidos a pesquisadores qualificados mediante licenciamento. Ao longo de mais de oito anos, o projeto foi conduzido pelo Instituto Virginia Tech (VT), envolvendo mais de 3.000 motoristas de diversas regiões dos Estados Unidos. Os motoristas selecionados receberam conjuntos de sensores instalados em seus veículos, responsáveis por coletar informações durante suas viagens. Os dados passam por uma fase de pré-processamento e classificação, ou seja, eles já vêm rotulados quanto a categorias como *crash*, *near-crash* ou *others* (outras manobras abruptas).

5.2 Preparação dos Dados

Na etapa de preparação dos dados, o primeiro passo foi determinar que tipo de informações seriam relevantes para a nossa pesquisa. Como já discutido na seção 4.2, os conjuntos de dados normalmente são agrupados de acordo com a sua representatividade. O projeto SHRP2 contém diversos tipos de dados, incluindo informações espaço-temporais, detalhes demográficos dos motoristas, dados sobre os veículos e informações sobre os próprios eventos, como *crash*, *near-crash* ou *others*. Para o nosso estudo, optamos por focar nas informações relacionadas aos eventos, que incluem dados de um conjunto de sensores como acelerômetros, giroscópios, GPS e informações do *CAN-Bus* do veículo. Os sinais dos sensores consistem em séries temporais que correspondem ao momento do evento crítico, como por exemplo, durante um acidente.

Nessa fase, a próxima atividade é a seleção das *features* que utilizaremos em nossa pesquisa. Os dados dos eventos recebidos contêm 50 *features*, coletadas por mais de cinco sensores distintos. Para nosso estudo, estamos interessados nos dados dos sensores de acelerômetro, giroscópio e GPS. A partir dessa seleção, optamos pelas seguintes *features*: *timestamp*, *longitudinal acceleration*, *lateral acceleration*, *yaw rate*, *heading GPS*, *speed* e *event type*. Essa seleção foi embasada em diversos estudos, como CASTIGNANI *et al.* (2015); SAIPRASERT *et al.* (2017); BA *et al.* (2017); YLIZALITURRI-SALCEDO *et al.* (2015); LONGHI e NANNI (2019).

Após a seleção das *features*, a fase de transformação dos dados é iniciada. Essa etapa é crucial, uma vez que os dados recebidos, apesar de já terem passado por tratamentos e validações, precisam ser ajustados para se adequarem aos objetivos da nossa pesquisa. Diversas transformações são aplicadas, incluindo a padronização dos *timesteps* para uma frequência de 10 Hz, ajustes nas unidades de velocidade e data, e a remoção de *features* nulas. Uma atividade fundamental foi a utilização de técnicas de janelas deslizantes para extrair segmentos de dados de comprimentos padronizados a partir dos dados brutos. O nosso objetivo era garantir que essas extrações contivessem o evento crítico, com tamanho de 15 segundos. Esse processo está relacionado à **identificação de eventos** mencionada anteriormente em nosso trabalho.

Para garantir tamanhos uniformes de janelas, adotamos uma abordagem de janela deslizante para detectar o início e o fim dos eventos, seguindo práticas recomendadas em estudos como BAGNALL *et al.* (2017, 2020); RUIZ *et al.* (2021); LINES *et al.* (2018). Por fim, convertemos nosso conjunto de dados para um formato conhecido como *nested format*, onde cada linha representa um evento (*crash*, *near-crash* ou *others*) e as colunas listas das *features* selecionadas, mantendo a ordem temporal do conjunto de dados original. Essa atividade foi realizada devido ao fato de que os **algoritmos específicos** utilizados na nossa pesquisa foram projetados para lidar com dados nesse formato, ao contrário do formato tabular utilizado nos **algoritmos tradicionais**.

5.3 Treinamento dos Classificadores

Nesta etapa, realizamos a tarefa central do nosso trabalho, a **classificação**. Essa tarefa é composta por duas atividades cíclicas: o treinamento e os testes dos classificadores selecionados para a pesquisa. O processo se inicia com a escolha do tipo de amostragem dos dados. Considerando o desbalanceamento dos dados recebidos, testamos duas abordagens de amostragem: a estratificada e a balanceada. Com base em análises empíricas, optamos pela amostragem balanceada, uma escolha respaldada pelo trabalho de ELAMRANI ABOU ELASSAD *et al.* (2020a). Nesse estudo, os autores justificam a amostragem balanceada face às características dos conjuntos de dados de acidentes, que geralmente apresentam uma quantidade significativamente maior de exemplos não-acidentes. Essa disparidade pode afetar a performance dos classificadores.

Em seguida, procedemos ao treinamento e teste dos classificadores selecionados para a nossa pesquisa. Avaliamos algoritmos tradicionais, algoritmos específicos e também a nossa proposta de algoritmos de aprendizado de máquina. Os algoritmos específicos são utilizados em problemas nos quais a ordem dos atributos é relevante

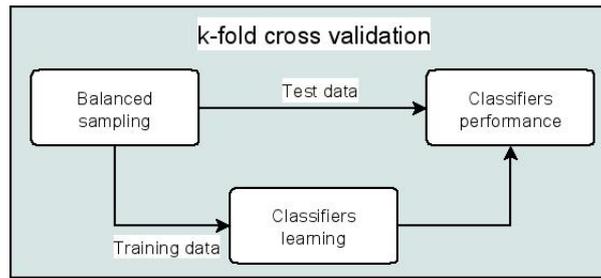


Figura 5.2: Processo de validação da metodologia

BAGNALL *et al.* (2017). Empregamos uma variedade de soluções desse tipo, uma vez que eles são recomendados para lidar tanto com tarefas de classificação quanto de regressão em séries temporais. Já os algoritmos tradicionais são os mais conhecidos e consolidados no campo do aprendizado de máquina. Todos os algoritmos foram submetidos a um processo de otimização a partir de centenas de testes considerando diversas variações. A escolha dos melhores parâmetros se baseou nos resultados de maior acurácia. No processo de otimização, também utilizamos a orientação de dois trabalhos, principalmente no caso dos algoritmos específicos, que nos forneceram um ponto de partida para ajustar os classificadores. Os estudos de BAGNALL *et al.* (2017); RUIZ *et al.* (2021) apresentam listas abrangentes de parâmetros previamente testados e as melhores configurações.

A última atividade consistiu nos testes dos algoritmos em relação ao conjunto de dados, ou seja, a execução da classificação propriamente dita. Desenvolvemos um procedimento de teste que englobou todos os algoritmos, incluindo a nossa proposta, utilizando cinco conjuntos de dados univariados distintos, cada um contendo apenas uma *feature*. Para melhorar a robustez dos resultados, empregamos a técnica de validação cruzada k-fold, repetindo o processo várias vezes para evitar problemas de *overfitting* (Figura 5.2).

5.4 Avaliação de Desempenho

Nesta etapa, nosso foco é avaliar o desempenho dos classificadores. Conforme ilustrado na Figura 5.1, essa avaliação ocorre de maneira cíclica e integrada à etapa de treinamento dos classificadores. Após o treinamento, é neste estágio que avaliamos o desempenho dos classificadores em relação ao conjunto de dados de teste. Caso o desempenho não seja satisfatório, o algoritmo de classificação é retroalimentado para a fase de treinamento. Esse ciclo é repetido até atingir o máximo de acurácia possível.

A principal medida de avaliação em nosso estudo é a acurácia, uma métrica amplamente reconhecida e considerada essencial na análise de classificação RUIZ *et al.*

(2021). Além da acurácia, nossas avaliações também consideram outras métricas, como precisão, sensibilidade, especificidade, taxa de falsos positivos (FPR) e tempo de processamento. Essas medidas são aplicadas a todos os casos (cinco conjuntos de dados) apresentados na Seção 5.3. Importante ressaltar que o objetivo dessa etapa é encontrar o melhor modelo para cada algoritmo e, adicionalmente, comparar a efetividade entre eles.

Capítulo 6

WALE-a: Camadas Adaptativas Baseada em Wavelets

Nossa proposta é dividida em duas tarefas principais: **identificação** e **classificação**. Na etapa de identificação, desenvolvemos um processo que compreende a coleta e o tratamento dos dados dos sensores, bem como a extração da parte relevante, ou seja, o evento crítico. Quanto à classificação, apresentamos o algoritmo **WALE-a**, responsável por receber a coleção de eventos críticos, submetê-los a técnicas de decomposição de dados e, em seguida, classificá-los de acordo com sua criticidade. No contexto deste estudo, consideramos três classes: acidente (*crash*), quase-acidente (*near-crash*) e manobra abrupta (*others*). Enquanto esta seção focaliza a apresentação do algoritmo **WALE-a**, as demais técnicas relacionadas à identificação serão abordadas no capítulo dedicado aos experimentos.

A principal fonte de inspiração para o algoritmo **WALE-a** foi a arquitetura em camadas das redes neurais convolucionais (CNN), reconhecidas por sua aplicação em problemas complexos de classificação, categorização de texto, visão computacional, processamento de imagens e reconhecimento de fala GUTIERREZ-OSORIO e PEDRAZA (2020). Nas CNNs, os dados de entrada percorrem a rede e são transformados até sua classificação final. Nosso algoritmo emprega uma estrutura que incorpora métodos de transformação de séries temporais, seguidos pela etapa de classificação. Como mencionado anteriormente, os dados brutos dos sensores correspondem a séries temporais. Portanto, nosso algoritmo se vale de uma variedade de técnicas para transformação e classificação dessas séries temporais. A classificação de séries temporais é amplamente estudada na literatura, conforme evidenciado por trabalhos como BAGNALL *et al.* (2017); ISMAIL FAWAZ *et al.* (2019); HU *et al.* (2013); AGGARWAL (2014); BIER *et al.* (2022); SALLES *et al.* (2019). Nosso algoritmo se dedica à classificação de séries temporais univariadas, ou seja, empregamos apenas uma *feature* coletada dos sensores. Abordaremos essa questão de forma mais detalhada no capítulo dedicado aos experimentos.

6.1 Classificação de Séries Temporais

Os dados coletados dos sensores internos e externos dos veículos no contexto do CPS são representados como séries temporais em formato bruto. Esses dados são fundamentais para a análise e previsão de possíveis acidentes. Do ponto de vista do aprendizado de máquina, o CPS é caracterizado como um problema de classificação de séries temporais. A tarefa de classificação consiste em empregar abordagens baseadas em dados para aplicar algoritmos de aprendizado de máquina à série temporal em questão. O objetivo é determinar se cada dado corresponde a um acidente, quase-acidente ou manobra abrupta.

Diversas técnicas de aprendizado de máquina estão disponíveis para a classificação de séries temporais (conforme detalhado no capítulo 4). As técnicas mais utilizadas são amplamente conhecidas e são consideradas abordagens tradicionais neste campo. No entanto, nos últimos anos, tem havido um crescente interesse na aplicação de algoritmos específicos projetados para lidar com tarefas de classificação de séries temporais. Tais algoritmos se destacam por considerarem a ordem dos atributos como fator crucial BAGNALL *et al.* (2017). Eles se revelam particularmente eficazes em situações em que a natureza temporal das séries desempenha um papel significativo. De fato, Bagnal et al. BAGNALL *et al.* (2017) demonstram que algoritmos de classificação de séries temporais frequentemente superam os classificadores tabulares tradicionais ao lidar com problemas desse tipo. No entanto, é importante observar que algoritmos específicos para séries temporais podem apresentar um desempenho mais lento, conforme apontado por BAGNALL *et al.* (2017, 2020); DEMPSTER *et al.* (2020); MIDDLEHURST *et al.* (2021). Neste contexto, nosso trabalho compara nossa proposta com algoritmos tradicionais e específicos, equilibrando a precisão e a eficiência, fatores imprescindíveis na previsão de acidentes BA *et al.* (2017).

6.2 Transformação de Séries Temporais

A eficácia de um sistema de previsão de acidentes depende, em grande parte, do desempenho dos algoritmos de aprendizado de máquina que o compõem. Nesse contexto, as técnicas de pré-processamento desempenham um papel fundamental. No âmbito do CPS, os dados utilizados em análises provêm de várias fontes e passam por fases de limpeza, transformação e pré-processamento. Essas etapas englobam a remoção de dados inválidos, o preenchimento de valores ausentes, a normalização, a padronização, a seleção de atributos e a transformação dos dados YADAV *et al.* (2019); WANG *et al.* (2017); AIT-MLOUK e AGOUTI (2019).

As técnicas de transformação de dados, como as transformadas de *Fourier* e

wavelet, são empregadas para analisar séries temporais. Os dados transformados são, dessa forma, empregados em tarefas de previsão e classificação. A adoção de diferentes transformações e representações permite a identificação de padrões mais significativos para a discriminação entre séries temporais de diferentes grupos, contribuindo para o aprimoramento do desempenho de modelos TAN *et al.* (2022). Na revisão de SALLES *et al.* (2019), são discutidos variados tipos de transformações de séries temporais utilizadas na previsão, incluindo transformações do tipo mapeamento e do tipo divisão. Os autores ressaltam que a transformação adequada dos dados é essencial para garantir a qualidade e a precisão dos resultados. Em KLUGER *et al.* (2016), os pesquisadores aplicam a transformada de *Fourier* (FT) às séries temporais originadas de acelerômetros, resultando em amplitudes para frequências que compõem um novo conjunto de dados. No entanto, o FT apresenta limitações, fornecendo apenas informações sobre as frequências presentes, mas carecendo de detalhes sobre a localização e o tempo. Essas limitações são superadas pela transformada de *wavelet* (WT), que se destaca por capturar, com mais precisão, a localização de cada frequência TAN *et al.* (2022).

A WT oferece uma solução flexível no âmbito tempo e frequência, o que a torna especialmente útil para a análise de séries temporais não lineares, ruidosas e não estacionárias TOO *et al.* (2019a). A WT decompõe uma série temporal ao correlacioná-la com uma família de *wavelets*, resultando em uma representação de tempo-frequência flexível SUBASI (2020). Segundo SUBASI (2020), a WT é uma ferramenta excelente para auxiliar na classificação de sinais. Isso é evidenciado em TOO *et al.* (2019a), onde os autores aplicam a WT para classificar dados de eletromiografia (EMG). Similarmente, AL OMARI *et al.* (2014) classifica oito movimentos da mão utilizando sinais EMG, e avalia o desempenho em relação a diferentes famílias de *wavelets* em combinação com a técnica de aprendizado de máquina SVM. Em PHINYOMARK *et al.* (2012), os autores exploram a utilidade da decomposição *wavelet* multinível para remover o ruído de séries temporais e extrair atributos eficazes, visando melhorar a separabilidade das classes no espaço das *features*. Esses exemplos deixam claro que a análise em várias resoluções, proporcionadas pela WT, pode ser uma opção valiosa para tarefas de classificação.



Figura 6.1: Aplicação de filtros **passa-alta** e **passa-baixa** ao sinal

6.3 Estrutura do Algoritmo WALE-a

Desenvolvemos um algoritmo composto por quatro camadas, onde as transformações dos dados ocorrem nas três primeiras camadas e a classificação na última. A entrada do algoritmo são os eventos críticos (acidentes, quase-acidentes e manobras abruptas) em formato bruto, ou seja, representados como séries temporais. Para realizar as transformações nos dados, utilizamos decomposições *wavelet* em conjunto com métricas estatísticas.

As transformações *wavelet* são amplamente empregadas na análise de séries temporais. Elas desdobram uma série temporal (sinal), ao correlacioná-la com uma família de *wavelets*, fornecendo uma representação flexível em termos de tempo e frequência. Isso é essencial para extrair informações de alta qualidade SALLES *et al.* (2019); TAN *et al.* (2022); TOO *et al.* (2019a). A transformada *wavelet* é uma abordagem que descreve as características tempo-frequência de um sinal, tornando-se crucial para otimizar o desempenho em tarefas de classificação SUBASI (2020); PHINYOMARK *et al.* (2012).

A transformada *wavelet* é categorizada em métodos discretos (DWT) e contínuos (CWT). Em nosso trabalho, optamos pelo DWT devido à sua adequação e ampla utilização em cenários reais, como também observado por PHINYOMARK *et al.* (2012). O DWT divide o sinal em vários subconjuntos de coeficientes por meio da aplicação de filtros **passa-alta** e **passa-baixa** ao sinal original. Esses filtros são determinados pelos coeficientes $h = \{h_n\}_{n \in \mathbb{Z}}$, correspondente ao filtro *passa-alta*, e $g = \{g_n\}_{n \in \mathbb{Z}}$, correspondente ao filtro **passa-baixa** (Figura 6.1).

A aplicação dos filtros resulta na decomposição do sinal em duas faixas de frequência. Podemos encadear uma série de bancos de filtros, usando a operação de *downsampling* para dividir a frequência por 2 (conforme ilustrado na figura 6.2) a cada novo banco de filtros encadeado. Dessa forma, obtemos um sinal de detalhe específico para cada faixa de frequência em nossa análise do sinal.

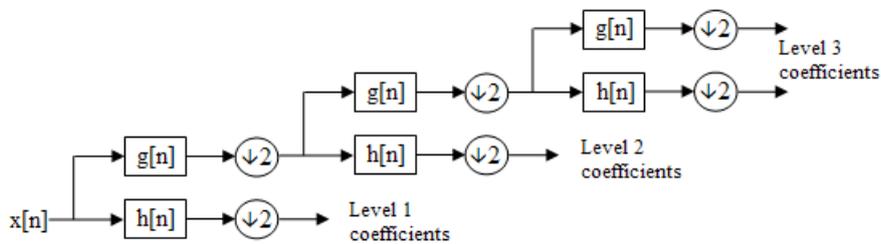


Figura 6.2: Encadeamento dos bancos de filtro na decomposição do sinal

A aplicação dos filtros h e g resulta em dois conjuntos de coeficientes: o conjunto de coeficientes de detalhe (cD) e o conjunto de coeficientes de aproximação (cA) PHINYOMARK *et al.* (2012). A saída dos primeiros filtros é representada por cA1

e $cD1$ (nível 1). O processo de decomposição é repetido até que o número desejado de níveis seja alcançado AMIN *et al.* (2015). A figura 6.3 ilustra a decomposição de uma série temporal.

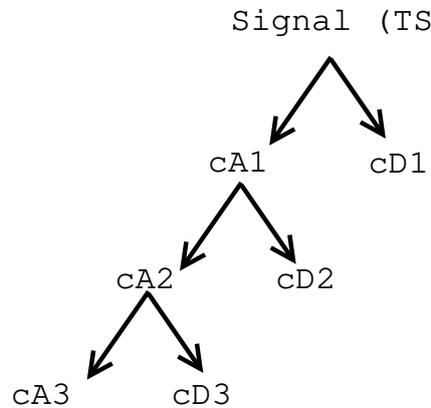


Figura 6.3: Decomposição de séries temporais

Os resultados dos coeficientes de detalhe dependem da função *wavelet* utilizada. A escolha da função *wavelet* está diretamente relacionada ao desempenho da classificação, como evidenciado nos estudos de SALLES *et al.* (2019); TAN *et al.* (2022); PHINYOMARK *et al.* (2012); AMIN *et al.* (2015). Em nosso trabalho, optamos por utilizar diversas funções *wavelet* em paralelo, visando maximizar a capacidade de reconhecimento de padrões. O número de níveis de decomposição varia de acordo com o tamanho da janela da série temporal de entrada e a função *wavelet* utilizada. A figura 6.4 proporciona uma visão geral da estrutura do nosso algoritmo por camada.

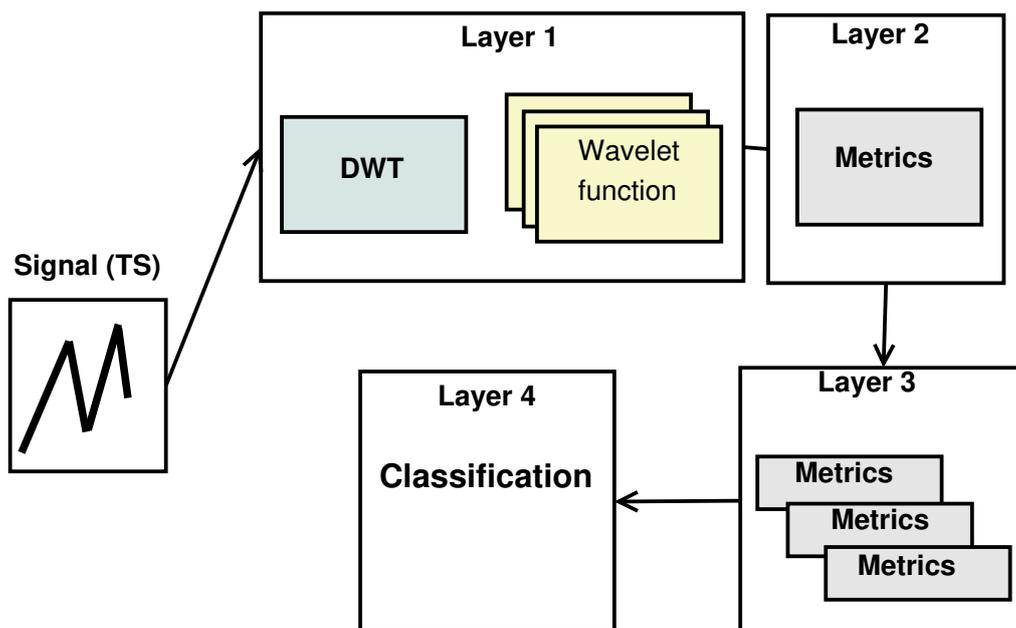


Figura 6.4: Visão geral das camadas do WALE-a

6.4 As Camadas do WALE-a

Nosso algoritmo é estruturado em camadas que realizam um processo progressivo de transformação dos dados. O evento crítico passa por essas camadas, sendo submetido a transformações e decomposições, para então ser classificado com base em sua criticidade. Iniciaremos a descrição de cada camada do **WALE-a**.

6.4.1 Camada de Decomposição

A primeira etapa do nosso algoritmo é a **Camada de Decomposição**. Nessa fase, o evento (representado como uma série temporal), é submetido a uma técnica de janela deslizante. O objetivo disso é criar segmentos distintos do dado original. O tamanho da janela, que é o **primeiro parâmetro** do algoritmo, determina a quantidade e o comprimento dos segmentos gerados. Posteriormente, aplicamos a função *wavelet* em cada segmento de dados obtidos da janela deslizante. Importante destacar que o algoritmo pode trabalhar com uma única função *wavelet* ou uma lista delas, determinadas pelo **segundo parâmetro**. Isso resulta em uma aplicação cíclica das funções *wavelets*.

A aplicação de cada função *wavelet* em cada segmento de dado resulta nas decomposições mencionadas na seção 6.2. Para realizar essas decomposições, implementamos três estratégias que auxiliam na criação de diferentes visões do mesmo evento. Essas estratégias consistem em variações das aplicações da função *wavelet* nos segmentos de dados, originados da janela deslizante mencionada anteriormente. As estratégias são as seguintes:

1. **Estratégia de Decomposição 1:** Nessa estratégia, os segmentos de dados gerados pelo processo da janela deslizante são mantidos integralmente, ou seja, não é aplicada a função *wavelet*.
2. **Estratégia de Decomposição 2:** Nessa abordagem, os segmentos de dados são decompostos por meio da função *wavelet*, resultando nos coeficientes cA (aproximação) e todos os cDs (detalhes). Essa estratégia emprega o máximo de níveis de decomposição possível, conforme ilustrado na Figura 6.3.
3. **Estratégia de Decomposição 3:** Essa estratégia reconstrói cada segmento de dado, com base nas decomposições da estratégia anterior (**Decomposição 2**). Para essa reconstrução, o algoritmo utiliza os coeficientes cA e o último cD.

Ao final dessa camada, o algoritmo armazena todas as estratégias de decomposição para cada segmento de dado gerado pela janela deslizante. Essas informações

são enviadas para a próxima camada. A Figura 6.5 proporciona uma visão geral da **Camada de Decomposição**.

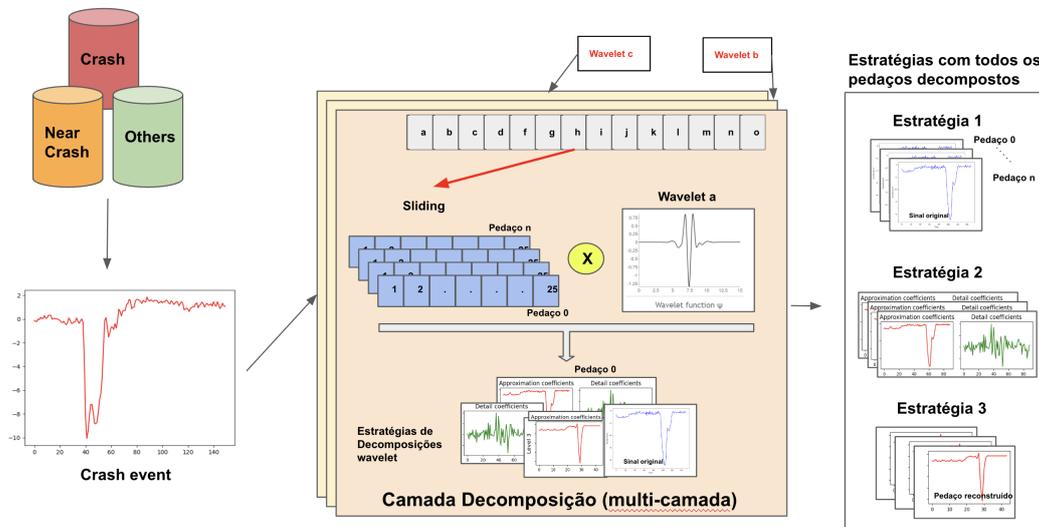


Figura 6.5: Visão geral da **Camada de Decomposição** do **WALE-a**

Na próxima seção, detalharemos a **Camada de Extração**, que é a segunda etapa essencial do nosso algoritmo **WALE-a**.

6.4.2 Camada de Extração

A **Camada de Extração** recebe as três estratégias compostas pelas decomposições *wavelets* de cada segmento de dado, originado a partir do processo de janela deslizante aplicado ao evento. Nesse estágio, esse novo conjunto de dados passa por várias métricas estatísticas, como média, desvio padrão, máximo, mínimo e outras. Essas métricas foram selecionadas com base em diversos estudos já mencionados, tais como ARAÚJO *et al.* (2012); CARMONA *et al.* (2015); TOO *et al.* (2019a); AL OMARI *et al.* (2014); PHINYOMARK *et al.* (2012); MURPHEY *et al.* (2009). O objetivo é criar um novo conjunto de dados composto por métricas estatísticas concatenadas. Nesse novo conjunto, cada segmento e suas decomposições *wavelets* são transformados em métricas estatísticas. As métricas utilizadas em nosso estudo estão detalhadas na Tabela 6.1. A Figura 6.6 proporciona uma visão da **Camada de Extração**.

Ao término desta camada, o novo conjunto de dados é formatado como uma lista de estratégias, contendo todas as métricas concatenadas, derivadas das decomposições dos segmentos de dados. Essa lista é então enviada para a próxima camada. A próxima seção abordará a terceira camada, a **Camada de Pooling**.

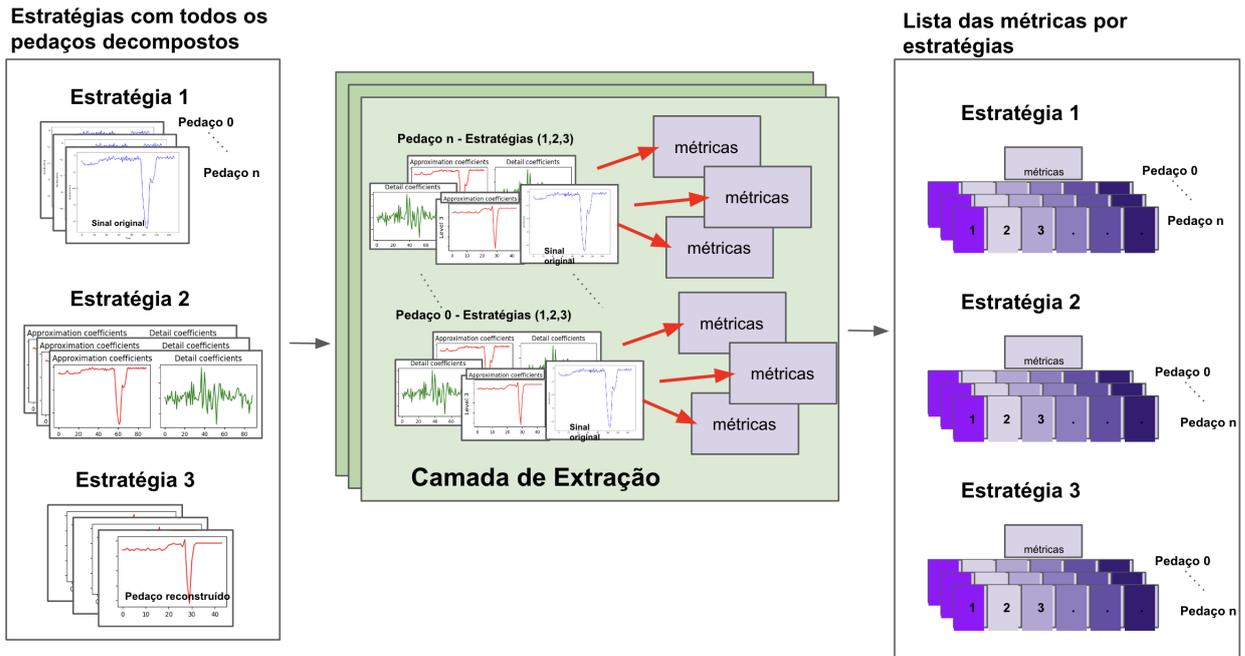


Figura 6.6: Visão da **Camada de Extração** do **WALE-a**

6.4.3 Camada de *Pooling*

A **Camada de *Pooling*** tem como objetivo reduzir a lista de estratégias provenientes da **Camada de Extração**. Para alcançar essa redução de dados, implementamos cinco abordagens de *pooling*. Essas abordagens estão em sintonia com vários estudos da literatura científica, como, por exemplo, *DEMPSTER et al. (2020)*; *TAN et al. (2022)*; *DEMPSTER et al. (2021)*; *COATES et al. (2011)*. As abordagens de *pooling* são: média, proporção de valores positivos (PPV), média de valores positivos (MPV), máximo e soma. Ao entrar na camada, a lista de estratégias, agora representando **métricas estatísticas** para cada segmento de dado, é consolidada de acordo com a abordagem de *pooling* selecionada.

Por exemplo, se a abordagem escolhida for a média, a camada de *pooling* organiza os dados por estratégia e calcula a média de todos os segmentos de dados provenientes da janela deslizante. Esse processo resulta em um novo conjunto de dados, no qual cada evento é representado por um conjunto de métricas resumidas pela média. Esse novo conjunto de dados é uma nova série temporal que representa o evento crítico após passar pela **Camada de Decomposição**. A Figura 6.7 oferece uma visão geral da **Camada de Pooling**.

Ao término dessa camada, um novo conjunto de dados é encaminhado para a classificação. É importante destacar que o tamanho desse conjunto varia de acordo com o número de decomposições *wavelet* aplicadas, multiplicado pelo número de métricas utilizadas.

Tabela 6.1: Lista das métricas usadas

Metrics	Mathematical definition
Mean	$\frac{1}{N} \sum_{n=1}^N x_n$
STD	$\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$
Maximum	$\max\{x_n\}$
Minimum	$\min\{x_n\}$
Trapezoidal rule	$\int_a^b f(x) dx$
Root Mean Square	$\sqrt{\frac{1}{N} \sum_{n=1}^N x_n^2}$
Mean absolute value	$\frac{1}{N} \sum_{n=1}^N x_n $
Zero crossing	$\sum_{n=1}^{N-1} [\text{sgn}(x_n \times x_{n+1}) \cap x_n - x_{n+1} \geq 0;]$ $\text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$
Average power	$\frac{1}{N} \sum_{n=1}^N x_n^2$
Willison amplitude	$\sum_{n=1}^{N-1} f(x_n - x_{n+1})$ $f(x) = \begin{cases} 1 & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
Wavelength	$\sum_{n=1}^{N-1} x_n - x_{n+1} $

6.4.4 Camada de Classificação

A **Camada de Classificação** é o estágio onde ocorre a classificação do evento crítico em uma das categorias: *crash*, *near-crash* ou *others*. Essa camada é uma integração em nossa arquitetura, não sendo desenvolvida em nosso trabalho, mas utilizando algoritmos de classificação estabelecidos. Nessa etapa, escolhemos empregar um classificador linear baseado na técnica de *ridge*, inspirados nas pesquisas de Dempster et al. DEMPSTER *et al.* (2020, 2021).

Optamos por utilizar o classificador linear *ridge* devido às descobertas dos autores, que demonstraram sua eficácia em conjuntos de dados de pequeno porte, devido à sua notável capacidade de generalização. Para nosso contexto, essa escolha se mostrou satisfatória. O classificador *ridge* aborda problemas de mínimos quadrados ordinários introduzindo um termo de penalidade. Os coeficientes do *ridge* são ajustados para minimizar a soma dos quadrados residuais penalizados. Um parâmetro crucial, representado pela letra a , controla o grau de encolhimento, onde valores mais altos resultam em coeficientes mais robustos.

Para nossa pesquisa, consideramos dez valores para o parâmetro a , selecionados de um espaço logarítmico variando de -3 a 3 na base 10, alinhado com o trabalho de DEMPSTER *et al.* (2020). A **Camada de Classificação** está ilustrada na Figura 6.7.

Desta forma, nosso algoritmo, o **WALE-a**, completo em suas quatro camadas, realiza uma série de transformações e etapas de processamento para classificar even-

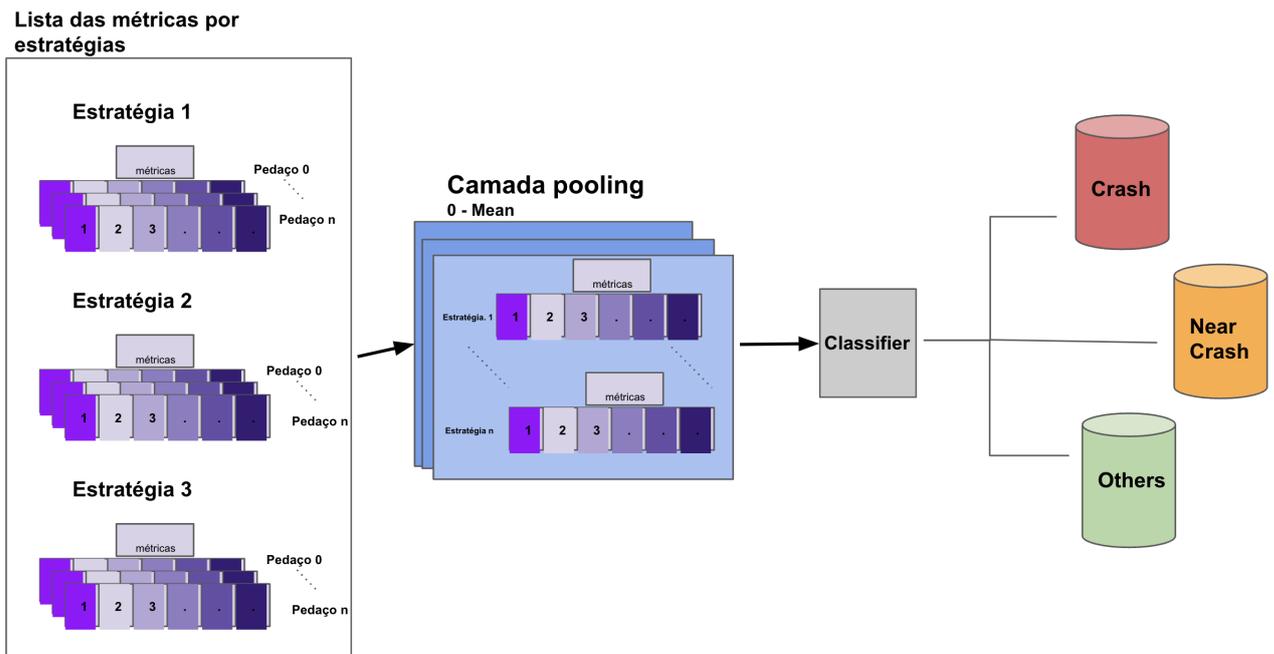


Figura 6.7: Visão da **Camada de Pooling** do **WALE-a**

tos críticos com base em séries temporais. Sua arquitetura progressiva e o uso de técnicas de transformação de dados, extração de métricas estatísticas e classificação, visam melhorar a precisão na identificação de eventos críticos de segurança veicular.

Capítulo 7

Experimentos

Neste capítulo, exploraremos a implementação prática das técnicas discutidas no Capítulo 5. Apresentaremos os passos principais do nosso experimento, detalhando as atividades planejadas no processo metodológico (Figura 5.1).

7.1 Configuração do Experimento

O primeiro passo deste capítulo consiste na configuração do nosso experimento. Desenvolvemos algoritmos para cada um dos subprocessos apresentados na Figura 5.1, desde a coleta até a avaliação dos classificadores. Esse conjunto de algoritmos é denominado de “*Classification Constructor*”. Nele, cada subprocesso é representado por um algoritmo independente, como ilustrado na Figura 7.1. O “*Classification Constructor*” engloba todos os códigos utilizados em nosso experimento, garantindo a execução em conformidade com o planejamento metodológico. Realizamos centenas de testes em um único notebook, utilizando um MacBook Pro com dez núcleos e 32 GB de RAM, executando *Python 3.10*, *Pandas 1.5.2*, *Scikit-learn 1.2* e *Sktime 0.15.0*.

7.2 Coleta de Dados

Conforme previamente demonstrado, nossos dados foram fornecidos pelo instituto *Virginia Tech Transportation*, responsável pelo projeto *SHRP2 Naturalistic Driving Study* (NDS) DAVOODI *et al.* (2020). Os dados foram transferidos por meio de serviços de armazenamento público, após um processo de negociações, autorizações e configurações do ambiente necessário para receber os dados do projeto. Foram mais de 60 dias dedicados a esse processo. No total, mais de 300GB de dados brutos referentes a eventos de *crash*, *near-crash* e *others* (outras manobras abruptas) foram transferidos, provenientes de diversos sensores. Os dados do projeto SHRP2-NDS

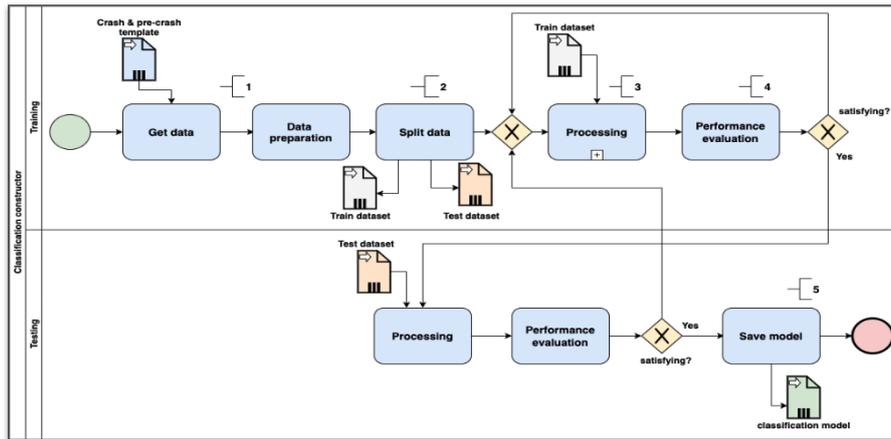


Figura 7.1: Sistema de execução dos experimentos *Classification Constructor*

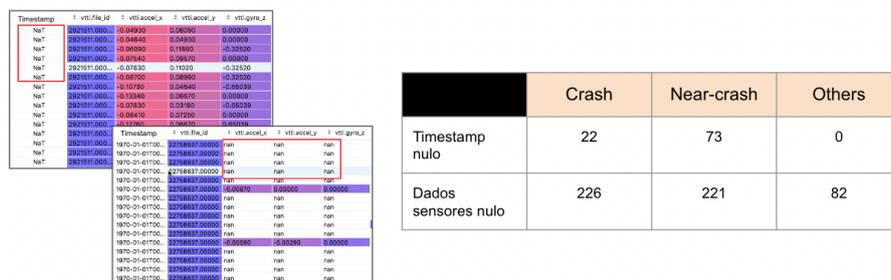


Figura 7.2: Visualização de registros nulos e a quantidade removida

DAVOODI *et al.* (2020) englobaram 19.339 ocorrências de manobras abruptas, 6.858 *near-crash* e 1.938 *crash*. Todos esses eventos passaram por um processo de validação realizado pelos engenheiros do projeto.

7.3 Preparação dos Dados

Com os dados em mãos, seguimos para o processo de preparação. O primeiro passo foi a escolha dos sensores e das *features* que seriam utilizados em nossa pesquisa. Baseando-nos em inúmeros trabalhos na área, tais como CASTIGNANI *et al.* (2015); SAIPRASERT *et al.* (2017); WU *et al.* (2016); BA *et al.* (2017); YLIZALITURRI-SALCEDO *et al.* (2015); LONGHI e NANNI (2019), selecionamos os sensores acelerômetro, giroscópio e GPS. Esses sensores não só possuem uma extensa literatura a respeito, como também são de fácil acesso e possuem custo reduzido, o que é uma preocupação central de nosso trabalho, para aprimorar sistemas de predição de acidentes (CPS). Quanto às *features*, optamos por incluir *timestamp*, *longitudinal acceleration*, *lateral acceleration*, *yaw rate*, *heading GPS*, *speed* e *event type*. Essas *features* representam os requisitos mínimos para a análise de dados em CPS e estão alinhadas aos objetivos deste trabalho.



Figura 7.3: Número de eventos selecionados e os tamanhos das janelas

Após uma análise minuciosa dos dados, identificamos os primeiros desafios a serem superados no conjunto de dados brutos. Notamos que as bases de eventos apresentavam tamanhos de janelas variáveis; por exemplo, as janelas da base de *crash* variavam entre 10 e 60 segundos, enquanto as da base de *others* variavam de 14 a 20 segundos. Observamos também a presença de muitos registros nulos em campos como *timestamp* dos sensores acelerômetro e giroscópio. Além disso, identificamos diferentes unidades de medida para variáveis como a velocidade, que estava em *km/h* para o GPS, e *m/s* para o acelerômetro. Nesta etapa, nosso algoritmo primeiro padronizou as unidades para *m/s* e removeu todos os registros com campos nulos, conforme ilustrado na Figura 7.2. Ao todo, mais de 600 registros foram removidos.

Com os dados limpos, concentramos nossos esforços na padronização dos tamanhos das janelas. Nosso objetivo era criar uma base de eventos com janelas de tamanho uniforme. Essa atividade é um subprocesso de extração de dados que envolve a seleção de partes relevantes das séries temporais. Para determinar os tamanhos das janelas, baseamo-nos em trabalhos como BAGNALL *et al.* (2017, 2020); RUIZ *et al.* (2021); LINES *et al.* (2018). Antes de executar a extração, analisamos os tamanhos mais comuns das janelas em todos os tipos de eventos, incluindo *crash*, *near-crash* e *others*. Nesta análise inicial, utilizamos o princípio de **Pareto** 80/20, que se mostrou adequado, pois a maioria dos eventos estava concentrada em um grupo reduzido de tamanhos de janela. A Figura 7.3 exibe o número de eventos selecionados e seus respectivos tamanhos de janela.

Após a pré-seleção mencionada anteriormente, nosso algoritmo de extração padronizou os tamanhos das janelas em toda a nossa base de dados. Na classificação de séries temporais, a questão da padronização do tamanho das janelas é muito discutida. Idealmente, deseja-se que todas as séries tenham o mesmo tamanho, como mencionado anteriormente. No entanto, não existe consenso absoluto sobre o tamanho ideal para as janelas, mas sim sugestões e boas práticas. Para este trabalho, dada sua natureza de classificação de séries temporais que representam eventos críticos de direção, optamos por janelas de 15 segundos, fundamentados em estudos como WU *et al.* (2016); BA *et al.* (2017); YANG ZHENG *et al.* (2014).

7.3.1 Identificação de Eventos

Dando continuidade ao processo de preparação dos dados, nosso algoritmo foi aplicado em todas as séries temporais (*crash*, *near-crash* e *others*), com o objetivo de extrair janelas de 15 segundos. Vale lembrar que, até aqui, tínhamos janelas variando entre 15 e 38 segundos, conforme selecionado na primeira etapa de extração. A etapa de **identificação** representa um dos pilares fundamentais deste trabalho, pois consiste na extração da parte relevante de um evento com tamanho fixo e pré-determinado. Nesse contexto, a identificação significa a seleção dos dados que compõem eventos potenciais, sejam eles *crash*, *near-crash* ou *others*. Para alcançar esse objetivo, desenvolvemos um algoritmo denominado “Seleção de Janelas” (“*Windows Selection*”), que visa extrair o início e o fim de um evento crítico com duração de 15 segundos.

O algoritmo “Seleção de Janelas” foi estruturado com quatro atividades principais, a saber:

1. Criação de janelas deslizantes para cada evento;
2. Cálculo de métricas para cada janela deslizante, gerando novas séries temporais;
3. Estabelecimento de um limite para cada métrica, a ser usado na detecção dos dados relevantes;
4. Seleção dos dados relevantes, ou seja, os que ultrapassam os limites estabelecidos pelas métricas.

Janelas Deslizantes

Nessa atividade, o algoritmo utiliza a técnica de janelas deslizantes em todos os eventos da nossa base de dados. Considerando os diferentes tamanhos dos eventos (variando de 15s a 38s), optamos por uma janela deslizante com tamanho de 1s. Essa janela percorre toda a série temporal, gerando novas séries temporais que correspondem a segmentos consecutivos. Por exemplo, em um evento de 35s, foram produzidas 34 novas séries (janelas) de 1s cada. A Figura 7.4 ilustra o funcionamento da técnica de janelas deslizantes em um evento do tipo *crash*.

Cálculo das Métricas

Com as novas séries temporais formadas pela aplicação da técnica de janelas deslizantes, nosso algoritmo utiliza três métricas reconhecidas para criar uma nova perspectiva sobre o evento, estabelecendo as condições para a extração da parte relevante de 15 segundos do evento. As métricas aplicadas são: **desvio padrão**,

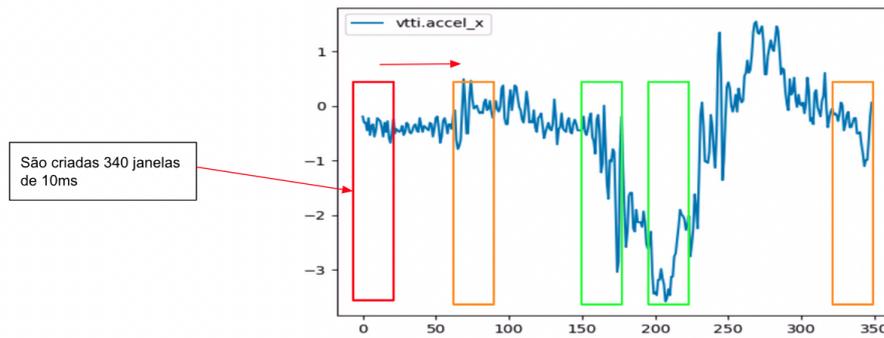


Figura 7.4: Funcionamento da técnica de janela deslizante

quadrado da média móvel simples e quadrado do *jerk*. Essas métricas geram três grupos distintos de séries temporais. A Figura 7.5 ilustra o cálculo dessas métricas para cada uma das 34 janelas em um exemplo.

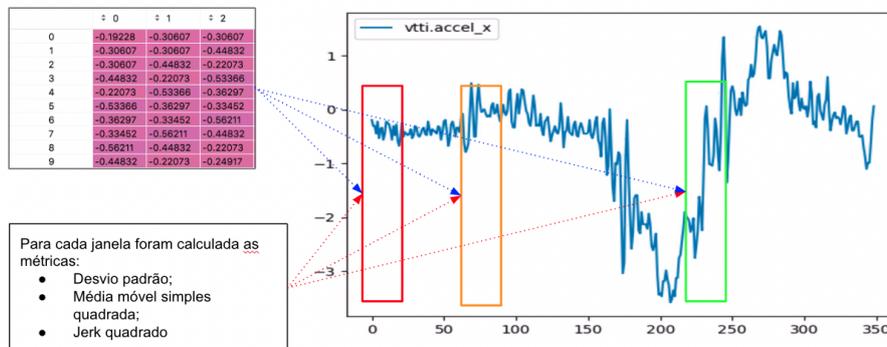


Figura 7.5: Cálculo das três métricas nas janelas

Definição dos Limites

Até este ponto, dispomos de três grupos de séries temporais, cada um representando uma métrica. A partir dessas séries, estabelecemos um limite para cada métrica e aplicamos esse limite às séries correspondentes. O objetivo é capturar os dados relevantes das séries, ou seja, aqueles que ultrapassam os limites estabelecidos pelas métricas. Para definir cada limite (um para cada métrica), utilizamos o método *InterQuartile Range*¹. Em cada uma das novas séries, os dados que são considerados *outliers* passam a ser considerados limites daquela série, ou seja, o limite daquela métrica. A Figura 7.6 apresenta uma visão das novas séries por métrica, já com os limites calculados.

¹Esse método define um intervalo e considera qualquer ponto de dados fora desse intervalo como um *outlier*.

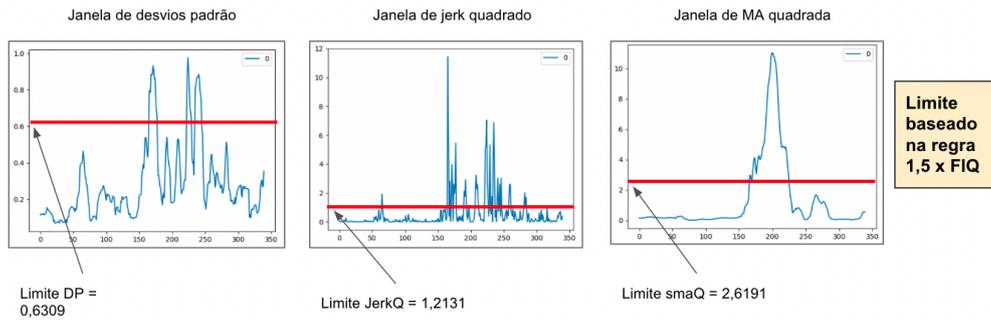


Figura 7.6: Cálculo dos limites para cada métrica

Seleção dos Dados Relevantes

Com os limites definidos, aplicamos esses limites aos três grupos de séries temporais para selecionar apenas os dados que os ultrapassam. É importante observar que cada série teve seu próprio limite correspondente aplicado, ou seja, a série derivada do desvio padrão teve o limite dessa métrica aplicado. O resultado consiste em um conjunto de dados, correspondente a cada série, que ultrapassaram os respectivos limites. Dentre esse conjunto resultante, estamos interessados nos pontos de tempo (*timepoint*) que são comuns a cada série. Portanto, coletamos os pontos de tempo das séries de desvio padrão, *jerk* e média móvel simples, onde os dados ultrapassaram seus respectivos limites. Esse processo é ilustrado na Figura 7.7.

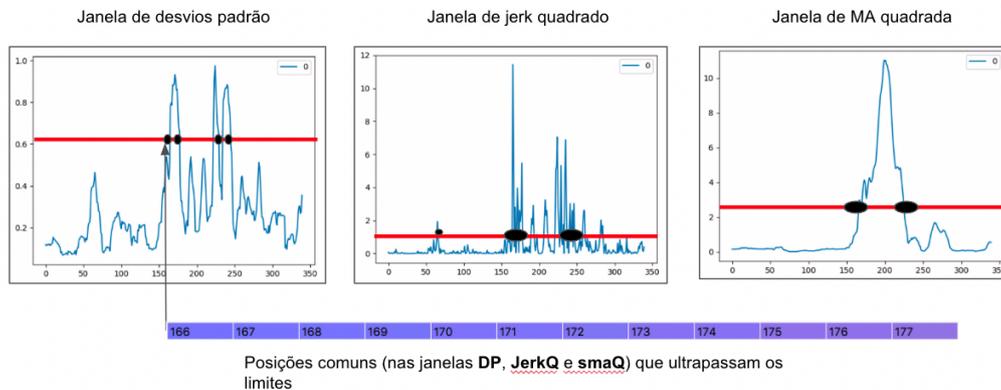


Figura 7.7: Seleção dos dados relevantes de cada série das métricas

Com o conjunto de pontos de tempo em mãos, onde os valores dos dados ultrapassaram os limites, os empregamos para selecionar apenas esses segmentos na série original, antes da aplicação da janela deslizante. Esse segmento de dados representa o evento em si. A técnica permite a seleção apenas dos dados que correspondem a distúrbios intensos na série temporal, extraíndo um segmento de dados que representa variações significativas nos valores, indicando prováveis eventos de *crash*, *near-crash* ou manobras abruptas. Limitamos o comprimento dos segmentos extraídos a 15s. Quando os segmentos foram menores, preenchemos os espaços com dados

anteriores ao início e posteriores ao fim do evento. Esse foi o cenário mais comum. Nos casos em que o segmento ultrapassou os 15s, cortamos o segmento no fim do evento, removendo pontos de tempo. A Figura 7.8 ilustra o funcionamento geral do algoritmo “Seleção de Janelas”, onde a partir de um evento de *crash* com duração de 35s, extraímos um segmento de 15s que representa a parte relevante do evento. Ao final desta etapa, a quantidade de eventos foi reduzida para: 1.497 *crash*, 5.473 *near-crash* e 10.475 manobras abruptas (*others*).

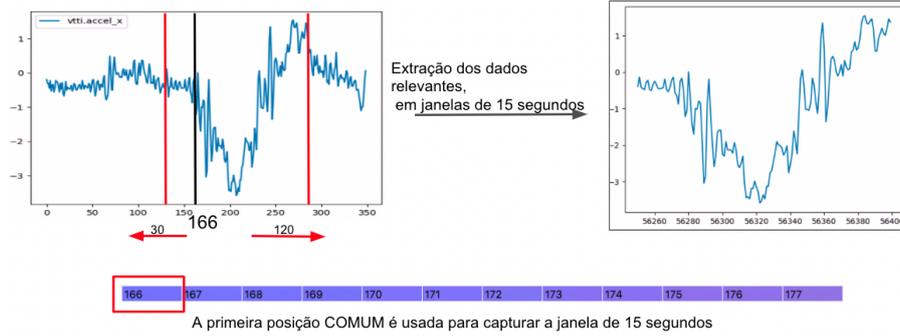


Figura 7.8: Visão geral do algoritmo “Seleção de Janelas”

7.4 Amostragem

Continuando o processo de preparação dos dados, empregamos uma técnica de seleção amostral no conjunto de dados. Esta etapa envolve a extração de uma amostra com o propósito de reduzir custos, seguida pela definição da seleção amostral. Optamos pela técnica de amostragem balanceada devido à considerável improbabilidade de eventos de *crash*, e à tendência dos classificadores de apresentarem alta precisão na classe majoritária, porém deficiências na classe minoritária ELAMRANI ABOU ELASSAD *et al.* (2020a); ZHU *et al.* (2018). A realização desse processo de amostragem resultou em um conjunto de dados contendo 107 *crash*, *near-crash* ou *others*, que foi usado nas etapas subsequentes do experimento.

Após configurarmos o ambiente, iniciamos o ciclo de execução do experimento, abrangendo as fases de treinamento, avaliação e construção do modelo, conforme previamente delineado na Seção 5.

Tabela 7.1: WALE-a versions

WALE-a	Windows size	Wavelet list	Pooling type
version 1	90	'bior3.1', 'bior1.5', 'db11'	mean
version 2	90	'coif6', 'sym6', 'bior2.2', 'coif4', 'db11', 'sym5'	mean
version 3	90	'rbio4.4', 'sym5', 'bior3.1', 'coif6', 'sym8', 'coif7', 'bior2.2', 'db11', 'coif4', 'db1', 'rbio3.5', 'bior1.5'	max

7.5 Ajuste de Algoritmos

Antes de avaliar os algoritmos, conduzimos centenas de experimentos para aprimorar os parâmetros envolvidos. No caso dos algoritmos tradicionais e específicos, utilizamos tanto os princípios da literatura quanto a abordagem de tentativa e erro. Quanto ao nosso algoritmo proposto, denominado **WALE-a**, a abordagem foi semelhante, porém com particularidades na configuração dos parâmetros.

Três parâmetros fundamentais foram otimizados no nosso algoritmo **WALE-a**: **tamanho da janela**, **lista de wavelets** e **tipo de pooling**. O **tamanho da janela** estabelece o comprimento da janela da série temporal, variando entre 0 e 15s. A **lista de wavelets** se refere às 146 *wavelets* discretas disponíveis, como listado no site PyWavelets². Quanto ao *pooling*, exploramos cinco estratégias distintas. Com tantos parâmetros e combinações possíveis, realizar testes exaustivos não era viável, dada a enormidade de bilhões de combinações possíveis. No capítulo 6 apresentamos os detalhes do nosso algoritmo.

Para restringir as possibilidades, empregamos técnicas de amostragem que permitiram obter conclusões substanciais e estabelecer critérios. Após conduzirmos mais de quinhentos testes, selecionamos **tamanhos de janela** entre 80 e 100. Em relação à **lista de wavelets**, determinamos inicialmente o tamanho da lista. Optamos por três grupos, cada um contendo 3, 6 e 12 *wavelets*, respectivamente, com base em estudos realizados por SALLES *et al.* (2019); TOO *et al.* (2019a); AL OMARI *et al.* (2014); PHINYOMARK *et al.* (2012); TOO *et al.* (2019b). Posteriormente, conduzimos cerca de duzentos testes com cada família de *wavelets* (sete famílias de *wavelets*), selecionando as que apresentaram melhor desempenho em cada uma para formar os três grupos. Por fim, identificamos as estratégias de *pooling* de **média** e **máxima** como as mais eficazes.

Após completar a fase de ajuste do nosso algoritmo, obtivemos as três versões exibidas na Tabela 7.1.

7.6 Aprendizado

Prosseguimos com a execução dos algoritmos de aprendizado de máquina, conforme listados na Tabela 7.2. Essa seleção foi fundamentada na análise dos resultados mais promissores apresentados em diversos estudos como XIONG *et al.* (2019); HALIM *et al.* (2016a); BA *et al.* (2017); ELAMRANI ABOU ELASSAD *et al.* (2020a); QU *et al.* (2017); GUTIERREZ-OSORIO e PEDRAZA (2020); CASTRO e KIM (2016); KLUGER *et al.* (2016); ZHU *et al.* (2018); BAGNALL *et al.* (2017); WANG *et al.* (2017); LONGHI e NANNI (2019); AKIN e AKBA (2010); REN *et al.* (2018); RUIZ

²API Wavelet: <https://pywavelets.readthedocs.io/>

Tabela 7.2: Algorithms list.

Class	Algorithm
Traditional	Naive Bayes
	k-nearest neighbors (K-nn)
	Random forest
	Support vector machine (SVM)
Specific	K-Neighbors Time Series (1-nn DTW)
	Time series forest (TSF)
	Rocket classifier
	MiniRocket classifier
	MultiRocket classifier
	Hierarchical Vote Collective of Transformation (Hive-COTE v2)
	Canonical Interval Forest (CIF)
Our proposal	WALE-a (version 1)
	WALE-a (version 2)
	WALE-a (version 3)

et al. (2021).

Categorizamos os algoritmos em três grupos distintos: **tradicionais**, **específicos** e **nossa proposta**. Antes de iniciar os experimentos, organizamos cinco conjuntos de dados, cada um contendo uma única *feature*, como já detalhado na seção 5.3. Para garantir a solidez dos resultados, submetemos o nosso experimento a processos de ajuste e validação cruzada. Avaliamos os quatorze algoritmos (quatro tradicionais, sete específicos e três versões do **WALE-a**) frente aos cinco conjuntos de dados, utilizando a técnica de validação cruzada *10-fold* e repetindo o procedimento cinco vezes para cada caso. Ao final do experimento, após a realização de todos os processos de treinamento, avaliação, ajustes e testes, acumulamos mais de 5.000 iterações.

Capítulo 8

Análise dos Resultados

Neste capítulo, apresentamos os resultados derivados dos nossos experimentos de avaliação. Realizamos uma análise do desempenho dos classificadores selecionados (conforme detalhado na Tabela 7.2), utilizando diversas métricas, tais como: acurácia, precisão, *recall*, especificidade, taxa de falso positivo (FPR) e tempo de processamento dos algoritmos. Essas métricas foram escolhidas com base em estudos anteriores, como BAGNALL *et al.* (2017, 2020); DEMPSTER *et al.* (2020); MIDDLEHURST *et al.* (2021).

A métrica primária para avaliar o desempenho do nosso estudo é a **acurácia**. Ela mede a proporção de previsões corretas, feitas por um modelo em relação ao número total de previsões de todas as classes. A acurácia oferece uma visão geral do desempenho do nosso modelo e é frequentemente considerada uma métrica universal para selecionar classificadores eficazes, conforme observado por BA *et al.* (2017); ELAMRANI ABOU ELASSAD *et al.* (2020a). No entanto, a acurácia por si só pode ser enganosa quando a distribuição de classes é desequilibrada, como no caso dos sistemas de previsão de acidentes, onde os eventos de não acidentes são muito mais numerosos que os eventos de acidentes. Em nosso trabalho, seguimos a abordagem de PAEFGEN *et al.* (2013), considerando a classe minoritária (*crash*) como a classe positiva, enquanto as manobras abruptas e os *near-crash* foram considerados como classe negativa.

Além da acurácia, empregamos outras métricas para aprimorar nossa análise dos modelos. Incluímos **precisão**, *recall* (sensibilidade), **especificidade**, **taxa de falso positivo** (FPR) e o tempo de processamento para obter uma compreensão mais abrangente do desempenho dos nossos classificadores, e identificar áreas de melhoria. O *recall*, também conhecido como taxa de verdadeiro positivo (TPR) ou sensibilidade, permite avaliar a capacidade dos modelos de identificar com exatidão eventos de acidentes (*crash*). Por sua vez, a precisão mede a proporção dos eventos *crash* previstos corretamente em relação a todos os eventos previstos, fornecendo uma medida da qualidade ao modelo em prever acidentes. A especificidade reflete a

taxa de verdadeiros negativos (*non-crash*) corretamente identificados pelo modelo, oferecendo informações sobre a capacidade dos classificadores em relação a eventos de *near-crash* e manobras abruptas. Além disso, a taxa de falso positivo ajuda a entender a proporção de erro dos classificadores em relação aos eventos de *acidentes*, fornecendo *insights* sobre as fragilidades do modelo. Por fim, também incorporamos o tempo de processamento dos classificadores, que engloba tempo de treinamento e teste. Embora menos comum, essa métrica é relevante para sistemas de previsão de acidentes, que não apenas necessitam de acurácia, mas também de rapidez, sendo considerada em algumas das nossas análises.

O centro das análises deste capítulo está na comparação dos catorze classificadores (conforme a Tabela 7.2), em relação a cinco conjuntos de dados univariados. Cada conjunto contém 107 eventos críticos, cada um representado por uma única *feature* extraída dos sensores. Inicialmente, examinamos a performance geral dos algoritmos, com foco na avaliação dos algoritmos tradicionais, específicos e daqueles constantes em nossa própria proposta.

Na Figura 8.1, apresentamos um panorama dessa análise, revelando claramente a superioridade da nossa abordagem. Observa-se que nosso algoritmo obteve uma dispersão dos dados e valores de acurácia mais consistentes. Por exemplo, no intervalo entre os Quartis 1 e 3, que engloba 50% dos dados, nossa proposta alcançou acurácia variando de 75% a 90%. Em contraste, os algoritmos tradicionais exibiram dispersão maior, com acurácia variando de 30% a 85%, o que é esperado dada sua ampla aplicabilidade sujeitos a diferentes configurações. Por outro lado, os algoritmos específicos apresentaram um desempenho muito bom, alinhados com achados de estudos como BAGNALL *et al.* (2017); RUIZ *et al.* (2021); DEMPSTER *et al.* (2020); TAN *et al.* (2022); DEMPSTER *et al.* (2021).

Prosseguindo na análise, examinamos os classificadores em relação aos diferentes conjuntos de dados, considerando a média das acurácias. A Figura 8.2 apresenta a comparação da acurácia dos classificadores frente aos cinco conjuntos de dados, destacando a superioridade do nosso algoritmo **WALE-a** em quase todos os casos, à exceção de *gyro_z*.

Mais uma vez, observa-se a estabilidade nas variações das acurácias, juntamente com valores consistentes em todos os conjuntos. Os conjuntos de dados mais favoráveis para o algoritmo **WALE-a** foram *accel_x* e *accel_y*, também observado nos algoritmos de classes específicas. Tal resultado decorre da origem desses conjuntos serem o acelerômetro, que se mostra mais eficaz na captura de eventos bruscos de direção, como constatado em estudos prévios HALIM *et al.* (2016a); ELAMRANI ABOU ELASSAD *et al.* (2020a); KLUGER *et al.* (2016); LONGHI e NANNI (2019). Nosso algoritmo superou os específicos em *accel_x* e *accel_y*, alcançando acurácias entre 89% e 88%, em comparação com 86% e 85% dos classificadores específicos.

Por outro lado, os classificadores tradicionais mais uma vez apresentaram as piores e mais variadas performances. Um dado interessante revela que, a melhor acurácia foi de apenas 68% com o conjunto *accel_x*. Já o segundo melhor desempenho foi obtido com o conjunto *speed_gps*, atingindo 67%, diferente da nossa solução e dos algoritmos específicos.

Outro aspecto notável é a análise da dispersão dos dados, não somente em termos das variações mínimas e máximas, mas também considerando o desvio padrão. A Figura 8.2 proporciona essa análise, evidenciando que nossa abordagem **WALE-a** e os classificadores específicos, foram mais consistentes em comparação com os tradicionais. Nas três versões do nosso algoritmo, a acurácia média mostrou um desempenho excelente e consistente, com valores de desvio padrão variando de 0.05 a 0.07 em todos os conjuntos de dados.

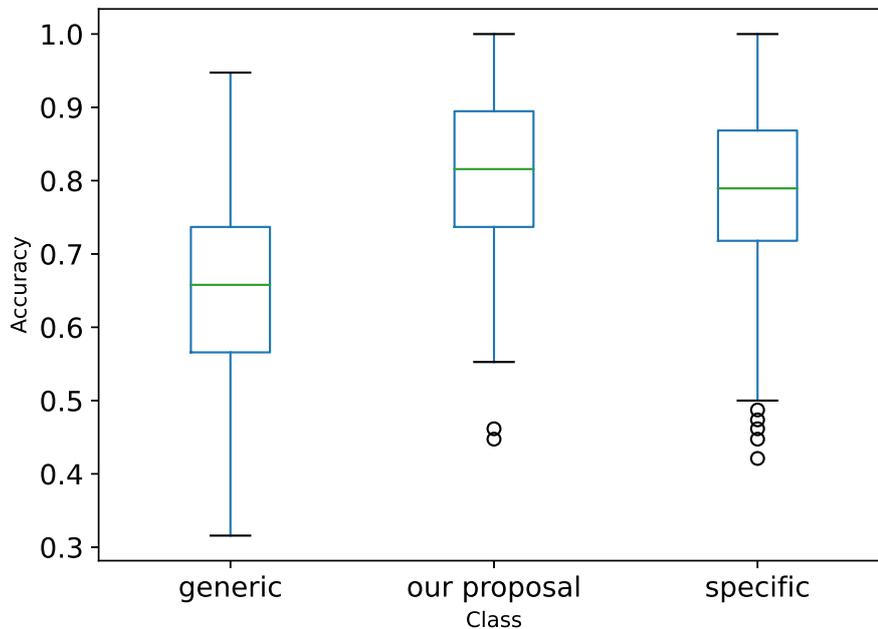


Figura 8.1: Desempenho da acurácia entre classificadores

Na terceira análise, nosso foco é identificar os algoritmos mais eficazes, os quais devem maximizar algumas métricas e minimizar outras. Para esta finalidade, adotamos a técnica conhecida como *Weighted Sum Model* (WSM), proposta por TRIANTAPHYLLOU (2000). Essa abordagem é aplicada em tomadas de decisão, quando diversas alternativas são avaliadas com base em múltiplos critérios. No contexto do nosso trabalho, empregamos o WSM para determinar os melhores algoritmos segundo critérios como alta acurácia, precisão e *recall* elevadas, bem como baixa taxa de falsos positivos.

Os resultados são apresentados na Tabela 8.1, onde se destaca o classificador es-

pecífico **CIF** com o melhor desempenho. Contudo, é importante notar que em nossa proposta, o algoritmo **WALE-a** apresentou uma diferença mínima de desempenho em relação ao **CIF**. Além disso, o **WALE-a** posiciona-se em seis das dez primeiras colocações do *ranking*, evidenciando sua robustez e eficácia. É válido destacar que tanto o **WALE-a** quanto o **CIF** demonstraram excelente capacidade de previsão (**precisão**), assertividade na identificação de acidentes (**recall**) e baixa taxa de erro (**FPR**). Especificamente, o **WALE-a** atingiu a melhor taxa de acurácia, com 90,03%. Além disso, ao analisarmos os conjuntos de dados no *ranking*, é evidente que os 10 melhores algoritmos utilizaram os conjuntos *accel_x* e *accel_y*, corroborando com as conclusões anteriores destacadas na Figura 8.2.

Mais uma vez, constatamos o excelente desempenho dos algoritmos específicos. Além do **CIF**, podemos citar o **MultiRocket** TAN *et al.* (2022) e o **HiveCote2** MIDDLEHURST *et al.* (2021), que alcançaram médias de acurácia superiores a 89% e, da mesma forma, apresentaram posições bem colocadas no ranking. Em contrapartida, os algoritmos tradicionais mostraram melhor desempenho com o **K-nn**, obtendo resultados satisfatórios com pontuações acima de 75%, porém sem conseguir figurar na lista dos dez primeiros classificados.

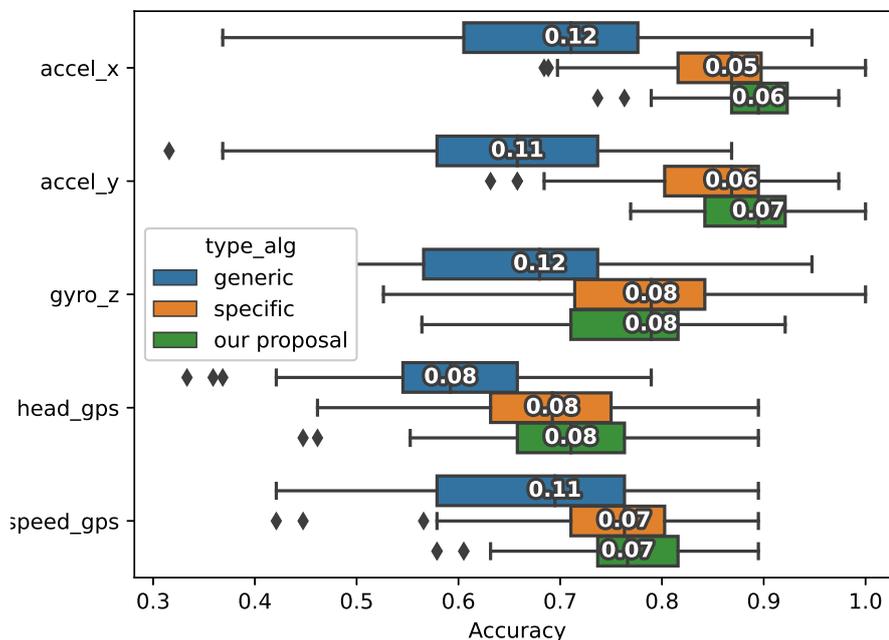


Figura 8.2: Desempenho dos classificadores entre conjuntos de dados

Em nossa próxima análise, trataremos da eficácia de um sistema de previsão de acidentes além da necessidade da pontualidade e precisão em suas informações. Em geral, sistemas desse tipo precisam ser altamente precisos para impossibilitar os falsos alarmes. Contudo, é igualmente desejável que o modelo possua a habilidade

Tabela 8.1: The top ten specific classifiers

Classifiers - Datasets	Accuracy	Precision	Recall	FPR
CIF - accel_y	89.51%	89.38%	77.61%	4.43%
WALE_v1 - accel_x	89.59%	88.45%	77.47%	4.45%
WALE_v3 - accel_x	88.88%	86.39%	75.18%	4.77%
MultiRocket - accel_y	89.11%	87.03%	77.51%	4.98%
WALE_v3 - accel_y	87.33%	87.33%	72.32%	4.83%
WALE_v1 - accel_y	90.03%	87.73%	81.38%	5.54%
HiveCote2 - accel_y	89.33%	86.54%	80.22%	6.00%
WALE_v2 - accel_x	89.05%	85.29%	78.77%	6.23%
MultiRocket - accel_x	89.76%	85.13%	82.09%	6.58%
WALE_v2 - accel_y	88.72%	84.77%	80.47%	6.88%

de discriminar eventos acidentes de não acidentes. Para avaliar essa capacidade, as métricas *recall* e *especificidade* são extremamente relevantes, pois medem a qualidade da assertividade dos modelos.

Na Figura 8.3, apresentamos o desempenho de vários classificadores sob essa perspectiva. Na parte superior esquerda, exibimos os melhores valores de *recall*, indicando a taxa de acertos para eventos de acidentes (*crash*). Já na parte inferior esquerda, exibimos as melhores *especificidades*, que refletem a taxa de acertos para eventos não acidentes, abrangendo tanto os *near-crash* quanto as manobras abruptas.

Mais uma vez, o algoritmo **WALE-a** demonstrou um desempenho notável, conforme evidenciado na Figura 8.3. Notamos que o **WALE-a_v1**, com o conjunto de dados *accel_x*, apresentou excelente desempenho em eventos não acidentes, enquanto no conjunto *accel_y*, o **WALE-a_v1** mostrou maior sensibilidade em eventos de acidentes. Essa versatilidade em diferentes cenários reforça a consistência de nossa abordagem. Vale mencionar que os algoritmos **Multirocket** e **HiveCote2**, também alcançaram altas taxas de *recall* e boas *especificidades*. Além disso, ressaltamos novamente que os conjuntos de dados *accel_x* e *accel_y* se destacaram como eficazes na classificação de eventos críticos, reafirmando as análises anteriores.

Em nossa última análise, concentramo-nos no tempo de processamento (treinamento e teste) dos algoritmos, uma consideração importante dado o crescimento contínuo dos dados de eventos críticos. À medida que os sistemas de predição se expandem, algoritmos eficientes e ágeis são essenciais para manter a pontualidade e precisão. Embora os algoritmos específicos ofereçam um desempenho superior em termos de precisão e acurácia, em comparação aos classificadores tradicionais, frequentemente apresentam desvantagem por serem computacionalmente complexos. Essa complexidade pode impactar negativamente a velocidade tanto do treinamento quanto do teste, fato observado por BAGNALL *et al.* (2017); ISMAIL FAWAZ *et al.* (2019). A figura 8.4 ilustra essa relação, mostrando como a acurácia de um classifi-

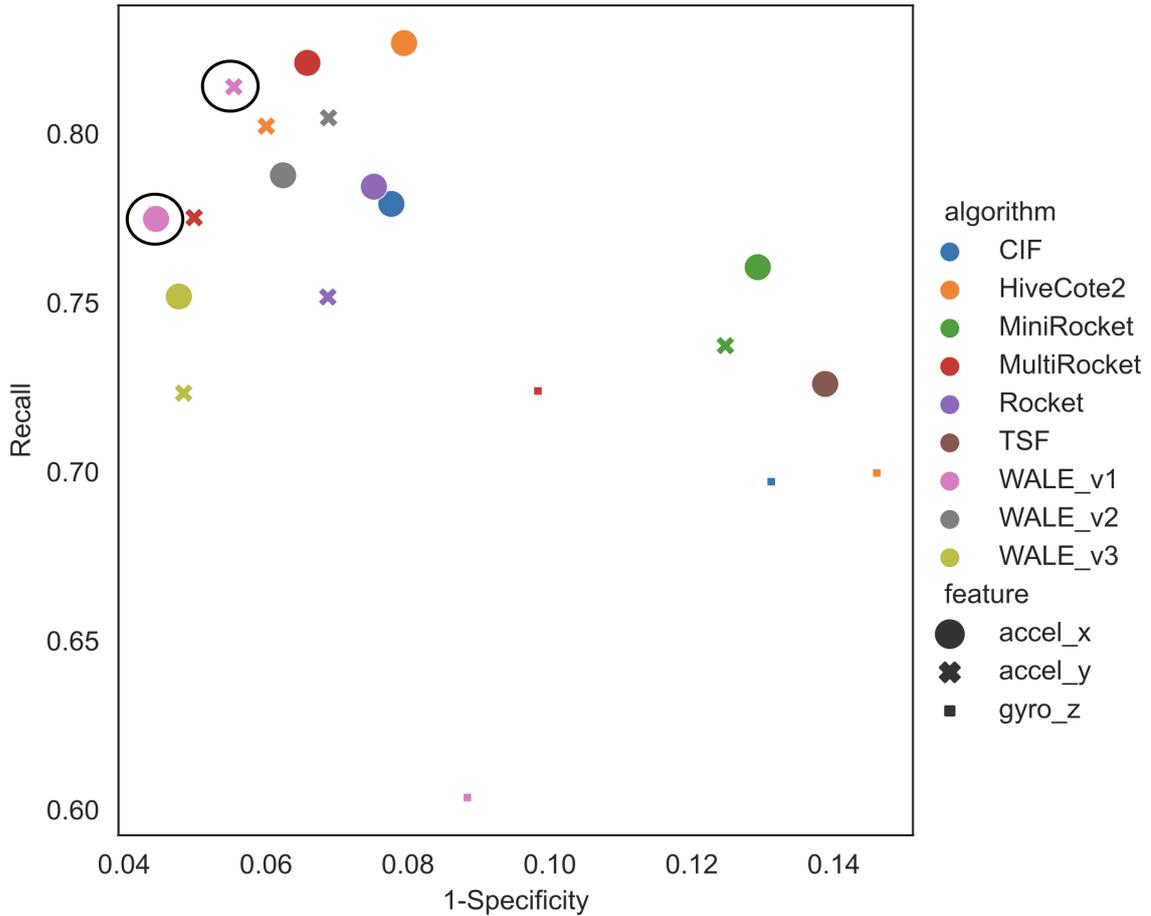


Figura 8.3: Relação entre a *recall* e especificidade dos classificadores

cadador se correlaciona com seu tempo de processamento, onde alguns classificadores podem demandar tempo superior a 500 segundos.

Em resumo, os classificadores específicos geralmente têm desempenho inferior em termos de tempo, mas exibem boa acurácia. Por outro lado, os classificadores tradicionais são notadamente mais rápidos, porém com acurácia menor. No entanto, existem algumas exceções, destacadas na Figura 8.4: dois classificadores tradicionais, **K-nn** e **RandomForest**, apresentam uma razoável acurácia média de 70% e 72%, com tempos de processamento inferiores a 1 segundo (indicados como "a" e "b"); dois classificadores específicos, **RandomForest** e **MiniRocket**, demonstram boa acurácia de 76% e 78% e, respectivamente, tempos de processamento de 1 e 7 segundos (indicados como "c" e "d").

Nossos classificadores **WALE-a** (versões 1, 2 e 3) se destacaram mais uma vez, mostrando um desempenho excepcional tanto em termos de acurácia quanto no tempo de processamento. Conforme evidenciado na Figura 8.4, todas as versões do **WALE-a** alcançaram uma acurácia média acima de 78%, mantendo tempos de processamento abaixo de 20 milissegundos (rótulos "e", "f" e "g"). Esses resultados atestam a qualidade de nossa proposta, que não apenas exhibe alta acurácia, estabi-

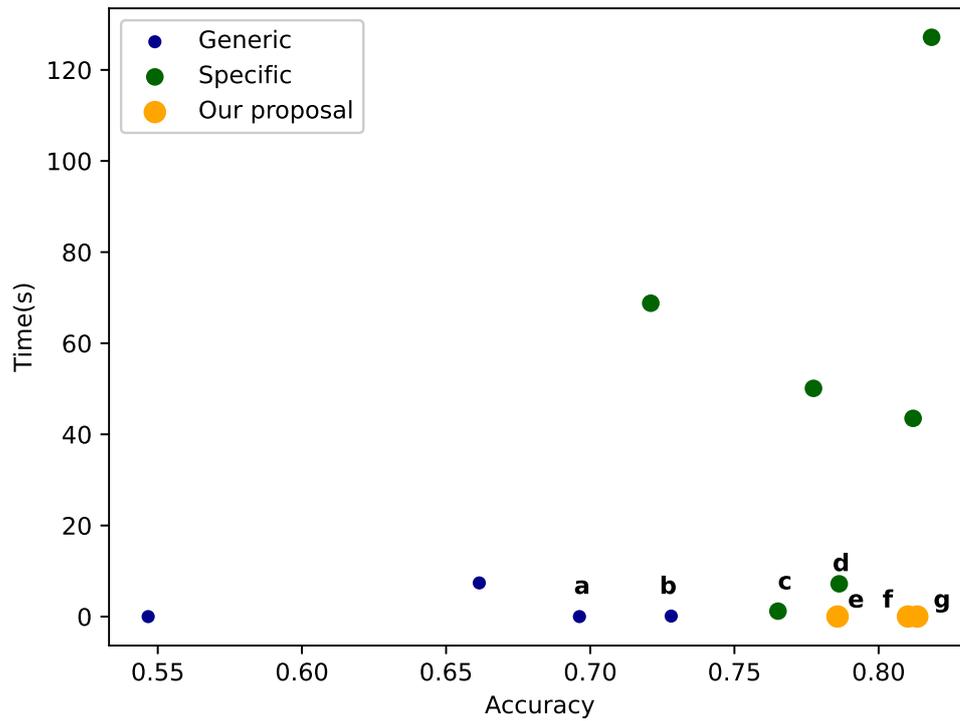


Figura 8.4: Relação entre a acurácia e o tempo de processamento dos classificadores

lidade e habilidade de distinguir eventos críticos, mas também é caracterizada por uma notável rapidez de execução.

Capítulo 9

Conclusão

No decorrer de nossa pesquisa, desenvolvemos um algoritmo inovador com o propósito de fortalecer os Sistemas de Previsão de Acidentes (CPS) e, conseqüentemente, aumentar a segurança viária por meio da aplicação de técnicas de inteligência artificial. Denominado de **WALE-a**, nosso modelo adota uma abordagem de decomposição *wavelet* multinível, que se mostrou promissora na análise dos dados dos sensores veiculares para identificação de eventos críticos. Com base no objetivo central de aprimorar a classificação de incidentes rodoviários, avaliamos extensivamente a eficácia do algoritmo proposto, por meio da comparação com outros algoritmos de aprendizado de máquina.

Para a avaliação do desempenho dos algoritmos, aplicamos uma variedade de métricas que abrangem desde a acurácia até a taxa de falsos positivos, bem como o tempo de processamento. Essas métricas foram selecionadas com base em estudos prévios, tais como BAGNALL *et al.* (2017, 2020); DEMPSTER *et al.* (2020); MIDDLEHURST *et al.* (2021), a fim de garantir uma análise abrangente e precisa. Além disso, optamos por empregar cinco conjuntos de dados univariados, cada um representando 107 eventos críticos. Adotamos também a validação cruzada *k*-fold, para garantir uma escolha eficiente do modelo, e uma comparação justa do desempenho dos algoritmos.

Nosso algoritmo **WALE-a** foi estruturado com base em princípios inspirados na arquitetura de redes neurais convolucionais (CNN). Com quatro camadas distintas, o modelo foi projetado para executar a classificação de eventos críticos de direção, tais como *crash*, *near-crash* e manobras abruptas. As três primeiras camadas englobam o pré-processamento dos dados, onde ocorrem transformações utilizando *wavelets* e métricas estatísticas. A última camada é responsável pela classificação propriamente dita. Mediante extensas avaliações, ficou claro que nosso algoritmo **WALE-a** superou os classificadores tradicionais e específicos em termos de resultados.

Os resultados obtidos revelaram a notável eficácia do algoritmo **WALE-a** em variados conjuntos de dados, destacando-se, nos cenários de *accel_x* e *accel_y*, que

se originam do acelerômetro. Com acurácias superiores a 88%, e alcançando mais de 90% no conjunto de dados *accel_y*, o **WALE-a** demonstrou sua capacidade de distinguir e classificar eventos críticos de forma precisa e consistente, superando outros classificadores. A força desse algoritmo foi reiterada ao assegurar seis das dez primeiras posições no ranking de classificação. Além disso, foi comprovado que o algoritmo **WALE-a** destaca-se ao equilibrar acurácia e tempo de processamento. Ao alcançar acurácia acima de 80% e manter tempos de processamento inferiores a 20 milissegundos nas três versões do algoritmo. Neste sentido, fica evidente o seu potencial em contribuir para os CPS, reduzindo acidentes rodoviários e promovendo um ambiente de direção mais seguro. Por fim, nossa pesquisa evidenciou o papel crucial da inteligência artificial na melhoria da segurança viária e na prevenção de acidentes.

Referências Bibliográficas

- MESEGUER, J. E., CALAFATE, C. T., CANO, J. C., et al. “DrivingStyles: A smartphone application to assess driver behavior”. In: *2013 IEEE Symposium on Computers and Communications (ISCC)*, pp. 000535–000540, jul. 2013. doi: 10.1109/ISCC.2013.6755001.
- CASTIGNANI, G., DERRMANN, T., FRANK, R., et al. “Driver Behavior Profiling Using Smartphones: A Low-Cost Platform for Driver Monitoring”, *IEEE Intelligent Transportation Systems Magazine*, v. 7, n. 1, pp. 91–102, 2015. ISSN: 1939-1390. doi: 10.1109/MITS.2014.2328673.
- AL-DOORI, M., PALUSZCZYSZYN, D., ELIZONDO, D., et al. “Range extended for electric vehicle based on driver behaviour recognition”. In: *5th IET Hybrid and Electric Vehicles Conference (HEVC 2014)*, pp. 1–7, nov. 2014. doi: 10.1049/cp.2014.0944.
- HAN, J., PEI, J., KAMBER, M. *Data Mining: Concepts and Techniques*. New York, Elsevier, jun. 2011. ISBN: 978-0-12-381480-7. Google-Books-ID: pQws07tdpjoC.
- BANERJEE, D., BANERJEE, N., CHAKRABORTY, D., et al. “How’s My Driving? A Spatio-Semantic Analysis of Driving Behavior with Smartphone Sensors”. In: Stojmenovic, I., Cheng, Z., Guo, S. (Eds.), *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pp. 653–666. Springer International Publishing, 2014. ISBN: 978-3-319-11569-6.
- KALSOOM, R., HALIM, Z. “Clustering the driving features based on data streams”. In: *INMIC*, pp. 89–94, dez. 2013. doi: 10.1109/INMIC.2013.6731330.
- SAIPRASERT, C., PHOLPRASIT, T., THAJCHAYAPONG, S. “Detection of Driving Events using Sensory Data on Smartphone”, *International Journal of Intelligent Transportation Systems Research*, v. 15, n. 1, pp. 17–28, jan. 2017. ISSN: 1868-8659. doi: 10.1007/s13177-015-0116-5. Disponível em: <<https://doi.org/10.1007/s13177-015-0116-5>>.

- WU, M., ZHANG, S., DONG, Y. “A Novel Model-Based Driving Behavior Recognition System Using Motion Sensors”, *Sensors (Basel, Switzerland)*, v. 16, n. 10, out. 2016. ISSN: 1424-8220. doi: 10.3390/s16101746.
- MARTINEZ, C. M., HEUCKE, M., WANG, F., et al. “Driving Style Recognition for Intelligent Vehicle Control and Advanced Driver Assistance: A Survey”, *IEEE Transactions on Intelligent Transportation Systems*, v. 19, n. 3, pp. 666–676, mar. 2018. ISSN: 1524-9050. doi: 10.1109/TITS.2017.2706978.
- FAZEEN, M., GOZICK, B., DANTU, R., et al. “Safe Driving Using Mobile Phones”, *IEEE Transactions on Intelligent Transportation Systems*, v. 13, n. 3, pp. 1462–1468, set. 2012. ISSN: 1524-9050. doi: 10.1109/TITS.2012.2187640.
- YADAV, P., JUNG, S., SINGH, D. “Machine learning based real-time vehicle data analysis for safe driving modeling”. In: *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, SAC '19*, pp. 1355–1358, Limassol, Cyprus, abr. 2019. Association for Computing Machinery. ISBN: 978-1-4503-5933-7. doi: 10.1145/3297280.3297584. Disponível em: <<https://doi.org/10.1145/3297280.3297584>>.
- WANG, J., XU, W., GONG, Y. “Real-time driving danger-level prediction”, *Engineering Applications of Artificial Intelligence*, v. 23, n. 8, pp. 1247–1254, dez. 2010. ISSN: 0952-1976. doi: 10.1016/j.engappai.2010.01.001. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0952197610000175>>.
- GERDES, A. “Automatic Maneuver Recognition in the Automobile: the Fusion of Uncertain Sensor Values using Bayesian Models”. In: *Proceedings of the 3rd International Workshop on Intelligent Transportation (WIT 2006)*, pp. 129–133, Hamburg, Deutschland, jan. 2006. Disponível em: <<https://elib.dlr.de/22833/>>.
- HÜLNHAGEN, T., DENGLER, I., TAMKE, A., et al. “Maneuver recognition using probabilistic finite-state machines and fuzzy logic”. In: *2010 IEEE Intelligent Vehicles Symposium*, pp. 65–70, jun. 2010. doi: 10.1109/IVS.2010.5548066.
- SATHYANARAYANA, A., SADJADI, S. O., HANSEN, J. H. L. “Leveraging sensor information from portable devices towards automatic driving maneuver recognition”. In: *2012 15th International IEEE Conference on Intelligent*

Transportation Systems, pp. 660–665, set. 2012. doi: 10.1109/ITSC.2012.6338717.

FUGIGLANDO, U., MASSARO, E., SANTI, P., et al. “Driving Behavior Analysis through CAN Bus Data in an Uncontrolled Environment”, *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2018. ISSN: 1524-9050. doi: 10.1109/TITS.2018.2836308.

JÚNIOR, J. F., CARVALHO, E., FERREIRA, B. V., et al. “Driver behavior profiling: An investigation with different smartphone sensors and machine learning”, *PLOS ONE*, v. 12, n. 4, pp. e0174959, abr. 2017. ISSN: 1932-6203. doi: 10.1371/journal.pone.0174959. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174959>>.

ORGANIZATION, W. H. *Road traffic injuries*. Relatório Técnico ISBN:9789241565684, WHO, 2018.

ENGELBRECHT, J., BOOYSEN, M. J., ROOYEN, G.-V., et al. “Performance Comparison of Dynamic Time Warping (DTW) and a Maximum Likelihood (ML) Classifier in Measuring Driver Behavior with Smartphones”. In: *2015 IEEE Symposium Series on Computational Intelligence*, pp. 427–433, dez. 2015. doi: 10.1109/SSCI.2015.70.

MOHAMAD, I., ALI, M. A. M., ISMAIL, M. “Abnormal driving detection using real time Global Positioning System data”. In: *Proceeding of the 2011 IEEE International Conference on Space Science and Communication (IconSpace)*, pp. 1–6, jul. 2011. doi: 10.1109/IConSpace.2011.6015840.

SAIPRASERT, C., THAJCHAYAPONG, S., PHOLPRASIT, T., et al. “Driver behaviour profiling using smartphone sensory data in a V2I environment”. In: *2014 International Conference on Connected Vehicles and Expo (IC-CVE)*, pp. 552–557, nov. 2014. doi: 10.1109/ICCVE.2014.7297609. ISSN: 2378-1297.

WAKITA, T., OZAWA, K., MIYAJIMA, C., et al. “Driver identification using driving behavior signals”. In: *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, pp. 396–401, set. 2005. doi: 10.1109/ITSC.2005.1520171.

DÖRR, D., GRABENGIESSER, D., GAUTERIN, F. “Online driving style recognition using fuzzy logic”. In: *17th International IEEE Conference on*

Intelligent Transportation Systems (ITSC), pp. 1021–1026, out. 2014. doi: 10.1109/ITSC.2014.6957822.

WAHAB, A., KEONG, T. C., ABUT, H., et al. “Driver Recognition System Using FNN and Statistical Methods”. In: Abut, H., Hansen, J. H. L., Takeda, K. (Eds.), *Advances for In-Vehicle and Mobile Systems: Challenges for International Standards*, Springer US, pp. 11–23, Boston, MA, 2007. ISBN: 978-0-387-45976-9. doi: 10.1007/978-0-387-45976-9_2. Disponível em: <https://doi.org/10.1007/978-0-387-45976-9_2>.

JOHNSON, D. A., TRIVEDI, M. M. “Driving style recognition using a smartphone as a sensor platform”. In: *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1609–1615, out. 2011. doi: 10.1109/ITSC.2011.6083078.

XIONG, X., WANG, M., CAI, Y., et al. “A forward collision avoidance algorithm based on driver braking behavior”, *Accident Analysis & Prevention*, v. 129, pp. 30–43, ago. 2019. ISSN: 0001-4575. doi: 10.1016/j.aap.2019.05.004. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0001457519300703>>.

HALIM, Z., KALSOOM, R., BASHIR, S., et al. “Artificial intelligence techniques for driving safety and vehicle crash prediction”, *Artificial Intelligence Review*, v. 46, n. 3, pp. 351–387, out. 2016a. ISSN: 1573-7462. doi: 10.1007/s10462-016-9467-9. Disponível em: <<https://doi.org/10.1007/s10462-016-9467-9>>.

BA, Y., ZHANG, W., WANG, Q., et al. “Crash prediction with behavioral and physiological features for advanced vehicle collision avoidance system”, *Transportation Research Part C: Emerging Technologies*, v. 74, pp. 22–33, jan. 2017. ISSN: 0968-090X. doi: 10.1016/j.trc.2016.11.009. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0968090X1630225X>>.

ELAMRANI ABOU ELASSAD, Z., MOUSANNIF, H., AL MOATASSIME, H. “A real-time crash prediction fusion framework: An imbalance-aware strategy for collision avoidance systems”, *Transportation Research Part C: Emerging Technologies*, v. 118, pp. 102708, set. 2020a. ISSN: 0968-090X. doi: 10.1016/j.trc.2020.102708. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0968090X20306239>>.

QU, X., WANG, W., WANG, W.-F., et al. “Real-time rear-end crash potential prediction on freeways”, *Journal of Central South University*,

v. 24, n. 11, pp. 2664–2673, nov. 2017. ISSN: 2227-5223. doi: 10.1007/s11771-017-3679-2. Disponível em: <<https://doi.org/10.1007/s11771-017-3679-2>>.

GUTIERREZ-OSORIO, C., PEDRAZA, C. “Modern data sources and techniques for analysis and forecast of road accidents: A review”, *Journal of Traffic and Transportation Engineering (English Edition)*, jul. 2020. ISSN: 2095-7564. doi: 10.1016/j.jtte.2020.05.002. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S209575642030101X>>.

CASTRO, Y., KIM, Y. J. “Data mining on road safety: factor assessment on vehicle accidents using classification models”, *International Journal of Crashworthiness*, v. 21, n. 2, pp. 104–111, mar. 2016. ISSN: 1358-8265. doi: 10.1080/13588265.2015.1122278. Disponível em: <<https://doi.org/10.1080/13588265.2015.1122278>>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13588265.2015.1122278>.

KLUGER, R., SMITH, B. L., PARK, H., et al. “Identification of safety-critical events using kinematic vehicle data and the discrete fourier transform”, *Accident Analysis & Prevention*, v. 96, pp. 162–168, nov. 2016. ISSN: 0001-4575. doi: 10.1016/j.aap.2016.08.006. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0001457516302846>>.

DAVOODI, M., SAVIO, R., SUDWEEKS, J. “Naturalistic Driving Study Data for Machine Learning Model Training”. 2020. Disponível em: <<https://doi.org/10.15787/VTT1/RIXJCO>>.

KITCHENHAM, B. *Procedures for performing systematic reviews*. Relatório Técnico 0400011T.1, Keele University, Keele, UK, jul. 2004. Disponível em: <<http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>>.

DYBA, T., DINGSOYR, T., HANSSEN, G. K. “Applying Systematic Reviews to Diverse Study Types: An Experience Report”. In: *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*, pp. 225–234, set. 2007. doi: 10.1109/ESEM.2007.59. ISSN: 1949-3789.

CENTRE FOR REVIEWS AND DISSEMINATION. *Systematic Reviews*. University of York, CRD, University of York, 2009. ISBN: 1-900640-47-3. Disponível em: <<https://www.york.ac.uk/crd/guidance/>>.

- KITCHENHAM, B., PEARL BRERETON, O., BUDGEN, D., et al. “Systematic literature reviews in software engineering – A systematic literature review”, *Information and Software Technology*, v. 51, n. 1, pp. 7–15, jan. 2009a. ISSN: 0950-5849. doi: 10.1016/j.infsof.2008.09.009. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950584908001390>>.
- KITCHENHAM, B., BRERETON, P., TURNER, M., et al. “The impact of limited search procedures for systematic literature reviews — A participant-observer case study”. In: *2009 3rd International Symposium on Empirical Software Engineering and Measurement*, pp. 336–345, out. 2009b. doi: 10.1109/ESEM.2009.5314238.
- TRAINA, A. J. M., TRAINA JR, C. “Como fazer pesquisa bibliográfica”, *SBC Horizontes*, v. 2, n. 2, pp. 30–35, 2009.
- JAAKKOLA, H., THALHEIM, B. “Architecture-Driven Modelling Methodologies”. In: *Proceedings of the 2011 Conference on Information Modelling and Knowledge Bases XXII*, pp. 97–116, Amsterdam, The Netherlands, The Netherlands, 2011. IOS Press. ISBN: 978-1-60750-689-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=1972754.1972761>>.
- RECHTIN, E. “The art of systems architecting”, *IEEE Spectrum*, v. 29, n. 10, pp. 66–69, out. 1992. ISSN: 0018-9235. doi: 10.1109/6.158642.
- EREN, H., MAKINIST, S., AKIN, E., et al. “Estimating driving behavior by a smartphone”. In: *2012 IEEE Intelligent Vehicles Symposium*, pp. 234–239, jun. 2012. doi: 10.1109/IVS.2012.6232298.
- ARAÚJO, R., IGREJA, A., CASTRO, R. D., et al. “Driving coach: A smartphone application to evaluate driving efficient patterns”. In: *2012 IEEE Intelligent Vehicles Symposium*, pp. 1005–1010, jun. 2012. doi: 10.1109/IVS.2012.6232304.
- VASCONCELOS, I., VASCONCELOS, R. O., OLIVIERI, B., et al. “Smartphone-based outlier detection: a complex event processing approach for driving behavior detection”, *Journal of Internet Services and Applications*, v. 8, n. 1, pp. 13, set. 2017. ISSN: 1869-0238. doi: 10.1186/s13174-017-0065-0. Disponível em: <<https://doi.org/10.1186/s13174-017-0065-0>>.
- CARMONA, J., GARCÍA, F., MARTÍN, D., et al. “Data Fusion for Driver Behaviour Analysis”, *Sensors*, v. 15, n. 10, pp. 25968–25991, out. 2015.

doi: 10.3390/s151025968. Disponível em: <<https://www.mdpi.com/1424-8220/15/10/25968>>.

YLIZALITURRI-SALCEDO, M. A., TENTORI, M., GARCIA-MACIAS, J. A. “Detecting Aggressive Driving Behavior with Participatory Sensing”. In: García-Chamizo, J. M., Fortino, G., Ochoa, S. F. (Eds.), *Ubiquitous Computing and Ambient Intelligence. Sensing, Processing, and Using Environmental Information*, Lecture Notes in Computer Science, pp. 249–261. Springer International Publishing, 2015. ISBN: 978-3-319-26401-1.

MADNI, H. A., ANWAR, Z., SHAH, M. A. “Data mining techniques and applications - A decade review”. In: *2017 23rd International Conference on Automation and Computing (ICAC)*, pp. 1–7, set. 2017. doi: 10.23919/ICoNAC.2017.8082090.

TAYLOR, P., GRIFFITHS, N., BHALERAO, A., et al. “Data Mining for Vehicle Telemetry”, *Applied Artificial Intelligence*, v. 30, n. 3, pp. 233–256, mar. 2016. ISSN: 0883-9514. doi: 10.1080/08839514.2016.1156954. Disponível em: <<https://doi.org/10.1080/08839514.2016.1156954>>.

MA, X., ANDREASSON, I. “Behavior Measurement, Analysis, and Regime Classification in Car Following”, *IEEE Transactions on Intelligent Transportation Systems*, v. 8, n. 1, pp. 144–156, mar. 2007. ISSN: 1524-9050. doi: 10.1109/TITS.2006.883111.

CHONG, L., ABBAS, M. M., MEDINA FLINTSCH, A., et al. “A rule-based neural network approach to model driver naturalistic behavior in traffic”, *Transportation Research Part C: Emerging Technologies*, v. 32, pp. 207–223, jul. 2013. ISSN: 0968-090X. doi: 10.1016/j.trc.2012.09.011. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0968090X12001210>>.

LIN, N., ZONG, C., TOMIZUKA, M., et al. “An Overview on Study of Identification of Driver Behavior Characteristics for Automotive Control”. 2014. Disponível em: <<https://www.hindawi.com/journals/mpe/2014/569109/>>.

CARPATOREA, I., NOWACZYK, S., RÖGNVALDSSON, T., et al. “Learning of Aggregate Features for Comparing Drivers Based on Naturalistic Data”. In: *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1067–1072, dez. 2016. doi: 10.1109/ICMLA.2016.0194.

- ALVAREZ, A. D., GARCIA, F. S., NARANJO, J. E., et al. “Modeling the Driving Behavior of Electric Vehicles Using Smartphones and Neural Networks”, *IEEE Intelligent Transportation Systems Magazine*, v. 6, n. 3, pp. 44–53, 2014. ISSN: 1939-1390. doi: 10.1109/MITS.2014.2322651.
- MEIRING, G. A. M., MYBURGH, H. C. “A Review of Intelligent Driving Style Analysis Systems and Related Artificial Intelligence Algorithms”, *Sensors (Basel, Switzerland)*, v. 15, n. 12, pp. 30653–30682, dez. 2015. ISSN: 1424-8220. doi: 10.3390/s151229822. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4721742/>>.
- CASTIGNANI, G., FRANK, R., ENGEL, T. “Driver behavior profiling using smartphones”. In: *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pp. 552–557, out. 2013. doi: 10.1109/ITSC.2013.6728289.
- BAGDADI, O. “Assessing safety critical braking events in naturalistic driving studies”, *Transportation Research Part F: Traffic Psychology and Behaviour*, v. 16, pp. 117–126, jan. 2013. ISSN: 1369-8478. doi: 10.1016/j.trf.2012.08.006. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1369847812000770>>.
- PHOLPRASIT, T., CHOOCHAIWATTANA, W., SAIPRASERT, C. “A comparison of driving behaviour prediction algorithm using multi-sensory data on a smartphone”. In: *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pp. 1–6, jun. 2015. doi: 10.1109/SNPD.2015.7176249.
- HALIM, Z., KALSOOM, R., BAIG, A. R. “Profiling drivers based on driver dependent vehicle driving features”, *Applied Intelligence*, v. 44, n. 3, pp. 645–664, abr. 2016b. ISSN: 0924-669X, 1573-7497. doi: 10.1007/s10489-015-0722-6. Disponível em: <<http://link.springer.com/10.1007/s10489-015-0722-6>>.
- GUOYING, C. “Study on Identification of Driver Steering Behavior Characteristics Based on Pattern Recognition”, *International Robotics & Automation Journal*, v. 1, out. 2016. doi: 10.15406/iratj.2016.01.00005.
- ZHU, M., LI, Y., WANG, Y. “Design and experiment verification of a novel analysis framework for recognition of driver injury patterns: From a multi-class classification perspective”, *Accident Analysis & Prevention*, v. 120,

pp. 152–164, nov. 2018. ISSN: 0001-4575. doi: 10.1016/j.aap.2018.08.011. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0001457518304524>>.

PAEFGEN, J., STAAKE, T., THIESSE, F. “Evaluation and aggregation of pay-as-you-drive insurance rate factors: A classification analysis approach”, *Decision Support Systems*, v. 56, pp. 192–201, dez. 2013. ISSN: 0167-9236. doi: 10.1016/j.dss.2013.06.001. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167923613001723>>.

BAGNALL, A., LINES, J., BOSTROM, A., et al. “The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances”, *Data Mining and Knowledge Discovery*, v. 31, n. 3, pp. 606–660, maio 2017. ISSN: 1573-756X. doi: 10.1007/s10618-016-0483-9. Disponível em: <<https://doi.org/10.1007/s10618-016-0483-9>>.

WAHLSTRÖM, J., SKOG, I., HÄNDEL, P. “Smartphone-Based Vehicle Telematics: A Ten-Year Anniversary”, *IEEE Transactions on Intelligent Transportation Systems*, v. 18, n. 10, pp. 2802–2825, out. 2017. ISSN: 1524-9050. doi: 10.1109/TITS.2017.2680468.

ISMAIL FAWAZ, H., FORESTIER, G., WEBER, J., et al. “Deep learning for time series classification: a review”, *Data Mining and Knowledge Discovery*, v. 33, n. 4, pp. 917–963, jul. 2019. ISSN: 1573-756X. doi: 10.1007/s10618-019-00619-1. Disponível em: <<https://doi.org/10.1007/s10618-019-00619-1>>.

LU, J., FILEV, D., PRAKAH-ASANTE, K., et al. “From vehicle stability control to intelligent personal minder: Real-time vehicle handling limit warning and driver style characterization”. In: *2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems*, pp. 43–50, mar. 2009. doi: 10.1109/CIVVS.2009.4938722.

CASTIGNANI, G., FRANK, R., ENGEL, T. “An evaluation study of driver profiling fuzzy algorithms using smartphones”. In: *2013 21st IEEE International Conference on Network Protocols (ICNP)*, pp. 1–6, out. 2014. ISBN: 978-1-4799-1270-4. doi: 10.1109/ICNP.2013.6733681. Disponível em: <<doi.ieeecomputersociety.org/10.1109/ICNP.2013.6733681>>.

MITROVIC, D. “Reliable method for driving events recognition”, *IEEE Transactions on Intelligent Transportation Systems*, v. 6, n. 2, pp. 198–205, jun. 2005. ISSN: 1524-9050. doi: 10.1109/TITS.2005.848367.

- LÄNGKVIST, M., KARLSSON, L., LOUTFI, A. “A review of unsupervised feature learning and deep learning for time-series modeling”, *Pattern Recognition Letters*, v. 42, pp. 11–24, jun. 2014. ISSN: 0167-8655. doi: 10.1016/j.patrec.2014.01.008. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865514000221>>.
- YANG ZHENG, WANG, J., XIAOFEI LI, et al. “Driving risk assessment using cluster analysis based on naturalistic driving data”. In: *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 2584–2589, out. 2014. doi: 10.1109/ITSC.2014.6958104. ISSN: 2153-0017.
- HU, B., CHEN, Y., KEOGH, E. “Time Series Classification under More Realistic Assumptions”. In: *Proceedings of the 2013 SIAM International Conference on Data Mining (SDM)*, Proceedings, Society for Industrial and Applied Mathematics, pp. 578–586, USA, maio 2013. ISBN: 978-1-61197-262-7. doi: 10.1137/1.9781611972832.64. Disponível em: <<https://epubs.siam.org/doi/abs/10.1137/1.9781611972832.64>>.
- JUN, J., OGLE, J., GUENSLER, R. “Relationships between Crash Involvement and Temporal-Spatial Driving Behavior Activity Patterns: Use of Data for Vehicles with Global Positioning Systems”, *Transportation Research Record*, v. 2019, n. 1, pp. 246–255, jan. 2007. ISSN: 0361-1981. doi: 10.3141/2019-29. Disponível em: <<https://doi.org/10.3141/2019-29>>. Publisher: SAGE Publications Inc.
- WANG, Y., XU, W., ZHANG, Y., et al. “Machine Learning Methods for Driving Risk Prediction”. In: *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Emergency Management using, EM-GIS’17*, pp. 1–6, Redondo Beach, CA, USA, nov. 2017. Association for Computing Machinery. ISBN: 978-1-4503-5493-6. doi: 10.1145/3152465.3152476. Disponível em: <<https://doi.org/10.1145/3152465.3152476>>.
- LONGHI, L., NANNI, M. “Car telematics big data analytics for insurance and innovative mobility services”, *Journal of Ambient Intelligence and Humanized Computing*, dez. 2019. ISSN: 1868-5145. doi: 10.1007/s12652-019-01632-4. Disponível em: <<https://doi.org/10.1007/s12652-019-01632-4>>.
- PAEFGEN, J., KEHR, F., ZHAI, Y., et al. “Driving Behavior Analysis with Smartphones: Insights from a Controlled Field Study”. In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia, MUM ’12*, pp. 36:1–36:8, New York, NY, USA, 2012. ACM.

ISBN: 978-1-4503-1815-0. doi: 10.1145/2406367.2406412. Disponível em: <<http://doi.acm.org/10.1145/2406367.2406412>>. event-place: Ulm, Germany.

TOLEDO, T., MUSICANT, O., LOTAN, T. “In-vehicle data recorders for monitoring and feedback on drivers’ behavior”, *Transportation Research Part C: Emerging Technologies*, v. 16, n. 3, pp. 320–331, jun. 2008. ISSN: 0968-090X. doi: 10.1016/j.trc.2008.01.001. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0968090X08000041>>.

HAQUE, M. M., OHLHAUSER, A. D., WASHINGTON, S., et al. “Decisions and actions of distracted drivers at the onset of yellow lights”, *Accident; Analysis and Prevention*, v. 96, pp. 290–299, nov. 2016. ISSN: 1879-2057. doi: 10.1016/j.aap.2015.03.042.

VAIANA, R., IUELE, T., ASTARITA, V., et al. “Driving Behavior and Traffic Safety: An Acceleration-Based Safety Evaluation Procedure for Smartphones”, *Modern Applied Science*, v. 8, n. 1, pp. p88, jan. 2014. ISSN: 1913-1844. doi: 10.5539/mas.v8n1p88. Disponível em: <<http://www.ccsenet.org/journal/index.php/mas/article/view/29037>>.

BERGASA, L. M., ALMERÍA, D., ALMAZÁN, J., et al. “DriveSafe: An app for alerting inattentive drivers and scoring driving behaviors”. In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*, pp. 240–245, jun. 2014. doi: 10.1109/IVS.2014.6856461.

ZALDIVAR, J., CALAFATE, C. T., CANO, J. C., et al. “Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones”. In: *2011 IEEE 36th Conference on Local Computer Networks*, pp. 813–819, out. 2011. doi: 10.1109/LCN.2011.6115556.

ALVESSON, M., SANDBERG, J. “Generating Research Questions Through Problematization”, *The Academy of Management Review*, v. 36, n. 2, pp. 247–271, 2011. ISSN: 0363-7425. Disponível em: <<https://www.jstor.org/stable/41318000>>.

ELLISON, A., DANIELS, R. “Profiling Drivers’ Risky Behaviour Towards All Road Users”. In: *Australasian College of Road Safety Conference*, 2012.

NAKAOKA, M., RAKSINCHAROENSAK, P., NAGAI, M. “Study on forward collision warning system adapted to driver characteristics and road environment”. In: *2008 International Conference on Control, Automation and Systems*, pp. 2890–2895, out. 2008. doi: 10.1109/ICCAS.2008.4694250.

- AKIN, D., AKBA, B. “A neural network (NN) model to predict intersection crashes based upon driver, vehicle and roadway surface characteristics 1”. 2010. Disponível em: </paper/A-neural-network-(NN)-model-to-predict-intersection-Akin-Akba/abb7c70caa5f2dc9de4ba3b33afa518509922fd3>. Library Catalog: www.semanticscholar.org.
- AIT-MLOUK, A., AGOUTI, T. “DM-MCDA: A web-based platform for data mining and multiple criteria decision analysis: A case study on road accident”, *SoftwareX*, v. 10, pp. 100323, jul. 2019. ISSN: 2352-7110. doi: 10.1016/j.softx.2019.100323. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352711019301438>.
- REN, H., SONG, Y., WANG, J., et al. “A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction”. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 3346–3351, nov. 2018. doi: 10.1109/ITSC.2018.8569437. ISSN: 2153-0017.
- ELAMRANI ABOU ELASSAD, Z., MOUSANNIF, H., AL MOATASSIME, H., et al. “The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review”, *Engineering Applications of Artificial Intelligence*, v. 87, pp. 103312, jan. 2020b. ISSN: 0952-1976. doi: 10.1016/j.engappai.2019.103312. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0952197619302672>.
- BAGNALL, A., FLYNN, M., LARGE, J., et al. “A tale of two toolkits, report the third: on the usage and performance of HIVE-COTE v1.0”, *arXiv:2004.06069 [cs, stat]*, abr. 2020. Disponível em: <http://arxiv.org/abs/2004.06069>. arXiv: 2004.06069.
- RUIZ, A. P., FLYNN, M., LARGE, J., et al. “The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances”, *Data Mining and Knowledge Discovery*, v. 35, n. 2, pp. 401–449, mar. 2021. ISSN: 1573-756X. doi: 10.1007/s10618-020-00727-3. Disponível em: <https://doi.org/10.1007/s10618-020-00727-3>.
- LINES, J., TAYLOR, S., BAGNALL, A. “Time Series Classification with HIVE-COTE: The Hierarchical Vote Collective of Transformation-Based Ensembles”, *ACM Transactions on Knowledge Discovery from Data*, v. 12,

n. 5, pp. 52:1–52:35, jul. 2018. ISSN: 1556-4681. doi: 10.1145/3182382. Disponível em: <<https://doi.org/10.1145/3182382>>.

Aggarwal, C. C. (Ed.). *Data Classification: Algorithms and Applications: 35*. 1^a edição ed. Boca Raton, CRC Press, jul. 2014. ISBN: 978-1-4665-8674-1.

BIER, A., JASTRZEBSKA, A., OLSZEWSKI, P. “Variable-Length Multivariate Time Series Classification Using ROCKET: A Case Study of Incident Detection”, *IEEE Access*, v. 10, pp. 95701–95715, 2022. ISSN: 2169-3536. doi: 10.1109/ACCESS.2022.3203523. Conference Name: IEEE Access.

SALLES, R., BELLOZE, K., PORTO, F., et al. “Nonstationary time series transformation methods: An experimental review”, *Knowledge-Based Systems*, v. 164, pp. 274–291, jan. 2019. ISSN: 0950-7051. doi: 10.1016/j.knosys.2018.10.041. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950705118305343>>.

DEMPSTER, A., PETITJEAN, F., WEBB, G. I. “ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels”, *Data Mining and Knowledge Discovery*, v. 34, n. 5, pp. 1454–1495, set. 2020. ISSN: 1573-756X. doi: 10.1007/s10618-020-00701-z. Disponível em: <<https://doi.org/10.1007/s10618-020-00701-z>>.

MIDDLEHURST, M., LARGE, J., FLYNN, M., et al. “HIVE-COTE 2.0: a new meta ensemble for time series classification”, *arXiv:2104.07551 [cs]*, abr. 2021. Disponível em: <<http://arxiv.org/abs/2104.07551>>. arXiv: 2104.07551.

TAN, C. W., DEMPSTER, A., BERGMEIR, C., et al. “MultiRocket: Multiple pooling operators and transformations for fast and effective time series classification”. fev. 2022. Disponível em: <<http://arxiv.org/abs/2102.00457>>. arXiv:2102.00457 [cs, stat].

TOO, J., ABDULLAH, A. R., SAAD, N. M. “A Comparative Analysis of Wavelet Families for the Classification of Finger Motions”, *International Journal of Advanced Computer Science and Applications (IJACSA)*, v. 10, n. 4, 2019a. ISSN: 2156-5570. doi: 10.14569/IJACSA.2019.0100426. Disponível em: <<https://thesai.org/Publications/ViewPaper?Volume=10&Issue=4&Code=IJACSA&SerialNo=26>>. Number: 4 Publisher: The Science and Information (SAI) Organization Limited.

SUBASI, A. “Chapter 2 - Data preprocessing”. In: Subasi, A. (Ed.), *Practical Machine Learning for Data Analysis Using Python*, Academic

Press, pp. 27–89, UK, jan. 2020. ISBN: 978-0-12-821379-7. doi: 10.1016/B978-0-12-821379-7.00002-3. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780128213797000023>>.

AL OMARI, F., HUI, J., MEI, C., et al. “Pattern Recognition of Eight Hand Motions Using Feature Extraction of Forearm EMG Signal”, *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, v. 84, n. 3, pp. 473–480, set. 2014. ISSN: 2250-1762. doi: 10.1007/s40010-014-0148-2. Disponível em: <<https://doi.org/10.1007/s40010-014-0148-2>>.

PHINYOMARK, A., NUIDOD, A., PHUKPATTARANONT, P., et al. “Feature Extraction and Reduction of Wavelet Transform Coefficients for EMG Pattern Classification”, *Elektronika ir Elektrotechnika*, v. 122, n. 6, pp. 27–32, jun. 2012. ISSN: 2029-5731. doi: 10.5755/j01.eee.122.6.1816. Disponível em: <<https://eejournal.ktu.lt/index.php/elt/article/view/1816>>. Number: 6.

AMIN, H. U., MALIK, A. S., AHMAD, R. F., et al. “Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques”, *Australasian Physical & Engineering Sciences in Medicine*, v. 38, n. 1, pp. 139–149, mar. 2015. ISSN: 1879-5447. doi: 10.1007/s13246-015-0333-x. Disponível em: <<https://doi.org/10.1007/s13246-015-0333-x>>.

MURPHEY, Y. L., MILTON, R., KILIARIS, L. “Driver’s style classification using jerk analysis”. In: *2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems*, pp. 23–28, mar. 2009. doi: 10.1109/CIVVS.2009.4938719.

DEMPSTER, A., SCHMIDT, D. F., WEBB, G. I. “MINIROCKET: A Very Fast (Almost) Deterministic Transform for Time Series Classification”. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 248–257, ago. 2021. doi: 10.1145/3447548.3467231. Disponível em: <<http://arxiv.org/abs/2012.08791>>. arXiv:2012.08791 [cs, stat].

COATES, A., NG, A., LEE, H. “An Analysis of Single-Layer Networks in Unsupervised Feature Learning”. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 215–223. JMLR Workshop and Conference Proceedings, jun. 2011. Disponível

em: <<https://proceedings.mlr.press/v15/coates11a.html>>. ISSN: 1938-7228.

TOO, J., ABDULLAH, A. R., MOHD SAAD, N., et al. “EMG Feature Selection and Classification Using a Pbest-Guide Binary Particle Swarm Optimization”, *Computation*, v. 7, n. 1, pp. 12, mar. 2019b. ISSN: 2079-3197. doi: 10.3390/computation7010012. Disponível em: <<https://www.mdpi.com/2079-3197/7/1/12>>. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

TRIANANTAPHYLLOU, E. *Multi-criteria Decision Making Methods: A Comparative Study*, v. 44, *Applied Optimization*. Boston, MA, Springer US, 2000. ISBN: 978-1-4419-4838-0 978-1-4757-3157-6. doi: 10.1007/978-1-4757-3157-6. Disponível em: <<http://link.springer.com/10.1007/978-1-4757-3157-6>>.