



UNIVERSIDADE FEDERAL DO PIAUÍ

CENTRO DE TECNOLOGIA

CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

WILSON BORBA DA ROCHA NETO

**SISTEMA DE TELEMETRIA AUTOMOTIVO PARA CLASSIFICAÇÃO DO ESTILO
DE DIREÇÃO**

TERESINA

2018

WILSON BORBA DA ROCHA NETO

SISTEMA DE TELEMETRIA AUTOMOTIVO PARA CLASSIFICAÇÃO DO ESTILO DE
DIREÇÃO

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia Elétrica do
Centro de Tecnologia da Universidade Federal
do Piauí, como requisito parcial à obtenção do
grau de bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. José Maria Pires
de Menezes Jr.

TERESINA

2018

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Piauí
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

D11s Rocha Neto, Wilson Borba da.
 Sistema de telemetria automotivo para classificação do estilo de direção / Wilson Borba da Rocha Neto. –
 2018.
 112 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Piauí, Centro de Tecnologia,
Curso de Engenharia Elétrica, Teresina, 2018.

Orientação: Prof. Dr. José Maria Pires de Menezes Júnior.

1. INTELIGÊNCIA ARTIFICIAL. 2. MICROELETRÔNICA. 3. VEÍCULOS AUTOMOTORES. 4.
CLASSIFICAÇÃO. 5. DIREÇÃO VEICULAR. I. Título.

CDD 621.3

WILSON BORBA DA ROCHA NETO

SISTEMA DE TELEMETRIA AUTOMOTIVO PARA CLASSIFICAÇÃO DO ESTILO DE
DIREÇÃO

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia Elétrica do
Centro de Tecnologia da Universidade Federal
do Piauí, como requisito parcial à obtenção do
grau de bacharel em Engenharia Elétrica.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. José Maria Pires de Menezes
Jr. (Orientador)
Universidade Federal do Piauí (UFPI)

Prof. Dr. XXXXXXXX XXXXXX XXXXXXXX
Universidade Federal do Piauí (UFPI)

Esp. Pedro Henrique Alves Ribeiro
Universidade Federal do Piauí (UFPI)

Este trabalho é dedicado à Francisca Célia Leitão Pinheiro Borba. Aquela que me emprestou um pouco de seu bom coração, me ensinando a ser uma pessoa responsável, sensata e de caráter. Aquela que me permitiu ser quem eu sou e me apoiou em todas minhas decisões. Aquela que põe minha felicidade em frente à própria. Aquela que é minha Mãe.

AGRADECIMENTOS

Ao meu orientador José Maria Pires de Menezes Jr. por tudo que me ensinou, incentivou e apoiou durante toda a graduação.

À minha família, em especial minha irmã pelos abraços e palavras carinhosas.

À Francisca Ramos, minha segunda mãe.

À Andreia Leal, minha namorada, pelo carinho, apoio incondicional e incentivo nas horas difíceis.

Aos colegas do Laboratório de Controle, pelas discussões ricas em conhecimento que me fizeram crescer como pesquisador.

Àqueles professores do Departamento de Engenharia Elétrica da Universidade Federal do Piauí que contribuíram com minha formação.

Aos alunos Ramon Pierote e José Borges, amigos e companheiros, os únicos que estiveram ao meu lado durante toda a graduação.

"Se eu vi mais longe, foi por estar sobre ombros
de gigantes"

(Sir. Isaac Newton)

RESUMO

Este trabalho apresenta o desenvolvimento de um sistema de telemetria para captura dos dados eletromecânicos de um veículo através do sistema de diagnóstico automotivo *On-Board Diagnostic II* (OBDII), obrigatório em todos os veículos produzidos atualmente. Após capturados, os dados são armazenados em um banco de dados online o qual pode ser posteriormente acessado pelo usuário. O sistema de telemetria desenvolvido apresenta vantagens quando comparado as outras soluções para captura de informações do sistema automotivo, incluindo elevada qualidade e quantidade de dados. Estes dados foram posteriormente utilizados na classificação do estilo de direção de motoristas, com o intuito de identificar o estilo agressivo destes, o maior causador de acidentes no trânsito. Dessa forma, o sistema é caracterizado como um *Driver Assistance System* (DAS), indicando ao condutor seu estilo agressivo para impedir que o mesmo crie situações de risco. Uma extensa discussão é apresentada para situar este trabalho dentre a literatura relacionada. Para a classificação utilizou-se o algoritmo *K-Nearest Neighbors* (KNN), o qual, apesar de não apresentar aplicações recentes em classificação de estilos de direção, possui uma boa acurácia em outras aplicações no sistema automotivo. Duas abordagens, instantânea e janela temporal, foram utilizadas com o classificador KNN, as quais apresentaram divergências nos resultados. Estas divergências são explicadas para que a abordagem mais apropriada seja selecionada. Os resultados obtidos sugerem que a escolha de uma abordagem depende da magnitude do banco de dados utilizado na classificação e do poder computacional do ambiente no qual o classificador é implementado. Além disso, o sistema de telemetria proposto foi testado e apresentou estabilidade e velocidade de captura de dados de forma satisfatória, confirmando as vantagens esperadas.

Palavras-chave: Sistema de Telemetria. Classificação de Estilos de Direção. On Board Diagnostic. Classificador KNN.

ABSTRACT

In this work, a telemetry system is developed to capture the electromechanical data from a vehicle through the On-Board Diagnostic II (OBDII) automotive diagnostic system, mandatory in all vehicles currently produced. Once captured, the data was stored in an online database where it can be retrieved by the user. The telemetry system offers advantages when compared to other solutions for automotive system data retrieval, including high quality and quantity of data. This data was used to classify a driver's driving style, in order to identify the aggressive style, the biggest cause of traffic crashes. In this way, the system is characterized as a Driver Assistance System (DAS), which indicates to the driver his aggressive style to prevent the creation of risky situations. An extensive discussion is presented to situate this work among related literature. For the classification, the algorithm K-Nearest Neighbors (KNN) was utilized, which, despite not presenting recent applications in driver's style classification, has a good accuracy in other applications in the automotive system. Two approaches, instantaneous and time window, were used with the KNN classifier, which showed divergences in the results. These divergences are explained so that the most appropriate approach is selected. The results obtained suggest that the selection of a approach depends on the length of the classifier's database and the computational power of the environment where the classifier is implemented. Moreover, the proposed telemetry system was tested and presented satisfactory stability and data capture speed, confirming the expected advantages.

Palavras-chave: Telemetry System. Driving Styles Classification. On-Board Diagnostic II. KNN classifier.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação dos tópicos desenvolvidos nesta Monografia	27
Figura 2 – Composição do perfil de velocidade de determinado motorista	31
Figura 3 – O sistema automotivo como um sistema de controle	32
Figura 4 – Composição do perfil de velocidade de determinado motorista	33
Figura 5 – Nuvem de palavras indicando quais categorias são as mais utilizadas na literatura.	34
Figura 6 – Representação das categorias que dividem os fatores que influenciam o ato de dirigir.	36
Figura 7 – Representação de como o ato de dirigir é construído.	36
Figura 8 – Representação da relação entre os termos que definem Direção.	41
Figura 9 – O <i>iDriveSafe</i> , um exemplo de aplicativo para classificação de motoristas. . .	46
Figura 10 – O <i>Racer Simulator</i> , um exemplo de simulador de direção.	48
Figura 11 – Exemplo da tela de um simulador de direção.	48
Figura 12 – As condições dos experimentos em um simulador de direção.	48
Figura 13 – Parametrização de um motorista simulado em software pelo CarMaker. . .	49
Figura 14 – Motorista como um usuário no software CarMaker.	49
Figura 15 – Estrutura de uma estrada desenhada no software SUMO.	50
Figura 16 – Possíveis locais de instalação do terminal OBDII e sua pinagem.	52
Figura 17 – Representação da distribuição da CAN em um automóvel.	53
Figura 18 – Exemplo de captura de informação utilizando o terminal OBDII.	55
Figura 19 – Sistema de telemetria desenvolvido	57
Figura 20 – Faces da placa de circuito impresso que contém o ELM327.	58
Figura 21 – Microcontrolador PIC18F4550	59
Figura 22 – Módulo Bluetooth Hc-05	60
Figura 23 – Display LCD 20x4	60
Figura 24 – SIM 808	61
Figura 25 – SIM 808	62
Figura 26 – Módulo Cartão de memória	62
Figura 27 – Sistema de telemetria prototipado.	63
Figura 28 – Representação de um pacote de captura enviado ao servidor pelo sistema. .	64
Figura 29 – Fluxograma de funcionamento do sistema	65

Figura 30 – Representação da relação entre os termos que definem Direção.	66
Figura 31 – Exemplos de dados clusterizados em diferentes classes.	68
Figura 32 – Ilustração do Algoritmo KNN.	73
Figura 33 – Ilustração do Algoritmo KNN com número de vizinhos pares.	74
Figura 34 – Percurso utilizado para a captura de informações.	77
Figura 35 – Possíveis problemas de comunicação do sistema de telemetria.	78
Figura 36 – Matriz de Correlação da captura Calma representada graficamente.	79
Figura 37 – Matriz de Correlação da captura Normal representada graficamente.	79
Figura 38 – Matriz de Correlação da captura Agressiva representada graficamente	80
Figura 39 – Dinâmica temporal do atributo MAF para as classes Calma e Agressiva. . . .	82
Figura 40 – Dinâmica temporal do atributo RPM para as classes Calma e Agressiva. . . .	83
Figura 41 – Agrupamento dos atributos pertencentes as três classes.	83
Figura 42 – Representação da distribuição do atributo Vel para cada classe.	84
Figura 43 – Algoritmo KNN com uma abordagem instantânea.	85
Figura 44 – Algoritmo KNN com uma abordagem de janelas temporais.	86
Figura 45 – Variação da máxima acurácia para diferentes números de vizinhos mais próximos	88
Figura 46 – Variação da mínima acurácia para diferentes números de vizinhos mais próximos	88
Figura 47 – Variação da Acurácia média para diferentes números de vizinhos mais próximos	89
Figura 48 – Variação do Desvio Padrão da acurácia para diferentes números de vizinhos mais próximos	89
Figura 49 – Variação da acurácia Média para diferentes janelas temporais	90
Figura 50 – Variação do Desvio Padrão da acurácia para diferentes janelas temporais . .	90
Figura 51 – Variação da acurácia Média para diferentes centroides de janelas temporais .	91
Figura 52 – Variação do Desvio Padrão da acurácia para diferentes centroides de janelas temporais	92
Figura 53 – Agrupamento dos centroides dos grupos de atributos pertencentes as três classes.	94
Figura 54 – Representação da distribuição do atributo Vel para cada classe.	95
Figura 55 – Agrupamento dos centroides dos grupos de atributos pertencentes as três classes.	96
Figura 56 – Representação da distribuição do atributo Vel para cada classe.	96

Figura 57 – Representação da diferença entre as abordagens instantânea e janela temporal	97
Figura 58 – Variação na média da acurácia conforme o aumento da janela temporal . . .	98
Figura 59 – Custo computacional para a abordagem instantânea.	99
Figura 60 – Custo computacional para as variações da abordagem de janela temporal. . .	99

LISTA DE TABELAS

Tabela 2 – Fatores categorizados e a sua descrição	35
Tabela 3 – Fatores categorizados e a sua descrição	37
Tabela 4 – Definição de termos com base em diferentes autores	39
Tabela 5 – Definição de termos menos utilizados na literatura	39
Tabela 6 – Variáveis relacionadas aos fatores da categoria Motorista.	44
Tabela 7 – Variáveis relacionadas aos fatores da categoria Ambiente.	45
Tabela 8 – Comparação entre as soluções para captura de informação.	51
Tabela 9 – Sistemas de captura utilizados por outros trabalhos.	51
Tabela 10 – Pinagem do terminal OBDII e sua descrição.	52
Tabela 11 – Os dez serviços disponíveis no sistemas OBDII.	54
Tabela 12 – Exemplo de decodificação da informação recebida do OBDII.	55
Tabela 13 – Componentes da placa de circuito impresso que contém o ELM327.	58
Tabela 14 – Especificações do vetor de informações recebido do GPS.	62
Tabela 15 – Exemplos de classes de motoristas	69
Tabela 16 – Exemplo de matriz de confusão	70
Tabela 17 – Contrastes entre diferentes algoritmos	71
Tabela 18 – Alguns algoritmos utilizados pela literatura	72
Tabela 19 – Informações capturadas do sistema OBDII e o sensor interno correspondente.	76
Tabela 20 – Matriz de Correlação da captura Calma	80
Tabela 21 – Matriz de Correlação da captura Normal	80
Tabela 22 – Matriz de Correlação da captura Agressiva	81
Tabela 23 – Descrição do conjunto de dados	81
Tabela 24 – Descrição das classes de estilo de direção	84
Tabela 25 – Matriz de confusão para $K = 1$ abordagem instantânea	93
Tabela 26 – Matriz de confusão para $K = 1$ e $t = 6$ abordagem janela temporal	93
Tabela 27 – Matriz de confusão para $K = 1$ e $t = 200$ abordagem janela temporal	93
Tabela 28 – Matriz de confusão para $K = 1$ e $t = 6$ abordagem centroide	93
Tabela 29 – Matriz de confusão para $K = 1$ e $t = 200$ abordagem centroide	94
Tabela 30 – Comparação entre as diferentes abordagens utilizadas com o classificador KNN.	101
Tabela 31 – Comparação entre os resultados obtidos	101

LISTA DE ABREVIATURAS E SIGLAS

Acl	Posição do pedal de aceleração
CAN	<i>Controller Area Network</i>
DAS	<i>Driver Assistance Systems/Sistemas de Auxílio ao Motorista</i>
DBQ	<i>Driver Behaviour Questionnaire</i>
DTW	<i>Dynamic Time Warp</i>
ECU	<i>Engine Central Unit</i>
Elo	Carga no motor
FIS	<i>Fuzzy Inference System/Sistema de Inferência Fuzzy</i>
FTP	<i>File Transfer Protocol/Protocolo de Transferência de Dados</i>
GPRS	<i>General Packet Radio Service</i>
GPS	<i>GlobalPositioning System/Sistema de Posicionamento Global</i>
GSM	<i>Global System for Mobile communications</i>
GUI	<i>Graphic User Interface/Interface Gráfica do Utilizador</i>
HMM	<i>Hidden Markov Model/Modelo Oculto de Markov</i>
I2C	<i>Inter-Integrated Circuit</i>
ITS	<i>Intelligent Transport System/Sistemas de Transporte Inteligentes</i>
KNN	<i>K-Nearest Neighbors/K-Vizinhos mais Próximos</i>
MAC	<i>Media Access Control</i>
MAF	<i>Mass Air Flow/Fluxo de massa ar-combustível</i>
NHTSA	<i>National Highway Traffic Safety Administration</i>
NMEA	<i>National Marine Electronics Association</i>
NN	<i>Neural Networks/Redes Neurais</i>
OBDII	<i>On-Board Diagnostic II/Sistema de Diagnóstico Automotivo II</i>
ONU	Organização das Nações Unidas
PEP	<i>Python Enhance Proposals/Propostas de Aperfeiçoamento do Python</i>
RMC	<i>Recommended Minimum Data for GPS</i>
RPM	Rotações por minuto
SPI	<i>Serial Peripheral Interface</i>
SVM	<i>Support Vecto Machines</i>
UART	<i>Universal Asynchronous Receiver/Trasmitter</i>
Vel	Velocidade do veículo

VHQ *Vehicule Integrity Questionnaire*

LISTA DE SÍMBOLOS

\vec{x} e \vec{y}	Vetores algébricos de dimensão l
D_{min}	Menor distância entre dois vetores
l	Dimensão dos vetores no cálculo da distância Euclidiana
\overline{RPM}	Valor máximo do atributo RPM em uma coleta
\underline{RPM}	Valor mínimo do atributo RPM em uma coleta
K	Número de vizinhos mais próximos
t	Tamanho da janela temporal
$ttRatio$	Proporção dos dados utilizados para treino
m	Número de grupos de vetores de atributos

SUMÁRIO

1	INTRODUÇÃO	22
1.1	Objetivos Principais e Secundários	25
1.2	Histórico de Desenvolvimento da Monografia	26
1.3	Produção Científica	27
1.4	Organização Geral do Restante do Trabalho	28
2	SISTEMA AUTOMOTIVO: CONCEITOS, FATORES E SEGURANÇA	30
2.1	Introdução	30
2.2	Segurança	30
2.3	Sistema Automotivo	32
2.3.1	<i>A Direção e seus fatores</i>	33
2.3.2	<i>A importância do motorista</i>	37
3	SISTEMA AUTOMOTIVO: INFORMAÇÕES E SISTEMAS DE CAPTURA	42
3.1	Introdução	42
3.2	Sensores e Sinais	42
3.2.1	<i>Motorista</i>	42
3.2.2	<i>Ambiente e Veículo</i>	44
3.3	Sistemas para obtenção de dados	44
3.3.1	<i>Smartphones</i>	45
3.3.2	<i>Simuladores Instrumentados</i>	46
3.3.3	<i>Softwares como ferramentas de simulação</i>	47
3.3.4	<i>Sistemas Embarcados</i>	49
3.4	On Board Diagnostic II	51
4	COMUNICAÇÃO COM O VEÍCULO	56
4.1	Introdução	56
4.2	Sistema de Telemetria	56
4.2.1	<i>Desenvolvimento do hardware</i>	57
4.2.1.1	<i>ELM327</i>	57
4.2.1.2	<i>Microcontrolador</i>	59
4.2.1.3	<i>Comunicação Bluetooth</i>	59

4.2.1.4	<i>Transmissão remota</i>	61
4.2.1.5	<i>Armazenamento de dados</i>	62
4.2.1.6	<i>Esquemático e Protótipo</i>	62
4.2.2	Desenvolvimento do Software	63
5	SISTEMA AUTOMOTIVO: CLASSIFICAÇÃO DO MOTORISTA	66
5.1	Introdução	66
5.2	Algoritmos de Classificação e Clusterização	66
5.2.1	<i>Clusterização</i>	67
5.2.2	<i>Classificação</i>	68
5.2.2.1	<i>Validação da classificação</i>	69
5.2.2.2	<i>Escolha do algoritmo</i>	70
5.2.2.3	<i>K-Nearest Neighbors</i>	72
6	METODOLOGIA UTILIZADA E PRÉ-PROCESSAMENTO DOS DADOS	75
6.1	Abordagem Utilizada	75
6.2	Dados	76
6.3	Coleta	76
6.4	Pré-processamento dos dados	77
6.4.1	<i>Preparação e Limpeza dos dados</i>	77
6.4.2	<i>Seleção e Preparação dos Atributos</i>	78
6.5	Análise	82
6.6	Algoritmo de Classificação	85
7	RESULTADOS	87
7.1	Introdução	87
7.1.1	<i>Abordagem instantânea</i>	87
7.1.2	<i>Abordagem de janela temporal</i>	88
7.1.2.1	<i>Janela temporal</i>	90
7.1.2.2	<i>Centroides</i>	91
7.1.3	<i>Análise dos resultados da classificação</i>	92
7.1.3.1	<i>Centroides</i>	94
7.1.3.2	<i>Janelas temporais</i>	95
7.1.4	<i>Oscilação</i>	97

7.2	Custo Computacional	99
7.3	Resumo dos resultados	100
8	CONCLUSÕES E TRABALHOS FUTUROS	102
8.1	Propostas de Trabalhos Futuros	104
	REFERÊNCIAS	105
	ANEXOS	111
	ANEXO A – Esquematico do Sistema de Telemetria	112

1 INTRODUÇÃO

Segurança no trânsito sempre foi um tópico bastante discutido. Atualmente, devido à expansão do avanço tecnológico aos meios de transporte, a qual lapidou conceitos como carros autônomos, inteligentes e conectados, esse tópico tem crescido ainda mais. A importância da segurança é relevante em qualquer desenvolvimento ou estudo relacionado à melhoria do ambiente automotivo, como pode ser visto em instituições especializadas no ensino de engenharia automotiva, nas quais a segurança nas estradas compõe o primeiro, e considerado mais importante, conjunto de conceitos que qualquer candidato a engenheiro automotivo deve dominar (*Automotive Engineering Master's Program at Chalmers University of Technology*¹).

A importância dada à segurança no trânsito é justificada pelo seu impacto nos meios sociais e econômicos. Segundo a Organização Mundial da Saúde (World Health Organization, 2015), acidentes no trânsito chegam a causar 1,25 milhão de mortes por ano, sendo a maior parte destas entre adolescentes de 15 a 29 anos. Apenas no Brasil, foram 662.219 mortes entre os anos de 1998 e 2015 (Associação Brasileira de Prevenção dos Acidentes de Trânsito, 2017). Entretanto, esse custo não é só em vidas humanas, os débitos hospitalares, previdenciários e de prevenção somaram 12,11 bilhões de reais apenas no ano de 2012 (Observatório Nacional de Segurança Viária, 2015). O impacto é tão grande que a Organização das Nações Unidas (ONU) (United Nations Road Safety Collaboration and others, 2011) declarou esta década (2011-2020) como a década de ação pela segurança no trânsito, instituindo um Plano de Ação Global, que conta com participação do Brasil, resultando na criação do projeto nacional Vida no Trânsito².

O aumento do número de veículos, de uma forma lógica e conceitual, influencia nas estatísticas de violência no trânsito. Entretanto, o veículo não é o único componente levado em conta na segurança. Podemos definir o sistema automotivo como sendo um sistema de controle no qual as decisões tomadas e executadas dependem da influência de três principais componentes: Motorista, Veículo e Ambiente. Logo, a segurança no trânsito também concerne, por exemplo, ao estado psicológico, atenção, habilidade e preferências pessoais do motorista (ABUALI; ABOU-ZEID, 2016). Além disso, parâmetros como o clima no momento da direção e a condição da estrada, influenciam no estilo de direção e, consequentemente, na segurança.

Essa distinção é importante, pois cada um desses componentes influencia de maneira diferente os acidentes de trânsito. O motorista, por exemplo, é o causador substancial dessas

¹ <https://www.chalmers.se/en/education/programmes/masters-info/Pages/Automotive-Engineering.aspx>

² <http://portalms.saude.gov.br/saude-de-a-z/acidentes-e-violencias/41896-projeto-vida-no-transito>

estatísticas, sendo o elemento que apresenta a maior porcentagem de falhas críticas na análise da cadeia de eventos de acidentes, somando um total de 93% dos casos. Essa estimativa é reafirmada pela *National Highway Traffic Safety Administration* (NHTSA), uma entidade do Departamento de Transporte dos Estados Unidos, em dois relatórios técnicos de 2008 e 2015 (SINGH, 2015; NHTSA, 2008), os quais foram baseados em investigações minuciosas de acidentes de trânsito autorizadas e financiadas pelo governo, provando-se uma fonte confiável para análise.

Um dos causadores substanciais dessas estatísticas é o comportamento agressivo do motorista (AAA - Foundation for Trafic Safety, 2016). Este comportamento pode ser identificado por (i) características comportamentais do próprio, como utilização excessiva da buzina, propensamente mantendo-se próximo ao veículo a sua frente, discutindo ou gritando; ou por (ii) análise de eventos do veículo, como acelerações e desacelerações bruscas. Logo, a identificação de momentos de agressividade e a comunicação destes ao motorista pode ser utilizada como uma ferramenta para diminuição destas causalidades. Estes momentos de agressividade compõem o estilo de direção, isto é, a maneira com a qual um determinado motorista lida com um evento de direção, como aceleração ou ultrapassagens. Dessa forma, a classificação do estilo de direção é uma alternativa para a identificação de momentos de agressividade do motorista.

A análise de características comportamentais dos motoristas geralmente é baseada em investigações policiais, depoimentos e questionários preenchidos pelo próprio condutor, informações que apresentam limitada acurácia (DINGUS *et al.*, 2006). Por conta disso, informações obtidas do motorista de maneira impessoal, como a utilização de dados capturados diretamente do veículo ou captura de imagem e vídeo, mostra-se uma alternativa viável para analisar e identificar o estilo agressivo do motorista e, consequentemente, utilizar tal identificação em uma aplicação de segurança, auxiliando o motorista a gerenciar situações de risco.

No tocante aos dados capturados do próprio veículo, diversas soluções já foram propostas. Algumas delas (ALJAAFREH; ALSHABATAT; AL-DIN, 2012; ARAÚJO *et al.*, 2012; CASTIGNANI *et al.*, 2015) aproveitam-se da gama de sensores presentes nos smartphones, como acelerômetros, giroscópios, magnetômetros e o *Global Positioning System* (Sistema de Posicionamento Global, GPS). O sistema proposto por ARAÚJO *et al.* (2012), por exemplo, utiliza estes sensores para analisar um total de 12 variáveis, derivadas de medidas estatísticas da velocidade e aceleração do veículo. Os autores informam que a escolha deste número de variáveis e de quais valores utilizar foram necessários devido ao ruído na medição sofrido pelo smartphone, causado pelas vibrações do veículo.

Outras alternativas (FERNANDEZ; ITO, 2016; FUGIGLANDO *et al.*, 2017; DER-BEL *et al.*, 2015) capturam informações diretamente do sistema eletromecânico do veículo, através do *On Board Diagnostic II* (Sistema de Diagnóstico Automotivo II, OBDII) (TSBEC, 2004), o que garante uma maior variedade e confiabilidade nos dados obtidos, incluindo velocidade, aceleração, rotações por minuto, posição do pé do acelerador e carga no motor. Esta premissa pode também ser utilizada na detecção e classificação do consumo de combustível do automóvel (PEREIRA; ALVES; MACEDO, 2016) e na predição de falhas no veículo (JARAMILLO; NARVAEZ, 2014).

Os dados capturados pelos métodos descritos podem ser armazenados no próprio sistema de captura ou enviados para outro sistema, no qual seriam processados. Esta última opção é denominada telemetria, isto é, um sistema eletrônico de monitoramento construído para rastrear, medir ou enviar comandos a algum equipamento que esteja distante ou em constante movimento. Quanto maior for a quantidade e variedade dos dados obtidos, mais interessante será o envio destes a um ponto de maior poder de processamento e, também, melhor será a descoberta de conhecimento. Por conta disso, a utilização de um sistema de telemetria automotivo conectado ao OBDII é uma opção interessante para capturar e analisar dados eletromecânicos de um automóvel. Estes dados, podem então ser utilizados para entendermos a dinâmica do componente mais influente no sistema automotivo: o Motorista; do qual a extração e compreensão de momentos de agressividade são importantes para manter o sistema automotivo dentro de uma margem de segurança.

Para identificar o estilo agressivo com os dados capturados, um sistema de classificação baseado em aprendizagem de máquinas é utilizado. O algoritmo *K-Nearest Neighbors* (K-Vizinhos mais Próximos, KNN) apresenta uma boa acurácia em diferentes campos de aplicação relacionados ao sistema automotivo, como na criação de um padrão do estilo de direção (KWAK; WOO; KIM, 2016), integridade dos sistemas do veículo (SRINIVASAN, 2018) e modelagem de emissão do veículo (LI; QIAO; YU, 2016). O bom desempenho do algoritmo KNN nesses casos motivou a sua utilização na classificação de estilos de direção com o intuito de verificar sua performance também nesta situação; sendo tal classificação realizada utilizando dados capturados do veículo através de um sistema de telemetria. Esta ideia guiou a elaboração do estudo aqui apresentado, o qual possui seus objetivos discutidos a seguir.

1.1 Objetivos Principais e Secundários

Dada a importância do monitoramento, do diagnóstico e da evolução dos sistemas ao longo do tempo, um dos objetivos principais deste estudo é desenvolver um sistema embarcado microcontrolado integrado ao automóvel, o qual fará uso do OBDII para captura de dados dos sistemas eletromecânicos de um veículo automotivo pessoal. Este sistema será capaz de conectar-se a um servidor web, utilizando a rede de telefonia móvel, o que possibilita a criação de um banco de dados online com as informações capturadas. Apesar de já existirem opções no mercado de dispositivos de telemetria automotivos que possam ser utilizados para captura de informações semelhantes, optou-se por desenvolver um sistema autoral, já que o controle do fluxo de dados, a flexibilidade de operação e possibilidade de incrementar-se aplicações do sistema se dão de forma livre em servidores e protótipos próprios.

Além disso, para identificar momentos de agressividade no estilo de direção do motorista do veículo, já que essa abordagem concerne à segurança no trânsito, este trabalho também apresenta como objetivo principal a aplicação de um algoritmo de aprendizagem de máquinas para efetuar a classificação dos dados capturados em classes que descrevam o estilo de direção de motoristas.

Para montar padrões e classificar o estilo de condução é necessário, então, um conjunto de observações. Dada a indefinição da classificação a ser feita, já que o estilo agressivo de um motorista pode ser influenciado por diferentes fatores relacionados ao próprio, ao veículo e ao ambiente, é necessária uma discussão capaz de definir com clareza a abordagem deste estudo. Portanto, com o intuito de satisfazer os objetivos principais, este trabalho também possui os objetivos secundários listados a seguir.

- Levantar uma discussão em torno da literatura já existente, para que seja possível compreender a aplicação exata da metodologia desenvolvida;
- Analisar o sistema automotivo para discernir e validar a utilização correta dos dados capturados;
- Implementar o classificador KNN utilizando a linguagem de programação Python;
- Discutir profundamente a performance e utilização do classificador KNN no estudo proposto.

1.2 Histórico de Desenvolvimento da Monografia

No ano de 2015, deu-se início ao processo de estudo da dinâmica do sistema automotivo. Através do estudo introdutório à eletrônica automotiva, desenvolveu-se um, então nomeado, sistema embarcado automotivo utilizando a recém-lançada plataforma Raspberry Pi³, uma plataforma embarcada que roda um sistema operacional próprio. Utilizando-se da linguagem de programação Python⁴, o objetivo inicial deste sistema era a identificação de acelerações agressivas por parte do motorista, as quais resultariam em um alarme sonoro para que o mesmo pudesse manter-se dentro de uma faixa previamente estabelecida. Além disso, uma *Graphic User Interface* (Interface Gráfica para o Usuário, GUI) foi desenvolvida com o intuito de mostrar ao motorista a dinâmica de sua velocidade e aceleração. O estudo do Raspberry Pi resultou na criação de relatórios técnicos para utilização de *features* da plataforma⁵, incluindo o protocolo Bluetooth, utilizado para captura das informações do veículo, e aplicações com *Near Field Communication*, que objetivava a segurança do sistema, para que apenas o motorista do veículo fosse capaz de acioná-lo.

Em 2016/2017, pretendendo-se embarcar o sistema previamente desenvolvido no automóvel, desenvolveu-se um novo sistema microcontrolado, utilizando-se a linguagem de programação C aliada ao PIC18F4550, o qual é capaz de enviar informações do veículo a um servidor online utilizando-se de um módulo de conexão com a rede de telefonia móvel, requisições HTTP e o *File Transfer Protocol* (Protocolo de Transferência de Dados). Dessa forma, além de otimizar a velocidade de captura, mais informações puderam ser incluídas, já que todo o processamento destas seria feito remotamente. Assim, um sistema de telemetria mais compacto, veloz e dinâmico que o anterior tornou-se a base para permitir captura das informações do veículo, características que fizeram deste um trabalho agraciado com o 1º Lugar do prêmio Destaque de Iniciação Científica UFPI - 2017⁶. O sistema de telemetria será discutido no capítulo 4.

Nos anos 2017/2018, o sistema de telemetria desenvolvido foi utilizado para a captura de informações em percursos previamente definidos, procurando-se identificar a dinâmica do automóvel e determinar que informações seriam úteis para uma análise do sistema automo-

³ <https://www.raspberrypi.org/>

⁴ <https://www.python.org/>

⁵ <https://github.com/wborbaneto/research/tree/master/JTC>

⁶ <https://ufpi.br/ultimas-noticias-ufpi/20647-trabalhos-submetidos-pelos-alunos-do-programa-de-iniciacao-cientifica-sao-premiados-no-xxvi-seminario-de-iniciacao-cientifica>

tivo. Após esta análise, os dados selecionados foram utilizado juntamente ao algoritmo KNN para classificação do estilo de direção de motoristas cujos resultados compõem o estudo aqui apresentado. A metodologia utilizada na captura, análise e classificação das informações serão discutidas nos Capítulo 6 e 7. Para melhor representar as contribuições abordadas nesta seção, a Figura 1 ilustra o que foi discutido.

Figura 1 – Representação dos tópicos desenvolvidos nesta Monografia



Fonte – Elaborado pelo Autor.

1.3 Produção Científica

Uma lista completa das premiações e trabalhos publicados ou aceitos com os estudos discutidos na seção anterior é fornecida a seguir.

- ROCHA NETO, W. B.; MENEZES JUNIOR, J. M. P. Classificação de dados de um sistema de telemetria automotiva. *Anais do V Congresso Brasileiro de Sistemas Fuzzy(CBSF'2018)*, Fortaleza-CE, 2018.
- ROCHA NETO, W. B.; MENEZES JUNIOR, J. M. P. Perfil de condução de motoristas: Análise por meio de algoritmos de reconhecimentos de padrões. *Anais do XXII Congresso Brasileiro de Automática(CBA'2018)*, João Pessoa-PB, 2018.
- ROCHA NETO, W. B.; MENEZES JUNIOR, J. M. P. Classificação de dados de um sistema de telemetria automotiva utilizando a rede ELM e o KNN. Aceito no *XIII Brazilian Congress on Computational Intelligence(CBIC'2017)*, Niterói-RJ, 2017.

- ROCHA NETO, W. B.; MENEZES JUNIOR, J. M. P.; SOUSA, R. V. L. C. Análise de dados obtidos através de um sistema de telemetria automotivo utilizando K-NN. *Anais do XIV Encontro Nacional de Inteligência Artificial e Computacional(ENIAC'2017)*, Uberlândia-MG, 2017.
- 1º Lugar no Prêmio Destaque na Iniciação Científica UFPI - 2017 com o trabalho intitulado "Sistema de Telemetria Automotivo Conectado à Internet."

1.4 Organização Geral do Restante do Trabalho

O restante deste trabalho está dividido conforme exposto a seguir. O Capítulo 2 é destinado a construir, de forma gradual e lógica, blocos de definições que serão utilizados para desenvolver a proposta pretendida. São tratados os tópicos de segurança no trânsito e como os sistemas de segurança atuam no veículo, fatores que influenciam o sistema automotivo e a definição dos termos relacionados à direção.

O Capítulo 3 objetiva discutir as diversas abordagens encontradas na literatura em relação à captura de dados do sistema automotivo. Sensores, sinais e sistemas utilizados são apresentados e comparados. Também neste capítulo discute-se o sistema de diagnóstico automotivo OBDII, um dos meios de aquisição de informações do sistema automotivo. Além disso, os algoritmos geralmente utilizados na classificação de estilos de direção são discutidos e o algoritmo KNN é apresentado.

No Capítulo 4 o desenvolvimento do sistema de telemetria autoral é discutido, com detalhamento do hardware e software utilizados. Seu funcionamento é explicado e esquemáticos do protótipo construído são apresentados. Os códigos e esquemáticos do sistema são disponibilizados⁷ sob a *MIT License*⁸, possibilitando livre acesso e utilização destes.

O Capítulo 5 introduz a ideia de que as informações capturadas com o sistema de telemetria proposto podem ser utilizados na classificação de estilos de direção. Esta ideia já foi trabalhada na literatura e aplicada em diversas situações, inclusive na própria segurança no trânsito. Por conta disso, este capítulo procura discutir soluções para a classificação do estilo de direção de motoristas.

No Capítulo 6 a abordagem utilizada neste trabalho é discutida, situando a aplicação aqui empregada dentro da literatura previamente discutida nos Capítulos 2, 3 e 5. A metodologia

⁷ <https://github.com/wborbaneto/research>

⁸ <https://opensource.org/licenses/MIT>

de captura, pré-processamento dos dados e análise é apresentada. O funcionamento e performance do sistema de telemetria desenvolvido são validados, apresentando resultados satisfatórios para a aplicação desejada.

No Capítulo 7 o algoritmo de classificação utilizado é discutido. Os dados analisados no Capítulo 6 são utilizados como entrada do classificador KNN, utilizando, para isso, duas abordagens. Os resultados são discutidos e a melhor abordagem é apresentada.

Por fim, o Capítulo 8 apresenta um resumo do que foi discutido, os resultados obtidos e evidencia os próximos passos para o estudo proposto.

2 SISTEMA AUTOMOTIVO: CONCEITOS, FATORES E SEGURANÇA

2.1 Introdução

Como comentado, os objetivos deste trabalho são o desenvolvimento de um sistema de telemetria automotivo e a utilização das informações capturadas por este na classificação de estilos de direção do motorista. Antes de propor uma solução que possa ser empregada em um veículo, é necessário entender com mais clareza as principais características do mesmo, de forma que uma abordagem eficaz possa ser escolhida para o desenvolvimento. Por conta disso, este capítulo destina-se a construir, de forma gradual e lógica, blocos de definição pertencentes ao sistema automotivo que serão utilizados para situar a proposta pretendida dentro de uma área específica na literatura relacionada.

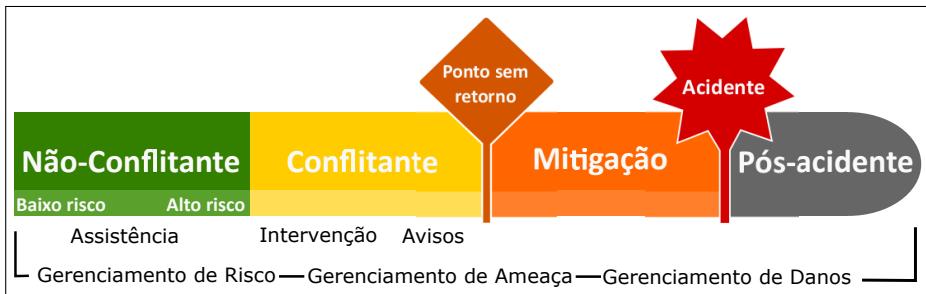
Inicialmente a segurança no sistema automotivo será abordada, explicando-se como sistemas de segurança podem atuar para evitar ou amenizar as casualidades de um acidente. Logo após, uma definição de sistema automotivo é fundamentada, levando em conta os fatores e elementos importantes em sua construção.

2.2 Segurança

Dentre os componentes do sistema automotivo, o motorista é o maior culpado na causa de acidentes, sendo responsável por cerca de 93% das falhas críticas na análise da cadeia de eventos de acidentes (NHTSA, 2008). Sob o olhar do motorista ou de algum ocupante do veículo, uma linha do tempo de segurança (CUI *et al.*, 2016; MASCIO *et al.*, 2011; BETANCUR *et al.*, 2018) pode ser descrita, a qual é apresentada na Figura 2. Nesta, a cadeia de eventos que resulta em um acidente pode ser analisada, na qual os eventos são divididos em fases. Primeiramente, duas seções distintas podem ser identificadas, chamadas de pré-acidente e pós-acidente.

A seção pré-acidente é dividida em fases, chamadas de não-conflitante, conflitante e mitigação. A fase não-conflitante pode ainda ser seccionada em (i) direção de baixo risco, na qual as margens de segurança são altas e não há nenhum risco eminente de acidentes; e (ii) direção de alto risco, na qual situações perigosas podem ser geradas por alguma ação do motorista ou daqueles que o cercam. Um exemplo de ação do motorista que o colocaria em uma situação perigosa, e consequentemente classificaria sua direção como de alto risco, é manter uma distância muito pequena entre o seu veículo e o veículo da frente.

Figura 2 – Composição do perfil de velocidade de determinado motorista



Fonte – Elaborado pelo Autor.

Os sistemas de segurança que estão ativos na fase não-conflitante são chamados de *Driver Assistance Systems* (Sistemas de Auxílio ao Motorista, DAS), já que o objetivo desses sistemas é fornecer informações ou executar tarefas para que o motorista mantenha-se dentro de uma margem de segurança. Um exemplo de DAS é o controle da aceleração máxima do carro de forma automática para mantê-la dentro de um limite considerado seguro em relação ao veículo logo a frente, aplicação conhecida como *Adaptive Cruise Control*.

Na fase conflitante, o motorista sai de sua margem de segurança, entrando em uma situação que pode resultar em um acidente. Nesta fase os sistemas de segurança procurarão alertar ao motorista que este está em uma situação de risco ou intervirão no automóvel para tirá-lo desta situação, como por exemplo um alerta sonoro para que o motorista reduza a aceleração do veículo. Separando a fase conflitante da fase de mitigação existe o ponto sem retorno, a partir do qual um acidente não pode mais ser evitado, apenas a sua gravidade pode ser amenizada ou mitigada, caracterizando a fase de mitigação.

Na fase de mitigação os sistemas de segurança tentarão evitar maiores complicações no acidente, como por exemplo disparando air-bags para diminuir o impacto sofrido pelos ocupantes do veículo. Após as consequências do acidente inevitável terem sido amenizadas, o motorista e veículo encontram-se na seção pós-accidente, na qual os sistemas de segurança podem comunicar-se com as agências responsáveis pelo resgate, informando-as o local, gravidade do acidente e situação prévia dos ocupantes do veículo, auxiliando o socorro às vítimas.

Uma das grandes dificuldades enfrentadas pelo sistemas de segurança é identificar e tratar os inúmeros eventos que podem criar uma situação de risco. Por conta disso, a linha do tempo de segurança também pode ser seccionada em fases de gerenciamento do risco, gerenciamento da ameaça e gerenciamento do dano (CUI *et al.*, 2016). Com essa divisão, a linha do tempo pode servir para a definição de utilização de barreiras de segurança. Estas barreiras

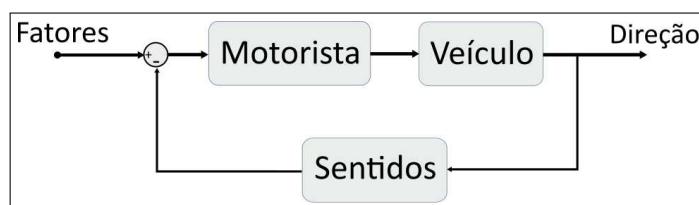
são utilizadas pelos sistemas de segurança para identificar quando devem atuar, com o intuito de impedir que o motorista e seu veículo avancem na linha do tempo, mantendo-se em situações de menor risco.

Apesar da criação de barreiras ser uma abordagem interessante para definir a atuação de sistemas de segurança, esta não consegue abranger toda e qualquer possível situação que um automóvel possa encontrar. Isso porque, o sistema automotivo é um sistema em tempo real, complexo, distribuído e influenciado por uma infinidade de fatores relacionados não só ao motorista, mas à integridade do veículo dirigido e ao ambiente que cerca o sistema. Por conta disso, os fatores que influenciam neste sistema são discutidos na próxima seção.

2.3 Sistema Automotivo

Quando um motorista entra em um carro, parte o motor e começa a se locomover, dando-se início ao processo de tomada de decisão conhecido como sistema automotivo. O sistema automotivo pode ser comparado a um sistema de controle, como é ilustrado na Figura 3. Assim como o controlador tem a função de atuar em uma planta ou processo de acordo com o estímulo de entradas, sua configuração pré-estabelecida e a resposta obtida através de sensores, o motorista tem como função atuar no veículo de acordo com o estímulo de entradas, as quais podem ser chamadas de fatores, sua experiência e a resposta informada pelos sensores do veículo e percebida pelos seus sentidos. Quando o motorista controla o veículo, ele o está dirigindo. Dessa forma, para que se possa definir o sistema automotivo também é importante discutir-se a definição de direção e os fatores que a influenciam, da mesma forma que se discute um sistema de controle e suas entradas.

Figura 3 – O sistema automotivo como um sistema de controle



Fonte – Elaborado pelo Autor.

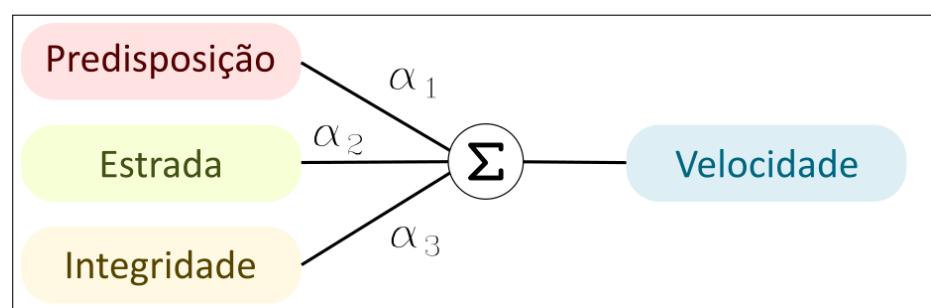
2.3.1 A Direção e seus fatores

O ato de dirigir é tido como um complexo processo de tomada de decisão (CHHABRA; VERMA; KRISHNA, 2017), do qual se procura encontrar um modelo ou descrição macroscópica que inclua todos os motoristas (ZHANG; ZHAO; RONG, 2014), apesar de tal singularidade ainda não ter sido encontrada.

A complexidade do processo de direção surge na determinação de quais fatores devem ser levados em conta e como estes influenciam individualmente sua definição. Por exemplo, o perfil de velocidade de determinado motorista é influenciado por diferentes fatores. A predisposição à velocidade de cada motorista é importante para descrever esse perfil, porém este é também influenciado pelo tipo de estrada no qual o veículo circula e pela integridade de suas partes mecânicas, como pode ser visto na Figura 4. Nela, observamos de forma simplificada que os pesos α_1 , α_2 e α_3 representam a intensidade com a qual cada fator influencia o perfil de velocidade de um motorista, sendo que a quantização da influência de cada um destes é incerta, dada a complexidade da análise.

Esta complexidade é tamanha que alguns estudos procuram unicamente modelar as influências que levam ao motorista escolher determinada velocidade. Esta abordagem inclui a criação de modelos matemáticos que abarquem diferentes características do motorista, como sua idade e sexo (SADIA; BEKHOR; POLUS, 2018), ou através da influência da geometria e condição da estrada (XU; SHU; SHAO, 2018).

Figura 4 – Composição do perfil de velocidade de determinado motorista



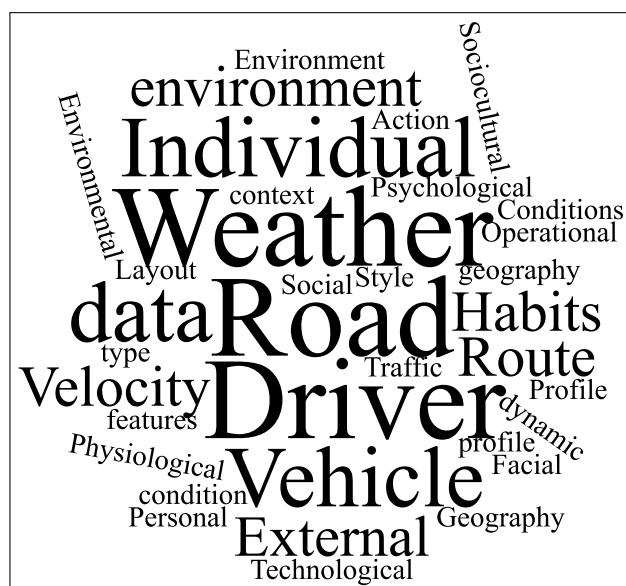
Fonte – Elaborado pelo Autor.

Devido à esta complexidade na escolha de fatores, a literatura deste assunto mostra-se bastante diversa, com diferentes trabalhos optando por separar estes fatores em categorias que dependem do objetivo pretendido pelo autor. LU *et al.* (2016), com uma abordagem de modelagem matemática, mencionam que a caracterização do motorista pode ser definida

por quatro conjuntos principais de fatores, os quais são (i) perfil de velocidade; (ii) hábitos de direção; (iii) geografia da rota percorrida; e (iv) clima do ambiente. Já SAGBERG *et al.* (2015), procurando ver o motorista sob um olhar pessoal, tomam a descrição da caracterização do motorista mais a fundo em sua construção psicológica, afirmando que seus hábitos são derivados de características (i) individuais; (ii) socioculturais e (iii) tecnológicas. Por outro lado, BOLOVINOU *et al.* (2014) julgam que determinado perfil de direção apenas pode ser totalmente descrito levando em conta o ambiente no qual o motorista se encontra, enumerando a estrada, o tráfego e a interação com outros motoristas elementos principais nesta construção. Para uma melhor visualização deste conjunto de fatores, um sumário de diversos trabalhos é apresentado na Tabela 2. Nesta, as exatas expressões empregadas por cada um dos autores são utilizadas.

Trabalhar com o grande número de categorias expostas na Tabela 2 seria confuso. Para resolver este problema, os fatores relatados podem ser divididos em categorias maiores, procurando-se um ponto comum entre todas aquelas descritas. Analisando-as, observa-se que as que aparecem com mais frequência são *Driver* (Motorista), *Road* (Estrada), *Weather* (Clima), *Individual* e *Vehicle* (Veículo), como é ilustrado graficamente na Figura 5. Dentre estas, as categorias Motorista e Individual são semelhantes, pois referem-se ao condutor do veículo. Além disso, as categorias Clima e Estrada são ambas características do ambiente no qual o sistema automotivo se encontra. Dessa forma, a análise pode ser simplificada através da divisão dos fatores nas categorias (i) Motorista; (ii) Veículo e (iii) Ambiente.

Figura 5 – Nuvem de palavras indicando quais categorias são as mais utilizadas na literatura.



Fonte – Elaborado pelo Autor.

Tabela 2 – Fatores categorizados e a sua descrição

Autor	Categoria de fatores	Relaciona-se com
(FERNANDEZ; ITO, 2016)	<i>External Personal Action</i>	estrada; ambiente; condições do clima gênero; experiência na direção e estado físico velocidade e aceleração em determinados eventos
(SAGBERG <i>et al.</i> , 2015)	<i>Individual Sociocultural Technological</i>	habilidade; personalidade e influência biológica Contexto social e nacional do motorista Tecnologia do veículo
(JACHIMCZYK <i>et al.</i> , 2018)	<i>Style Environmental</i>	velocidade; aceleração; curvas e instabilidade luz natural; clima e tipo de estrada e sua condição
(ZHANG; ZHAO; RONG, 2014)	<i>Individual</i>	características da direção e hábitos de direção
(ERICSSON, 2000)	<i>External Conditions Driver</i>	tipo de estrada; número de faixas e condições do tráfego características do motorista
(BOLOVINOU <i>et al.</i> , 2014)	<i>Road environment Individual</i>	disposição da estrada; condições do tráfego e ações de motoristas próximos experiência e comportamento atrás do volante
(LU <i>et al.</i> , 2016)	<i>Velocity Profile Driver Habits Route Geography Weather Environment</i>	velocidade e seu padrão aceleração tipo de estrada e condição do tráfego condições do clima
(ARAÚJO <i>et al.</i> , 2012) (SHARMA <i>et al.</i> , 2015)	<i>Vehicle data</i>	velocidade; RPM; aceleração e consumo
(CASTIGNANI; FRANK; ENGEL, 2013)	<i>Vehicle data</i>	velocidade e aceleração
(DÖRR; GRABENGIESSER; GAUTERIN, 2014)	<i>Vehicle data Road type</i>	velocidade e aceleração terra; urbana; rural; rodovia
(LIN <i>et al.</i> , 2014)	<i>Facial features Physiological Psychological Operational Driver dynamic</i>	movimento da cabeças e expressões Conforto performance homem-máquina direção anormal e comportamento na direção sistema Motorista-Veículo-Ambiente
(LU <i>et al.</i> , 2016)	<i>Velocity profile Driver Habits Route geography Weather environment</i>	velocidade ao longo do dia aceleração e velocidade distância; elevação; semáforos e curvas condicionadores de ar e clima
(VAIANA <i>et al.</i> , 2014)	<i>Road condition Weather</i>	atraito e instabilidade nível de umidade
(BRAMBILLA; MASCETTI; MAURI, 2017)	<i>Road Layout Traffic Weather Social context</i>	tipo de estrada; curvas interação com outros carros condições do clima características pessoais

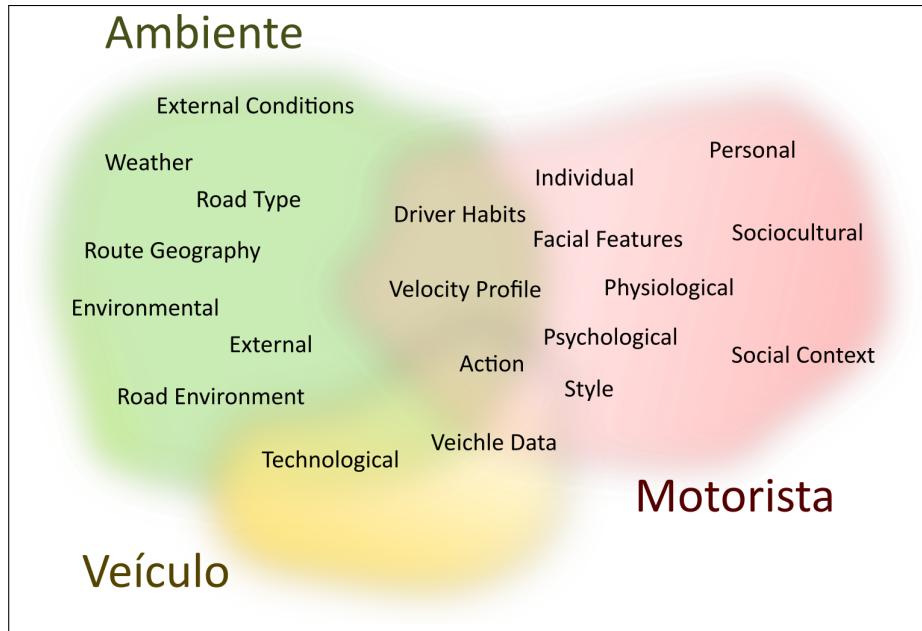
Fonte – Elaborado pelo Autor

Nota – Esta tabela apresenta categorias de fatores nomeados por diferentes autores. Note que tais categorias são utilizadas na problematização do problema proposto, porém nem todos as utilizam em processos de classificação, já que diferentes categorias correspondem a diferentes abordagens.

Apesar desta separação, cada fator não compõe uma única categoria específica, estas são, na verdade, influenciadas com diferente intensidade ou relevância por cada um dos fatores. Esse comportamento é representado na Figura 6, que enumera e divide cada um dos fatores, mas também ilustra a influência destes em cada categoria através de uma gradual mudança de coloração. Essa análise holística possibilita a descrição de uma forma mais definida e sucinta as variáveis que influenciam o ato de dirigir. Agrupando todos os fatores da Tabela 2 constrói-se, então, as definições das categorias apresentadas na Tabela 3, as quais serão utilizadas a partir de

então neste trabalho.

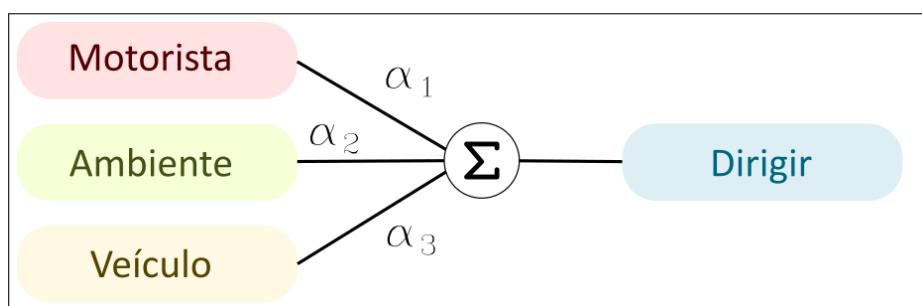
Figura 6 – Representação das categorias que dividem os fatores que influenciam o ato de dirigir.



Fonte – Elaborado pelo Autor.

Dessa forma, definimos o ato de dirigir como influenciável por estas categorias de fatores, nos fornecendo uma visão macroscópica da construção de determinada ação tomada por um motorista. Podemos, então, retomar o exemplo inicial do perfil de velocidade ilustrado na Figura 4, e generalizá-lo para construirmos uma nova representação do ato de dirigir. Essa representação é ilustrada na Figura 7, na qual observa-se que o ato de dirigir é influenciado pelos grupos de fatores pertencentes ao Ambiente, Veículo e Motorista. Assim, definimos *direção* como o ato de estar no comando de um veículo, sendo tal ato influenciado por fatores de diferentes intensidades referentes ao (i) Ambiente, no qual o ato é realizado; (ii) Veículo, com qual o ato é realizado e (iii) Motorista, o sujeito da ação.

Figura 7 – Representação de como o ato de dirigir é construído.



Fonte – Elaborado pelo Autor.

Tabela 3 – Fatores categorizados e a sua descrição

Categoría	Descrição	Constituída por
Ambiental	Essa categoria resume todos os fatores que são relacionados ao ambiente externo, isto é, que não dependam da interação Motorista-Veículo.	<ul style="list-style-type: none"> • Interações com a estrada (tipo, condição, distância, inclinação, sinalização, atrito, curvas e redutores de velocidade). • Interações com o clima (sua condição, visibilidade, nível de ruído, pressão do ar, temperatura e umidade). • Interações com outros veículos (nível de tráfego e ações ou proximidades de veículos ao redor do motorista).
Motorista	Essa categoria é relacionado unicamente ao motorista e a razão de suas ações sem considerar influências externas, apenas a predisposição pessoal de cada pessoa.	<ul style="list-style-type: none"> • Personalidade (como esta foi construída durante a vida do motorista, incluindo fatores socioculturais, psicológicos e fisiológicos). • Estado (como o motorista está se sentindo ou agindo no atual momento, cansaço, distrações, influência de químicos e stress). • Perfil de direção (como a velocidade, aceleração, arranque, consumo de combustível e outras características cinéticas são tratadas pelo motorista).
Veículo	Esta categoria descreve o quanto o veículo influencia na interação Motorista-Veículo.	<ul style="list-style-type: none"> • Tecnologia (como o carro auxilia o motorista, controle de direção adaptativo, alertas e conforto do automóvel). • Integridade (estado do veículo, problemas e mal funcionamentos que este possa apresentar, integridade de suas partes e como estes fatores influenciam as ações do motorista).

Fonte – Elaborado pelo Autor

2.3.2 A importância do motorista

A construção da definição de *dirigir* é um importante passo para entendermos com mais clareza a abordagem que é dada quando se avalia este ato. Entretanto, tal termo não é constantemente utilizado sozinho na literatura. Isso ocorre pois é preferido um conjunto de termos mais próximo do sujeito da ação - o motorista - já que, como discutido no Seção 2.2, este é o componente mais importante do sistema automotivo. Assim, termos como *Comportamento do Motorista* e *Estilo de Direção* prevalecem em textos de outros autores, que apesar de derivados da mesma ideia, aquela trabalhada na Seção 2.3.1, são utilizados sob diferentes pretextos, divergindo em significado ou aplicação. Logo, assim como as categorias de fatores, tais termos também não possuem uma definição comum na literatura.

Em alguns trabalhos (MEIRING; MYBURGH, 2015; SAGBERG *et al.*, 2015) *Comportamento do Motorista* é especificado como um conjunto de ações de direção praticadas pelo motorista, enquanto *Estilo de Direção* é abordado como uma subclassificação do anterior. Em outros (DÖRR; GRABENGIESSER; GAUTERIN, 2014) explica-se que *Estilo de Direção* e *Tipo de Direção* são termos intercambiáveis, definindo-se o primeiro como uma descrição mais generalista das ações do motoristas. Alguns autores relacionam o termo *Comportamento do Motorista* a diferentes eventos, como aceleração, desaceleração e velocidade do veículo (ALJAAFREH; ALSHABATAT; AL-DIN, 2012; FERNANDEZ; ITO, 2016; ARAÚJO *et al.*, 2012), ou ao estado atual do veículo e movimentos de cabeça e expressões faciais do motorista (LIN *et al.*, 2014).

CHHABRA; VERMA; KRISHNA (2017) argumentam que as características únicas de cada motorista são conhecidas como *Estilo de Direção*, ponto com o qual WITT *et al.* (2018) concordam, afirmando que idade, gênero, experiência em direção, personalidade, fadiga e stress são importantes na definição deste termo. FENG *et al.* (2018) focam na definição de *Estilo de Direção*, definindo-o como as escolhas de manobras feitas pelo motorista, as quais consistem em suas ações. Este pensamento também é seguido por ELANDER; WEST; FRENCH (1993), explicando que este mesmo termo (*Estilo de Direção*) refere-se ao modo como uma pessoa prefere ou habitualmente dirige, e por VAIANA *et al.* (2014) que o define como o modo com o qual o motorista dirige. Para facilitar a visualização de todas essas diferentes definições, a Tabela 4 é montada.

Algumas ideias destoam-se das apresentadas, como em ZHANG; ZHAO; RONG (2014), trabalho no qual o termo *Status de Direção* é utilizado referindo-se a *Comportamento do Motorista*, ou em LU *et al.* (2016), no qual os termos *Padrão de Direção* e *Ciclos de Direção* são utilizados para especificar as atitudes tomadas por um motorista e o seccionamento destas, respectivamente. Estes e outros termos menos utilizados na literatura foram organizados na Tabela 5.

SAGBERG *et al.* (2015) apresentam uma pesquisa relacionando o *Estilo de Direção* ao *Comportamento do Motorista*, incluindo neste estudo definições e diferenças entre os termos. De acordo com este autor, o *Estilo de Direção* é definido como uma "forma habitual de dirigir, a qual é característica de um motorista ou de um grupo de motoristas", enquanto *Comportamento do Motorista* é definido como "todas as ações que o motorista executa durante o ato de dirigir"e, por causa disso, o *Estilo de Direção* é uma subcategoria do *Comportamento do Motorista*.

Tabela 4 – Definição de termos com base em diferentes autores

Termo	Descrição
<i>Comportamento do Motorista</i>	<ul style="list-style-type: none"> Todas as ações que o motorista executa durante o ato de dirigir (SAGBERG <i>et al.</i>, 2015). Ações tipicamente executadas pelo motorista em diferentes situações, tais como aceleração e desaceleração em pistas inclinadas (FERNANDEZ; ITO, 2016). Características operacionais escolhidas pelo motorista que melhor se encaixam em determinada condição de direção(LIN <i>et al.</i>, 2014). Padrões comportamentais refentes a como o motorista se porta em determinadas situações, como aceleração e desaceleração do veículo(SUN <i>et al.</i>, 2017). Comportamento do motorista da estrada, o qual pode ser analisado para criação de um modelo do motorista, permitindo que este seja categorizado(MIYAJIMA <i>et al.</i>, 2007). Acúmulo de informações pelo motorista, o qual acaba desenvolvendo um hábito de direção, e estes hábitos resumem-se ao comportamento do motorista(HWANG <i>et al.</i>, 2018).
<i>Estilo de Direção</i>	<ul style="list-style-type: none"> Forma habitual de dirigir, a qual é característica de um motorista ou de um grupo de motoristas(SAGBERG <i>et al.</i>, 2015). Definição generalista das ações tomadas por um motorista (DÖRR; GRABENGISSER; GAUTERIN, 2014). Como o motorista escolhe realizar habitualmente manobras com o veículo (FENG <i>et al.</i>, 2018). Modo como a pessoa dirige habitualmente ou por preferência(ELANDER; WEST; FRENCH, 1993). Relacionado a como o motorista dirige, dependendo de sua experiência e habilidade(VAIANA <i>et al.</i>, 2014). Síntese do processamento mental e tendência comportamental relativamente estável, crônica e intrínseca do motorista(SUN <i>et al.</i>, 2017).

Fonte – Elaborado pelo Autor

Tabela 5 – Definição de termos menos utilizados na literatura

Termo	Descrição
<i>Tipo de direção</i>	Parte constituinte do <i>Estilo de Direção</i> , englobando razões para dirigir, situações e condições de direção (DÖRR; GRABENGISSER; GAUTERIN, 2014).
<i>Intenção do Motorista</i>	Conjunto formado por uma série de pequenos <i>Comportamentos do Motorista</i> (LIN <i>et al.</i> , 2014).
<i>Padrão de Direção</i>	Constituído por fatores que modelam como as variáveis do veículo se comportam, como consumo de combustível e condição do motorista (LU <i>et al.</i> , 2016)
<i>Hábito de Direção</i>	Ação sistemática efetuada pelo motorista desenvolvida pela sua aplicação ao longo do tempo (HWANG <i>et al.</i> , 2018)

Fonte – Elaborado pelo Autor

MARTINEZ *et al.* (2018), com outro pensamento, argumentam que o *Estilo de Direção* é classificado como o local onde o processo de tomada de decisão de todo o sistema Motorista-Veículo-Ambiente ocorre, enquanto *Comportamento do Motorista* é uma ação particular do motorista que pode ocorrer após ter sido tomada a decisão da ação, isto é, o *Comportamento do Motorista* é uma caracterização do *Estilo de Direção*.

Como pode-se ver, SAGBERG *et al.* e MARTINEZ *et al.* apresentam duas bem argumentadas, porém distintas, definições dos termos *Comportamento do Motorista* e *Estilo do Motorista*. Com base nessas bem estruturadas discussões e nas argumentações propostas pela literatura (Tabelas 4 e 5), pode-se então conceber uma definição que delimita a utilização dos termos para que estes sejam melhor aproveitados por este trabalho. Define-se, então, *Comportamento do Motorista* como um conjunto de ações cuja decisão foi tomada pelo motorista dado os estímulos recebidos por fatores pertencentes as categorias Motorista, Veículo e Ambiente. Já *Estilo do Motorista* é definido como a forma com a qual uma dada ação do motorista é executada, dependendo dos mesmos fatores pertencentes as categorias Motorista, Veículo e Ambiente. Caso, por exemplo, em determinada situação um motorista decida que é uma boa opção realizar uma ultrapassagem, essa decisão é característica de seu comportamento, já a forma com a qual essa ultrapassagem será executada, incluindo sua velocidade e agressividade, são características do estilo do motorista.

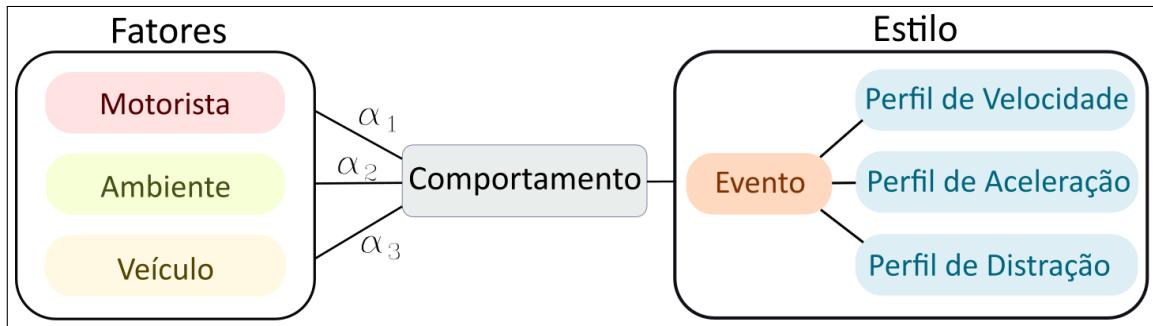
Além destes termos, existem outros que merecem ser esclarecidos. *Eventos de Direção* são eventos disparados pelo motorista dado um determinado comportamento escolhido, como por exemplo os eventos de aceleração e desaceleração experienciados em uma ultrapassagem. *Perfis de direção* são as dinâmicas desempenhadas por determinados parâmetros durante a ocorrência de um evento, como por exemplo o perfil de velocidade ao longo de uma ultrapassagem.

Dessa forma, pode-se finalmente refinar a ideia trabalhada desde o início do capítulo, obtendo a Figura 8. Nela, é ilustrada uma representação do sistema automotivo, no qual o processo de tomada de decisão do motorista, isto é, a forma como ele vai se comportar, é influenciado por fatores pertencentes as categorias Motoristas, Ambiente e Veículo. Dada a decisão tomada, o motorista executará a ação escolhida de determinada maneira, enfatizada pelo seu estilo de direção.

Como discutido, o motorista é geralmente relacionado ao seu *Comportamento* e *Estilo*. Quando se busca uma modelagem do sistema automotivo, prevendo como este agirá frente a diferentes ações, estamos nos referindo ao *Comportamento*, isto é, o bloco central do diagrama da Figura 8. Quando se busca uma classificação do *Comportamento*, o qual é expressado em *Perfis de Direção*, nos referimos ao *Estilo*, que compõe as consequências do comportamento escolhido.

Com a definição da construção qualitativa do sistema automotivo, percebe-se a im-

Figura 8 – Representação da relação entre os termos que definem Direção.



Fonte – Elaborado pelo Autor.

portância do motorista neste sistema, o qual desempenha um papel central de tomada de decisão, influenciado por diversos fatores. Apesar destes fatores terem sido descritos qualitativamente na Seção 2.3.1, deve-se buscar representações destes em variáveis que possam ser capturadas ou calculadas e, então, serem utilizadas para a descrição quantitativa do sistema automotivo, ideia trabalhada no próximo capítulo.

3 SISTEMA AUTOMOTIVO: INFORMAÇÕES E SISTEMAS DE CAPTURA

3.1 Introdução

Os fatores pertencentes as categorias Motorista, Ambiente e Veículo, discutidas na Seção 2.3.1, podem influenciar de diferentes maneiras o comportamento e estilo do motorista, como foi discutido na Seção 2.3.2. A grande questão é como obter estes valores e a intensidade com a qual estes influenciam as definições anteriormente destacadas. Diversos autores possuem diferentes respostas e abordagens para esta questão, variando tanto na forma de obtenção das informações necessárias, como a forma com a qual estas são tratadas. Portanto, o objetivo deste capítulo é discutir as diversas abordagens para seleção de variáveis relacionadas aos fatores e sua captura que são encontradas na literatura. Também neste capítulo discute-se o sistema de diagnóstico automotivo OBDII, um dos meios de aquisição de informações do sistema automotivo.

3.2 Sensores e Sinais

Os fatores que influenciam o sistema automotivo já foram discutidos, porém também é importante relacioná-los as variáveis que possam ser capturadas e utilizadas em um processo de modelagem (comportamento) ou classificação (estilo). Dessa forma, as variáveis relacionadas as categorias de fatores Motorista, Veículo e Ambiente são discutidas a seguir.

3.2.1 *Motorista*

As variáveis relativas ao Motorista são relacionadas à sua personalidade, estado físico e psicológico, e estilo de direção, como discutido na Tabela 3. Dentre estas três divisões, o estilo de direção mostra ser o meio mais prático de obtenção de informações, já que este é composto por variáveis como velocidade e aceleração, as quais podem ser obtidas de diferentes maneiras, tais como sensores introduzidos pelo usuário ou embarcados em smartphones e no próprio carro.

Além disso, no tocante à classificação do estilo de um motorista, nada faz mais sentido do que utilizar para tal variáveis oriundas de ações do próprio condutor. Por conta disso, grande parte da literatura que trata deste tipo de classificação faz, de alguma forma, uso de fatores pertencentes à categoria Motorista e relacionados ao seu estilo, como pode ser visto na

Tabela 6. Esses fatores incluem aceleração, velocidade, rotações por minuto, ângulo do volante, ângulo de abertura do pedal de aceleração, carga no motor e marcha atual do veículo.

Pesquisas que utilizam o estado do motorista para classificação de seu estilo são mais escassas, já que estes fatores são pessoais e diferentes em cada motorista, sendo de difícil determinação. Uma solução seria utilizar veículos instrumentados com sistemas de detecção destes fatores, como câmeras e medidores de sinais vitais (KAPLAN *et al.*, 2015). Outra abordagem seria a utilização de um questionário respondido pelo próprio motorista, conhecido como *Driver Behaviour Questionnaire* (DBQ), desenvolvido para relacionar determinados comportamentos do motorista ao risco de acidentes na direção (REASON *et al.*, 1990). Perguntas também podem ser feitas ao motorista, inquirindo como ele avalia sua direção antes e depois dos testes (WITT *et al.*, 2018).

Estes métodos já se provaram válidos na determinação das características que aumentam a probabilidade de comportamentos que apresentam maior risco de segurança (HELMAN; REED, 2015) e estão, inclusive, sendo aperfeiçoados com medidas de relação entre as informações do DBQ (ROWE *et al.*, 2015), com o intuito de simplificar e potencializar este método. Além disso, são empregados conjuntamente com outras dados, tais como acidentes (HASSAN *et al.*, 2017) e fatores ambientais (SADIA; BEKHOR; POLUS, 2018), para complementar análises complexas, como a determinação do estilo de direção relacionado à velocidade.

Outras abordagens incluem a observação do ato de dirigir por terceiros dentro do veículo (AMADO *et al.*, 2014), em ambientes de simulação (CHEN; FANG; TIEN, 2013; MEULENERS; FRASER, 2015) ou por meio de veículos instrumentados. Porém, esse método influencia o processo de decisão do motorista, já que este sabe que está sendo monitorado, causando mudanças comportamentais no mesmo. Os dados naturalistas, que serão discutidos na Seção 3.3, apesar de onerosos e morosos, provam-se menos invasivos, pois há tempo suficiente para o motorista se acostumar com a observação.

Apesar de não comumente utilizadas em processos de classificação do estilo de direção, o estado do motorista pode ser utilizado como peso na classificação de estilos semelhantes. Por exemplo, caso dois motoristas possuam o mesmo perfil de velocidade, aquele que desvie o olhar da estrada por um maior tempo ou que comete um maior número de erros e ações perigosas terá seu estilo de direção classificado como mais agressivo do que o outro.

Tabela 6 – Variáveis relacionadas aos fatores da categoria Motorista.

Relacionados ao Estilo de Direção	Autores
Aceleração latitudinal e longitudinal	(CASTIGNANI; FRANK; ENGEL, 2013),(SHARMA <i>et al.</i> , 2015)
Aceleração e Velocidade	(DÖRR; GRABENGIESSER; GAUTERIN, 2014)
Pedal de aceleração e Volante	(BOLOVINOU <i>et al.</i> , 2014)
Diferentes variáveis eletromecânicas	(MARTINELLI <i>et al.</i> , 2018),(WITT <i>et al.</i> , 2018)
Velocidade, Aceleração, Freio, Volante e Marcha	(ZHANG; ZHAO; RONG, 2014)
Velocidade, Aceleração e Freio	(VAIANA <i>et al.</i> , 2014)
Velocidade, RPM, Aceleração, Volante, Pedais, Marcha e Freio	(BOUHOUTE <i>et al.</i> , 2018)
Pedal de aceleração e Freio	(SUN <i>et al.</i> , 2017)
Velocidade, RPM e Carga no Motor	(NIRMALI <i>et al.</i> , 2017)
Velocidade, Aceleração, Carga no Motor, RPM, Temperatura e Tensão na Bateria	(HWANG <i>et al.</i> , 2018)
Relacionados ao Estado do Motorista	Autores
Idade, Experiência e Ansiedade	(WITT <i>et al.</i> , 2018)
Expressões faciais, características físicas e psicológica	(LIN <i>et al.</i> , 2014)

Fonte – Elaborado pelo Autor

3.2.2 Ambiente e Veículo

Variáveis relacionadas ao ambiente são mais comuns do que aquelas referentes ao estado e personalidade do motorista. Geralmente englobam o clima e luminosidade do ambiente no qual o motorista dirige; o tipo e estrutura da estrada e a relação do motorista com outros veículos, como pode ser visto na Tabela 7. Em contrapartida, os fatores relacionados ao veículo sempre são relacionados à alguma das outras categorias, como foi discutido na Figura 6. Os estudos de classificação do estilo geralmente não consideram fatores como (i) o modelo, (ii) atual integridade das peças, (iii) histórico de mal funcionamentos e (iv) eficiência do motor. Estes fatores poderiam ser obtidos, por exemplo, com um *Vehicule Integrity Questionnaire* (VHQ) semelhante ao aplicado aos motoristas (DBQ). Essa ausência é explicada pelo aumento da complexidade que seria causado pela adição de mais fatores, os quais necessitariam de novos estudos para relacioná-los ao estilo do motorista.

3.3 Sistemas para obtenção de dados

Com a determinação de quais variáveis estão relacionadas aos fatores que influenciam o sistema automotivo, deve-se quais métodos podem ser utilizados para obtê-las. Por conta disso, esta seção trata dos sistemas para obtenção de dados comumente utilizados por outros autores.

Tabela 7 – Variáveis relacionadas aos fatores da categoria Ambiente.

Relacionados à estrada	Autores
Tipo de estrada	(DÖRR; GRABENGIESSER; GAUTERIN, 2014),(LU <i>et al.</i> , 2016)
Estrutura da Estrada	(BOLOVINOU <i>et al.</i> , 2014),(BRAMBILLA; MASCETTI; MAURI, 2017)
Impacto da estrada e pneus no atrito	(VAIANA <i>et al.</i> , 2014)
Relacionados ao Clima	Autores
Clima em geral	(BRAMBILLA; MASCETTI; MAURI, 2017)
Impacto do clima no atrito	(VAIANA <i>et al.</i> , 2014)
Luminosidade e visibilidade	(VIROJBOONKIATE; VATEEKUL; ROJVIBOONCHAI, 2017)
Relacionados a outros motoristas	Autores
Condições do Tráfego e Comportamento dos motoristas próximos	(BOLOVINOU <i>et al.</i> , 2014),(LU <i>et al.</i> , 2016)

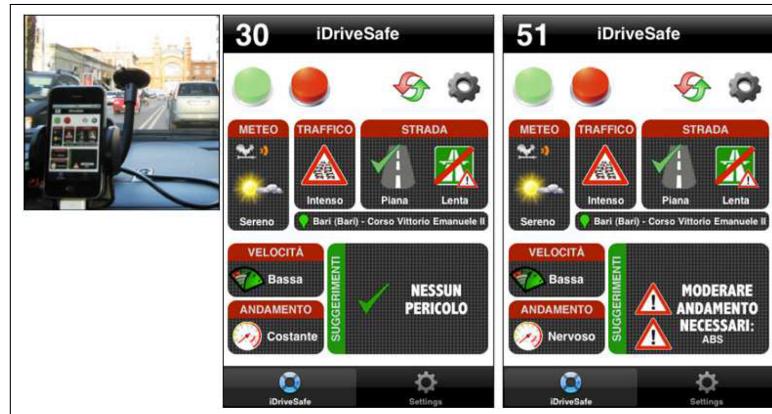
Fonte – Elaborado pelo Autor

3.3.1 Smartphones

Alguns autores procuram simplificar a captura de informações, utilizando o mínimo de sensores possível ou apenas aqueles que estejam disponíveis. Isso é muito comum em aplicações que utilizam o smartphone como ferramenta de captura, já que este é uma plataforma bastante difundida atualmente, contando com acelerômetro, magnetômetro e GPS em sua estrutura (CASTIGNANI *et al.*, 2015; SHARMA *et al.*, 2015; VAIANA *et al.*, 2014). Assim, ao invés da necessidade de compra e instalação de equipamentos no veículo, um simples aplicativo seria o suficiente para criação de um dispositivo de captura, o qual já conta, inclusive, com uma saída de vídeo para interação com o usuário (a tela do smartphone). Um exemplo desta aplicação pode ser vista na Figura 9, a qual ilustra o aplicativo *iDriveSafe* (RUTA *et al.*, 2010).

O grande problema dessa solução é o fato do smartphone ser um sistema instável e constantemente manipulado, o qual pode mudar de posição a qualquer momento, criando ruídos que interferem na captura de dados. Uma simples mudança na posição do aparelho, causados pelo veículo ou pelo próprio usuário, já exige um novo processo de calibração de seu acelerômetro (CASTIGNANI *et al.*, 2015). Ruídos também estão presentes no não tão robusto magnetômetro desta plataforma, já que campos magnéticos externos o influenciam. Por isso, alguns autores recorrem aos menos ruidosos dados do GPS. Além disso, por ser um dispositivo

Figura 9 – O *iDriveSafe*, um exemplo de aplicativo para classificação de motoristas.



Fonte – Retirado de (RUTA *et al.*, 2010)

pessoal não exclusivo para captura, sua não padronização pode refletir em divergências nos dados, já que os sensores e suas características são diferentes entre modelos. Por conta disso, a utilização de smartphones é útil e prática, mas possui limitações na quantidade de sensores e na qualidade dos dados capturados. Em um cenário padronizado, como na utilização de frotas de veículos de uma mesma empresa, no qual variáveis como o posicionamento e modelo do aparelho possam ser padronizadas, essa abordagem ganha mais força.

3.3.2 Simuladores Instrumentados

Apesar de sua rede de sensores, somente agora com o advento dos *Intelligent Transport System* (Sistemas de Transporte Inteligentes, ITS) e carros autônomos, os dados gerados por veículos viraram um ponto de interesse comum para a comunidade científica. Além disso, por serem objetos pessoais, catalogar dados de determinado veículo também envolve questões éticas e morais. Por conta disso, o número de bancos de dados considerados naturalistas é escasso. Naturalista refere-se ao fato das informações capturadas corresponderem ao ambiente natural dos motoristas, adquiridas pelo monitoramento contínuo de automóveis e sem intervenções de um pesquisador. Um exemplo muito famoso deste tipo de estudo é o *The 100-car naturalistic driving study* (DINGUS *et al.*, 2006), elaborado pela NHTSA, órgão pertencente ao Departamento de Transporte dos Estados Unidos. Nele, 100 carros foram instrumentados para fornecer informações sobre o veículo e motorista de maneira não intrusiva, contendo mais de 43.000 horas de informações capturada durante aproximadamente 13 meses, incluindo variáveis cinéticas, como velocidade, e captura de vídeos do rosto do motorista, interior e exterior do veículo.

Outro exemplo é o *Driver Distractions in Commercial Vehicle Operation*, financiado

e aplicado pela *Virginia Tech Transportation Institute* (OLSON *et al.*, 2009), que possuiu como objetivo coletar dados naturalistas de frotas de caminhões. Cem motoristas participaram e nove caminhões instrumentados de forma semelhante ao *The 100-car naturalistic driving study* foram utilizados. Cerca de 14.500 horas de dados foram coletadas, perfazendo mais de 1 milhão de quilômetros rodados. Estudos desse tipo dispendem de um grande investimento com material e pessoal para instalação, manutenção do sistema de captura, recuperação dos dados periodicamente e posterior processamento de todo o material.

Logo, adquirir dados para determinados eventos específicos apenas presentes em dados naturalista, como os de acidentes, e que possam ser usados na classificação do estilo de motoristas é uma atividade restrita. Uma outra opção é a utilização de simuladores instrumentados, no qual o motorista dirige em uma pista virtual, capaz de recriar situações reais vivenciadas pelos motoristas mas das quais não se tem muita informação sobre sua dinâmica, como eventos repentinos, batidas, mal funcionamento do veículo, embriaguez e stress (WITT *et al.*, 2018; LIN *et al.*, 2014). Alguns exemplos de implementação de tais simuladores envolvem (i) classificação de estilos de motoristas (LÓPEZ; PINILLA *et al.*, 2012), como pode ser visto na Figura 10 ; (ii) Determinação do estado mental do motorista, com medições de sinais biológicos e movimentos dos olhos (IIZUKA; SAITO; YAMADA, 2015), ilustrado nas Figuras 11 e 12; e (iii) Auxílio para ajuste do comportamento de cada motorista, com o crescimento do uso dos DAS (OGITSU; MIZOGUCHI, 2015). Estas soluções conseguem abranger diferentes situações, motoristas e estados destes de forma muito eficaz, tudo ao alcance de uma simples mudança de parâmetros da simulação. Porém, a disponibilidade e o custo de desenvolvimento deste método torna-o restrito a determinados grupos de pesquisadores. Além disso, apesar de próximos da realidade, não há como garantir que os motoristas envolvidos na simulação agem da mesma forma que em uma situação real.

3.3.3 Softwares como ferramentas de simulação

A utilização de softwares como ferramentas de simulação é uma abordagem próxima da abordagem com simuladores instrumentados. No entanto, menor fidelidade dos dados ao comportamento real do motorista é o preço a ser pago pelo menor custo envolvido e maior flexibilidade encontrada nessa solução. Softwares como o licenciado CarMaker¹, o qual pode simular o carro e seu motorista como um sistema de controle, incluindo variáveis mecânicas; ou

¹ <https://ipg-automotive.com/products-services/simulation-software/carmaker/>

Figura 10 – O *Racer Simulator*, um exemplo de simulador de direção.



Fonte – Retirado de (LÓPEZ; PINILLA *et al.*, 2012)

Figura 11 – Exemplo da tela de um simulador de direção.



Fonte – Retirado de (IIZUKA; SAITO; YAMADA, 2015)

Figura 12 – As condições dos experimentos em um simulador de direção.

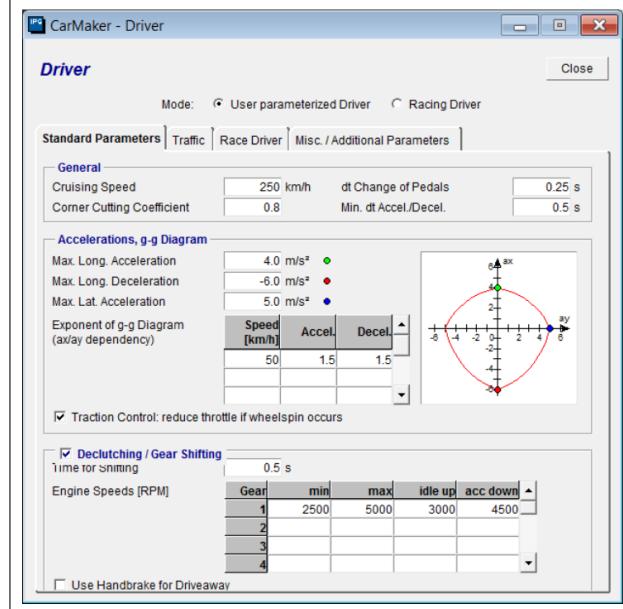


Fonte – Retirado de (IIZUKA; SAITO; YAMADA, 2015)

o open-source e flexível SUMO², o qual permite simulações com um grande número de veículos, são utilizados por inúmeros trabalhos com o intuito de classificação. A Figura 13 mostra uma aplicação do CarMaker parametrizando o comportamento de um motorista, já a Figura 14 mostra uma aplicação na qual o motorista foi substituído pela entrada de um usuário, transformando o sistema em um simulador de direção instrumentado mais simples do que os discutidos na seção anterior. A Figura 15 apresenta uma estrutura de estrada que pode ser criada com o SUMO.

² <http://sumo.dlr.de/index.htm>

Figura 13 – Parametrização de um motorista simulado em software pelo CarMaker.



Fonte – Retirado de (OLOFSSON; PETTERSSON, 2015)

Figura 14 – Motorista como um usuário no software CarMaker.

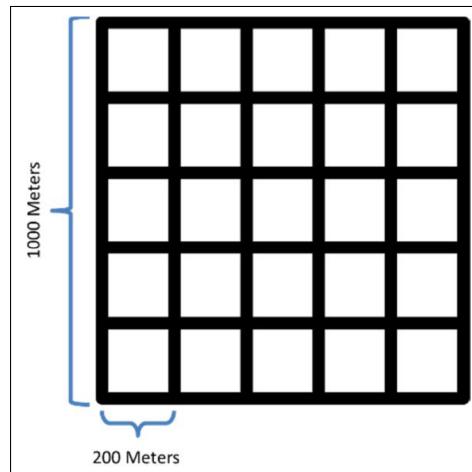


Fonte – Retirado de (ZHAO; ZHAO; TIAN, 2018)

3.3.4 Sistemas Embarcados

O desenvolvimento de sistemas embarcados específicos para o automóvel procura estabelecer um meio termo entre as soluções anteriores. Dependendo da complexidade do sistema embarcado este pode realizar o processamento de informações utilizando recursos do próprio sistema ou pode enviar tais informações para um ponto de maior poder de processamento, o que caracteriza um sistema embarcado de telemetria isto é, trata-se de um sistema com captura local e processamento remoto. Esta última abordagem diminui a complexidade do sistema de

Figura 15 – Estrutura de uma estrada desenhada no software SUMO.



Fonte – Retirado de (KU *et al.*, 2014)

captura e apresenta maior flexibilidade com a manipulação de dados, já que esta é realizada em um ponto com maior quantidade de recursos.

Os sistemas embarcados geralmente são utilizados para capturar dados do sistema eletromecânico do veículo através do OBDII ou de módulos de sensores integrados ao próprio sistema. A primeira opção necessita de um maior entendimento do sistema eletromecânico do veículo e, por conta disso, exige uma explicação mais aprofundada, como será feito na Seção 3.4.

O desenvolvimento de um sistema embarcado e inserção deste em um veículo possui um custo inerente menor do que o de um simulador instrumentado. Apesar de não possibilitar capturas em ambiente controlado de batidas, eventos repentinos e direções com embriaguez e cansaço, por motivos de segurança, estes sistemas permitem que o motorista participe de uma dinâmica real, dando maior confiabilidade aos dados capturados. Além disso, o mesmo sistema pode ser utilizado futuramente para capturas naturalistas.

Quando comparado ao smartphone, o sistema embarcado exige maior investimento financeiro e maior tempo gasto com o desenvolvimento da plataforma. Entretanto, a maior capacidade de processamento e por permitir a escolha dos sensores utilizados, o que também significa uma padronização, os dados capturados são de maior qualidade, sofrendo menor influência de ruídos. Para melhor ilustrar as diferenças aqui comentadas a Tabela 8 foi elaborada, na qual podemos observar as principais disparidades entre as soluções mais utilizadas pela literatura: smartphones, simuladores físicos, ferramentas de simulação e sistemas embarcados. Já a Tabela 9 traz exemplos de aplicação destas soluções encontradas na literatura.

Apesar da possibilidade de inserção de sensores próprios em sistema embarcados,

Tabela 8 – Comparaçao entre as soluções para captura de informação.

Característica	Simuladores Instrumentados	Softwares de Simulação	Sistemas Embarcados	Smartphones
Custo Total	●●●	●○○	●●○	●○○
Fidelidade dos dados	●●○	●○○	●●●	●●○
Quantidade de dados	●●●	●●○	●●●	●○○
Diversidade de eventos	●●●	●●○	●●○	●●○
Presença de Ruído	●○○	●○○	●○○	●●●

Fonte – Elaborado pelo Autor

Tabela 9 – Sistemas de captura utilizados por outros trabalhos.

Dados	Autores
Smartphone	(CASTIGNANI <i>et al.</i> , 2015),(VAIANA <i>et al.</i> , 2014),(SHARMA <i>et al.</i> , 2015), (GĂVRUTĂ; MARCU; BOGDAN, 2018),(NIRMALI <i>et al.</i> , 2017)
Simulador Instrumentado	(LIN <i>et al.</i> , 2014),(WITT <i>et al.</i> , 2018),(LÓPEZ; PINILLA <i>et al.</i> , 2012)(IZUKA; SAITO; YAMADA, 2015)(OGITSU; MIZOGUCHI, 2015)
Software de simulação	(DÖRR; GRABENGIESSER; GAUTERIN, 2014),(ZHANG; ZHAO; RONG, 2014),(SUN <i>et al.</i> , 2017)
Sistema Embarcados	(MARTINELLI <i>et al.</i> , 2018),(FENG <i>et al.</i> , 2018),(BOUHOUTE <i>et al.</i> , 2018),(HWANG <i>et al.</i> , 2018), (VIROJBOONKIADE; VATEEKUL; ROJVIBO-ONCHAI, 2017)
Banco de dados de terceiros	(BRAMBILLA; MASCETTI; MAURI, 2017)

Fonte – Elaborado pelo Autor

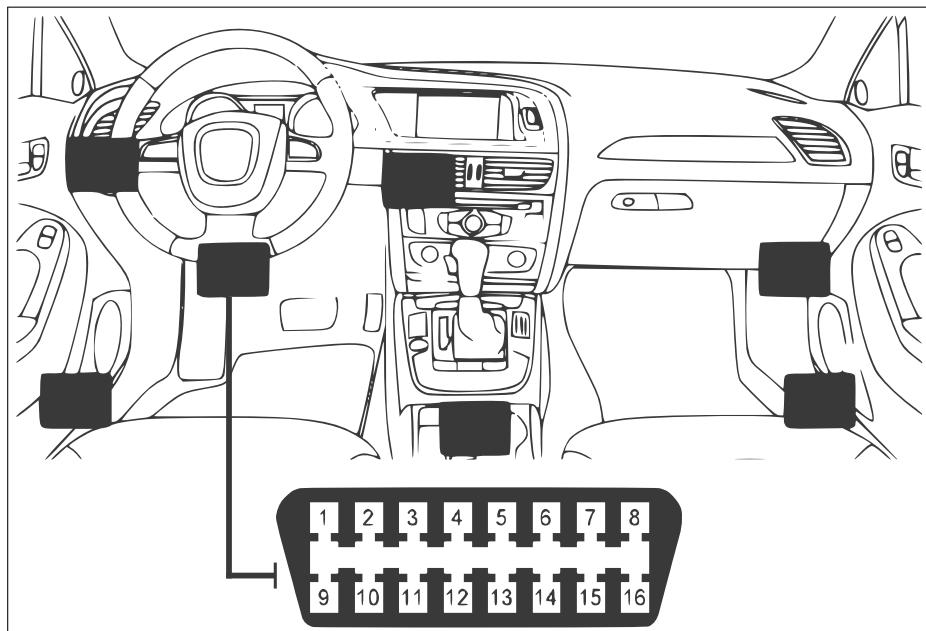
como acelerômetros, magnetômetros e GPS (JACHIMCZYK *et al.*, 2018; DERBEL *et al.*, 2015), uma das abordagens bastante interessantes presentes na literatura é a utilização da já extensa rede de sensores que o veículo possui, necessária para que o mesmo possa gerenciar seu funcionamento. Essa rede de sensores pode ser acessada por um agente externo através do sistema de diagnóstico automotivo OBDII, fornecendo um leque de informações dos mais diversos sistemas eletromecânicos do automóvel. Por conta disso, essa abordagem é discutida mais a fundo na próxima seção.

3.4 On Board Diagnostic II

Para que se possa receber informações do veículo, este deve possuir uma rede de comunicação interna capaz de monitorar seus sistemas eletromecânicos e entender as requisições realizadas por um agente externo. Isto é possível graças ao sistema de diagnóstico automotivo OBDII, definido pela TSBEC (2004) como um padrão de *Data Link Conector* (Conector de Enlace de Dados, DLC) com conexão e pinos especificados para cada função, protocolos eletrônicos, códigos de diagnósticos de problemas e listas de parâmetros a serem monitorados ou armazenados.

As especificações do OBDII incluem a padronização do conector e da função de seus pinos. O conector, que deve estar posicionado até 0,61m do volante, é composto por duas colunas de oito pinos, como pode ser visto na Figura 16, na qual as possíveis localizações do conector são ilustradas. Já a Tabela 10 exemplifica a função de cada pino presente na camada física do sistema, evidenciando que os pinos possuem diferentes papéis dependendo do protocolo utilizado e de opções implementadas pelo fabricante.

Figura 16 – Possíveis locais de instalação do terminal OBDII e sua pinagem.



Fonte – Elaborado pelo Autor.

Tabela 10 – Pinagem do terminal OBDII e sua descrição.

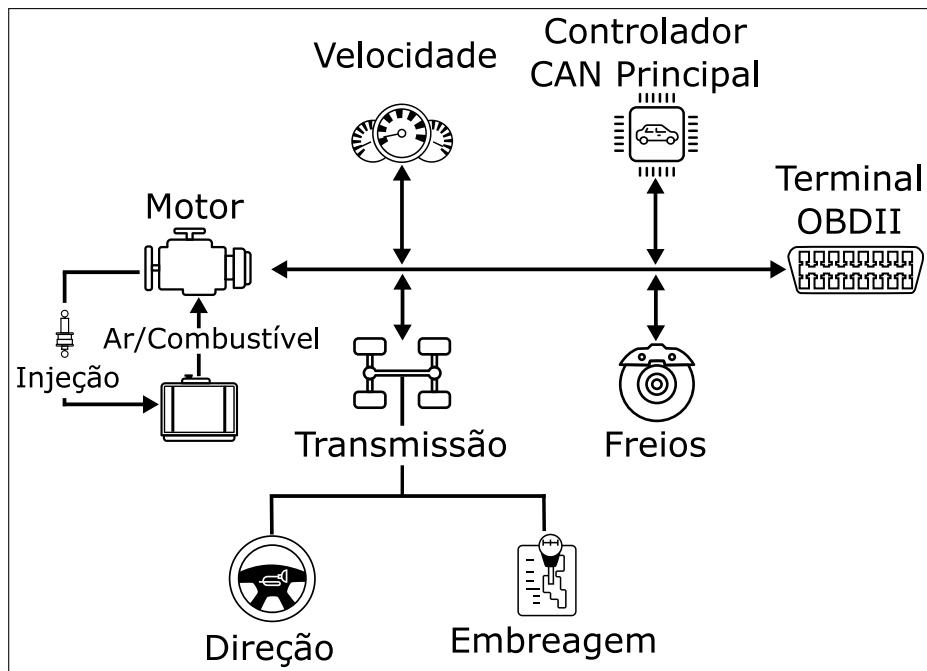
Pino	Descrição	Pino	Descrição
1	Opção do vendedor	9	Opção do vendedor
2	SAE J1850 Bus+	10	SAE J1850 Bus-
3	Opção do vendedor	11	Opção do vendedor
4	Terra do Chassis	12	Opção do vendedor
5	Terra do Sinal	13	Opção do vendedor
6	CAN (ISO 15765-4 e SAE J2284) High	14	CAN ISO (15765-4 e SAE J2284) Low
7	ISO 9141-2 K-Line	15	ISO 9141-2 L-Line
8	Opção do vendedor	16	Bateria do automóvel

Fonte – Adaptado de obdii.com

Para ser capaz de monitorar os sistemas eletromecânicos do veículo, o OBDII aproveita-se da rede de comunicação já existente no interior do automóvel, a qual é responsável

por transportar dados do complexo e distribuído sistema automotivo, composto por até 60 micro-controladores, nomeados de *Engine Central Unit* (ECU), e que possui, aproximadamente, 2500 sinais de diferentes tipos (STEINBACH; KORF; SCHMIDT, 2011). Tais redes de comunicação podem ser construídas com a utilização de diferentes protocolos, já que cada fabricante de automóvel opta por uma diferente abordagem. Exemplos destes protocolos são o (i) SAE J1850 VPW (SAE, 1988), utilizado principalmente pela General Motors (GM); (ii) ISO9141 (ISO, 1994) adotado pela montadora Chrysler e a maioria das montadoras asiáticas; e (iii) J1950 PWM (SAE, 1988), amplamente utilizada pela Ford Motors. Entretanto, a *Controller Area Network* (CAN), desenvolvida pela Bosch, é uma das expoentes mais famosas, sendo escolhida como protocolo padrão para todos os veículos fabricados a partir de 2008³. A Figura 17 ilustra como os diferentes ECU de cada sistema do carro trocam informações e interagem utilizando a CAN. Um terminal OBDII é ilustrado na extremidade direita, representando a possibilidade de um dispositivo externo conectar-se à rede e, consequentemente, aos ECUs para obtenção de informação destes.

Figura 17 – Representação da distribuição da CAN em um automóvel.



Fonte – Elaborado pelo Autor.

Caso queira requerer algum dado ao sistema OBDII, o dispositivo compatível conectado ao terminal do carro deve enviar comandos de diagnóstico, conhecidos como *Diagnostic*

³ <http://obdii.com/>

Trouble Codes (Códigos de Diagnóstico de Problemas) (DTC), ou parâmetros de informações, conhecidos como *Parameter Identification* (PID), os quais, assim como os protocolos discutidos anteriormente, são padronizados (SAE, 2012; ISO, 2015). Os PIDs são divididos em um total de 10 serviços, ilustrados na Tabela 11. Cada um destes possui uma série de códigos que podem ser utilizados para atividades específicas. O serviço \$01, por exemplo, contém mais de 200 PIDs utilizados para requisitar informações em tempo real dos sistemas eletromecânicos do veículo, como velocidade e quantidade de combustível. O serviço \$03 apresenta códigos de problemas salvos pelo próprio sistema OBDII quando este identifica uma falha nos sistemas eletromecânicos do veículo, os quais podem ser resgatados para diagnosticar problemas no automóvel. Outro serviço interessante é o \$09, capaz de fornecer informações como o número do chassi do veículo, nome da ECU do veículo e efetuar a calibração de alguns protocolos.

Tabela 11 – Os dez serviços disponíveis no sistemas OBDII.

Serviços	Descrição
\$01	Apresenta dados atuais
\$02	Apresenta dados congelados
\$03	Apresenta os DTCs salvos
\$04	Limpia os DTCs salvos
\$05	Resultados do teste do Sensor de Oxigênio ¹
\$06	Resultados do teste de outros componentes ¹
\$07	Apresenta DTCs pendentes (atual ou último ciclo de direção)
\$08	Controle de operação dos sistemas e componentes de bordo
\$09	Requisita informações do veículo
\$0A	DTCs permanentes (limpos em comandos anteriores)

Fonte – Adaptado de obdii.com.

Cada PID enviado ao OBDII é identificado e respondido pelo sistema na forma de um conjunto de hexadecimais, o qual possui tamanho variável e diferente forma de decodificação dependendo da informação solicitada. No PID **010D**, exemplificado na Figura 18, os dois primeiros dígitos (**01**), os quais chamaremos de cabeçalho, informam que este pertence ao serviço \$01, isto é, requisição de informação em tempo real, já os dois últimos (**0D**) indicam que a informação requerida é a velocidade. A resposta dada pelo OBDII a este PID é um conjunto de 3 hexadecimais, como por exemplo, **410D28**. Os quatro primeiros dígitos que formam o cabeçalho (**410D**), assim como os do PID enviado, indicam que a resposta pertence ao serviço \$01 e é relacionada à velocidade. Já os dois últimos (**28**), os quais são informação útil, podem ser diretamente convertidos em decimal para obter-se a velocidade atual do veículo em Km/h, a qual, nesse caso, é dada por 40 Km/h.

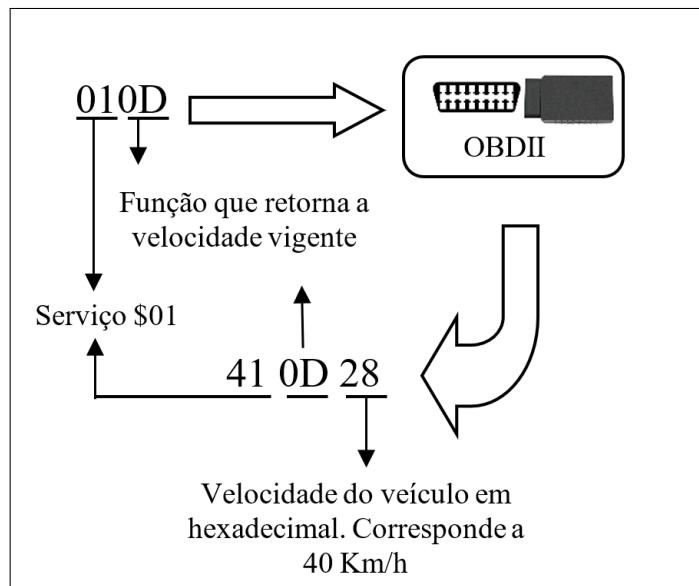
O maior número de hexadecimais com informação útil que uma resposta pode possuir é quatro. Logo, caso utilizemos **ABCD** para representar a informação útil das possíveis respostas dadas pelo OBDII e **abcd** para seus respectivos valores decimais, pode-se montar a Tabela 12 que exemplifica a decodificação de algumas informações que podem ser obtidas através do serviço \$01.

Tabela 12 – Exemplo de decodificação da informação recebida do OBDII.

PIDs	Informação requisitada	Cálculo
010C	Rotações por Minuto	$\frac{256a+b}{4}$
010D	Velocidade	a
0110	Fluxo de Massa Ar-Combustível	$\frac{256a+b}{100}$
0111	Aceleração	$\frac{100}{255}a$
0104	Carga no Motor	$\frac{100}{255}a$

Fonte – Elaborado pelo Autor.

Figura 18 – Exemplo de captura de informação utilizando o terminal OBDII.



Fonte – Elaborado pelo Autor.

4 COMUNICAÇÃO COM O VEÍCULO

4.1 Introdução

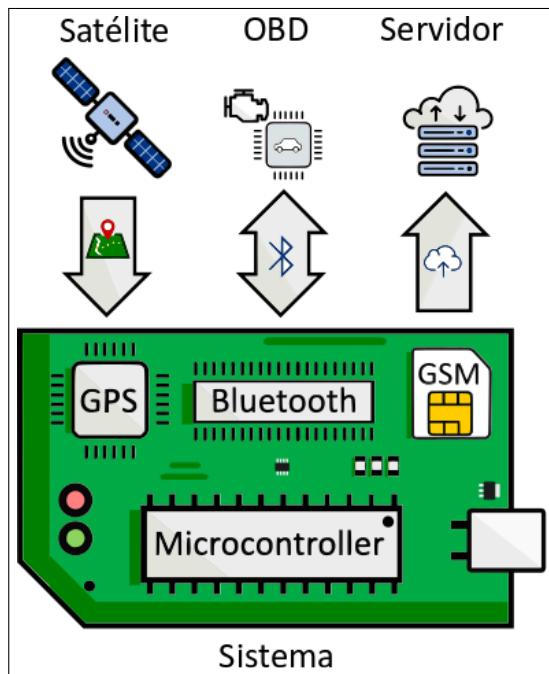
Como foi visto no Capítulo 3, as variáveis relacionadas ao Motorista-Veículo são as mais utilizadas para descrever o sistema automotivo. Estas podem ser capturas de diversas maneiras, como por um sistema embarcado de telemetria (a partir de agora referido apenas como sistema de telemetria), que apresenta uma boa relação entre custo de desenvolvimento, qualidade/quantidade dos dados capturados e flexibilidade no tratamento de informações. Além disso, quando aliado ao OBDII, a possibilidade da captura de diversas informações relacionadas aos sistemas eletromecânicos do veículo fazem o sistema de telemetria uma solução ainda mais interessante. Dessa forma, este capítulo procura detalhar o desenvolvimento do sistema de telemetria autoral que é utilizado no processo de captura dos dados do sistema automotivo.

4.2 Sistema de Telemetria

Para capturar e salvar as informações do sistema automotivo optou-se pelo desenvolvimento de um sistema de telemetria conectado à internet, capaz de comunicar-se com o sistema de diagnóstico OBDII de um veículo, para captura de dados dos sistemas eletromecânicos do mesmo. O sistema desenvolvido é capaz de enviar requisições pertencentes a qualquer um dos serviços do OBDII (\$01 - \$10). As repostas obtidas das requisições são enviadas pelo sistema a um servidor online através da rede de telefonia móvel. As informações são, então, organizadas pelo servidor em um banco de dados online. Além disso, o uso de um módulo compatível com o protocolo GPS permite a identificação e localização das capturas, o que adiciona ainda mais variáveis ao processamento. A Figura 19 apresenta um esquemático do sistema desenvolvido, que evidencia todas as características discutidas.

Apesar de já existirem opções no mercado que supram as necessidade de captura de dados do OBDII, optou-se por projetar um sistema próprio, já que o controle do fluxo de dados e modificações do sistema se dão de forma livre em servidores e protótipos próprios. As próximas seções discutem o desenvolvimento de hardware e software para este sistema. Um dos pontos importantes desse desenvolvimento é a forma com a qual o sistema desenvolvido se conectará ao terminal OBDII discutido na Seção 3.4. Para esta conexão, utilizou-se uma plataforma comercial conhecida como ELM327.

Figura 19 – Sistema de telemetria desenvolvido



Fonte – Elaborado pelo Autor

4.2.1 Desenvolvimento do hardware

4.2.1.1 ELM327

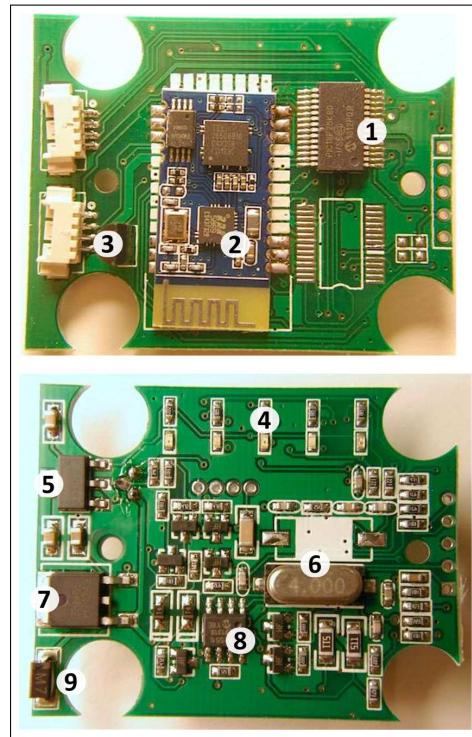
Devido aos protocolos de comunicação utilizados pelo OBDII e da necessidade de uma conexão física ao terminal, optou-se por utilizar uma pequena interface de comunicação desenvolvida pela empresa ELM Electronics, a qual possui o ELM327¹ como microcontrolador principal, baseado no PIC18F2480. Tal dispositivo funciona como uma ponte entre as requisições realizadas por um usuário e as respostas obtidas pelo padrão OBDII, já que possui suporte a todos os protocolos discutidos na Seção 3.4. Para isso, ele possui uma interface Universal Asynchronous Receiver/Transmitter (UART), capaz de comunicar-se serialmente com qualquer dispositivo compatível. Nesse caso, tal interface comunicar-se-á com um transmissor Bluetooth integrado ao circuito da mesma. A Figura 20 ilustra de forma mais clara a placa de circuito impresso dupla face na qual o ELM327 é montado. Nela podemos observar o módulo Bluetooth que possui o controlador Beken BK3231², representado pelo número 2, responsável pela compatibilidade do ELM327 com o protocolo Bluetooth RFCOMM. Assim, basta que um sistema possua suporte a tecnologia Bluetooth para que uma conexão entre o mesmo e a interface

¹ <https://www.elmelectronics.com/ic/elm327/>

² <https://oss.aliyuncs.com/netmarket/f49abd45-0b36-460b-a160-e836d9da18b2.pdf>

OBDII seja concretizada por intermédio do ELM327.

Figura 20 – Faces da placa de circuito impresso que contém o ELM327.



Fonte – Adaptado de amazon.com.

Tabela 13 – Componentes da placa de circuito impresso que contém o ELM327.

Número	Componente
1	ELM327 (PIC PIC18F25K80)
2	Beken BK3231
3	LM317
4	Leds de funcionamento
5	AMS1117 (semelhante ao LT117)
6	Cristal 4 MHZ
7	Regulador 78M05
8	MCP2551 CAN Transceiver
9	M7 (smd semelhante ao diodo 1N4007)

Fonte – Elaborado pelo Autor.

A Figura 20 também ilustra o próprio ELM327 e alguns componentes que dão suporte à tarefa por ele desempenhada, como os Leds de indicação de funcionamento e os reguladores AMS1117 e 78M05. A Tabela 13 complementa a Figura 20 definindo cada um dos elementos que compõem o restante da placa. Alguns componentes são importantes para o funcionamento do microcontrolador, como por exemplo o CI MCP2551, responsável pela

compatibilidade do ELM327 ao protocolo CAN. O interfaceamento com os outros protocolos (SAE J1850 PWM/VPW e ISO9141) é feito através da utilização de circuitos de componentes discretos, como indicado no datasheet do ELM327³ e cujo funcionamento detalhado foge ao objetivo deste trabalho. No restante deste trabalho a presença do ELM327 no terminal OBDII, mesmo que não citada, deverá ser subentendida.

Definido o método com o qual o sistema de telemetria pode se comunicar com o OBDII, deu-se início ao desenvolvimento do sistema em si, discutido nas próximas seções.

4.2.1.2 Microcontrolador

Com o intuito de projetar um sistema economicamente acessível, de dimensões reduzidas e estável, para integrá-lo ao ambiente do veículo, optou-se pela utilização do Microcontrolador (Microcontroller, MCU) PIC 18F4550⁴, fabricado pela Microchip e ilustrado na Figura 21. Este MCU suporta os protocolos seriais UART, *Serial Peripheral Interface* (SPI) e *Inter-Integrated Circuit* (I2C), necessários para regular e comunicar-se com os outros dispositivos do projeto. Além disso possui um número considerável de pinos Entrada/Saída, uma grande quantidade de memória de programa e de dados e funciona com uma frequência de cristal elevada (48MHz). Estes fatores garantem a capacidade de processamento necessária no sistema desenvolvido, já que se espera uma alta taxa de bits por segundo (bps) na captura.

Figura 21 – Microcontrolador PIC18F4550



Fonte – Microchip

4.2.1.3 Comunicação Bluetooth

O módulo HC-05, escolhido para o desenvolvimento das atividades, ilustrado na Figura 22, funciona como um transmissor e receptor Bluetooth, podendo ser controlado e configurado através de uma interface UART embutida no módulo. O HC-05 pode atuar como

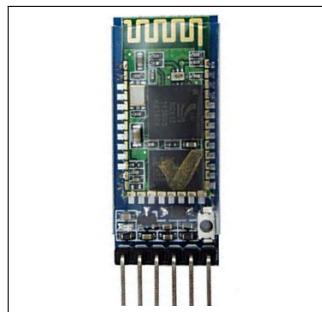
³ <https://www.elmelectronics.com/wp-content/uploads/2016/07/ELM327DSF.pdf>

⁴ <https://www.microchip.com/wwwproducts/en/PIC18F4450>

mestre ou escravo da comunicação Bluetooth, bastando para isso a devida configuração via terminal serial. Além disso, este módulo permite selecionar o Media Access Control (MAC) ao qual é feita a conexão Bluetooth. Essa função é utilizada para que haja conexão apenas à interface da qual o ELM327 faz parte, impedindo que outros dispositivos Bluetooth tentem se conectar ao MCU.

Com o intuito de averiguar a funcionalidade do sistema, foi instalado um visor LCD 20x4 soldado ao Circuito integrado CI PCF8574, ambos representados na Figura 23, capaz de externar respostas e variáveis do sistema por meio de uma comunicação I2C. Como todos os dispositivos operam com a tensão de 5 volts não houve quaisquer problemas em sua interconexão, permitindo o trabalho com apenas uma linha de alimentação.

Figura 22 – Módulo Bluetooth Hc-05



Fonte – [amazon.com](https://www.amazon.com)

Figura 23 – Display LCD 20x4



Fonte – [amazon.com](https://www.amazon.com)

A escolha do HC-05 é justificada por (i) facilmente encontrado nas mais diversas lojas de material eletrônico para sistemas embarcados; (ii) economicamente viável, com preços compatíveis com os outros módulos do sistema; (iii) literatura disponível e acessível; e (iv) possibilidade de programação da dupla função (Mestre/Escravo).

4.2.1.4 Transmissão remota

O SIM 808 é um módulo desenvolvido pela empresa SIMCOM Wireless Solutions⁵ capaz de conectar-se à rede de comunicação móvel, também conhecida como *Global System for Mobile communications* (GSM), utilizando o protocolo *General Packet Radio Service* (GPRS). O mesmo ainda utiliza a tecnologia GPS para navegação via satélite. Dessa forma, tal módulo é capaz de realizar o rastreio do sistema e gerenciar o envio de dados. O SIM 808 utilizado, exposto na Figura 24, é integrado a uma placa que possibilita o acoplamento de antenas, para aumentar a confiabilidade da rede, e terminais para comunicação UART, os quais são a fonte de controle utilizada pelo microcontrolador.

O módulo segue os padrões da *National Marine Electronics Association* (NMEA), sendo capaz de requisitar alguns tipos de dados previstos por esta organização. No projeto, utilizou-se o *Recommended Minimum Data for GPS* (RMC), tipo no qual obtém-se a latitude, longitude e horário mundial do requerimento da forma mais compacta possível (HEWERDINE, 2005). A Figura 25 ilustra uma representação truncada dos dados coletados pelo GPS do SIM808, contendo a informação útil considerada para o projeto, a qual é especificada na Tabela 14. O SIM808 suporta *File Transport Protocol* (FTP), protocolo que será utilizado para enviar as informações coletadas em arquivos no servidor online.

Figura 24 – SIM 808



Fonte – amazon.com

⁵ <http://www.simcomm2m.com/En/module/detail.aspx?id=137>

Figura 25 – SIM 808

1	2	3	4	5	6
1	1	20170319152025	-5.051797	-42.817238	67.600

Fonte – NMEA

Tabela 14 – Especificações do vetor de informações recebido do GPS.

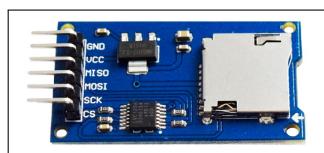
Número	Componente
1	GPS Ligado
2	Localização obtida com sucesso
3	Data e Hora (YYYY/MM/DD hh:mm:ss)
4	Latitude
5	Longitude
6	Altitude em metros

Fonte – Elaborado pelo Autor.

4.2.1.5 Armazenamento de dados

Para se ter segurança no caso de possíveis problemas no envio, integrou-se ao MCU um módulo de cartão de memória fabricado pela empresa CATALEX⁶, para armazenamento de informações em arquivos de backup. O módulo, apresentado na Figura 26, pode ser alimentado e comunicado junto aos outros componentes de 5 volts, pois possui um regulador de tensão interno e um conversor de nível lógico responsáveis por entregar ao cartão 3,3 Volts. Com a utilização de comunicação SPI e a implementação da biblioteca adequada ao programa do controlador, iniciou-se a captura de informação por longos períodos de tempo.

Figura 26 – Módulo Cartão de memória



Fonte – amazon.com

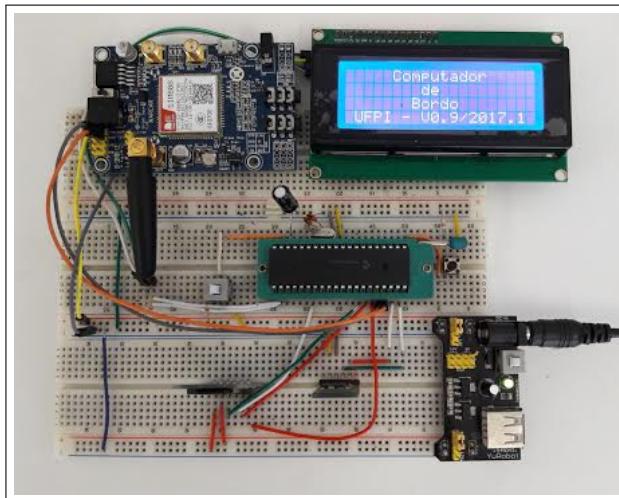
4.2.1.6 Esquemático e Protótipo

O sistema final prototipado utilizando os componentes descritos é apresentado na Figura 27. Com o intuito de apresentar de forma mais clara a conexão entre os dispositivos, um

⁶ <https://catalex.nz/>

esquemático do sistema criado utilizando o software Proteus⁷ é apresentado no Anexo A .

Figura 27 – Sistema de telemetria prototipado.



Fonte – Elaborado pelo Autor

4.2.2 Desenvolvimento do Software

Para o software, utilizou-se a linguagem de programação C, escolhida pela sua elevada relação performance/produtividade, e o Ambiente de Desenvolvimento Integrado (Integrated Development Environment, IDE) MikroC⁸, escolhido pela grande quantidade de bibliotecas e desenvolvedores que o mesmo possui. Dada a natureza estruturada da linguagem C, o funcionamento do sistema pode ser descrito em passos.

O módulo HC-05 e SIM808 possuem uma interface UART para comunicação com um dispositivo de controle externo, nesse caso o PIC18F4550. O módulo de cartão de memória também possui uma interface serial, entretanto do tipo SPI. Dessa forma, a biblioteca *SoftUART* e a biblioteca *MMC*, já incluídas no software MikroC, foram utilizadas para comunicação e controle dos módulos HC-05/SIM808 e cartão de memória, respectivamente.

O sistema é iniciado logo após a partida do motor do automóvel, sendo alimentado pela bateria do veículo. Primeiramente, a interface que possui o ELM327 prepara-se, após um *Power On Reset*, inicia sua rotina principal, identifica os protocolos suportados pelo carro e aguarda a conexão via Bluetooth. Já o PIC 18F4550 passa por uma rotina de inicialização, verificando se todos os componentes estão devidamente conectados, configura-os e certifica a

⁷ <https://www.labcenter.com/>

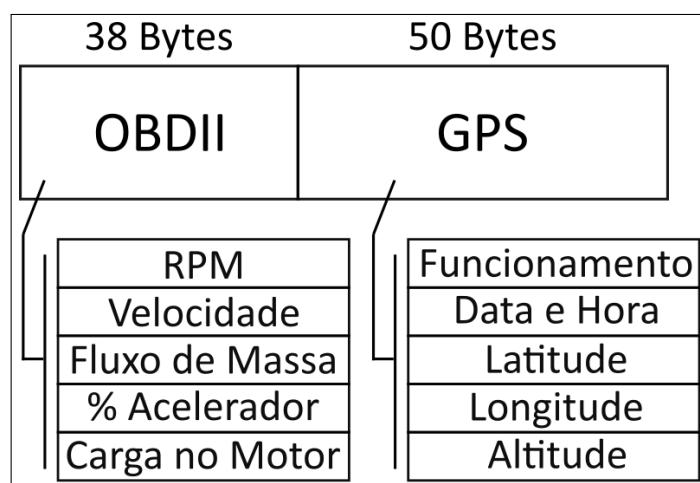
⁸ <https://www.mikroe.com/mikroc>

não existência de problemas de comunicação. Caso qualquer empecilho seja detectado, o mesmo será exibido no display LCD e um log de erros será salvo no sistema de armazenamento.

Enquanto o MCU é inicializado, o HC-05, funcionando como mestre na comunicação Bluetooth, procura pelo MAC do ELM327, o qual foi previamente configurado no módulo, e concretiza uma conexão. Com o vínculo estabelecido, dá-se início a rotina principal. Nesta, o Microcontrolador requisitará informações dos sistemas do veículo, utilizando-se do OBDII e por intermédio do ELM327. Cada ciclo dessa rotina é denominado de captura. Os PIDs desejados e as respostas esperadas estão implementadas no código do sistema. Assim, os parâmetros são enviados individualmente e as respostas são armazenadas em um vetor de caracteres.

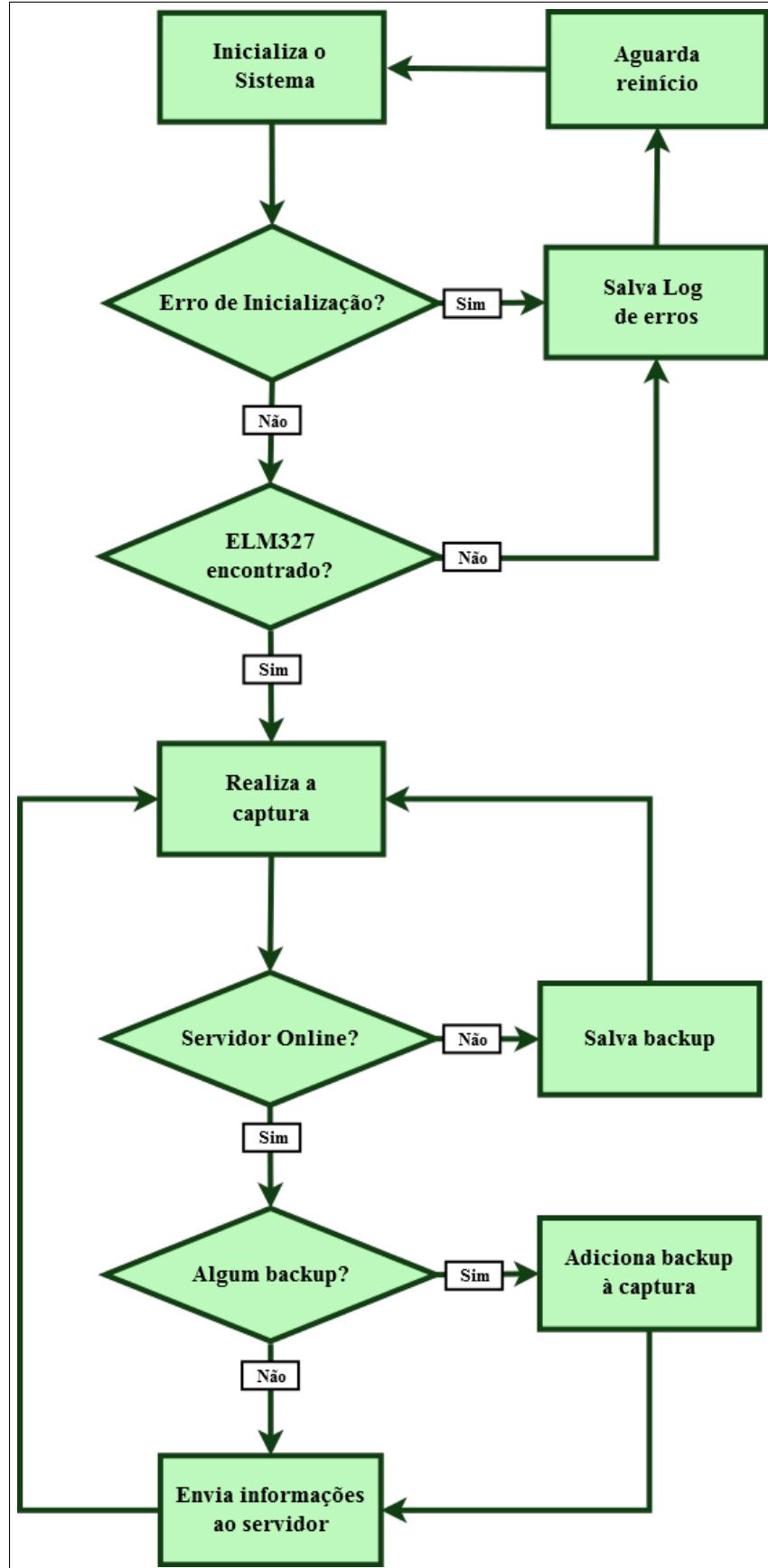
Em seguida, o MCU requisita os dados RMC do sistema através do SIM808, os quais são adicionados ao mesmo vetor anterior, gerando um pacote de completo de captura, exemplificado na Figura 28. Logo após, o sistema verifica se há sinal GSM. Caso positivo, o MCU, utilizando o SIM808 como transmissor, salva, em um arquivo previamente selecionado no servidor correspondente, os dados da captura. Caso não exista, o vetor é armazenado em um arquivo de texto no Cartão de memória, no qual cada linha corresponde a uma captura. Nos próximos ciclos, caso detectada a presença de sinal, o sistema passa a adicionar informação salva no backup à capturada, enviando pares de dados em cada ciclo, cujo horário está presente na captura, fato que impede má interpretação dos dados. Uma vez no servidor, as informações salvas no arquivo de texto podem ser utilizadas conforme desejado pelo usuário. Todo o processo de captura é apresentado no fluxograma da Figura 29.

Figura 28 – Representação de um pacote de captura enviado ao servidor pelo sistema.



Fonte – Elaborado pelo Autor

Figura 29 – Fluxograma de funcionamento do sistema



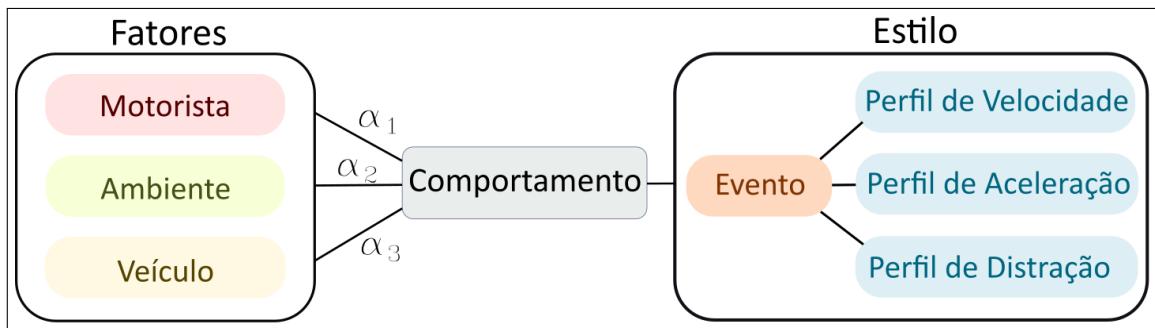
Fonte – Elaborado pelo Autor

5 SISTEMA AUTOMOTIVO: CLASSIFICAÇÃO DO MOTORISTA

5.1 Introdução

As informações adquiridas pelos diversos sistemas de captura, como o sistema de telemetria desenvolvido neste trabalho, possibilitam uma descrição qualitativa e quantitativa do sistema automotivo, o qual é recapitulado na Figura 30.

Figura 30 – Representação da relação entre os termos que definem Direção.



Fonte – Elaborado pelo Autor.

A descrição do sistema automotivo pode envolver a modelagem e previsão de atuação do sistema (comportamento) ou a classificação das respostas do sistema (estilo). A classificação de respostas pode ser utilizada como uma solução de segurança, com o intuito de alertar ao condutor sobre o seu estilo de direção, para que este possa ser evitado caso crie situações de risco à segurança. A ideia de classificação de estilos de direção já foi trabalhada na literatura e aplicada em diversas situações, inclusive na própria segurança no trânsito. Por conta disso, este capítulo procura discutir soluções para a classificação do estilo de direção de motoristas.

5.2 Algoritmos de Classificação e Clusterização

Os fatores discutidos na Seção 2.3.1, capturados com os sistemas descritos na Seção 3.3, são fundamentais para o processo de classificação do estilo de direção. Assim, a escolha do algoritmo de classificação deve ser feita considerando as características do sistema desenvolvido. De fato, diferentes algoritmos possuem diferentes aproveitamentos dependendo dos dados e objetivos com o qual este é construído. Como previamente discutido, o estilo do motorista é influenciado por fatores pertencentes à três principais categorias, porém também é importante perguntar quantos estilos de motorista existem, de forma que seja possível efetuar

uma classificação destes.

Classificações e clusterizações investigam o desenvolvimento de técnicas que permitem a construção de sistemas automáticos de aprendizado. Normalmente este sistema é construído em duas fases, conhecidas como (i) treinamento, onde é aprendida uma descrição geral dos conceitos; e (ii) teste, que utiliza a descrição aprendida para categorização de novos dados passados ao sistema. Dentre os principais tipos de aprendizado pode-se citar o supervisionado e não-supervisionado. No primeiro, a classe de dados é fornecida por uma fonte externa ao processo de aprendizado, muitas vezes por um especialista, isto é, o desenvolvedor do sistema. Já o aprendizado não-supervisionado utiliza a identificação de subconjuntos de dados que compartilham certas similaridades (REAL; NICOLETTI, 2014).

5.2.1 Clusterização

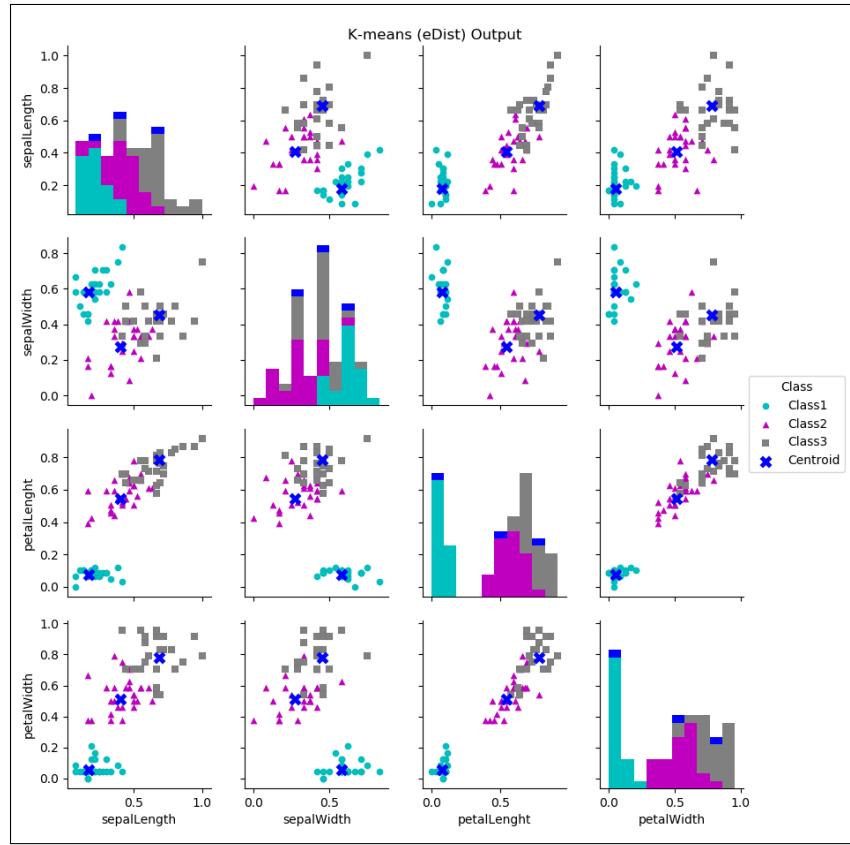
Clusterização é um sistema de aprendizado não-supervisionado baseado na organização de dados que seguem determinados padrões de similaridade em grupos com um fronteira definida, denominados clusters (JAIN; MURTY; FLYNN, 1999; HAVENS; BEZDEK; PALANISWAMI, 2013). A clusterização é tipicamente utilizada em situações nas quais se possui pouca informação da relação entre os dados coletados, sendo apropriada para analisar de forma preliminar estas relações (JAIN; MURTY; FLYNN, 1999). Datada do século passado, a clusterização é um campo em evolução, dado o entusiasmo da academia relacionado ao aprendizado de máquina, mineração de dados (SHIRKHORSHIDI *et al.*, 2014) e dados fracamente classificados ou não classificados (SCHWENKER; TRENTIN, 2014).

A Figura 31, por exemplo, apresenta clusters para os diferentes parâmetros que compõem o famoso banco de dados Iris¹. Esse banco de dados contém informações de dimensões das flores de diferentes espécies de planta pertencentes à família Iris. As informações podem ser utilizadas como entrada de um algoritmo para criação de clusters, cada um representando uma das três espécies contidas no banco de dados.

Por conta de suas características, a clusterização possui várias aplicações no sistema Motorista-Veículo-Ambiente. Os dados do DBQ podem ser utilizados para encontrar padrões no comportamento do motorista, de forma a classificá-los quanto ao critério da segurança (YUEPENG *et al.*, 2007). Padrões ou eventos podem ser clusterizados para tornar possível sua detecção, como em dados naturalistas, os quais necessitam de tratamento para identificar-se quais

¹ <https://archive.ics.uci.edu/ml/datasets/iris>

Figura 31 – Exemplos de dados clusterizados em diferentes classes.



Fonte – Elaborado pelo Autor.

entradas são úteis (LI; MISU; MIRANDA, 2016). Estes eventos podem então ser novamente clusterizados para identificação de estilos atribuídos ao motorista (HIGGS; ABBAS, 2015) ou de fatores que podem influenciá-lo, como as condições do tráfego (YANGUO; LUNHUI; YA, 2011) e a distração no trânsito (LIU *et al.*, 2015). A verificação da similaridade entre os fatores que possam influenciar comportamentos perigosos do motorista pode ser utilizada para determinar a importância de cada um na classificação (YUAN *et al.*, 2017).

O estilo dos motoristas pode ser classificado em diferentes grupos. Geralmente, os grupos são relacionados ao risco de acidentes que o comportamento do motorista representa ou ao seu estilo, considerando seu perfil de aceleração ou consumo do veículo. A Tabela 15 apresenta alguns exemplos de classes encontradas na literatura.

5.2.2 Classificação

O reconhecimento de padrão envolve, essencialmente, as etapas de processamento conhecidas como (i) pré-processamento, que tem o objetivo de eliminar dados desnecessários fazendo com que a entrada apresente apenas dados relevantes para o reconhecimento do objeto

Tabela 15 – Exemplos de classes de motoristas

Estilos	Autores
<i>Confortable, Normal e Sporty</i>	(DÖRR; GRABENGIESSER; GAUTERIN, 2014)
<i>Mild, Normal e Agres- sive</i>	(WANG; LUKIC, 2011)
<i>Defensive, Normal, Aggressive</i>	(ZHANG; ZHAO; RONG, 2014)
<i>Non-agressive e Agres- sive</i>	(VAIANA <i>et al.</i> , 2014)
<i>Low, Moderate e High risk</i>	(HIGGS; ABBAS, 2015)
<i>Moderate e Aggressive</i>	(WANG; XI, 2016)

Fonte – Elaborado pelo Autor

em análise; (ii) extração de parâmetros, também denominada etapa de seleção de atributos, consiste em análise dos dados de entrada a fim de extrair e derivar informações úteis para o processo de reconhecimento; e (iii) classificação, a etapa de tomada de decisão, na qual se analisa as características extraídas dos dados do objeto de entrada e o declara pertencente a uma determinada categoria ou classe (MATOS, 2008). Dessa forma a classificação pode ser considerada como a etapa na qual cria-se um modelo que procura se ajustar as características da informação extraída, o qual deve ser validado para comprovar sua eficácia na modelagem.

5.2.2.1 Validação da classificação

Para avaliar a qualidade do modelo produzido na etapa de classificação, verificando o quanto bem este se adequa ao conjunto de dados, alguns métodos podem ser utilizados. O banco de dados formado com os atributos selecionados pode ser utilizado completamente na validação do modelo, calculando-se o erro do modelo para cada ponto do banco a fim de obter-se a média do erro absoluto. Este método é conhecido como resíduos (PICARD; COOK, 1984).

Além da média do erro absoluto, existem outras formas de avaliar o desempenho do modelo criado, como a utilização de uma matriz de confusão. Na matriz de confusão, em especial com a utilização de classificadores, as classes de determinado conjunto de dados são utilizadas como rótulos das linhas e colunas de uma matriz, como pode ser visto na Tabela 16, na qual três classes são utilizadas.

Neste caso, além do acerto da classificação, os falso-positivos também são computados, isto é, os momentos nos quais o classificador classificou erroneamente uma determinada entrada. Dessa forma, dada as entradas pra uma determinada classe, a matriz de confusão contém

Tabela 16 – Exemplo de matriz de confusão

		Classe Verdadeira		
		Classe 1	Classe 2	Classe 3
Classe Estimada	Classe 1	75%	0.0%	0.0%
	Classe 2	25%	100%	25%
	Classe 3	0%	0.0%	75%

não só a acurácia correta da classificação para esta classe, mas também o quanto o classificador "confundiu" estas entradas como pertencentes às outras classes. No exemplo da Tabela 16, o classificador "confundiu" as entradas pertencentes à Classe 1 com a Classe 2 em 25 % dos casos, entretanto conseguiu classificar todos os pontos da Classe 2 corretamente.

Diferentemente dos resíduos, preocupando-se em avaliar como o modelo desempenhará com informações nunca apresentadas a este, no método de validação cruzada parte dos dados são retirados do banco para que possam ser utilizados futuramente em uma validação com informações "novas". O método *holdout* é uma forma de validação cruzada na qual os dados são separados em treino e teste seguindo uma determinada proporção. Os dados de teste serão utilizados como entrada do modelo criado com o intuito de validá-lo, enquanto os dados de treino serão utilizados como um banco de dados para criação do modelo.

5.2.2.2 Escolha do algoritmo

A escolha do algoritmo que realizará a classificação deve ser realizada de forma conjunta à escolha das variáveis que serão utilizadas e do seu método de obtenção. Uma determinada situação ou abordagem utilizará diferentes algoritmos com resultados diversos, sendo o trabalho do pesquisador encontrar aquele mais congruente com o proposto. Quando se trata do sistema Motorista-Veículo-Ambiente já sabemos que diferentes propostas de classificação podem surgir do mesmo, como identificação de eventos, estilos de motorista e fatores do ambiente. Por conta disso, é importante verificar quais são os pontos positivos e negativos de cada algoritmo e como estes podem ser aplicados nestas diferentes abordagens.

Alguns autores já buscaram evidenciar a importância desta definição e elucidaram a performance de diferentes algoritmos de classificação no sistema Motorista-Veículo-Ambiente. Apesar de estar focado na identificação de sistemas, LIN *et al.* (2014) realizaram uma comparação entre os algoritmos *Neural Networks* (Redes Neurais, NN), *Hidden Markov Model* (Modelo oculto de Markov, HMM) e *Fuzzy Inference System* (Sistema de Inferência Fuzzy, FIS). Os autores reforçam a ideia que cada algoritmo possui seus pontos positivos e negativos, sendo

melhor utilizados em casos específicos.

Outros autores também se preocuparam em diferenciar as aplicações dentre os algoritmos. Descrevendo aplicações que variam de identificação de inattenção, fadiga e distração até assistência na direção, avisos de segurança e monitoração das situações de direção, MEIRING; MYBURGH (2015) focaram seu estudo na revisão de algoritmos de inteligência computacional aplicados na análise do estilo do motorista. Já JÚNIOR *et al.* (2017) optaram por realizar um estudo em aplicações que utilizam smartphones como ferramentas de captura, utilizando diferentes algoritmos com um determinado conjunto de dados para avaliar as diferenças de desempenho entre estes. A Tabela 17 evidencia alguns dos pontos discutidos pelos autores que são relevantes para este trabalho. Já a Tabela 18 informa algoritmos utilizados pelos trabalhos já discutidos em seções anteriores.

Tabela 17 – Contrastes entre diferentes algoritmos

Algoritmo	Descrição	Aplicação	Custo computacional	Desvantagens
K-Means	Capaz de organizar, com uma rápida convergência, um grande conjunto de dados em clusters para possibilitar a identificação de padrões	Organizar em clusters os parâmetros que refletem características do comportamento do motorista	●○○	Sensível a erros nas medições e incertezas
NN	Capaz de identificar diversos tipos de comportamentos de motoristas em um grande numero de dados armazenados ao longo do tempo. A qualidade da captura dos parâmetros é importante para acurácia desse algoritmo.	Pode ser utilizado em reconhecimento de padrões nas quais os parâmetros são de fácil captura e identificação.	●●●	Não há um método definido para ajuste dos parâmetros, como por exemplo a quantidade de neurônios na camada escondida
HMM	Descrever estatisticamente propriedades de processos estocásticos e identificar estados invisíveis inertes através de uma sequencia de observações externas.	Pode ser utilizado em reconhecimento de padrões em séries temporais.	●○○	Não é tão eficaz em previsões de longo prazo e requer hipóteses artificiais para a sequência de distribuição dos estados atuais.
FIS	Formular regras Fuzzy baseadas em experiência prévia e, então, desenvolver um modelo que atenda às expectativas do pesquisador.	Pode ser utilizado em reconhecimento de padrões nos quais os limites dos parâmetros são de difícil definição	●○○	Já que é baseada em conhecimento prévio, os resultados alcançados podem desviar-se dos reais.

Fonte – Adaptado de (LIN *et al.*, 2014),(MEIRING; MYBURGH, 2015) e (JÚNIOR *et al.*, 2017)

Apesar de pouco presente em trabalhos recentes que referem-se diretamente ao comportamento e estilo do motorista, o algoritmo de classificação de métricas *K-Nearest Neighbors* (K-Vizinhos mais Próximos, KNN) apresenta boa performance em diversas aplicações relacionadas ao sistema automotivo. Em trabalhos nos quais o objetivo é criar uma impressão

Tabela 18 – Alguns algoritmos utilizados pela literatura

Algoritmos	Autores
Fuzzy	(CASTIGNANI <i>et al.</i> , 2015),(DÖRR; GRABENGIESSER; GAUTERIN, 2014),(FENG <i>et al.</i> , 2018)
Hidden Markov Model	(ZHANG; ZHAO; RONG, 2014)
Área de modelo gráfico	(VAIANA <i>et al.</i> , 2014),(BOUHOUTE <i>et al.</i> , 2018)
Support Vector Machines	(SHARMA <i>et al.</i> , 2015)
Neural Networks	(VIROJBOONKIADE; VATEEKUL; ROJVIBOONCHAI, 2017)
Dynamic Time Warp	(ZHANG; ZHAO; RONG, 2014)
K-means	(NIRMALI <i>et al.</i> , 2017)

Fonte – Elaborado pelo Autor

digital do o motorista do veículo com o intuito de identificar possíveis invasões ou o roubo do veículo, o KNN está entre os algoritmos de melhor performance. KWAK; WOO; KIM (2016), considerando as informações eletromecânicas de um automóvel, utilizaram quatro diferentes algoritmos com esta proposta, obtendo uma acurácia de (i) 0,939 com Árvores de Decisão; (ii) 0,844 com KNN; (iii) 0,961 com *Random Forest*; e (iv) 0,747 utilizando a NN *Multi-Layer Perceptron*.

Outra aplicação, focada na determinação da segurança do sistema automotivo utilizando para isso informações de temperatura e pressão de componentes do veículo, mostrou que o KNN é um algoritmo que pode ser utilizado nestas ocasiões atingindo uma acurácia de 0,93 (SRINIVASAN, 2018). O KNN também apresenta bom resultados na estimativa de emissão de gases em veículos ligados porém parados, fenômeno que apesar de representar apenas uma parte das emissões diárias tem grande impacto quando multiplicado pelo enorme contingente de veículos (FARZANEH *et al.*, 2014), estando entre os melhores dentre cinco modelos diferentes (LI; QIAO; YU, 2016). Dada a sua boa acurácia em diferentes campos de aplicação relacionados ao sistema automotivo, o algoritmo KNN é especialmente discutido na próxima seção.

5.2.2.3 K-Nearest Neighbors

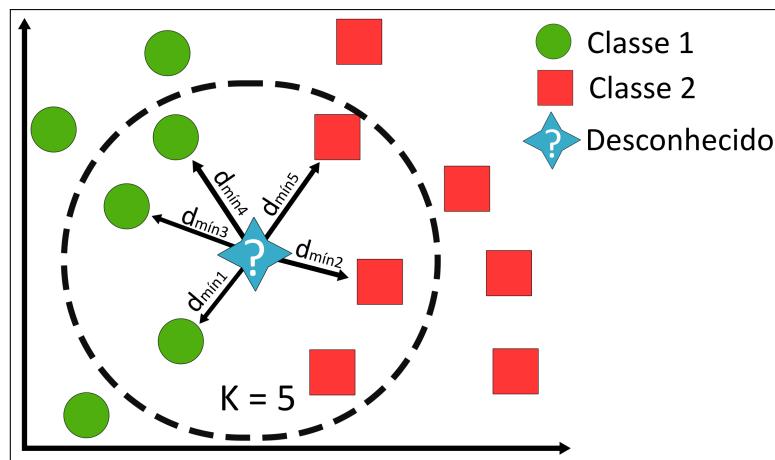
O classificador de K-vizinhos mais próximos, proposto por FUKUNAGA; NAREN-DRA (1975), é uma classificador que cria fronteiras de decisões a partir de um conjunto de dados utilizados para treinamento, sendo estes com classes conhecidas a priori (MATOS, 2008). Mais precisamente, para um dado vetor característico no conjunto teste, os vetores de K vizinhos mais próximos no conjunto de treino são selecionados de acordo com alguma medida de distância, e o vetor de características de teste é atribuído à classe mais representada dentre os K vizinhos.

Classificar uma entrada desconhecida com o algoritmo KNN pode ser resumido nos passos a seguir.

- Calcular a distância entre o exemplo desconhecido e os outros exemplos do conjunto de treinamento;
- Identificar os K vizinhos mais próximos;
- Utilizar as classe dos K vizinhos mais próximos para determinar a classe do exemplo desconhecido em uma espécie de votação.

Dessa forma, como pode ser visualizado na Figura 32, a distância do dado de entrada representado em azul a cada um dos dados que compõem o banco de treino seria calculada. Neste exemplo, com a utilização de cinco vizinhos mais próximos, as cinco menores distâncias seriam utilizadas para decidir a qual classe o novo dado pertence. Como três dos quatro vizinhos mais próximos pertencem à Classe 1, o novo dado será classificado como pertencente a esta classe. A métrica utilizada para o cálculo de distância geralmente é a Euclidiana, cuja teoria é baseada na geometria analítica. A distância Euclidiana entre vetores é apresentada na Equação 5.1, na qual \vec{x} e \vec{y} são vetores de dimensão l .

Figura 32 – Ilustração do Algoritmo KNN.



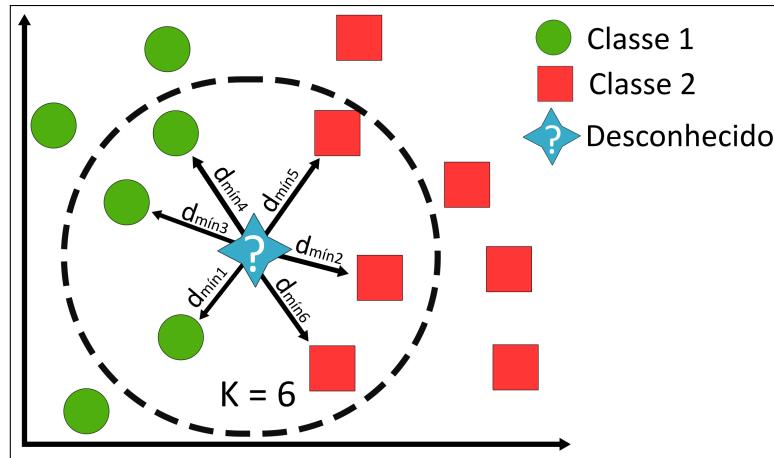
Fonte – Elaborado pelo Autor.

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=0}^l (x_i - y_i)^2} \quad (5.1)$$

Deve-se ter cuidado com o número de escolha dos vizinhos mais próximos. Número de vizinhos pares ou múltiplos das classes podem fazer com que haja empate no número de vizinhos pertencentes as classes. Na Figura 33, o número de vizinhos mais próximo foi

aumentado para $K = 6$. Nessa situação, três vizinhos mais próximos pertencem a cada uma das classes, o que causa uma indefinição no classificador. Essas situações podem ser evitadas caso regras adicionais sejam colocadas no classificador. Por exemplo, caso um empate entre classes ocorra, a entrada será associada àquela classe cuja soma das distâncias for a menor.

Figura 33 – Ilustração do Algoritmo KNN com número de vizinhos pares.



Fonte – Elaborado pelo Autor.

6 METODOLOGIA UTILIZADA E PRÉ-PROCESSAMENTO DOS DADOS

6.1 Abordagem Utilizada

O estilo de direção corresponde ao modo com o qual o motorista executa ações previamente definidas pelo seu comportamento. Dessa forma, a captura de variáveis relacionadas aos fatores das categorias Motorista, Veículo e Ambiente pode ser útil para identificar o estilo de um motorista através de seus perfis de direção. A boa relação entre custo de desenvolvimento e estabilidade/aplicabilidade de um sistema de telemetria torna-o uma escolha interessante para a criação de um banco de dados com estas variáveis.

O banco de dados criado pode então ser utilizado conjuntamente a algoritmos de aprendizado de máquina permitindo a realização da premissa inicial de classificação do estilos de direção de motoristas. Dessa forma, este sistema de classificação pode auxiliar o motorista informando-o se seu estilo de direção é ou não adequado para estar dentro de uma margem de segurança. Isto é, este sistemas pode ser utilizado como uma barreira de segurança, atuando como um DAS na fase não-conflitante da linha do tempo de segurança, evitando que o motorista crie situações de risco no que concerne aos perfis utilizados na classificação de seus estilos.

A boa acurácia do classificador KNN em diferentes campos de aplicação relacionados ao tema deste trabalho motivou a sua utilização na classificação de estilos de direção com o intuito de verificar sua performance também nesta situação. Em concordância com outros estudos já destacados neste trabalho, como pode ser visto na Tabela 15, optou-se pela utilização de três estilos para classificação com o KNN, nomeados como classes (i) Calma; (ii) Normal e (iii) Agressiva.

Por fim, o intuito deste trabalho é coletar informações referentes as categorias Motorista-Veículo de um sistema automotivo, utilizando, para isso, um sistema de telemetria conectado ao sistema de diagnóstico OBDII do veículo, e utilizar tais informações para a classificação do estilo de direção de motoristas em três classes, estas Calma, Normal e Agressiva, visando aplicações de segurança que auxiliem o motorista atuando na zona não-conflitante da linha do tempo de segurança.

6.2 Dados

Neste trabalho, utiliza-se majoritariamente o serviço \$01, o qual contém os PIDs utilizados para requisitar informações em tempo real dos sistemas eletromecânicos do veículo. Aproximadamente 200 diferentes informações podem ser requisitadas, variando em número, tipo e especificações entre diferentes modelos e fabricantes, já que informações adicionais ou mais precisas necessitam da instalação de sensores mais onerosos para as montadoras. Exemplos de variáveis que podem ser capturadas com o serviço \$01 incluem Velocidade, Fricção no motor, Nível de combustível e Temperatura do catalisador, sendo a lista completa disponibilizada pela norma que trata desta padronização (SAE, 2012).

Dentre estas, há aquelas que são essenciais em todos os carros, já que são importantes no processo de ignição, como Nível percentual de álcool no combustível e Fluxo de massa ar-combustível, e na verificação de funcionamento e integridade do motor, como Carga no motor e Rotações por minuto. Portanto, as informações escolhidas para captura devem estar dentre estas essenciais, garantindo que qualquer veículo possa fornecê-las. Foram selecionadas dentre estas aquelas capazes de fornecer uma avaliação da situação do veículo e do estilo de condução do motorista, considerando os fatores pertencentes as categorias Motorista e Veículo, pois estes são capazes de modelar o funcionamento do automóvel. Estas informações são ilustradas na Tabela 19 juntamente de seus acrônimos.

Tabela 19 – Informações capturadas do sistema OBDII e o sensor interno correspondente.

PIDs	Dados	Acrônimo	Sensor
010C	Velocidade do Motor (Rotações por minuto, RPM)	RPM	Sensor de efeito Hall
010D	Velocidade do veículo (Km/h)	Vel	Sensor de velocidade
0110	Fluxo de massa	MAF	Sensor mássico
0111	Posição do acelerador	Acl	Potenciômetro
0104	Carga no motor	Elo	Obtido por cálculo indireto

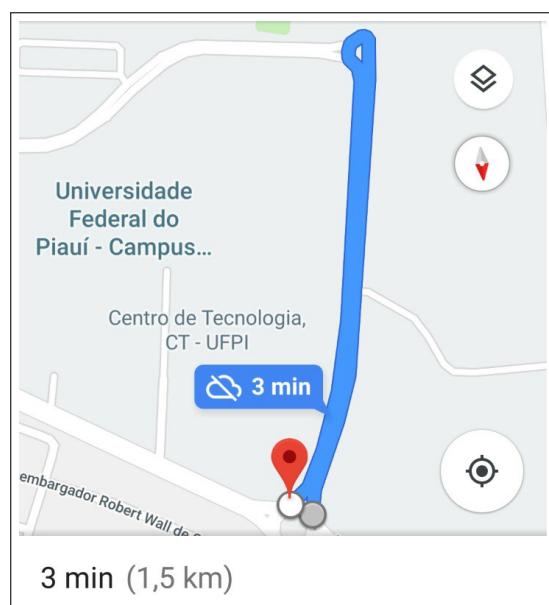
Fonte – Elaborado pelo Autor.

6.3 Coleta

A validação do funcionamento do sistema de telemetria e a coleta das informações especificadas na Seção 6.2 ocorreram em um percurso predefinido. O percurso escolhido (i) pertence a um perímetro urbano, em pista asfaltada sem interferência de outros veículos; (ii) foi percorrido com o tempo ensolarado e alta visibilidade; (iii) possui um redutor de velocidade e

uma curva acentuada; (iv) e tem extensão total de 1,5 quilômetros. O motorista que realizou os testes foi orientado a percorrê-lo de maneira (i) calma, respeitando o desenvolvimento do veículo ; (ii) normal, como dirigiria respeitando o limite de velocidade urbano; (iii) e agressiva, com o objetivo de terminar o percurso o mais rápido possível. O conjunto de dados gerados para cada uma dessas orientações foi nomeado de coleta. Observou-se que a maior taxa de captura que pode ser realizada com o sistema é de duas capturas por segundo. O teste foi realizado até se obter ao menos 3600 capturas de informação para cada uma das coletas. A Figura 34 ilustra o percurso utilizado, localizado ao lado do Centro de Tecnologia da Universidade Federal do Piauí.

Figura 34 – Percurso utilizado para a captura de informações.



Fonte – Retirado de *Google Maps*.

6.4 Pré-processamento dos dados

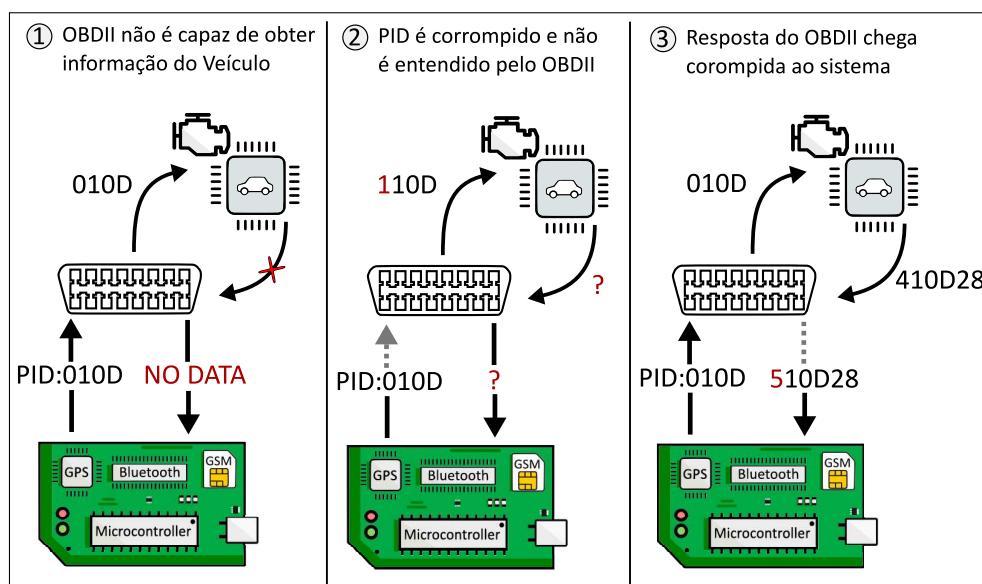
6.4.1 Preparação e Limpeza dos dados

Após coletados, os dados recebidos no servidor foram transferidos para um computador onde pudessem ser analisados. Inicialmente realizou-se a limpeza dos dados, na qual pacotes de captura iguais e pacotes inválidos foram descartados. Pacotes de capturas iguais podem ocorrer já que o sistema de telemetria captura informações a cada 0,5 segundo, o que pode gerar capturas semelhantes caso o veículo esteja dinamicamente estável.

Pacotes inválidos podem ocorrer, correspondendo a problemas de comunicação entre

sistema de telemetria e OBDII, resultando no envio ou recebimento de caracteres inválidos, como pode ser visto nos exemplos da Figura 35. Primeiramente, o sistema OBDII pode ter sido incapaz de coletar a informação do veículo devido a problemas internos, o que o faz enviar ao sistema de telemetria a *string* "NO DATA". Caso o PID recebido pelo sistema OBDII esteja corrompido, o OBDII responderá com uma *string* "?". Caso a resposta recebida pelo sistema de telemetria esteja corrompida, o cabeçalho de identificação apresentará valores inconsistentes ou sua conversão não será possível. Além disso, os dados do GPS foram também identificados para serem utilizados apenas no rastreamento do veículo.

Figura 35 – Possíveis problemas de comunicação do sistema de telemetria.



Fonte – Elaborado pelo Autor.

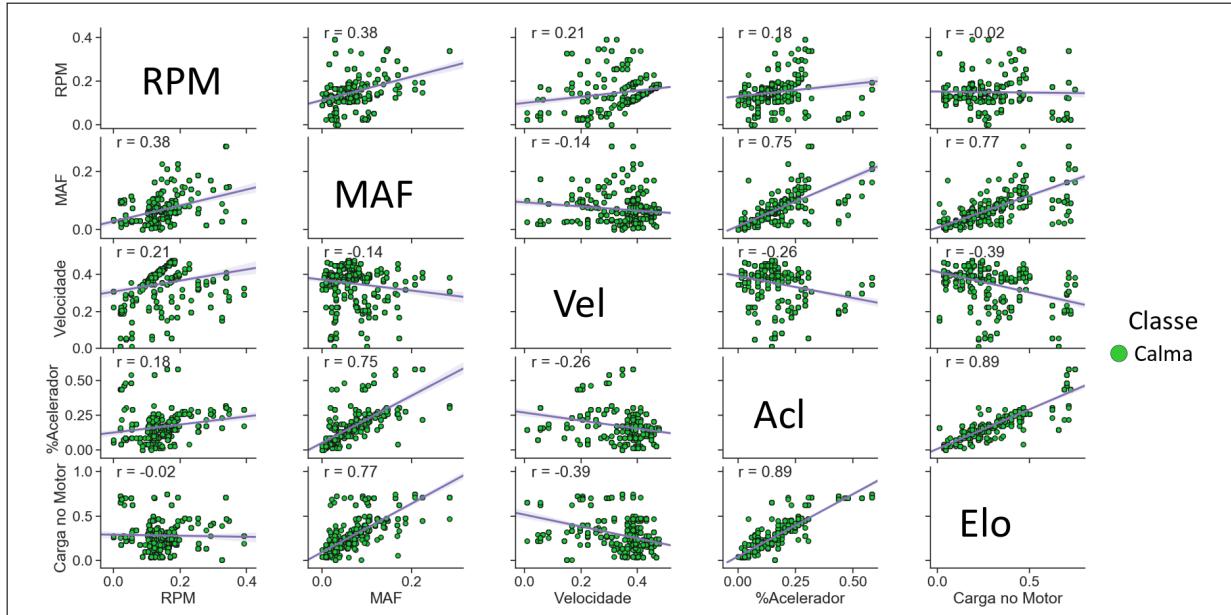
Em todos os casos relatados os pacotes foram retirados do banco de dados e salvos em outro arquivo para que estes não causem problemas no processamento e possam ser posteriormente analisados. Apenas 0,3% de todos os pacotes de captura recebidos no servidor foram retirados durante o processo de limpeza, o que corresponde a 4,5 segundos de captura, mostrando que o sistema operou de maneira estável.

6.4.2 Seleção e Preparação dos Atributos

Informações semelhantes produzirão conhecimento redundante quando utilizadas em um sistema de aprendizado, o que tende a aumentar o custo computacional empregado sem um igual incremento na conhecimento produzido. Para evitar que isso aconteça, os dados obtidos

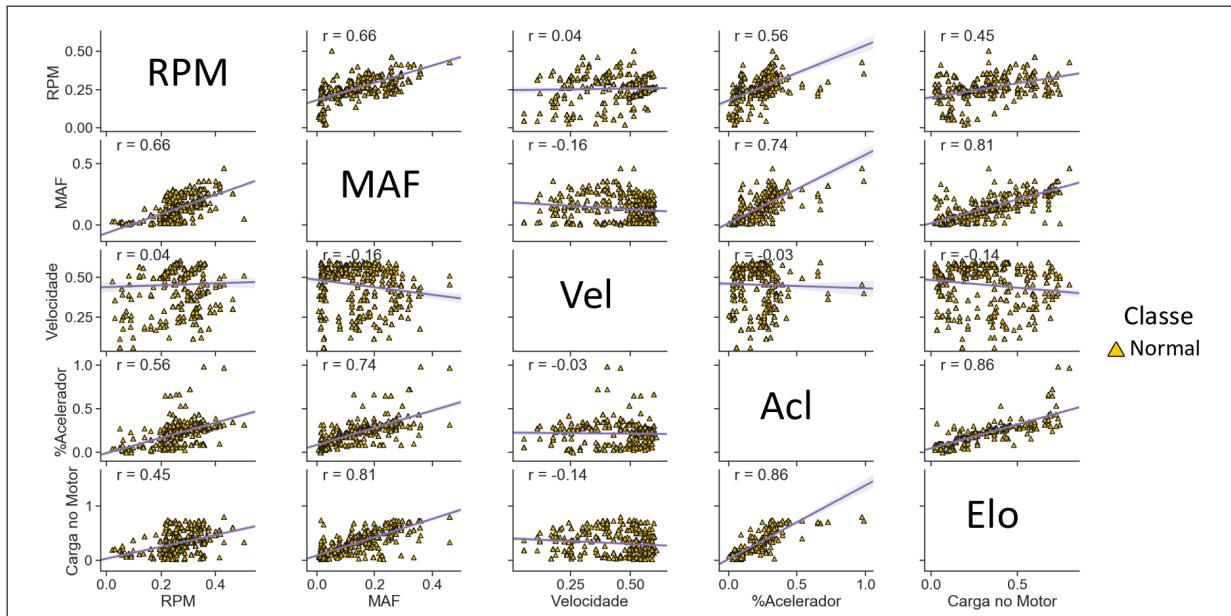
nos testes foram analisados para decidir a validade de sua utilização no sistema de classificação. Essa análise foi realizada com o cálculo da correlação entre todas as variáveis capturadas para cada uma das coletas. Os resultados obtidos são apresentados graficamente nas Figuras 36, 37 e 38 e numericamente nas Tabelas 36, 37 e 38.

Figura 36 – Matriz de Correlação da captura Calma representada graficamente.



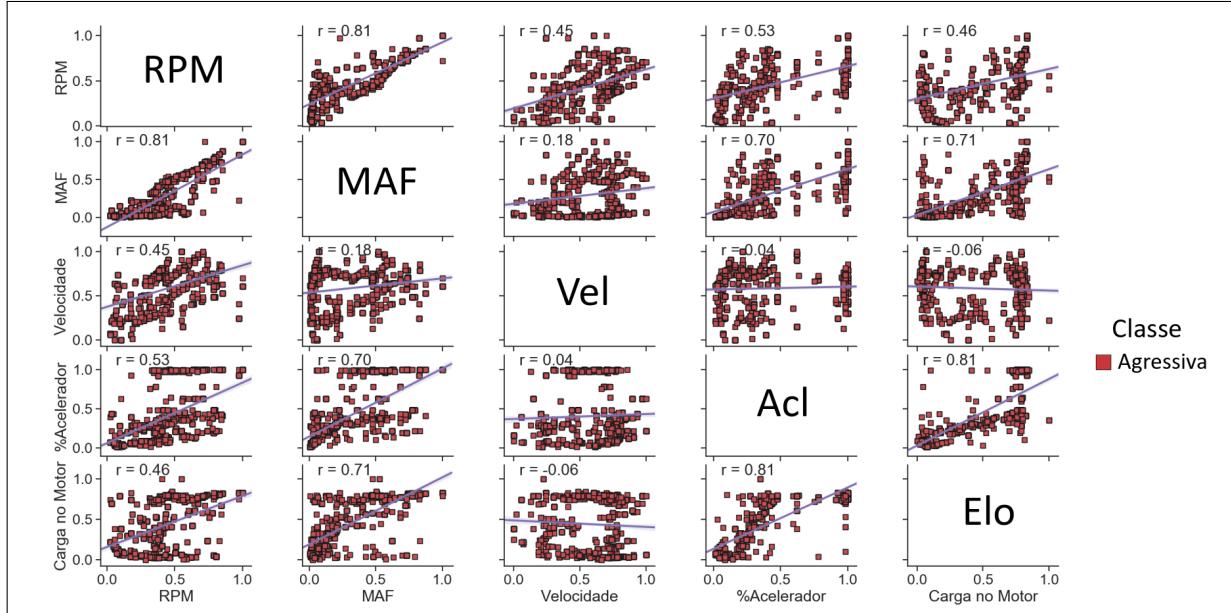
Fonte – Elaborado pelo Autor.

Figura 37 – Matriz de Correlação da captura Normal representada graficamente.



Fonte – Elaborado pelo Autor.

Figura 38 – Matriz de Correlação da captura Agressiva representada graficamente .



Fonte – Elaborado pelo Autor.

Tabela 20 – Matriz de Correlação da captura Calma

-	RPM	MAF	Vel	Acl	Elo
RPM	-	0.38	0.21	0.18	-0.02
MAF	0.38	-	-0.14	0.75	0.77
Vel	0.21	-0.14	-	-0.26	-0.39
Acl	0.18	0.75	-0.26	-	0.89
Elo	-0.02	0.77	-0.39	0.89	-

Fonte – Elaborado pelo Autor

Tabela 21 – Matriz de Correlação da captura Normal

-	RPM	MAF	Vel	Acl	Elo
RPM	-	0.66	0.04	0.56	-0.45
MAF	0.66	-	-0.16	0.74	0.81
Vel	0.04	-0.16	-	-0.03	-0.14
Acl	0.56	0.74	-0.03	-	0.86
Elo	-0.45	0.81	-0.14	0.86	-

Fonte – Elaborado pelo Autor

Primeiramente, observou-se que as variáveis Acl, Elo e MAF são fortemente correlacionadas em todas as coletas, comportamento explicado pela transferência de energia ao motor dada pela relação mecânica entre a abertura da válvula de passagem da mistura ar/combustível com a pressão exercida pelo motorista no pedal de aceleração(KALITA, 2017). Verifica-se também que as variáveis Acl e Elo possuem uma correlação semelhante com todas as outras variáveis, mostrando que impactam praticamente da mesma forma o conjunto de dados, com-

Tabela 22 – Matriz de Correlação da captura Agressiva

-	RPM	MAF	Vel	Acl	Elo
RPM	-	0.81	0.45	0.53	0.46
MAF	0.81	-	0.18	0.70	0.71
Vel	0.45	0.18	-	0.04	-0.06
Acl	0.53	0.70	0.04	-	0.81
Elo	0.46	0.71	-0.06	0.81	-

Fonte – Elaborado pelo Autor

portamento menos evidente entre as variáveis Acl e MAF. Como o pedal de aceleração é um estímulo mais próximo do motorista, optou-se por utilizar a variável Acl em detrimento da Elo. Dessa forma, as informações (i) RPM; (ii) MAF; (iii) Vel e (iv) Acl foram selecionadas como atributos a serem utilizados no processo de classificação. Os atributos selecionados de cada pacote de captura formam um vetor de atributos. Este vetor possui quatro dimensões, uma para cada atributo, como pode ser visto na Tabela 23, na qual a descrição do conjunto de dados está detalhada.

Tabela 23 – Descrição do conjunto de dados

Tipo	Numérico(quantitativo)
Classes	3 (Agressiva, Normal e Calma)
Atributos	4 (RPM, MAF, Vel e Acl)
Quantidade de vetores de dados	11.520 (3.840 para cada classe)
Tempo total	96 minutos (32 para cada classe)

Fonte – Elaborado pelo Autor

Para efetuar uma análise, os valores dos atributos foram então normalizados entre o intervalo [0,1], com o intuito permitir uma melhor visualização destes e reduzir o consumo de processamento do algoritmo de análise. Esse processo é exemplificado na Equação 6.1 com o atributo RPM, na qual \overline{RPM} e \underline{RPM} são, respectivamente, os valores mínimo e máximo do atributo.

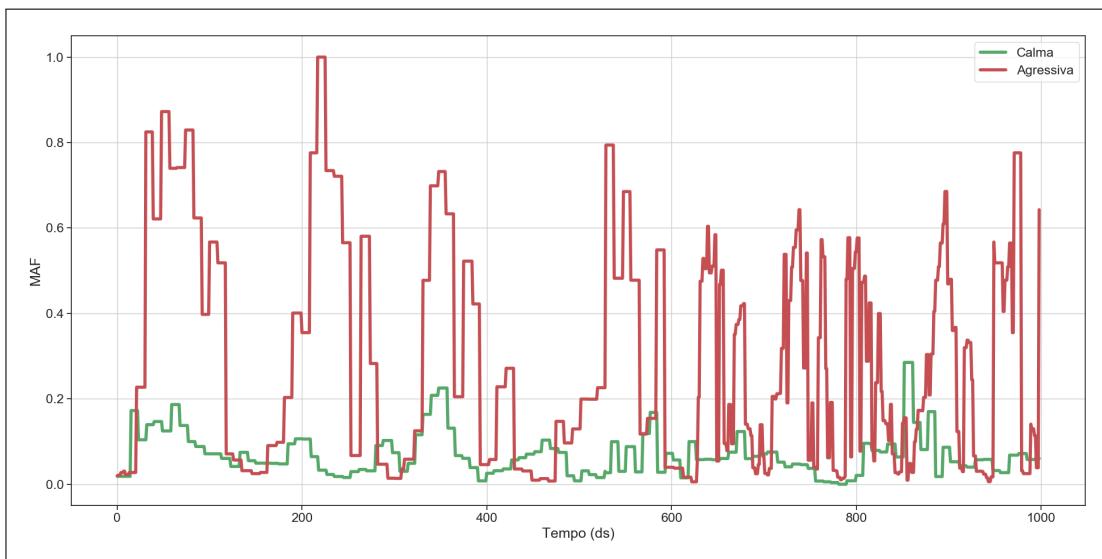
$$RPM := \frac{RPM - \underline{RPM}}{\overline{RPM} - \underline{RPM}} \quad (6.1)$$

Os processos de limpeza, seleção de atributos e normalização geraram um banco de dados propício para que se iniciasse a análise das informações coletadas. A análise preliminar destas variáveis permite a criação de uma expectativa inicial de como se dará a classificação destas.

6.5 Análise

Sabe-se que a direção é uma atividade expressada em uma janela temporal, isto é, a variação dos atributos que a compõem apresentam uma dinâmica (perfil) e é esta que classificará o estilo de determinado motorista. As Figuras 39 e 40 apresentam uma comparação entre a dinâmica dos atributos MAF e RPM para o estilo de direção Calma e Agressiva. Nestas, pode-se observar uma nítida diferença entre os valores capturados, com picos de MAF na direção agressiva. Além disso, o comportamento dos atributos possibilitam a determinação de eventos como aceleração e troca de marcha. A Aceleração, por exemplo, pode ser identificada em momentos nos quais o motorista pressiona o Pedal de Aceleração (Acl). Por outro lado, regiões com grandes variações negativas deste atributo e pequenas variações do atributo RPM indicam a troca de marcha.

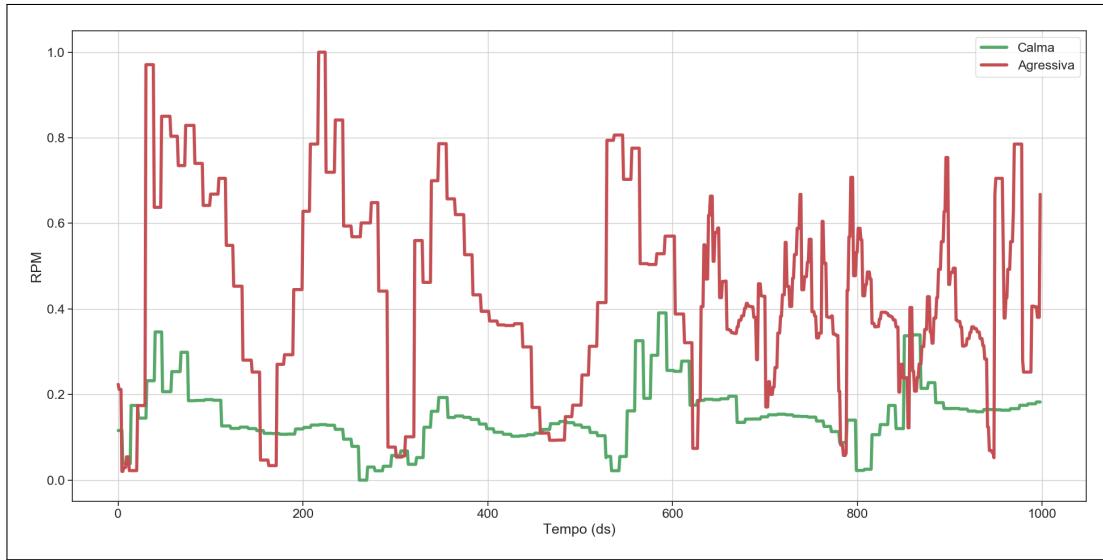
Figura 39 – Dinâmica temporal do atributo MAF para as classes Calma e Agressiva.



Fonte – Elaborado pelo Autor.

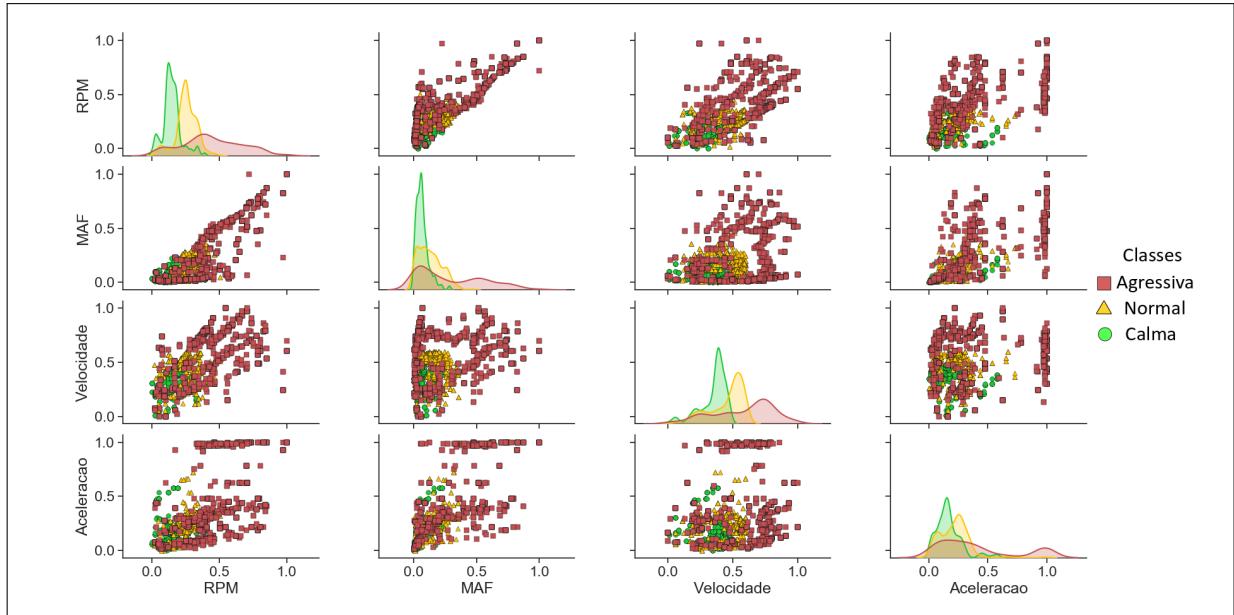
A Figura 41 apresenta um gráfico matricial no qual todos os atributos são plotados agrupadamente por coleta. Neste mesmo gráfico, a distribuição dos atributos por classe é plotada na diagonal principal. Através da Figura 41 observamos que os atributos relacionados à coleta Calma estão exclusivamente concentrados em pontos próximos da origem, região na qual a separação entre as classes é de menor intensidade. Essa região descreve condições de parada, redução de desenvolvimento do automóvel, curvas e eventos que exigem cautela; comportamento esse, presente em todos as classes.

Figura 40 – Dinâmica temporal do atributo RPM para as classes Calma e Agressiva.



Fonte – Elaborado pelo Autor.

Figura 41 – Agrupamento dos atributos pertencentes as três classes.

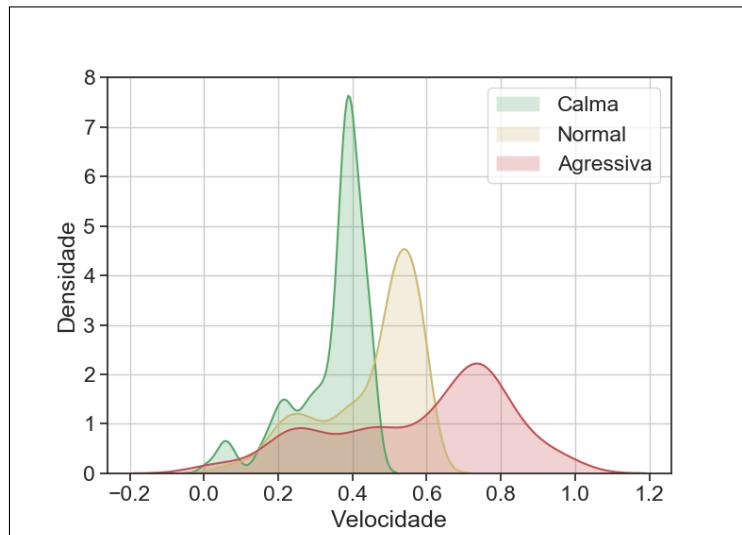


Fonte – Elaborado pelo Autor.

Podemos concluir, então, que o estilo de direção de determinado motorista é composto pelas classes Calma, Normal e Agressiva, sendo que uma destas é predominante dentre as outras. Esse pensamento é ilustrado na Figura 42, na qual a distribuição do atributo Vel apresenta regiões que poderiam pertencer as outras duas classes, comportamento representado pelas intersecções das distribuições. Esse comportamento pressupõe que a utilização de pontos isolados como entradas na classificação pode incorrer em decisões erradas por parte do classificador, já que mesmo a classe Agressiva possui momentos na qual é mais parecida com a classe Calma

ou Normal. Com estas análises a Tabela 24 é montada, descrevendo as três classes de estilo de direção utilizadas neste trabalho.

Figura 42 – Representação da distribuição do atributo Vel para cada classe.



Fonte – Elaborado pelo Autor.

Tabela 24 – Descrição das classes de estilo de direção

Classe	Características
Calma	Classe caracterizada por uma aceleração gradual; velocidade média baixa e praticamente invariante; RPM bastante estável mas com picos de valores próximos da média de todas as classes.
Normal	Classe caracterizada por um RPM estável e bem próximo da média de todas as classes; velocidade variável mas de valores baixos; aceleração gradual bastante semelhante à classe calma.
Agressiva	Classe com RPM altamente variável representada por variações bruscas de aceleração; velocidade bastante variável, contrária ao observado nas outras classes.

Fonte – Elaborado pelo Autor

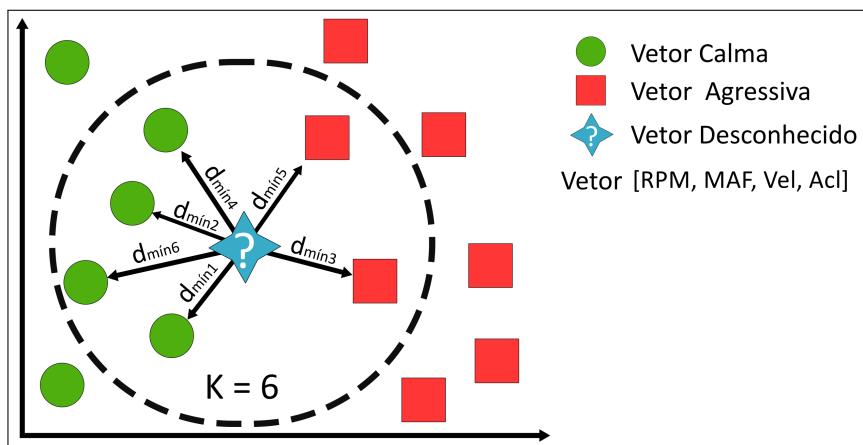
Com a análise dos atributos observa-se que apesar das interseções existentes, cada uma das classes apresenta pontos característicos, os quais podem ser utilizados para diferenciá-las. Dessa forma, a próxima seção aborda a utilização do classificador KNN nos atributos selecionados e analisados.

6.6 Algoritmo de Classificação

Com as classes e atributos definidos, implementou-se o algoritmo de classificação KNN, o qual utiliza como métrica a distância euclidiana. Este algoritmo foi implementado utilizando a linguagem de programação orientada ao objeto Python v3.6 e seguindo as diretrizes da *Python Enhance Proposals* (Propostas de Aperfeiçoamento do Python, PEP), em especial a PEP8. Dessa forma o código resultante é autoral, livre e extensamente comentado para que o mesmo possa ser reproduzido. Todo os dados e códigos utilizados neste trabalho estão publicados sob a *MIT License* (Lincença Open-Source), podendo ser acessados de forma livre¹.

Como discutido na Seção 6.5, podemos fazer a consideração inicial de que a utilização de janelas temporais para a classificação dos estilos parece ser uma abordagem mais lógica, enquanto que a utilização de vetores isolados aparentemente criará situações nas quais o classificador acusará falso-positivos (classificar o vetor de atributos de uma classe como pertencente a outra classe). Por conta disso, para averiguar estas premissas decidimos abordar os dados de duas maneiras: instantânea e janela temporal. Em uma abordagem instantânea cada pacote de captura é utilizado isoladamente como entrada do algoritmo KNN, o que resultaria em 11.520 possíveis entradas, já que cada captura seria representada em um vetor, como é ilustrado na Figura 43.

Figura 43 – Algoritmo KNN com uma abordagem instantânea.



Fonte – Elaborado pelo Autor.

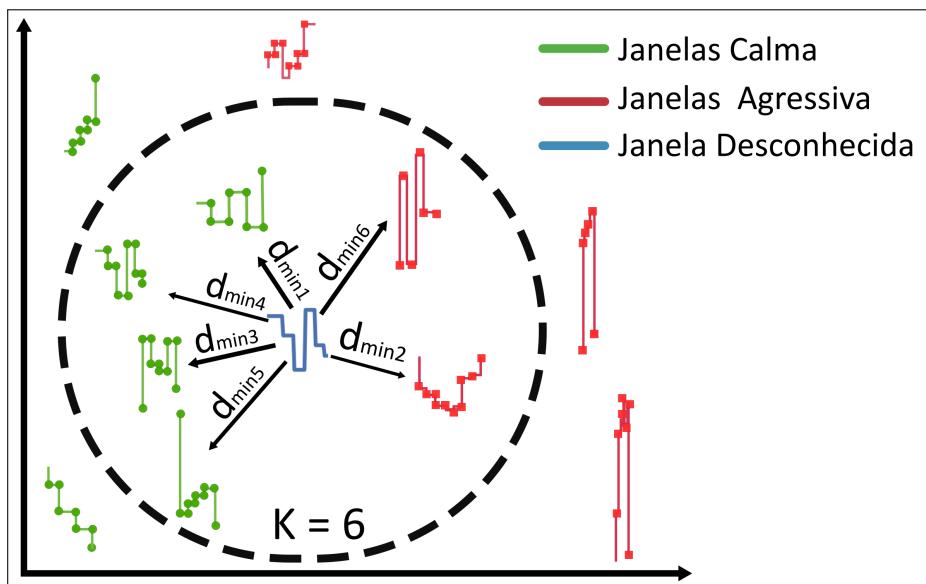
Em uma abordagem de janela temporal, a dinâmica dos atributos seria considerada. Dessa forma, ao invés de pontos isolados, pontos sucessivos do conjunto de dados seriam utilizados para compor uma janela, representando uma seção da curva de cada atributo. Assim, o conjunto de dados seria dividido em m grupos de t vetores, os quais poderiam ser utilizados

¹ <https://github.com/wborbaneto/research/tree/master/TCC>

como entradas do algoritmo. Por exemplo, caso seja escolhida uma janela temporal de $t = 5$ capturas, os 11.520 vetores iniciais seriam divididos em 2.304 grupos de 5 vetores. A utilização de medidas estatísticas dos grupos também é uma possibilidade com esta abordagem, como, por exemplo, a utilização das médias dos atributos que compõem um grupo de vetores. Esta última variação da abordagem de janelas temporais é interessante, já que o valor médio dos atributos selecionados são usualmente utilizados em sistemas automotivos, como, por exemplo, a velocidade média, aceleração média e o consumo médio, sendo que este último pode ser derivado do valor do atributo MAF (ROCHA NETO; MENEZES JUNIOR, 2018).

A abordagem de janela temporal é ilustrada na Figura 44, na qual os pontos de um grupo de vetores formam uma seção da curva total de um atributo, como aquelas ilustradas nas Figuras 39 e 40. Os pontos desta curva são então comparados com outras seções que formam o banco de treino. Assim, o algoritmo procura aquelas seções que, ponto a ponto, mais se aproximam da entrada.

Figura 44 – Algoritmo KNN com uma abordagem de janelas temporais.



Fonte – Elaborado pelo Autor.

A abordagem de janela temporal parece interessante pois procura manter a identidade da dinâmica dos atributos, o que pode resultar em um maior discernimento das características de cada classe, impedindo que pontos isolados sejam classificados erroneamente pelo algoritmo, dada a interseção entre as distribuições dos atributos das três classes, comportamento que foi discutido na Seção 6.5. Com a definição de quais abordagens serão utilizadas, o próximo capítulo apresenta os testes e resultados obtidos com cada abordagem

7 RESULTADOS

7.1 Introdução

Com os dados capturados, pré-processados e analisados, o próximo passo é a utilização destes no algoritmo de classificação KNN. Este capítulo objetiva analisar e comentar detalhadamente todo o processo de pensamento que guiou os experimentos com o classificador KNN e apresentar a melhor abordagem para a classificação do estilo de direção com os dados capturados.

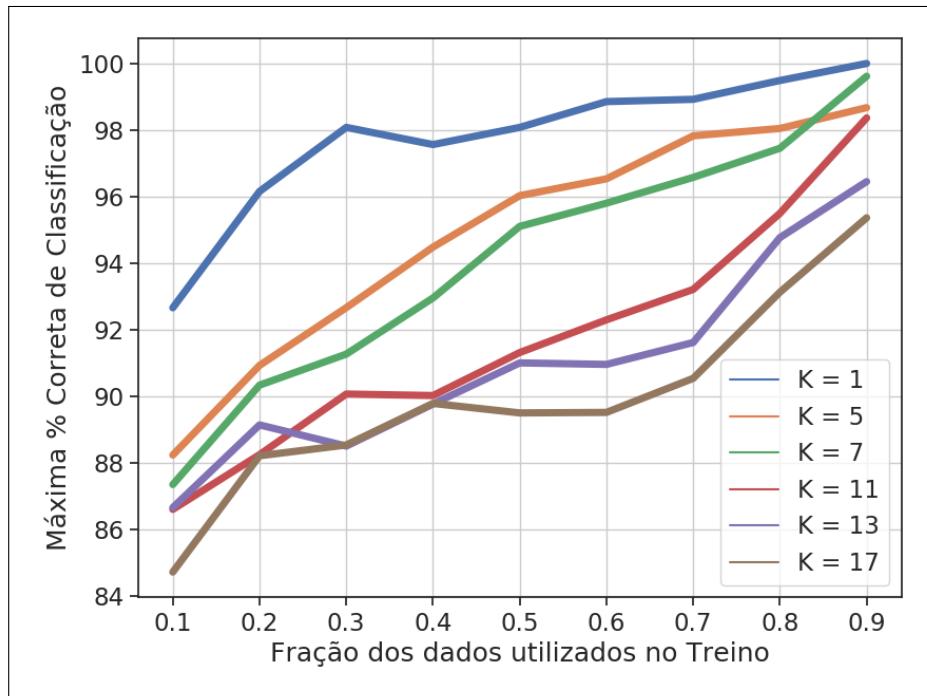
O método de validação *holdout* foi escolhido para verificar o desempenho da classificação, isto é, os dados das coletas foram separados em treino e teste seguindo uma determinada proporção (*ttRatio*). Os dados de teste serão os vetores de entrada, enquanto os dados de treino serão utilizados como um banco de dados para que o algoritmo possa calcular a distância euclidiana e determinar os vizinhos mais próximos. As próximas seções discutem os testes aplicados com cada uma das abordagens, iniciando com a instantânea.

7.1.1 Abordagem instantânea

Cada rodada de classificação foi repetida 100 vezes, sendo que os dados foram embaralhados aleatoriamente em cada rodada, permitindo diferentes resultados de classificação. Assim, obteve-se resultados médios de acertos e erros do classificador. Primeiramente foram feitos testes de variação do número de vizinhos mais próximos do classificador KNN, utilizando a abordagem instantânea. Os gráficos apresentados nas Figura 45, 46, 47 e 48 mostram o efeito da variação da relação entre treino e teste na média, máximo, mínimo e desvio padrão do acerto da classificação para diferentes números de vizinhos mais próximos. Nos gráficos apresentados, a letra K representa o número de vizinhos mais próximos utilizado. Os números de vizinhos $K = [1,5,11,13,17]$ foram escolhidos por serem ímpares e não serem múltiplos de 3 (o número de classes) impedindo que haja um empate no número de vizinhos mais próximos.

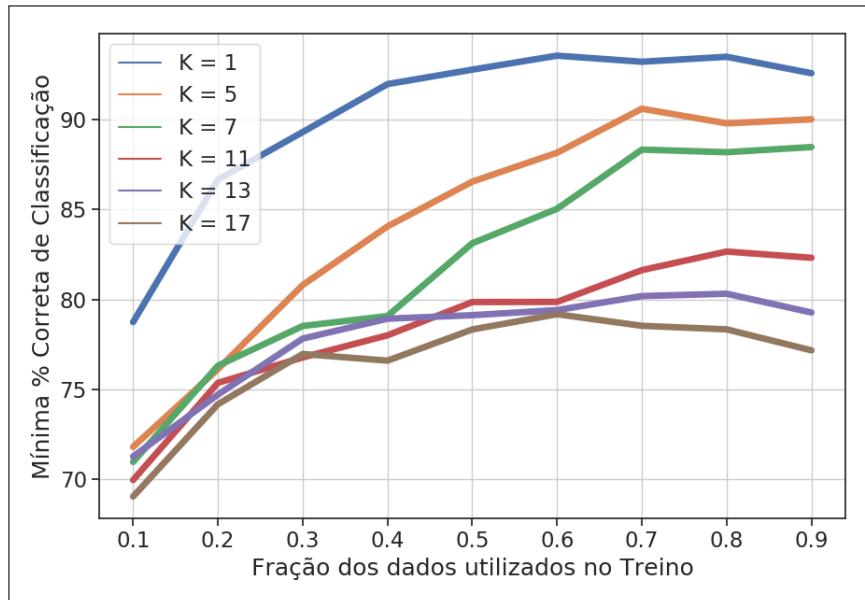
Observamos que em todas as situações o algoritmo obteve melhores resultados utilizando apenas um vizinho mais próximo. Mesmo na situação na qual apenas 10% dos dados foram utilizados para treino (1.152 capturas), o algoritmo obteve uma média de acurácia de 86,7% e a segunda menor variação. O segundo menor desvio padrão e a segunda maior acurácia foram obtidos com $K = 1$ e 80% dos dados utilizados para treino (*ttRatio* = 0,8), sendo considerado, então, como o melhor resultado do classificador.

Figura 45 – Variação da máxima acurácia para diferentes números de vizinhos mais próximos



Fonte – Elaborado pelo Autor.

Figura 46 – Variação da mínima acurácia para diferentes números de vizinhos mais próximos

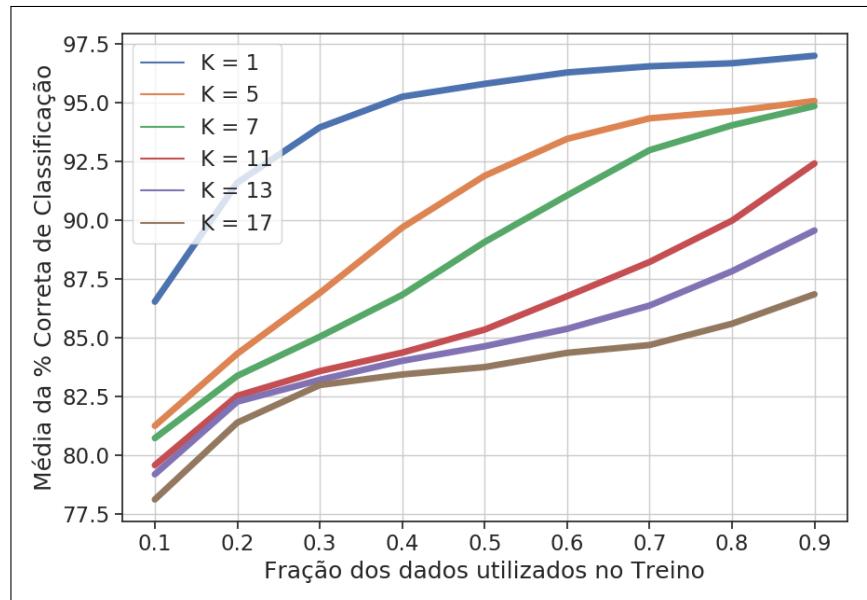


Fonte – Elaborado pelo Autor.

7.1.2 Abordagem de janela temporal

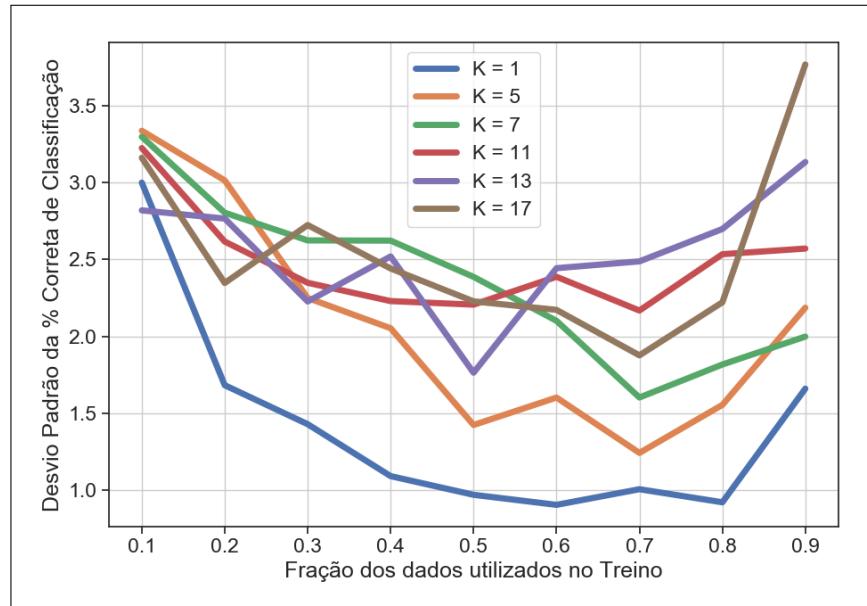
Os mesmos testes foram repetidos para a abordagem de janela temporal. Os mesmo parâmetros foram utilizados para cada variação, sendo cada teste repetido 100 vezes e com embaralhamento dos grupos de vetores em cada repetição. Duas variações foram utilizadas,

Figura 47 – Variação da Acurácia média para diferentes números de vizinhos mais próximos



Fonte – Elaborado pelo Autor.

Figura 48 – Variação do Desvio Padrão da acurácia para diferentes números de vizinhos mais próximos



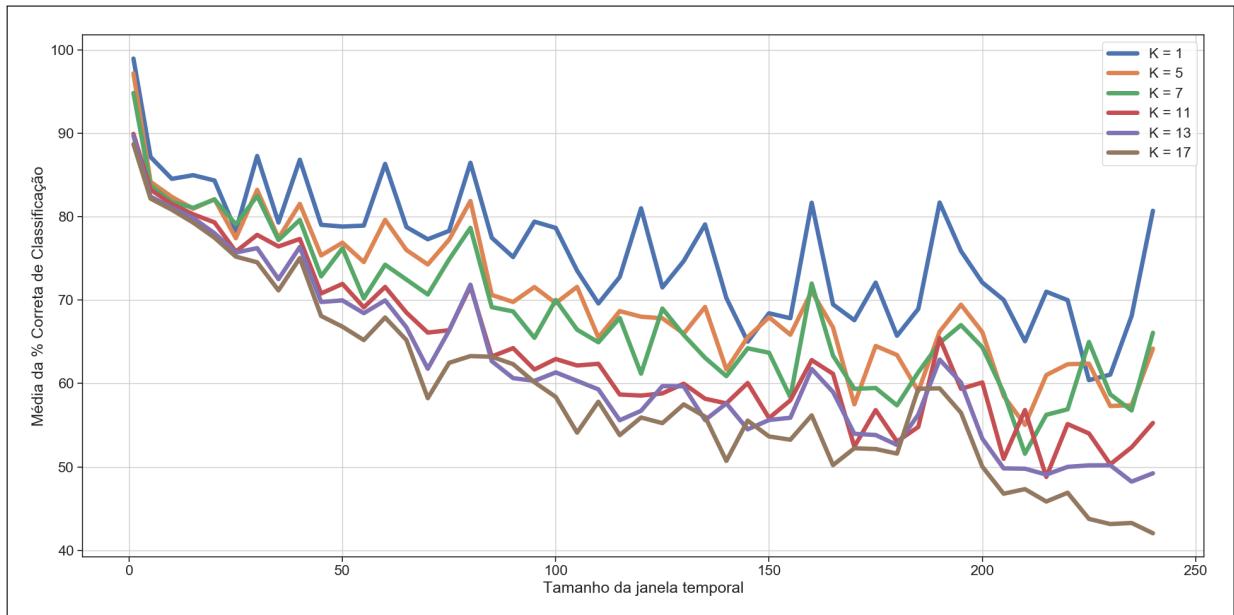
Fonte – Elaborado pelo Autor.

ambas com janelas temporais de tamanhos t pertencentes ao intervalo $t = [1, 240]$ capturas, isto é, entre 0,5 e 120 segundos. Na primeira variação a entrada do algoritmo KNN são os próprios grupos de vetores. Na segunda utilizou-se como entrada o centroide de cada grupo (formado pela média dos valores de cada atributo), dessa forma o algoritmo calculará a distância entre os centroides dos grupos e não entre vetores.

7.1.2.1 Janela temporal

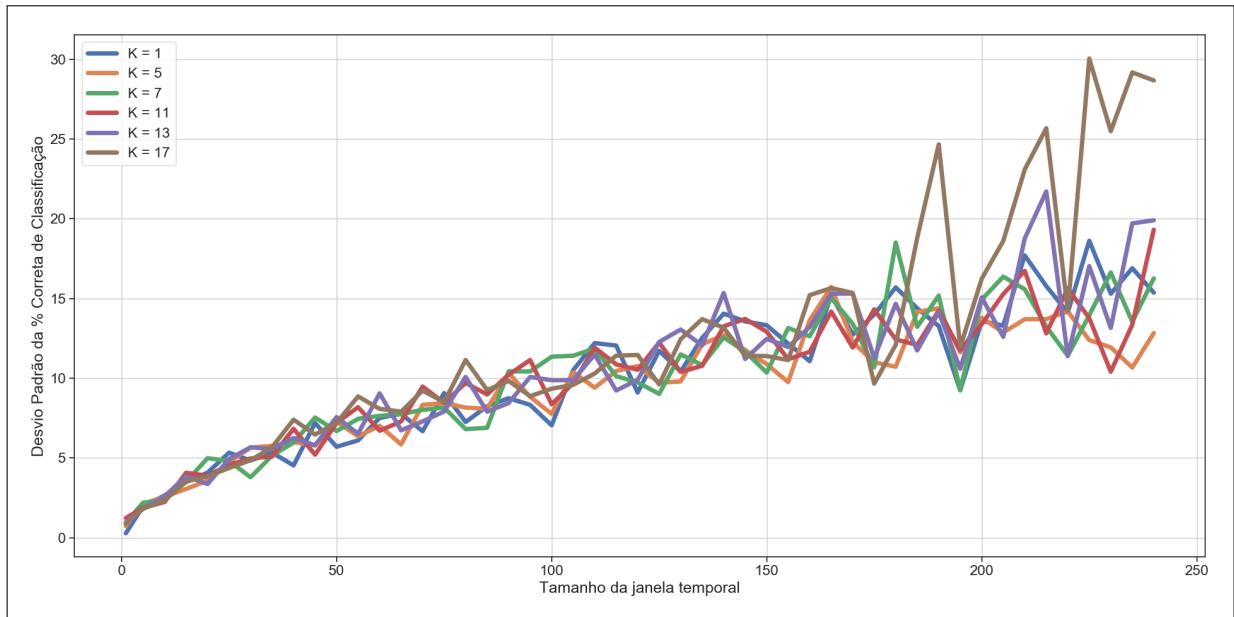
As Figuras 49 e 50 apresentam os resultados para grupos de vetores completos como entrada. Utilizou-se a proporção de treino igual a 0.8 ($ttRatio = 0,8$), pois neste valor foram encontrados ótimos resultados no teste com abordagem instantânea.

Figura 49 – Variação da acurácia Média para diferentes janelas temporais



Fonte – Elaborado pelo Autor.

Figura 50 – Variação do Desvio Padrão da acurácia para diferentes janelas temporais



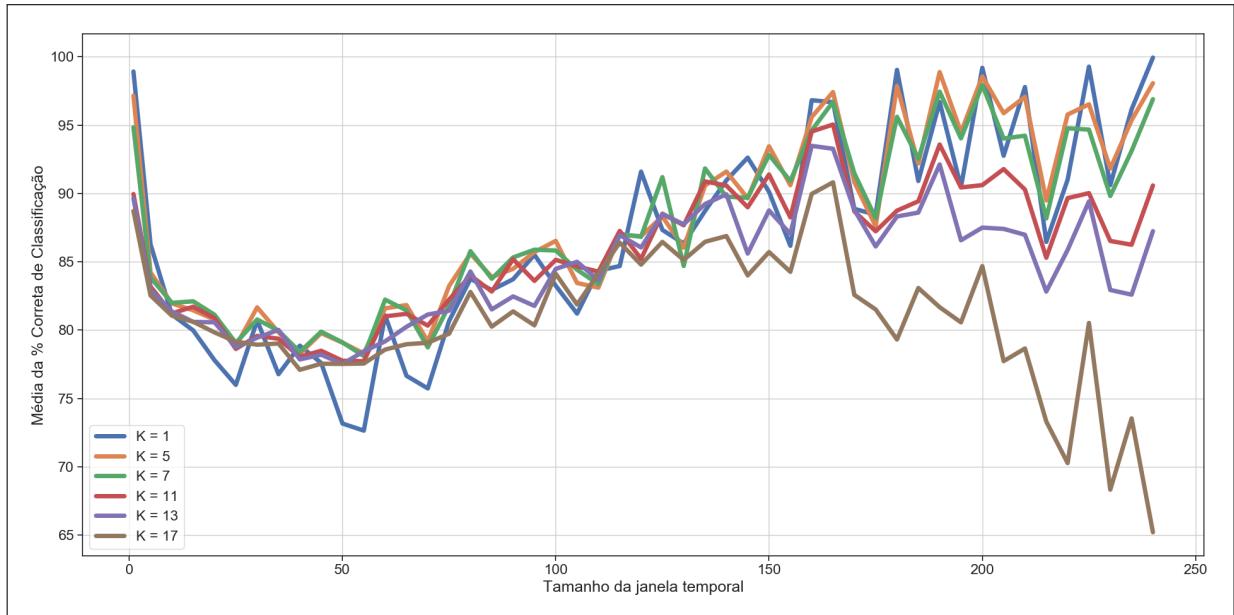
Fonte – Elaborado pelo Autor.

Apesar de não ter mantido a melhor acurácia média em todos os casos, o classificador com apenas um vizinho teve o melhor desempenho de uma maneira geral. Observa-se uma queda de desempenho do classificador, o qual obtém variações negativas da acurácia e positivas do desvio padrão praticamente proporcionais ao aumento do tamanho da janela temporal. Os gráficos oscilam bastante apesar do grande número de iterações por teste, sugerindo que o tamanho da janela influencia bastante o classificador.

7.1.2.2 *Centroides*

Assim como a variação anterior, as Figuras 51 e 52 apresentam os resultados para os centroides dos grupos como entrada, com os mesmos parâmetros para teste. Quanto à média de acertos, observa-se uma queda brusca de desempenho inicial do classificador até uma janela temporal de tamanho $t = 55$ capturas. Logo após, a média da acurácia sobe de maneira oscilante até atingir valores próximos da abordagem instantânea. Esse comportamento sugere que os valores médios dos atributos durante o percurso podem ser bons indicadores para a classificação do estilo de direção, desde que a janela seja suficientemente grande.

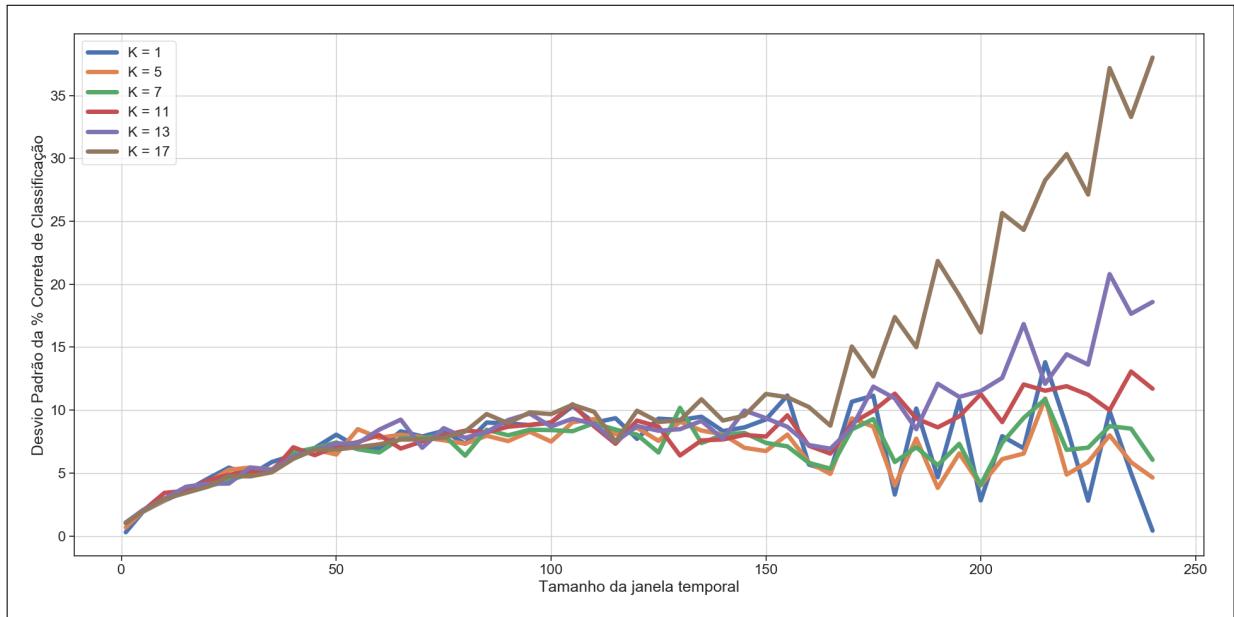
Figura 51 – Variação da acurácia Média para diferentes centroides de janelas temporais



Fonte – Elaborado pelo Autor.

Apesar dos resultados apresentarem-se semelhantes independente do número de vizinhos, o classificado com $K = 17$ apresenta uma queda de desempenho gradual e aumento do desvio padrão a partir da janela $t = 175$ capturas. Esse comportamento é explicado pela

Figura 52 – Variação do Desvio Padrão da acurácia para diferentes centroides de janelas temporais



Fonte – Elaborado pelo Autor.

diminuição do número de grupos de atributos conforme o aumento da janela, o que restringe o número de vizinhos que podem ser utilizados pelo algoritmo.

Quanto à variação, observa-se um valor estável de desvio padrão, excluindo-se o comportamento com o uso de um grande números de vizinhos, os quais sofrem com a redução do número de grupos comentada no parágrafo anterior. Os valores de desvio são maiores do que a abordagem instantânea e menores do que a primeira variação de janela temporal.

7.1.3 Análise dos resultados da classificação

Nos resultados encontrados observa-se que quando empregamos a primeira variação da abordagem de janelas temporais a média da acurácia tende a cair e o desvio padrão aumentar de acordo com o aumento da janela, comportamento oposto a premissa inicial que sugeria um melhor desempenho da abordagem de janela temporal em detrimento da instantânea. No entanto, a variação com centroides apresenta uma queda de desempenho apenas para pequenos valores, com uma recuperação conforme o aumento da janela. Considerando o caso com apenas um vizinho, a variação utilizando centroides supera a acurácia da abordagem instantânea em casos com uma grande janela temporal. Porém, a variação utilizando centroides possui, em média, um desvio padrão maior do que a abordagem instantânea.

As diferenças na eficácia do classificador ficam evidentes quando comparamos as matrizes de confusão (Tabelas 25, 26, 27, 28 e 29) para o melhor caso da abordagem instantânea ($K = 1$ e $ttRatio = 0.8$) com as duas variações da abordagem janela temporal considerando uma janela de tamanho $t = 6$ e $t = 200$ capturas.

Tabela 25 – Matriz de confusão para $K = 1$ abordagem instantânea

Classes	Calma	Normal	Agressiva
Calma	97,75	2,24	1,86
Normal	1,23	95,72	3,00
Agressiva	1,02	2,04	95,14

Fonte – Elaborado pelo Autor

Tabela 26 – Matriz de confusão para $K = 1$ e $t = 6$ abordagem janela temporal

Classes	Calma	Normal	Agressiva
Calma	83,17	12,92	9,71
Normal	10,05	81,29	17,27
Agressiva	6,78	5,79	73,02

Fonte – Elaborado pelo Autor

Tabela 27 – Matriz de confusão para $K = 1$ e $t = 200$ abordagem janela temporal

Classes	Calma	Normal	Agressiva
Calma	90,67	19,13	6,51
Normal	9,33	80,87	40,57
Agressiva	0,00	0,00	52,92

Fonte – Elaborado pelo Autor

Tabela 28 – Matriz de confusão para $K = 1$ e $t = 6$ abordagem centroide

Classes	Calma	Normal	Agressiva
Calma	82,36	11,67	7,21
Normal	10,23	79,89	15,34
Agressiva	7,41	8,44	77,45

Fonte – Elaborado pelo Autor

Observa-se que a variação da janela temporal para grupo de vetores obteve um desempenho inferior aos dos outros dois testes. A variação com uso de centroides obteve um resultado pior do que a abordagem instantânea com um baixo tamanho da janela, mas superior

Tabela 29 – Matriz de confusão para $K = 1$ e $t = 200$ abordagem centroide

Classes	Calma	Normal	Agressiva
Calma	100	0,30	0,00
Normal	0,00	98,53	0,33
Agressiva	0,00	1,157	99,67

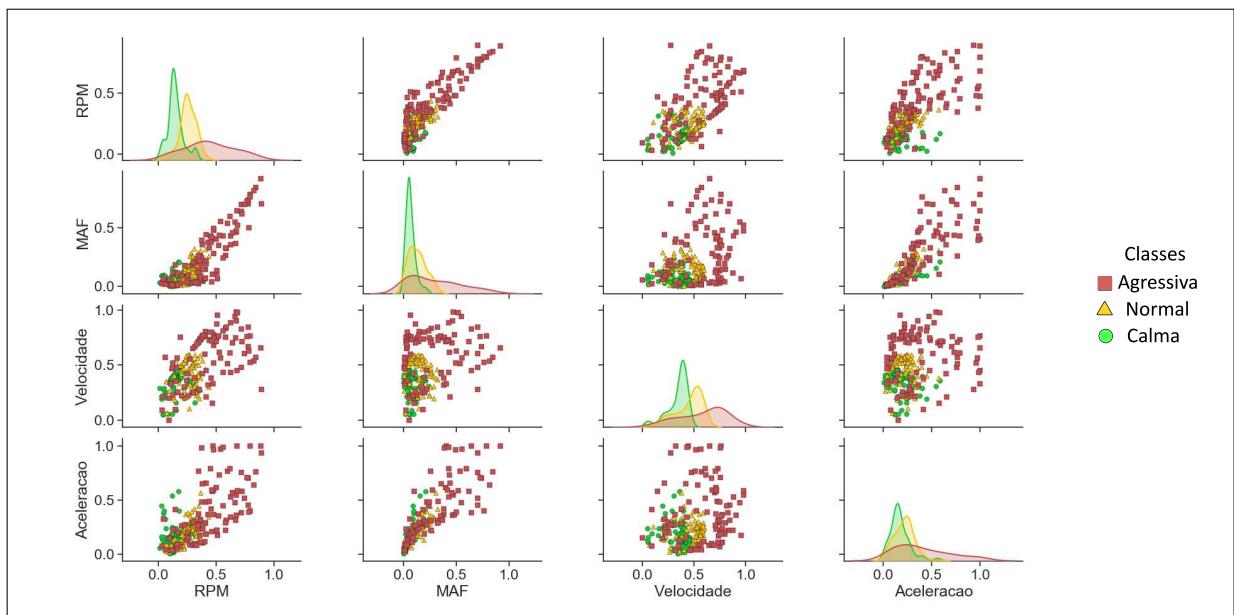
Fonte – Elaborado pelo Autor

com tamanhos de janela expressivos. Os resultados para cada uma das variações da abordagem de janela temporal são discutidos com mais detalhes nas seções a seguir.

7.1.3.1 Centroides

Como podemos observar na variação de janela temporal com centroide, em um baixo valor de janela temporal (Tabela 28) a proporção de falso-positivos foi semelhante para cada classe, indicando que esta abordagem possivelmente evidenciou pontos comuns entre todas as classes, os quais foram classificados erroneamente pelo algoritmo KNN. Como uma forma de compreender de maneira mais clara o resultado obtido, a disposição do conjunto de dados nesse caso é apresentada no gráfico da Figura 53. Neste, semelhante ao da Figura 41, todos os centroides dos grupos de atributos para $t = 6$ são plotados agrupadamente.

Figura 53 – Agrupamento dos centroides dos grupos de atributos pertencentes as três classes.

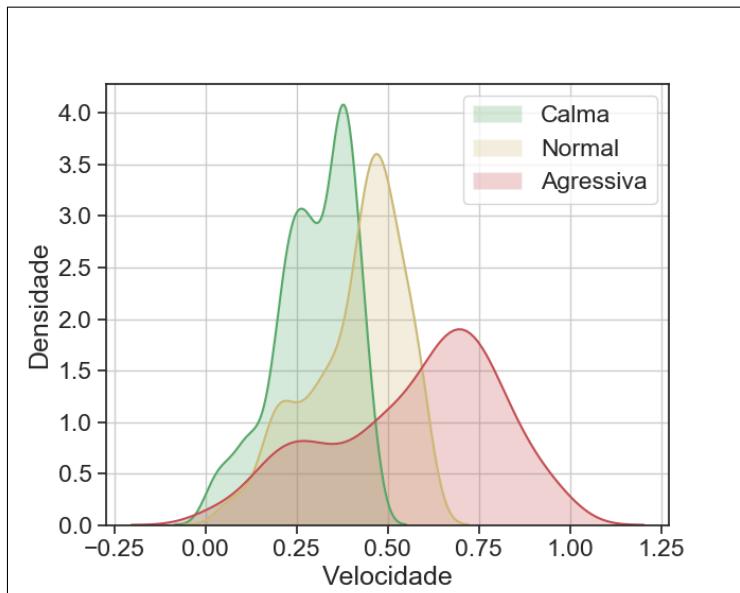


Fonte – Elaborado pelo Autor.

Observa-se que apesar da disposição dos pontos para cada classe estarem semelhantes

aos pontos da Figura 41, as distribuições dos atributos, plotadas na diagonal principal, mudaram. Como pode ser visto, os centroides obtiveram distribuições menos distintas do que no caso dos vetores isolados, com interseções mais expressivas entre as classes, comportamento evidente quando comparamos a Figura 42, distribuição para o atributo velocidade vista no Capítulo 6, com a Figura 54, que ilustra a distribuição do atributo velocidade para a variação centroide. Esse comportamento fez com que a acurácia do algoritmo diminuisse.

Figura 54 – Representação da distribuição do atributo Vel para cada classe.



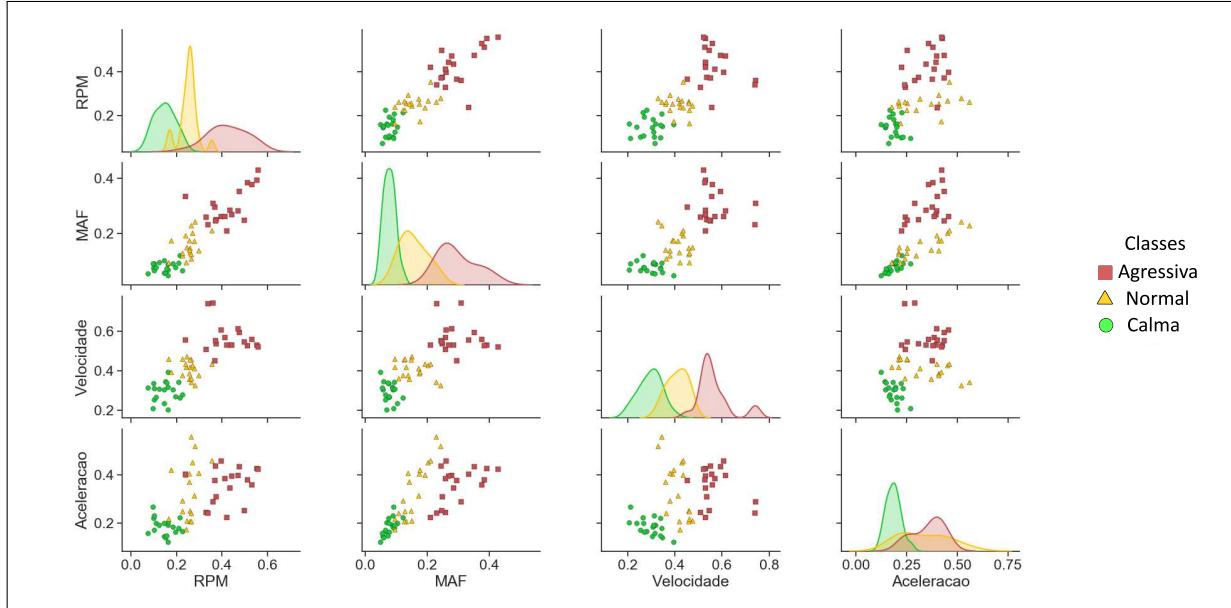
Fonte – Elaborado pelo Autor.

Quando a janela temporal aumenta, os centroides dos grupos distanciam-se, formando clusters com fronteiras bem definidas, o que permite uma melhor taxa de acerto por parte do classificador. A Figura 55 apresenta a disposição dos centroides dos grupos para uma janela $t = 200$ capturas, já a Figura 56 contém o gráfico da distribuição do atributo velocidade para este caso. Nestas, observamos a maior distinção entre as classes, justificando os resultados da Tabela 29.

7.1.3.2 Janelas temporais

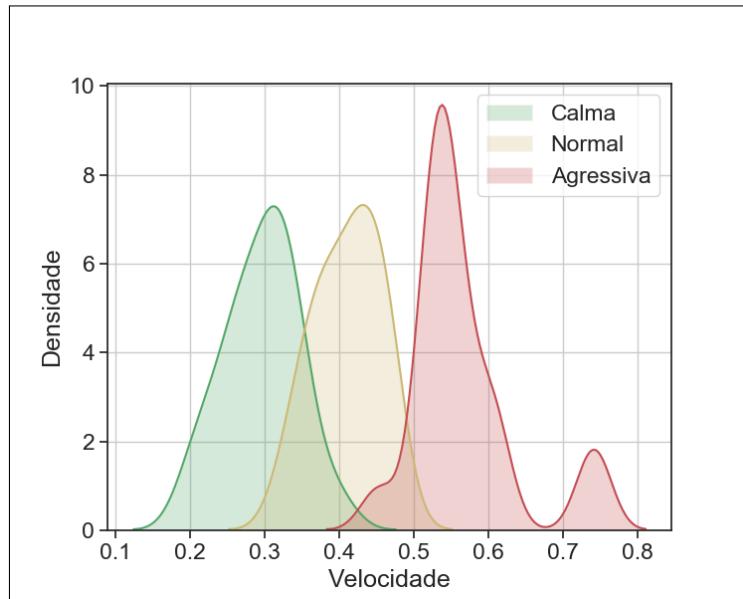
Na abordagem de janela temporal com a utilização de grupo de vetores, diferentemente da variação com centroides, as distribuições são semelhantes à abordagem instantânea. Então o que poderia causar a queda de desempenho observada? Para responder esta pergunta uma observação menos óbvia deve ser discutida. Na abordagem de janelas temporais o classificador

Figura 55 – Agrupamento dos centroides dos grupos de atributos pertencentes as três classes.



Fonte – Elaborado pelo Autor.

Figura 56 – Representação da distribuição do atributo Vel para cada classe.



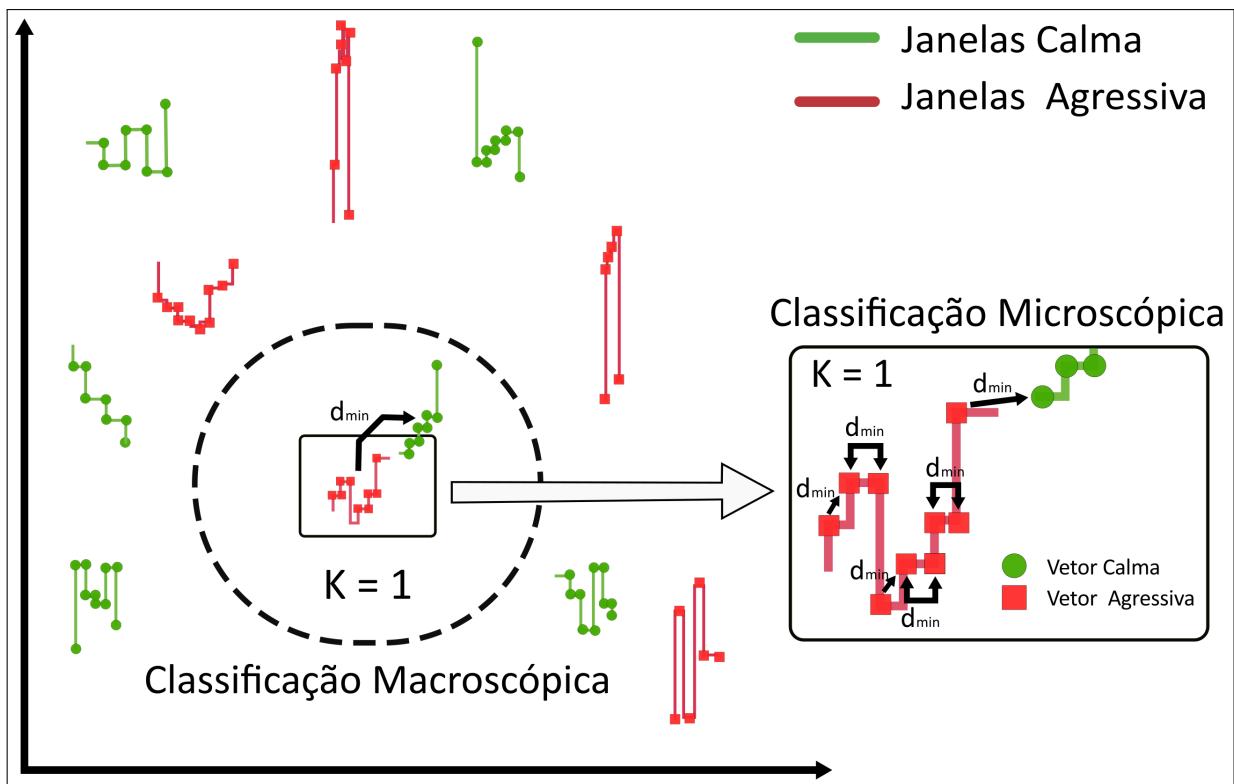
Fonte – Elaborado pelo Autor.

tem uma visão macroscópica dos atributos, já que estes estão formando um grupo, isto é, uma seção da curvas de atributos de uma classe. Já a abordagem instantânea possui uma visão microscópica sobre os dados, identificando cada valor dos atributos que compõem os grupos.

Tomando como exemplo a classe Agressiva, sabemos que esta também apresenta pontos de comportamento Calmo, porém ainda é predominantemente agressiva. O problema surge quando os classificadores com a abordagem instantânea e de janela temporal observam

os pontos da classe Agressiva próximos da Calma. A Figura 57 ilustra essa situação para $K = 1$. O classificador de janela temporal observa macroscopicamente uma seção da curva do estilo Agressivo e o classifica erroneamente como Calmo. Já o classificador instantâneo observa microscopicamente os valores de cada atributo e, dada a mesma seção do estilo Agressivo, realiza oito classificações corretas e apenas uma errada. Neste exemplo, o uso do primeiro classificador resultou em 100% das classificações como falso-positivos, já o uso do segundo resultou em 88,89% de classificações corretas e 11,11% de falso-positivos. Esse comportamento tende a ficar mais crítico quanto maior for o tamanho da janela, pois o classificador de janela temporal perderá detalhes dos atributos que são visíveis para o classificador instantâneo.

Figura 57 – Representação da diferença entre as abordagens instantânea e janela temporal



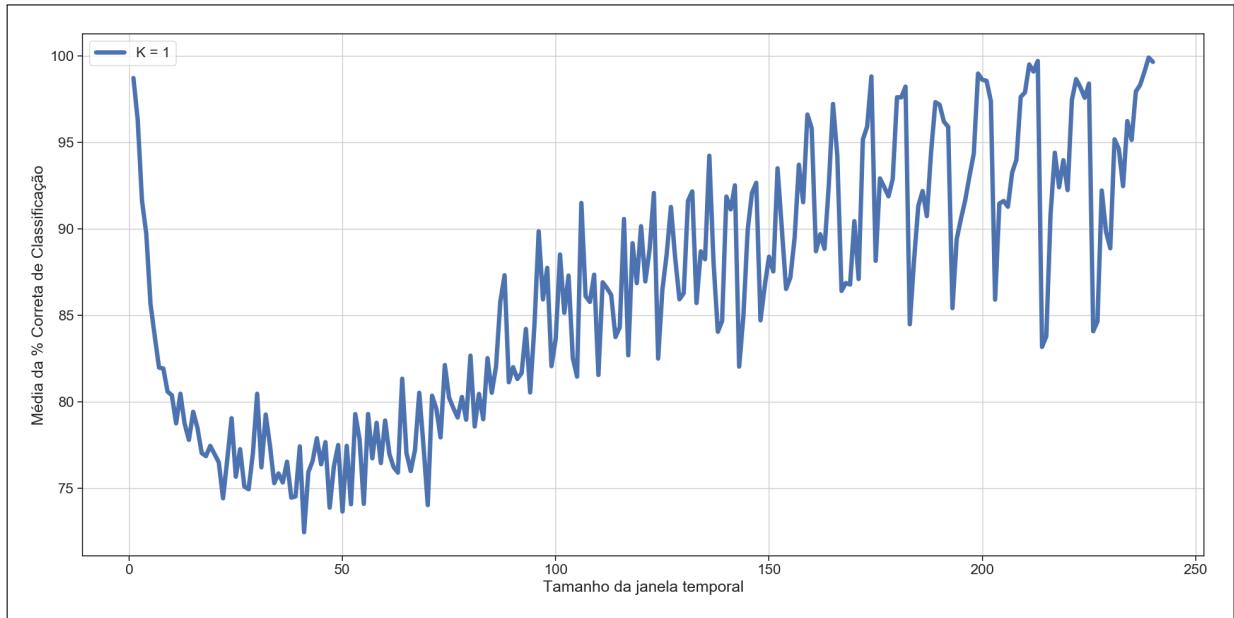
Fonte – Elaborado pelo Autor.

7.1.4 Oscilação

Apesar da tendência de aumento ou diminuição da média de acertos conforme a utilização de janelas temporais maiores, observa-se que esta média também torna-se mais instável. Para ilustrar esse comportamento, o gráfico da média de acertos para o classificador com apenas um vizinho na variação utilizando centroide é apresentado na Figura 58. Utilizou-se um maior

número de pontos do intervalo $t = [1, 240]$ capturas, para obter-se uma melhor resolução da curva e mostrar-se com mais clareza as variações.

Figura 58 – Variação na média da acurácia conforme o aumento da janela temporal



Fonte – Elaborado pelo Autor.

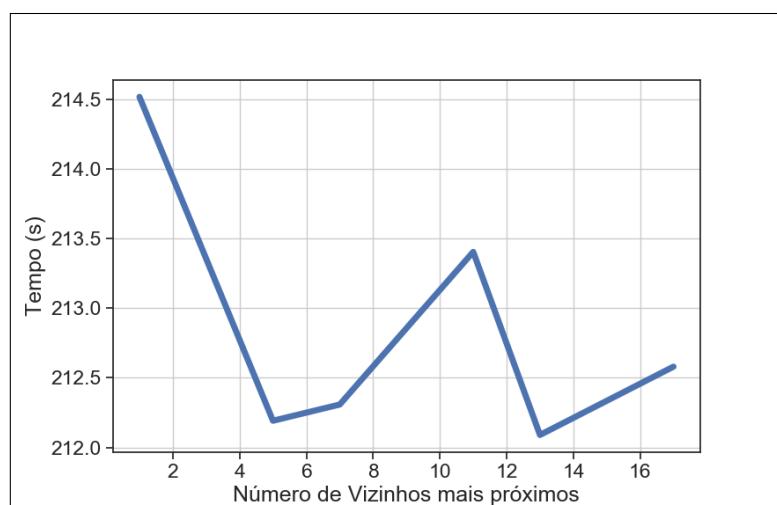
A variação ocorre devido à diminuição do número de pontos que podem ser utilizados pelo algoritmo para a classificação conforme a janela temporal aumenta. Com $t = 200$ capturas, por exemplo, os 11.520 pontos inciais são transformados em apenas 57 centroides, cada um calculado de um grupo de 200 vetores. Assim, com um banco de dados formado por apenas 57 pontos distintos, separados em 46 pontos para treino e 11 para teste, uma pequena variação no conjunto de dados pode influenciar bastante o resultado do classificador. Como o número de centroides diminui e seus valores mudam para cada janela temporal utilizada, a oscilação é mais presente em altos valores desta janela.

Como visto, a variação visualizada dependerá não só do tamanho da janela temporal mas também da magnitude do banco de dados. Dessa forma, caso o banco possua um tamanho fixo, como o utilizado neste trabalho, a melhor abordagem seria verificar o desempenho do classificador para diversas extensões de janela temporal, assim como foi feito, a fim de selecionar o ponto ótimo de atuação.

7.2 Custo Computacional

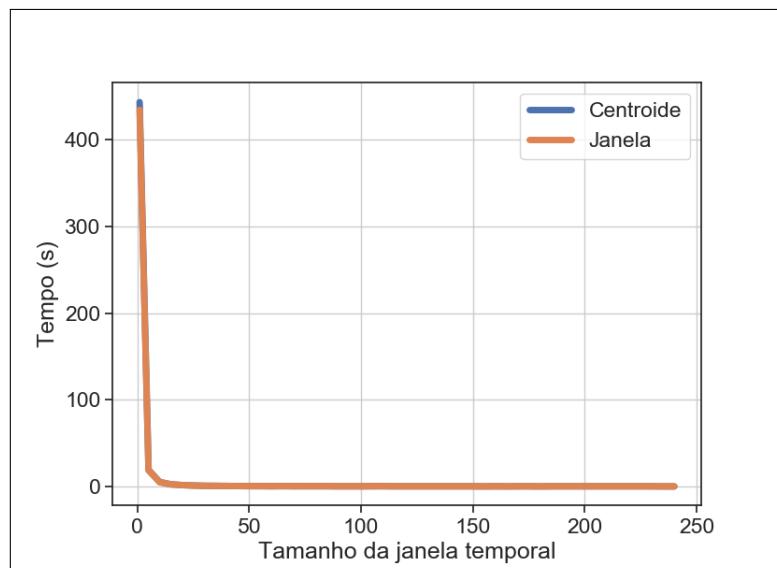
Por fim, os custos computacionais para as 100 repetições em cada um dos testes foram comparados. As Figuras 59 e 60 ilustram os gráficos de custo computacional em segundos para cada uma das abordagens. Percebemos que a abordagem instantânea apresenta um desempenho semelhante independente do número de vizinhos mais próximos. Dessa forma, o aumento de número de vizinho tende a diminuir a acurácia do classificador, causando uma queda da relação entre acurácia e custo computacional.

Figura 59 – Custo computacional para a abordagem instantânea.



Fonte – Elaborado pelo Autor.

Figura 60 – Custo computacional para as variações da abordagem de janela temporal.



Fonte – Elaborado pelo Autor.

Para a abordagem de janela temporal, por outro lado, observamos uma diminuição significativa do custo computacional com o aumento da janela, já que a divisão dos dados em grupos fará com que o classificador tenha que realizar uma menor quantidade de cálculos de distância. Dessa forma, a relação entre acurácia e custo computacional tende a crescer com o aumento da janela temporal. Entretanto, o aumento do desvio padrão em ambas variações de janela temporal e o decréscimo da taxa de acerto na primeira variação são os preços a serem pagos pelo melhor desempenho computacional. Dessa forma, as abordagens podem ser utilizadas em diferentes situações. Naquelas que o tempo é um fator primordial, a abordagem de janela temporal ganha força.

Para este trabalho, o envio dos dados para um servidor com grande poder computacional possibilita a utilização da abordagem instantânea. Porém, em situações nas quais pretende-se embarcar o classificador em um sistema mais simples, como no próprio PIC18f4550 ou a plataforma RaspberryPi, a abordagem de janela temporal apresenta uma melhor relação entre acurácia e custo computacional. A diferença entre o custo computacional da abordagem de janela temporal variação centroide com a abordagem instantânea é tão grande que a utilização de várias iterações para uma classificação é possível, reduzindo o problema do desvio elevado encontrado nas janelas temporais.

7.3 Resumo dos resultados

De acordo com os resultados obtidos, algumas considerações podem ser levantadas no tocante as características de cada abordagem, as quais são resumidas na Tabela 30. A abordagem instantânea apresentou uma ótimo desempenho, baixo desvio padrão, porém com custo computacional muito elevado. Além disso, apresenta uma boa estabilidade na classificação, como pôde ser visto, mesmo com um pequeno banco de treino ($ttRatio = 0,1$) esta apresentou bons resultados. A variação de janela temporal com grupo de vetores apresentou os piores resultados, com baixa acurácia e alto desvio padrão, porém com custo computacional baixo. Já a variação com centroides, juntando o melhor dos dois mundos, apresentou um ótimo desempenho e baixo custo computacional, porém necessita de um banco de dados extenso para sua aplicação, caso contrário sua acurácia média sofrerá de grandes oscilações com o aumento da janela, dificultando a escolha de um ponto ótimo de atuação, conforme foi constatado.

A Tabela 31 resume os resultados para os casos discutidos de cada abordagem. Considerando apenas os resultados de média de acurácia, a melhor abordagem dentre as testadas

Tabela 30 – Comparação entre as diferentes abordagens utilizadas com o classificador KNN.

Característica	Instantânea	Janela temporal	Centroide
Acurácia média	●●●	●○○	●●●
Desvio padrão	●○○	●●●	●●○
Custo computacional	●●●	●○○	●○○
Magnitude do banco de dados	●○○	●●●	●●●

Fonte – Elaborado pelo Autor

é a variação centroide. O custo computacional desta abordagem é muito baixo, significando que mesmo que seja necessário diversas classificações de uma mesma entrada para compensar o alto desvio padrão em alguns casos, ela ainda apresentará uma relação custo computacional/acurácia maior do que a abordagem instantânea. Entretanto, o banco de dados montado não é suficientemente grande para impedir a oscilação de sua acurácia conforme o aumento da janela temporal. Por conta disso, todo um espectro de possíveis janelas temporais e número de vizinhos foi varrido, para que um ponto ótimo de atuação do classificador pudesse ser selecionado, o qual mostrou-se ser para $K = 1$ e $t = 200$, no qual obteve-se uma acurácia de 99,39 %.

Tabela 31 – Comparação entre os resultados obtidos

Abordagem	Máximo (%)	Mínimo (%)	Médio (%)	Desvio Padrão	Custo por iteração (ms)
Instantânea ($K = 1$)	99,526	92,574	96,808	1,085	2147
Janela ($K = 1$ e $t = 6$)	92,76	54,545	82,70	2,10	18,61
Janela ($K = 1$ e $t = 200$)	95,2	42,41	73,50	13,65	1
Centroide ($K=1$ e $t=6$)	90,82	81,01	86,36	2,30	18,98
Centroide ($K=1$ e $t=200$)	100	92,54	99,39	1,78	1

Fonte – Elaborado pelo Autor

8 CONCLUSÕES E TRABALHOS FUTUROS

Este capítulo discute as conclusões, considerações finais, dificuldades encontradas e resume os resultados obtidos durante o desenvolvimento deste trabalho. Além disso, propostas de trabalhos futuros são elencadas.

No Capítulo 4, um sistema de telemetria automotivo autoral conectado à internet foi desenvolvido. O sistema é capaz de capturar informações do sistemas eletromecânicos de um automóvel utilizando o sistema de diagnóstico OBDII. Todo o desenvolvimento de hardware, software e esquemáticos foram discutidos, sendo estes também disponibilizados para livre acesso e desenvolvimento¹. O sistema de telemetria mostrou-se estável e eficiente, com erros de comunicação em apenas 0,3% de todos os dados capturados. Este sistema de telemetria corresponde à primeira contribuição deste trabalho.

Através deste sistema, como discutido no Capítulo 6, um banco de dados online foi criado. Das 200 possíveis variáveis disponibilizadas pelo OBDII, as quais dependem do modelo e fabricante do veículo, foram escolhidas àquelas que possibilitariam a modelagem da dinâmica de um sistema automotivo. Além disso, data, hora e localização do veículo também foram capturadas, permitindo uma correspondência temporal entre as variáveis. Dado que o sistema de telemetria é autoral, a quantidade de informações e taxa de captura podem ser alteradas pelo usuário, possibilitando a expansão deste banco de acordo com uma dada necessidade. Neste trabalho o banco utilizado corresponde à 11.520 vetores de capturas dos atributos RPM, MAF, Vel, Acl e Elo para três classes de estilo de direção: Calma, Normal e Agressiva.

Ainda no Capítulo 6, as variáveis capturadas foram analisadas. Percebeu-se que as variáveis Elo e Acl são bastante correlacionadas e relacionam-se com todos os outros atributos de forma semelhante, logo decidiu-se descartar Elo e utilizar apenas a variável Acl por esta ser diretamente relacionada ao motorista. Os atributos selecionados RPM, MAF, Vel e Acl foram analisados para cada classe. Notou-se que o estilo de direção de um motorista é, na verdade, composto por todas as classes, porém com predominância de uma delas. Esse comportamento é explicado pela existência de eventos nos quais o estilo de direção é limitado, como curvas, redutores de velocidade e situações de parada. Por conta disso, as distribuições das classes possuem interseções. O banco de dados criado com os atributos selecionados é a segunda contribuição deste trabalho.

No Capítulo 7, o algoritmo KNN foi utilizado para classificação do estilo de direção,

¹ <https://github.com/wborbaneto/research>

utilizando, para isso, o banco de dados criado como entrada. Duas abordagens foram utilizadas: instantânea e janela temporal; cujos melhores resultados são resumidos na Tabela 32.

Tabela 32 – Comparação entre os resultados obtidos

Abordagem	Máximo (%)	Mínimo (%)	Médio (%)	Desvio Padrão	Custo por iteração (ms)
Instantânea	99,52	92,57	96,81	1,08	201,7
Janela	92,76	54,545	82,70	2,10	186,1
Centroide	100	92,54	99,39	1,78	1

Fonte – Elaborado pelo Autor

A premissa inicial de que a abordagem de janela temporal apresentaria um melhor resultado na classificação, dado que esta preserva informações da dinâmica dos atributos, foi invalidada apenas para a variação de janelas utilizando grupos de vetores, a qual apresentou o pior desempenho dentre os testes realizados. Dessa forma, para classificação de estilos de direção com o algoritmo KNN os melhores resultados foram obtidos com a abordagem janela temporal variação centroide, com a acurácia média 99,39 % para $K = 1$ e $t = 200$ no ponto ótimo da classificação, como pode ser visto na Tabela 32, que compara os valores médios, máximos, mínimos e variação para cada uma das abordagens.

Apesar do ótimo desempenho da variação centroide, é necessário um banco de dados de grande magnitude para garantir estabilidade no desempenho desta abordagem conforme o aumento da janela temporal. Caso o banco não seja suficientemente grande, inicialmente deve ser feita uma varredura do desempenho do classificador com a variação do tamanho da janela, para que um ponto ótimo de atuação seja escolhido. Neste estudo, a janela de tamanho $t = 200$ capturas foi escolhida, dentro do intervalo que vai de 1 até 240 capturas.

Além disso, observou-se, também na Tabela 32, que o custo computacional por iteração da abordagem instantânea é muito superior ao da janela temporal, fazendo com que esta última ganhe força quando aplicada em sistemas de poder computacional limitado. Dessa forma, mesmo que a classificação da janela temporal apresente um valor de variância maior do que a abordagem instantânea, isto pode ser compensado por um maior número de testes sem comprometer a performance do classificador.

A divergência dos resultados entre as abordagens é explicada pela forma com a qual o classificador KNN compara os dados de entrada com o banco de treino. Uma visão macroscópica, observada na abordagem de janela temporal, compromete a classificação em pontos localizados nas interseções entre as classes. Já uma visão microscópica, característica da

abordagem instantânea, é menos influenciada por esse fenômeno. A discussão de utilização do KNN e a discussão levantada sobre as abordagens instantânea e janela temporal são a terceira contribuição deste trabalho.

8.1 Propostas de Trabalhos Futuros

Qualquer que seja o estilo de condução do motorista, este apresentará características não condizentes com sua classificação. Por mais agressivo que um motorista possa ser, ele sempre passará por eventos, como curvas e redutores de velocidade, que o forçam a agir como um motorista de outra classe, como da classe Normal, por exemplo. Esses desvios de classificação podem ser evitados caso essas situações sejam identificadas e aplicadas em processos de classificação específicos para cada uma, o que deixa o processo como um todo mais preciso.

Além disso, para identificação de padrões, quanto maior for a quantidade de dados e situações capturadas, mais refinado será o processo. Por isso, os trabalhos futuros envolvem aumentar abrangência das capturas, com um maior número de veículos, motoristas e percursos distintos, possibilitando a criação de uma lógica que permita a identificação de diversas situações pelas quais o motorista passa - como engarrafamentos, curvas e buracos na via – melhorando, assim, a eficácia da classificação.

Pretende-se, também, estender o número de classificadores utilizados e realizar uma nova análise comparativa entre estes. O Sistema de Inferência Fuzzy (FIS), em especial, aparenta apresentar uma abordagem interessante para esta análise, já que a complexidade da análise evidenciada neste trabalho mostra que a classificação do estilo de um motorista pode depender de diversos fatores e não possui uma representação matemática exata, demonstrada pela evidente interseção entre as distribuições das classes. Por conta desta indefinição, o FIS se apresenta como um promissor método de classificação para esta abordagem.

REFERÊNCIAS

- AAA - Foundation for Trafic Safety. *Prevalence of self-reported Aggressive Behaviour: United States, 2014*. United States, 2016.
- ABUALI, N.; ABOU-ZEID, H. Driver behavior modeling: Developments and future directions. **International journal of vehicular technology**, Hindawi, v. 2016, 2016.
- ALJAAFREH, A.; ALSHABATAT, N.; AL-DIN, M. S. N. Driving style recognition using fuzzy logic. In: IEEE. **Vehicular Electronics and Safety (ICVES), 2012 IEEE International Conference on**. [S.l.], 2012. p. 460–463.
- AMADO, S. *et al.* How accurately do drivers evaluate their own driving behavior? an on-road observational study. **Accident Analysis & Prevention**, Elsevier, v. 63, p. 65–73, 2014.
- ARAÚJO, R. *et al.* Driving coach: A smartphone application to evaluate driving efficient patterns. In: IEEE. **Intelligent Vehicles Symposium (IV), 2012 IEEE**. [S.l.], 2012. p. 1005–1010.
- Associação Brasileira de Prevenção dos Acidentes de Trânsito. **Estatísticas do Ministério da Saúde**. Brasil, 2017.
- BETANCUR, J. A. *et al.* Research topics and implementation trends on automotive head-up display systems. **International Journal on Interactive Design and Manufacturing (IJIDeM)**, Springer, v. 12, n. 1, p. 199–214, 2018.
- BOLOVINOU, A. *et al.* Driving style recognition for co-operative driving: A survey. In: CITESEER. **The Sixth International Conference on Adaptive and Self-Adaptive Systems and Applications**. [S.l.], 2014. p. 73–78.
- BOUHOUTE, A. *et al.* Advanced driving behavior analytics for an improved safety assessment and driver fingerprinting. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, n. 99, p. 1–14, 2018.
- BRAMBILLA, M.; MASCETTI, P.; MAURI, A. Comparison of different driving style analysis approaches based on trip segmentation over gps information. In: IEEE. **Big Data (Big Data), 2017 IEEE International Conference on**. [S.l.], 2017. p. 3784–3791.
- CASTIGNANI, G. *et al.* Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. **IEEE Intelligent Transportation Systems Magazine**, IEEE, v. 7, n. 1, p. 91–102, 2015.
- CASTIGNANI, G.; FRANK, R.; ENGEL, T. Driver behavior profiling using smartphones. In: IEEE. **Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on**. [S.l.], 2013. p. 552–557.
- CHEN, S.-W.; FANG, C.-Y.; TIEN, C.-T. Driving behaviour modelling system based on graph construction. **Transportation research part C: emerging technologies**, Elsevier, v. 26, p. 314–330, 2013.
- CHHABRA, R.; VERMA, S.; KRISHNA, C. R. A survey on driver behavior detection techniques for intelligent transportation systems. In: IEEE. **Cloud Computing, Data Science & Engineering-Confluence, 2017 7th International Conference on**. [S.l.], 2017. p. 36–41.

- CUI, Y. X. *et al.* Interactive safety analysis framework of autonomous intelligent vehicles. In: EDP SCIENCES. **MATEC Web of Conferences**. [S.I.], 2016. v. 44, p. 01029.
- DERBEL, O. *et al.* Driving style assessment based on the gps data and fuzzy inference systems. In: IEEE. **Systems, Signals & Devices (SSD), 2015 12th International Multi-Conference on**. [S.I.], 2015. p. 1–8.
- DINGUS, T. A. *et al.* **The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment**. California, 2006.
- DÖRR, D.; GRABENGIESSER, D.; GAUTERIN, F. Online driving style recognition using fuzzy logic. In: IEEE. **Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on**. [S.I.], 2014. p. 1021–1026.
- ELANDER, J.; WEST, R.; FRENCH, D. Behavioral correlates of individual differences in road-traffic crash risk: An examination of methods and findings. **Psychological bulletin**, American Psychological Association, v. 113, n. 2, p. 279, 1993.
- ERICSSON, E. Variability in urban driving patterns. **Egyptian Journal of Medical Human Genetics**, Elsevier, p. 337–354, May 2000.
- FARZANEH, M. *et al.* Texas-specific drive cycles and idle emissions rates for using with epa's moves model—final report. **Texas Department of Transportation, Austin**, Citeseer, 2014.
- FENG, Y. *et al.* Driving style modelling with adaptive neuro-fuzzy inference system and real driving data. In: SPRINGER. **International Conference on Applied Human Factors and Ergonomics**. [S.I.], 2018. p. 481–490.
- FERNANDEZ, S.; ITO, T. Driver classification for intelligent transportation systems using fuzzy logic. In: IEEE. **Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on**. [S.I.], 2016. p. 1212–1216.
- FUGIGLANDO, U. *et al.* Driving behavior analysis through can bus data in an uncontrolled environment. **arXiv preprint arXiv:1710.04133**, 2017.
- FUKUNAGA, K.; NARENDRA, P. M. A branch and bound algorithm for computing k-nearest neighbors. **IEEE transactions on computers**, IEEE, v. 100, n. 7, p. 750–753, 1975.
- GĂVRUTĂ, A.; MARCU, M.; BOGDAN, R. Software solution for monitoring and analyzing driver behavior. In: IEEE. **2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI)**. [S.I.], 2018. p. 000345–000350.
- HASSAN, H. M. *et al.* Investigation of drivers' behavior towards speeds using crash data and self-reported questionnaire. **Accident Analysis & Prevention**, Elsevier, v. 98, p. 348–358, 2017.
- HAVENS, T. C.; BEZDEK, J. C.; PALANISWAMI, M. Scalable single linkage hierarchical clustering for big data. In: IEEE. **Intelligent Sensors, Sensor Networks and Information Processing, 2013 IEEE Eighth International Conference on**. [S.I.], 2013. p. 396–401.
- HELMAN, S.; REED, N. Validation of the driver behaviour questionnaire using behavioural data from an instrumented vehicle and high-fidelity driving simulator. **Accident Analysis & Prevention**, Elsevier, v. 75, p. 245–251, 2015.

- HEWERDINE, W. **NMEA Reference Manual**. [S.l.]: SIRF Technology, Inc, San Jose, 2005.
- HIGGS, B.; ABBAS, M. Segmentation and clustering of car-following behavior: Recognition of driving patterns. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 16, n. 1, p. 81–90, 2015.
- HWANG, C.-p. *et al.* Apply scikit-learn in python to analyze driver behavior based on obd data. In: IEEE. **2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)**. [S.l.], 2018. p. 636–639.
- IIZUKA, G.; SAITO, Y.; YAMADA, M. Development of 4k driving simulator and measurement of eye movement and physiological data during the driving. In: IEEE. **Consumer Electronics (GCCE), 2015 IEEE 4th Global Conference on**. [S.l.], 2015. p. 187–188.
- ISO. **ISO 9141-2:Road vehicles — Diagnostic systems — Part 2: CARB requirements for interchange of digital information**. 1994.
- ISO. **ISO 15031:Road vehicles – Communication between vehicle and external equipment for emissions-related diagnostics – Part 5: Emissions-related diagnostic services**. 2015.
- JACHIMCZYK, B. *et al.* Iot on-board system for driving style assessment. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 18, n. 4, p. 1233, 2018.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM computing surveys (CSUR)**, Acm, v. 31, n. 3, p. 264–323, 1999.
- JARAMILLO, D. M. G.; NARVAEZ, C. Vehicle online monitoring system based on fuzzy classifier. **Proc. Vehicular**, p. 33–38, 2014.
- JÚNIOR, J. F. *et al.* Driver behavior profiling: An investigation with different smartphone sensors and machine learning. **PLoS one**, Public Library of Science, v. 12, n. 4, p. e0174959, 2017.
- KALITA, P. Power train and emission control: allocation procedure by obd-ii system for automotive technology. In: IOP PUBLISHING. **IOP Conference Series: Materials Science and Engineering**. [S.l.], 2017. v. 211, n. 1, p. 012019.
- KAPLAN, S. *et al.* Driver behavior analysis for safe driving: A survey. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 16, n. 6, p. 3017–3032, 2015.
- KU, I. *et al.* Towards software-defined vanet: Architecture and services. In: IEEE. **2014 13th annual Mediterranean ad hoc networking workshop (MED-HOC-NET)**. [S.l.], 2014. p. 103–110.
- KWAK, B. I.; WOO, J.; KIM, H. K. Know your master: Driver profiling-based anti-theft method. In: IEEE. **2016 14th Annual Conference on Privacy, Security and Trust (PST)**. [S.l.], 2016. p. 211–218.
- LI, N.; MISU, T.; MIRANDA, A. Driver behavior event detection for manual annotation by clustering of the driver physiological signals. In: IEEE. **Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on**. [S.l.], 2016. p. 2583–2588.

LI, Q.; QIAO, F.; YU, L. A machine learning approach for light-duty vehicle idling emission estimation based on real driving and environmental information. **Environ Pollut Climate Change**, v. 1, n. 106, p. 2, 2016.

LIN, N. *et al.* An overview on study of identification of driver behavior characteristics for automotive control. **Mathematical Problems in Engineering**, Hindawi, v. 2014, 2014.

LIU, T. *et al.* Cluster regularized extreme learning machine for detecting mixed-type distraction in driving. In: IEEE. **Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on**. [S.l.], 2015. p. 1323–1326.

LÓPEZ, J. O.; PINILLA, A. C. C. *et al.* Driver behavior classification model based on an intelligent driving diagnosis system. In: IEEE. **Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on**. [S.l.], 2012. p. 894–899.

LU, Y. *et al.* Influence analysis of driving pattern factors on bus fuel economy by the sem. In: IEEE. **Control Conference (CCC), 2016 35th Chinese**. [S.l.], 2016. p. 8920–8925.

MARTINELLI, F. *et al.* Human behavior characterization for driving style recognition in vehicle system. **Computers & Electrical Engineering**, Elsevier, 2018.

MARTINEZ, C. M. *et al.* Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 19, n. 3, p. 666–676, 2018.

MASCIO, P. D. *et al.* Innovative guidelines and tools for vulnerable road users safety in india and brazil [saferbrain]. d3. 2 [v3] analysis of innovative methodologies. European (7th RTD Framework Programme), 2011.

MATOS, F. M. d. S. Reconhecimento de faces utilizando a transformada cosseno discreta. **Masther Thesis**, 2008.

MEIRING, G. A. M.; MYBURGH, H. C. A review of intelligent driving style analysis systems and related artificial intelligence algorithms. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 15, n. 12, p. 30653–30682, 2015.

MEULENERS, L.; FRASER, M. A validation study of driving errors using a driving simulator. **Transportation research part F: traffic psychology and behaviour**, Elsevier, v. 29, p. 14–21, 2015.

MIYAJIMA, C. *et al.* Driver modeling based on driving behavior and its evaluation in driver identification. **Proceedings of the IEEE**, IEEE, v. 95, n. 2, p. 427–437, 2007.

NHTSA. **National motor vehicle crash causation survey: Report to congress**. United States, 2008.

NIRMALI, B. *et al.* Vehicular data acquisition and analytics system for real-time driver behavior monitoring and anomaly detection. In: IEEE. **Industrial and Information Systems (ICIIS), 2017 IEEE International Conference on**. [S.l.], 2017. p. 1–6.

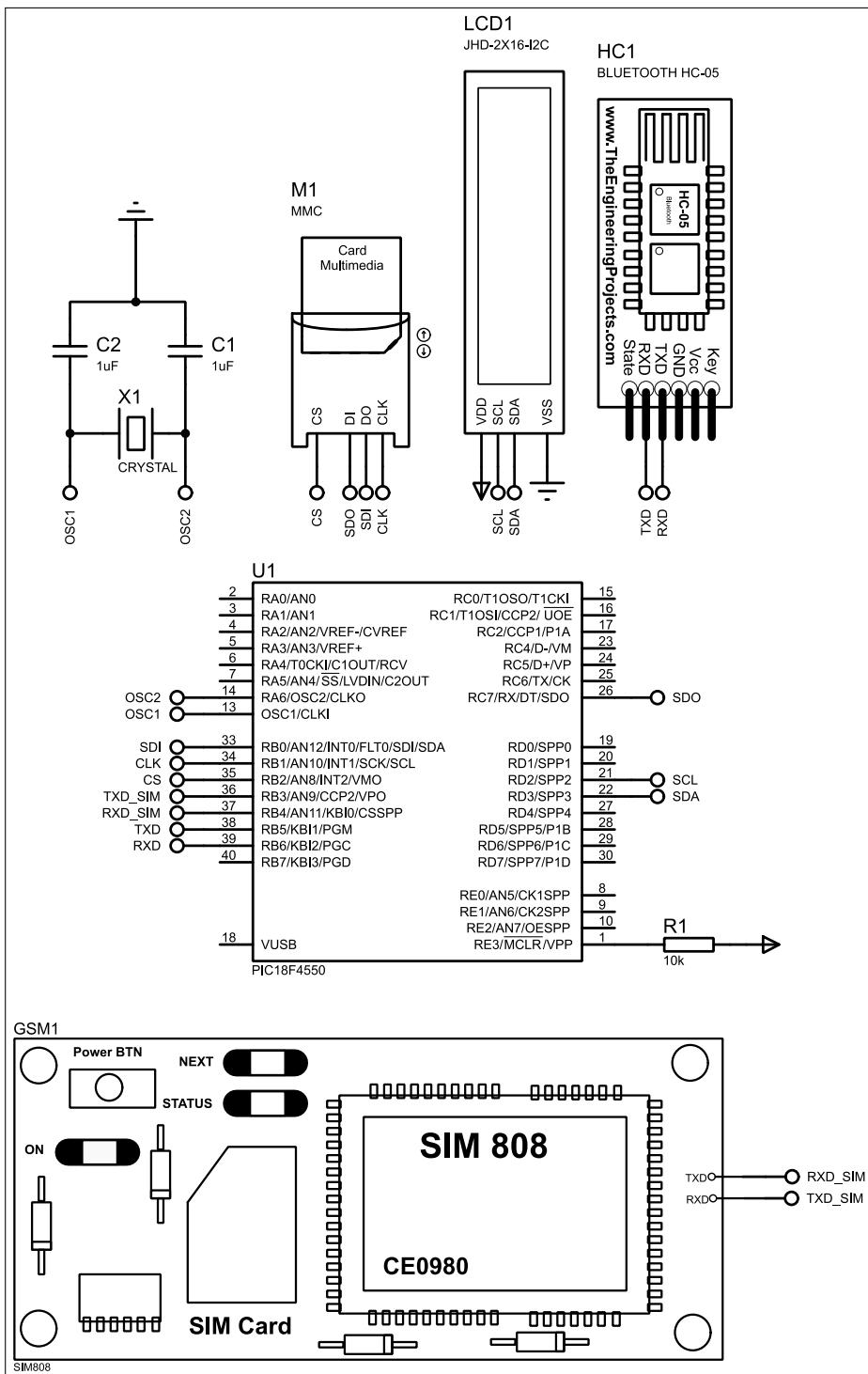
Observatório Nacional de Segurança Viária. **Atualização do custo total dos acidentes de trânsito no Brasil**. Brasil, 2015.

- OGITSU, T.; MIZOGUCHI, H. A study on driver training on advanced driver assistance systems by using a driving simulator. In: IEEE. **Connected Vehicles and Expo (ICCVE), 2015 International Conference on**. [S.l.], 2015. p. 352–353.
- OLOFSSON, M.; PETTERSSON, J. Parameterization and validation of road and driver behavior models for carmaker simulations and transmission hil-rig. **Master's thesis, Department of Applied Mechanics, Chalmers University of Technology, Sweden**, 2015.
- OLSON, R. L. *et al.* **Driver distraction in commercial vehicle operations**. [S.l.], 2009.
- PEREIRA, A.; ALVES, M.; MACEDO, H. Vehicle driving analysis in regards to fuel consumption using fuzzy logic and obd-ii devices. In: IEEE. **Telematics and Information Systems (EATIS), 2016 8th Euro American Conference on**. [S.l.], 2016. p. 1–4.
- PICARD, R. R.; COOK, R. D. Cross-validation of regression models. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 79, n. 387, p. 575–583, 1984.
- REAL, E. M.; NICOLETTI, M. do C. Investigação de algoritmos sequenciais de agrupamento com pré-processamento de dados em aprendizado de máquina. **Faculdade Campo Limpo Paulista**, p. 21, 2014.
- REASON, J. *et al.* Errors and violations on the roads: a real distinction? **Ergonomics**, Taylor & Francis, v. 33, n. 10-11, p. 1315–1332, 1990.
- ROCHA NETO, W. B.; MENEZES JUNIOR, J. M. P. Perfil de condução de motoristas: Análise por meio de algoritmos de reconhecimentos de padrões. In: **Congresso Brasileiro de Automática (CBA'2018)**. [S.l.: s.n.], 2018.
- ROWE, R. *et al.* Measuring errors and violations on the road: A bifactor modeling approach to the driver behavior questionnaire. **Accident Analysis & Prevention**, Elsevier, v. 74, p. 118–125, 2015.
- RUTA, M. *et al.* A mobile knowledge-based system for on-board diagnostics and car driving assistance. In: CITESEER. **UBICOMM 2010, The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies**. [S.l.], 2010. p. 91–96.
- SADIA, R.; BEKHOR, S.; POLUS, A. Structural equations modelling of drivers' speed selection using environmental, driver, and risk factors. **Accident Analysis & Prevention**, Elsevier, v. 116, p. 21–29, 2018.
- SAE. **J1850: Class B Data Communication Network Interface**. 1988.
- SAE. **SAE J1979: E/E Diagnostic Test Modes**. 2012.
- SAGBERG, F. *et al.* A review of research on driving styles and road safety. **Human factors**, Sage Publications Sage CA: Los Angeles, CA, v. 57, n. 7, p. 1248–1275, 2015.
- SCHWENKER, F.; TRENTIN, E. Pattern classification and clustering: A review of partially supervised learning approaches. **Pattern Recognition Letters**, Elsevier, v. 37, p. 4–14, 2014.
- SHARMA, H. *et al.* S-road assist: Road surface conditions and driving behavior analysis using smartphones. In: IEEE. **Connected Vehicles and Expo (ICCVE), 2015 International Conference on**. [S.l.], 2015. p. 291–296.

- SHIRKHORSHIDI, A. S. *et al.* Big data clustering: a review. In: SPRINGER. **International Conference on Computational Science and Its Applications**. [S.I.], 2014. p. 707–720.
- SINGH, S. **Critical reasons for crashes investigated in the national motor vehicle crash causation survey**. Washington, DC, 2015.
- SRINIVASAN, A. IoT cloud based real time automobile monitoring system. In: IEEE. **2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)**. [S.I.], 2018. p. 231–235.
- STEINBACH, T.; KORF, F.; SCHMIDT, T. C. Real-time ethernet for automotive applications: A solution for future in-car networks. In: IEEE. **Consumer Electronics-Berlin (ICCE-Berlin), 2011 IEEE International Conference on**. [S.I.], 2011. p. 216–220.
- SUN, B. *et al.* Research on the classification and identification of driver's driving style. In: IEEE. **Computational Intelligence and Design (ISCID), 2017 10th International Symposium on**. [S.I.], 2017. v. 1, p. 28–32.
- TSBEC. On-Board Diagnostics II (OBDII) and Light-Duty Vehicle Emission related Inspection (I/M) Programs**. 2004.
- United Nations Road Safety Collaboration and others. Global plan for the decade of action for road safety 2011-2020. Geneva, 2011.
- VAIANA, R. *et al.* Driving behavior and traffic safety: an acceleration-based safety evaluation procedure for smartphones. **Modern Applied Science**, v. 8, n. 1, p. 88, 2014.
- VIROJBOONKIADE, N.; VATEEKUL, P.; ROJVIBOONCHAI, K. Driver identification using histogram and neural network from acceleration data. In: IEEE. **Communication Technology (ICCT), 2017 IEEE 17th International Conference on**. [S.I.], 2017. p. 1560–1564.
- WANG, R.; LUKIC, S. M. Review of driving conditions prediction and driving style recognition based control algorithms for hybrid electric vehicles. In: IEEE. **Vehicle Power and Propulsion Conference (VPPC), 2011 IEEE**. [S.I.], 2011. p. 1–7.
- WANG, W.; XI, J. A rapid pattern-recognition method for driving styles using clustering-based support vector machines. In: IEEE. **American Control Conference (ACC), 2016**. [S.I.], 2016. p. 5270–5275.
- WITT, M. *et al.* Cognitive driver behavior modeling: Influence of personality and driver characteristics on driver behavior. In: SPRINGER. **International Conference on Applied Human Factors and Ergonomics**. [S.I.], 2018. p. 751–763.
- World Health Organization. **Global status report on road safety 2015**. Geneva: World Health Organization, 2015.
- XU, J.; SHU, H.-B.; SHAO, Y.-M. Modeling of driver behavior on trajectory-speed decision making in minor traffic roadways with complex features. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, n. 99, p. 1–13, 2018.
- YANGUO, H.; LUNHUI, X.; YA, G. The urban road traffic state identification method based on fcm clustering. In: IEEE. **Transportation, Mechanical, and Electrical Engineering (TMEE), 2011 International Conference on**. [S.I.], 2011. p. 596–600.

- YUAN, Q. *et al.* Cluster and factor analysis on data of fatal traffic crashes in china. In: IEEE. **Transportation Information and Safety (ICTIS), 2017 4th International Conference on**. [S.I.], 2017. p. 211–224.
- YUEPENG, C. *et al.* Cluster analysis on evaluation indicators of driver characteristics. In: IEEE. **Intelligent Transportation Systems Conference, 2007. ITSC 2007**. IEEE. [S.I.], 2007. p. 70–75.
- ZHANG, X.; ZHAO, X.; RONG, J. A study of individual characteristics of driving behavior based on hidden markov model. **Sensors & Transducers**, IFSA Publishing, SL, v. 167, n. 3, p. 194, 2014.
- ZHAO, F.; ZHAO, Y.; TIAN, Y. Driver lane changing intention recognition based on multi-class information. In: IEEE. **2018 Chinese Control And Decision Conference (CCDC)**. [S.I.], 2018. p. 3439–3444.

ANEXO A – ESQUEMÁTICO DO SISTEMA DE TELEMETRIA



Fonte – Elaborado pelo Autor.