



Pós-Graduação em Ciência da Computação

ERONIDES FELISBERTO DA SILVA NETO

Um Sistema de Detecção de Anomalias em Sensor Veicular baseado em Classificadores One-Class



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
<http://cin.ufpe.br/~posgraduacao>

Recife
2019

ERONIDES FELISBERTO DA SILVA NETO

**Um Sistema de Detecção de Anomalias em Sensor Veicular baseado em
Classificadores One-Class**

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Sistemas Embarcados.

Orientador: Prof. Dr. Abel Guilhermino da Silva Filho

Recife
2019

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S586s Silva Neto, Eronides Felisberto da
Um sistema de detecção de anomalias em sensor veicular baseado em classificadores *one-class* / Eronides Felisberto da Silva Neto. – 2019.
108 f.: il., fig., tab.

Orientador: Abel Guilhermino da Silva Filho.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2019.
Inclui referências e apêndice.

1. Ciência da computação. 2. Sistemas embarcados. 3. Telemetria veicular.
I. Silva Filho, Abel Guilhermino da (orientador). II. Título.

004

CDD (23. ed.)

UFPE- MEI 2019-116

Eronides Felisberto da Silva Neto

“Um Sistema de Detecção de Anomalias em Sensor Veicular Baseado em Classificadores One-Class”

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 30/05/2019.

BANCA EXAMINADORA

Prof. Dr. Divanilson Rodrigo de Sousa Campelo
Centro de Informática / UFPE

Prof. Dr. Max Mauro Dias Santos
Departamento de Eletrônica / UTFPR-PG

Prof. Dr. Abel Guilhermino da Silva Filho
Centro de Informática/UFPE
(Orientador)

AGRADECIMENTOS

Felicidade é poder agradecer por todas oportunidades que a vida me deu.

Agradeço pelo apoio incondicional, carinho e exemplo de tudo que uma pessoa pode conquistar, pessoalmente e profissionalmente, a você que sempre esteve comigo, minha mãe, Amélia. Para o abraço, de tão forte, não tem separação, para você quem me ensinou o valor da raiz quadrada de dois, "*um vírgula quatrocentos e quatorze*", sem eu mesmo saber o quanto isso seria importante na minha vida, a você pai! Também agradeço minha família e amigos por apoiarem mais esta etapa da minha vida acadêmica.

Durante o período deste mestrado também pude realizar pesquisas no laboratório de inovação veicular (LIVE), na UFPE. Agradeço a *Fiat Chrysler Automobiles* (FCA) pelo incentivo à pesquisa. Durante o período correspondente a participação nos projetos do laboratório, foi possível se aprofundar nos desafios e problemas da indústria automotiva, no contexto de inovação veicular. Todo o período de colaboração no LIVE foi de intenso aprendizado e muito importante para a elaboração deste trabalho. Agradeço também ao meu orientador, Prof. Abel, pela paciência, inúmeros rabiscos nos papéis em nossas conversas sobre o direcionamento da pesquisa e confiança na realização da mesma. Também agradeço pela colaboração do Prof. George Darminton durante a execução do trabalho.

Por fim, a todos aqueles que participaram juntos nesta jornada de alguma forma, obrigado por me fazer aprender com vocês. Aos colegas do DES/UFPE e CIn/UFPE, com agradecimento especial para Paulo F. e Rafaela. Aos meus colegas do LIVE, em especial a Allan e Marcus, e aos colegas do *Magic Lab* do C.E.S.A.R.

RESUMO

Com o aumento das diversas aplicações de veículos conectados, como sistemas de navegação automotiva e sistemas de transporte inteligentes, o setor de telemetria veicular torna-se estratégico para as montadoras e consumidor. Novas soluções, baseadas na manutenção preditiva e no diagnóstico de veículos, são habilitadas de maneira mais fácil. Além da redução dos preços dos dispositivos eletrônicos, outro fator que também impulsionou o aumento no número de veículos conectados é a ampla cobertura de redes celulares e tecnologias habilitadoras da comunicação móvel. A partir de um mapeamento sistemático, este trabalho apresenta as diversas arquiteturas e aplicações da telemetria veicular. Graças à combinação das novas tecnologias de informação e comunicação (TICs), os veículos são equipados com unidade telemática que adiciona inteligência às tradicionais funções veiculares. A partir da proposta de um sistema de telemetria veicular, este trabalho avaliou a presença de anomalias na operação do sensor de Temperatura do Líquido Arrefecimento do Motor (ECT) por meio da coleta de dados de um único carro, em dois modos operacionais diferentes. A abordagem proposta avaliou dez classificadores de one-class (*one-class classifiers*) diferentes, em cinco níveis de anomalia, definidos a partir da faixa dinâmica de potência do sinal do sensor. Com base nos resultados dos dados experimentais, a avaliação baseada na métrica *F2-score* mostrou: a técnica OC-SVM com função de núcleo polinomial de terceiro grau como a melhor técnica de detecção de anomalias para a operação do veículo no modo em movimento e a técnica *k-NN* com o melhor resultado para o modo *idle*, onde o carro permanece com o motor ligado, porém parado.

Palavras-chaves: Telemetria veicular. Classificadores *one-class*. Detecção de anomalias. Sensor ECT.

ABSTRACT

With the increase in the number of the variety of connected vehicle applications, such as automotive navigation systems and intelligent transportation systems, the vehicle telemetry sector becomes strategic both for automakers and consumer. New solutions based on predictive maintenance and vehicle diagnostics are enabled in an easier way. In addition to the electronic devices cost reduction, another factor that has promoted the increase in the number of connected vehicles is the wide coverage of cellular networks and technologies that enables mobile communication. From a systematic mapping study, this work presents the various architectures and applications of vehicular telemetry. Thanks to the combination of new information and communication technologies (ICTs), vehicles are equipped with a telematics unit that adds intelligence to the basic vehicle functions. Based on a proposal of a vehicle telemetry system, this work has evaluated the presence of anomalies in the operation of the Engine Coolant Temperature (ECT) sensor by collecting data from a single car in two different operational modes. The proposed approach has evaluated ten different one-class classifiers of in five anomaly levels, defined based on the dynamic range of the sensor signal power. Based on the results of experimental data, OC-SVM technique with a third-degree polynomial core function performs the best result based on the F2-score metric. Considering the vehicle on idle mode, the k-NN technique presents the best F2-score results for the case when the car remains with the engine on, but stationary.

Keywords: Vehicular telemetry. One-class classifiers. Outlier detection. ECT sensor.

LISTA DE FIGURAS

Figura 1 – Resultado de pesquisa baseada na experiência do consumidor ao usar o carro conectado	17
Figura 2 – Arquitetura básica de um sistema embarcado	22
Figura 3 – Organização dos componentes de uma ECU	23
Figura 4 – Ilustração do barramento CAN	24
Figura 5 – Representação dos diferentes níveis lógicos do barramento CAN	25
Figura 6 – Formato de uma mensagem CAN	25
Figura 7 – Formato de um <i>frame</i> LIN	27
Figura 8 – Evolução do número médio de ECUs por tipo de carro	28
Figura 9 – Representação da divisão de diferentes barramentos intraveiculares instalados em um automóvel	29
Figura 10 – Representação do funcionamento da interface OBD-II	29
Figura 11 – Formato e descrição dos barramentos disponíveis no conector OBD-II comercial	30
Figura 12 – Processo de requisição de mensagens no padrão OBD-II	30
Figura 13 – Descrição dos tipos de <i>frame</i> OBD	31
Figura 14 – Evolução dos barramentos intraveiculares	31
Figura 15 – Tipos de anomalias (a) Pontual (b) Contextual (c) Coletiva.	33
Figura 16 – Definição de limiar para detecção de anomalia em uma distribuição Gaussiana	36
Figura 17 – Exemplo ilustrativo do funcionamento do LOF	38
Figura 18 – Representação da transformação linear entre o espaço de entrada e o de atributos para a técnica SVM	42
Figura 19 – Exemplo do cálculo da distância entre os hiperplanos	43
Figura 20 – Fronteiras de separação dos diferentes tipos de <i>kernel</i> (a) Linear (b) Linear com SVM Linear (c) RBF (d) Polinomial	44
Figura 21 – Taxonomia dos sistemas de telemetria veicular	46
Figura 22 – Número de trabalhos classificados no tipo da pesquisa	49
Figura 23 – Número de trabalhos classificados em cada arquitetura de <i>hardware</i>	49
Figura 24 – Número de trabalhos classificados em cada setor de interesse da telemetria veicular	50
Figura 25 – Número de trabalhos classificados em cada tema de pesquisa da nuvem veicular	50
Figura 26 – Sistemas de Aquisição de Dados: a) baseado em <i>smartphone</i> b) combinado com SBC, transceptor 3G / 4G e <i>dongle</i> OBD	51

Figura 27 – Divisão dos sensores utilizados por aplicação final do sistema de telemetria veicular	51
Figura 28 – Sensores utilizados nas diferentes aplicações da telemetria veicular . . .	52
Figura 29 – Visão geral do sistema proposto	56
Figura 30 – Placa de desenvolvimento <i>Particle Electron</i> com a plataforma Carloop integrada	57
Figura 31 – Arquitetura de <i>hardware</i> do sistema embarcado utilizado	58
Figura 32 – Fluxograma para os <i>firmwares</i> utilizados para o desenvolvimento do trabalho (a) detecção de padrão intraveicular (b) para captura dos dados	59
Figura 33 – Diagrama das possibilidades para montagem da base de dados a partir do sistema de aquisição	62
Figura 34 – Mapa do trajeto do veículo no campus da UFPE	62
Figura 35 – Representação de ciclo com informações intraveiculares capturadas pelo dispositivo	63
Figura 36 – Sinal ECT com anomalia Tipo I, desconexão do cabo do sensor com ECU	66
Figura 37 – Circuitos eletrônico do sensor de temperatura do líquido de arrefecimento do motor (ECT)	66
Figura 38 – Cinco diferentes níveis de anomalia aplicados ao sinal do sensor ECT .	67
Figura 39 – Faixa dinâmica de potência dos sinais utilizados nos modos (a) carro em movimento (b) modo <i>idle</i>	68
Figura 40 – Exemplo de sinal veicular com componente de ruído	69
Figura 41 – Etapas do sistema de detecção de anomalias, considerando as etapas de treinamento e teste	70
Figura 42 – Gráfico da Matriz de Correlação entre sete parâmetros coletados na base de dados	72
Figura 43 – Histograma dos parâmetros utilizados para treinamento no modo em movimento (a) variável do sensor ECT (b) número de rotações por minuto (RPM) e modo <i>idle</i> (c) variável do sensor ECT (d) número de rotações por minuto (RPM)	76
Figura 44 – Histograma dos parâmetros utilizados para treinamento no modo do veículo em movimento com trajetos específicos. Variável do sensor ECT (a) em trajeto fixo (c) trajeto em campus da universidade (e) trajetos urbanos (g) vias expressas. Variável do número de rotações por minuto (RPM) (b) em trajeto fixo (d) trajeto em campus da universidade (f) trajetos urbanos (h) vias expressas.	77
Figura 45 – Seleção do tamanho da janela para extração de características	79

Figura 46 – Tempo médio de processamento na etapa de teste das técnicas utilizadas nos experimentos (a) em modo <i>idle</i> (b) em modo do veículo em movimento	81
Figura 47 – Detalhamento das etapas de treinamento e teste para a avaliar o processo de detecção de anomalias	83
Figura 48 – <i>Linear Outlier Factor</i> para sinal (a) com funcionamento normal do ECT (b) sinal com desconexão do cabo	84
Figura 49 – Resultado da técnica <i>Linear Outlier Factor</i> em sinal do sensor ECT (a) sinal completo representado pela caracterização das amostras (b) destaque para a parte da desconexão do sensor	84
Figura 50 – Matriz de confusão para a técnica OC-SVM (a) Nível de anomalia 4 e (b) Nível de anomalia 5.	88
Figura 51 – Resultado do teste de Friedman para diferenças estatísticas entre as técnicas utilizadas no modo em movimento (a) Nível de Anomalia 4 (b) Nível 5. O índice indica o numero da técnica analisada, seguindo a ordem mostrada nas tabelas.	89
Figura 52 – <i>Boxplot</i> do F2-score para (a) Nível 4 de anomalia do veículo em movimento (b) Nível 5 de anomalia, respectivamente.	89
Figura 53 – Matriz de confusão para a técnica k-NN (a) Nível de anomalia 4 e (b) Nível de anomalia 5.	92
Figura 54 – Comparação do sinal ECT para modo <i>idle</i> e modo do veículo em movimento.	93
Figura 55 – <i>Boxplot</i> do F2-score para (a) Nível 4 de anomalia do veículo em modo <i>idle</i> (b) Nível 5 de anomalia, respectivamente.	94
Figura 56 – <i>Boxplot</i> de comparação entre os modos <i>Idle</i> e em movimento para o Nível de anomalia 4.	95
Figura 57 – <i>Boxplot</i> de comparação entre os modos <i>Idle</i> e em movimento para Nível de anomalia 5.	96
Figura 58 – Resultado do teste de Friedman para diferenças estatísticas entre as técnicas utilizadas no modo <i>idle</i> (a) Nível de Anomalia 4 (b) Nível 5. O índice indica o numero da técnica analisada, seguindo a ordem mostrada nas tabelas.	96

LISTA DE TABELAS

Tabela 2 – Comparação das contribuições entre trabalhos que abordam telemetria veicular	52
Tabela 3 – Comparação das contribuições entre trabalhos que abordam detecção de anomalias veiculares	54
Tabela 4 – Descrição dos parâmetros coletados a partir da OBD-II	61
Tabela 5 – Descrição da de base de dados construída a partir dos parâmetros da OBD-II	62
Tabela 6 – Exemplo de uma Matriz de Confusão	74
Tabela 7 – Descrição da base de dados usado nos experimentos do carro em modo <i>idle</i> e em movimento para anomalia Tipo II.	76
Tabela 8 – Descrição da base de dados usado nos experimentos considerando os quatro cenários do carro em movimento	76
Tabela 9 – Tamanho da janela de seleção de atributos para técnicas utilizadas no trabalho	80
Tabela 10 – Tamanho da janela de seleção de atributos para técnicas utilizadas no trabalho - Trajetos específicos	80
Tabela 11 – Tempo médio de processamento, em segundos, na etapa de teste para o modo <i>idle</i>	81
Tabela 12 – Tempo médio de processamento, em segundos, na etapa de teste para o modo do veículo em movimento	82
Tabela 13 – Especificações do Computador utilizado para a realização dos experimentos	82
Tabela 14 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo I	85
Tabela 15 – Parâmetros específicos utilizadas em cada técnica	85
Tabela 16 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 1	86
Tabela 17 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 2	86
Tabela 18 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 3	87
Tabela 19 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 4	87
Tabela 20 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 5	88
Tabela 21 – Resultado do teste <i>post-hoc</i> de Nemenyi para o modo em movimento	90

Tabela 22 – Comparação dos resultados do F2- <i>score</i> para o modo do veículo em movimento	90
Tabela 23 – Média e desvio padrão das métricas para os experimentos do veículo em modo <i>idle</i> para anomalia Tipo I	91
Tabela 24 – Média e desvio padrão das métricas para os experimentos do veículo em modo <i>Idle</i> para anomalia Tipo II - Nível 1	92
Tabela 25 – Média e desvio padrão das métricas para os experimentos do veículo em modo <i>Idle</i> para anomalia Tipo II - Nível 2	92
Tabela 26 – Média e desvio padrão das métricas para os experimentos do veículo em modo <i>Idle</i> para anomalia Tipo II - Nível 3	93
Tabela 27 – Média e desvio padrão das métricas para os experimentos do veículo em modo <i>Idle</i> para anomalia Tipo II - Nível 4	94
Tabela 28 – Média e desvio padrão das métricas para os experimentos do veículo em modo <i>Idle</i> para anomalia Tipo II - Nível 5	95
Tabela 29 – Resultado do teste <i>post-hoc</i> de Nemenyi para o modo <i>idle</i>	97
Tabela 30 – Comparação dos resultados do F2- <i>score</i> para o modo <i>idle</i>	97
Tabela 31 – Resultado da melhor técnica para cada trajeto e nível de anomalia, com seu respectivo valor de F2- <i>score</i>	98

LISTA DE ABREVIATURAS E SIGLAS

CRC	<i>Cyclic Redundancy Check</i>
CSMA/CA	<i>Carrier Sense Multiple Access Collision Avoidance</i>
DTC	<i>Diagnostic Trouble Code</i>
ECT	<i>Engine Coolant Temperature</i>
ECU	<i>Electronic Control Unit</i>
IoT	Internet das Coisas
ITS	<i>Intelligent Transportation System</i>
LDF	<i>Lin Description File</i>
LIN	<i>Local Interconnect Network</i>
LODA	<i>Lightweight On-Line Detector of Anomalies</i>
LPWAN	<i>Low Power Wide Area Network</i>
MQTT	<i>MQ Telemetry Transport</i>
OBD-II	<i>On-Board Diagnostic II</i>
OC-SVM	<i>One Class Support Vector Machine</i>
OSI	<i>Open System Interconnection</i>
SBC	<i>Single Board Computer</i>
SOM	<i>Self-organizing map</i>
SVDD	<i>Support Vector Data Description</i>
SVM	<i>Support Vector Machine</i>
TCP/IP	<i>Transmission Control Protocol/Internet Protocol</i>
TICs	Tecnologias da Informação e Comunicação
V2I	<i>Vehicle to Infrastructure</i>
V2V	<i>Vehicle to Vehicle</i>
V2x	<i>Vehicle to Anything</i>

LISTA DE SÍMBOLOS

γ	Letra grega Gama
\in	Pertence
δ	Delta
θ	Teta
σ	Sigma
μ	Mi

SUMÁRIO

1	INTRODUÇÃO	16
1.1	PROBLEMA DE PESQUISA	18
1.2	OBJETIVOS	19
1.2.1	Objetivo Geral	19
1.2.2	Objetivos Específicos	19
1.3	ORGANIZAÇÃO DO TRABALHO	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	O SISTEMA DE AQUISIÇÃO DE DADOS	21
2.1.1	Sistemas de telemetria	21
2.1.2	Sistemas embarcados	22
2.2	REDES INTRAVEICULARES	22
2.2.1	A rede CAN	23
2.2.2	A rede LIN	26
2.2.3	Outras redes automotivas	27
2.2.4	O padrão de diagnóstico OBD-II	28
2.3	DETECÇÃO DE ANOMALIAS	32
2.3.1	Tipos de anomalias	32
2.3.2	Técnicas de detecção de anomalias	34
2.3.2.1	Análise Estatística	34
2.3.2.2	Sistemas baseados em regras	37
2.4	CLASSIFICADORES <i>ONE-CLASS</i>	39
2.4.1	<i>Self-organized map (SOM)</i>	39
2.4.2	<i>Support Vector Data Description (SVDD)</i>	39
2.4.3	k-NN	40
2.4.4	OC-SVM	41
3	REVISÃO BIBLIOGRÁFICA	45
3.1	SISTEMAS DE TELEMETRIA VEICULAR	45
3.2	DETECÇÃO DE ANOMALIAS EM REDES INTRAVEICULARES	52
4	SISTEMA PROPOSTO	56
4.1	VISÃO GERAL	56
4.2	AQUISIÇÃO DE DADOS DA REDE INTRAVEICULAR	57
4.2.1	<i>Firmware Embarcado</i>	58

4.3	BASE DE DADOS COM SINAIS INTRAVEICULARES	59
4.4	PROCESSO DE INSERÇÃO DE ANOMALIAS EM SINAIS VEICULARES	63
4.4.1	Anomalias no sensor de temperatura do líquido de arrefecimento do motor (ECT)	65
4.5	SISTEMA DE DETECÇÃO DE ANOMALIAS VEICULARES	69
4.5.1	Sistema de aquisição de dados	71
4.5.2	Pré-Processamento: extração e seleção de características	71
4.5.3	Construção do modelo	73
4.5.4	Modelo	73
5	RESULTADOS E DISCUSSÃO	74
5.1	MÉTRICAS DE DESEMPENHO	74
5.2	BASE DE DADOS UTILIZADAS	75
5.3	DEFINIÇÃO DA TÉCNICA DE DETECÇÃO DE ANOMALIA	78
5.4	EXTRAÇÃO DE CARACTERÍSTICAS: DEFINIÇÃO DO TAMANHO DA JANELA	78
5.5	EXPERIMENTOS COM O VEÍCULO EM CICLOS DE DIREÇÃO	82
5.5.1	Anomalia Tipo I	82
5.5.2	Anomalia Tipo II	85
5.6	EXPERIMENTOS COM O VEÍCULO EM MODO <i>IDLE</i>	90
5.6.1	Anomalia Tipo I	90
5.6.2	Anomalia Tipo II	91
5.7	EXPERIMENTOS COM O VEÍCULO EM CICLOS DE DIREÇÃO EM TRAJETOS ESPECÍFICOS	98
6	CONCLUSÃO	99
6.1	CONTRIBUIÇÕES	100
6.2	DIFICULDADES	101
6.3	TRABALHOS FUTUROS	101
	REFERÊNCIAS	102
	APÊNDICE A – FIRMWARE EMBARCADO E BASE DE DADOS	108

1 INTRODUÇÃO

O atual cenário do desenvolvimento tecnológico se caracteriza pela junção de Tecnologias da Informação e Comunicação (TICs), aplicadas ao desenvolvimento de novas soluções inovadoras e sustentáveis. Considerando o cenário industrial, pode-se definir o conceito da Indústria 4.0, também conhecida como a Quarta Revolução Industrial. A Indústria 4.0 se caracteriza, por exemplo, pelo uso de manufatura aditiva, prototipagem rápida e inteligência artificial nos diferentes setores industriais. Em rápida comparação com a terceira revolução industrial, caracterizada pela inserção da microeletrônica, telecomunicações e processo de automatização na indústria, a quarta revolução amplia as possibilidades e aplicações baseadas nas novas tecnologias digitais.

Entre essas tecnologias a Internet das Coisas (IoT) se destaca, considerando a interconexão de diferentes objetos e o surgimento de novas funcionalidades a partir das diferentes coisas conectadas. Assim, em um futuro próximo espera-se que quase todos objetos do cotidiano estejam conectados. A estimativa é de 20 bilhões de objetos conectados à internet até 2020 (GARTNER, 2017). Seguindo esta tendência, os veículos também tendem a estabelecer algum tipo de conexão em um futuro próximo (DEVI; RUKMINI, 2016).

Como um dos principais motivadores para o estabelecimento de algum tipo de conexão veicular, o iminente surgimento do carro autônomo também demandará uma alta capacidade de processamento computacional (RODRIGUES, 2017). De uma maneira geral, todas as fabricantes automotivas começarão a estabelecer e fornecer algum tipo de serviço automotivo que fornece uma análise inteligente a partir de dados capturados. Um dos exemplos mais populares no mercado brasileiro é o fornecido pela Chevrolet, o sistema *OnStar* (CHEVROLET, 2018).

Considerando as diferentes propostas do estabelecimento de algum tipo de comunicação veicular, estas são habilitadas diretamente pelos fabricantes ou por diferentes fornecedores a partir de dispositivos acessórios, produtos do setor automotivo *aftermarket*. De maneira geral, o veículo conectado objetiva a troca de dados entre veículos, comunicação *Vehicle to Vehicle* (V2V), ou entre um veículo e alguma infraestrutura, *Vehicle to Infrastructure* (V2I), proporcionando novos serviços e experiência ao condutor (DEVI; RUKMINI, 2016). Cinco diferentes setores podem se beneficiar do veículo conectado: sistemas de gestão de frotas, sistemas de rastreamento veicular, seguradoras, montadoras e o novo segmento, Sistemas de Transportes Inteligentes (*Intelligent Transportation System* (ITS)) (NAVET; SIMONOT-LION, 2008a).

No setor dos fabricantes, sejam estes as próprias montadoras ou fornecedores de peças veiculares, o monitoramento e posterior análise dos dados veiculares tornam-se etapas fundamentais para sistemas de diagnóstico avançado e agendamento de manutenção preditiva. Em rápida comparação com os atuais modelos comerciais, todo veículo possui um

sistema de alerta de falhas, alguns modelos com notificações no painel a partir de um *Diagnostic Trouble Code* (DTC). O funcionamento de um DTC é baseado no armazenamento de um código relacionado ao problema em sua memória, assim, quando uma unidade de controle veicular detecta que um componente ou sistema não está operando dentro de seu limite aceitável, o erro é reportado na rede veicular.

Desta forma, a principal diferença no uso dos dados veiculares coletados a partir de um sistema de telemetria veicular é a possibilidade da implementação de adaptativos, isto é, sistemas que aprendam o funcionamento normal e possam identificar a presença de um comportamento anômalo de uma componente veicular. Apesar de fornecer benefícios para o consumidor final, o processo de coleta de dados do veículo pode se tornar um problema, devido ao grande volume de dados provenientes dos diversos sensores.

Com mercado estimado em 45 bilhões de dólares em 2019 (WAHLSTRÖM; SKOG; HÄNDEL, 2017), diversos serviços baseados em sistemas de telemetria veicular são propostos. A partir de uma pesquisa realizada com 396 pessoas, baseada na experiência do consumidor, ao usar o carro como um produto ou serviço, pode-se listar as principais funcionalidades desejadas de um carro conectado. A Figura 1 ilustra as principais perguntas da pesquisa, e o resultado.

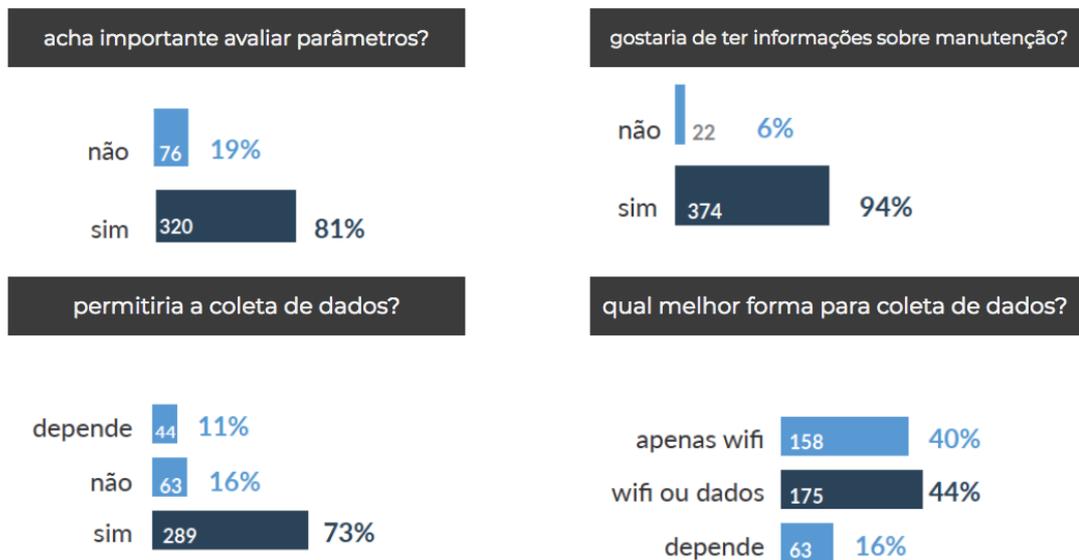


Figura 1 – Resultado de pesquisa baseada na experiência do consumidor ao usar o carro conectado

Fonte: (CIN-UFPE/FCA, 2018)

Uma das principais vantagens do carro conectado, para o consumidor, é a possibilidade da oferta de novos serviços relacionados à manutenção preditiva. Em geral, estes serviços começam a partir de uma análise específica de peças ou partes do veículo (THEISSLER; DEAR, 2013a). Assim, torna-se importante determinar os sistemas mais críticos para o funcionamento do veículo, para que em seguida, seja realizado o processo de detecção de

mau funcionamento, ou comportamento anômalo.

Entre os problemas mais comuns enfrentados pelos motoristas está o superaquecimento do motor (STATUSSEGUROS, 2018). Além da temperatura externa, diversos fatores influenciam para o superaquecimento, e a maioria destes podem ser evitados a partir do acompanhamento periódico utilizando ferramentas de diagnóstico, incluindo os DTCs DTC. Em específico ao funcionamento do motor, existe um sensor para controlar a temperatura do líquido de arrefecimento do motor. O sensor de temperatura do líquido de arrefecimento, *Engine Coolant Temperature* (ECT), retorna com precisão o valor exato da temperatura do motor.

É recomendada a manutenção preventiva do sensor de temperatura e de todo sistema de arrefecimento do veículo a cada 30.000 km rodados (MTE-THOMSON, 2018). Além de muitos condutores não seguirem a recomendação, o sensor ECT pode apresentar falhas em seu funcionamento, ocasionando assim, riscos diretos ao motor. Como elemento principal de qualquer veículo, os sistemas auxiliares ao motor necessitam de um sistema diagnóstico preciso.

Apesar de possuir DTCs específicos para o sensor ECT, como verificar o valor da temperatura de partida fria e valores fora da escala normal do seu funcionamento (OBD-CODES, 2018), os veículos ainda não possuem um sistema que identifique um funcionamento do sensor ECT fora do esperado, isto é, um sensor com variações bruscas e medições inconstantes.

Desta forma, a proposta deste trabalho consiste em apresentar uma arquitetura de sistema de telemetria veicular e a partir desta desenvolver um sistema de detecção de anomalias do sensor ECT. Para tal objetivo, torna-se necessário o uso de uma base de dados veiculares com o funcionamento de um veículo em operação normal. A partir dos dados, uma análise das diferentes técnicas de detecção de anomalias é realizada. Por fim, são apresentadas as melhores técnicas para os respectivos cenários de uso do veículo e tipos de anomalias.

1.1 PROBLEMA DE PESQUISA

A detecção de anomalias no funcionamento de diferentes sensores veiculares tem sido estudada a partir da avaliação de técnicas de análise de limiar e aprendizado máquina. Algumas propostas avaliam diferentes técnicas e modos de operação veiculares, como apresentado por (THEISSLER, 2017). Este trabalho objetiva responder a seguinte pergunta:

É possível detectar anomalias em sensores veiculares considerando diferentes níveis de mau funcionamento?

1.2 OBJETIVOS

Propor e avaliar um sistema de detecção de anomalias veiculares para o sensor ECT a partir da presença de diferentes níveis de anomalias. A proposta é baseada na avaliação de diferentes técnicas de detecção de anomalias, a serem avaliadas usando base de dados com informações da rede intraveicular captadas a partir de sistema de telemetria.

1.2.1 Objetivo Geral

Avaliar o desempenho do sistema de detecção de anomalias veiculares do sensor ECT proposto, a partir de diferentes níveis de anomalias presentes no sinal no sensor veicular.

1.2.2 Objetivos Específicos

Este trabalho possui os seguintes objetivos específicos:

1. Identificar as principais arquiteturas de sistemas de telemetria veicular e propor um sistema para a coleta de dados de maneira simples.
2. Apresentar uma ou mais técnicas para a detecção de anomalias no sensor ECT, baseada nos dados adquiridos pelo sistema proposto.

1.3 ORGANIZAÇÃO DO TRABALHO

O presente trabalho está organizado da seguinte maneira:

- **Capítulo 2 - Fundamentação Teórica:** A fundamentação teórica deste trabalho é apresentada no Capítulo 2. Conceitos relacionados aos tipos de redes intraveiculares e padrão de diagnóstico automotivos são apresentados. Além destes conceitos, fundamentos do processo de detecção de anomalias também são apresentados.
- **Capítulo 3 - Revisão Bibliográfica:** No terceiro capítulo é apresentada uma revisão da literatura acerca dos trabalhos relacionados aos temas de telemetria veicular e detecção de anomalias em redes veiculares. As principais arquiteturas propostas de telemetria veicular, bem como suas aplicações são apresentadas. Relacionado a detecção de anomalias em redes veiculares, os diferentes descrevem as técnicas e parâmetro analisado.
- **Capítulo 4 - Proposta:** Neste capítulo é apresentada a proposta deste trabalho, um sistema de telemetria veicular para detecção de anomalias de um sensor da rede veicular. A visão geral do sistema é apresentada a partir do módulo de captura de dados e do sistema responsável pelo processamento e detecção das anomalias.

- **Capítulo 5 - Resultados e Discussão:** No Capítulo 5 são apresentados os resultados obtidos durante o desenvolvimento deste trabalho, seguidos por uma discussão.
- **Capítulo 6 - Conclusão:** Neste capítulo são apresentadas as dificuldades, conclusões gerais, contribuições e possíveis trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a base teórica dos temas abordados ao longo desta dissertação. Na Seção 2.1 é apresentado o conceito de sistema embarcado, elemento fundamental em um sistema de telemetria veicular. A Seção 2.2 descreve a operação básica de uma rede veicular, detalhada a partir das redes CAN, LIN e o padrão de diagnóstico OBD-II, utilizado para captura dos dados utilizados no desenvolvimento deste trabalho.

2.1 O SISTEMA DE AQUISIÇÃO DE DADOS

O objetivo de um sistema de telemetria veicular é a captura de parâmetros relevantes ao funcionamento dos diversos subsistemas veiculares. Assim, a subseção 2.1.1 discorre brevemente a história dos sistemas de telemetria. Em seguida, na subseção 2.1.2, é definido o conceito de sistema embarcado, que caracteriza o funcionamento de um sistema de telemetria veicular.

2.1.1 Sistemas de telemetria

O processo de telemetria é estabelecido como a comunicação automatizada, com o objetivo da medição de informações de diferentes sistemas. O sistema que controla o processo de envio de dados também torna-se o principal responsável pelo processo de aquisição dos dados. Em geral, os dados são enviados para estações de recepção e posteriormente analisados e processados. Apesar dos sistemas de comunicação sem fio serem associados aos dispositivos de telemetria, um meio de conexão física também pode ser utilizado para tal fim (BARNES; TONEY; JAROMCZYK, 2016).

Como descrito por (QUEIRÓS, 2011) a proposta do primeiro sistema de telemetria surgiu em 1845 com fins militares. Como seguimento da proposta inicial, a operação de telemetria foi utilizada para medir informações climáticas. Ao longo da Segunda Guerra Mundial, foram desenvolvidas diversas técnicas para transmissão de dados através de técnicas de modulação.

Com o surgimento de diversas tecnologias para conexão dos dispositivos, o processo de telemetria pode ser definido como a operação de sensoriamento automatizado e o controle de dispositivos de maneira remota (COMATRELECO, 2018). No contexto da IoT, a necessidade de padronização entre os diversos dispositivos de telemetria é tão grande, que levou o surgimento de um protocolo específico para tal.

O protocolo *MQ Telemetry Transport* (MQTT) foi projetado especificamente para a troca de mensagens de forma segura entre dispositivos móveis baseados em redes *Transmission Control Protocol/Internet Protocol* (TCP/IP) (YASSEIN et al., 2017). No contexto das diversas coisas conectadas, a telemetria é o sensoriamento, medição e controle dos da-

dos de forma automatizada. A ênfase está na transmissão de dados a partir de diferentes dispositivos para uma unidade de controle central.

2.1.2 Sistemas embarcados

Além de requisitos como tamanho reduzido e operação de maneira auto suficiente, os sistemas de telemetria possui funções específicas e pré-definidas. Como consequência, estas características definem um sistema de telemetria como um sistema embarcado. Na sua maior parte, os sistemas embarcados são caracterizados por um microcontrolador ou microprocessador, a unidade de controle responsável pelas operações (NAVET; SIMONOT-LION, 2008a).

Outro dispositivo fundamental na estrutura de um sistema embarcado, além da unidade de controle, é o responsável pela conversão das grandezas físicas em sua representação digital, o sensor. De maneira contrária ao processo de medição, os atuadores são elementos que produzem movimento (ação mecânica), ao serem estimulados. A partir de comandos pré-definidos, sendo estes manuais ou elétricos, os atuadores permitem o controle de outros dispositivos. Ao integrar os elementos listados nesta seção, em conjunto com uma fonte de alimentação, defini-se a arquitetura básica de um sistema embarcado, ilustrada na Figura 2.

Com a evolução da microeletrônica e a consequente redução nos preços dos dispositivos semicondutores, o preço médio dos microcontroladores de propósito geral favoreceu o surgimento de vários produtos e objetos baseados no funcionamento de sistemas embarcados.

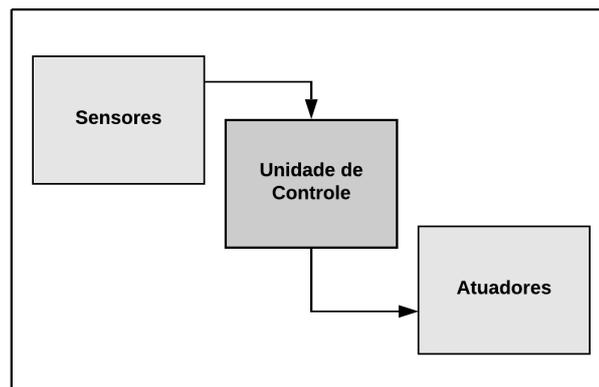


Figura 2 – Arquitetura básica de um sistema embarcado

2.2 REDES INTRAVEICULARES

Além do avanço da eletrônica embarcada, o desenvolvimento de novas funcionalidades e melhorias no desempenho, conforto e segurança dos automóveis motivaram o aumento do

numero de sistemas embarcados automotivos (NAVET; SIMONOT-LION, 2009). Com uma grande quantidade de microcontroladores no carro, tornou-se necessário definir unidades de controle responsáveis pelo gerenciamento das principais funções veiculares.

Assim, cada unidade de controle eletrônico, *Electronic Control Unit* (ECU), é definida como o dispositivo de controle de um grupo de funções específicas do veículo. A organização dos componentes de uma ECU, ilustrada na Figura 3, é idêntica a de um sistema embarcado, possuindo a captura de dados a partir de sensores, unidade de processamento e acionamento de atuadores. Apesar das principais funções do veículo estarem distribuídas em diferentes unidades de controle, esta abordagem começou a se tornar ineficiente, devido ao aumento das interconexões das diferentes ECUs, por meio de conexão ponto-a-ponto.

A evolução dos diferentes tipos de automóveis também levou a uma definição do sistema central responsável pelo funcionamento do veículo. Os componentes responsáveis pela geração e distribuição da energia ao longo do veículo são classificados como parte do chamado sistema *powertrain*. Em uma visão de alto nível, os componentes deste sistema estão sujeitos a melhoria do desempenho do veículo, eficiência e controle de emissões de poluentes. Dessa forma todas informações relacionadas ao motor, transmissão e estrutura do veículo são relacionadas ao sistema *powertrain* (CROLLA; MASHADI, 2011).

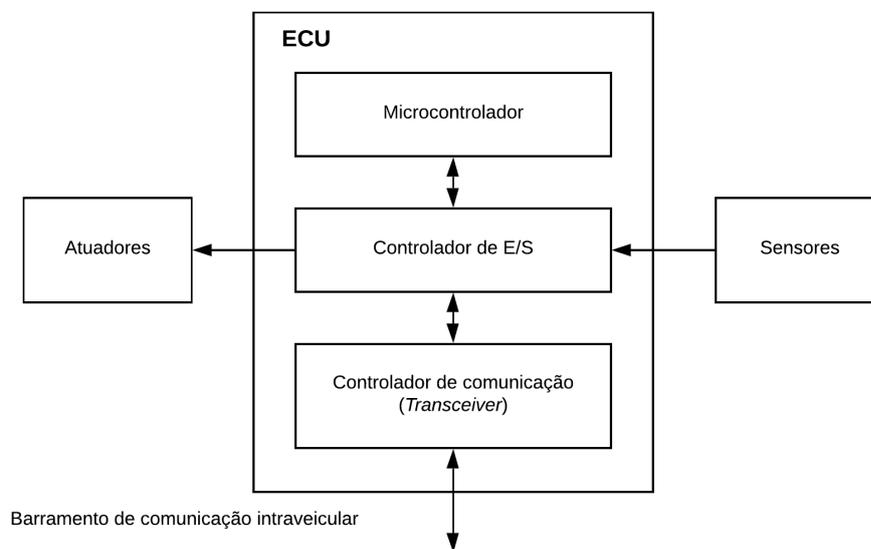


Figura 3 – Organização dos componentes de uma ECU

2.2.1 A rede CAN

Como proposta de solução para o problema, foi desenvolvido na metade da década de 1980 um protocolo de comunicação intra-veicular. A arquitetura de funcionamento do protocolo é baseada em barramento específico para a comunicação das diferentes ECUs, a rede CAN (*Controller Area Network*) (NAVET; SIMONOT-LION, 2008b). Além de objetivar

a troca de informação de maneira rápida e robusta, o uso do barramento CAN reduz o comprimento dos fios dedicados a interconexão das ECUs, visto que não necessita da comunicação ponto a ponto. Outra vantagem no uso do barramento é o compartilhamento de sensores e atuadores entre as diferentes unidades de controle.

O barramento CAN é um sistema de transmissão mensagens *broadcast*, não precisando de um controlador central e é orientado a mensagens. Tornando-se um protocolo baseado em mensagens com uma taxa máxima de transmissão de 1 megabit por segundo (1 Mbps). Para o estabelecimento do protocolo, torna-se necessária a interconexão de no mínimo dois nós conectados através de um barramento de dois fios. Uma característica da conexão pela rede CAN é a presença de um par trançado de resistores de $120\ \Omega$ entre os sinais CAN *high* e CAN *low*, mostrado na Figura 4.

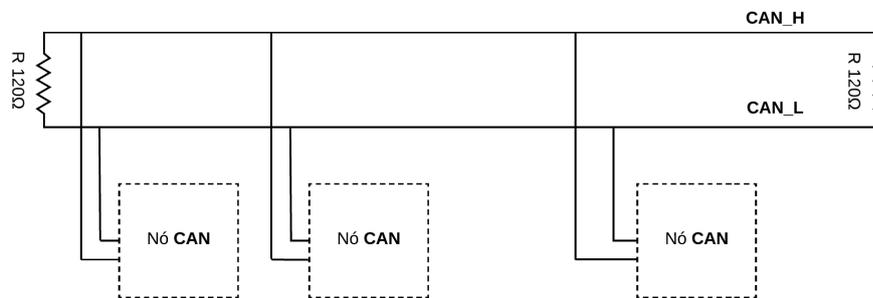


Figura 4 – Ilustração do barramento CAN

Os dois sinais analógicos CAN *high* e CAN *low* definem o nível dominante, representado por **0** (nível lógico baixo), ou recessivo, representado por **1** (nível lógico alto), a partir da diferença de tensão dos respectivos sinais. A depender da velocidade utilizada no barramento, definindo assim o uso de *high-speed* CAN ou *low-speed* CAN, a tensão do nível dominante é definida quando a diferença for maior que 2.3V (*low-speed*) ou 2V (*high-speed*) e o recessivo quando a diferença entre as tensões for menor que 0.6V (*high-speed*) ou diferença de 5V (*low-speed*). A Figura 5 ilustra a definição dos níveis lógicos para rede *high-speed* CAN.

O processo de transmissão de dados na CAN é realizado segmentando um dado em vários pacotes, podendo estes serem transmitidos de forma periódica ou por requisições, no caso de uma comunicação entre ECUs da forma cliente/servidor. Outra característica que deve ser ressaltada do barramento é que a transmissão dos dados entre os nós na rede CAN é realizada sem um sinal de relógio (*clock signal*).

A sincronização do barramento é realizada a partir da arbitragem de todos os nós da rede CAN pelo *bit-timing*. Dessa forma, cada bit transmitido no barramento CAN é, para fins de temporização, dividido em quatro intervalos de tempo, chamados de *time quanta*. Para ajustar o relógio do barramento no *transceiver*, o controlador CAN pode encurtar ou prolongar o comprimento de um bit por um número inteiro de *quanta* (KVASER, 2018).

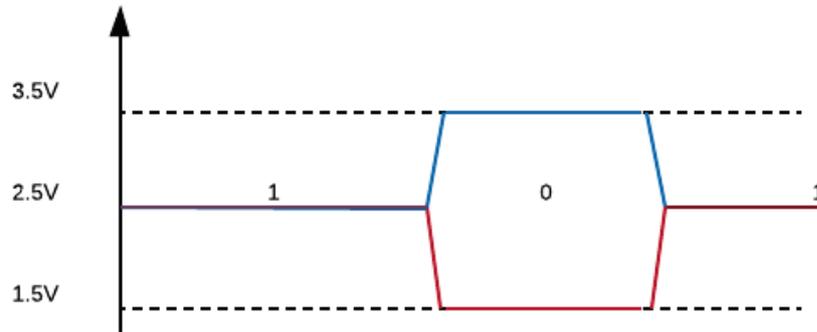


Figura 5 – Representação dos diferentes níveis lógicos do barramento CAN

A prioridade do envio de mensagens no barramento é realizada pelo identificador da mensagem, que pode ter o tamanho de 11 bits ou 29 bits (caracterizando o padrão CAN 2.0B). Na parte dos bits mais significativos, o nó com o maior prioridade deve ter uma maior sequência de bits dominantes durante a parte de arbitragem do *frame*. O formato padrão de uma mensagem CAN pode ser vista na Figura 6.

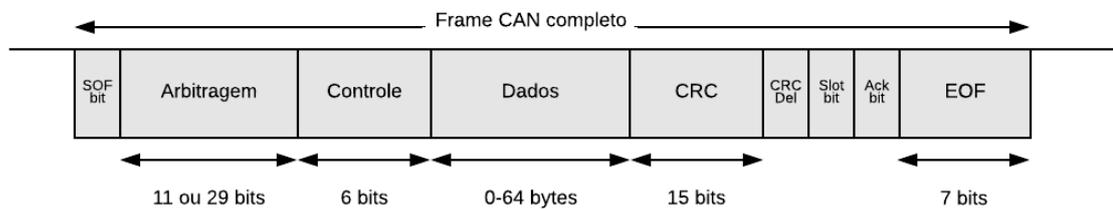


Figura 6 – Formato de uma mensagem CAN

A rede CAN possui quatro tipos de mensagens (*frames*):

- **Data frame:** mensagem com dados para transmissão entre nós.
- **Remote frame:** mensagem que solicita a transmissão de dados de um nó específico.
- **Error frame:** qualquer mensagem transmitida por um nó pode conter um erro, assim, este tipo de *frame* indica o erro após a transmissão de um *data frame*, sendo detectado por uma ECU. Este tipo de *frame* é usado para notificar outros nós da ocorrência do erro.
- **Overload frame:** tipo de mensagem para inserir atraso entre mensagens de dados e requisições.

Assim como outras redes de comunicação, o protocolo CAN pode ser decomposto nas seguintes camadas de abstração do padrão *Open System Interconnection* (OSI):

- **Camada Física:** uma camada física define os níveis de tensão e formato de conexão do barramento. Considerando sua evolução ao longo do tempo, existem diferentes padrões para o barramento CAN. O tipo mais comum, usado nas aplicações automotivas, é aquele definido pelo padrão CAN ISO 11898-2, representado pela ligação a partir de um par de fios, com taxas de transmissão de até 1Mbps (KVASER, 2017). Torna-se importante ressaltar que existe um limite ao valor da taxa de transmissão a depender do comprimento do barramento. Parte dos outros padrões CAN são adaptados como alternativas para comunicação serial de diferente dispositivos.
- **Camada de Enlace de Dados:** as características da camada de ligação, ao considerar o barramento CAN, envolvem a detecção de erro das mensagens, processo de validação, arbitragem e taxa de transmissão. Este último atributo está diretamente relacionado com o comprimento do barramento, em metros. Considerando os valores aproximados, um barramento CAN de 100 metros consegue velocidades de 500 kbit/s. Outros atributos podem ser detalhados a partir da composição da mensagem CAN. A primeira parte da mensagem torna-se responsável pela arbitragem. Outro exemplo é a parte responsável pelo controle de erro, a operação do *Cyclic Redundancy Check* (CRC), a partir de 15 bits em cada mensagem.

2.2.2 A rede LIN

Apesar de ser a rede de intraveicular mais utilizada para a interconexão das ECUs, o barramento CAN tornou-se caro demais para ser utilizado em todas as interconexões dos componentes do carro. Dessa forma, diversos fabricantes começaram a desenvolver diferentes protocolos de comunicação serial para conectar componentes de sistemas auxiliares do carro, como os controladores da abertura de janelas e portas. No início dos anos 2000 a rede serial *Local Interconnect Network* (LIN) começou a ser utilizada (TUOHY et al., 2015).

O barramento LIN pode ser classificado como um sub-barramento da rede CAN, com suas principais características resumidas em:

- Camada física LIN é baseada na ISO 9141 (*K-line*).
- Barramento de baixa velocidade com velocidades de até 20 kbps.
- Sem necessidade de arbitragem e comunicação serial orientada pelo envio de *bytes*.

A comunicação serial LIN permite transmissão de mensagens entre até 16 nós, considerando a topologia em que 15 destes são escravos e a presença de um mestre. A partir de sua topologia, todo processo de troca de mensagens é iniciado pelo mestre com cada escravo respondendo com seu respectivo identificador da mensagem. Diferente da CAN, que usa uma versão modificada da técnica *Carrier Sense Multiple Access Collision Avoidance* (CSMA/CA) para evitar colisão de mensagens, o uso da rede LIN não necessita a implementação da detecção de colisão de mensagens.

Uma das características da camada física LIN é o uso de apenas um sinal para transmissão dos dados, com limitação do comprimento do barramento em até 40 metros. O formato das mensagens é baseado em arquivo descrito no software dos nós da rede LIN, o arquivo *Lin Description File* (LDF). Toda troca de informação é baseada em mensagens que contém um cabeçalho, uma resposta e espaço de resposta para que o escravo tenha tempo de responder. Cada quadro é enviado em um campo determinado pelo arquivo LDF. O formato dos *frames* é ilustrado na Figura 7.

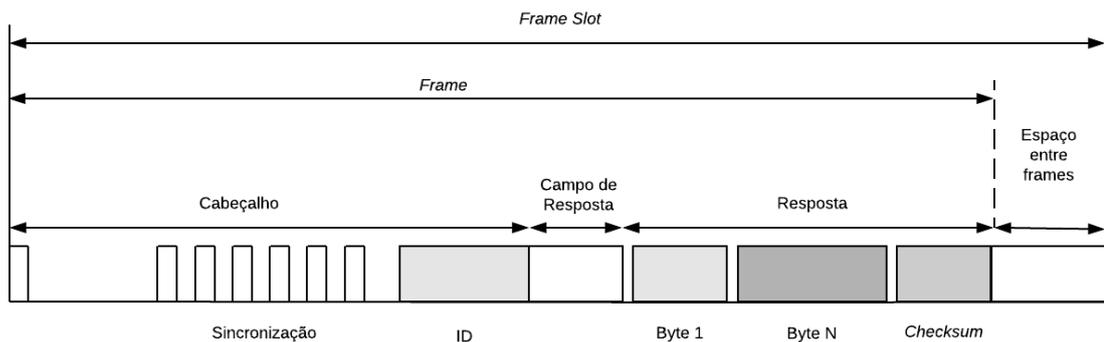


Figura 7 – Formato de um *frame* LIN

Ao considerar aspectos de *hardware* para a implementação da rede LIN, cada ECU necessita apenas de um *transceiver* característico, que possui um menor custo quando comparado a um *transceiver* CAN.

2.2.3 Outras redes automotivas

Apesar de padrões já estabelecidos, como as redes CAN e LIN, o número de ECUs nos veículos modernos continua crescendo (NXP, 2018), como mostrado na Figura 8. O foco atual das pesquisas em redes intraveiculares se concentra no desenvolvimento e estabelecimento de padrões para redes intra-veiculares destinadas as aplicações de entretenimento e informação veicular (*infotainment*) e redes capazes de uma alta taxa de transmissão de dados, com uma latência mínima (TUOHY et al., 2015).

Outras novas funcionalidades automotivas, que definem o uso dos novos padrões de redes intraveiculares, são implementadas pelos sistemas avançados de assistência ao motorista (ADAS). Em conjunto com o uso da *Ethernet* automotiva, outras redes como *FlexRay*, *Media Oriented Systems Transport* (MOST) foram desenvolvidos para satisfazer diferentes necessidades de confiabilidade, tempo de resposta, flexibilidade, e largura de banda no carro (SOARES et al., 2015). O aumento na taxa de transmissão é motivado pela tendência de possíveis carros autônomos em um futuro próximo.

Considerando as diferentes redes intraveiculares presente em um carro, a Figura 9 ilustra a divisão dos barramentos, seu respectivo uso e integração.

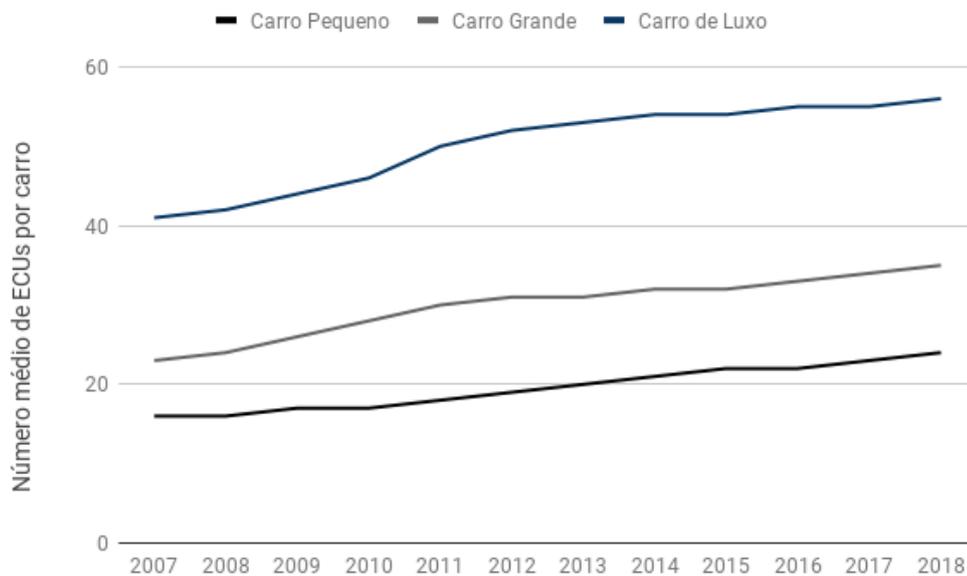


Figura 8 – Evolução do número médio de ECUs por tipo de carro

Fonte: (NXP, 2018)

2.2.4 O padrão de diagnóstico OBD-II

Ainda na década de 80, após o estabelecimento de padrões de redes intraveiculares, também foi definido um padrão para diagnóstico automotivo. A motivação inicial do padrão era tornar disponíveis informações relacionadas ao consumo de combustível. A partir da interface de comunicação com dispositivos externos, conclui-se que a mesma poderia ser estendida para também fornecer informações de diagnóstico do veículo. Considerando a arquitetura veicular descrita anteriormente, o padrão OBD funciona como uma ponte que liga as leituras de sensores a dispositivos externos. A Figura 10 ilustra sua conexão.

Outra funcionalidade, adicionada a partir do padrão OBD, é a presença da lâmpada indicadora de mau funcionamento (MIL), permitindo ao condutor um alerta para gerenciamento de algum defeito em seu veículo. A ativação da MIL pode ser realizada a partir da presença de um DTC fornecido pela interface OBD. Outras funcionalidades ao usuário decorrentes da OBD também podem ser encontradas no painel do veículo, como a temperatura do líquido de arrefecimento do motor.

Com a evolução das redes intraveiculares e dos automóveis, foi estabelecido no início dos anos 2000 o padrão *On-Board Diagnostic II* (OBD-II). Em essência, além de uma maior compatibilidade com os atuais modelos de automóveis comerciais, o padrão OBD-II também permite a captura de informações com maior taxa de transmissão nos diferentes padrões automotivos (CAN, K-line). A Figura 11 mostra o conector e pinagem correspondente das interfaces OBD-II disponíveis nos atuais veículos comerciais.

Outra maneira de descrever o padrão OBD-II é a partir do formato de sua mensagem,

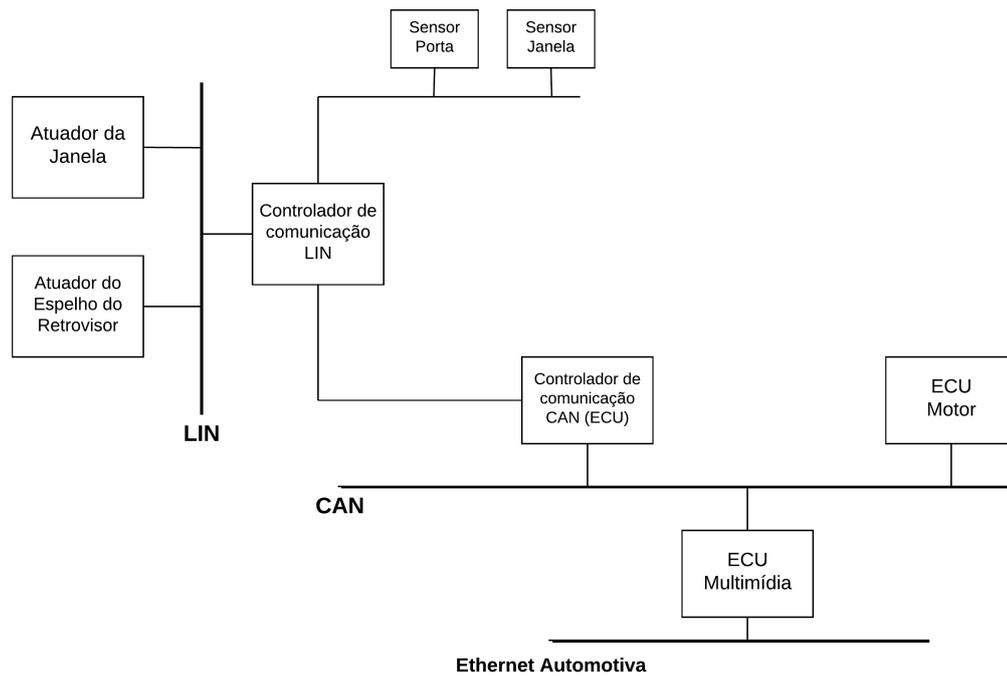


Figura 9 – Representação da divisão de diferentes barramentos intraveiculares instalados em um automóvel

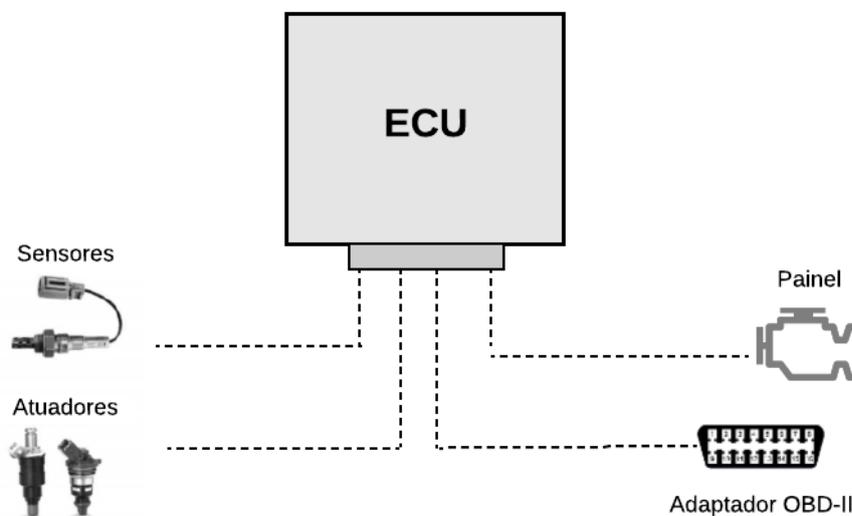


Figura 10 – Representação do funcionamento da interface OBD-II

definida após o envio de um código PID (*Parameter IDs*). O envio deste código em formato de numeração hexadecimal define qual informação do veículo está sendo requisitada naquele momento pelo controlador externo a interface OBD-II. O padrão dos códigos PIDs é definido pelo padrão SAE J1979 (SAE, 2017). O tamanho de uma mensagem com informações da OBD varia entre 1 e 4 *bytes*. A ilustração do processo de requisição e troca de mensagens pode ser visto na Figura 12.

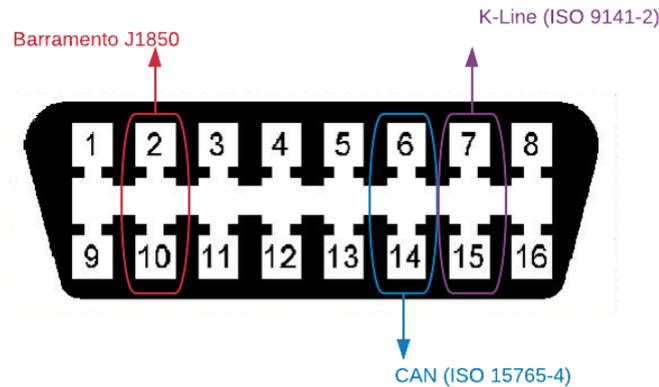


Figura 11 – Formato e descrição dos barramentos disponíveis no conector OBD-II comercial

Fonte: adaptado de (OBDII.COM, 2019)

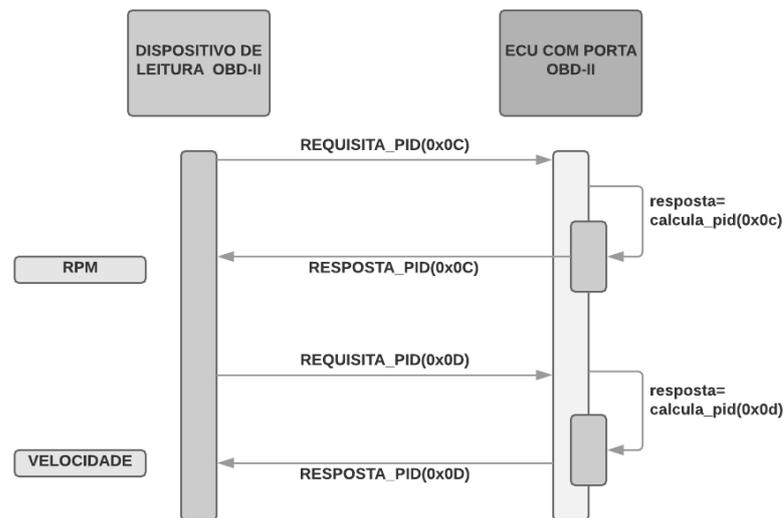


Figura 12 – Processo de requisição de mensagens no padrão OBD-II

Diferentemente dos protocolos CAN e LIN, as informações na OBD encontram-se disponíveis e as mensagens são trocadas quando requisitadas. A origem dos dados disponíveis na OBD-II são em sua maior parte provenientes da ECM (*Engine Control Module*) e PCM (*Powertrain Control Module*), duas ECUs específicas com informações do motor e sistema *powertrain* (SAE, 2017).

A divisão das informações disponíveis é classificada em dez categorias, chamadas de serviços, descritos pelo padrão SAE J1979. A distinção entre estas está relacionada ao fornecimento de dados em tempo real, DTCs e mensagens salvas em situações específicas (ELECTRONICS, 2018). Considerando o conteúdo da mensagem, esta é composta por um identificador e até 8 *bytes*, considerando a inclusão de campo de verificação de erro (CRC e *checksum*). A Figura 13 detalha o formato de uma mensagem OBD e seus possíveis dois

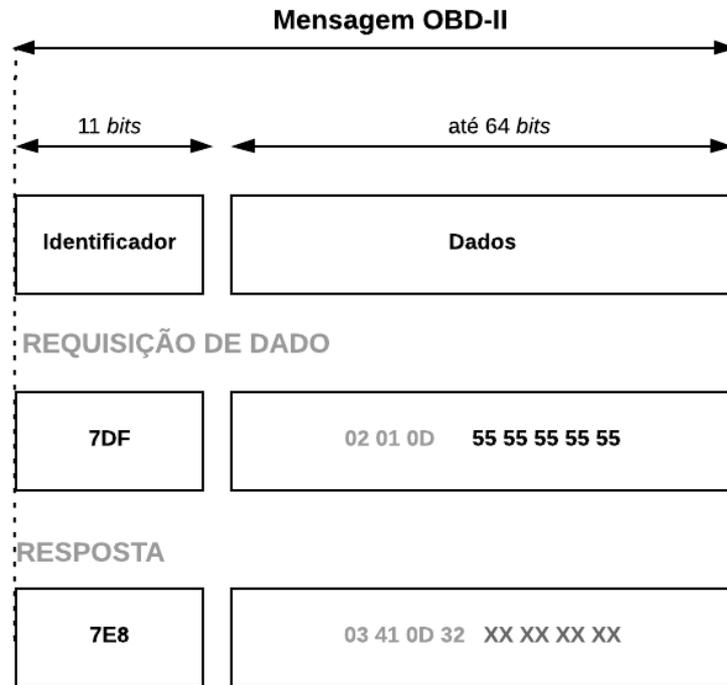


Figura 13 – Descrição dos tipos de *frame* OBD

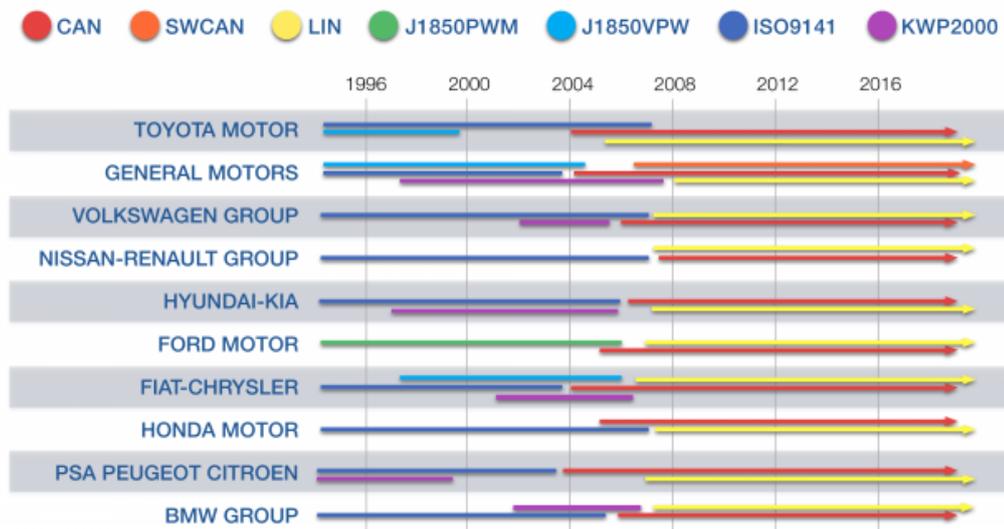


Figura 14 – Evolução dos barramentos intraveiculares

Fonte: (MACCHINA, 2018)

diferentes tipos, quando requisita o dado e com a resposta.

Uma comparação baseada na linha do tempo, e diferentes montadoras e modelos de carro, a Figura 14 mostra a presença dos diferentes tipos de redes intraveiculares ao longo das últimas duas décadas.

2.3 DETECÇÃO DE ANOMALIAS

O processo de detecção de anomalias pode ser resumido como encontrar padrões em dados que não estão seguindo um comportamento esperado, (CHANDOLA; BANERJEE; KUMAR, 2009). A proposta deste trabalho é detectar anomalias a partir das informações intraveiculares disponíveis através da interface OBD-II, em específico, a detecção de anomalias do sensor ECT.

No cenário dos sensores da rede intraveicular, as possíveis anomalias a serem detectadas são provenientes de possíveis falhas de componentes da estrutura veicular, como o *mau* funcionamento de sensores. Independente do componente analisado, o processo para detecção envolve uma análise baseada em informações de uma base de dados. Cada conjunto de dados pode ser descrito usando diferentes atributos (variáveis características). Além da caracterização do conjunto de dados, também torna-se importante o tipo de anomalia a ser detectada.

2.3.1 Tipos de anomalias

Em geral as anomalias podem ser classificadas em três categorias diferentes:

- **Anomalia Pontual:** é o tipo de anomalia mais simples e pode ser detectada quando uma instância de dados individual for anômala em relação ao restante. O agrupamento de dados mostrado na Figura 15 (a) mostra a presença de anomalias pontuais, representadas pelo conjunto O .
- **Anomalia Contextual:** quando uma instância de dados é considerada anômala em um contexto específico. Neste tipo, não é necessária a comparação com todas as outras instâncias do sinal. Torna-se importante ressaltar que uma instância de dados pode ser uma anomalia contextual em um determinado cenário, mas a mesma instância pode ser considerada normal em um contexto diferente. A Figura 15 (b) mostra um intervalo com presença de anomalia contextual.
- **Anomalia Coletiva:** quando uma coleção de instâncias de dados for anômala com relação a todo conjunto de dados. Neste tipo de anomalia as instâncias individuais podem não ser anomalias, mas a ocorrência em conjunto como coleção é caracterizada como tal, como ilustrado na Figura 15 (c).

Alguns fundamentos do aprendizado de máquina (*machine learning*) definem as etapas e processos da análise de anomalias em sinais e bases de dados. Segundo (MITCHELL et al., 1997), o aprendizado de máquina é o estudo de algoritmos que melhoram o desempenho P de alguma tarefa T com experiência E . Contextualizando a definição para a detecção de anomalias em sinais veiculares, tem-se:

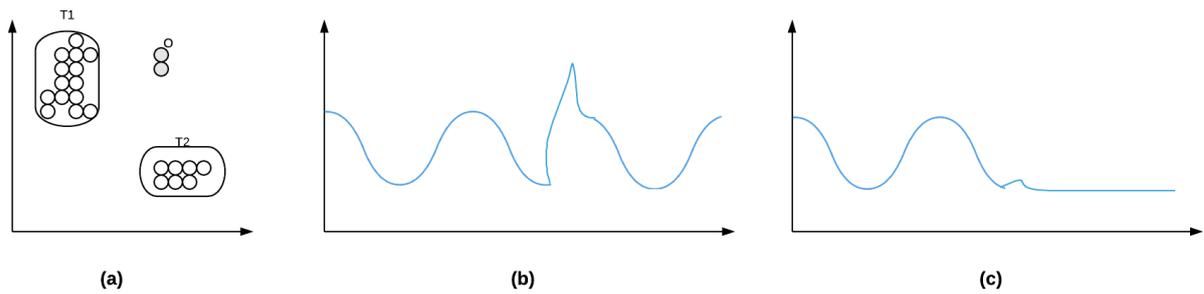


Figura 15 – Tipos de anomalias (a) Pontual (b) Contextual (c) Coletiva.

Fonte: adaptado de (CHANDOLA; BANERJEE; KUMAR, 2009)

- **Medida de desempenho P:** porcentagem de instâncias classificadas de maneira correta (funcionamento em perfeitas condições ou presença de anomalia).
- **Tarefa T:** reconhecer a operação em condição normal de um sistema ou sensor da rede intraveicular.
- **Experiência de treinamento E:** o processo de aprendizado é realizado a partir de uma base de dado com informações da rede intraveicular.

De maneira resumida, o aprendizado de máquina utiliza ferramentas matemáticas e estatísticas para uma análise automatizada dos dados. Sendo um dos ramos da inteligência artificial, é baseado na premissa que os sistemas podem aprender a partir dos dados fornecidos de um conjunto e conseqüentemente, identificar padrões a partir de decisão própria, com mínima intervenção humana.

Dessa forma, a solução para um novo problemas é definida a partir da adaptação de soluções dadas a problemas similares. O processo de treinamento, etapa em que exemplos previamente rotulados são recuperados e adaptados para a solução de um novo problema, define o uso de uma instância de dados.

A formalização de fundamentos do aprendizado de máquina para a detecção de anomalias pode ser vista com o uso de rótulos (*data labels*) associados a uma instância de dados, indicando assim se a mesma é normal ou anômala. Outra parte fundamental do processo é a identificação, na instância de dados, de quais atributos podem ser utilizados no processo de detecção, sendo esta etapa chamada de seleção de características.

O aprendizado pode ser baseado em instâncias, isto é, cada vez que um novo conjunto de dados é recebido pela máquina de aprendizado, o modelo construído, se computa uma função objetivo com base no conhecimento oferecido pela base composta pelos exemplos de treinamento. A estimativa da classe da nova instância é então baseada em comportamentos locais.

Ao considerar a presença, ou ausência, de rótulos nos dados disponíveis, o processo de detecção de anomalias também pode ser dividido em:

- **Detecção de anomalia supervisionada:** as técnicas utilizadas nesta categoria assumem a presença de dados anômalos (*outliers*) e dados de operação normal (*target*). É similar ao processo de classificação binária em aprendizagem de máquina.
- **Detecção de anomalia semi-supervisionada:** as técnicas desta categoria utilizam apenas instâncias de treinamento rotuladas com a classe de operação normal. Assim, um classificador construído a partir de uma técnica deste tipo pode classificar a presença de um comportamento normal ou anômalo. Apesar de classificar uma instância como anomalia, o classificador não possui capacidade para detectar qual tipo de anomalia ocorreu.
- **Detecção de anomalia não supervisionada:** os dados disponíveis para classificação não possuem nenhum tipo de rótulo. Assim, a saída do classificador é baseada em grupos de dados (*clusters*).

Além do uso de técnicas aprendizado de máquina, as técnicas de detecção de anomalias mais utilizadas são, em geral, baseadas em processos estatísticos, análise de limiar (*thresholding*) e técnicas de análise de densidade e agrupamentos (*clustering*).

2.3.2 Técnicas de detecção de anomalias

Além de diversas áreas de pesquisa, outros fatores influenciam para o uso de diferentes técnicas para detecção de anomalias. A natureza e volume dos dados também habilitam diferentes desafios nesta área. Assim, fundamentos das diferentes técnicas podem ser apresentados.

2.3.2.1 Análise Estatística

O uso da análise estatística no processo de detecção de anomalias é baseado na observação de que uma anomalia não segue o modelo estocástico assumido para o sinal, ou grupo de sinais, em análise (ANSCOMBE, 1960). Dessa maneira, as técnicas de detecção de anomalias baseadas em processos estatísticos são baseadas em hipóteses:

- **Hipótese:** as instâncias de dados dos sinais normais (*target*) são característicos de regiões de alta probabilidade em um modelo. De outro lado, as anomalias são caracterizadas nas regiões de baixa probabilidade do modelo estocástico.

O uso de testes de hipótese ajusta um modelo para os respectivos tipos de dados analisados. Assim, para detectar um comportamento anômalo, os dados em questão podem ser submetidos a um teste para determinar se o conjunto pertence ao modelo ou não. Considerando a aplicação das diferentes técnicas, tanto o uso das técnicas paramétricas, aquelas onde os dados são considerados provenientes de um tipo de distribuição específica, ou das não paramétricas, são utilizadas na detecção de anomalias.

Técnicas Paramétricas: assume uma distribuição normal dos dados a partir de uma função de densidade de probabilidade $f(x, \Theta)$. De maneira alternativa as técnicas que são baseadas em manipulações a partir função de densidade de probabilidade, testes de hipótese também podem ser usados para a detecção de *outliers* estatísticos.

De acordo com (TARASSENKO et al., 1995), ao considerar a análise estatística, o método mais direto para a construção de um classificador *one-class* é a partir da estimação dos dados de treinamento e a consequente definição de um limiar a partir destes. Segundo (TAX, 2001), quando o tamanho da amostra dos dados é suficientemente alto, um modelo de densidade pode ser usado, para sua estimativa e consequente detecção de *outlier*.

As técnicas paramétricas mais utilizadas se baseiam na distribuição normal (Gaussiana), sendo esta uma das mais utilizadas em modelos estatísticos. Como proposto por (TAX, 2001), as técnicas paramétricas mais simples que podem ser usadas como classificadores *one-class* são: *Gauss Data Description* e o *Extreme Value Analysis*.

No classificador *one-class* Gauss, a classe que representa os dados de funcionamento normal é modelada como distribuição gaussiana. A estimativa de densidade é calculada utilizando a distância de Mahalanobis:

$$f(x) = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (2.1)$$

O classificador é então definido como:

$$h(x) = \begin{cases} target & \text{se } f(x) \geq \theta \\ outlier & \text{se } f(x) < \theta \end{cases} \quad (2.2)$$

A média μ e a matriz de covariância Σ são provenientes da amostra analisada. O *threshold* θ é definido de acordo com conjunto de dados contendo funcionamento normal fornecido no conjunto de treinamento. O modelo ilustrado por uma distribuição Gaussiana pode ser representado na Figura 16.

Ainda considerando a Figura 16, a distribuição de probabilidade para um objeto d -dimensional \mathbf{x} é dada por:

$$p_{\mathcal{N}}(z; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right) \quad (2.3)$$

De maneira semelhante, outra técnica que também utiliza a distância de Mahalanobis é o classificador *one-class* de mesmo nome, conhecido como Mahalanobis *classifier*, para problemas que envolvem uma classe. A técnica também usa a métrica da distância de Mahalanobis como parâmetro de decisão para as instâncias. De maneira diferente, a média considerada no cálculo da distância é a do conjunto de treinamento como um todo. Comparando com a técnica Gauss, o classificador de Mahalanobis retorna em geral, resultados melhores para espaços de dimensões mais altas, visto que evita a estimativa Gaussiana de cada instância.

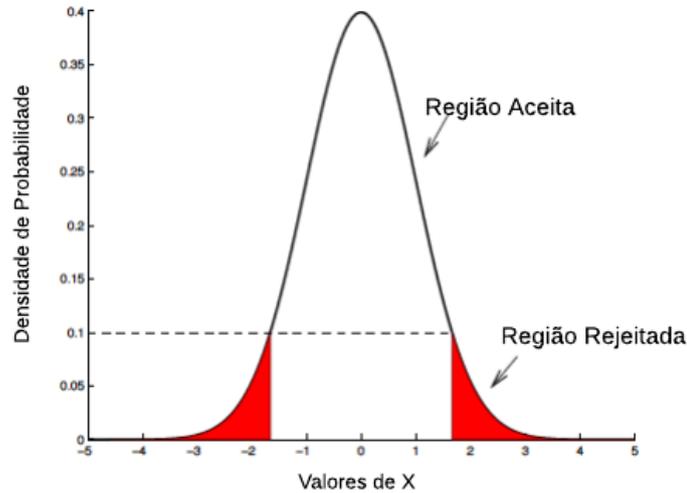


Figura 16 – Definição de limiar para detecção de anomalia em uma distribuição Gaussiana

Fonte: Adaptada de (TAX, 2001)

Além do uso do *Gauss Data Description*, a outra forma mais básica utilizada para detecção de *outliers* é a análise de valor extremo, *Extreme-Value Analysis*, de dados unidimensionais. No caso de uma distribuição normal assimétrica, ou seja, um gráfico similar ao da Figura 16, porém com áreas diferentes nas duas caudas. Nesse caso específico, o *threshold* é definido a partir da cauda com maior densidade calculada. Embora a análise a partir da técnica de valor extremo seja projetada para conjuntos de dados unidimensionais, também é possível generalizá-la para conjuntos com dados em mais de uma dimensão a partir de análise em campo multidimensional (AGGARWAL, 2015).

Técnicas Não Paramétricas: considerando que as técnicas baseadas em modelos estatísticos não paramétricos não possuem um modelo definido, se faz necessário o uso de uma função *kernel* para aproximação do modelo real. Assim, uma técnica baseada em estimativa de densidade mais flexível, visto que não depende de modelo Gaussiano, é a *Parzen density estimator*. A estimativa de densidade pode ser calculada como:

$$f(x) = \sum_{i=1}^N \exp(-(\mathbf{x} - \mathbf{x}_i)^T h^{-2}(\mathbf{x} - \mathbf{x}_i)) \quad (2.4)$$

O parâmetro h é otimizado utilizando a verossimilhança a partir do método *leave-one-out*. A regra de decisão se torna a mesma descrita em (2.2). Para o bom funcionamento deste método, torna-se necessário um conjunto de treinamento razoável (TAX, 2001), o que corresponde a uma base de dados descrevendo diferentes possibilidades do evento a

ser analisado. As limitações da aplicabilidade da técnica são relacionadas a necessidade de armazenamento das instâncias utilizadas na construção do modelo.

Outra proposta a partir da técnica é baseada de maneira similar a abordagem da técnica *Naive Bayes*, para a estimativa da classe a partir de diferentes atributos. A técnica *Naive Parzen* propõe uma combinação a partir de cada atributo, sendo esta caracterizada pela estimativa de Parzen, e a densidade total calculada multiplicando as densidades por característica (TAX; DUIN, 2008):

$$f(x) = \prod_{l=1}^p \sum_{i=1}^N \exp(-(\mathbf{x}_l - \mathbf{x}_{li})^T h^{-2}(\mathbf{x}_l - \mathbf{x}_{li})) \quad (2.5)$$

Em uma comparação com (2.3), a densidade total das classes na técnica *Naive Parzen* é calculada como:

$$\hat{p}(x|\omega_i) = \prod_{l=1}^p \frac{1}{n_i} \sum_{i=1}^{n_i} \mathcal{N}(x_l; x_{il}, h_l^2) \quad (2.6)$$

2.3.2.2 Sistemas baseados em regras

Outras técnicas para detecção de anomalias não necessariamente utilizam fundamentos de estatística. Em geral, estas técnicas são baseadas no cálculo de um limiar para decisão. Segundo (AGGARWAL, 2015), a ideia principal dos métodos baseados em limiar de proximidade (*proximity base*) é modelar os *outliers* como pontos que são isolados dos dados restantes, *target*, com base em funções de similaridade ou distância. Uma das técnicas mais simples que utilizam a ideia de verificar apenas em uma vizinhança restrita a presença de *outliers* é o *Linear Outlier Factor* (LOF).

Linear Outlier Factor (LOF):

A técnica é baseada na detecção de dados discrepantes em relação a densidade em torno de um objeto, uma observação, do conjunto de dados. Proposta por (BREUNIG et al., 2000), o funcionamento pode ser resumido a partir da definição do número de vizinhos a ser analisados, k , e cálculo da métrica de densidade e por fim, o valor do *local outlier factor*. A aplicação numa vizinhança é ilustrada na Figura 17.

Em seguida a definição do número de vizinhos, a distância até o k -ésimo vizinho mais próximo, definida como $dist_k$, é calculada baseada no grupo de dados de teste ($G(x_{test})$). O cálculo pode ser definido baseado nas distâncias da instância analisada e o grupo de testes, definido k cálculos a partir da distância de alcance (x_{test}, x_i).

$$d_r(x_{test}, x_i) = \max(dist_k(x_i), dist(x_{test}, x_i)) \quad (2.7)$$

Em seguida é determinado um valor médio para o grupo de testes, contendo os vizinhos mais próximos:

$$d_r(x_{test})_{avg} = \frac{\sum d_r(x_{test}, x_i)}{\|G(x_{test})\|} \quad (2.8)$$

O inverso desta distância média também é utilizada para ao cálculo do LOF, podendo ser interpretada como uma densidade do grupo de teste.

$$d_{dens}(x_{test}) = \frac{\|G(x_{test})\|}{\sum d_r(x_{test}, x_i)} \quad (2.9)$$

O cálculo do LOF pode ser então determinado pelo somatório dos índices de densidade de cada instância vizinha no grupo $G(x_{test})$ e a densidade $d_{dens}(x_{test})$ no número de instâncias vizinhas.

$$LOF(x_{test}) = \frac{\sum \frac{d_{dens}(x_i)}{d_{dens}(x_{test})}}{\|G(x_{test})\|} \quad (2.10)$$

Ao considerar o uso do LOF na detecção de anomalias, deve-se usar o valor máximo de LOF dentro do conjunto de treinamento como limiar para a decisão.

$$thrsh_{LOF} = \max(LOF(x_i \forall i)) \quad (2.11)$$

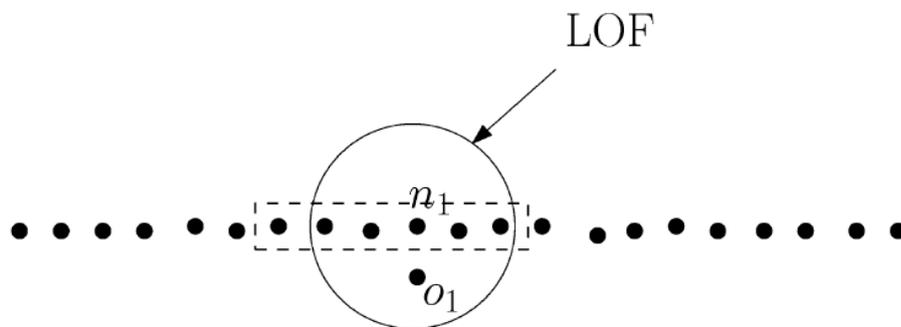


Figura 17 – Exemplo ilustrativo do funcionamento do LOF

Fonte: adaptado de (CHANDOLA; BANERJEE; KUMAR, 2009)

Além dos métodos baseados em limiares e proximidade, técnicas baseadas em agrupamento (*clustering*) também podem ser aplicadas para o problema de detecção de uma classe, um exemplo é o uso do *k-means data description*. Pode ser realizada uma divisão de algumas das técnicas utilizadas, divididas em suas respectivas categorias:

- **Técnicas baseadas em instâncias:** k-nearest neighbour (k-NN) e suas variações, Linear Outlier Factor (LOF).
- **Redes neurais:** *Self-organizing Map* (SOM), *Multilayer Perceptrons* e redes baseadas em funções RBF (*Radial Basis Function*).

- **Baseado em regras:** *Minimum Spanning Tree* (MST) e distância de Mahalanobis.
- **Máquina de vetores de suporte:** *Support Vector Data Description* (SVDD) and *One-class SVM* (OC-SVM).

Além das técnicas listadas acima, separadas por categorias, outras técnicas utilizadas em aprendizado de máquina também possuem sua versão adaptada para o problema detecção de uma classe.

2.4 CLASSIFICADORES ONE-CLASS

Esta subseção apresenta algumas técnicas populares em projetos que envolvem aprendizado de máquina, adaptadas para a solução de problemas que envolvem a detecção de apenas uma classe.

2.4.1 *Self-organized map* (SOM)

O método baseado em *Self-organized map* (SOM) para detecção de anomalias utiliza um mapa auto-organizável para a descrição dos dados. A técnica SOM é um método de *clustering*, sendo assim uma técnica de aprendizado não-supervisionada. A construção dos centros dos *clusters* é baseada no objetivo de que um objeto no espaço obtenha o máximo de informação possível, relacionada a sua vizinhança (TAX, 2001). Uma etapa importante do método é um mapeamento realizado a partir de rede neural definida com uma regra específica.

Exemplificando, o mapeamento de um espaço de características d -dimensional em um k -dimensional, com $k < d$. É definido então, uma grade ortogonal $K \times K$ onde em cada ponto da grade é estabelecido um neurônio. A cada neurônio associado, um vetor d -dimensional com informações do centro de um *cluster* é registrado. Ao definir uma rede para os neurônios, cada neurônio não possui apenas um vizinho no espaço, mas também um neurônio vizinho na grade. A etapa de treinamento os neurônios vizinhos na grade também são vizinhos no espaço de medição. Na etapa de classificação, a avaliação de um objeto a pertencer ou não ao modelo, a um *cluster* mais próximo de um neurônio, é realizada a partir de :

$$f(x) = \min \|x - x_{SOM}\|^2 \quad (2.12)$$

2.4.2 *Support Vector Data Description* (SVDD)

A técnica *Support Vector Data Description* (SVDD) utiliza como região de decisão uma hipersfera, que pode ser entendida como uma esfera generalizada para um espaço multi-dimensional, com a correspondência a um círculo em duas dimensões e a uma esfera em três dimensões. Os parâmetros que determinam a hipersfera são seu raio R e seu centro

a. O problema solucionado pela técnica SVDD é determinar a escolha do centro e o valor do R com o menor valor do erro na classe normal e a chance de classificar erroneamente os dados anômalos (TAX; DUIN, 2004).

2.4.3 k-NN

Considerando o processo de aprendizado baseado em instâncias, uma das técnicas mais simples é a do k vizinhos mais próximos, o k -NN. A técnica proposta por (FUKUNAGA; NARENDRA, 1975), assume que todas as instâncias correspondem a pontos em um espaço n -dimensional (\mathbb{R}^n) e que sua vizinhança pode, então, ser definida por uma função de distância ou similaridade. Dessa maneira, a etapa de classificação, a definição do rótulo de uma nova instância, é definida a partir dos vizinhos mais próximos.

Definindo então x_i como a instância descrita pelo vetor $\langle a_1(x_i), a_2(x_i), \dots, a_n(x_i) \rangle$ e $f(x_i)$ a classe da instância x_i , etapas do processo de aprendizado de máquina podem ser definidas para o algoritmo k -NN. A etapa de treinamento para esta técnica consiste em armazenar os exemplos $\langle x_i, f(x_i) \rangle$. Apesar de suficientes para a aplicação da técnica, é recomendada uma etapa de pré-processamento. Assim, para o uso da técnica é realizada uma transformação dos dados a partir da normalização dos mesmos, determinando os atributos com valores entre 0 e 1. Como consequência do armazenamento das instâncias de dados, o k -NN é um dos métodos de aprendizado baseado em instâncias mais elementar, utilizando a distância Euclidiana como métrica para classificação.

Considerando duas instâncias x_i e x_j , a distância Euclidiana entre as duas é definida como:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (2.13)$$

A equação 2.13 mostra matematicamente a necessidade pela normalização dos dados ao utilizar a técnica, visto que atributos com valores em diferentes unidades decimais podem influenciar no resultado final. De maneira direta, a aplicação da técnica é baseada em determinar o limiar na etapa de treinamento, a partir das distâncias armazenadas em um conjunto de distâncias $\|T\|$. A escolha pelo limiar pode ser feita definindo o número dos vizinhos mais próximos para o cálculo. Os passos a seguir resumem sua aplicação:

1. Armazenamento de todas as instâncias de treinamento, definindo assim $\|T\|$.
2. Definição do número k de vizinhos.
3. Para uma instância desconhecida x_i , os k vizinhos mais próximos em $\|T\|$ são definidos a partir do cálculo da distância euclidiana.
4. A classe da instância x_i é definida a partir da classe da maioria dos k vizinhos.

Apesar de ser bastante utilizado, o roteiro descrito anteriormente é utilizado em problemas de classificação multi-classe. Para os problemas de classificação (*one-class*, o que inclui detecção de anomalias, pequenas diferenças podem ser observadas na técnica k-NN. De maneira similar ao funcionamento do LOF, a distância entre os k vizinhos mais próximos pode ser vista como uma métrica de estimativa de densidade local, e assim, ser usada para detectar anomalias.

Assim, para o uso do k-NN para problemas de detecção de uma classe, a distância média definida a partir da distância Euclidiana dos k vizinhos mais próximos é definida a partir de sua média.

$$dist_{avg}(x_{test}) = \frac{1}{k} \sum dist(x_{test}, x_i) \quad (2.14)$$

Como consequência, o estabelecimento do *threshold* se dá pelo valor máximo da média das distâncias entre k vizinhos.

$$threshold = max(dist_{avg} \forall i) \quad (2.15)$$

De forma similar ao problema multi-classe, a determinação do limiar ocorre no período de treinamento a partir das distâncias salvas no conjunto $\|T\|$. Para os casos em que $k = 1$, o k-NN se torna 1-NN. Além disso, como discutido por (THEISLER, 2013), em sistemas que são baseados em *thresholding*, o k-NN torna-se uma boa alternativa para definição inicial do limiar.

2.4.4 OC-SVM

Diferente ao aprendizado baseado em instâncias, as Máquinas de Vetores de Suporte, do inglês *Support Vector Machine* (SVM), se baseiam em fronteiras lineares para a separação de diferentes conjuntos de dados. Apesar de também ser uma técnica de aprendizado baseada em processos estatísticos, as SVMs não fornecem estimativas de probabilidade diretamente, visto que estas são calculadas usando validação cruzada (PEDREGOSA et al., 2011).

Assim, os dados de entrada de uma máquina de vetor de suporte são mapeados de um espaço de alta dimensão para um hiperplano em que melhor descreve a separação das características, como ilustrado na Figura 18. Considerando a fundamentação matemática da técnica, esta etapa é estabelecida a partir do uso funções denominadas *kernel*, o núcleo de uma transformação linear. No caso específico para a detecção de anomalias, a técnica é apresentada como *One Class Support Vector Machine* (OC-SVM).

De maneira geral, a técnica OC-SVM pode ser vista como uma SVM regular para classificação binária. Diferente do caso multi-classe, a etapa de treinamento contém ape-

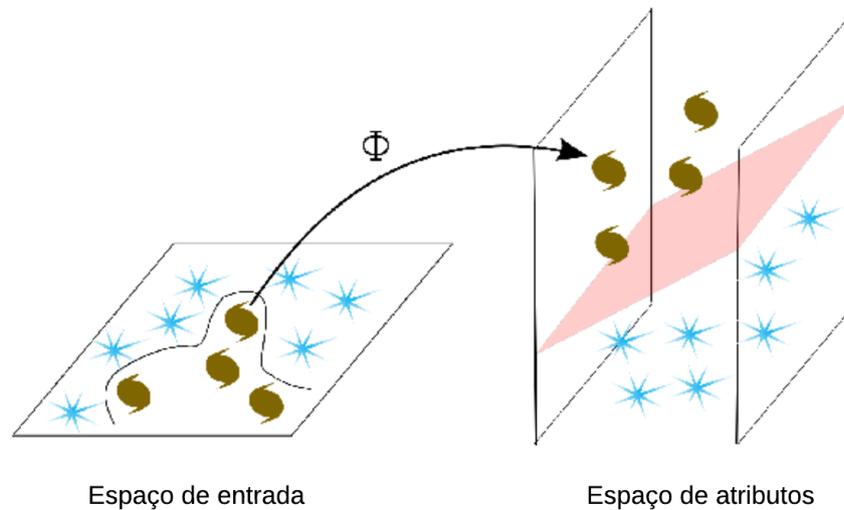


Figura 18 – Representação da transformação linear entre o espaço de entrada e o de atributos para a técnica SVM

Fonte: adaptada de (MALEK et al., 2013)

nas dados de uma das classes. O hiperplano de separação das classes pode ser descrito matematicamente como:

$$f(x) = \langle w, x \rangle + b \quad (2.16)$$

Ainda considerando a Figura 18, mais de um plano de separação pode ser definido para separar as classes. A técnica é baseada na solução de um problema de otimização para encontrar o hiperplano ótimo, ou com margem geométrica máxima. Além disso, os pontos mais próximos do hiperplano de separação são os vetores suporte, com w sendo o vetor normal e b o fator de polarização, *bias term*. Para encontrar os valores para os vetores de suporte a técnica utiliza o método dos multiplicadores de Lagrange como solução para o problema de otimização.

Considerando uma fronteira de separação, dois hiperplanos podem ser definidos, o superior e inferior, sendo escalonados como:

$$\begin{cases} x_1 \cdot w + b = 1 \\ x_2 \cdot w + b = -1 \end{cases} \quad (2.17)$$

Ao considerar vetores de suporte dos dois hiperplanos, sendo estes x_1 e x_2 , a diferença entre ambas as funções define a margem, $\frac{2}{\|w\|}$. O problema de otimização busca maximizar este valor, visto que a distância mínima entre o hiperplano separador (ótimo) e possíveis dados de treinamento pode ser calculada por $\frac{1}{\|w\|}$.

O problema de otimização a ser solucionado pode ser resumido como maximizar $\frac{2}{\|w\|}$ sujeito a $y_i \cdot (x_i \cdot w + b) - 1 \geq 0$. Para encontrar os valores para os vetores de suporte a técnica

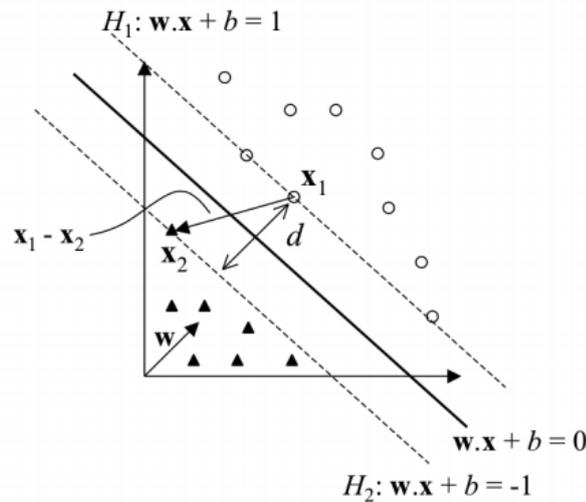


Figura 19 – Exemplo do cálculo da distância entre os hiperplanos

Fonte: adaptada de (LORENA; CARVALHO, 2007)

utilizada o método dos multiplicadores de Lagrange para o problema de otimização. Como exemplo, a regra de classificação para a técnica é realizada a partir do cálculo do de $f(x_i)$ para uma instância de dados x_i .

$$h(x) = \begin{cases} target & \text{se } f(x_i) \geq 0 \\ outlier & \text{se } f(x_i) < 0 \end{cases} \quad (2.18)$$

Durante o processo de otimização a partir dos multiplicadores de Lagrange, outro parâmetro importante do sistema é a função *kernel*. O *kernel* projeta os vetores do espaço de entrada para o espaço de atributos, permitindo assim, limites de decisão não-lineares. Considerando $\phi(x)$ a transformação que leva o espaço de entrada X em \mathbb{R}^N , uma função *kernel* pode ser definida por:

$$K(x, y) = \langle \phi(x), \phi(y) \rangle \quad (2.19)$$

Diferentes algoritmos de SVM usam diferentes tipos de funções *kernel*. Alguns dos tipos mais utilizados são:

- Linear: $K(x, y) = (x \cdot y)$
- Polinomial: $K(x, y) = (x \cdot y + 1)^d$
- Gaussiano: $K(x, y) = e^{-\|x-y\|^2/(2\sigma^2)}$
- RBF (*Radial Basis Function*): $K(x, y) = \exp(-\frac{\|x-y\|^2}{2\sigma^2})$
- Sigmoid: $K(x, y) = \tanh(\gamma \cdot x^T \cdot y + r)$

De forma a ilustrar algum destes tipos, a Figura 20 mostra as fronteiras de separação ao utilizar cada respectivo tipo.

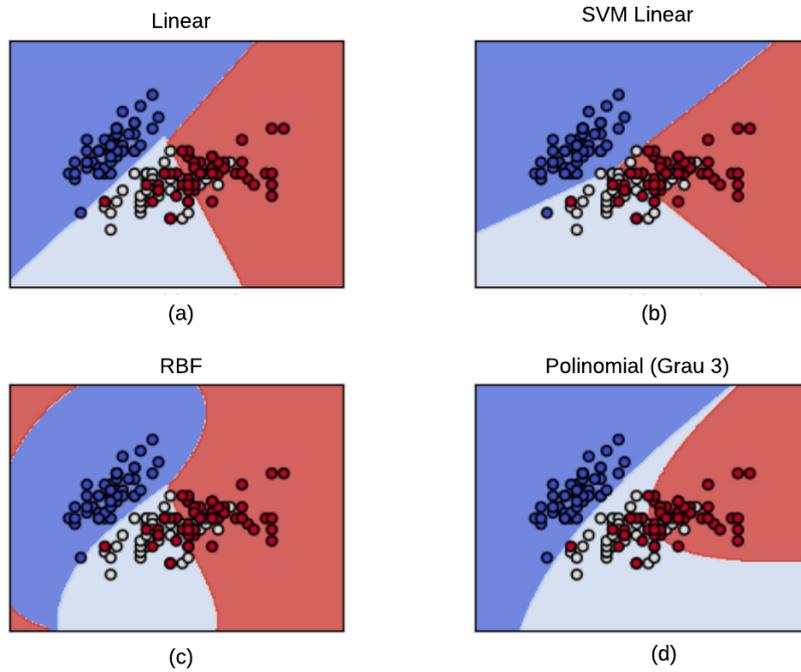


Figura 20 – Fronteiras de separação dos diferentes tipos de *kernel* (a) Linear (b) Linear com SVM Linear (c) RBF (d) Polinomial

Fonte: adaptada de (PEDREGOSA et al., 2011)

3 REVISÃO BIBLIOGRÁFICA

Trabalhos relacionados a diferentes sistemas de telemetria veicular e a detecção de anomalias em redes veiculares são apresentadas neste capítulo. Na Seção 3.1 é apresentada uma divisão do tema telemetria veicular em diferentes áreas, mostrando subáreas de interesse e a organização dos diferentes sistemas propostos. Em seguida, a Seção 3.2 aborda a detecção de anomalias em redes veiculares, com diferentes trabalhos propostos pela literatura.

3.1 SISTEMAS DE TELEMETRIA VEICULAR

Foi realizado um mapeamento sistemático para definir o estado da arte em telemetria veicular. Em geral, um mapeamento tem como objetivo definir um método para classificação em uma determinada área de interesse. Os resultados mostram as quantidades e frequências de publicações por categorias pré-definidas. Assim, torna-se necessário elaborar perguntas de pesquisa para estruturar as possíveis divisões do processo de telemetria veicular.

Focando em aspectos de implementação, as perguntas de pesquisa foram elaboradas para dividir os sistemas de telemetria em categorias referentes a arquitetura de *hardware* utilizada, tecnologia de conectividade e área de aplicação. Para selecionar repositórios que envolvam tanto artigos de engenharia, como também artigos de ciência da computação, as seguintes bibliotecas digitais são utilizadas: *ACM Digital Library*, *IEEEExplore* e *Science Direct*. A busca inicial por artigos foi realizada em novembro de 2017, com trabalhos adicionais incluídos em dezembro de 2017, resultando em um total de 2309 artigos. Artigos duplicados e não relacionados ao tema foram removidos, resultando em 123 artigos selecionados como base.

Analisando os artigos selecionados, outra possível divisão do tema de pesquisa é relacionada a qual técnica é pesquisada ou melhorada ao considerar o tema de computação em nuvem, relacionada a área veicular. Assim, torna-se possível a divisão em quatro categorias diferentes, contendo ramificações, baseadas nos diferentes artigos selecionados. A fim de fornecer uma visão geral do atual estado da arte dos sistemas de telemetria veicular, as diferentes categorias representadas na Figura 21 são detalhadas a seguir:

- **Organização do sistema de aquisição de dados:** em geral, os artigos pesquisados com propostas de sistemas de aquisição de dados para telemetria veicular podem ser divididos a partir de sua arquitetura de *hardware*. Como consequência, foram identificadas duas possíveis áreas para a divisão dos trabalhos.

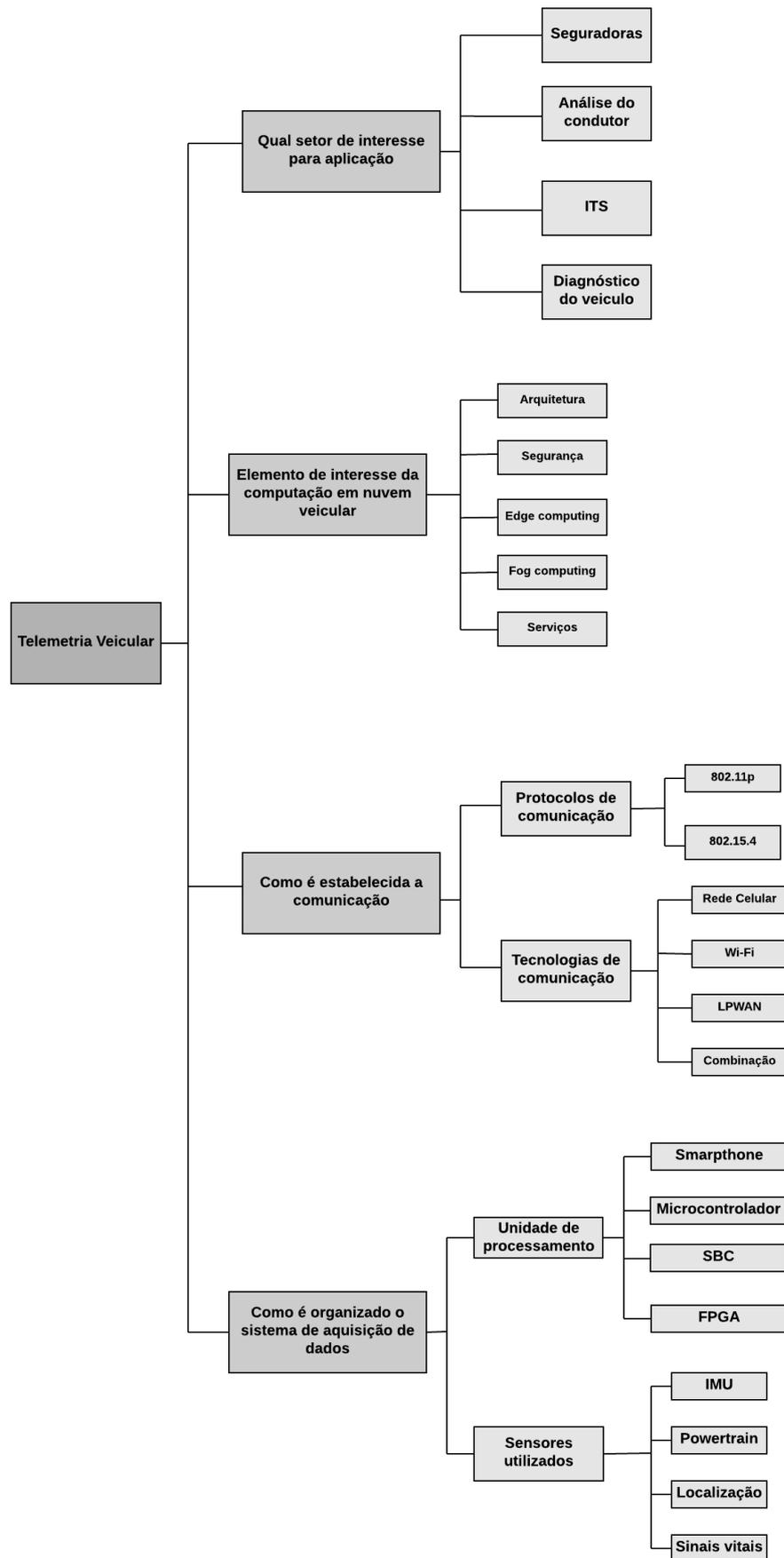


Figura 21 – Taxonomia dos sistemas de telemetria veicular

- **Unidade de processamento:** são os dispositivos da computação ubíqua ou sistemas embarcados capazes de gerenciar todo o processo de telemetria, incluindo a comunicação com os sensores, envio das informações e inteligência computacional.
 - **Sensores:** são os elementos responsáveis por converter uma grandeza física em sua representação digital. Assim, no processo de telemetria, diferentes tipos de sensores são utilizados a fim de monitoramento ou medição de conjuntos específicos de variáveis.
- **Estabelecimento da comunicação:** ao definir um objeto conectado muitas vezes não se sabe a melhor maneira para enviar as informações coletadas. Entre as possibilidades, estão a conexão entre os dispositivos de uma mesma classe, ou comunicação direta do dispositivo à internet, permitindo assim, o uso de diferentes tecnologias de comunicação. Além da tecnologia, os protocolos tornam-se importantes ao controlar a transferência de dados entre os diferentes sistemas.
 - **Tecnologias de comunicação:** cada trabalho pôde ser classificado em relação à tecnologia que foi usada para realizar o processo de telemetria, ou qual tecnologia seria usada na implementação do sistema proposto. As tecnologias para realizar a telemetria veicular podem ser agrupadas em: baseadas em redes celulares (GPRS/2G/3G/4G), Wi-Fi, rede de longa distância com baixo consumo, *Low Power Wide Area Network* (LPWAN), e o uso combinado de tecnologias.
 - **Protocolos de comunicação:** mesmo considerando a divisão anterior, uma mesma tecnologia pode ser usada de diferentes maneiras para estabelecer a comunicação para o processo de telemetria veicular. Assim, os diferentes protocolos são listados como ferramenta auxiliar no processo de aquisição de dados.
 - **Elemento de interesse da computação em nuvem veicular:** após a etapa de coleta de dados e estabelecimento da comunicação, tornam-se necessário o processamento e armazenamento de dados. Com o surgimento do conceito de computação em nuvem, que se tornou uma alternativa para modificar serviços de software e habilitar novos serviços para proprietários de veículos (HE; YAN; XU, 2014), também surgem novos temas de interesse da pesquisa dos artigos relacionados. Os elementos a seguir foram escolhidos por serem os mais discutidos nos trabalhos.
 - **Arquitetura:** estabelecimento de uma arquitetura dividida em camadas que estabeleça comunicação e serviços baseada nas tecnologias de comunicação veicular.
 - **Segurança:** o trabalho apresenta alguma contribuição nesta categoria. A mesma ganha destaque devido ao aumento no número de carros conectados e os di-

ferentes serviços fornecidos ao condutor. Além da segurança, a privacidade também é discutida em parte dos trabalhos.

- **Computação de borda:** trabalhos que abordam o conceito de computação de borda para redes de veículos conectados ou como técnica análise e processamento para reduzir o envio de dados para a nuvem. O objetivo principal com o uso é a redução do tamanho de dados gerados, diminuir a latência da rede e melhorar o tempo de resposta para aplicações de segurança crítica.
 - **Serviços:** considerando o estabelecimento de novos modelos de negócio baseado em serviços a partir do carro conectado, e a sua respectiva plataforma em nuvem, os diferentes trabalhos discutem como estes podem ser oferecidos ao motorista.
- **Qual setor de interesse para aplicação do sistema de telemetria veicular:** esta categoria considera quais são as principais aplicações dos sistemas de telemetria veicular. A partir das áreas de uso, também torna-se possível definir quais sensores e variáveis foram monitoradas. As seguintes áreas foram abordadas nos trabalhos.
 - **Seguradoras:** as diferentes informações coletadas a partir do veículo, como consumo de combustível, funcionamento do motor e informações do condutor, tornam-se de interesse para companhias de seguros. A partir dos dados, novos modelos de negócio e serviços tornam-se possíveis.
 - **Análise do condutor:** como a segurança é uma grande preocupação nos sistemas de transporte, o estado do motorista e, como consequência, seus hábitos de direção tornam-se essenciais para tornar o trânsito mais seguro. Assim, diferentes sistemas de monitoramento são propostos baseados nos hábitos do condutor, estilo de condução e comportamento.
 - **Sistemas de transporte inteligentes:** esta área também usa informações da telemetria veicular para habilitar diferentes aplicações. Além da comunicação entre veículos, baseada na comunicação *Vehicle to Anything* (V2x), dados da telemetria veicular são para análise geral de mobilidade urbana, oferecendo monitoramento em tempo real de rodovias e ocorrência de acidentes.
 - **Diagnóstico do veículo:** a partir de informações de específicas partes do veículo, como bateria, motor e sistema de transmissão, tornam-se possíveis análises relacionadas ao seu atual estado de funcionamento e comportamento das peças. Os trabalhos em geral, focam em sistemas específicos do veículo.

O estabelecimento das palavras chave de pesquisa: telemetria veicular, diagnóstico do veículo e nuvem veicular, direcionam o tipo dos trabalhos relacionados apresentarem em sua maior parte propostas de soluções para implementação de sistemas de telemetria e

temas de pesquisa da nuvem veicular. A divisão do tipo de pesquisa apresentado pelos artigos selecionados pode ser vista na Figura 22.

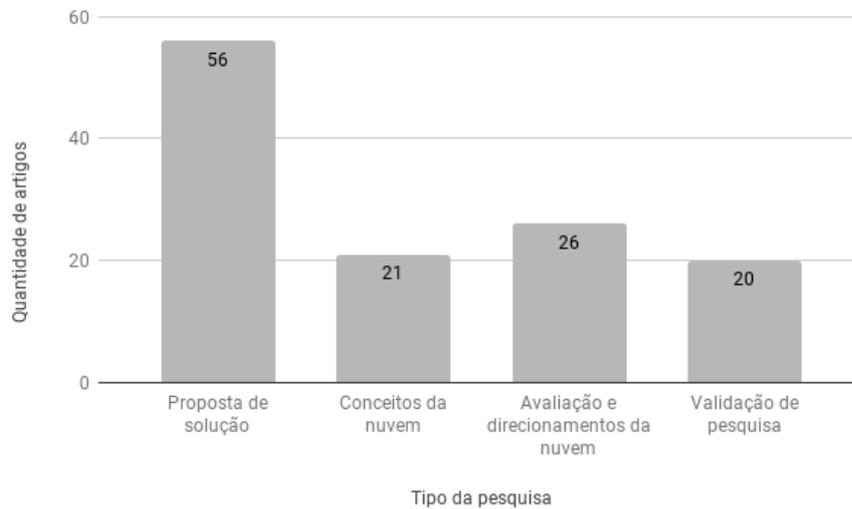


Figura 22 – Número de trabalhos classificados no tipo da pesquisa

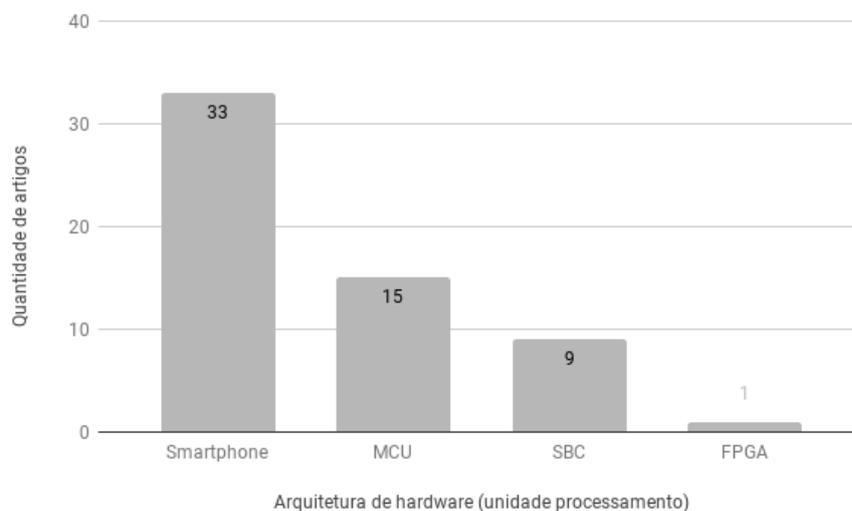


Figura 23 – Número de trabalhos classificados em cada arquitetura de *hardware*

Considerando a divisão em categorias diferentes, a Figura 23 apresenta a classificação entre as unidades de processamento utilizadas no processo de telemetria veicular, as arquiteturas de *hardware*. De maneira simples, as arquiteturas utilizadas, representadas na Figura 26, podem ser resumidas ao uso do *smartphone* conectado com *dongle* OBD-II e uso de uma ECU baseada em microcontrolador ou *Single Board Computer* (SBC). Considerando o objetivo do processo de captura de dados, a Figura 24 apresenta os setores de interesse da aplicação da telemetria veicular. Os elementos de interesse nas pesquisas

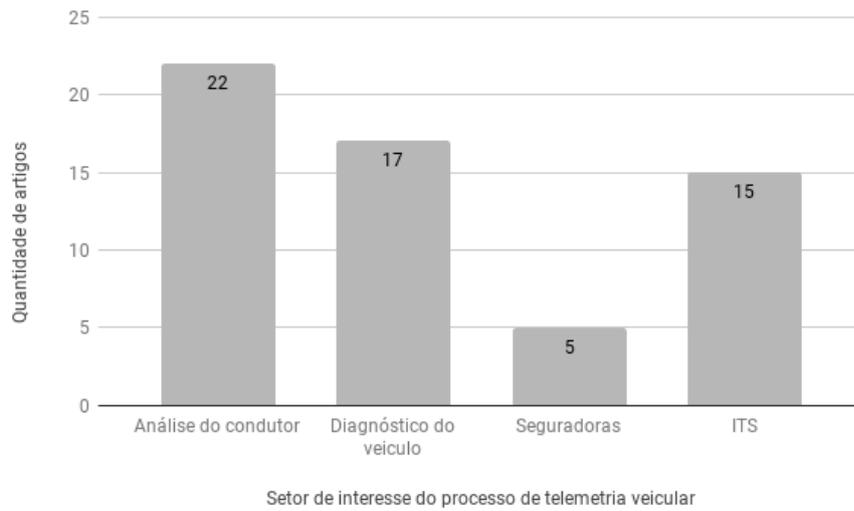


Figura 24 – Número de trabalhos classificados em cada setor de interesse da telemetria veicular

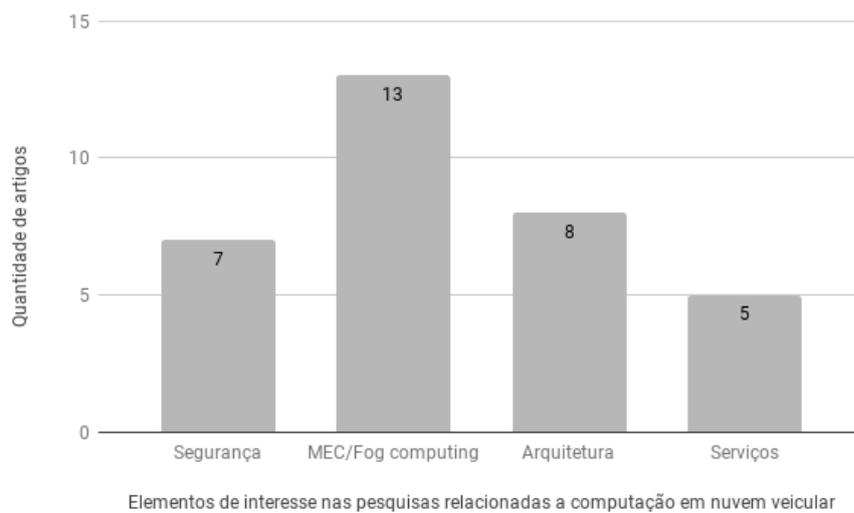


Figura 25 – Número de trabalhos classificados em cada tema de pesquisa da nuvem veicular

relacionadas a computação em nuvem veicular são detalhados na Figura 25. A Figura 26 mostra o resumo das arquiteturas.

Outra divisão importante é com relação aos tipos de sensores utilizados em cada setor de aplicação da telemetria veicular. Considerando a possibilidade de uso de sensores externos e integração de sensores internos dos veículos através da informação disponível na rede intraveicular, uma possível divisão pode ser feita. Assim, a Figura 27 mostra a quantidade trabalhos e os seus respectivos tipos de sensores utilizados em cada setor da telemetria veicular.

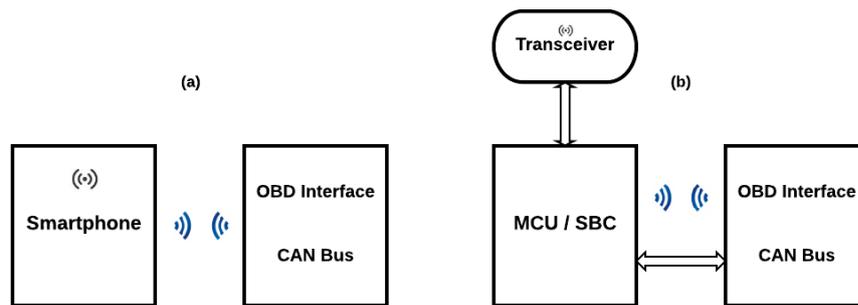


Figura 26 – Sistemas de Aquisição de Dados: a) baseado em *smartphone* b) combinado com SBC, transceptor 3G / 4G e *dongle* OBD

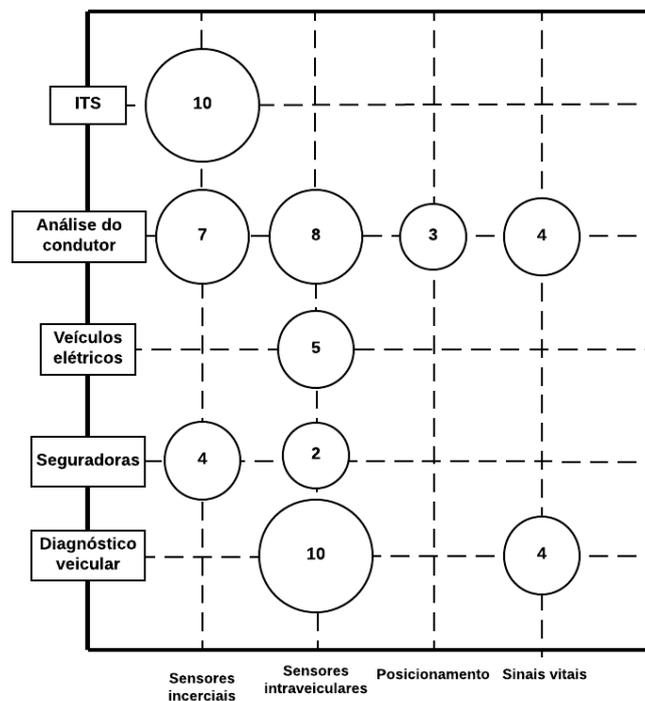


Figura 27 – Divisão dos sensores utilizados por aplicação final do sistema de telemetria veicular

Como resultado do mapeamento, a maioria dos artigos apresentam propostas de sistemas de aquisição de dados. Entre as diferentes soluções propostas, (XIE et al., 2017) propõe um sistema de aquisição de dados com arquitetura baseada em um microcontrolador e captura de dados através do padrão OBD-II, para disseminar os veículos conectados e com funcionalidades inteligentes para a chamada *Internet-of-Vehicles* (IoV) (BARCELOS et al., 2014) desenvolve um dispositivo de aquisição de dados e comunicação capaz de operar no padrão IEEE 802.11p e também com suporte a comunicação 3G. De maneira similar, mas com a aplicação da telemetria veicular definida, (SIEGEL et al., 2014) propõe um sistema de aquisição de dados para caracterização do óleo do carro. A partir desta informação,

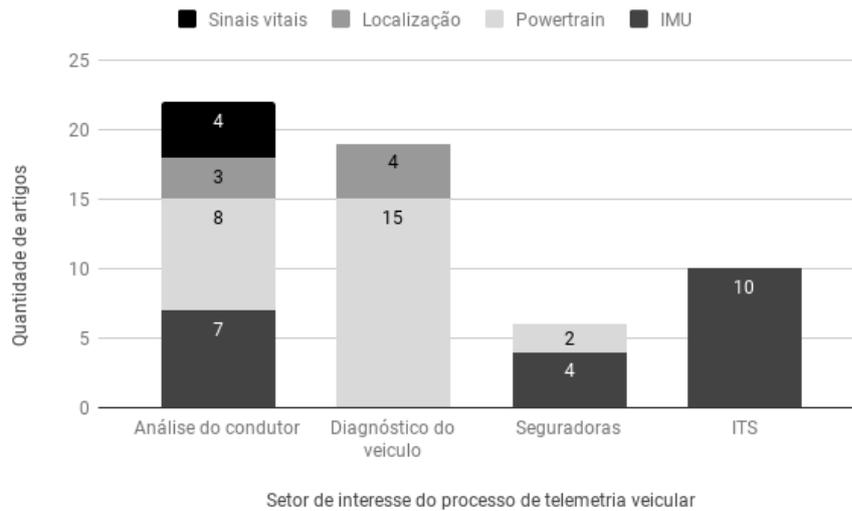


Figura 28 – Sensores utilizados nas diferentes aplicações da telemetria veicular

falhas e danos ao motor do carro podem ser evitados. (VASUDEVAN et al., 2017) cria uma base de dados a partir de sistema de aquisição de dados para criar um sistema de classificação do estado de sonolência do motorista. Seguindo a proposta da Criação de uma base de dados, (HUSNI et al., 2016) integra as informações coletadas através da plataforma de nuvem *IBM Bluemix*, habilitando assim uma série de serviços para carros conectados.

A partir das propostas de sistemas de telemetria veicular mais relevantes, a Tabela 1 apresenta uma comparação das contribuições dos trabalhos citados nesta seção.

Tabela 2 – Comparação das contribuições entre trabalhos que abordam telemetria veicular

Contribuição	Xie	Siegel	Vasudevan	Barcelos	Husni	Este trabalho
Proposta de sistema de aquisição de dados	X			X	X	X
Uso de aprendizado de máquina			X			X
Criação de base de dados	X	X	X		X	X

3.2 DETECÇÃO DE ANOMALIAS EM REDES INTRAVEICULARES

De forma diferente ao tema telemetria veicular, a detecção de anomalias em redes veiculares torna-se mais específica, e assim não se faz necessário um processo de busca de referências por meio de um mapeamento. Assim, de forma a complementar os trabalhos que abordam sistemas de telemetria veicular, também foram pesquisadas referências relacionadas à detecção de anomalias em redes intraveiculares.

Ao considerar este tema de pesquisa, o surgimento do carro conectado contribuiu para o aumento de trabalhos que abordam o processo de detecção de anomalias veiculares

(RAJBAHADUR et al., 2018). Os resultados da busca por trabalhos indicam anomalias nas redes intraveiculares como consequência da alteração da integridade da mensagem no barramento intraveicular ou por falha no sistema responsável pelo conteúdo da mensagem.

Relacionados ao setor de segurança, a preocupação com ataques e troca de mensagens não autorizadas na rede interna do veículo é abordada nos trabalhos de (WANG et al., 2018), (MüTER; GROLL; FREILING, 2010), (SARGOLZAEI et al., 2016) e (MARCHETTI et al., 2016).

De maneira similar aos trabalhos que objetivam a detecção de mensagens suspeitas na rede intraveicular, o desenvolvimento e aplicação de ferramentas de análise de dados na detecção de anomalias em parâmetros da rede veicular também torna-se possível. Assim, trabalhos relacionados também apresentam propostas de detecção de anomalias de variáveis específicas do veículo, como nível da bateria, funcionamento do motor e sistemas de resfriamento.

No contexto de detecção de anomalias em parâmetros da rede veicular, torna-se importante definir o que é considerado uma anomalia da rede intraveicular. Seguindo a definição de (CHANDOLA; BANERJEE; KUMAR, 2009), também utilizada em (THEISLER, 2013), uma anomalia é caracterizada como consequência de um erro, indicando assim a possível presença de uma falha no funcionamento do sensor ou sistema do veículo. Considerando uma informação que é uma variável de um sensor, o dado anômalo é um valor fora de um intervalo esperado. É importante observar que um erro sempre caracteriza uma anomalia, mas o contrario não é verdadeiro.

Ao considerar a detecção de anomalias em sinais veiculares como um problema de classificação, (THEISLER; DEAR, 2012) apresenta uma série de trabalhos. O primeiro trabalho, propõe uma máquina de aprendizado para detecção de anomalias de um motor DC. A metodologia é validada a partir de base de dados criada a partir de um equipamento teste de motores DC.

Como continuação do trabalho, (THEISLER; DEAR, 2013b) propõe uma abordagem para encontrar o conjunto ideal de parâmetros para a máquina de aprendizado baseada na técnica *Support Vector Data Description* (SVDD). Como sequência da pesquisa, (THEISLER; DEAR, 2013a), usa máquina de aprendizado para auxiliar especialistas na área veicular para detecção anomalias. A última contribuição, (THEISLER, 2017), apresenta uma abordagem com detecção de anomalias conhecidas e desconhecidas, além de não requerer configuração de parâmetros prévios. Diferentemente dos trabalhos anteriores, é utilizado uma combinação de classificadores.

Realizando o processo de coleta de dados veiculares durante três anos, a partir de diferentes modelos de veículos, (PRYTZ et al., 2015) apresenta um modelo de previsão para a manutenção de compressores veiculares. A proposta é baseada no uso do *Random Forest* como algoritmo de classificação dos dados, além disso, dois métodos para seleção de atributos são usados e os respectivos resultados comparados a ação de especialista humano.

O pré-diagnóstico de problemas veiculares também foi abordado por (SUWATTHIKUL; MCMURRAN; JONES, 2011), onde a aplicação de um Sistema de Inferência *Fuzzy* Adaptável em Rede (ANFIS) para o pré-diagnóstico de problemas que possivelmente poderiam causar falhas no veículo é apresentado. Na proposta de (SVENSSON; BYTTNER; ROGNVALDSSON, 2008), a técnica *Self-organizing map* (SOM) é utilizada para detecção de diferentes níveis do radiador veicular entupido, a partir de uma análise no sistema refrigeração do motor.

Tabela 3 – Comparação das contribuições entre trabalhos que abordam detecção de anomalias veiculares

Trabalho	Detecção em tempo real	Variáveis usadas para análise de anomalia	Técnica para detecção da anomalia
THEISSLER; DEAR, 2012	Não	Velocidade, Corrente, saída PWM	SVDD
THEISSLER; DEAR, 2013b	Não	Diferentes base dados	SVDD com kernel RBF
THEISSLER; DEAR, 2013a	Não	STFT, RPM, velocidade, posição do corpo de borboleta e tempo de injeção	SVDDsubseq: nova proposta baseada em séries temporais
THEISSLER, 2017	Não	STFT e ECT	Combinação de classificadores: SVM, Random Forest, Extreme value, OC-SVM, SVDD, MOG
ZHANG et al., 2017	Sim	Velocidade, RPM, número da marcha e troca de faixas	Gráficos de estado
NARAYANAN; MITTAL; JOSHI, 2016	Sim	RPM	Hidden Markov Model
NIRMALI et al., 2017	Sim	RPM, velocidade e sensores IMU	Clusterização k-means
NAIR, V. V.; KOUSTUBH, 2017	Não	Velocidade, velocidade do motor elétrico, Torque e Status KL	SVDD combinado com k-means
WEBER et al., 2018	Não	-	iForest, LODA

Outros artigos também abordam a detecção de anomalias a partir de sinais intraveiculares. (ZHANG et al., 2017) apresenta uma abordagem para reconhecimento em tempo real de anomalias relacionadas ao comportamento do motorista baseado em gráficos de estado. O sistema proposto, denominado de *SafeDrive*, utiliza informações baseadas na velocidade, RPM, desvio de rota e mudança de marcha. Baseada na metodologia proposta por (THEISSLER, 2017), (NAIR; KOUSTUBH, 2017) explora máquinas de aprendizado aplicadas a detecção de anomalias em veículos híbridos e elétricos. Outra contribuição que também relacionada a detecção de anomalia no comportamento do condutor, utilizando a clusterização k-Means é apresentada por (NIRMALI et al., 2017).

Outra abordagem, a partir da captura de dados através do padrão OBD-II, (NARAYANAN; MITTAL; JOSHI, 2016) utiliza modelos ocultos de Markov para detectar injeção de mensagens não autorizadas e anomalias nos sinais. Os resultados evidenciam que tal técnica pode ser usada com sucesso para detectar anomalias nos veículos, com análise baseada em um ou dois parâmetros. Apesar de ser uma interface de baixo custo e presente em todos os carros, o padrão OBD-II não fornece todas informações disponíveis no barramento CAN. Dessa maneira, (WEBER et al., 2018) apresenta um sistema de detecção de anomalias híbridas para as ECUs do barramento CAN. A proposta combina máquina de aprendizado

baseado em OC-SVM com a técnica *Lightweight On-Line Detector of Anomalies* (LODA).

A fim de detalhar os trabalhos relacionados a detecção de anomalias nos sinais intra-veiculares citados nesta seção, a Tabela 3 apresenta uma comparação das contribuições presentes nos trabalhos mais relevantes.

4 SISTEMA PROPOSTO

Neste capítulo é apresentado o sistema de telemetria para detecção de anomalia em sinal intraveicular proposto. A Seção 4.1 apresenta uma visão geral do sistema, considerando sua arquitetura em diagrama de blocos. A descrição do processo de aquisição de dados, bem como os cenários considerados para a construção da base de dados, são apresentados na seção 4.2. Considerando a presença de anomalias nos sinais intraveiculares na etapa de teste do sistema, a Seção 4.3 apresenta o procedimento para a inserção de anomalias. Na Seção 4.4 detalhes do sistema de detecção de anomalias são apresentados.

4.1 VISÃO GERAL

A proposta deste trabalho é apresentar um sistema de telemetria para a detecção de anomalias em sinal intraveicular, o sinal do sensor ECT. Assim, torna-se importante definir: o sistema de aquisição de dados e a arquitetura do detector de anomalias, detalhando suas etapas.

A primeira etapa, a responsável pela captura das informações, é o sistema embarcado que deve ser integrado ao veículo. Ao considerar os diferentes atributos de qualidade para a definição do sistema, a tecnologia de comunicação e custo da unidade de processamento devem ser analisadas. Como resultado do processo de captura de dados, as diferentes informações coletadas pelo dispositivo devem ser armazenadas. Esta pode ser considerada a etapa intermediária, onde os dados serão armazenadas para posterior análise. Com o objetivo definido, a detecção de anomalias, outra parte essencial do sistema é a aplicação da técnica de detecção de anomalia, o sistema de detecção, a partir dos dados disponíveis.

A Figura 29 ilustra a visão geral do sistema, apresentada com três blocos principais: o sistema de aquisição de dados veicular, os dados obtidos e o módulo responsável pela detecção de anomalias.

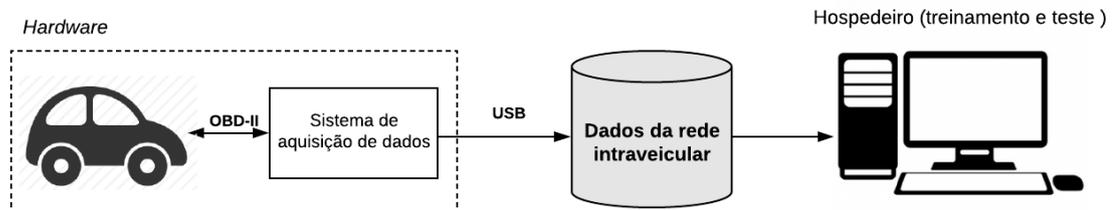


Figura 29 – Visão geral do sistema proposto

As próximas seções deste capítulo detalham as principais etapas do módulo responsável pela captura de dados, a forma como as anomalias são inseridas e o sistema de detecção

de anomalias.

4.2 AQUISIÇÃO DE DADOS DA REDE INTRAVEICULAR

Um dos maiores desafios no desenvolvimento de um trabalho que envolve o uso de dados de redes intraveiculares é a utilização de uma base de dados apropriada. A maioria das informações geradas pela rede CAN de qualquer veículo são confidenciais, mesmo considerando o uso de equipamentos específicos para leitura do barramento a partir do mapa de mensagens do veículo. Diante deste cenário, o uso da interface OBD-II torna-se uma opção limitada, porém viável, para captura de dados relacionados aos diferentes sistemas do veículo.

Assim, os dados utilizados neste trabalho foram obtidos a partir de um sistema embarcado conectado à interface OBD-II de um único veículo, dirigido pelo mesmo condutor, durante um período de 6 meses (entre Junho de 2018 e Novembro de 2018). O modelo do veículo utilizado foi o **Toyota Etios 2014 1.5 16v Flex**. Diferente da maioria das arquiteturas propostas pela literatura (WAHLSTRÖM; SKOG; HÄNDEL, 2017), o sistema utilizado na execução deste trabalho não necessita de um *smartphone* integrado ao conector OBD-II.

A arquitetura proposta, o sistema embarcado para captura de dados veiculares, é composto por um microcontrolador conectado a um *transceiver* CAN (TJA1049T), além de módulo de comunicação 3G para o envio das informações intraveiculares. A placa de desenvolvimento utilizada para implementação do sistema foi a *Particle Electron* em conjunto com o adaptador *Carloop*, que contém o conector OBD-II com *transceiver* CAN integrado, ilustrados na Figura 30. Além de uma configuração específica para a coleta de dados da OBD-II, a *Carloop* oferece parte do *firmware open-source* para a coleta de informações da OBD.



Figura 30 – Placa de desenvolvimento *Particle Electron* com a plataforma *Carloop* integrada

Fonte: (CARLOOP.IO, 2018)

Apesar da montagem do sistema de aquisição baseado em uma placa de desenvolvimento, o sistema embarcado possui algumas funções interessantes como a programação

de *firmware* embarcado através da rede 3G e integração direta com diferentes plataformas de nuvem. As especificações técnicas do módulo embarcado são:

- **Unidade de processamento:** microcontrolador STM32F205 (ARM Cortex M3) com 120MHz;
- **Memória:** 1MB *flash*, 128KB RAM;
- **Interfaces:** 30 pinos de E/S, SPI, UART, I2C e CAN RX/TX conectado ao transceiver TJA1049;
- **Outras características:** compatibilidade com sistema operacional de tempo real (RTOS) e certificação FCC.

Outra apresentação do sistema, em termos de suas funcionalidades, pode ser feita a partir da arquitetura de *hardware*, ilustrada na Figura 31. Com o objetivo de adquirir o maior volume possível de dados, o sistema de aquisição deve ser utilizado em diferentes cenários de uso do veículo.

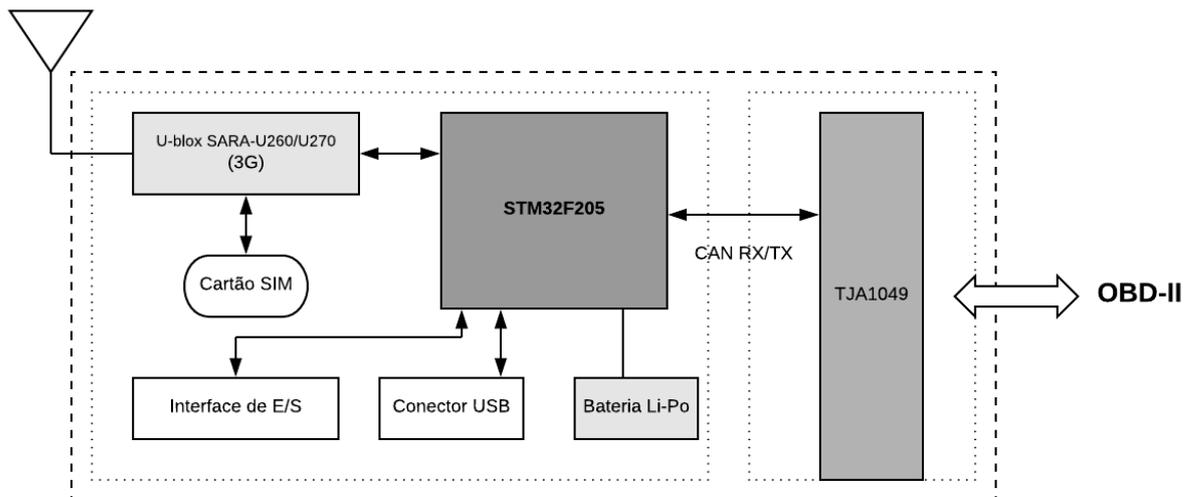


Figura 31 – Arquitetura de *hardware* do sistema embarcado utilizado

4.2.1 *Firmware* Embarcado

Como citado anteriormente, dois *firmwares* específicos foram utilizados ao longo deste trabalho. Como a placa de desenvolvimento *Particle Electron* e *Carloop* são partes de projetos *open-source*, uma parte funcional do *firmware* disponibilizado pelo *Carloop* pôde ser utilizado.

A primeira parte da implementação com o sistema embarcado utilizou um *firmware* específico para descobrir o padrão da rede intraveicular do veículo utilizado, representado

pela Figura 32 (a). O *firmware* propriamente utilizado para a construção da base de dados, representado na Figura 32 (b), utiliza as informações retornadas dos PIDs disponíveis no carro para sua consequente requisição de informação.

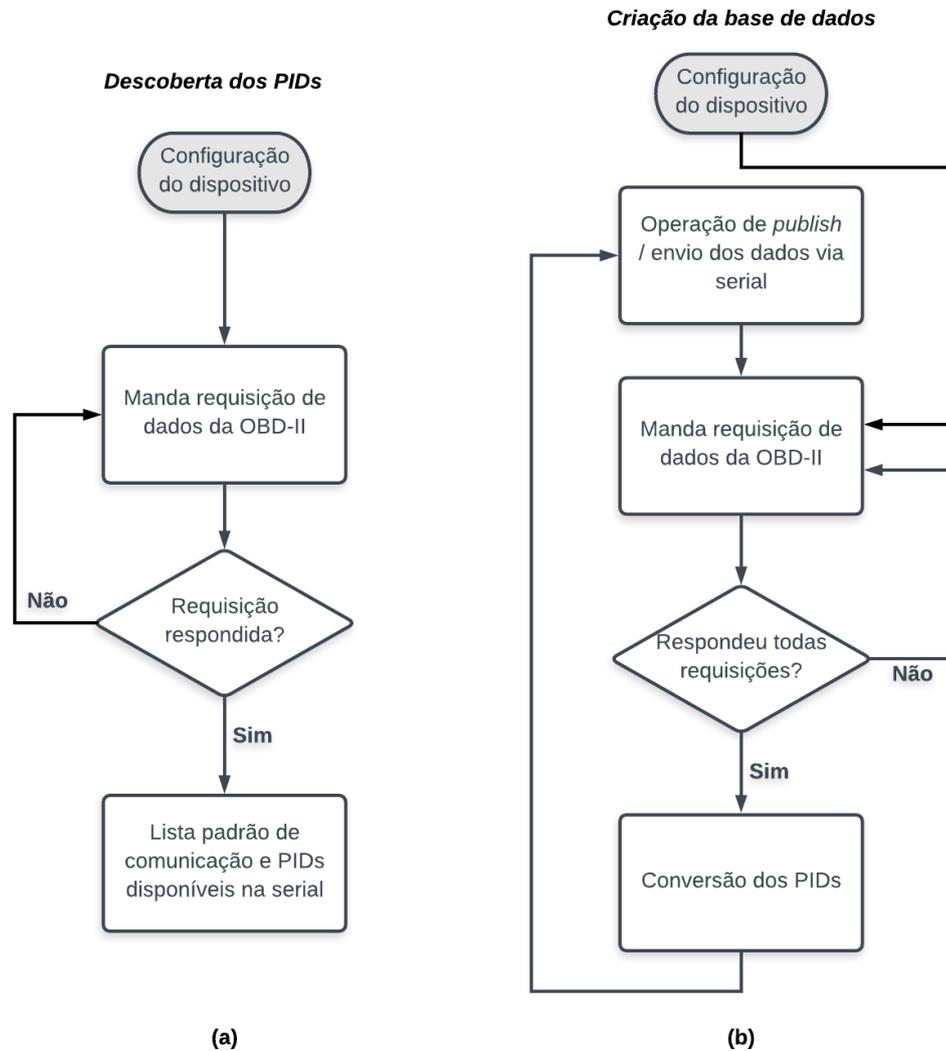


Figura 32 – Fluxograma para os *firmwares* utilizados para o desenvolvimento do trabalho (a) detecção de padrão intraveicular (b) para captura dos dados

4.3 BASE DE DADOS COM SINAIS INTRAVEICULARES

O sistema proposto na seção anterior foi utilizado para coletar as informações da interface OBD-II do veículo Toyota Etios 2014. Como descrito na Seção 2.2.4, a interface utilizada possui diferentes padrões de barramentos, definindo diferentes velocidades de transmissão de dados e protocolo de comunicação. Com o objetivo de determinar o padrão, foi utilizado um *firmware* específico para descobrir as informações disponíveis na OBD-II do veículo, bem como seu padrão.

Como resultado, o veículo escolhido possui sua interface configurada com a **ISO 15765-2**, correspondendo ao uso da rede CAN com identificador de 11-bit e velocidade de transmissão de 500 kbit/s. A Tabela 4 apresenta as informações disponíveis, as variáveis da base de dados, bem como seu código PID (hexadecimal), requisitado pelo microcontrolador. Além dos parâmetros intraveiculares obtidos, torna-se importante definir os modos de operação veicular registrados. Seguindo a nomenclatura de (THEISSLER, 2013), têm-se:

- **Veículo em modo *Idle***: corresponde ao estado em que o veículo está parado, porém com o motor ligado.
- **Veículo em movimento**: cenário de uso do veículo em trajetória, seja em vias urbanas ou vias expressas.

Diferente do modo *Idle*, o uso do veículo em movimento considerado neste trabalho possui quatro cenários distintos, com diferentes tipos de trajetória do veículo: trajetória urbana, trajetória em vias expressas, trajetória em campus de universidade e trajetória fixa envolvendo área urbana e via expressa. Todos os dados coletados em ambos modos, e considerando também todos os cenários, foram obtidos a partir de uma taxa de amostragem de 1Hz. A fim de separar as medições, a cada percurso realizado é definido um ciclo de condução. Os ciclos do modo *idle* são considerados cada vez em que o carro é ligado no respectivo modo. A Tabela 5 detalha a composição da base dados construída ao longo de seis meses, com o número de ciclos em cada modo, cenário considerado e tamanho.

Além da comunicação por meio da tecnologia 3G, o sistema de aquisição também apresenta uma saída USB para coleta das informações. Durante a maior parte da construção da base de dados deste trabalho foi usada a saída USB. O motivo para a escolha da opção cabeada é que a placa de desenvolvimento *Particle Electron* possui um serviço de nuvem associado ao seu uso, onde a cobrança é feita por tamanho de dados enviados e armazenados em modelo *SaaS (storage-as-a-service)*.

Diferente da operação utilizando comunicação sem fio, o processo de coleta de dados a partir da porta USB exige o uso de um computador para processamento e armazenamento das informações. Foi escrito um código na linguagem de programação Python para salvar os 27 parâmetros coletados a cada segundo da interface OBD-II. A Figura 33 ilustra o diagrama para o processo de montagem da base com o serviço *Particle publish* e o com o uso de um computador.

Entre os trajetos do veículo em movimento, o que pode caracterizar de maneira mais precisa o seu funcionamento em uma determinada faixa de operação do motor é o trajeto em campus de universidade. Ilustrado na Figura 34, a caracterização dessa base de dados foi realizada no campus da Universidade Federal de Pernambuco (UFPE), onde o limite de velocidade não passa de 40 *km/h*. Assim, os dados provenientes deste tipo de percurso descrevem o funcionamento do veículo em baixas velocidades, diferente do que ocorre em parte em parte dos dados obtidos de trajetórias em vias expressas.

Tabela 4 – Descrição dos parâmetros coletados a partir da OBD-II

Parâmetro	Código PID	Descrição	Tamanho (bytes)
<i>Engine Run Time</i>	0x1F	Tempo (s) desde que o motor foi ligado	2
<i>Engine RPM</i>	0x0C	Rotações por minuto do motor	2
<i>Vehicle Speed</i>	0x0D	Velocidade (km/h)	1
<i>Throttle</i>	0x11	Porcentagem de abertura do corpo de borboleta	1
<i>Engine Load</i>	0x04	Porcentagem da carga do motor	1
<i>Coolant Temperature</i>	0x05	Temperatura do líquido de arrefecimento do motor	1
<i>Long Term Fuel Trim Bank 1</i>	0x07	Fator de correção usado pelo sistema de controle de combustível nos modos de operação de malha aberta e fechada	2
<i>Short Term Fuel Trim Bank 1</i>	0x06	Fator de correção usado pela PCM para manter uma mistura de combustível equilibrada	2
<i>Intake Manifold Pressure</i>	0x0B	Pressão do coletor de admissão	1
<i>Fuel Tank</i>	0x2F	Nível do tanque de combustível	1
<i>Absolute Throttle B</i>	0x47	Posição atual da abertura do corpo de borboleta	1
<i>Pedal Position D</i>	0x49	Posição D do acelerador	1
<i>Pedal Position E</i>	0x4A	Posição E do acelerador	1
<i>Commanded Throttle Actuator</i>	0x4C	Valor de abertura do corpo de borboleta, 0% para fechado e 100% para totalmente aberto	1
<i>Fuel Air Equivalence Ratio Absolute</i>	0x44	Proporção combustível / ar.	2
<i>Barometric Pessure Relative</i>	0x33	Pressão atmosférica	1
<i>Throttle Position Intake</i>	0x45	Posição relativa da abertura do corpo de borboleta	1
<i>Air Temperature</i>	0x0F	Temperatura do ar no coletor de admissão	1
<i>Timing Advance</i>	0x0E	Graus de avanço da ignição	1
<i>Catalyst Temperature Bank 1 Sensor 1</i>	0x3C	Temperatura dentro do conversor catalítico.	2
<i>Catalyst Temperature Bank 1 Sensor 2</i>	0x3E	Temperatura dentro do conversor catalítico.	2
<i>Control Module Voltage</i>	0x42	Nível tensão da bateria	2
<i>Commanded Evaporative Purge</i>	0x2E	Porcentagem normalizada do escape (sistemas de limpeza)	1
<i>Time Run with MIL on</i>	0x4D	Tempo (s) desde que o carro está com MIL ativo	2
<i>Time Since DTC Cleared</i>	0x4E	Tempo (s) desde que os DTCs foram zerados	4
<i>Distance Traveled with MIL on</i>	0x21	Distância viajada (km) com luz da MIL ligada	2
<i>Warm Ups Since codes cleared</i>	0x30	Número de vezes que o carro foi ligado desde que os DTCs foram zerados	1

Tabela 5 – Descrição da de base de dados construída a partir dos parâmetros da OBD-II

Modo de operação	Trajeto	Duração total (s)	Distância percorrida	Número de ciclos (<i>trips</i>)
<i>Idle</i>	-	35309	-	47
Movimento	Áreas urbanas	29958	717 km	12
Movimento	Via expressa	19493	458 km	13
Movimento	Universidade	15873	387 km	18
Movimento	Fixo	12561	1104 km	37

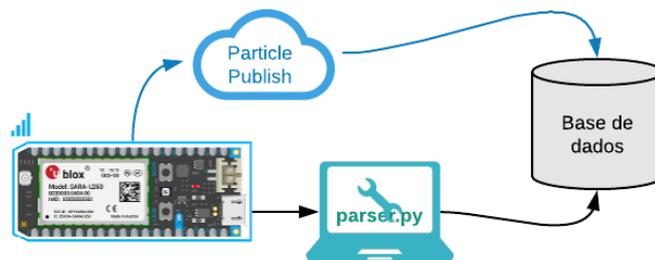


Figura 33 – Diagrama das possibilidades para montagem da base de dados a partir do sistema de aquisição

Fonte: adaptado de (PARTICLE, 2018)

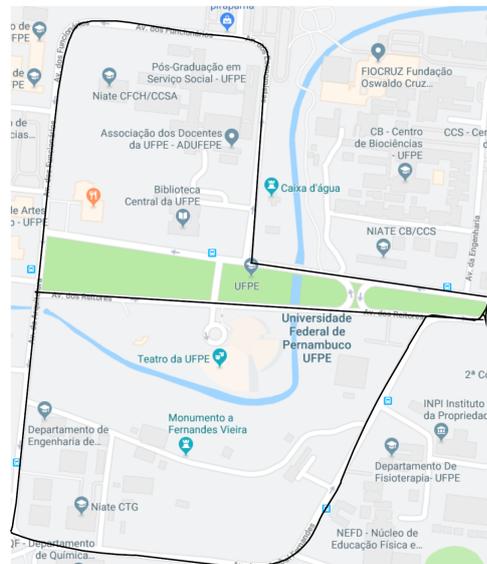


Figura 34 – Mapa do trajeto do veículo no campus da UFPE

Como forma de ilustrar os dados capturados, e a respectiva base construída, a Figura 35 ilustra sete parâmetros da rede intraveicular capturados durante um ciclo de 2157 segundos em trajeto de via expressa.

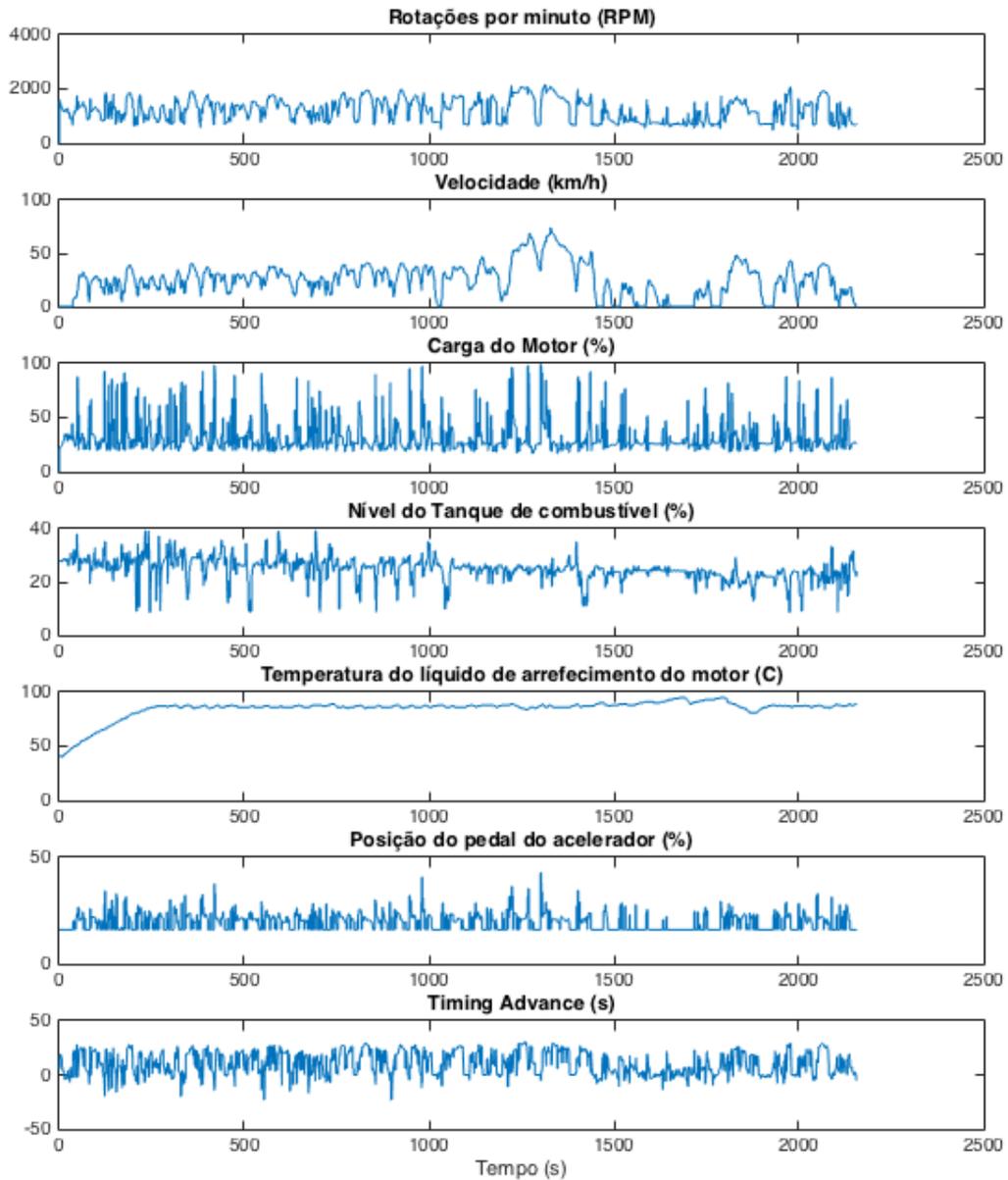


Figura 35 – Representação de ciclo com informações intraveiculares capturadas pelo dispositivo

4.4 PROCESSO DE INSERÇÃO DE ANOMALIAS EM SINAIS VEICULARES

Outro desafio, além da aquisição de dados provenientes da rede intraveicular, é a captura de dados que possuam algum tipo de anomalia. Como definido na Seção 3.2, um dos tipos de anomalias a ser abordada neste trabalho deriva da possível falha de sensor veicular. Baseado nas informações veiculares coletadas e da relevância do sensor para o funcionamento do veículo, é então realizado processo de detecção de anomalias no sensor

do líquido de arrefecimento.

O sensor de temperatura do líquido de arrefecimento do motor, normalmente conhecido como sensor ECT, pelo seu nome em inglês *Engine Coolant Temperature* (ECT), é usado para medir a temperatura da mistura do líquido de arrefecimento e anticongelante no sistema de arrefecimento do veículo (PRESTONE, 2018). Sua principal função é indicar quanto calor o motor está emitindo. Em geral a informação retornada por este sensor fica visível no painel do motorista.

Uma característica do sensor ECT que é comum a outros sensores veiculares é o fato do mesmo ser baseado em elementos resistivos como fonte da conversão da grandeza física em tensão analógica para posterior conversão analógica digital. Com o desafio de obter um comportamento anômalo destes sensores, e respectivos sistemas veiculares, durante o processo de captura de dados da rede veicular, é proposta uma técnica de inserção de anomalias baseada na inserção de um tipo de ruído característico a possível falha resistiva.

Apesar de relacionadas, o processo de detecção de anomalias é distinto da remoção de ruídos em sinais (CHANDOLA; BANERJEE; KUMAR, 2009). O ruído pode ser definido como um fenômeno não desejado no sinal de interesse, com um valor instantâneo impreciso. Diferentes tipos de ruído são associados na análise de circuitos eletrônicos. O tipo de ruído associado em uma possível falha resistiva é o ruído em excesso (*flicker noise*).

O ruído em excesso, também chamado de ruído rosa, é encontrado em todos os componentes eletrônicos ativos, bem como alguns elementos passivos como resistores de carbono (SANTOS, 2011). A origem física deste tipo de ruído em um circuito eletrônico, seja considerando um sinal de tensão ou corrente, origina de armadilhas associadas a contaminação do retículo cristalino do dispositivo. Uma característica do ruído em excesso, é que o comportamento de sua densidade espectral fica inferior ao do ruído branco. Assim, a possível falha no sensor, corresponde diretamente na presença de um ruído em excesso com alta potência no sinal considerado.

A partir do comportamento aleatório do ruído, a fundamentação matemática mais adequada para análise de seu em circuitos eletrônicos se baseia na teoria de probabilidade. Considerando um sinal eletrônico, e sua notação, (4.1) mostra a representação utilizada ao considerar a presença de ruído em um sinal.

$$V_v = V_V + v_v + v_{ruído} \quad (4.1)$$

Como abordado por (SANTOS, 2011) e (GRAY et al., 2008), torna-se importante estudar a influência do ruído em circuito eletrônico separadamente e posteriormente considerar sua influencia no sinal de interesse final a partir da relação sinal-ruído (SNR). Embora o ruído possua um valor médio nulo, o valor esperado do ruído elevado ao quadrado não é nulo. Assim, torna-se possível analisar a influência do ruído em um circuito eletrônico a partir de uma análise baseada na densidade espectral.

Considerando o elemento eletrônico que é a fonte da medição do sinal de interesse nos circuitos dos sensores veicular, o resistor, (4.2) caracteriza a influência do ruído em excesso. Este tipo em específico, está sempre associado a um fluxo de corrente contínua.

$$\langle i_{\text{ruído}}^2 \rangle = K_1 \frac{I_{CC}}{f^b} \Delta f \quad (4.2)$$

Na equação I_{CC} é a corrente elétrica média, proporcional ao número médio de elétrons que chegam a um ponto do condutor com velocidade específica, K_1 é uma constante do resistor e b define o tipo de ruído ($b = 1$ define ruído rosa e $b = 2$ ruído vermelho). Após a avaliação das fontes de ruído em um circuito eletrônico, pode-se realizar sua análise baseada no teorema da superposição.

Como forma de detalhar a inserção de ruído no sinal do sensor do líquido de arrefecimento e sua consequente caracterização como anomalia, a próxima seção descreve a criação de sinais anômalos do sensor de temperatura do líquido de arrefecimento do motor a partir da inserção de ruído em excesso.

4.4.1 Anomalias no sensor de temperatura do líquido de arrefecimento do motor (ECT)

O sensor de temperatura do líquido de arrefecimento do motor é um dos sensores mais importantes do veículo. Com a função de monitorar a troca de calor no motor do veículo, seu valor pode indicar possíveis estados e mau funcionamento de algum subsistema do motor veicular. Assim, dois tipos diferentes de anomalias são associadas a este sensor e avaliadas neste trabalho.

1. **Anomalia Tipo I (valor não disponível):** esta anomalia é derivada da desconexão do cabo que interliga o sensor e a ECU responsável pelo seu monitoramento. Em geral, ao acontecer a desconexão do cabo a temperatura do sensor é fixada em um valor fundo de escala, na Figura 36 representado pelo valor nulo. A depender do veículo, esta pode ser fixada em $0^\circ C$ ou $-40^\circ C$. Este tipo de anomalia pode ser considerada uma anomalia coletiva, caso a desconexão do cabo for definitiva. Nos casos em que a desconexão do cabo aconteça de forma momentânea, a anomalia Tipo I pode ser considerada contextual.
2. **Anomalia Tipo II (falha resistiva):** este tipo de anomalia é caracterizada por um mau funcionamento do sensor veicular. Como descrito na seção anterior, esta falha é consequência da estrutura do sensor resistivo. A Figura 37 ilustra o circuito eletrônico do sensor ECT, representado pelo termistor (ou termistor), um resistor que possui sua resistência definida baseada na sensibilidade à temperatura. Assim, torna-se possível medir a temperatura do sistema de arrefecimento a partir de uma tensão analógica. Caso este resistor apresente um comportamento fora do esperado,

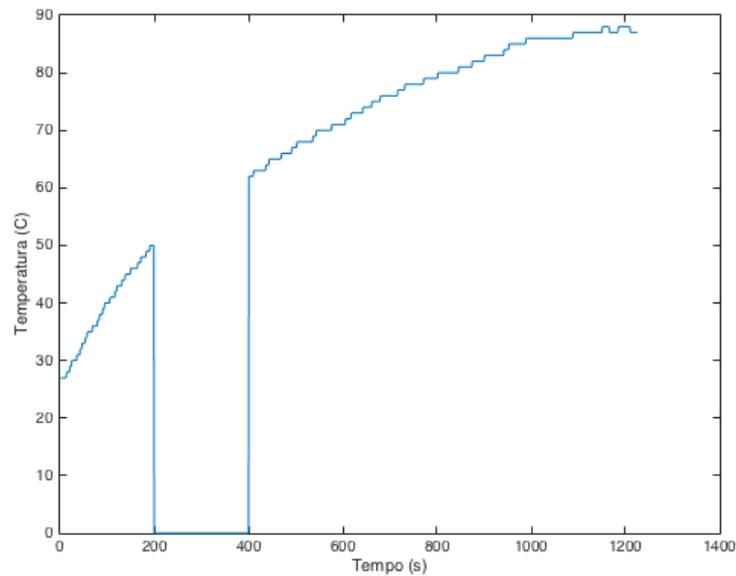


Figura 36 – Sinal ECT com anomalia Tipo I, desconexão do cabo do sensor com ECU

a consequente medição da temperatura do líquido de arrefecimento estará comprometida.

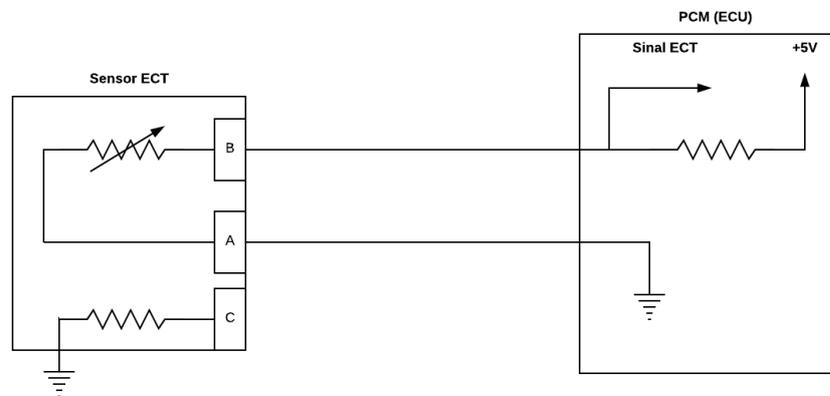


Figura 37 – Circuitos eletrônico do sensor de temperatura do líquido de arrefecimento do motor (ECT)

Apesar de os circuitos dos sensores veiculares possuírem proteção contra a presença de ruído nos circuitos de sensoriamento, como mostrado em (Exponent Inc., 2012), a presença de um sinal com alta relação sinal-ruído pode afetar diretamente o resultado da temperatura calculada pela ECU correspondente. Assim, foi feita uma análise no valor da potência do sinal considerado normal, o do sensor ECT, e o respectivo valor da relação sinal-ruído dos sinais considerado normais para a inserção do ruído em excesso. Foram definidos cinco valores distintos para a potência dos sinais de ruído adicionados ao sinal de ECT. A Figura 38 mostra os diferentes sinais

com os níveis de anomalia.

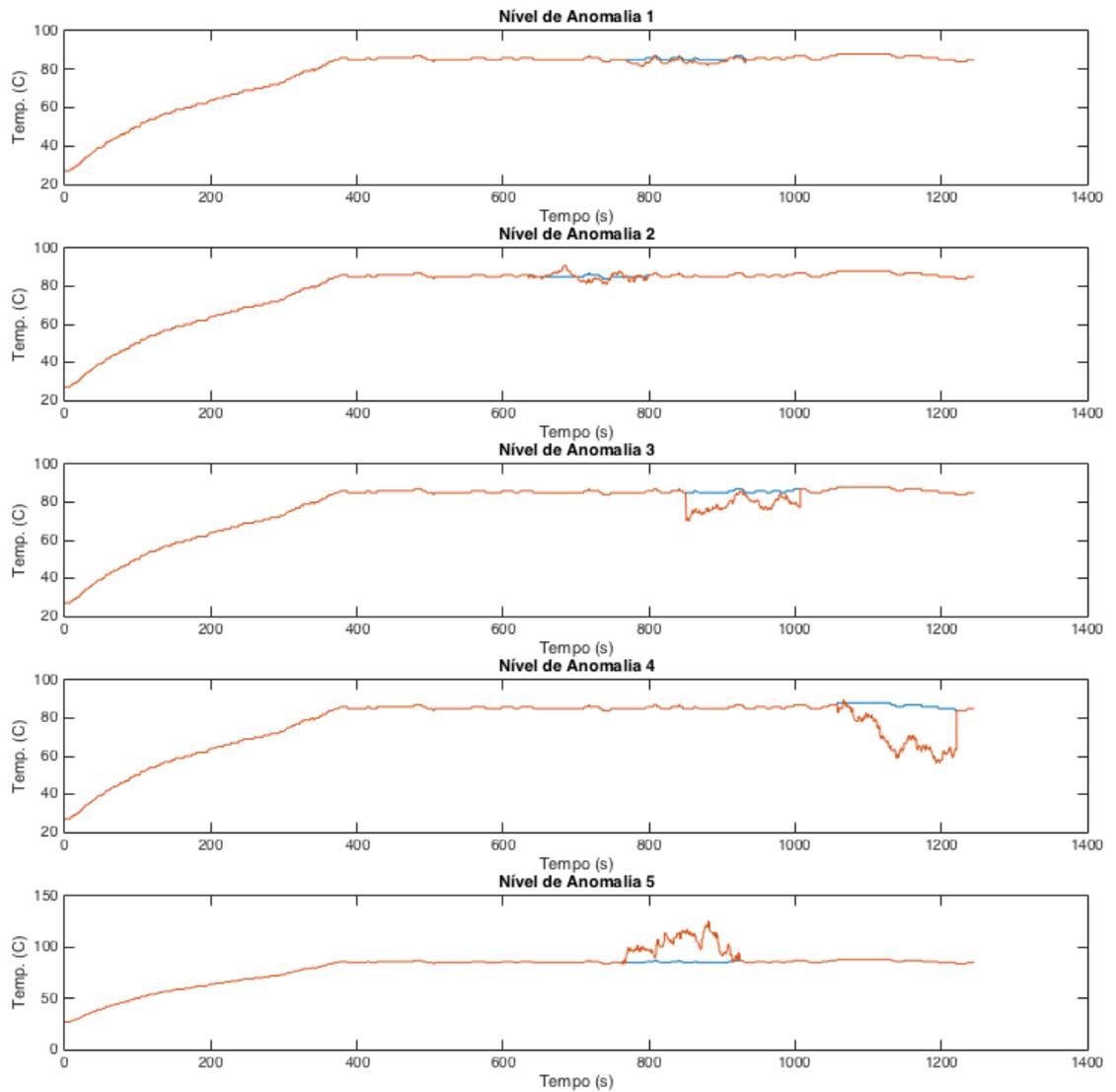


Figura 38 – Cinco diferentes níveis de anomalia aplicados ao sinal do sensor ECT

Pelo fato de as métricas de SNR possuírem uma faixa dinâmica, em razão dos valores da potência do sinal, os sinais são geralmente descritos usando a escala logarítmica de decibel. Como mostrado em (4.3), a relação sinal-ruído é o termo para a razão entre as potências de um sinal contendo algum tipo de informação e ruído.

$$SNR_{db} = 10 \cdot \log_{10} \left(\frac{P_{sinal}}{P_{ruído}} \right) \quad (4.3)$$

A relação sinal-ruído também pode ser medida em termos das amplitudes do sinal e ruído em questão, assim, considerando que o sinal e o ruído são medidos na mesma

impedância do circuito, o SNR pode ser calculado a partir do valor quadrático médio (RMS) das amplitudes:

$$SNR = \left(\frac{A_{\text{signal}}}{A_{\text{ruído}}}\right)^2 \quad (4.4)$$

Considerando a relação de potência entre os sinais provenientes do processo de coleta de dados intraveiculares, são definidos cinco diferentes níveis diferentes de potência para o sinal anômalo, chamados de níveis de anomalia, para serem analisados pelo sistema de detecção de anomalias.

- **Nível de Anomalia 1:** relação sinal-ruído de **18dB**.
- **Nível de Anomalia 2:** relação sinal-ruído de **13dB**.
- **Nível de Anomalia 3:** relação sinal-ruído de **8dB**.
- **Nível de Anomalia 4:** relação sinal-ruído de **3dB**.
- **Nível de Anomalia 5:** relação sinal-ruído de **0dB**. Este nível pode ser tratado uma situação extrema onde o mau funcionamento do sensor resistivo é tao grande que têm-se a mesma potência para o sinal do sensor e a fonte de ruído.

Os valores descritos acima foram baseados na faixa de potência dinâmica do sinal original proveniente do sensor ECT, ilustrados na Figura 39. No modo *idle* a potência do sinal também variou no mesmo intervalo, apesar de possuir alguns valores fora do intervalo. Assim, os níveis de anomalias foram considerados para ambos modos veiculares de maneira idêntica.

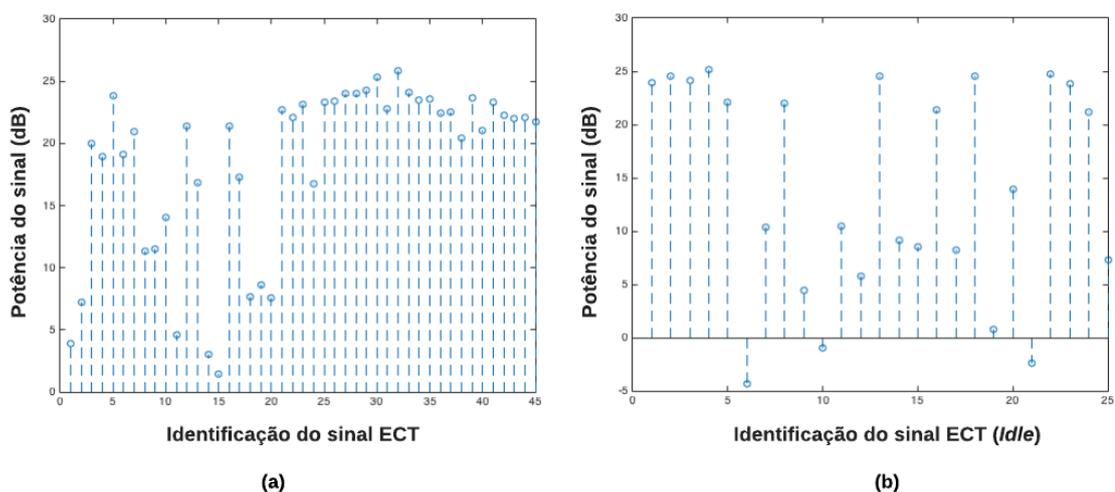


Figura 39 – Faixa dinâmica de potência dos sinais utilizados nos modos (a) carro em movimento (b) modo *idle*

Como forma de justificar a presença do ruído como anomalia no sinal do sensor ECT, uma comparação pode ser realizada com um sinal intraveicular que contém uma componente de ruído proveniente de seu sensor, o sinal do nível de tanque de combustível. A Figura 40 ilustra o sinal ao longo de um ciclo de direção. Apesar da imprecisão, a maioria dos veículos que ainda usam sensores de nível para cálculo da porcentagem gasolina no tanque usam algoritmo baseado valor inicial de partida e atualizações periódicas. Como pode ser visto na figura, o sinal possui valor inicial por volta de 81, alguns pontos anômalos indicam valor 55 ou 90 para o nível do tanque.

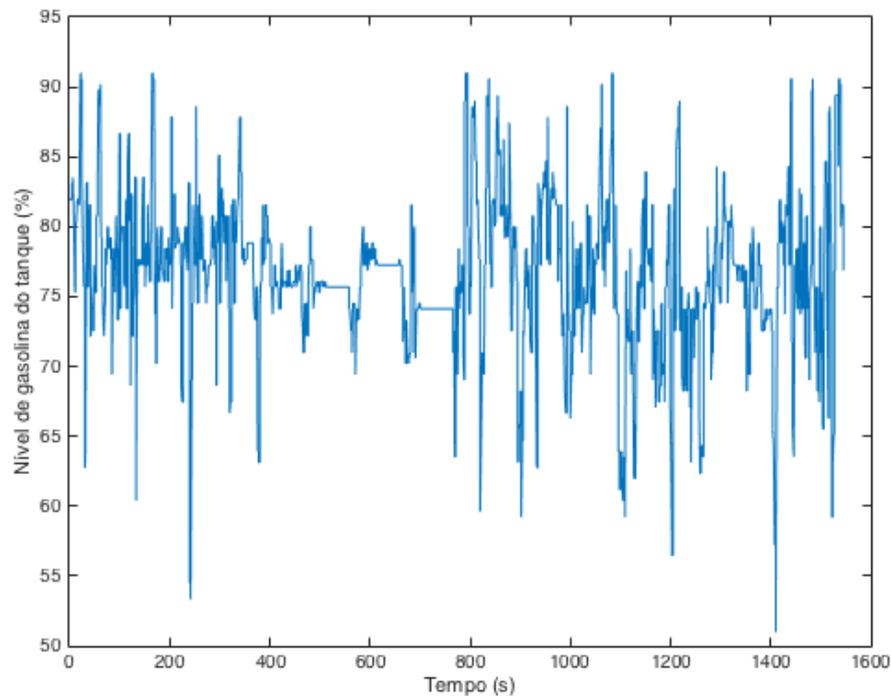


Figura 40 – Exemplo de sinal veicular com componente de ruído

Para o processo de teste do sistema de detecção de anomalias se faz necessária um conjunto contendo apenas dados anômalos. Assim, o módulo gerador de anomalias deste trabalho, para cada um dos tipos de anomalias, é definido como a aplicação da anomalia em 300 amostras subsequentes de maneira aleatória. Diferentes bases são criadas, correspondendo a cada tipo de anomalia. Além da divisão em tipos, as anomalias do Tipo II são separadas pelos diferentes níveis.

4.5 SISTEMA DE DETECÇÃO DE ANOMALIAS VEICULARES

O sistema de detecção de anomalias proposto neste trabalho utiliza um computador como hospedeiro (*host*), ou seja, como unidade de processamento para treinar e validar o modelo de aprendizado. Como mostrado anteriormente na arquitetura do sistema, os dados

utilizados pelas diferentes técnicas de detecção de anomalias avaliadas são provenientes do sistema embarcado veicular. Desta maneira, o computador hospedeiro com o sistema de detecção de anomalias recebe as informações através de conexão USB e armazena os detalhes dos ciclos de direção (*trips*) em arquivo no formato CSV, tipo de arquivo de texto com ordenação de valores separados por vírgulas.

Visto que o sistema de detecção de anomalias utiliza classificadores *one-class* para o processo de detecção de anomalias veiculares, assim, o mesmo pode ser caracterizado como um sistema de aprendizado de máquina. Em geral, os sistemas de aprendizado utilizam as seguintes etapas para o processo de classificação: aquisição de dados, pré-processamento, transformação, extração de características, seleção de características, construção do modelo (etapa de treinamento) e o teste do modelo (etapa de testes).

Seguindo as etapas da criação de uma máquina de aprendizado, e de sua validação, todo o fluxo de dados realizado pelo sistema de detecção de anomalias veiculares pode ser detalhado na Figura 41. Os submódulos de cada uma das etapas serão detalhados a seguir:

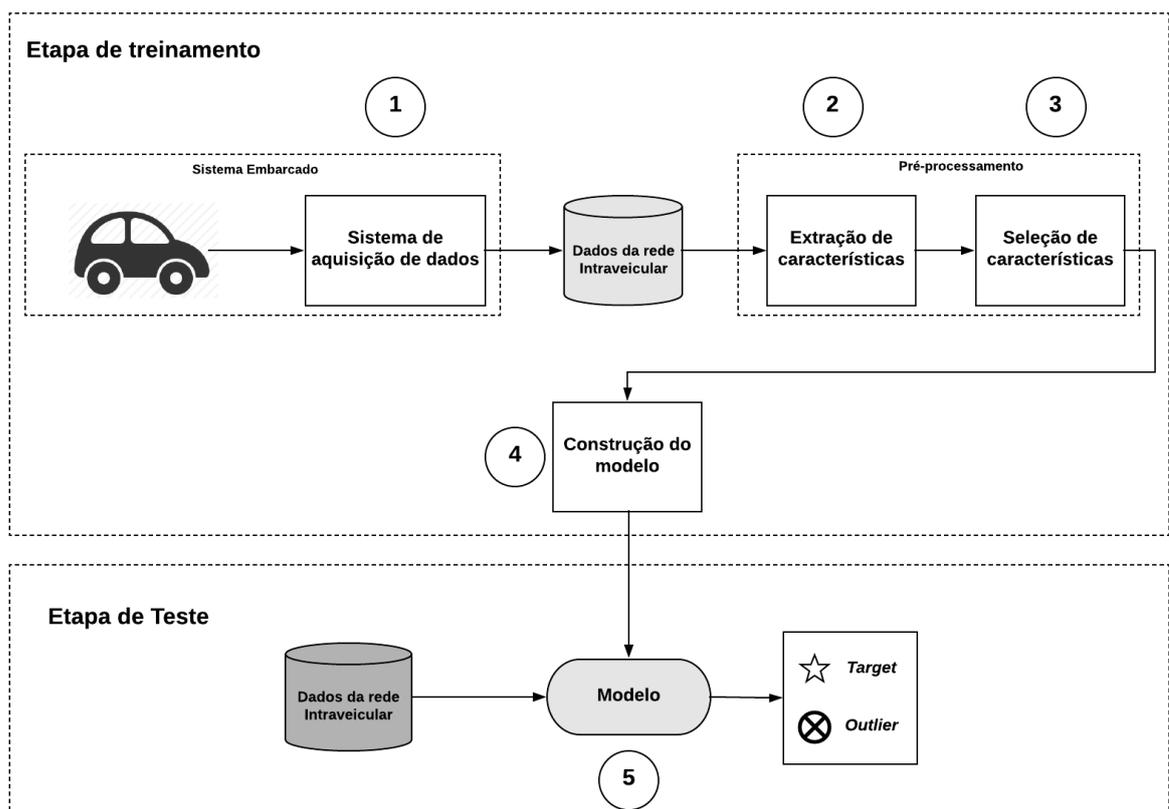


Figura 41 – Etapas do sistema de detecção de anomalias, considerando as etapas de treinamento e teste

4.5.1 Sistema de aquisição de dados

Este submódulo é representado pelo sistema embarcado, descrito na Seção 4.2, conectado a interface OBD-II do veículo para a captura das informações intraveiculares. Quando conectado, o dispositivo captura cada um dos 27 parâmetros intraveiculares disponíveis a uma taxa de uma amostra por segundo (1 Hz).

4.5.2 Pré-Processamento: extração e seleção de características

A primeira operação matemática (4.5) realizada por este módulo é a normalização das variáveis em valores entre 0 e 1, para as técnicas de detecção de anomalias que necessitam de dados normalizados, como exemplo o k -NN. Com a presença de diversos sistemas em um veículo moderno, nem todas as 27 variáveis capturadas pelo sistema de aquisição de dados estão diretamente relacionadas ao sensor ECT.

$$y_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (4.5)$$

Analisando as variáveis disponíveis, além da própria temperatura do sensor ECT, apenas a carga do motor e o número de rotações por minuto são parâmetros diretamente relacionadas ao sistema de resfriamento do motor.

Em uma comparação a nível de funcionamento do sistema e correlação entre os sinais, foi escolhida a variável que representa o número de rotações por minuto (RPM) do motor para que em conjunto com o valor da temperatura do sensor ECT, são definidas como atributos do sistema. A correlação média entre o sinal de ECT e outros sinais que poderiam ser considerados para o sistema é mostrada na Figura 42, onde *Load* é a variável da carga do motor, *Tank* o nível do tanque de combustível, *MAP* a pressão absoluta do coletor, *Ltft* sensor de oxigenação da mistura e *C. Temp* sensor de temperatura do catalisador. Apesar da figura apresentar o valor médio, sinais específicos apresentam valores maiores, como um valor de correlação de 0.2718 entre o valor do ECT e RPM.

Com o objetivo da detecção de um comportamento anômalo do sensor ECT, sensor puramente resistivo, é utilizada uma janela de tamanho \mathbf{N} para extração da média, variância e desvio padrão de cada atributo (valor do ECT e RPM) naquele intervalo.

$$X_{trip_1} = \begin{pmatrix} x_{1,t_1} & x_{2,t_1} \\ x_{1,t_2} & x_{2,t_2} \\ \vdots & \vdots \\ x_{1,t_k} & x_{2,t_k} \end{pmatrix} \quad (4.6)$$

Assim, para um sinal de uma *trip* (de k segundos), representado por (4.6), o janelamento pode ser aplicado para obtenção de um novo conjunto de dados. Com x_1 represen-

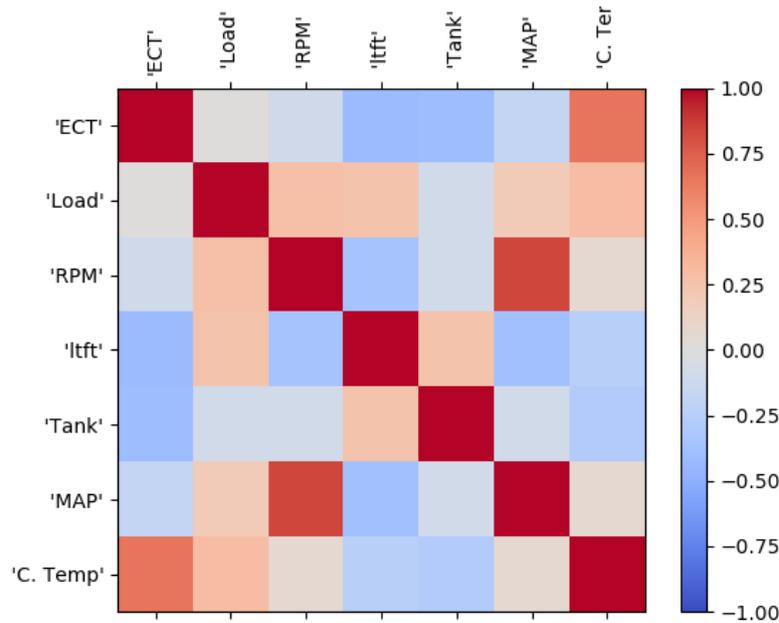


Figura 42 – Gráfico da Matriz de Correlação entre sete parâmetros coletados na base de dados

tando o valor do ECT e x_2 o do RPM, (4.6) detalha a saída da janela que é aplicada na parte $X_{t_j \dots t_{j+N-1}}$ do sinal.

$$Y_{t_j \dots t_{j+N-1}} = \left(\bar{x}_1 \quad s_{x_1}^2 \quad s_{x_1} \quad \bar{x}_2 \quad s_{x_2}^2 \quad s_{x_2} \right) \quad (4.7)$$

A estratégia do uso de uma janela nos dados é escolhida devido a dificuldade na análise de anomalias pontuais. Quatro novas características são criadas, a partir do valor do ECT e RPM. Inicialmente são calculadas as médias, desvio padrão e variância de cada uma das variáveis no intervalo da janela. A justificativa para a criação de novas características é a análise em partes do ciclo de direção veicular, considerando a presença de uma anomalia contextual, como descrito na Seção 2.3, Por fim, são selecionadas apenas quatro características de (4.7) para a composição final do módulo de pré-processamento. Uma análise inicial realizada nos dados mostrou que a extração de novas características a partir do RPM não retorna nenhuma relação direta com a possível presença de anomalias. Dessa forma, (4.8) descreve uma instância de entrada para o sistema de detecção de anomalias.

$$F_i = \left(\bar{x}_1 \quad s_{x_1}^2 \quad s_{x_1} \quad \bar{x}_2 \right) \quad (4.8)$$

O novo conjunto dentre dados resultante do módulo de pré-processamento é formado pela concatenação destas instancias, resultante de uma *trip* de M segundos.

4.5.3 Construção do modelo

Nesta etapa a técnica de detecção de anomalias escolhida é utilizada para construir o modelo, a máquina de aprendizado. Assim, a partir dos dados fornecido pela base, um modelo específico é construído para ser aplicado para resolver o problema em questão. No caso particular deste trabalho, apenas dados referentes ao funcionamento normal do veículo são utilizados para a construção do modelo.

O sistema de detecção de anomalias é construído a partir da base de dados criada com o sistema de aquisição de informações intraveiculares. Com a maior parte dos dados descrevendo o funcionamento do carro em seu funcionamento normal, o modelo construído se caracteriza como um modelo de detecção de novidade (*novelty detection*). No caso específico do sistema veicular relacionado ao sensor ECT, uma novidade é caracterizada como anomalia.

Com a proposta do trabalho definida na detecção de dois diferentes tipos de anomalia, a construção do modelo do sistema é a mesma para ambos os tipos de anomalias (Tipo I e II). A etapa de teste, onde a base de dados contém informações do funcionamento normal e as caracterizadas como anomalias, retorna o quão bom é o sistema proposto. Apesar de ser um problema de classificação binária, a detecção de anomalias, o procedimento de teste do sistema para diferentes tipos anomalias ocorre de maneira separada.

4.5.4 Modelo

Como ultimo submódulo do sistema, a aplicação do modelo define a caracterização de uma sequência de observações como comportamento normal ou anômalo do veículo. Parte da base veicular coletada é utilizada aqui, sem nenhuma sobreposição com o conjunto de construção do modelo. As anomalias são provenientes do procedimento de inserção de anomalias descrito anteriormente.

5 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados os experimentos realizados durante o desenvolvimento deste trabalho. A Seção 5.1 apresenta as métricas de desempenho para comparação dos resultados, seguido pela descrição da base de dados utilizadas na Seção 5.2. Em seguida as técnicas de detecção de anomalias utilizadas são listadas na Seção 5.3. De forma a realizar os experimentos com os melhores parâmetros das técnicas, a Seção 5.4 detalha a etapa de extração de características. Os resultados para o sistema com o veículo em ciclos direção e em modo *idle* são apresentados nas Seções 5.5 e 5.6, respectivamente. Por fim, a Seção 5.7 apresenta os melhores resultados para trajetos individuais do veículo em modo em movimento.

5.1 MÉTRICAS DE DESEMPENHO

A avaliação da performance de um classificador pode ser realizada a partir dos resultados de sua matriz de confusão. Considerando o caso do uso de classificadores *one-class*, utilizados no processo de detecção de anomalias, sua matriz de confusão pode ser representada na Tabela 6.

A matriz de confusão consiste nos seguintes elementos:

- TP (verdadeiros positivo): a quantidade de elementos positivos previstos como positivos;
- FP (falsos positivo): a quantidade de elementos negativos previstos como positivos;
- FN (falsos negativo): a quantidade de elementos positivos previstos como negativos;
- TN (verdadeiro negativo): a quantidade de elementos negativos previstos como negativos;

A avaliação das técnicas deve considerar o desempenho tanto da classe de dados anômalos (*outliers*), quanto do funcionamento normal (*target*). Assim, as métricas de desempenho escolhidas para comparação neste trabalho são:

Tabela 6 – Exemplo de uma Matriz de Confusão

Rótulo da classe	Resultado da classificação	
	Normal	Anomalia
Normal	TP	FN
Anomalia	FP	TN

1. **Acurácia:** a taxa de decisões corretas realizadas pelo classificador ou técnica de detecção de anomalia.

$$acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (5.1)$$

2. **Sensibilidade:** também chamado de taxa de acerto ou *recall*, mede quanto um classificador pode reconhecer exemplos positivos.

$$sens = \frac{TP}{TP + FN} \quad (5.2)$$

3. **Especificidade:** também chamado de *true negative rate* (TNR), mede quanto um classificador pode reconhecer exemplos negativos.

$$spec = \frac{TN}{TN + FP} \quad (5.3)$$

4. **Precisão:** é a taxa de acerto dos exemplos positivos (*target*).

$$prec = \frac{TP}{TP + FP} \quad (5.4)$$

Esta métrica também pode ser expandida para os dados do tipo *outlier*, sendo definida por:

$$prec_{out} = \frac{TN}{TN + FN} \quad (5.5)$$

5. **F2-score:** é uma medida da precisão de um classificador. Assim, o F2-score considera tanto a precisão, quanto a sensibilidade em sua fórmula. Por ser uma métrica baseada na média harmônica ponderada da sensibilidade e precisão, é definida a partir do *F-measure* com o valor de $\beta = 2$ (ESPÍNDOLA; EBECKEN, 2005).

$$F_2 = \frac{5 \times prec \times sens}{4 \times prec + sens} \quad (5.6)$$

5.2 BASE DE DADOS UTILIZADAS

Nem todos os dados veiculares coletados durante os seis meses de construção da base foram utilizados. Definindo $\|T\|$ como o tamanho do conjunto de treinamento em segundos, $\|TT\|$ o tamanho do conjunto de teste referente ao tipo de dado *target* também em segundos e $\|OT\|$ o tamanho dos *outliers* presentes no conjunto de teste, a Tabela 7 detalha os tamanhos dos conjuntos.

Em específico ao modo do veículo em movimento, os testes podem também ser divididos em quatro cenários de diferentes tipos de trajetos realizados pelo veículo. A Tabela 8

Tabela 7 – Descrição da base de dados usado nos experimentos do carro em modo *idle* e em movimento para anomalia Tipo II.

Modo	$\ T\ $	$\ TT\ $	$\ OT\ $	Número de <i>trips</i>
Idle	16430	8400	1875	25
Movimento	38321	19840	3375	45

detalha o tamanho das bases utilizadas no treinamento e teste dos diferentes trajetos no modo do veículo em movimento. De maneira a detalhar as duas informações utilizadas no sistema, a Figura 43 ilustra os histogramas das duas variáveis para o conjunto de treinamento, nos modos do veículo em movimento e *idle*. A Figura 44 apresenta os histogramas dos quatro diferentes trajetos do veículo em movimento.

Tabela 8 – Descrição da base de dados usado nos experimentos considerando os quatro cenários do carro em movimento

Cenário	$\ T\ $	$\ TT\ $	$\ OT\ $	Número de <i>trips</i>
Fixo	13198	3512	750	10
Universidade campus	4881	3837	750	10
Via Expressa	12584	3480	450	6
Áreas urbanas	8301	2768	375	5

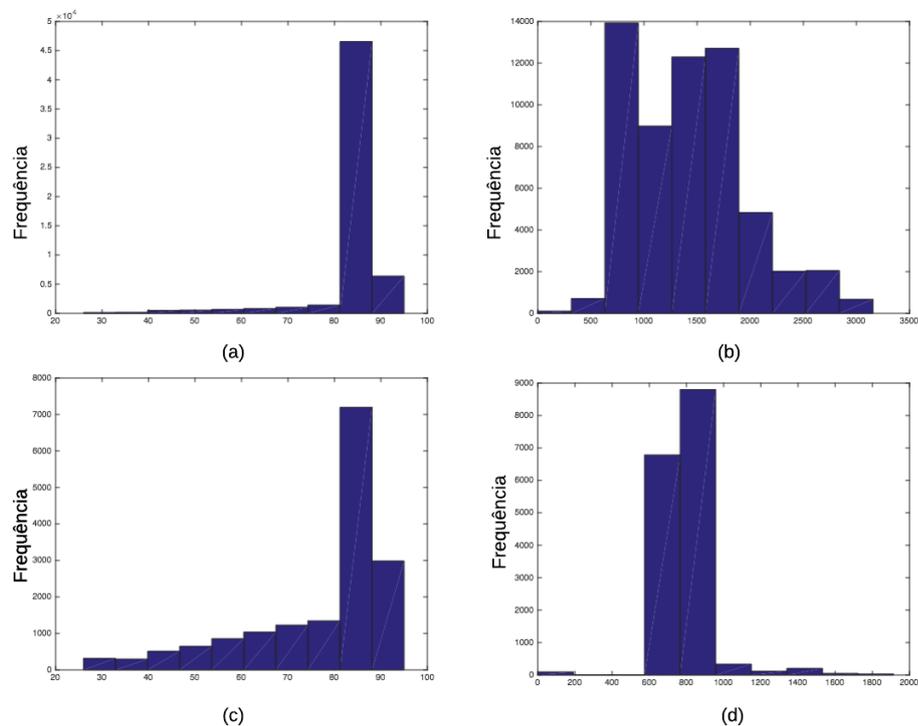


Figura 43 – Histograma dos parâmetros utilizados para treinamento no modo em movimento (a) variável do sensor ECT (b) número de rotações por minuto (RPM) e modo *idle* (c) variável do sensor ECT (d) número de rotações por minuto (RPM)

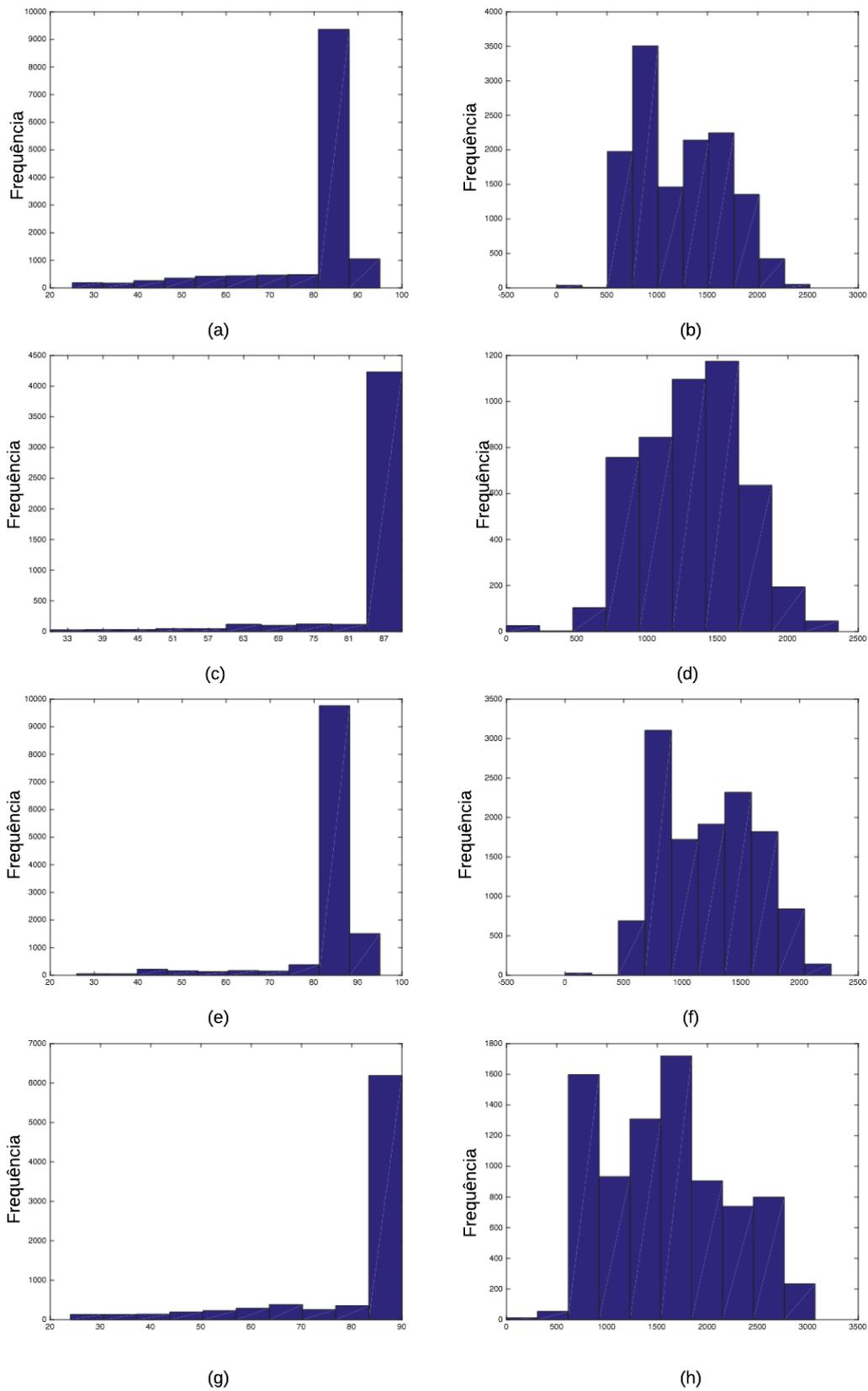


Figura 44 – Histograma dos parâmetros utilizados para treinamento no modo do veículo em movimento com trajetos específicos. Variável do sensor ECT (a) em trajeto fixo (c) trajeto em campus da universidade (e) trajetos urbanos (g) vias expressas. Variável do número de rotações por minuto (RPM) (b) em trajeto fixo (d) trajeto em campus da universidade (f) trajetos urbanos (h) vias expressas.

5.3 DEFINIÇÃO DA TÉCNICA DE DETECÇÃO DE ANOMALIA

A avaliação das diferentes técnicas de detecção de anomalias neste trabalho objetiva selecionar a melhor para as situações em que o sistema proposto deve atuar, a detecção de anomalias Tipo I e II. Comparando com outros trabalhos propostos na literatura, (THEISSLER, 2017) apresenta a detecção de anomalias no sensor ECT através de um *ensemble*. Apesar de a proposta ser baseada em combinação de técnicas, seu trabalho apresenta uma comparação individual das mesmas, ao analisar a anomalia Tipo II.

Na avaliação proposta por este trabalho, além da avaliação do sistema em diferentes níveis de anomalias, o objetivo é definir se outras diferentes técnicas também podem ser aplicadas para a solução do problema, em avaliação separada das anomalias Tipo II. De maneira inicial, foi replicada a metodologia proposta por (THEISSLER, 2013), incluindo a nomenclatura correspondente aos tipos das anomalias. Assim, ao considerar anomalias Tipo I, foi utilizada a técnica de detecção de limiar *Linear Outlier Factor* para replicar o resultado a partir dos dados obtidos pelo sistema de aquisição.

Como anomalia Tipo I deriva da desconexão do cabo que conecta a ECU ao sensor, a presença do valor de escala do sensor no sinal o torna mais fácil de detecção, quando comparado a outros tipos de anomalia. A técnica LOF retorna ótimo resultado com baixa complexidade de implementação, fazendo com que a mesma já seja suficiente para a detecção de anomalias Tipo I.

Em relação as anomalias Tipo II, o trabalho de (THEISSLER, 2017) compara individualmente apenas as técnicas: *Extreme Value*, OC-SVM, SVDD e Mahalanobis, a proposta aqui apresentada visa incluir outras técnicas de detecção de limiar e classificadores *one-class*. Além das alternativas listadas anteriormente, também são avaliadas as seguintes técnicas:

- Classificadores *one-class*: *k-NN*, *distance k-NN* (é assumida a entrada como uma matriz de distância ou dissimilaridade), *Minimum Spanning Tree* (MST), *Self-organized map* (SOM)
- Análise Estatística: *Gauss Data Description*, *Naive Parzen*, *Parzen*

As técnicas escolhidas são fundamentadas na possibilidade de comparação de diferentes ferramentas matemáticas, além do seu uso em trabalhos relacionados. Dessa forma algoritmos baseados em redes neurais, limiares, índices de proximidade e máquinas de vetor de suporte são avaliados.

5.4 EXTRAÇÃO DE CARACTERÍSTICAS: DEFINIÇÃO DO TAMANHO DA JANELA

No processo de extração de atributos, etapa utilizada no fluxo do sistema detector de anomalia Tipo II, a seleção do tamanho da janela a ser utilizado em cada técnica avaliada

e respectivo modo de operação é feita a partir da análise do valor do $F2-score$. A justificativa da escolha dessa métrica é baseada em que o sistema deve garantir boas taxas de acerto para ambos tipos de dado. Considerando (5.6), o cálculo do $F2-score$ neste trabalho considera o uso da precisão no acerto dos *outliers*.

De forma a simplificar os resultados propostos, desde que são avaliados cinco diferentes níveis de anomalia, é definido um único tamanho de janela para todos os cinco níveis. Assim, outro parâmetro que também é considerado, além do $F2-score$, é o tempo de processamento a cada técnica, quando utilizado aquele tamanho de janela. A Figura 45 exemplifica o processo de escolha de janela para a técnica SVDD no modo de operação *idle*. O conjunto de dados utilizado para a escolha do tamanho da janela é um subgrupo do conjunto de teste $\|TT\|$.

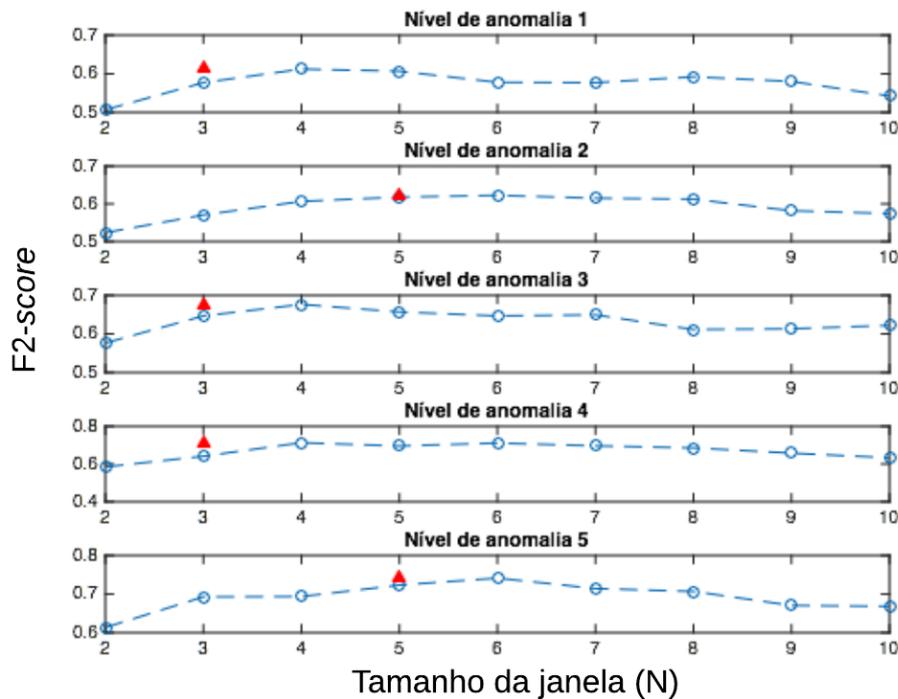


Figura 45 – Seleção do tamanho da janela para extração de características

Como pode ser visto, nem todos os níveis de anomalia retornam o mesmo tamanho de janela como o melhor valor, representado pelo triângulo destacado. Assim, também é realizada uma análise no valor do tempo de processamento para os valores de $N = 3$, $N = 4$ e $N = 5$ para a seleção do tamanho da janela. Em particular a técnica SVDD, quanto maior o tamanho da janela, menor o tempo de processamento. Como a diferença de valores entre os resultados das janelas $N = 4$ e $N = 5$ para os níveis de anomalia 4 e 5 é muito tempo, é escolhido o valor de $N = 5$ para ser o tamanho da janela de extração de atributos da técnica SVDD para o modo *idle*.

De forma a documentar o tamanho das janelas nos experimentos realizados com as bases de dados descritas na seção anterior, a Tabela 9 apresenta os respectivos valores

do tamanho da janela de extração de atributos utilizada em cada modo operação. Além de experimentos nos dois principais modos de operação do veículo, são realizados experimentos em trajetos específicos do modo do veículo em movimento. Assim, a Tabela 10 apresenta o tamanho das janelas para os respectivos trajetos específicos.

Tabela 9 – Tamanho da janela de seleção de atributos para técnicas utilizadas no trabalho

Tamanho da janela selecionado para extração de atributos		
Técnica	Modo IDLE	Modo Movimento
Gauss	10	10
<i>k-NN</i>	5	6
Mahalanobis	10	6
MST	6	7
<i>Naive Parzen</i>	6	10
SOM	7	6
Parzen	6	10
SVDD	10	4
<i>Extreme Value</i>	10	10
OC-SVM	3	9

Observando os valores apresentados na Tabela 9, nota-se a seleção de tamanhos idênticos para as técnicas Gauss e *Extreme Value*, para ambos modos de operação do veículo. Apesar de diferentes, os tamanhos de janela para as técnicas *k-NN*, MST e SOM são bastante próximos. Por outro lado, as outras técnicas possuem valores distintos do tamanho da janela para os dois modos.

Tabela 10 – Tamanho da janela de seleção de atributos para técnicas utilizadas no trabalho - Trajetos específicos

Tamanho da janela selecionado para extração de atributos				
Técnica	Trajeto Fixo	Trajeto Campus Universidade	Trajeto áreas urbanas	Trajeto vias expressas
Gauss	9	10	9	10
<i>k-NN</i>	10	8	10	10
Mahalanobis	6	6	6	8
MST	6	3	6	5
<i>Naive Parzen</i>	6	2	5	3
SOM	10	4	10	10
Parzen	3	5	2	3
SVDD	5	2	2	2
<i>Extreme Value</i>	10	10	10	10
OC-SVM	3	10	10	10

Apesar de um maior tamanho da janela resultar em um menor tempo para criação do modelo, os tamanhos de janela para o modo *idle* são em sua maioria menores que os do modo em movimento. Analisando as duas características para a construção do modelo, a observação pode ser explicado a partir que o valor da variável RPM não varia tanto no modo *idle*, quando comparado ao veículo em movimento.

Na Figura 46 (a), encontram-se os tempos médios de processamento, considerando o uso do conjunto de treinamento $\|TT\|$, para as técnicas avaliada neste trabalho para cada tamanho de janela no modo *idle*. Por sua vez, a Figura 46 (b) representa os tempos

Tabela 11 – Tempo médio de processamento, em segundos, na etapa de teste para o modo *idle*

Técnica	N = 2	N = 3	N = 4	N = 5	N = 6	N = 7	N = 8	N = 9	N = 10
Gauss	0.013	0.012	0.009	0.009	0.009	0.010	0.009	0.002	0.004
k-NN	2.305	1.075	0.501	0.303	0.210	0.151	0.109	0.089	0.074
Mahalanobis	0.024	0.023	0.022	0.019	0.019	0.019	0.018	0.018	0.019
MST	2.008	0.879	0.496	0.313	0.217	0.159	0.121	0.098	0.078
Naive Parzen	0.651	0.309	0.187	0.118	0.088	0.062	0.056	0.038	0.035
SOM	0.015	0.011	0.038	0.012	0.010	0.018	0.010	0.017	0.009
Parzen	0.561	0.343	0.189	0.103	0.071	0.057	0.042	0.041	0.032
SVDD	339.800	162.267	134.533	131.133	65.667	58.733	55.333	49.372	36.732
Extreme Value	0.087	0.085	0.079	0.056	0.05	0.032	0.028	0.023	0.019
OC-SVM	3.202	2.432	2.102	1.978	1.802	1.02	0.089	0.013	0.010

médios para o modo do veículo em movimento. Apesar de ser proporcional para todas as técnicas avaliadas, o tamanho da base influencia de uma maneira maior as técnicas k-NN, MST, *Naive Parzen* e SVDD.

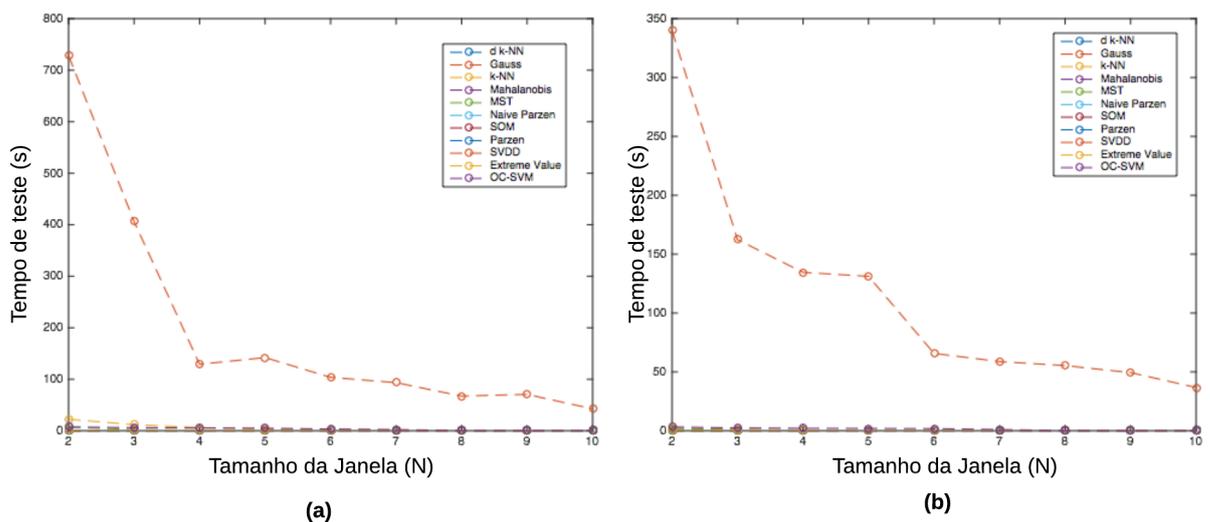


Figura 46 – Tempo médio de processamento na etapa de teste das técnicas utilizadas nos experimentos (a) em modo *idle* (b) em modo do veículo em movimento

De posse dos valores do tempo de processamento na etapa de teste para cada técnica, nos diferentes tamanhos de janela, a Tabela 11 detalha os resultados. Como discutido anteriormente, métodos como *Extreme Value*, *Mahalanobis* e Gauss possuem valores muito próximos, o que permite afirmar que o tamanho da base não influencia tanto, como nas outras técnicas.

Da mesma forma, a Tabela 12 mostra os resultados do tempo médio de processamento para o modo em movimento. De maneira semelhante, as mesmas técnicas citadas no modo *idle* também possuem valores médios próximos, o que corrobora com a observação anterior.

Tabela 12 – Tempo médio de processamento, em segundos, na etapa de teste para o modo do veículo em movimento

Técnica	N = 2	N = 3	N = 4	N = 5	N = 6	N = 7	N = 8	N = 9	N = 10
Gauss	0.013	0.011	0.010	0.010	0.010	0.0098	0.0096	0.009	0.009
k-NN	22.341	11.683	6.095	3.416	2.295	1.6695	1.1988	0.943	0.751
Mahalanobis	0.029	0.025	0.025	0.023	0.022	0.0207	0.0205	0.019	0.019
MST	7.242	2.135	1.175	0.749	0.539	0.389	0.292	0.237	0.188
Naive Parzen	1.157	0.587	0.408	0.315	0.252	0.148	0.119	0.089	0.075
SOM	0.0216	0.013	0.012	0.011	0.015	0.011	0.010	0.010	0.010
Parzen	1.052	0.645	0.079	0.048	0.149	0.089	0.074	0.064	0.051
SVDD	728.2333	406.133	142.0	129.066	103.660	93.400	67.200	70.533	42.467
Extreme Value	0.102	0.092	0.089	0.086	0.085	0.0075	0.08	0.071	0.07
OC-SVM	7.172	6.212	5.452	4.92	3.302	2.072	0.19	0.0353	0.028

5.5 EXPERIMENTOS COM O VEÍCULO EM CICLOS DE DIREÇÃO

Os experimentos realizados neste trabalho foram implementados em sua maioria na linguagem MATLAB. Foi escolhido o MATLAB devido a parte das técnicas de classificadores *one-class* utilizados já estarem implementadas pela *toolbox* DDTools versão 2.1.3 (TAX, 2018). Além da DDTools, a técnica Mahalanobis e *Extreme Value* foram implementadas. O processo de execução do sistema e avaliação das técnicas foram realizadas em computador hospedeiro com as especificações descritas na Tabela 13.

Tabela 13 – Especificações do Computador utilizado para a realização dos experimentos

Processador	
Modelo	Intel(R) Core(TM) i5
Número de cores	2
Número de Threads	2
Frequência	2.5 GHz
Frequência Turbo Boost	3.1 GHz
Tamanho da Cache	4MB L3
RAM	
Tamanho da RAM	10 GB - 600MHz DDR3

A Figura 47 detalha as etapas de treinamento e teste para o sistema de detecção de anomalias. Os dois grupos, $\|A\|$ e $\|B\|$, são totalmente separados a partir dos conjuntos de *trips*. Para cada *trip* do grupo $\|B\|$ o gerador de anomalias insere um determinado número de amostras consecutivas com comportamento anômalo. O restante dos dados de cada ciclo é considerado como dado de funcionamento normal. Considerando os dois grupos $\|A\|$ e $\|B\|$, o objetivo é obter uma divisão próxima de relação 80% dos dados do conjunto de testes descreverem o funcionamento normal e 20% com presença de anomalias.

As duas próximas seções descrevem os experimentos realizados para o modo do carro em movimento e a análise dos respectivos resultados.

5.5.1 Anomalia Tipo I

Seguindo a nomenclatura proposta por (THEISSLER, 2013), o primeiro tipo de anomalia a ser detectado em um sensor veicular é a desconexão do seu cabo ligado a ECU responsável

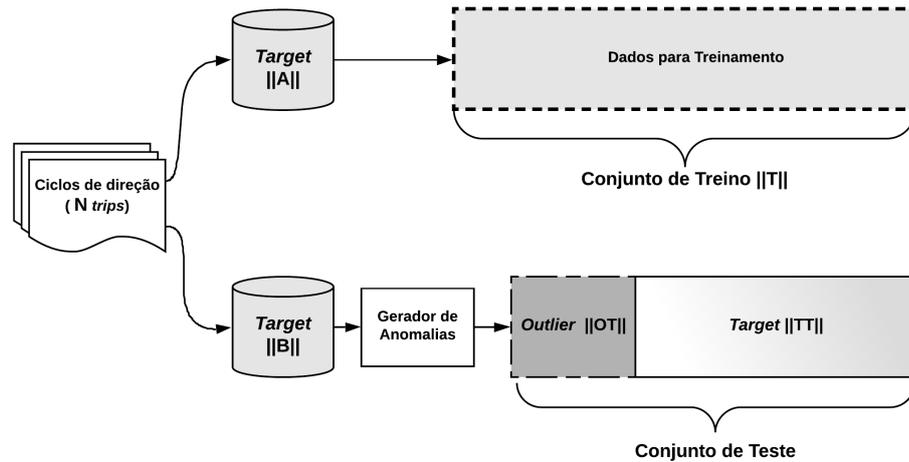


Figura 47 – Detalhamento das etapas de treinamento e teste para a avaliar o processo de detecção de anomalias

pelo seu controle/monitoramento. Assim, a anomalia Tipo I se caracteriza de forma simples pela ausência do sinal elétrico do sensor ECT na ECU correspondente. Na proposta deste trabalho, o valor de fundo de escala correspondente ao processo de desconexão é definido como zero.

O experimento realizado utiliza os dados *target* da base do veículo em movimento descrito na Tabela 7. O conjunto $\|A\|$ contém 30 *trips* e o conjunto descrito $\|B\|$ contém 15 *trips*, que são transformadas em sinais com anomalias a partir do gerador de anomalias do Tipo I. As anomalias são inseridas individualmente a cada uma das 15 *trips*, sendo caracterizadas por uma sequência de **300** amostras seguidas com o valor de fundo de escala.

A etapa de treinamento para a técnica LOF é realizada de maneira simples, visto que a mesma objetiva definir o *threshold* do comportamento do sinal do funcionamento normal. O estabelecimento dos parâmetros do LOF, valor do limite para comparação com k vizinhos, foi definido a partir da análise de toda base treinamento. Definindo o valor de $k = 10$, para uma análise que envolva o contexto, o valor para o limiar para a base do veículo em movimento foi de 1.15. De maneira a ilustrar o procedimento, a Figura 48 mostra o valor do *threshold* para uma (a) *trip* normal e (b) com a presença de uma desconexão do cabo do sensor ECT.

Foi possível observar que a aplicação da técnica LOF no início e final de uma *trip* retornou resultados não satisfatórios, como a presença de falsos negativos (FN). Como solução, a técnica LOF deve iniciar o processo a partir da décima amostra, devido ao valor de $k = 10$. A Figura 49 (a) mostra o resultado da técnica LOF e em (b) é possível ver o momento da desconexão do cabo e o respectivo resultado da técnica.

Ainda como resultado mostrado na Figura 49, a técnica LOF caracteriza os valores

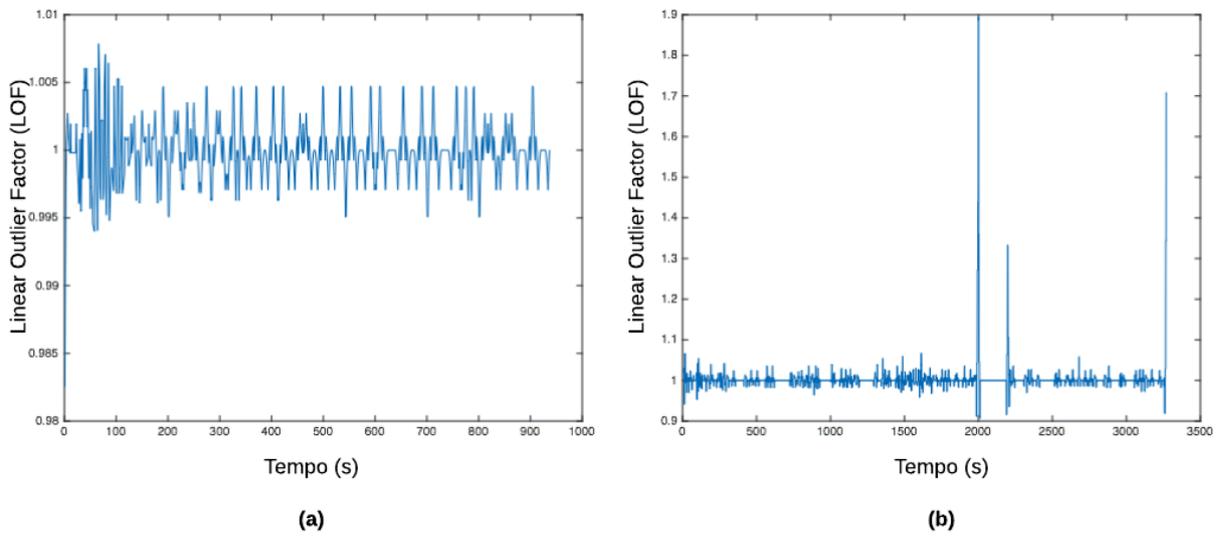


Figura 48 – *Linear Outlier Factor* para sinal (a) com funcionamento normal do ECT (b) sinal com desconexão do cabo

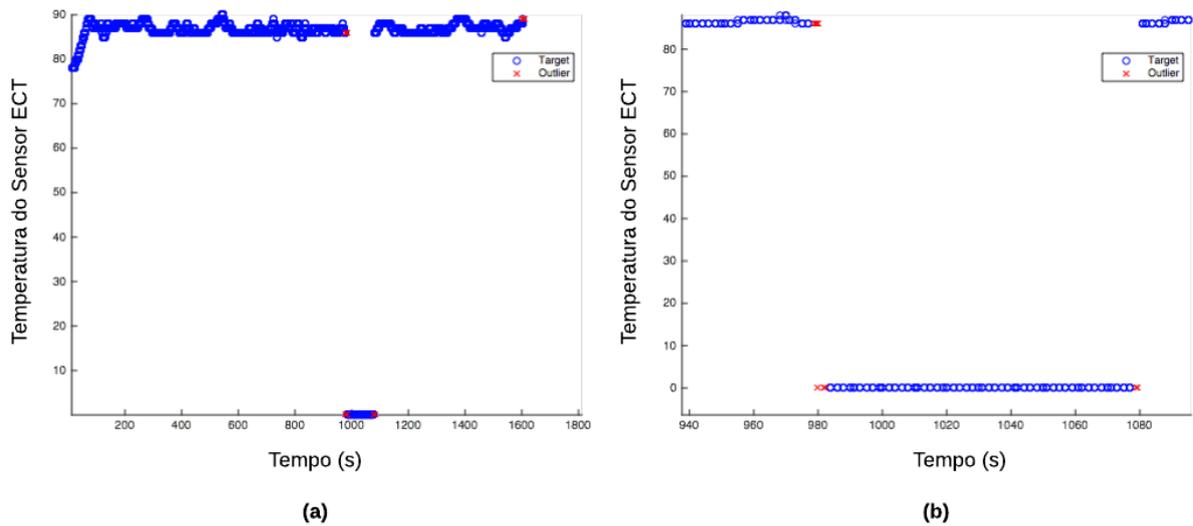


Figura 49 – Resultado da técnica *Linear Outlier Factor* em sinal do sensor ECT (a) sinal completo representado pela caracterização das amostras (b) destaque para a parte da desconexão do sensor

de fundo de escala como dados do tipo *target*. Assim, a proposta da avaliação do LOF nesse trabalho objetiva a detecção da ruptura do cabo de comunicação, sendo a avaliação dos dados do tipo *outlier* o valor subsequente a desconexão. Totalizando todos os dados presentes na base de dados, e os 15 sinais utilizados na etapa de teste, a Tabela 14 mostra os resultados para a detecção de anomalias Tipo I.

O valor da métrica precisão *outlier* na Tabela 14 considera os falsos negativos do final de cada *trip*. Ao considerar o resultado sem as amostras finais, o valor do resultado é 1.00. Como a desconexão do cabo do sensor é um tipo de anomalia mais simples, as etapas 2 e

Tabela 14 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo I

Modo	Acurácia	Sensibilidade	Especificidade	Precisão <i>Target</i>	Precisão <i>Outlier</i>	F2-score
Movimento	0.997 ± 0.001	0.997 ± 0.001	1.000 ± 0.000	1.000 ± 0.000	0.971 ± 0.001	0.998 ± 0.001

Tabela 15 – Parâmetros específicos utilizadas em cada técnica

Técnica	Descrição do parâmetro
Gauss	Regularização, $R = 0.001$
k-NN	$k = 1$, $Threshold = \max(d)$
Mahalanobis	$Threshold = \max(d)$
MST	$Threshold = \max(d)$
Naive Parzen	h otimizado
SOM	Rede de tamanho 5×5
Parzen	h otimizado
SVDD	<i>Kernel</i> RBF, $\sigma = 5$
Extreme Value	$Threshold = \sigma$
OC-SVM	<i>Kernel</i> polinomial de terceiro grau

3 do fluxo proposto na Figura 41 não são realizados. A etapa de construção do modelo é resumida pela definição do *threshold*.

5.5.2 Anomalia Tipo II

Para a detecção das anomalias provenientes da possível falha resistiva também é utilizado o processo descrito na Figura 47 para divisão dos grupos de treinamento e teste. De maneira diferente, o gerador de anomalias para o Tipo II cria *outliers* a partir de 225 amostras consecutivas de cada *trip*. Utilizando a metodologia proposta na Seção 4.4, cinco diferentes níveis de anomalias são definidos para o sinal do sensor ECT, definindo assim a base utilizada nos experimentos, descrita na Tabela 7.

De maneira mais detalhada, os dados referentes ao funcionamento normal do veículo, são separados nos respectivos conjuntos de ciclos de direção $\|A\|$ e $\|B\|$. Para a anomalia tipo II o valor de N , número total de ciclos utilizados para criar a base em movimento é 45. Sendo 30 *trips* usadas para o grupo $\|A\|$ e 15 para o grupo $\|B\|$. Após o procedimento descrito na Figura 47, obtém-se os conjuntos $\|T\|$, $\|TT\|$ e $\|OT\|$. Diferente da anomalia Tipo I, o conjunto de dados anômalos é construído a partir da inserção de anomalia em 225 amostras consecutivas dos sinais selecionados.

Os experimentos realizados neste trabalho objetivam definir a(s) melhor(es) técnica(s) para os dois modos de operação definidos. Os parâmetros específicos para cada técnica são descritos na Tabela 15, válidos para ambos os modos do veículo.

Com o objetivo de realizar uma análise estatística do sistema de detecção de anomalias proposto, os resultados apresentados são baseados na construção dos modelos de cada técnica trinta vezes. Dessa maneira, os resultados são apresentados com valores médios e seus respectivos desvios padrões. As tabelas 16 à 20 resumem os valores obtidos para os cinco níveis de anomalias.

Tabela 16 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 1

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.590 ± 0.011	0.654 ± 0.005	0.335 ± 0.018	0.795 ± 0.019	0.189 ± 0.025	0.438 ± 0.008
k-NN	0.827 ± 0.017	0.939 ± 0.001	0.670 ± 0.022	0.799 ± 0.036	0.873 ± 0.020	0.925 ± 0.006
Mahalanobis	0.685 ± 0.001	0.685 ± 0.001	1.000 ± 0.000	1.000 ± 0.000	0.005 ± 0.001	0.024 ± 0.001
MST	0.649 ± 0.007	0.680 ± 0.003	0.474 ± 0.023	0.880 ± 0.011	0.109 ± 0.006	0.332 ± 0.005
Naive Parzen	0.662 ± 0.008	0.722 ± 0.004	0.510 ± 0.016	0.791 ± 0.019	0.336 ± 0.009	0.587 ± 0.007
SOM	0.576 ± 0.007	0.639 ± 0.004	0.297 ± 0.021	0.799 ± 0.019	0.185 ± 0.009	0.428 ± 0.006
Parzen	0.596 ± 0.009	0.669 ± 0.005	0.380 ± 0.015	0.760 ± 0.015	0.301 ± 0.008	0.537 ± 0.008
SVDD	0.552 ± 0.008	0.704 ± 0.005	0.401 ± 0.006	0.539 ± 0.017	0.379 ± 0.010	0.600 ± 0.007
Extreme Value	0.741 ± 0.012	0.831 ± 0.013	0.093 ± 0.007	0.869 ± 0.011	0.071 ± 0.008	0.264 ± 0.009
OC-SVM	0.901 ± 0.001	0.998 ± 0.000	0.226 ± 0.003	0.899 ± 0.002	0.957 ± 0.001	0.989 ± 0.002

Os resultados iniciais para o Nível I, mostrados na Tabela 16, mostram a técnica OC-SVM, com *kernel* polinomial de terceiro grau, como a melhor para o veículo em trajetória. Antes de definir o uso deste *kernel* também foram testados o RBF e o *kernel* linear. A técnica apresenta uma maior precisão para detectar dados anômalos do que os normais (*target*).

Ao considerar a técnica k-NN, a mesma apresentou uma configuração de parâmetros interessante, visto que foi utilizada uma janela de tamanho $N = 6$. O valor do k utilizado neste trabalho foi igual a um. Apesar do valor definir o uso da técnica como o 1-NN, este valor para comparação na vizinhança retornou o melhor resultado dentre valores ímpares de k testados entre 1 e 7. A Tabela 17 mostra os resultados para o segundo nível de anomalia.

Analisando a precisão de maneira separada, a técnica baseada na distância de Mahalanobis apresenta o melhor resultado, quando considerado os dados de funcionamento normal do veículo. Em relação a detecção das anomalias (*outliers*) a técnica com o melhor resultado foi a OC-SVM. É importante ressaltar que durante a análise, os níveis de anomalia quatro e cinco possuem uma maior influência na escolha dos resultados gerais, visto que as mesmas representam um maior grau de degeneração ou falha do sensor, podendo acarretar maiores danos ao motor do carro.

Tabela 17 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 2

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.573 ± 0.012	0.640 ± 0.005	0.264 ± 0.015	0.801 ± 0.017	0.134 ± 0.022	0.364 ± 0.004
k-NN	0.829 ± 0.015	0.943 ± 0.000	0.672 ± 0.021	0.799 ± 0.021	0.916 ± 0.002	0.930 ± 0.004
Mahalanobis	0.694 ± 0.003	0.691 ± 0.002	0.994 ± 0.014	1.000 ± 0.000	0.037 ± 0.001	0.018 ± 0.001
MST	0.677 ± 0.006	0.704 ± 0.003	0.554 ± 0.019	0.876 ± 0.011	0.244 ± 0.006	0.511 ± 0.005
Naive Parzen	0.709 ± 0.007	0.770 ± 0.004	0.580 ± 0.014	0.794 ± 0.014	0.494 ± 0.007	0.692 ± 0.007
SOM	0.601 ± 0.011	0.658 ± 0.005	0.377 ± 0.019	0.802 ± 0.017	0.207 ± 0.010	0.458 ± 0.009
Parzen	0.619 ± 0.012	0.691 ± 0.008	0.431 ± 0.021	0.761 ± 0.017	0.433 ± 0.010	0.617 ± 0.010
SVDD	0.579 ± 0.022	0.733 ± 0.006	0.430 ± 0.020	0.557 ± 0.043	0.394 ± 0.026	0.625 ± 0.020
Extreme Value	0.758 ± 0.013	0.839 ± 0.013	0.153 ± 0.009	0.881 ± 0.008	0.112 ± 0.011	0.366 ± 0.008
OC-SVM	0.900 ± 0.002	0.997 ± 0.001	0.221 ± 0.003	0.899 ± 0.002	0.933 ± 0.001	0.983 ± 0.002

Tabela 18 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 3

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.611 ± 0.012	0.670 ± 0.004	0.392 ± 0.020	0.803 ± 0.018	0.246 ± 0.002	0.498 ± 0.013
k-NN	0.822 ± 0.015	0.927 ± 0.008	0.669 ± 0.018	0.803 ± 0.032	0.847 ± 0.017	0.909 ± 0.011
Mahalanobis	0.730 ± 0.003	0.717 ± 0.002	0.997 ± 0.014	1.000 ± 0.000	0.129 ± 0.001	0.375 ± 0.001
MST	0.741 ± 0.006	0.763 ± 0.004	0.673 ± 0.016	0.879 ± 0.010	0.318 ± 0.005	0.596 ± 0.005
Naive Parzen	0.766 ± 0.011	0.848 ± 0.005	0.641 ± 0.021	0.786 ± 0.021	0.617 ± 0.011	0.788 ± 0.010
SOM	0.615 ± 0.009	0.672 ± 0.004	0.421 ± 0.017	0.796 ± 0.015	0.187 ± 0.009	0.442 ± 0.008
Parzen	0.695 ± 0.011	0.771 ± 0.007	0.554 ± 0.016	0.761 ± 0.014	0.448 ± 0.009	0.673 ± 0.010
SVDD	0.597 ± 0.013	0.771 ± 0.007	0.450 ± 0.011	0.543 ± 0.023	0.659 ± 0.016	0.745 ± 0.012
Extreme Value	0.797 ± 0.018	0.872 ± 0.011	0.352 ± 0.012	0.888 ± 0.005	0.319 ± 0.011	0.648 ± 0.010
OC-SVM	0.901 ± 0.002	0.998 ± 0.000	0.226 ± 0.003	0.899 ± 0.002	0.957 ± 0.001	0.989 ± 0.001

Tabela 19 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 4

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.681 ± 0.016	0.736 ± 0.004	0.543 ± 0.011	0.802 ± 0.020	0.435 ± 0.011	0.646 ± 0.007
k-NN	0.816 ± 0.014	0.922 ± 0.007	0.662 ± 0.019	0.799 ± 0.032	0.856 ± 0.019	0.908 ± 0.011
Mahalanobis	0.810 ± 0.012	0.783 ± 0.011	0.999 ± 0.003	1.000 ± 0.000	0.379 ± 0.006	0.645 ± 0.003
MST	0.814 ± 0.006	0.844 ± 0.004	0.749 ± 0.014	0.880 ± 0.001	0.584 ± 0.005	0.775 ± 0.006
Naive Parzen	0.814 ± 0.009	0.914 ± 0.005	0.683 ± 0.015	0.792 ± 0.015	0.767 ± 0.008	0.880 ± 0.006
SOM	0.579 ± 0.008	0.641 ± 0.005	0.308 ± 0.019	0.802 ± 0.020	0.201 ± 0.012	0.444 ± 0.006
Parzen	0.776 ± 0.008	0.883 ± 0.009	0.637 ± 0.019	0.760 ± 0.020	0.751 ± 0.012	0.853 ± 0.007
SVDD	0.652 ± 0.008	0.880 ± 0.004	0.501 ± 0.010	0.539 ± 0.021	0.739 ± 0.012	0.847 ± 0.004
Extreme Value	0.864 ± 0.025	0.942 ± 0.017	0.560 ± 0.019	0.893 ± 0.007	0.716 ± 0.021	0.886 ± 0.027
OC-SVM	0.902 ± 0.002	0.999 ± 0.000	0.230 ± 0.003	0.899 ± 0.002	0.983 ± 0.001	0.988 ± 0.001

O algoritmo baseado na distância de Mahalanobis apresenta o melhor resultado de precisão, quando analisada a classificação apenas dos dados tipo *target*. Apesar do bom resultado para os dados do funcionamento normal, a técnica apresenta o pior resultado da precisão para o tipo de dado anômalo. O resultado pode ser explicado pela não normalidade de todas as instâncias analisadas. Outra observação a partir dos resultados da Tabela 19 e 20 é relacionada à técnica *Extreme Value*. Observando os resultados, a mesma apresenta seus melhores resultados para os níveis 4 e 5. Ainda relacionado à técnica *Extreme Value*, das seis métricas analisadas a que apresenta uma mudança considerável, quando comparados os níveis de anomalia, é a precisão no reconhecimento das anomalias. Os resultados para os níveis 1, 2 e 3 foram abaixo de 0.32, apresentando um salto para 0.71 a partir do nível 4.

A técnica de detecção de anomalias baseada em análise estatística que apresentou o melhor resultado foi a *Naive Parzen*, seguida pela *Extreme Value*, que realiza uma análise de valores extremos de uma distribuição. O mau resultado da técnica Gauss pode ser explicado pela distribuição dos dados provenientes do sensor ECT, que não seguem uma distribuição normal.

Como mostrado ao longo desta seção, os resultados são descritos em seis métricas diferentes. Apesar da importância de cada uma, a métrica escolhida para uma análise geral do sistema proposto, é o *F2-score*. Interpretada como uma média harmônica das

Tabela 20 – Média e desvio padrão das métricas para os experimentos do veículo em movimento para anomalia Tipo II - Nível 5

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.737 ± 0.010	0.799 ± 0.004	0.618 ± 0.018	0.801 ± 0.019	0.616 ± 0.016	0.754 ± 0.009
k-NN	0.823 ± 0.014	0.934 ± 0.006	0.667 ± 0.019	0.798 ± 0.030	0.868 ± 0.017	0.920 ± 0.007
Mahalanobis	0.853 ± 0.008	0.823 ± 0.008	1.000 ± 0.000	1.000 ± 0.000	0.550 ± 0.000	0.748 ± 0.015
MST	0.834 ± 0.009	0.873 ± 0.003	0.761 ± 0.018	0.876 ± 0.013	0.718 ± 0.007	0.836 ± 0.004
Naive Parzen	0.829 ± 0.008	0.941 ± 0.004	0.691 ± 0.018	0.788 ± 0.020	0.873 ± 0.016	0.926 ± 0.005
SOM	0.593 ± 0.007	0.653 ± 0.004	0.356 ± 0.014	0.800 ± 0.009	0.227 ± 0.006	0.474 ± 0.008
Parzen	0.775 ± 0.009	0.884 ± 0.005	0.636 ± 0.014	0.758 ± 0.016	0.842 ± 0.008	0.875 ± 0.006
SVDD	0.663 ± 0.013	0.900 ± 0.005	0.510 ± 0.010	0.543 ± 0.020	0.724 ± 0.013	0.858 ± 0.005
Extreme Value	0.875 ± 0.021	0.921 ± 0.019	0.772 ± 0.015	0.902 ± 0.014	0.816 ± 0.025	0.897 ± 0.034
OC-SVM	0.902 ± 0.002	0.998 ± 0.001	0.235 ± 0.004	0.901 ± 0.002	0.943 ± 0.001	0.986 ± 0.001

métricas de precisão e sensibilidade, a mesma torna-se uma boa avaliadora, ao considerar acerto de ambos tipos de dado. Das diversas técnicas utilizadas, a melhor, para os cinco níveis de anomalias, foi a OC-SVM. Com o objetivo de detalhar o resultado da criação de um modelo do sistema proposto usando a técnica OC-SVM, a Figura 50 (a) mostra uma matriz de confusão para o nível de anomalia 4 e a (b) para o nível 5.

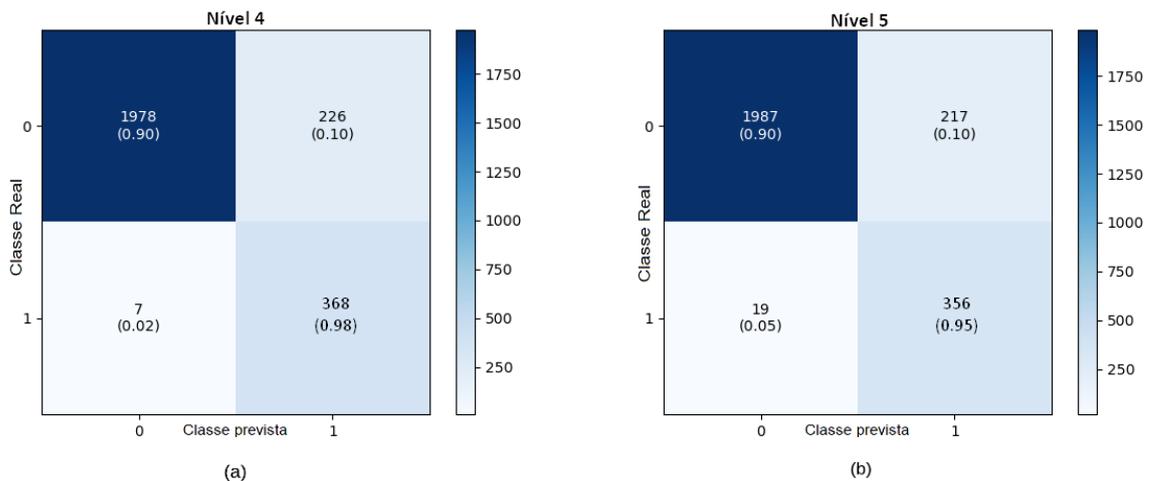


Figura 50 – Matriz de confusão para a técnica OC-SVM (a) Nível de anomalia 4 e (b) Nível de anomalia 5.

A Figura 52 mostra os *boxplots* da métrica F2-score para os níveis de anomalia 4 (a) e 5 (b), respectivamente. Como todos os resultados obtidos são descritos em termos de desvio padrão e valores médios, é utilizado o teste estatístico não-paramétrico de Friedman, para verificar os resultados. O teste de hipótese é realizado com 99% de confiança, para certificar se existe diferença estatística entre cada classificador, considerando o F2-score. A hipótese nula (H_0) do teste é que os dados são estatisticamente iguais. O teste realiza os testes de desempenho em pares.

O resultado, ilustrado na Figura 51, mostrou que no nível 4, apenas as técnicas SOM e OC-SVM apresentam diferenças estatísticas entre si. De maneira diferente, para o quinto nível existe diferença estatística entre as técnicas Mahalanobis e Naive Parzen, Naive

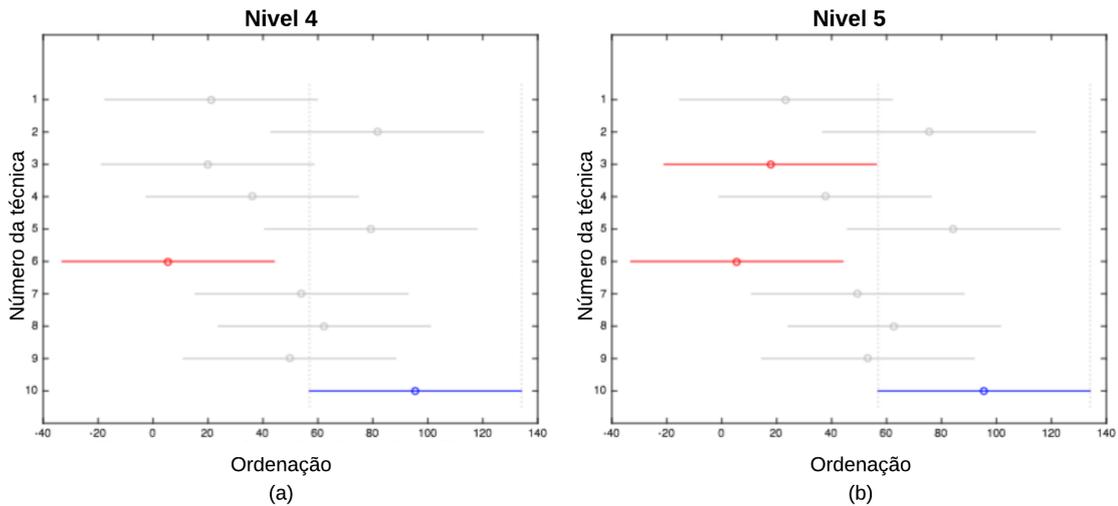


Figura 51 – Resultado do teste de Friedman para diferenças estatísticas entre as técnicas utilizadas no modo em movimento (a) Nível de Anomalia 4 (b) Nível 5. O índice indica o numero da técnica analisada, seguindo a ordem mostrada nas tabelas.

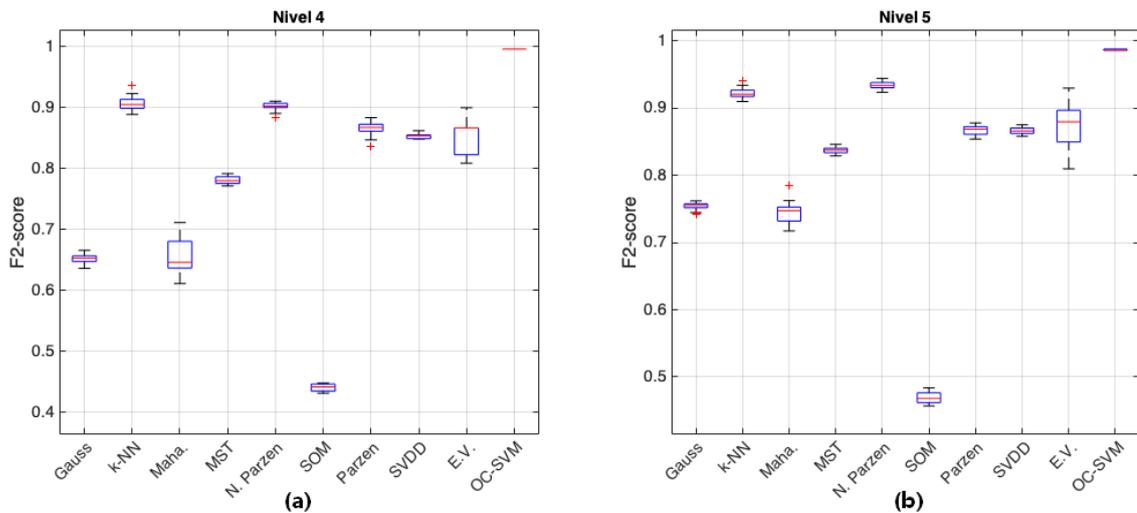


Figura 52 – *Boxplot* do F2-score para (a) Nível 4 de anomalia do veículo em movimento (b) Nível 5 de anomalia, respectivamente.

Parzem e SOM, e as técnicas SOM e Mahalanobis são estatisticamente diferentes da OC-SVM.

Para confirmar o resultado, o teste *post-hoc* de Nemenyi é usado para confirmar as melhores técnicas baseadas em somas classificadas, os *ranks*. A Tabela 21 mostra os resultados dos *ranks* para os níveis de anomalia 4 e 5 do modo do veículo em movimento. Os resultados comprovam a técnica OC-SVM como a melhor para ambos níveis de anomalia, seguidos da técnica k-NN, para o nível 4 e Naive Parzen no nível 5.

Na comparação dos resultados deste trabalho com os de (THEISSLER, 2017), ambos

Tabela 21 – Resultado do teste *post-hoc* de Nemenyi para o modo em movimento

Técnica	Rank Nível 4	Rank Nível 5
Gauss	21.100	23.333
k-NN	81.533	75.400
Mahalanobis	19.900	17.667
MST	36.133	37.600
Naive Parzen	79.233	84.400
SOM	5.500	5.500
Parzen	54.067	49.567
SVDD	62.333	62.800
Extreme Value	49.700	53.233
OC-SVM	95.500	95.500

resultados convergem para a OC-SVM ser a melhor técnica para o modo do veículo em movimento. O valor médio do *F2-score* obtido é 0.987 para os cinco diferentes níveis de anomalia, contra o valor de 0.806 apresentados pelo autor. A Tabela 22 compara os resultados dos quatro algoritmos utilizados pelo autor que também foram analisados neste trabalho. Para a comparação, os resultados deste trabalho são baseados na média dos cinco diferentes níveis de anomalia.

Tabela 22 – Comparação dos resultados do *F2-score* para o modo do veículo em movimento

Técnica	(THEISSLER, 2017)	Resultado deste trabalho
Mahalanobis	0.462	0.362
SVDD	0.806	0.735
One-class SVM	0.806	0.987
Extreme Value	0.658	0.423

Por considerar a média do *F2-score* dos cinco níveis de anomalia, o resultado mostrado na Tabela 22 determina apenas a técnica OC-SVM com resultado geral melhor. Porém, as técnicas SVDD, *Extreme Value* e Mahalanobis apresentam maior índice nos níveis 4 e 5.

Por fim, ao considerar a performance temporal dos algoritmos, os tempos de processamento para a base utilizada segue os resultados da Tabela 46. O tempo de treinamento, criação de modelo, também segue baseada na complexidade de cada técnica. Como exemplo, o algoritmo SVDD apresenta o maior tempo de criação do modelo para todos tamanhos de janela e níveis de anomalia.

5.6 EXPERIMENTOS COM O VEÍCULO EM MODO *IDLE*

5.6.1 Anomalia Tipo I

A mesma metodologia das etapas de treinamento e teste, mostrada na Figura 47, é utilizada nos experimentos em modo *idle*. De maneira diferente, o conjunto $\|A\|$ é caracterizado por 15 sinais e o $\|B\|$ por 10 sinais. Em análise similar, o valor de $k = 10$ determina

Tabela 23 – Média e desvio padrão das métricas para os experimentos do veículo em modo *idle* para anomalia Tipo I

Modo	Acurácia	Sensibilidade	Especificidade	Precisão <i>Target</i>	Precisão <i>Outlier</i>	F2-score
Movimento	0.998 ± 0.001	0.997 ± 0.001	1.000 ± 0.000	1.000 ± 0.000	0.971 ± 0.001	0.998 ± 0.001

um *threshold* de 1.12. A Tabela 23 resume os resultados do experimentos envolvendo a detecção de anomalia Tipo I no modo *idle*.

Numericamente os resultados são iguais aos do modo do veículo em movimento, mostrando a capacidade da técnica LOF para ambos os modos veiculares. As mesmas observações realizadas na Figura 49 valem para os resultados do modo *idle*. De maneira geral, a falha de desconexão do cabo utilizada neste trabalho considera o valor de fundo de escala como zero. Se considerado outro valor, visto que alguns sensores utilizam um valor de fundo de escala negativo, os mesmos resultados numéricos são esperados.

Por fim, tanto os resultados do modo em movimento quanto os do modo *idle* são idênticos aos obtidos por (THEISSLER, 2013), o que mostra que a metodologia para a criação da base de dados e inserção das anomalias foi utilizada com sucesso.

5.6.2 Anomalia Tipo II

Utilizando o mesmo procedimento da divisão da base dados descrito na Seção 5.5.2, com a diferença do grupo $\|A\|$ conter 15 *trips* e o $\|B\|$ com 10, seguindo o fluxo descrito na Figura 47, e os parâmetros específicos da Tabela 15, os resultados para o modo *idle* são apresentados nas Tabelas 24 a 28. Diferente do modo em movimento, a técnica com os melhores resultados, realizando a análise a partir do F2-score é o k-NN. Para os níveis de anomalia 2, 4 e 5 a técnica retorna o valor máximo, como mostrado nas Tabelas 25, 27 e 28.

Um resultado expressivo para o modo *idle* é referente a técnica de detecção de anomalias baseada na distância de Mahalanobis, quando a mesma apresenta valores superiores a 0.99 na precisão de detecção de dados do funcionamento normal do sensor ECT. Apesar do bom resultado na detecção de dados do tipo *target*, a técnica não repete os bons resultados para dados *outliers*, seguindo o resultado do modo do veículo em movimento.

Com exceção do quinto nível de anomalia, a técnica k-NN apresenta o melhor valor de acurácia, além do melhor F2-score em todos os níveis. Outra métrica com excelente resultado apresentado pelo k-NN é o valor da precisão dos dados anômalos, atingindo valor máximo para os níveis 2, 4 e 5. Assim como o modo anterior, o valor de *k* utilizado é igual a um. A Figura 53 (a) mostra uma matriz de confusão para o nível de anomalia 4 e a (b) para o nível 5.

Na comparação com o modo em movimento, a técnica *Extreme Value* apresenta uma melhora em sua performance geral a partir do terceiro nível, no modo em movimento esta melhora acontece no quarto nível. A melhora comum aos dois modos de operação

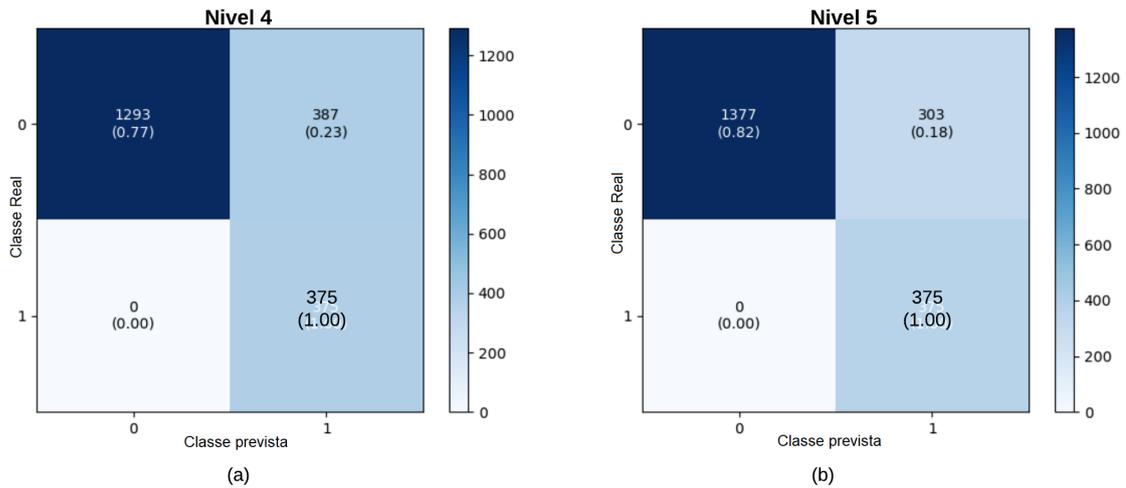


Figura 53 – Matriz de confusão para a técnica k-NN (a) Nível de anomalia 4 e (b) Nível de anomalia 5.

Tabela 24 – Média e desvio padrão das métricas para os experimentos do veículo em modo *Idle* para anomalia Tipo II - Nível 1

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.752 ± 0.016	0.920 ± 0.002	0.0886 ± 0.010	0.799 ± 0.018	0.225 ± 0.010	0.568 ± 0.020
k-NN	0.898 ± 0.017	0.957 ± 0.000	0.864 ± 0.022	0.801 ± 0.036	0.972 ± 0.000	0.960 ± 0.010
Mahalanobis	0.488 ± 0.003	0.487 ± 0.001	1.000 ± 0.000	1.000 ± 0.003	0.017 ± 0.002	0.008 ± 0.002
MST	0.653 ± 0.005	0.687 ± 0.003	0.340 ± 0.024	0.906 ± 0.009	0.109 ± 0.004	0.333 ± 0.005
Naive Parzen	0.642 ± 0.011	0.717 ± 0.004	0.414 ± 0.018	0.790 ± 0.015	0.325 ± 0.009	0.577 ± 0.010
SOM	0.623 ± 0.008	0.698 ± 0.004	0.339 ± 0.017	0.801 ± 0.010	0.217 ± 0.006	0.483 ± 0.007
Parzen	0.596 ± 0.009	0.692 ± 0.004	0.335 ± 0.012	0.740 ± 0.018	0.278 ± 0.010	0.533 ± 0.009
SVDD	0.612 ± 0.009	0.718 ± 0.005	0.386 ± 0.013	0.714 ± 0.012	0.383 ± 0.008	0.611 ± 0.008
Extreme Value	0.621 ± 0.036	0.669 ± 0.026	0.224 ± 0.310	0.877 ± 0.019	0.075 ± 0.035	0.259 ± 0.039
OC-SVM	0.880 ± 0.001	0.972 ± 0.000	0.195 ± 0.002	0.890 ± 0.000	0.488 ± 0.000	0.811 ± 0.001

é obtida por um maior valor da precisão na detecção de dados anômalos, o que pode ser explicado pelo nível de anomalia nos sinais.

Tabela 25 – Média e desvio padrão das métricas para os experimentos do veículo em modo *Idle* para anomalia Tipo II - Nível 2

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.746 ± 0.017	0.913 ± 0.002	0.0578 ± 0.007	0.800 ± 0.019	0.153 ± 0.011	0.458 ± 0.010
k-NN	0.915 ± 0.012	1.000 ± 0.000	0.869 ± 0.016	0.803 ± 0.027	1.000 ± 0.000	1.000 ± 0.000
Mahalanobis	0.503 ± 0.007	0.495 ± 0.003	0.969 ± 0.005	0.999 ± 0.002	0.028 ± 0.003	0.115 ± 0.007
MST	0.692 ± 0.008	0.719 ± 0.004	0.526 ± 0.030	0.904 ± 0.011	0.253 ± 0.006	0.525 ± 0.007
Naive Parzen	0.697 ± 0.012	0.772 ± 0.004	0.521 ± 0.021	0.791 ± 0.019	0.482 ± 0.011	0.689 ± 0.010
SOM	0.619 ± 0.014	0.694 ± 0.007	0.323 ± 0.027	0.803 ± 0.019	0.206 ± 0.012	0.470 ± 0.011
Parzen	0.647 ± 0.010	0.738 ± 0.006	0.437 ± 0.015	0.751 ± 0.019	0.433 ± 0.010	0.646 ± 0.009
SVDD	0.606 ± 0.024	0.718 ± 0.006	0.384 ± 0.018	0.701 ± 0.046	0.393 ± 0.027	0.616 ± 0.022
Extreme Value	0.631 ± 0.038	0.675 ± 0.010	0.275 ± 0.018	0.883 ± 0.015	0.094 ± 0.041	0.303 ± 0.043
OC-SVM	0.880 ± 0.001	0.972 ± 0.000	0.195 ± 0.001	0.890 ± 0.001	0.492 ± 0.001	0.813 ± 0.001

Apesar de um alto valor do F2-score, a técnica OC-SVM não repete os bons resultados devido a sua baixa taxa de acerto dos dados anômalos. O resultado pode ser explicado por um diferente comportamento, forma do sinal, do sensor ECT para o veículo em modo

idle e em movimento, como ilustrado na Figura 54. O tipo do sinal para o modo *idle* apresenta menor variação geral, tendendo a ter um valor com menor variação após um determinado intervalo de tempo, em geral 500 segundos.

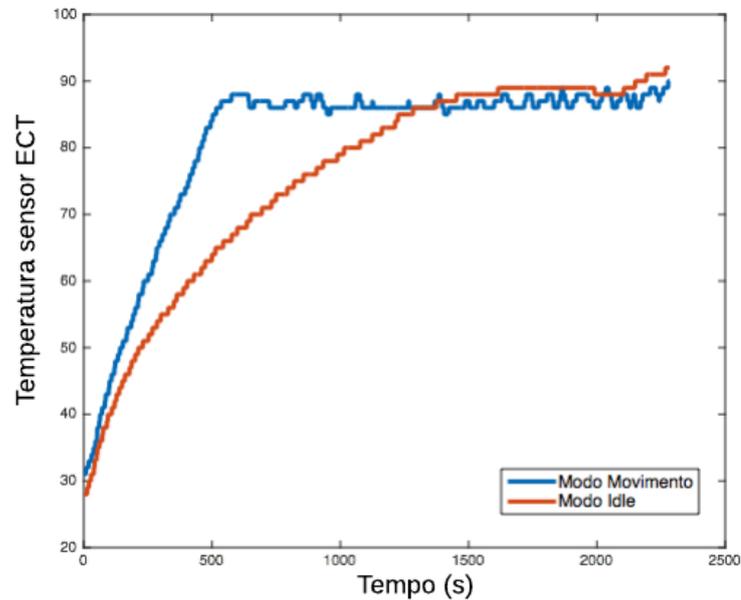


Figura 54 – Comparação do sinal ECT para modo *idle* e modo do veículo em movimento.

Tabela 26 – Média e desvio padrão das métricas para os experimentos do veículo em modo *Idle* para anomalia Tipo II - Nível 3

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.767 ± 0.016	0.935 ± 0.003	0.143 ± 0.013	0.801 ± 0.018	0.375 ± 0.010	0.720 ± 0.020
k-NN	0.895 ± 0.014	0.941 ± 0.019	0.867 ± 0.012	0.808 ± 0.031	0.983 ± 0.017	0.949 ± 0.002
Mahalanobis	0.572 ± 0.012	0.532 ± 0.007	0.993 ± 0.018	0.998 ± 0.004	0.142 ± 0.005	0.343 ± 0.007
MST	0.717 ± 0.008	0.738 ± 0.003	0.604 ± 0.029	0.91 ± 0.010	0.318 ± 0.006	0.583 ± 0.006
Naive Parzen	0.741 ± 0.009	0.821 ± 0.004	0.584 ± 0.017	0.796 ± 0.014	0.617 ± 0.008	0.770 ± 0.007
SOM	0.602 ± 0.013	0.682 ± 0.007	0.272 ± 0.030	0.796 ± 0.021	0.172 ± 0.011	0.428 ± 0.010
Parzen	0.653 ± 0.011	0.747 ± 0.006	0.451 ± 0.016	0.746 ± 0.018	0.462 ± 0.010	0.665 ± 0.008
SVDD	0.676 ± 0.041	0.814 ± 0.016	0.494 ± 0.040	0.683 ± 0.059	0.649 ± 0.040	0.774 ± 0.037
Extreme Value	0.787 ± 0.023	0.749 ± 0.017	0.855 ± 0.015	0.900 ± 0.011	0.663 ± 0.016	0.729 ± 0.022
OC-SVM	0.880 ± 0.001	0.972 ± 0.001	0.193 ± 0.002	0.901 ± 0.001	0.485 ± 0.001	0.809 ± 0.001

A técnica com o pior resultado geral, devido a baixa detecção de *outliers* em ambos modos veiculares, foi o SOM. Com funcionamento baseado em mapeamento a partir de rede neural, que transforma dados em espaços de dimensões diferentes, uma grade ortogonal $N \times N$ é definida para uma análise baseada em *clusters*. Como discutido por (TAX, 2001), o uso da técnica com um baixo número de nós se comporta de maneira semelhante ao *K-means*. Outro fator que também influencia os maus resultados é o processo de janelamento aplicado aos dados, dificultando a análise em vizinhança.

Na comparação entre as técnicas *Naive Parzen* e *Parzen*, a que apresentou o melhor resultado em todos os níveis de anomalias, em ambos os modos, foi a *Naive Parzen*. Na

Tabela 27 – Média e desvio padrão das métricas para os experimentos do veículo em modo *Idle* para anomalia Tipo II - Nível 4

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.766 ± 0.025	0.935 ± 0.002	0.144 ± 0.018	0.801 ± 0.028	0.325 ± 0.016	0.679 ± 0.024
k-NN	0.909 ± 0.011	1.000 ± 0.000	0.861 ± 0.013	0.790 ± 0.022	1.000 ± 0.000	1.000 ± 0.000
Mahalanobis	0.685 ± 0.013	0.608 ± 0.010	0.996 ± 0.007	0.999 ± 0.003	0.412 ± 0.012	0.555 ± 0.016
MST	0.801 ± 0.008	0.822 ± 0.006	0.737 ± 0.021	0.905 ± 0.010	0.580 ± 0.006	0.758 ± 0.009
Naive Parzen	0.782 ± 0.011	0.877 ± 0.004	0.627 ± 0.018	0.792 ± 0.016	0.767 ± 0.009	0.852 ± 0.005
SOM	0.613 ± 0.019	0.689 ± 0.006	0.305 ± 0.024	0.799 ± 0.020	0.212 ± 0.011	0.475 ± 0.022
Parzen	0.739 ± 0.012	0.855 ± 0.005	0.567 ± 0.017	0.745 ± 0.019	0.758 ± 0.011	0.833 ± 0.008
SVDD	0.713 ± 0.021	0.855 ± 0.007	0.534 ± 0.031	0.699 ± 0.048	0.732 ± 0.011	0.827 ± 0.006
Extreme Value	0.879 ± 0.018	0.872 ± 0.015	0.889 ± 0.008	0.905 ± 0.015	0.851 ± 0.018	0.867 ± 0.027
OC-SVM	0.882 ± 0.001	0.974 ± 0.001	0.204 ± 0.001	0.900 ± 0.002	0.518 ± 0.001	0.828 ± 0.003

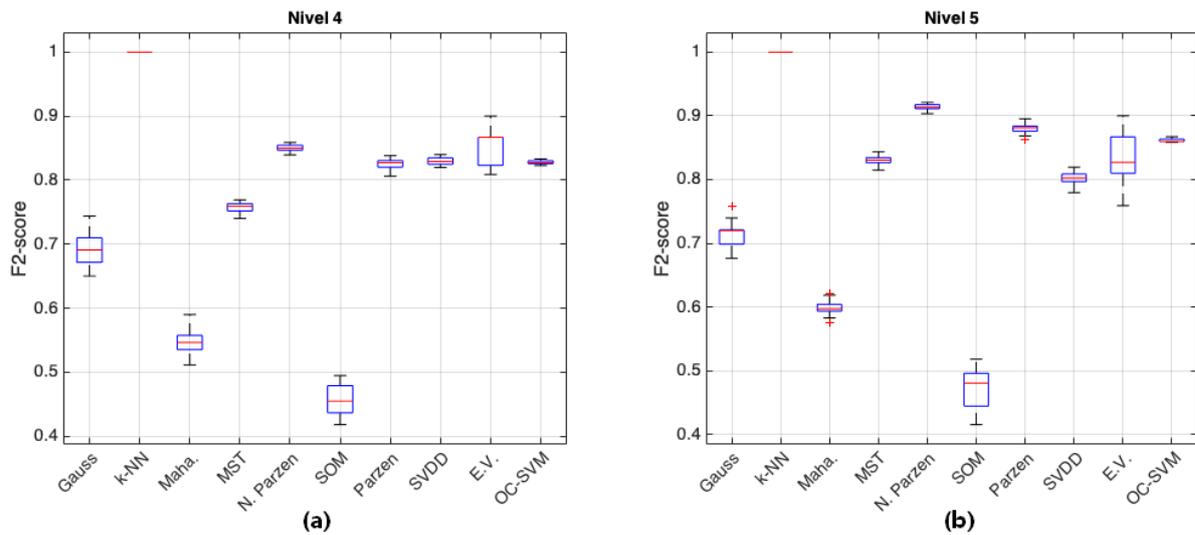


Figura 55 – Boxplot do F2-score para (a) Nível 4 de anomalia do veículo em modo *idle* (b) Nível 5 de anomalia, respectivamente.

análise dos diferentes níveis de anomalia, a métrica de precisão de acerto de dados do funcionamento normal é maior que a precisão de anomalias, com exceção para o Nível 5. Ao usar *Naive Parzen*, o melhor resultado pode ser explicado pela estimativa de densidade total, calculada pelas densidades de cada atributo.

Na análise individual da técnica MST, seu melhor resultado é apresentado para o modo do veículo em movimento, com o mesmo correndo também para as técnicas OC-SVM, *Extreme Value*, SVDD, Mahalanobis, SOM e Parzen. A comparação geral entre os dois modos de operação, nos Níveis 4 e 5, é mostrada na Figura 56 e 57.

Seguindo a metodologia da análise do modo do veículo em movimento, é realizado o teste estatístico não-paramétrico de Friedman, para analisar diferenças estatísticas entre os resultados do F2-score das técnicas. Como mostrado na Figura 58 (a), para o Nível 4 existe apenas diferença estatística entre as técnicas k-NN e Mahalanobis, k-NN e SOM. Para o quinto nível, além das diferenças apresentadas no nível 4, também ocorre diferença entre *Naive Parzen* e SOM.

Tabela 28 – Média e desvio padrão das métricas para os experimentos do veículo em modo *Idle* para anomalia Tipo II - Nível 5

Técnica	Acurácia	Sensibilidade	Especificidade	Precisão Target	Precisão Outlier	F2-score
Gauss	0.766 ± 0.010	0.935 ± 0.018	0.141 ± 0.002	0.801 ± 0.018	0.375 ± 0.021	0.720 ± 0.022
k-NN	0.916 ± 0.009	1.000 ± 0.000	0.871 ± 0.004	0.807 ± 0.012	1.000 ± 0.000	1.000 ± 0.000
Mahalanobis	0.727 ± 0.008	0.641 ± 0.007	0.997 ± 0.010	0.998 ± 0.005	0.474 ± 0.004	0.598 ± 0.010
MST	0.840 ± 0.005	0.869 ± 0.004	0.769 ± 0.019	0.902 ± 0.012	0.672 ± 0.004	0.820 ± 0.007
Naive Parzen	0.817 ± 0.009	0.927 ± 0.003	0.660 ± 0.015	0.794 ± 0.014	0.859 ± 0.008	0.912 ± 0.004
SOM	0.616 ± 0.024	0.692 ± 0.008	0.317 ± 0.033	0.799 ± 0.016	0.196 ± 0.010	0.459 ± 0.029
Parzen	0.767 ± 0.010	0.898 ± 0.005	0.596 ± 0.015	0.745 ± 0.017	0.813 ± 0.009	0.879 ± 0.007
SVDD	0.694 ± 0.015	0.832 ± 0.011	0.513 ± 0.033	0.691 ± 0.054	0.724 ± 0.036	0.807 ± 0.009
Extreme Value	0.934 ± 0.022	0.803 ± 0.013	0.974 ± 0.018	0.903 ± 0.010	0.942 ± 0.019	0.827 ± 0.033
OC-SVM	0.885 ± 0.002	0.977 ± 0.001	0.223 ± 0.001	0.890 ± 0.001	0.587 ± 0.000	0.862 ± 0.002

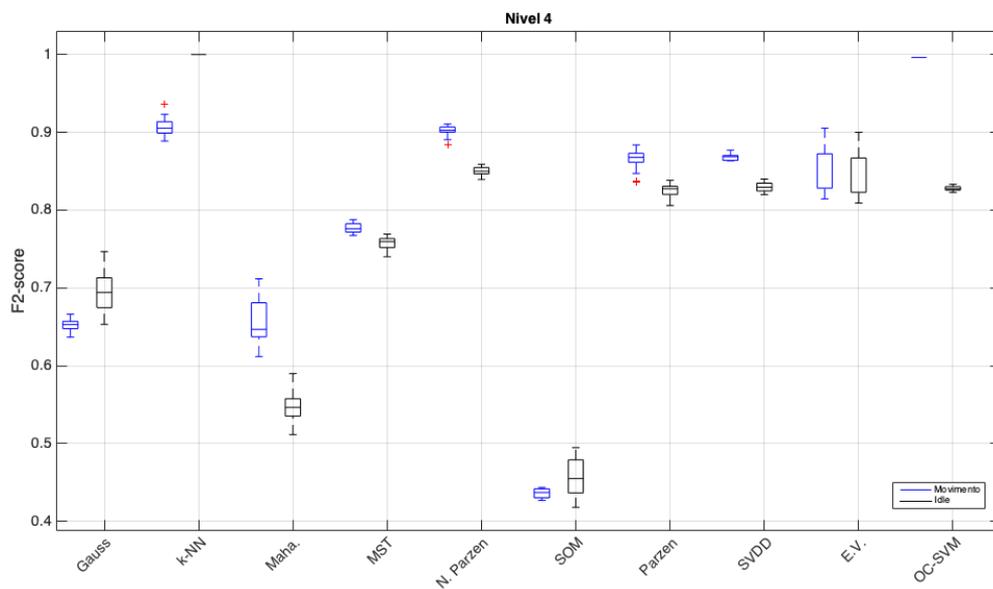


Figura 56 – *Boxplot* de comparação entre os modos *Idle* e em movimento para o Nível de anomalia 4.

Assim como no modo anterior, é realizado o teste *post-hoc* de Nemenyi para ordenar as técnicas. Com o resultado mostrado na Tabela 29, define-se a ordem das melhores técnicas. Na comparação direta da média dos resultados do modo *idle* deste trabalho com os de (THEISSLER, 2017), Tabela 30, as técnicas OC-SVM e SVDD apresentam melhores resultados. Se comparado de maneira individual, o algoritmo SVDD apresenta melhor *F2-score* a partir do terceiro nível, a OC-SVM em todos os níveis.

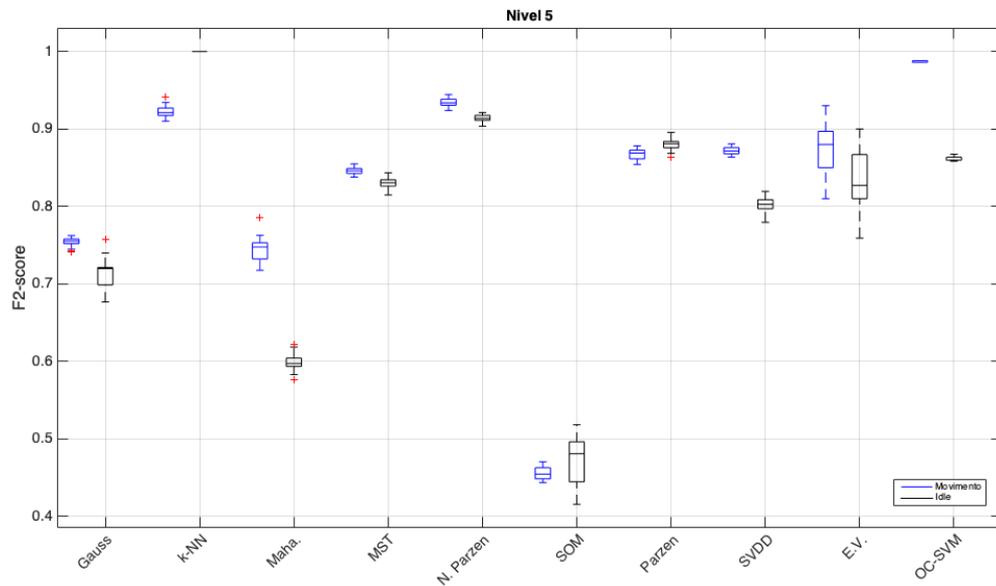


Figura 57 – *Boxplot* de comparação entre os modos *Idle* e em movimento para Nível de anomalia 5.

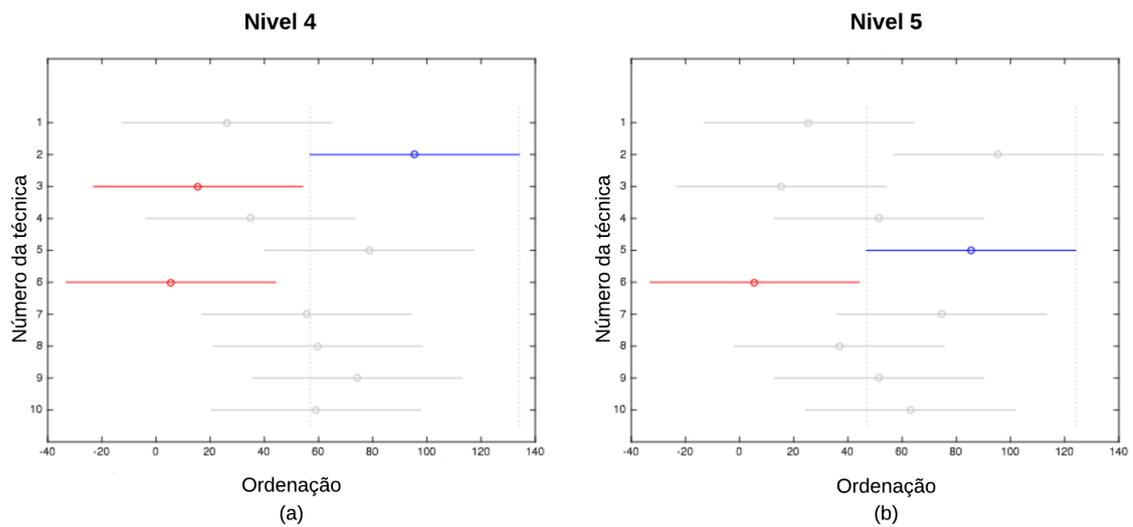


Figura 58 – Resultado do teste de Friedman para diferenças estatísticas entre as técnicas utilizadas no modo *idle* (a) Nível de Anomalia 4 (b) Nível 5. O índice indica o numero da técnica analisada, seguindo a ordem mostrada nas tabelas.

Tabela 29 – Resultado do teste *post-hoc* de Nemenyi para o modo *idle*

Técnica	Rank Nível 4	Rank Nível 5
Gauss	26.167	25.500
k-NN	95.500	95.500
Mahalanobis	15.500	15.500
MST	34.833	51.500
Naive Parzen	78.800	85.500
SOM	5.500	5.500
Parzen	55.633	74.600
SVDD	59.733	36.800
Extreme Value	74.2667	51.533
OC-SVM	59.067	63.067

Tabela 30 – Comparação dos resultados do F2-score para o modo *idle*.

Técnica	(THEISLER, 2017)	Resultado deste trabalho
Mahalanobis	0.818	0.324
SVDD	0.638	0.727
One-class SVM	0.600	0.825
Extreme Value	0.865	0.597

5.7 EXPERIMENTOS COM O VEÍCULO EM CICLOS DE DIREÇÃO EM TRAJETOS ESPECÍFICOS

Como descrito anteriormente na Tabela 8, os ciclos de direção do modo do veículo em movimento são separados, para posterior análise individual. Seguindo o mesmo processo de divisão das bases para anomalia Tipo II das seções anteriores, é realizada a avaliação do sistema para cada trajeto específico.

Na comparação global a partir do valor do *F2-score*, a Tabela 31 apresenta a melhor técnica para cada trajeto específico nos diferentes níveis de anomalia. Ao comparar os resultados do veículo em trajeto específicos com o modo em movimento, nota-se que para um mesmo trajeto, a técnica com melhor *F2-score* varia com o nível de anomalia. O resultado mostra que a escolha da melhor técnica envolve a determinação de um conjunto ótimo para a etapa de treinamento.

A análise dos resultados dos trajetos individuais mostra que existe uma diferença entre a melhor técnica para o quarto e quinto nível de anomalia, sendo o OC-SVM comum aos três primeiros níveis como a melhor técnica. Outra observação interessante relacionada ao resultado é a presença de trajetos mesclados, isto é, com partes de outros. Um exemplo de um trajeto como este é o fixo, onde parte de cada *trip* deste trajeto é realizada em área urbana e a outra parte em via expressa. Da mesma forma, partes iniciais dos ciclos dos trajetos de vias expressas contém dados de rápida passagem em via urbana.

Finalmente, os resultados dos trajetos específicos mostram a necessidade de uma ferramenta que realize uma classificação prévia do sinal veicular que entra no sistema de detecção de anomalias. Sendo assim, possível determinar qual sua respectiva faixa de potência e se aquele funcionamento corresponde ao veículo em baixas, médias ou altas velocidades.

Tabela 31 – Resultado da melhor técnica para cada trajeto e nível de anomalia, com seu respectivo valor de *F2-score*

Trajeto Especifico	Nível 1	Nível 2	Nível 3	Nível 4	Nível 5
Universidade Campus	OC-SVM (0.97)	OC-SVM (0.97)	OC-SVM (0.95)	OC-SVM (0.98)	k-NN (0.98)
Fixo	OC-SVM (0.91)	OC-SVM (0.92)	OC-SVM (0.95)	Naive Parzen (0.97)	Naive Parzen (0.99)
Vias Expressas	OC-SVM (0.98)	OC-SVM (0.98)	OC-SVM (0.98)	OC-SVM (0.98)	k-NN (0.96)
Areas Urbanas	OC-SVM (0.96)	OC-SVM (0.92)	OC-SVM (0.98)	OC-SVM (0.97)	k-NN (0.97)

6 CONCLUSÃO

A crescente demanda por funcionalidades veiculares inovadoras baseadas na experiência do carro como serviço evidenciam a necessidade de sistemas inteligentes automotivos, sendo estes capazes de detectar estados do funcionamento veicular. Dessa maneira, novos serviços podem ser fornecidos ao condutor com o objetivo de evitar danos aos sistemas veiculares. Com o objetivo de detectar comportamento anômalo de um dos sistemas automotivos mais importantes, o de arrefecimento veicular, este trabalho avaliou a classificação de sinais provenientes do sensor do líquido de arrefecimento, capturados a partir de um sistema de telemetria proposto. O processo de detecção de anomalias foi baseado em dois tipos diferentes de falhas associados ao sensor ECT.

A primeira abordagem para detecção de um comportamento não esperado foi associada à desconexão do cabo de conexão do sensor a sua ECU correspondente. Realizando uma comparação com a técnica mais simples e eficiente, apresentada pelo estado da arte, os resultados deste trabalho mostraram a eficácia da aplicação da técnica *Linear Outlier Factor* na base de dados gerada. Um resultado significativo para a técnica é a independência ao modo de operação veicular. A análise realizada de maneira separada, considerando o veículo parado e em movimento, apresentou praticamente o mesmo resultado em ambos modos.

Outro problema comum, associado a componentes automotivos, é a presença de falhas em sua estrutura eletrônica ou mecânica, ocasionando assim, um mau funcionamento. No caso de um sensor eletrônico o mau funcionamento é associado com uma medição errônea do sensor veicular. Ao considerar o problema neste trabalho, dez diferentes técnicas de detecção de anomalias baseadas em classificadores *one-class* são avaliadas, para os dois respectivos modos de operação do veículo. Como diferencial da proposta deste trabalho, a análise das técnicas foi realizada a partir de diferentes níveis de mau funcionamento, baseado em possível falha resistiva, do sensor veicular. Os resultados mostrados ao longo do trabalho mostram bons resultados apresentados nos cinco diferentes níveis de anomalias definidos.

Ao considerar o modo de operação *idle*, onde o veículo opera com motor ligado, porém parado, os resultados apresentados mostram a técnica *k-NN* como a de melhor desempenho. A técnica apresentou o melhor *F2-score*, índice que avalia o resultado de acerto de dados do funcionamento normal e anômalo como uma média geométrica. Outra métrica com excelente resultado para o modo é a precisão na detecção de anomalias, alcançando valor máximo em três dos cinco níveis de anomalia resistivas definidos.

Os resultados obtidos para o modo do veículo em movimento mostram a técnica OC-SVM com *kernel* polinomial de terceiro grau, com os melhores resultados para todos os cinco níveis de anomalia. Particular a técnica, a precisão de detecção de dados anômalos

possui média geral nos cinco níveis muito próxima a um. Outro resultado apresentado é referente a avaliação das diferentes técnicas para diferentes trajetos específicos do modo do veículo em movimento. Em contraste com os resultados do modo em movimento contendo os trajetos integrados, a avaliação em trajetos específicos apresentou diferentes técnicas com o melhor resultado para os dois últimos níveis de anomalia.

Ao considerar diferentes níveis de mau funcionamento, descritos pelos níveis de anomalias, a performance do sistema de detecção possui, de fato, melhor acurácia de detecção quanto maior for o seu correspondente nível de anomalia. Os resultados mostrados ao longo deste trabalho mostram a predominância da performance de uma mesma técnica para todos os níveis. A depender da métrica avaliada, o resultado da melhor técnica para o respectivo nível de anomalia pode variar. Além disso, os resultados mostram diferentes definições da melhor técnica, quando analisada a acurácia e *F2-score*. Ao final da avaliação do sistema, mesmo em sinais de sensor veicular com baixa componente de comportamento anômalo, torna-se possível a detecção do comportamento não esperado com bons resultados de acurácia e *F2-score*.

Quando comparada a mesma métrica para um modo em diferentes níveis de anomalia, apenas os resultados de trajetos específicos, mostrado na Tabela 31, mostram a melhor técnica diferente para os níveis de anomalia 4 e 5. Concluindo assim, o uso de uma única técnica para análise em diferentes níveis de mau funcionamento. Por fim, em comparação com os resultados gerais de (THEISSLER, 2017), para os dois modos veiculares, este trabalho apresentou resultados para outras técnicas de classificação de uma classe, além de melhores resultados na performance do sistema de detecção.

6.1 CONTRIBUIÇÕES

O presente trabalho apresenta as seguintes contribuições:

1. Um mapeamento sistemático sobre as diversas propostas de sistemas de telemetria veicular, suas arquiteturas de *hardware*, sensores utilizados e respectiva área de aplicação da proposta.
2. A criação de uma base de dados veicular com 27 diferentes parâmetros provenientes da interface OBD-II de um único veículo durante um período de seis meses. A base de dados é resultado de dados capturados através de dois diferentes modos veiculares. Ao considerar o veículo em movimento, quatro tipos diferentes de trajetórias são registradas. A base de dados encontra-se disponível em (NETO, 2019).
3. A proposta de uma maneira de inserção de anomalias em sensores resistivos baseados na faixa dinâmica de potência do sinal do sensor. O processo é baseado na caracterização de uma anomalia como ruído em excesso com alta potência.

4. Avaliação de dez diferentes classificadores *one-class* para detecção de anomalias no sinal do sensor ECT em cinco diferentes níveis de anomalia.
5. Submissão de artigo para a conferência *Vehicular Technology Conference, VTC2019 Fall*.

6.2 DIFICULDADES

As dificuldades apresentadas neste trabalho estiveram presentes no processo e validar as técnicas de detecção de anomalias para os sinais veiculares. Na medida que não existiam base de dados correspondentes a sinais provenientes da OBD-II na época da elaboração da proposta, foi necessária a montagem de uma base dedicada ao objetivo. Além da não disponibilidade de uma base com informações básicas, de maneira organizada, da OBD-II, também não foi possível encontrar base de dados com funcionamento anômalo de peças e sistemas veiculares. Tal dificuldade justifica a necessidade da proposta de inserção de anomalias baseada em faixa dinâmica de potencia de sensor veicular. Por ultimo, a avaliação de algumas técnicas no *software* MATLAB tornou-se lenta. A etapa de treinamento para as técnicas SVDD e MST demandaram muito esforço computacional.

6.3 TRABALHOS FUTUROS

A partir da metodologia utilizada neste trabalho e dos resultados apresentados, alguns direcionamentos podem ser realizados para a sequência deste trabalho de mestrado.

Considerando os resultados apresentados pelos trajetos específicos do veículo, uma etapa de pré-classificação do sinal veicular pode ser realizada. O objetivo principal desta etapa é a identificação do estado do veículo, podendo esta análise ser baseada em informações do motor, considerando velocidade e rotações por minuto. Dessa forma, a aplicação de janelas de média móvel ou limiar podem simplificar a definição do tipo de trajeto percorrido pelo veículo numa determinada janela de tempo de um ciclo veicular.

A fim de analisar aspectos práticos de implementação, outra linha para a continuidade do trabalho é o processo de implementar as melhores técnicas de aprendizado em um sistema embarcado de tempo real, sendo este baseado em uma unidade micro-controlada mais próxima de uma ECU comercial. Diferentes técnicas da portabilidade dos melhores algoritmos de classificação de uma classe podem ser analisadas, além da análise temporal e de requisitos funcionais do processo. Por fim, também podem ser exploradas outras técnicas de uma classe baseadas em redes neurais.

REFERÊNCIAS

- AGGARWAL, C. C. Outlier analysis. In: SPRINGER. *Data mining*. [S.l.], 2015. p. 237–263.
- ANSCOMBE, F. J. Rejection of outliers. *Technometrics*, Taylor & Francis Group, v. 2, n. 2, p. 123–146, 1960.
- BARCELOS, V. P.; AMARANTE, T. C.; DRURY, C. D.; CORREIA, L. H. A. Vehicle monitoring system using iee 802.11p device and android application. In: *2014 IEEE Symposium on Computers and Communications (ISCC)*. [S.l.: s.n.], 2014. p. 1–7. ISSN 1530-1346.
- BARNES, C. R.; TONEY, E. G.; JAROMCZYK, J. W. Comparison of network architectures for a telemetry system in the solar car project. In: *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*. [S.l.: s.n.], 2016. p. 751–755.
- BREUNIG, M. M.; KRIEGEL, H.-P.; NG, R. T.; SANDER, J. Lof: identifying density-based local outliers. In: ACM. *ACM sigmod record*. [S.l.], 2000. v. 29, n. 2, p. 93–104.
- CARLOOP.IO. *Carloop: make apps for your car*. 2018. (acessado novembro 2018). Disponível em: <<https://www.carloop.io/>>.
- CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, ACM, v. 41, n. 3, p. 15, 2009.
- CHEVROLET. *Diagnóstico do Veículo com o sistema OnStar*. 2018. Disponível em: <<https://bit.ly/2LHFGAO>>.
- CIN-UFPE/FCA. *Pesquisa de experiência do consumidor ao usar um carro conectado*. [S.l.], 2018.
- COMATRELECO. *O que é um sistema de telemetria*. 2018. Disponível em: <<https://www.comatreleco.com.br/telemetria-sistema-telemetria/>>.
- CROLLA, D.; MASHADI, B. *Vehicle Powertrain Systems*. [S.l.]: John Wiley & Sons, 2011.
- DEVI, Y. U.; RUKMINI, M. S. S. Iot in connected vehicles: Challenges and issues — a review. In: *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPE)*. [S.l.: s.n.], 2016. p. 1864–1867.
- ELECTRONICS, C. *OBD2 EXPLAINED - A SIMPLE INTRO*. 2018. Disponível em: <<https://www.csselectronics.com/screen/page/simple-intro-obd2-explained/language/en>>.
- ESPÍNDOLA, R.; EBECKEN, N. On extending f-measure and g-mean metrics to multi-class problems. *WIT Transactions on Information and Communication Technologies*, WIT Press, v. 35, 2005.

Exponent Inc. *Analysis of Toyota ETCS-i System Hardware and Software*. [S.l.]: Exponent Inc., 2012.

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.

GARTNER. *Gartner Says 8.4 Billion Connected Things Will Be in Use in 2017, Up 31 Percent From 2016*. 2017. Disponível em: <<https://gtnr.it/2Mcqz56>>.

GRAY, P. R.; HURST, P. J.; MEYER, R. G.; LEWIS, S. H. *Analysis and design of analog integrated circuits*. [S.l.]: John Wiley & Sons, 2008.

HE, W.; YAN, G.; XU, L. D. Developing vehicular data cloud services in the iot environment. *IEEE Transactions on Industrial Informatics*, v. 10, n. 2, p. 1587–1595, May 2014. ISSN 1551-3203.

HUSNI, E.; HERTANTYO, G. B.; WICAKSONO, D. W.; HASIBUAN, F. C.; RAHAYU, A. U.; TRIAWAN, M. A. Applied internet of things (iot): Car monitoring system using ibm bluemix. In: *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*. [S.l.: s.n.], 2016. p. 417–422.

KVASER. *The CAN Protocol Tour: CAN Physical Layers*. 2017. Disponível em: <<https://www.kvaser.com/about-can/the-can-protocol/can-physical-layers/>>.

KVASER. *CAN Bit Timing*. 2018. Disponível em: <<https://www.kvaser.com/about-can/the-can-protocol/can-bit-timing/>>.

LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007.

MACCHINA. *Introduction to M2*. 2018. Disponível em: <<https://www.macchina.cc/m2-introduction>>.

MALEK, K.; SOLARZ, A.; POLLO, A.; FRITZ, A.; GARILLI, B.; SCODEGGIO, M.; IOVINO, A.; GRANETT, B.; ABBAS, U.; ADAMI, C. et al. The vimos public extragalactic redshift survey (vipers)-a support vector machine classification of galaxies, stars, and agns. *Astronomy & Astrophysics*, EDP Sciences, v. 557, p. A16, 2013.

MARCHETTI, M.; STABILI, D.; GUIDO, A.; COLAJANNI, M. Evaluation of anomaly detection for in-vehicle networks through information-theoretic algorithms. In: *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*. [S.l.: s.n.], 2016. p. 1–6.

MITCHELL, T. M. et al. *Machine learning*. WCB. [S.l.]: McGraw-Hill Boston, MA., 1997.

MTE-THOMSON. *Como funciona o sensor de temperatura*. 2018. Disponível em: <<https://bit.ly/2GKYz7e>>.

MüTER, M.; GROLL, A.; FREILING, F. C. A structured approach to anomaly detection for in-vehicle networks. In: *2010 Sixth International Conference on Information Assurance and Security*. [S.l.: s.n.], 2010. p. 92–98.

- NAIR, V. V.; KOUSTUBH, B. P. Data analysis techniques for fault detection in hybrid/electric vehicles. In: *2017 IEEE Transportation Electrification Conference (ITEC-India)*. [S.l.: s.n.], 2017. p. 1–5.
- NARAYANAN, S. N.; MITTAL, S.; JOSHI, A. Obd_securealert: An anomaly detection system for vehicles. In: IEEE. *Smart Computing (SMARTCOMP), 2016 IEEE International Conference on*. [S.l.], 2016. p. 1–6.
- NAVET, N.; SIMONOT-LION, F. *Automotive embedded systems handbook*. [S.l.]: CRC press, 2008.
- NAVET, N.; SIMONOT-LION, F. A review of embedded automotive protocols. *Automotive Embedded Systems Handbook, Industrial Information Technology Series*, pages, p. 4–1, 2008.
- NAVET, N.; SIMONOT-LION, F. *Automotive embedded systems handbook*. [S.l.]: CRC press, 2009.
- NETO, E. da S. *Base de dados carOBD*. 2019. Disponível em: <<https://github.com/eron93br/carOBD/tree/master/obdiidata>>.
- NIRMALI, B.; WICKRAMASINGHE, S.; MUNASINGHE, T.; AMALRAJ, C. R. J.; BANDARA, H. M. N. D. Vehicular data acquisition and analytics system for real-time driver behavior monitoring and anomaly detection. In: *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*. [S.l.: s.n.], 2017. p. 1–6.
- NXP. *Future Advances in Body Electronics*. 2018. Disponível em: <<https://bit.ly/2wDKpfY>>.
- OBD-CODES. *ECT (Engine Coolant Temperature) Sensor DTCs*. 2018. Disponível em: <<https://bit.ly/2Cifu6o>>.
- OBDII.COM. *Does My Car Have OBD-II?* 2019. Disponível em: <<http://www.obdii.com/connector.html>>.
- PARTICLE. *Particle Electron: global cellular iot*. 2018. Disponível em: <<https://www.particle.io/cellular/>>.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- PRESTONE. *What is a Coolant Temperature Sensor?* 2018. Disponível em: <<https://www.holtsauto.com/prestone/news/coolant-temperature-sensor/>>.
- PRYTZ, R.; NOWACZYK, S.; RÖGNVALDSSON, T.; BYTTNER, S. Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data. *Engineering Applications of Artificial Intelligence*, v. 41, p. 139 – 150, 2015. ISSN 0952-1976. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0952197615000391>>.

- QUEIRÓS, J. *Sistema de Sensorização e Telemetria de um VEC (Veículo Eléctrico de Competição)*. 2011. 117 f. Tese (Doutorado) — Dissertação (Mestrado Integrado em Engenharia Electrotécnica e de Computadores Major Automação)-Faculdade de Engenharia da Universidade do Porto. Cidade do Porto, 2011.
- RAJBAHADUR, G. K.; MALTON, A. J.; WALENSTEIN, A.; HASSAN, A. E. A survey of anomaly detection for connected vehicle cybersecurity and safety. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. [S.l.: s.n.], 2018. p. 421–426. ISSN 1931-0587.
- RODRIGUES, L. C. *Fundamentos, tecnologias e aplicações de veículos autónomos*. Dissertação (B.S. thesis) — Universidade Tecnológica Federal do Paraná, 2017.
- SAE. E/e diagnostic test modes. *Vehicle E E System Diagnostic Standards Committee, J1979*, p. 1–143, 2017.
- SANTOS, E. J. *Eletrônica Analógica Integrada e Aplicações*. [S.l.]: Livraria da Física, 2011.
- SARGOLZAEI, A.; CRANE, C. D.; ABBASPOUR, A.; NOEI, S. A machine learning approach for fault detection in vehicular cyber-physical systems. In: *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. [S.l.: s.n.], 2016. p. 636–640.
- SIEGEL, J.; BHATTACHARYYA, R.; DESHPANDE, A.; SARMA, S. Vehicular engine oil service life characterization using on-board diagnostic (obd) sensor data. In: *IEEE SENSORS 2014 Proceedings*. [S.l.: s.n.], 2014. p. 1722–1725. ISSN 1930-0395.
- SOARES, F. L.; CAMPELO, D. R.; YAN, Y.; RUEPP, S.; DITTMANN, L.; ELLEGARD, L. Reliability in automotive ethernet networks. In: *2015 11th International Conference on the Design of Reliable Communication Networks (DRCN)*. [S.l.: s.n.], 2015. p. 85–86.
- STATUSSEGUROS. *Sinais que o motor do carro está fundido*. 2018. Disponível em: <<https://bit.ly/2GLOWyY>>.
- SUWATTHIKUL, J.; MCMURRAN, R.; JONES, R. In-vehicle network level fault diagnostics using fuzzy inference systems. *Applied Soft Computing*, v. 11, n. 4, p. 3709 – 3719, 2011. ISSN 1568-4946. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1568494611000615>>.
- SVENSSON, M.; BYTTNER, S.; ROGNVALDSSON, T. Self-organizing maps for automatic fault detection in a vehicle cooling system. In: IEEE. *2008 4th International IEEE Conference Intelligent Systems*. [S.l.], 2008. v. 3, p. 24–8.
- TARASSENKO, L.; HAYTON, P.; CERNEAZ, N.; BRADY, M. Novelty detection for the identification of masses in mammograms. IET, 1995.
- TAX, D. *DDtools, the Data Description Toolbox for Matlab*. 2018. Version 2.1.3.
- TAX, D. M.; DUIN, R. P. Support vector data description. *Machine learning*, Springer, v. 54, n. 1, p. 45–66, 2004.
- TAX, D. M.; DUIN, R. P. Growing a multi-class classifier with a reject option. *Pattern Recognition Letters*, Elsevier, v. 29, n. 10, p. 1565–1570, 2008.

TAX, D. M. J. *One-class classification: concept-learning in the absence of counter-examples*. Dissertação (PhD thesis) — Delft University of Technology, 2001.

THEISSLER, A. *Detecting anomalies in multivariate time series from automotive systems*. Tese (Doutorado) — Brunel University School of Engineering and Design PhD Theses, 2013.

THEISSLER, A. Detecting known and unknown faults in automotive systems using ensemble-based anomaly detection. *Knowledge-Based Systems*, v. 123, p. 163 – 173, 2017. ISSN 0950-7051. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950705117301077>>.

THEISSLER, A.; DEAR, I. Detecting anomalies in recordings from test drives based on a training set of normal instances. In: *Proceedings of the IADIS International Conference Intelligent Systems and Agents 2012 and European Conference Data Mining*. [S.l.: s.n.], 2012. p. 124–132.

THEISSLER, A.; DEAR, I. An anomaly detection approach to detect unexpected faults in recordings from test drives. In: *Proceedings of the WASET International Conference on Vehicular Electronics and Safety*. [S.l.: s.n.], 2013. p. 1144–1151.

THEISSLER, A.; DEAR, I. Autonomously determining the parameters for svdd with rbf kernel from a one-class training set. In: *Proceedings of the WASET International Conference on Machine Intelligence*. [S.l.: s.n.], 2013. v. 7, n. 7, p. 1135–1143.

TUOHY, S.; GLAVIN, M.; HUGHES, C.; JONES, E.; TRIVEDI, M.; KILMARTIN, L. Intra-vehicle networks: A review. *IEEE Transactions on Intelligent Transportation Systems*, v. 16, n. 2, p. 534–545, April 2015. ISSN 1524-9050.

VASUDEVAN, K.; DAS, A. P.; B, S.; P, S. Driver drowsiness monitoring by learning vehicle telemetry data. In: *2017 10th International Conference on Human System Interactions (HSI)*. [S.l.: s.n.], 2017. p. 270–276.

WAHLSTRÖM, J.; SKOG, I.; HÄNDEL, P. Smartphone-based vehicle telematics: A ten-year anniversary. *IEEE Transactions on Intelligent Transportation Systems*, IEEE, v. 18, n. 10, p. 2802–2825, 2017.

WANG, C.; ZHAO, Z.; GONG, L.; ZHU, L.; LIU, Z.; CHENG, X. A distributed anomaly detection system for in-vehicle network using htm. *IEEE Access*, v. 6, p. 9091–9098, 2018.

WEBER, M.; KLUG, S.; SAX, E.; ZIMMER, B. Embedded hybrid anomaly detection for automotive can communication. In: *9th European Congress on Embedded Real Time Software and Systems (ERTS 2018)*. [S.l.: s.n.], 2018.

XIE, Y.; SU, X.; HE, Y.; CHEN, X.; CAI, G.; XU, B.; YE, W. Stm32-based vehicle data acquisition system for internet-of-vehicles. In: *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. [S.l.: s.n.], 2017. p. 895–898.

YASSEIN, M. B.; SHATNAWI, M. Q.; ALJWARNEH, S.; AL-HATMI, R. Internet of things: Survey and open issues of mqtt protocol. In: *2017 International Conference on Engineering MIS (ICEMIS)*. [S.l.: s.n.], 2017. p. 1–6. ISSN 2575-1328.

ZHANG, M.; CHEN, C.; WO, T.; XIE, T.; BHUIYAN, M. Z. A.; LIN, X. Safedrive: Online driving anomaly detection from large-scale vehicle data. *IEEE Transactions on Industrial Informatics*, v. 13, n. 4, p. 2087–2096, Aug 2017. ISSN 1551-3203.

APÊNDICE A – FIRMWARE EMBARCADO E BASE DE DADOS

A.1 CÓDIGO DO *FIRMWARE* EMBARCADO RESPONSÁVEL PELA DETECÇÃO DOS PIDS

O código se encontra disponível em <https://github.com/eron93br/carOBD>.

A.2 CÓDIGO DO *FIRMWARE* EMBARCADO RESPONSÁVEL PELA COLETA DA BASE DE DADOS

O código se encontra disponível em: <https://github.com/eron93br/carOBD>.

A.3 BASE DE DADOS UTILIZADA

A base de dados utilizada neste trabalho também encontra-se disponível na URL acima.

