

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS  
GERAIS  
CAMPUS BETIM**

**OTIMIZAÇÃO DE CLASSIFICADOR DE IMPERFEIÇÕES VIÁRIAS  
EM TRECHOS DA CIDADE DE BETIM**

**RODOLFO MOREIRA DINIZ LIMA**

**Betim, julho de 2021**

**RODOLFO MOREIRA DINIZ LIMA**

**OTIMIZAÇÃO DE CLASSIFICADOR DE IMPERFEIÇÕES VIÁRIAS EM TRECHOS  
DA CIDADE DE BETIM**

Trabalho de Conclusão de Curso apresentado junto ao Curso de Engenharia de Controle e Automação Bacharelado do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - Campus Betim, como requisito parcial à obtenção do título de Bacharel em Engenharia de Controle e Automação.

Orientadores:

Prof. Me. Virgil Del Duca Almeida.

Prof. Dr. Leandro Freitas de Abreu.

**Betim, julho de 2021**

---

L732o

Lima, Rodolfo Moreira Diniz.

Otimização de classificador de imperfeições várias em trechos da cidade de Betim. [manuscrito] / Rodolfo Moreira Diniz Lima. – Betim: Instituto Federal de Minas Gerais, 2021.

62 f.; il., color.

Orientadores: Prof. Virgil Del Duca Almeida.

Monografia (Graduação) – Instituto Federal de Minas Gerais – *Campus* Betim, Bacharelado em Engenharia de Controle e Automação.

1. Pavimentos – defeitos – Betim (MG). 2. Solução de problemas. 3. Aprendizado do computador. 4. Smartphones. 5. Transformadas integrais. I. Almeida, Virgil Del Duca. II. Instituto Federal de Minas Gerais – *Campus* Betim. III. Título

CDU 625.8

---

**Catálogo: Luciana Batista Neves - CRB-6/2000**

**RODOLFO MOREIRA DINIZ LIMA**

**OTIMIZAÇÃO DE CLASSIFICADOR DE IMPERFEIÇÕES VIÁRIAS EM TRECHOS  
DA CIDADE DE BETIM**

Trabalho de Conclusão de Curso apresentado junto ao Curso de Engenharia de Controle e Automação Bacharelado do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - Campus Betim, como requisito parcial à obtenção do título de Bacharel em Engenharia de Controle e Automação.

Orientadores:

Prof. Me. Virgil Del Duca Almeida.

Prof. Dr. Leandro Freitas de Abreu.

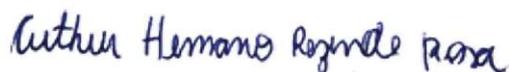
Aprovado em julho, 2021.



Prof. Me. Virgil Del Duca Almeida - Orientador



Prof. Dr. Leandro Freitas de Abreu - Coorientador



Prof. Dr. Arthur Hermano Rezende Rosa - IFMG/Campus Betim



Prof. Me. Welinton La Fontaine Lopes - IFMG/Campus Betim

## **AGRADECIMENTOS**

Agradeço em primeiro lugar a Deus, por sempre guiar meu caminho, e aos meus pais por não medirem esforços para que eu pudesse correr atrás dos meus sonhos e serem a minha base. Pelo amor, incentivo aos estudos e valores passados ao longo da vida. Grato também ao Arthur, meu irmão, pelo companherismo e amizade.

Agradeço ao Instituto Federal de Minas Gerais - IFMG, por proporcionar um ensino público, gratuito e de qualidade, e todos os professores que de alguma forma contribuíram para meu desenvolvimento acadêmico e profissional. Em especial à paciência infinita dos professores Virgil Del Duca Almeida e Leandro Freitas de Abreu, que desde o 1º período me suportaram em iniciações científicas, workshops, pesquisas aplicadas e aulas diferenciadas. Pelas orientações, por todo tempo dedicado e pelo compromisso em extrair sempre o melhor de seus alunos.

Gratidão especial à Thais, por todo apoio, amor e por compartilhar comigo os melhores momentos de minha vida. Aos meus amigos, que têm participado ativamente do meu desenvolvimento, em especial Letícia, Talita, Fabrício (Twuba) e Samantha, por terem me suportado e aconselhado desde o primeiro dia de faculdade, com meus questionamentos, dúvidas, medos. Vocês fizeram esta fase ser mais leve e divertida, e espero levá-los para além dos muros acadêmicos. A todos vocês, e aqueles que não citei, mas que participaram ativamente deste processo e não me deixaram jogar a toalha, meus sinceros agradecimentos!

## RESUMO

A massificação dos aparelhos celulares e, conseqüentemente, de seus respectivos sensores internos, possibilitou a introdução de novas tecnologias no cotidiano dos usuários. O porte destes aparelhos, no interior de veículos, permite sua utilização para classificar e detectar imperfeições nas vias onde se trafega é um claro exemplo de oportunidade às intervenções impulsionadas por métodos baseados em Inteligência Artificial em especial com o uso de Máquinas de Vetores de Suporte (SVM). Este trabalho propõe validar a funcionalidade e aplicabilidade do método de classificação de anomalias via SVM definido por (CARLOS et al., 2018) nas vias da cidade de Betim (MG, Brasil). Foi desenvolvido um estudo dos acelerômetros utilizados e as propostas que visam melhores resultados, considerando as metodologias referenciadas. Foi apresentado também as técnicas heurísticas do vetor de suporte e os possíveis ajustes no tratamentos de dados que contribuem para a assertividade da classificação, como o filtro passa-baixa, Transformada Rápida de Fourier e decomposição por Transformada de Wavelet. As saídas do algoritmo classificador detalham que o desempenho entregue pelo método é viável e aplicável nas condições de análise, aproximando-se dos resultados médios referenciados (percentual de acerto F1 de 78,5%), com assertividade encontrada de 74,1%, sendo validadas através de dados obtidos via pesquisa de campo.

**Palavras-chave:** Acelerômetro. Aparelho celular. Anomalias. Aprendizado em máquina. Classificação de vias. Filtro Passa-Baixa. SVM. Transformada de Wavelet. Transformada Rápida de Fourier.

## ABSTRACT

The massification of smartphones and, consequently, of their respective internal sensors, enabled the introduction of new technologies in the users daily lives. The use of smartphone accelerometers, allocated inside vehicles together with their drivers users, to detect and classify the routes and imperfections through which they travel is a clear example of the first steps towards interventions caused by studies and methods such as Artificial Intelligence, Machine Learning and Support Vector Machines (SVM). This work proposed a validation of the functionality and applicability of the anomaly classification via the SVM method defined by (CARLOS et al., 2018) on the roads in the city of Betim (MG, Brazil). A study of the variables used and proposed is developed aiming better results, considering the referenced methodologies. The heuristic techniques of the support vector are also presented, the possible adjustments in data processing that increase the accuracy of the classification, such as the low-pass filter, Fast Fourier Transform and decomposition made using the Wavelet Transform. The outputs details that the performance delivered by the method is feasible and applicable under the conditions of analysis, approaching the referenced average results (correct F1 percentage of 78.5%), with F1 score found as 74,1%, being validated through data obtained via real field research.

**Keywords:** Accelerometer. Anomalies. Fast Fourier Transform. Lowpass Filter. Machine learning. Roads Classification. Smartphones. SVM. Wavelet Transform.

## LISTA DE FIGURAS

Figura 1 – Exemplo de buraco . . . . .	19
Figura 2 – Exemplo de deformação . . . . .	20
Figura 3 – Exemplo de rachaduras . . . . .	20
Figura 4 – Exemplo de lombada . . . . .	21
Figura 5 – Relação dos eixos em um aparelho celular . . . . .	23
Figura 6 – Representação Inteligência Artificial . . . . .	26
Figura 7 – Representação em diagrama de blocos do sistema nervoso . . . . .	28
Figura 8 – Rede Neural x SVM . . . . .	29
Figura 9 – Orientação do aparelho celular no interior do veículo . . . . .	37
Figura 10 – Estrutura do algoritmo de nível de rugosidade do pavimento . . . . .	38
Figura 11 – Exemplo de posicionamento do aparelho celular no veículo . . . . .	40
Figura 12 – Trecho de estudo e imperfeição identificada . . . . .	40
Figura 13 – Fluxograma de Desenvolvimento . . . . .	41
Figura 14 – Exemplo de sinal em frequência [Hz] . . . . .	47
Figura 15 – Eixo Y de Rua boa . . . . .	49
Figura 16 – Eixo Y de Rua ruim . . . . .	49
Figura 17 – Eixo Y de Rua boa com buracos localizados . . . . .	50
Figura 18 – Eixo Y com o veículo parado. . . . .	51
Figura 19 – Registro de sinal de acelerômetro sem tratamento. . . . .	53
Figura 20 – Registro de sinal classificado sem filtro. . . . .	54
Figura 21 – Sinal original vs Sinal filtrado. . . . .	55
Figura 22 – Registro de sinal classificado com filtro passa-baixa. . . . .	56
Figura 23 – Espectrograma do sinal após filtro passa-baixa. . . . .	56

## LISTA DE TABELAS

Tabela 1 – Valores das Características do SVM . . . . .	42
Tabela 2 – Cálculo de Amostragem n . . . . .	52
Tabela 3 – Sensibilidade, Precisão e F1 - Medições Horizontais . . . . .	54
Tabela 4 – Sensibilidade, Precisão e F1 - Medições Verticais . . . . .	57
Tabela 5 – Sensibilidade, Precisão e F1 - Sinais Decompostos . . . . .	58

## LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
TCC	Trabalho de Conclusão do Curso
NBR	Norma Brasileira
SVM	Support Vector Machine
SVC	Support Vector Classification
IA	Inteligência Artificial
RNA	Redes Neurais Artificiais
FFT	Fast Fourier Transform
DWT	Discrete Wavelet Transform
CWT	Continuous Wavelet Transform
IoT	Internet of Things
IoV	Internet of Vehicles
CTB	Código de Trânsito Brasileiro
CONTRAN	Conselho Nacional de Trânsito

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
1.1	Objetivos	16
<b>1.1.1</b>	<b>Objetivos Específicos</b>	<b>16</b>
1.2	Estruturação Textual	17
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>18</b>
2.1	Tipos de pavimentação e suas imperfeições	18
2.2	Sensores e coleta de dados	21
2.3	IA	24
<b>2.3.1</b>	<b>Aprendizado de máquina</b>	<b>26</b>
<b>2.3.2</b>	<b>Redes neurais</b>	<b>27</b>
<b>2.3.3</b>	<b>SVMs e suas características</b>	<b>28</b>
<b>2.3.4</b>	<b>Biblioteca Scikit-learn</b>	<b>30</b>
2.4	Técnicas de Processamento de Sinais	32
<b>2.4.1</b>	<b>Filtros</b>	<b>32</b>
<b>2.4.2</b>	<b>Transformada de Fourier x Transformada de Wavelet</b>	<b>33</b>
2.5	Trabalhos Relacionados	34
2.6	Tratamentos de dados e suas características	39
<b>3</b>	<b>DESENVOLVIMENTO</b>	<b>41</b>
3.1	Escolha da ferramenta de coleta	41
3.2	Pesquisa de Campo	43
3.3	Conversor	44
3.4	Definição do eixo representativo	45
3.5	Tratamento de dados	45
<b>4</b>	<b>RESULTADOS</b>	<b>48</b>
4.1	Análise e aplicação do SVM	48
4.2	Cálculo de Amostragem	51

4.3	Resultados Encontrados e Validação . . . . .	52
5	<b>CONCLUSÕES . . . . .</b>	<b>60</b>
6	<b>TRABALHOS FUTUROS . . . . .</b>	<b>62</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>63</b>

# 1 INTRODUÇÃO

A utilização cada vez maior dos smartphones em todo o mundo trouxe oportunidades para otimizar as funcionalidades incumbidas a esses aparelhos em diversas áreas (GHOSE et al., 2012). Com crescimento populacional e a maior concentração de veículos em grandes centros urbanos, têm-se observado uma grande necessidade de aprimorar a identificação e rápida correção dos problemas nas vias urbanas dos problemas das grandes metrópoles.

O aumento da quantidade de veículos que circulam em um determinado local gera um desgaste maior na infraestrutura de trânsito como um todo (ruas, avenidas, rodovias, etc.), tendo como um grande desafio dos órgãos responsáveis a rápida atuação e manutenção destas vias de trânsito. Acidentes de trânsito estão relacionados diretamente com a qualidade de uma determinada via (TEDESCHI; BENEDETTO, 2017), e os governos responsáveis acabam sendo pressionados a garantir a manutenção e melhor análise desta infraestrutura. A identificação dos problemas nas vias é dificultada por, muitas vezes, necessitar que um agente do departamento de trânsito responsável por aquela via circule pelo local, identifique e registre o problema, para assim tomar as devidas providências e sanar a situação (GHOSE et al., 2012). Todo esse processo demanda muito tempo para ser concluído, o que gera mais insatisfação e riscos para os usuários daquele trecho.

Embora existam diversas iniciativas para informar o tráfego em tempo real (como aplicativos de smartphones e a imprensa - principalmente rádios e emissoras de televisão), poucas trazem informações sobre a qualidade atual da via durante o trajeto, aspecto que pode ser decisivo na tomada de decisão entre rotas pelos usuários. Existem protocolos de reconhecimento de trânsito (se há congestionamento ou não) que utilizam como base a velocidade média dos veículos em determinado local (MEDNIS et al., 2011). Recursos como este, disponíveis nos eletrônicos, ainda não são utilizados em larga escala para identificação de imperfeições de forma rápida e precisa.

Uma via pública de qualidade pode ser definida pelo conjunto de diversos fatores, dentre eles a quantidade de veículos que transitam por ela diariamente, se existe ou não valetas, buracos, rachaduras, lombadas, se a velocidade média dos veículos está de acordo com as leis locais.

Estes dados poderiam ser utilizados como informação para empresas de seguros automotivos e de aluguéis veiculares, aplicativos de determinação de rotas

para os usuários e seus tipos de veículos, dentre outros. Segundo o artigo “Classificação da Qualidade de Vias Urbanas baseado em Sensoriamento Participativo” (NUNES et al., 2018), é perceptível que, quando se há uma melhor análise das variáveis impactantes na qualidade final das vias, há um melhor planejamento dos órgãos responsáveis e efetiva correção dos problemas viários. Um melhor monitoramento das vias urbanas com informações mais precisas contribui indiretamente na qualidade de vida de todos que transitam por estas, além de auxiliar os órgãos responsáveis, fazendo com que os investimentos disponibilizados para revitalização destas seja melhor dimensionado onde realmente necessita.

A utilização dos recursos disponíveis nos aparelhos celulares atuais possibilita reduzir os custos de implementação de sensores nos veículos que trafegam e avaliam as vias (DOUANGPHACHANH; ONEYAMA, 2014), item no qual a grande maioria dos motoristas carregam consigo durante uso de algum meio de transporte. Apesar da tecnologia estar disponível há aproximadamente 10 anos (WAHLSTRÖM et al., 2017), a utilização dos dados gerados pelos sensores dos smartphones ainda é pouco difundida quando relacionada a identificação de fatores externos e de qualidade viária em questão.

Baseado na Internet das Coisas (do inglês IoT - *Internet of Things*), o conceito de Internet dos Veículos (do inglês IoV – *Internet of Vehicles*) têm ganhado importância em todo mercado com a expansão e implantação da internet 5G (Mario Gerla, Eun-Kyu Lee, Giovanni Pau; LEE, 2015). Estas definições têm como foco realizar conexões entre os veículos automotores para que possam entregar um deslocamento mais seguro e de qualidade para os passageiros e usuários destes.

Já existem estudos relatando os diversos protocolos que podem ser utilizados no conceito IoV, classificando-os quanto a desempenho, atraso, aplicabilidade em diferentes dimensões, testes já realizados e possíveis expansões em larga escala (CHENG et al., 2015). A tendência de integração dos dados gerados por cada aparelho celular localizado nos veículos tem potencial para formar um sistema na nuvem onde há centralização e geração de diversos tipos de informações que possam ser destinadas às partes interessadas.

A conexão destes entes está diretamente relacionada à expansão da Inteligência Artificial (IA), área que realiza o agrupamento de várias tecnologias, como Redes Neurais Artificiais (RNA), algoritmos e Máquinas de Vetores de Suporte (do inglês, Support Vector Machine - SVM), visando estabelecer padrões e soluções para os mais diversos problemas encontrados no dia-a-dia. Um SVM, por exemplo, pode ser preparado para receber um determinado agrupamento de dados e, a partir de um algo-

ritmo desenvolvido e treinado, realizar classificações e regressões destes em campos previamente definidos, memorizando padrões e desenvolvendo uma rede artificial de informações.

A necessidade de melhorias nos espaços públicos cada dia mais populosos, que se soma à grande quantidade de veículos nas ruas, torna-se visível a importância do desenvolvimento e estudos de estímulos que facilitem a identificação e manutenção destes espaços viários e, como consequência, contribui para melhorar a experiência dos usuários que transitam por estes.

As vias de trânsito da cidade de Betim carecem de informações precisas, tratadas e analisadas, que, quando inseridas em um sistema adequado, gerem contribuições para uma melhor manutenção e correção de seus problemas. Torna-se, então, necessário o estudo e desenvolvimento de uma aplicação que viabilize a utilização dos recursos disponíveis (sensores) nos aparelhos celulares atuais dos motoristas usuários, por meio da melhoria de um sistema de identificação de imperfeições e anomalias viárias adaptadas a funcionar nos atuais cenários de estudo.

## 1.1 Objetivos

- Desempenhar aplicabilidade do SVM classificador de (CARLOS et al., 2018) para a região;
- Realizar o tratamento adequado dos dados;
- Reduzir a variabilidade das características avaliativas;
- Validar as técnicas de processamento de dados utilizadas (Análise estatística, FFT, filtro passa-baixa e Transformada de Wavelet);
- Encontrar resultados compatíveis com trabalhos referenciados;
- Determinar viabilidade de aplicação real;

### 1.1.1 Objetivos Específicos

- Definir o eixo representativo e as características avaliativas, levando em consideração as particularidades locais;
- Utilizar a Transformada Rápida de Fourier para análise da frequência;
- Filtrar dados reais utilizando o filtro passa-baixa;

- Realizar a decomposição e reconstrução dos sinais via Transformada de Wavelet;
- Obter as métricas Precisão, Sensibilidade e Nota F1 próximas às encontradas por (CARLOS et al., 2018);
- Captar e classificar dados reais que certifiquem o funcionamento da aplicação na região;

## 1.2 Estruturação Textual

No Capítulo 2 está representado o referencial teórico no qual este trabalho se baseia, como os tipos de pavimentações analisados, os principais sensores utilizados atualmente para captação de dados, além de conceitos e denominações da inteligência artificial. Os trabalhos relacionados e as principais técnicas de tratamento e processamento de sinais existente também estão estruturados. O Capítulo 3 seguinte é focado no desenvolvimento do trabalho e quais foram os caminhos percorridos que visaram atingir os objetivos almejados, assim como os recursos utilizados para análise e formatação dos dados.

As respostas e resultados consequentes do desenvolvimento estão distribuídos no decorrer do Capítulo 4. Durante o Capítulo 5, as conclusões baseadas nos resultados obtidos são destacadas e discutidas. Por último, estão presentes no Capítulo 6 quais foram as principais melhorias identificadas no decorrer do trabalho, que podem ser desenvolvidas e implementadas em possíveis trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

As especificações do desenvolvimento do trabalho e os objetos de estudo selecionados estão descritos abaixo. Na Seção 2.1, contém os ambientes selecionados para a pesquisa e as respectivas especificações viárias determinadas para o estudo. Os diversos sensores disponíveis no mercado e os aplicativos mais usados e adequados para este tipo de aplicação localizam-se na Seção 2.2. Na terceira Seção é explanado as tecnologias relacionadas à Inteligência Artificial e suas características.

As técnicas de processamento de sinais mais relevantes para o tema são descritas na Seção 2.4. A sexta Seção analisa os tipos de tratamentos de dados geralmente usados e estudados, e os trabalhos relacionados para encontrar os resultados ambicionados encontram-se discutidos na Seção 2.5.

### 2.1 Tipos de pavimentação e suas imperfeições

A localização da via na qual vai ser avaliada e suas respectivas características de construção tendem a impactar no resultado final das avaliações. Um exemplo disso é dado por A. Tedeschi e F. Benedetto (TEDESCHI; BENEDETTO, 2017), onde é descrito que uma rua com um determinado tipo de pavimentação contém características de avaliação diferentes de tipo de material (como asfaltada, concretada ou calçada), variando de acordo com a região e o modo como é identificado. Este trabalho será focado na análise de trechos de pavimentação asfáltica, visto que é o tipo mais encontrado e usado na região de estudo. Com o objetivo de abranger diversos cenários e suas respectivas respostas geradas no classificador, é esperado a captação de testes em vias visualmente consideradas com boas condições de tráfego – com asfalto liso e sem interferências na maior parte do trajeto - e em situações que contenham imperfeições, como buracos, rachaduras, lombadas e valetas, sendo consideradas em condições inadequadas para tráfego.

A malha viária existente no município de Betim, estado de Minas Gerais (MG, Brasil), contém diversos níveis de qualidade e de anomalias, podendo identificar visualmente problemas na sinalização, adequação do espaço, existência de pavimentação e imperfeições. Este trabalho focará na análise e classificação das vias betinenses quanto às imperfeições encontradas nesta região.

É possível distinguir os tipos de imperfeições encontradas na malha em estudo, de acordo com o Catálogo de Estradas Defeitos - CORD (National Institute for

Phys Planning Constr Res, 1985) e o Manual de identificação de socorro (MILLER; BELLINGER, 2014), conforme mostrado abaixo, em três mais identificados na região:

- Buracos: são pedaços de pavimento ausentes, ou profundamente afundados; existem de vários tamanhos, envolvendo uma ou mais camadas da estrutura do pavimento asfáltico. Seu tamanho e profundidade podem aumentar com o acúmulo de água da chuva e demais detritos;
- Rachaduras: estão entre as anomalias mais comuns, podendo ser de diferentes tipos, como longitudinal, de fissuras transversais ou por fadiga. A longitudinal e transversal são fissuras paralelas e perpendiculares à linha central do pavimento, respectivamente. Enquanto a trinca por fadiga é composta por rachaduras interconectadas; são anomalias tipicamente comuns de vias pavimentadas por concretagem e massa asfáltica;
- Deformações de superfície: são alterações nas superfícies do pavimento original. Uma deformação típica é a depressão longitudinal, normalmente causada pelos caminhos das rodas; nesta categoria também estão presentes algumas variações de deformação, como as lombadas, valetas e ondulações de vias;

De acordo com o Conselho Nacional de Trânsito (CONTRAN) (CONSELHO NACIONAL DE TRÂNSITO - CONTRAN, 2016), as lombadas, ou ondulações transversais, podem ser inseridas intencionalmente pelos órgãos responsáveis pela via, com o intuito de reduzir a velocidade do veículo de forma imperativa em determinados trechos, seguindo as especificações predefinidas pelos Código de Trânsito Brasileiro (CTB). Para este trabalho, tais ondulações serão tratadas como imperfeições, se enquadrando como uma deformação de superfície. As Figuras 1, 2, 3 e 4 exemplificam os tipos de imperfeições citadas atrás encontradas na região de Betim.

**Figura 1 – Exemplo de buraco**



Fonte: Disponível em: <<https://www.gazetadopovo.com.br/abntemfoco>>. Acesso em: 24 de jan. de 2020.

**Figura 2 – Exemplo de deformação**



Fonte: Disponível em: <<https://www.gazetadopovo.com.br/abntemfoco>>. Acesso em: 24 de jan. de 2020.

**Figura 3 – Exemplo de rachaduras**



Fonte: Disponível em: <<https://www.gazetadopovo.com.br/abntemfoco>>. Acesso em: 24 de jan. de 2020.

**Figura 4 – Exemplo de lombada**

Fonte: Disponível em: <<https://www.gazetadopovo.com.br/abntemfoco>>. Acesso em: 24 de jan. de 2020.

## 2.2 Sensores e coleta de dados

Atualmente existem diversos sensores disponíveis no mercado que podem ser usados como parte do trabalho de identificação e classificação da qualidade das vias estudadas e suas imperfeições. Dentre os sensores existentes, é necessário destacar os seguintes: identificação por imagem (faz uso de uma câmera como sensor) e por giroscópio e acelerômetros (que podem ser referenciados em eixos, geralmente são três: X, Y e Z), usados em trabalhos desenvolvidos por A. Tadeschi (TEDESCHI; BENEDETTO, 2017) e A. Mednis (MEDNIS et al., 2011), respectivamente.

As identificações realizadas através de imagens captadas das vias têm sido utilizadas em pesquisas e experimentos realizados em países desenvolvidos, como os Estados Unidos, Canadá, Alemanha e China. Alguns carros presentes nestes mercados já possuem câmeras instaladas de fábrica, o que possibilita a aplicação e reduz os custos para o uso deste sensor. Além disso, o padrão das imperfeições encontradas nestas regiões possui características distintas das encontradas no Brasil: com variações de clima, intensidade de tráfego, sinalizações visuais que as diferenciam das brasileiras (TEDESCHI; BENEDETTO, 2017). De acordo com Tadeschi, para identificar todos estes dados, as câmeras, dispostas em sua maioria próximo ao para-brisa dos veículos, contribuem para uma melhor percepção, identificação e leitura das vias, utilizando de algumas funcionalidades como o foco e a profundidade para detecção.

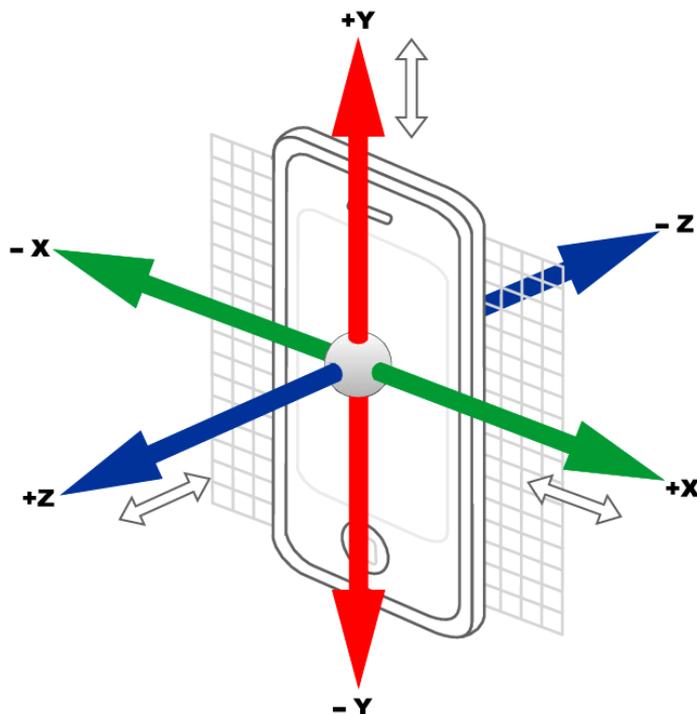
Característica comum das imperfeições encontradas nas vias brasileiras, os desníveis gerados devido aos constantes recapeamentos geram ondulações e vibrações que diferem dos buracos em si. O uso de câmeras para esse tipo de imperfeição torna-se instável, muitas vezes por não haver um contraste suficiente entre o asfalto e o

trecho recapeado ondulado. Outro fator que contribui para o descarte da câmera como sensor de captação das informações deste projeto está relacionado a sua disponibilidade nos veículos automotores brasileiros. A grande maioria dos carros comercializados no Brasil não possui câmera frontal instalada de fábrica. Além disso, o uso das câmeras presentes nos aparelhos celulares para este fim conduziria à necessidade de se ter um posicionamento físico fixo e pouco ergonômico do aparelho no interior dos veículos. A manipulação e tratamento de imagens geram, também, um grande volume de dados, se fazendo necessário realizar um considerável investimento incremental para armazenagem e processamento.

O acelerômetro pode ser descrito como um transdutor que converte o estresse mecânico (ou outra forma de energia) em um sinal elétrico (BELLO-SALAU et al., 2019). Como o reconhecimento da qualidade das vias pode ser dado pela vibração que o asfalto impele ao veículo automotor e, conseqüentemente, ao aparelho celular posicionado internamente no mesmo, tais tipos de sensores foram considerados para aplicação desta identificação.

Na Figura 5, é possível observar o comportamento de um acelerômetro medidor de três eixos. A principal diferença entre o acelerômetro e o giroscópio tem relação direta com as medições realizadas: enquanto o primeiro mede a aceleração de um corpo em relação à gravidade, o último mede a posição do corpo através da força da gravidade. Juntos, tais equipamentos podem complementar suas medições, deixando-as mais precisas e exatas.

Figura 5 – Relação dos eixos em um aparelho celular



Fonte: Disponível em: <<https://www.gazetadopovo.com.br/abntemfoco>>. Acesso em: 24 de jan. de 2015.

A democratização dos aparelhos celulares, e por consequência dos acelerômetros que representam os eixos X, Y e Z, possibilitou a determinação destes equipamentos como fontes de captação inicial dos dados para este trabalho. Ademais, os celulares contam hoje com o sistema de localização global (do inglês, *Global Position System - GPS*) que podem ser usados como recursos de localização do veículo, reunindo, em uma única ferramenta, os principais sensores de dados.

A escolha de usar os acelerômetros disponíveis nos aparelhos celulares atuais entra na sinergia por serem sensores que já estão presentes massivamente junto à população de motoristas, não sendo necessário a instalação de nenhum equipamento no veículo utilizado nos testes. Desse modo, o custo de aquisição dos dados é reduzido e há uma melhora na aplicabilidade do mesmo, visto que os sensores são portáteis e podem fornecer os dados para um aplicativo compatível disponível no aparelho celular.

A análise de vibrações mecânicas a partir de acelerômetros é um método bastante eficiente na detecção prematura de anomalias, através da observação da evolução do nível global de vibrações e de uma análise das respectivas frequências, na análise espectral. A análise de vibrações é um dos indicadores mais precisos do estado de funcionamento da máquina, permitindo identificar a fonte perturbadora. Estas

técnicas provaram ser extremamente fiáveis e bastante precisas na detecção de avarias (SEQUEIRA, 2013).

Primeira aplicação analisada para realização da captura de dados dos acelerômetros, o *Science Journal* é uma plataforma de dispositivos móveis desenvolvida pelo Google compatível com os principais sistemas operacionais usados nos aparelhos celulares atuais (como o Android e o IOS), que, depois de instalado, permite adequar os dados em relação ao domínio discreto do tempo, organizando-os para um futuro tratamento dos mesmos. Os dados mensurados e armazenados neste aplicativo podem ser exportados em arquivos planilhados (como em extensões do tipo csv e xls), que servirão de entrada nos softwares de simulação a serem usados. Porém, esta aplicação acabou sendo descontinuada pelo desenvolvedor, não recebendo mais suporte de atualizações.

O *Physics Toolbox Accelerometer* é outro aplicativo que também possibilita realizar leituras dos acelerômetros presentes nos celulares. Criado pela Vieyra Software (VIEYRA; VIEYRA, 2015), este aplicativo está presente em ambas plataformas *mobile* – Android e IOS, e, em uma comparação inicial, mostra-se mais simplificado que o *Science Journal*, principalmente durante a extração dos dados do aplicativo para alguma outra plataforma. Os dados são planilhados e gravados em csv, que podem ser transferidos para o computador via e-mail ou usando um cabo de dados.

Com os dados amostrais encontrados e o intuito de realizar um tratamento inicial, foi visada a utilização de, por exemplo, filtros e processamento dos sinais no domínio da frequência, com o objetivo de reduzir ou cancelar possíveis ruídos indesejáveis. Partindo da premissa que os dados tratados representam com mais fidelidade a realidade o trecho em questão, era esperado que o algoritmo tivesse melhor capacidade de identificar e apontar os trechos que possuem imperfeições, por meio de comparações com sinais pré-definidos.

## 2.3 IA

Termo que tem sido utilizado com notável frequência na atualidade, a inteligência artificial pode ser descrita como a tecnologia que busca aproximar as máquinas do pensamento dos seres humanos, onde são treinadas a processarem as informações recebidas, tomarem decisões e atuarem conforme aprendizado adquirido. Um dos principais objetivos dos sistemas de inteligência artificial (também comumente chamados de IA, ou pela sigla em inglês AI - *Artificial Intelligence*) pode ser definido como executar as funções que antes seriam necessárias a inteligência humana para realizá-las (VV

VASCONCELOS; PP MARTINS-JÚNIOR, 2004). A capacidade de raciocínio - aplicar regras lógicas a um conjunto de dados disponíveis para chegar a uma conclusão, aprendizagem - aprender com os erros e acertos de forma que no futuro agirá de maneira mais eficaz, reconhecer padrões (tanto visuais e sensoriais, como também padrões de comportamento) e inferência (capacidade de conseguir aplicar o raciocínio nas situações do cotidiano) são algumas das características básicas desses sistemas.

O uso de IA na sociedade tem ampliado as possibilidades em universos onde há uma necessidade de grande processamento de dados e tomadas de decisões rápidas, atuando com mais velocidade e menor taxa de erros que humanos nas mesmas funções. Conjuntos de dados que levariam horas, e, em alguns casos, dias para terem uma análise criteriosa geridas por seres humanos, podem ser realizadas em questão de segundos por um sistema bem treinado para tal. Além disso, as máquinas, ao contrário dos seres humanos, não necessitam de intervalos frequentes, conseguindo exercer várias horas de funcionamento sem interrupção, apenas seguindo a programação realizada.

Uma das grandes vantagens do uso de IA em determinado sistema está no fato de que, a cada novo funcionamento (novos dados analisados, tomadas de decisões concluídas), o próprio sistema realiza um refinamento e atualização de seus protocolos para melhorar o processamento seguinte, criando um ciclo de aprendizagem e padrões. Assim, quanto mais dados são processados, maior será o grau de acerto que se busca, melhorando o sistema como um todo.

O surgimento e massificação dos computadores modernos tem contribuído para a inteligência artificial ganhar meios e massa crítica para se estabelecer como ciência integral, com problemáticas e metodologias próprias. Seu desenvolvimento tem extrapolado programas básicos ou de conversão e envolvido áreas como visão computacional, análise e síntese da voz, lógica difusa, redes neurais artificiais e muitas outras (RICH; KNIGHT, 1994). A Figura 6 demonstra um posicionamento dos principais campos e conceitos centrados no mundo da inteligência artificial.

**Figura 6 – Representação Inteligência Artificial**

Fonte: Autoria própria.

Um dos campos mais promissores da IA, o aprendizado profundo (do inglês, *Deep Learning*) consiste em fazer dos sistemas entidades capazes de aprender evolutivamente. O uso de uma rede neural para realizar a classificação dos dados coletados através dos acelerômetros está diretamente ligado com um constante aprendizado do sistema classificador.

### 2.3.1 Aprendizado de máquina

O aprendizado de máquina (do inglês, *Machine Learning*), pode ser definido como o campo de estudo de reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial. Existem três principais métodos de aprendizagem que podem ser aplicados (PEDRO C T GOMES, 2019), são eles: aprendizagem supervisionada, quando o sistema já reconhece quais entradas estão associadas com quais saídas e necessita aprender um meio de entender esse vínculo, utilizando de detecção de padrões para assim estabelecer previsões. Ou seja, o objetivo é aprender uma regra geral que mapeia as entradas para suas respectivas saídas.

Segundo método, chamado de aprendizagem não-supervisionada, o sistema busca identificar padrões para rotular os dados, relacionando determinados conjuntos de dados. Desse modo, o algoritmo busca encontrar estrutura nas entradas fornecidas, descobrir novos padrões nos dados, ou um meio, para atingir um fim. Por último, na aprendizagem por reforço é utilizado da estratégia de tentativa e erro da máquina, ensinando o sistema a priorizar certos hábitos conforme a prática é realizada, otimizando o processo a cada novo ciclo.

Existem diversos modelos de aplicabilidade e abordagens para os métodos de aprendizagem em máquina e resolução de problemas, como o aprendizado baseado em árvores de decisão, redes neurais artificiais, aprendizado profundo, análise de *clusters*, máquinas de vetores de suporte (do inglês *Support Vector Machines* - SVMs)

e detecção de anomalias, para citar os principais. Cada um dos modelos se entrelaçam e podem ser usados em conjunto ou individualmente de acordo com a necessidade do universo de aplicação. Nos parágrafos abaixo, é realizada uma visão geral destes principais conceitos relacionados ao aprendizado de máquinas.

Os métodos baseados em árvores de decisão geram modelos preditivos de alta precisão, estabilidade e facilidade de interpretação. Ao contrário dos modelos lineares, eles mapeiam muito bem relações não-lineares, podendo ser adaptados para resolver vários tipos de problema (como, por exemplo, classificação ou regressão) (Analytics Vidhya Content Team, 2016).

### **2.3.2 Redes neurais**

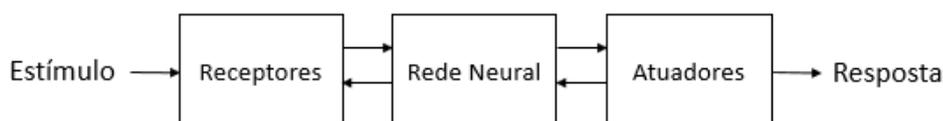
As Redes Neurais Artificiais (também conhecidas por RNAs) podem ser definidas como modelos computacionais inspirados pelo sistema nervoso central humano, capazes de realizar o aprendizado de máquina, bem como o reconhecimento de padrões. Segundo Simon Haykin, autor do livro *Redes Neurais: Princípios e Prática* (HAYKIN, 2001), de forma geral uma rede neural é uma máquina projetada para modelar a maneira como o cérebro humano realiza uma tarefa em particular ou função de interesse, normalmente implementada utilizando componentes eletrônicos ou é simulada por programação de um computador digital.

Haykin (HAYKIN, 2001) afirma que uma RNA se assemelha ao cérebro em dois aspectos: tem a capacidade de adquirir conhecimento de um ambiente através de um processo chamado fase de aprendizagem, onde são fornecidos dados como entrada, e, por sua vez, informando qual é a saída (resposta) que é esperada após o processamento. O segundo aspecto é possuir forças de conexão entre neurônios, conhecidos como pesos sinápticos, que são utilizados para armazenar o conhecimento adquirido durante o processo de aprendizagem. Outra característica importante das RNAs está relacionado a sua flexibilidade, podendo lidar com grandes mudanças na informação de entrada, tais como sinais ruidosos ou outras alterações imprevistas.

A relação de uma RNA com o cérebro humano se aproxima ainda mais quando analisados seus respectivos funcionamentos. Haykin (HAYKIN, 2001) realiza uma análise demonstrando que o sistema nervoso humano pode ser analisado como um sistema de três estágios (semelhante ao diagrama de blocos apresentado na Figura 7). No centro do sistema encontra-se o cérebro (representado pela rede neural), que recebe informações continuamente, e toma decisões apropriadas com base nas mesmas. Observa-se que existem dois conjuntos de setas representados na figura. Aquelas que apontam da esquerda para a direita indicam a transmissão para frente do sinal portador

de informação, através do sistema. Já as setas voltadas da direita para a esquerda indicam a presença de realimentação no sistema. Os receptores convertem estímulos do corpo humano ou do ambiente externo em impulsos elétricos que transmitem a informação à rede neural (cérebro). Já os atuadores convertem os impulsos elétricos gerados pela rede neural em respostas discerníveis como saídas do sistema.

**Figura 7 – Representação em diagrama de blocos do sistema nervoso**



Fonte: Autoria própria.

O procedimento utilizado para realizar o processo de aprendizagem de uma determinada rede neural é chamado de algoritmo de aprendizagem, cuja função é modificar os pesos sinápticos da rede de uma forma ordenada para alcançar o objetivo de projeto desejado (HAYKIN, 2001). O processo de aprendizagem conhecido como aprendizado profundo consiste no processamento de múltiplas camadas escondidas de determinada rede neural artificial, muito usado em aplicações como visão computacional e reconhecimento de fala (Honglak Lee et al., 2009).

A análise de *clusters* (também chamado de Clusterização, do inglês *Clustering*) pode ser definido como um conjunto de técnicas de organização de dados que tem como objetivo realizar agrupamentos automáticos, segundo o seu grau de semelhança, formando grupos ou conglomerados homogêneos. Desse modo, objetos que se encontram em um mesmo grupo possuem características próximas, e objetos de grupos diferentes estão mais distantes entre si. Cada um destes conjuntos de dados resultantes de um processo é chamado de *cluster*. Porém, conforme explicado no livro *Artificial Intelligence: A Modern Approach* (2010), embora o agrupamento torne possível usar um algoritmo de tempo linear, o tamanho da rede aumenta exponencialmente, pois as tabelas de probabilidade condicional para os clusters envolvem o produto cruzado dos domínios das variáveis, retardando o processamento. Apesar deste problema, as técnicas de agrupamento são atualmente a abordagem mais eficaz para a avaliação exata de redes multiplamente conectadas (RUSSELL et al., 2010).

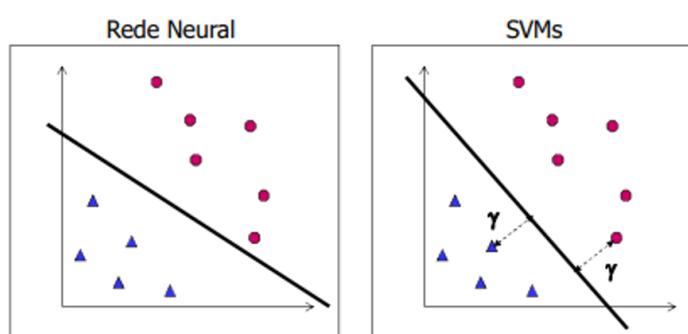
### 2.3.3 SVMs e suas características

Próximo modelo referência deste estudo, os SVMs podem ser definidos como conjuntos de métodos supervisionados usados para realização de reconhecimento

de padrões, classificações, regressões e outras detecções (AL, 2011). Algumas características se destacam nestes modelos, como a eficácia em espaços dimensionais elevados e onde os casos em que o número de dimensões é maior do que o número de amostras, fazem uso de subconjuntos de pontos de treinamento na função de decisão (chamados de vetores de suporte), tendo maior eficiência em termos de memória.

Em outras palavras, a estratégia básica de determinada máquina de vetor de suporte é encontrar um hiperplano - uma linha - que maximize a margem de separação, também chamada de margem larga, entre os dados observados, e com erro marginal baixo (poucos objetos próximos da margem) (KATTI FACELI et al., 2021). A Figura 8 ilustra a diferença entre uma rede neural comum e uma máquina de vetor de suporte durante uma classificação.

**Figura 8 – Rede Neural x SVM**



Fonte: Adaptado de (KATTI FACELI et al., 2021).

Segundo Haykin, os algoritmos de aprendizagem por vetor de suporte podem ser usados para construir diferentes tipos de máquinas de aprendizagem. Ou seja, para cada uma das redes alimentadas adiante pode-se usar o algoritmo de aprendizagem por vetor de suporte para implementar o processo de aprendizagem, com determinado conjunto de dados de treinamento. Desse modo, enquanto outros algoritmos são planejados especificamente para treinar um *perceptron* de múltiplas camadas, o algoritmo de aprendizagem por vetor de suporte é mais genérico, tendo uma aplicabilidade mais ampla (HAYKIN, 2001).

Existe uma classe de vetor de suporte capaz de realizar a classificação multi-classe em um determinado conjunto de dados, chamada de SVC (do inglês, *Support Vector Classification*), que faz uso da mesma biblioteca do algoritmo SVM (AL, 2011). A principal diferença entre SVM e SVC apontada está relacionada a esta classificação. Enquanto o primeiro realiza classificações binárias, o segundo é capaz de desempenhar classificações binárias e multiclasse em determinado conjunto de dados.

Os problemas multiclases tendem a ser mais complexos e de difícil resolução que os problemas binários, em grande parte se devem às suas fronteiras de decisão que tendem a ser mais complexas (GALAR et al., 2011). Diversos métodos já foram desenvolvidos na área de binarização de problemas multiclases. Uma das técnicas mais populares consiste em dividir o conjunto de dados original em subconjuntos de duas classes, onde para cada subconjunto é usado um método binário e a saída desses classificadores binários são combinadas para a decisão final.

Os critérios de avaliação dos algoritmos SVMs levam em consideração as características impostas durante a execução. Conforme descrito por (WU et al., 2020), a extração destes recursos é uma etapa essencial antes da classificação - para treinamento e teste - e afeta significativamente a eficiência do sistema de detecção de anomalias. A avaliação destas características foi através das métricas de classificação usadas neste estudo, incluindo a sensibilidade, precisão e a pontuação F1.

#### **2.3.4 Biblioteca Scikit-learn**

Scikit-learn é uma biblioteca de aprendizado de máquina de código aberto que oferece suporte ao aprendizado supervisionado e não supervisionado. Ele também fornece várias ferramentas para ajuste de modelo, pré-processamento de dados, seleção e avaliação de modelo e muitos outros utilitários. A biblioteca possui uma plataforma online que tem como objetivo ser um guia ilustrativo de alguns dos principais recursos que o scikit-learn oferece. Desse modo, fornece dezenas de modelos e algoritmos de aprendizado de máquina integrados, chamados estimadores. Cada estimador pode ser ajustado a alguns dados usando seu método de ajuste. Em geral, os métodos de ajustes recebem duas entradas (AL, 2011):

1. A chamada matriz de amostras (ou matriz de design)  $X$ . O tamanho de  $x$  ( $n$ -amostras,  $n$ -recursos), onde as amostras são representadas como linhas e os recursos são representados como colunas.
2. Os valores alvo  $y$ , que podem ser números reais para tarefas de regressão ou inteiros para classificação (ou qualquer outro conjunto discreto de valores). Para tarefas de aprendizagem não supervisionadas,  $y$  não precisa ser especificado, onde  $y$  é a  $i$ -ésima entrada corresponde ao destino da  $i$ -ésima amostra (linha) de  $x$ .

Ambos  $x$  e  $y$  são normalmente esperados que sejam matrizes tipo NumPy, ou tipos de dados semelhantes às matrizes equivalentes, embora alguns estimadores trabalhem com outros formatos, como matrizes esparsas. O NumPy é o pacote da

linguagem Python que permite trabalhar com arranjos, vetores e matrizes de N dimensões. Ele provê diversas funções e operações sofisticadas, incluindo objeto array para a implementação de arranjos multidimensionais, objeto matrix para o cálculo com matrizes, ferramentas para álgebra linear, transformadas de Fourier básicas, ferramentas sofisticadas para geração de números aleatórios (PYSCIENCE-BRASIL, 2021). Além disso, as classes criadas podem ser facilmente herdadas, permitindo a customização do comportamento, o que possibilita maior velocidade às operações realizadas.

No Scikit-learn (AL, 2011), a implementação do SVC é baseada no libsvm. O tempo de ajuste é dimensionado pelo menos quadraticamente com o número de amostras, podendo ser impraticável além de dezenas de milhares de amostras. Desse modo, para um grande conjunto de dados, é recomendado o uso da classe Linear Support Vector Classification, representado pelo código `sklearn.svm.LinearSVC`. Diferentemente do SVC, que trabalha com o esquema “um-contra-um” (do inglês, *one-vs-one* - OVO), o suporte multiclasse Linear SVC é tratado de acordo com um esquema “um-contra-todos” (*one-vs-the-rest* ou *one-vs-all*, OVA).

Diversas estratégias de decomposição podem ser encontradas na literatura. As duas mais comuns são chamadas “um-contra-um” e “um-contra-todos”. O primeiro consiste em dividir o problema inicial em vários problemas binários, gerando todas as combinações possíveis entre pares de classes, de modo que o classificador é ensinado a discriminar entre cada par e, em seguida, as saídas desses classificadores básicos são combinados para prever a classe de saída (KNERR et al., 1990). Já o OVA aprende um classificador para cada classe, onde a classe é distinta de todas as demais, desse modo o classificador base com uma resposta positiva indica a classe de saída (CLARK; BOSWELL, 1991).

Os algoritmos SVM e SVC usam um conjunto de funções matemáticas que são definidas como kernel, que tem como função básica pegar dados como entrada e transformá-los na forma necessária para processamento (DATA-FLAIR, 2021). O kernel aplicado ao SVM pode mapear pontos de dados para um espaço de dimensão superior, tornando viável a separação não linear (WU et al., 2020). A função kernel pode ser linear, polinomial, função de base radial (do inglês, *Radial Basis Function*, RBF) e sigmoideal. O RBF e o polinomial são os mais usados quando os pontos não são linearmente separáveis, quando não há conhecimento prévio sobre os dados. Além disso, o primeiro tem uma resposta localizada e finita ao longo de todo o eixo x, enquanto o kernel polinomial é geralmente usado para processamento de imagens.

Ao realizar o treinamento de um SVM com o kernel RBF, existem dois parâmetros que precisam ser considerados: Gamma e C (AL, 2011). O parâmetro Gamma

define até onde chega a influência de um único exemplo de treinamento. Ou seja, um Gamma alto considerará apenas os pontos mais próximos ao hiperplano plausível, enquanto o Gamma baixo faz consideração de pontos a uma distância maior, conseqüentemente crescendo a quantidade amostrada.

Já o parâmetro C, chamado de parâmetro de regularização, é comum a todos os kernels SVM, e compensa a classificação incorreta de exemplos de treinamento pela simplicidade da superfície de decisão (CZAKO ZOLTAN, 2018). Ou seja, o C diz à otimização do SVM o quanto se deseja evitar perder a classificação de cada exemplo de treinamento. Um C baixo torna a superfície de decisão mais suave, enquanto um C alto visa classificar todos os exemplos de treinamento corretamente, de modo que a taxa de classificação de perda de dados de treinamento será menor. Porém, mesmo que todos os dados de treinamento tenham sido classificados corretamente, isso não significa que aumentar o C sempre aumentará a precisão, devido ao sobre ajuste (do inglês, *overfitting*) - quando um modelo estatístico se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para prever novos resultados (CÁRDENAS-MONTES, 2006). A escolha adequada de C e Gamma é crítica para o desempenho adequado do algoritmo SVM.

Último parâmetro usado durante a implementação do SVC, o *tol* é descrito como tolerância para critério de parada (AL, 2011). Ele diz para o algoritmo parar de pesquisar um mínimo (ou máximo) quando alguma tolerância for alcançada, ou seja, quando estiver perto o suficiente. Se o valor de *tol* for muito grande, o algoritmo para antes que possa realizar a conversão. Possui uma entrada do tipo *float*, com uma variação padrão entre 1 e -3.

## 2.4 Técnicas de Processamento de Sinais

### 2.4.1 Filtros

Filtros digitais que definem uma determinada faixa de frequência de trabalho possibilitam eliminar certas frequências que possam distorcer o sinal e desestabilizar o algoritmo classificador. O filtro denominado passa-baixa é um dispositivo que permite a passagem de determinada banda inferior de frequência, restringindo as demais.

O desenvolvimento de um filtro passa por determinadas parametrizações, que o possibilitam se adequar melhor para as necessidades das quais precisam ser atendidas. A ideia principal do filtro se passa através da determinação de uma banda de passagem do sinal, para que realize a atenuação das frequências fora desta faixa desejada. Com isso, é esperado que os sinais representantes das anomalias asfálticas sejam mais

facilmente destacados em relação ao sinal como um todo, contribuindo para melhor resposta do algoritmo classificador.

#### 2.4.2 Transformada de Fourier x Transformada de Wavelet

Apesar de bastante utilizada, a transformada de Fourier tem capacidade de aplicação apenas no domínio da frequência. Já a transformada wavelet é uma função definida como uma ferramenta analítica desenvolvida para representar um sinal simultaneamente nos domínios da frequência e do tempo (ou distância). A análise da forma wavelet usa pequenas funções semelhantes a ondas conhecidas como wavelets, e cada uma delas pode ser movida para vários locais no sinal para ser transformada, podendo ser esticada (dilatada) ou comprimida para alterar a forma do sinal em uma escala e local específicos. Essa transformação é realizada em vários locais do sinal e em várias escalas da wavelet (WEI et al., 2005).

Ao selecionar uma wavelet, é possível indicar qual deve ser o nível de decomposição que o sinal receberá. Essa capacidade de decomposição tanto no domínio da frequência quanto no domínio do tempo faz com que as funções wavelet sejam ferramentas poderosas de processamento de sinais, podendo ser aplicadas na compressão de dados, eliminação de ruído, separação de componentes no sinal, identificação de singularidades e detecção de autossemelhança. O nível máximo de decomposição leva em consideração o comprimento do sinal de entrada e da wavelet (TASPINAR, 2018). Conforme o nível de decomposição do sinal aumenta, é expresso um incremento no número de amostras dessa wavelet.

A decomposição de um sinal com o uso da Transformada Wavelet possui duas variantes: contínua e discreta. Na forma discreta o sinal é dimensionado em um trecho de baixa frequência e outro de alta frequência. Já na variante contínua é suposto um deslocamento e escala do sinal único em toda reta real.

Os diferentes tipos de família que caracterizam os tipos de decomposição de uma wavelet contribuem para melhor adaptação aos resultados que se desejam encontrar. A família *Symlets* da análise discreta possui coeficientes de decomposição em passa-baixa e passa-alta, podendo realizar reconstrução do sinal em etapas ou por completo. Essas características afirmam a necessidade de realização experimental no presente objeto de estudo. A expectativa é de que o uso da transformada de wavelet, juntamente com suas características, venham aperfeiçoar os tratamentos dos sinais analisados, agregando robustez às já percorridas Transformada de Fourier.

## 2.5 Trabalhos Relacionados

É notável a grande variabilidade e aplicabilidade de dados e resultados obtidos através de acelerômetros nos diferentes cenários da atualidade, usando métodos de tratamento e processamento que melhor se adequam para cada uma das situações observadas. Um exemplo de aplicação está relacionado ao transporte viário de cargas especiais, como componentes de uma usina eólica, que necessitam de uma devida atenção quanto às condições do trajeto percorrido pelos caminhões transportadores e aos erros de planejamento de rota, onde os impactos podem gerar danos nas peças que compõem a estrutura dos aerogeradores, principalmente nas pás eólicas.

O trabalho desenvolvido por Guilherme Ferreira Horta (HORTA, 2020) propõe o uso de acelerômetros a fim de se mensurar os problemas que os impactos físicos podem causar a essas peças, registrando as oscilações de aceleração que estes componentes sofrem durante o seu deslocamento até o destino e instalação final. Para isso, Horta utilizou da função da Transformada de Fourier para avaliar as condições de uso da peça quando chega ao seu local de instalação. Por não impor restrições à classificação (dimensionalidade) de transformações multidimensionais, a FFTW (Biblioteca de Transformadas de Fourier) é executada visando encontrar a DFT (sigla em inglês para a expressão Transformada Discreta de Fourier).

Assim, após obter a aceleração medida pelo acelerômetro nos 3 eixos, eles são implementados pela transformada de Fourier para obtenção dos espectros de frequência, e suas amplitudes correspondentes, sendo estes espectros analisados e servidos como parâmetros para dizer a condição do objeto medido. Neste caso, a aplicação atingiu o objetivo de acelerar a análise de danos sofridos pela peça e pode culminar com o desenvolvimento de um modelo a ser seguido no transporte de cargas especiais.

Outro estudo de aplicação de acelerômetro, este voltado para as áreas da saúde e segurança, é relatado por Ning Jia (JIA, 2009) sobre a necessidade de identificar possíveis acidentes de queda dos seres humanos, como idosos (que podem sofrer quedas acidentais devido à fraqueza ou tontura relacionadas à idade avançada), e em outras condições e atividades para as quais um alerta imediato para uma possível queda, especialmente de altura substancial, seria bastante útil. Se enquadram como exemplos os montanhistas, operários da construção, lavadores de janelas e pintores de telhados.

Baseado em pesquisas sobre os princípios de detecção de queda para um corpo individual, Jia propõe uma nova solução para detecção destas situações, usando

um acelerômetro de 3 eixos. A técnica se baseia no princípio de detectar mudanças no movimento e na posição corporal de um indivíduo, usando um sensor, por meio do rastreamento das mudanças de aceleração em três direções ortogonais (JIA, 2009).

Para diferenciar movimentos do dia a dia, como descer e subir escadas, sentar-se e levantar-se de uma cadeira de uma possível queda, o autor faz comparações das variações encontradas nos três vetores de medição (X, Y e Z) e da força g (soma dos vetores), e com isso, obtêm características de um evento de queda que podem servir como critérios para detecção da mesma. Após dividir o evento em quatro etapas (início da queda, impacto, resultado e comparação antes e depois), Jia realiza combinações destas características para identificar o evento com um queda do paciente e, a partir disso, emita um alerta apropriado para tal.

A solução de detecção de queda proposta aproveita vantagens identificadas na utilização de acelerômetros (para este caso, o modelo ADXL345), tais como sensibilidade fixa de 4 mg / LSB, baixo consumo de energia (25  $\mu$ A a 130  $\mu$ A), interface digital serial padrão I2C® e SPI e armazenamento FIFO de 32 níveis (JIA, 2009). Além disso, faz-se uso de um microcontrolador analógico (ADuC7026) conectado ao acelerômetro. Pode-se observar a conveniência e flexibilização desta tecnologia para atender cada uma das situações observadas.

Já o trabalho desenvolvido por (ARIZA-LÓPEZ et al., 2019), apresenta um conjunto de dados que consiste em traços tridimensionais, capturados por técnicas do Sistema Global de Navegação por Satélite (GPS) com coordenadas tridimensionais, que podem servir como uma base de teste para pesquisas em aplicações de mineração de dados. Algumas ferramentas de informação geográfica usada por eles (como ArcGIS®, QGIS®, and GRASS®), fazem uso de linguagens de script (como o Python), por serem de grande ajuda para o processamento de conjuntos de dados de rastreamento e *multitrace*. A utilização de Python para estes casos demonstra a capacidade de compatibilidade da linguagem com os sistemas medidores mais usados atualmente.

Outro exemplo é relatado no artigo desenvolvido por Mircea Strutu e Dan Popescu (POPESCU; STRUTU, 2014), onde os acelerômetros 3D são usados em conjunto com módulos de captura de vídeo e processamento de dados para identificação de defeitos nas estradas chinesas. Além disso, aparelhos de notebook e tablets são conectados em rede wireless como centrais integradoras e de gerenciamento dos dados analisados. Definiram o eixo Z como mais importante para o trabalho, por estar perpendicular ao vetor do campo gravitacional da Terra. Para o tratamento de dados, os esforços foram centrados na identificação do método de filtragem mais adequado para detectar estradas e defeitos. O cálculo do desvio padrão foi demarcado para distinguir

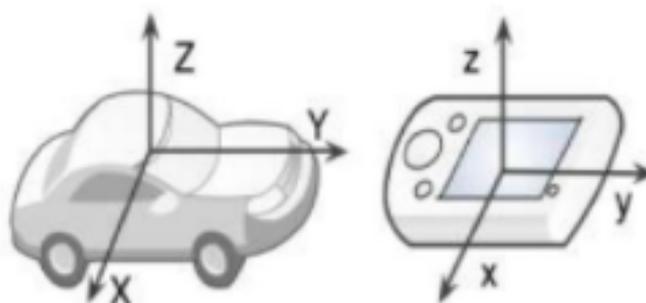
a granulação fina e áspera das estradas. A partir deste ponto, calculam a média móvel do sinal, e, inserindo limites como marcadores de defeito, fazem uso de filtros para distinguir e identificar os sinais de anomalias visados. Porém, com o uso de tal método, existe um grande risco de identificar falsos positivos e até mesmo de perder alguns defeitos.

Para mitigar estes possíveis problemas, Strutu e Popescu testaram diversos filtros nos sinais amostrados, tendo destacado no uso do filtro passa-baixa as estimativas mais precisas e menos falsos positivos identificados (POPESCU; STRUTU, 2014). A abordagem de filtrar o sinal do acelerômetro e definir um limite no sinal filtrado para encontrar os defeitos da estrada foi positiva, no entanto revela limitações quanto as possíveis oscilações que não seguem os padrões definidos de filtro dos sinais amostrados.

Trabalho semelhante foi desenvolvido pelo (ALLOUCH et al., 2017), que, a partir dos dados subtraídos do GPS, dos acelerômetros e giroscópio, fez uso do classificador de árvore de decisão denominado C4.5, na aplicação em dados de treinamento para classificar segmentos de estrada e construir o modelo, com precisão consistente de 98,6%. Durante o desenvolvimento, realizaram análises e comparações com três diferentes algoritmos de aprendizado de máquina: classificador C4.5 - árvore de decisão que cria uma árvore binária para modelar o processo de classificação, SVM e o algoritmo nomeado Naive Bayes - classificador que calcula um conjunto de probabilidades contando a frequência e as combinações de valores em um determinado conjunto de dados.

A técnica de (ALLOUCH et al., 2017) realizou o desenvolvimento de um aplicativo Android que permitiu a coleta e análise automática, em tempo real, dos dados do acelerômetro e do giroscópio para obter rótulos confiáveis da superfície da estrada. Além de dispor de um maior investimento para desenvolvimento da aplicação, a técnica impôs uma limitação a determinado sistema operacional, reduzindo abrangência de uma população mais diversificada. Após a coleta dos dados, conforme Figura 9, estes foram passados por um filtro passa-baixa visando a remoção de possíveis ruídos. A partir deste ponto, usando a FFT para transpor a amplitude ao longo do tempo em magnitude na frequência, foi realizado diversos armazenamentos de magnitudes junto às etiquetas fornecidas pela estrada. Assim, os dados processados foram recebidos no sistema, que extraiu características dos padrões das condições das estradas, e previu a qualidade da estrada com base no modelo classificador treinado, mostrando o traçado de localização da estrada em um mapa geográfico. Um histórico com as condições relatadas era salvo no próprio aplicativo.

**Figura 9 – Orientação do aparelho celular no interior do veículo**



Fonte: Adaptado de (ALLOUCH et al., 2017)

Para as comparações de desempenho entre os algoritmos classificadores, foram encontrados valores para sensibilidade, precisão e F1 (por lá denominados TP Rate, Precision e F-measure). Segundo (ALLOUCH et al., 2017), os melhores resultados foram encontrados durante o uso do classificador C4.5, apesar dos três terem registrados respostas bastante próximas entre si.

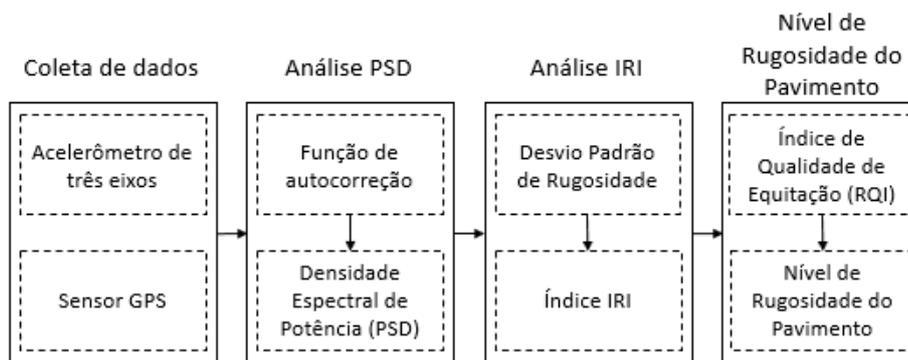
Se baseando nos níveis de rugosidade determinados e padronizados pela indústria chinesa, no trabalho de (KONGYANG CHEN et al., 2011) é mostrado como os recursos de captação de dados (acelerômetros de 3 eixos e o GPS) são relacionados à diferentes algoritmos com intuito de obter a correta classificação da rugosidade viária. Por meio da análise da densidade espectral de potência (do inglês, *Power Spectral Density* - PSD) da rugosidade do pavimento, é estimado o Índice Internacional de Rugosidade (do inglês, *International Roughness Index* - IRI) e, a partir deste ponto, a rugosidade do pavimento é definida e classificada.

Desse modo, o sistema desenvolvido obtém os dados dos acelerômetros, e os percorre em um algoritmo, relacionando-os com o PSD e o IRI, e assim obtendo a classificação desejada. Para captação de dados, um dos problemas relatados por eles está relacionado ao não funcionamento do GPS em algumas regiões da cidade, por interferência de prédios e túneis. Com isso, foi necessário deletar algumas zonas que não receberam o correto sinal do aparelho. Também foi necessário realizar a normalização dos dados para o Sistema Internacional de Medidas, sendo necessário definir o range de abrangência dos três eixos. Este processo foi identificado como função de auto correção.

Já a arquitetura do algoritmo foi desenvolvida modelando a métrica PSD na transformada de Fourier. Após, relacionam a métrica com o IRI, onde o desvio padrão da rugosidade é considerado como a variável intermediária. Utilizando da técnica de

análise de regressão, o IRI é determinado, sendo esta a variável que é condicionada à obtenção do Índice de Qualidade de Condução (do inglês, *Riding Quality Index - RQI*), métrica que compreende o índice de conforto ao dirigir - dividido em quatro níveis: excelente, bom, qualificado e não qualificado. A Figura 10 ilustra a estrutura do algoritmo.

**Figura 10 – Estrutura do algoritmo de nível de rugosidade do pavimento**



Fonte: Adaptado de (KONGYANG CHEN et al., 2011)

Após coletar dados de estradas reais com várias características, tais como: estradas lisas, estradas gerais (contendo pequenas partículas, como areia ou pedras), estradas com alguns buracos e bueiros e, por fim, estradas com algumas correias de desaceleração - especificidade regional, os resultados classificatórios obtidos foram satisfatórios, diferenciando os tipos de rugosidade quanto às condições apontadas acima. Além disso, observaram o eixo Z como mais representativo dos sensores, por mensurar o eixo vertical de acordo com o posicionamento do acelerômetro.

O trabalho desenvolvido por (CARLOS et al., 2018) propôs a criação de uma plataforma de base de dados viários que podem ser utilizados pela comunidade para geração de rodovias virtuais, a fim de desenvolver e aprimorar sistemas classificatórios baseados em acelerômetros. A plataforma web denomina-se Pothole Lab. Fazendo o uso da mesma, Carlos construiu diversas rodovias virtuais com intuito de usá-las como base de dados treinadoras para os algoritmos classificadores a serem comparados.

O estudo realizou comparações de diferentes métodos de classificação referenciados. Os trechos incluíram asfalto plano, buracos, lombadas normais e lombadas metálicas. Também foram utilizados diferentes veículos automotores para coleta dos dados reais. Contendo estas definições em mãos, os resultados calculados e apontados pelo trabalho de Carlos (CARLOS et al., 2018) demonstram que, através do algoritmo classificador SVM, foi possível encontrar a média de resposta para a nota

F1, semelhante à 0,785 (ou 78,5%), em sinais onde encontram-se diferentes tipos de rodovias e imperfeições.

## 2.6 Tratamentos de dados e suas características

Os tratamentos dos dados captados podem utilizar de diversas ferramentas existentes no mercado que contribuem para uma melhor visualização e entendimento do usuário perante à um dado bruto. A presença de características irrelevantes ou redundantes em uma base de dados pode afetar negativamente o desempenho de algoritmos classificadores (NUNES et al., 2018). Com isso, dentre os principais meios de tratamento deste tipo de dado atual, destacam-se três: análise estatística, machine learning e análise no domínio da frequência.

A análise estatística é bastante utilizada para encontrar resultados específicos e pontuais em relação a determinada observação. Faz uso de métodos como vetores resultantes, desvio padrão, variância e diferenças entre medidas das vibrações para detecção das anomalias viárias, por exemplo. Já a técnica de machine learning é baseada no desenvolvimento e / ou adequação de uma Rede Neural Artificial (RNA), que utiliza de vetores para treinamento da mesma e, assim, tendo sua base definida, realiza comparações e classificações de acordo com os parâmetros predefinidos pelo usuário.

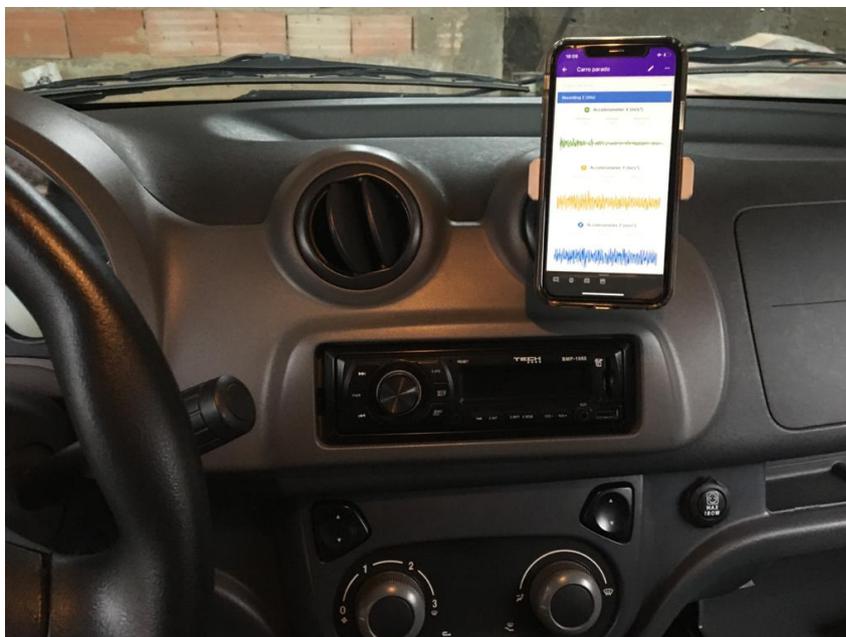
A utilização do machine learning contém vantagens quando diz respeito à assertividade de classificação, visto que quanto mais dados são processados pela rede neural, mais bem afinado e definido será o padrão de classificação da mesma. Ou seja, tal rede pode canalizar e se adaptar aos diferentes tipos variáveis, como as variações encontradas entre celulares e veículos, reduzindo os erros de classificação.

No terceiro tratamento de dados descrito, o método de domínio da frequência utiliza da transformada rápida de Fourier (do inglês, *Fast Fourier Transform*) para adequar o sinal na frequência desejada. É um caminho que tem sido bastante utilizado quando relacionado à uma melhora e alinhamento do sinal amostrado perante determinada frequência.

Para a realização das medições iniciais, foi utilizado como veículo de teste um Fiat Uno 2012 Vivace. Conectado por um suporte plástico no console central do veículo, o aparelho celular de medição é do modelo Apple Iphone XR (utiliza o IOS como sistema operacional), que contém os sensores necessários para aquisição dos dados, conforme Figura 11. Posteriormente, realizou-se variações dos aparelhos celulares (com sistema operacional Android) e dos veículos utilizados (com motores movidos à

óleo diesel e flexíveis - gasolina e etanol).

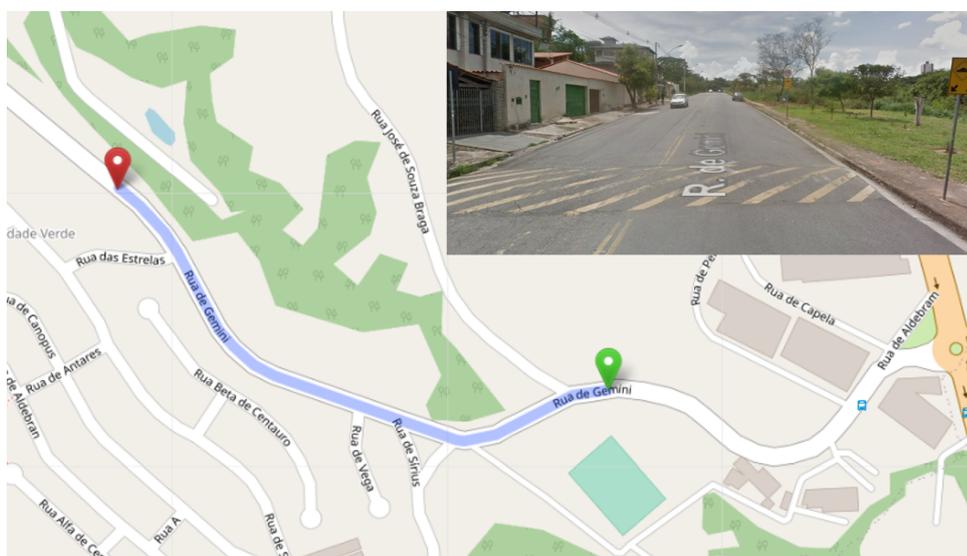
**Figura 11 – Exemplo de posicionamento do aparelho celular no veículo**



Fonte: Autoria própria.

As medições foram realizadas em diversos trechos selecionados conforme a identificação prévia da presença imperfeições. Na figura 12 é apresentado um exemplo dos trechos de estudo, que neste caso contém, em destaque, cinco lombadas conforme demonstrado no canto superior direito da figura. O trecho está localizado na rua Gemini, bairro Cidade Verde, na cidade de Betim - MG.

**Figura 12 – Trecho de estudo e imperfeição identificada**

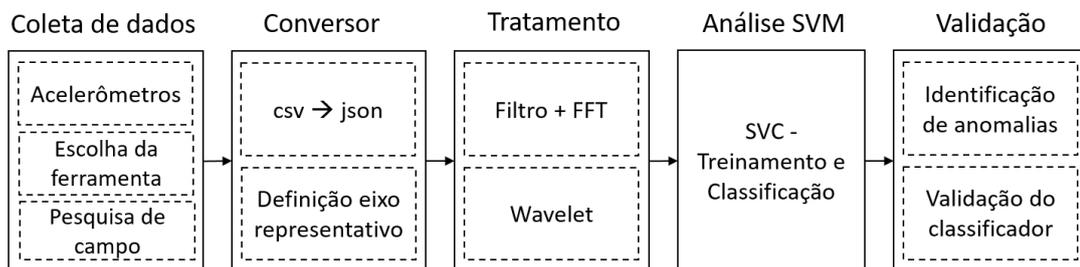


Fonte: Autoria própria.

### 3 DESENVOLVIMENTO

Levando em consideração o referencial teórico descrito no Capítulo 2, o desenvolvimento do trabalho foi planejado conforme mostrado na figura abaixo. O percurso pelas cinco vozes ilustradas e sequenciadas foi distribuído de forma a atender os objetivos almejados.

**Figura 13 – Fluxograma de Desenvolvimento**



Fonte: Autoria própria.

#### 3.1 Escolha da ferramenta de coleta

O primeiro passo tomado para validação da metodologia avaliada é referente a adequação dos dados captados pelo aplicativo Physics Toolbox Accelerometer, visando obter a correta interpretação destes pelo SVM. Através do *script* desenvolvido em python, foi possível utilizar os dados locais no sistema, sem a necessidade de alterar o código base.

Com isso, chega-se ao principal questionamento deste estudo, onde é esperado que o sistema leia a sequência de dados dos acelerômetros e identifique / processe a existência de uma anomalia ou não, através do treinamento do sistema realizado utilizando como vetores os dados gerados virtualmente no portal mencionado na Seção 2.5, o Pothole Lab. Para solucionar tal questão, antes é necessário demonstrar o uso de três cálculos de pontuações que avaliam este processo: Sensibilidade (TPR), Precisão (PPV) e a pontuação (F1).

$$TPR = \frac{TP}{TP + FN} \quad (1)$$

$$PPV = \frac{TP}{TP + FP} \quad (2)$$

$$F1 = \frac{2TP}{(2TP + 2FP + FN)} \quad (3)$$

De acordo com o trabalho de base (CARLOS et al., 2018), a Sensibilidade pode ser definida como taxa de verdadeiro positivo (do inglês, *True Positive Rate*, TPR), faz a medição da proporção de exemplos positivos verdadeiros que foram identificados corretamente. A Precisão (no inglês chamado de *Positive Predicted Value* - PPV), mede a proporção de exemplos verdadeiros positivos que ficaram de fora do número total de detecções, ou seja, não foram identificados. A pontuação F1 (do inglês, *F-measure*), é uma medida de precisão de um teste e pode ser interpretado como o harmônico de Precisão e Sensibilidade, onde uma pontuação F1 atinge seu melhor valor em 1 e pior em 0.

Demonstrado os valores, é idealizado que um Verdadeiro Positivo (TP) é uma detecção feita pelo algoritmo dentro de um intervalo conhecido por conter uma anomalia. Um Falso Negativo (FN) é determinado quando nenhuma detecção é feita pelo algoritmo dentro de um intervalo conhecido por conter uma anomalia. E um Falso Positivo (FP) é uma detecção feita onde nenhuma anomalia é conhecida no intervalo de análise.

Para realizar as comparações dos padrões almeçados com os dados de anomalias, foram zeradas as variáveis TP, FN e FP, onde são definidos que o TP receberia as verdadeiros positivos, o FN os falsos negativos e o FP os falsos positivos. Conforme o sinal era percorrido e contagem realizada, o contador que obtivesse a variação entre os limites inferiores e superiores era somado ao contador, identificando assim a presença de uma anomalia.

Uma das vantagens de se utilizar a metodologia SVM é a possibilidade de adaptação de suas propostas através das importações de bibliotecas mais adequadas para determinado projeto. Levando em consideração que este trabalho busca obter uma nota F1 que se aproxima dos valores no trabalho referenciado, as características do SVM também foram mantidas conforme originalmente proposto por (MEDNIS et al., 2011) na Tabela 1. A variável  $g$  está representando o valor da gravidade - 9,81 metros por segundo ao quadrado ( $m/s^2$ ).

**Tabela 1 – Valores das Características do SVM.**

Características	Valores
Média	$g*0,3$
Desvio Padrão	$g*0,15$
Coeficiente de Variação	$g*0,015$
Diferença (Max-Min)	$g*0,2$
Soma da pontuação de confiança	3

Fonte: Adaptado de (CARLOS et al., 2018).

## 3.2 Pesquisa de Campo

Para que pudesse ter uma maior abrangência de dados, foi desenvolvido um questionário virtual para que voluntários anônimos pudessem contribuir com medições realizadas, respectivamente, em seus aparelhos celulares e veículos, diversificando tais variáveis. Além disso, o uso dos dados gerados pelos voluntários tem como objetivo analisar a aplicabilidade do estudo em áreas e pavimentações dispostas além do perímetro citadino de Betim.

Os voluntários foram instruídos a instalar o aplicativo Physics Toolbox Accelerometer nos aparelhos celulares, e, durante as medições, realizassem o posicionamento do aparelho em uma das duas formas solicitadas: deitado / horizontal (podendo ser no banco, no console central ou mesmo no assoalho do veículo, desde que esteja com a tela virada para cima e esteja fixado nesta posição - evitando assim variações do ponto de referência do aparelho e, conseqüentemente, dos acelerômetros). Outra possibilidade divulgada aos voluntários foi de posicionar o aparelho disposto na vertical - como por exemplo, no painel de controles ou no console central, também fixado em algum suporte que evite sua locomoção durante as medições.

Foi proposto para que cada medição sempre iniciasse e finalizasse com o veículo totalmente parado, e em um local relativamente plano. O intuito é reduzir os possíveis distúrbios e ruídos causados pela aceleração e retardo do veículo, semelhante aos testes iniciais. Além disso, os voluntários foram instruídos a conduzirem o veículo de maneira cotidiana na qual o mesmo já está habituado, levando em consideração todas as regras de trânsito impostas para aquela determinada via (como placas e sinais de trânsito por exemplo) e as condições gerais da estrada. Durante o envio do arquivo para a pesquisa, os voluntários foram convidados a preencher um questionário, onde eram solicitadas algumas informações, tais como: em que posição o celular foi apresentado, qual a avaliação pessoal sobre o estado de conservação do trecho percorrido (se passou por lombadas, valetas, buracos, trepidações), além dos respectivos modelos do aparelho e do carro utilizados.

Contudo, os esforços para ampliar a varredura de cada uma das variáveis aplicadas não compreenderam todos os espectros desejados com níveis de satisfação esperado. Apesar da baixa arrecadação de dados de voluntários em relação ao tamanho e diversidade da malha viária existente na região, a implementação do questionário

voluntário possibilitou analisar a capacidade de aplicabilidade da plataforma de captura de dados selecionada, contendo funcionamento adequado em diversos modelos de celulares, sem registros de falhas durante o processamento.

Com os resultados desenvolvidos e exemplificados na seção Resultados, é possível adiantar a importância de se estabelecer e manter fixado determinados parâmetros para realização de medições que possuem envolvimento de voluntários, tais como velocidade média do veículo, posicionamento interno do aparelho celular, duração dos ensaios, conhecimento prévio das condições viárias. A correta denominação destas contribui para uma melhor compreensão e assertividade do algoritmo testado.

### 3.3 Conversor

A classificação das imperfeições realizada pelo sistema em estudo se dá pela leitura dos arquivos tipo Python (.json, ou JavaScript Object Notation). Estes arquivos são considerados modelos para armazenamento e transmissão de informações no formato texto (DEV MEDIA, 2021), com capacidade de estruturar informações de forma mais compacta do que a conseguida pelo modelo XML, tornando mais rápido a análise das informações ali contidas. Para realizar a conversão dos arquivos gerados em CSV para JSON, foi desenvolvido um *script* em python, realizando as padronizações das separações entre as colunas do arquivo, além dos seus respectivos nomes, visto que cada um dos aplicativos usados detinha um formato de organização diferente. Para conversão, inicia-se os contadores em zero, e cada um dos eixos de leitura (X, Y e Z) são definidos como classes listáveis (list).

Após, é realizado a leitura do arquivo de entrada, e a vírgula é definida como caractere delimitador de colunas. A cada leitura, os eixos X, Y e Z recebem os dados tipo float das colunas gFx, gFy e gFz, substituindo seus respectivos delimitadores decimais de virgula para ponto - seguindo o mesmo padrão utilizado pelos pesquisadores M. R. Carlos, M. E. Aragón (CARLOS et al., 2018).

O sistema também requer que o arquivo contenha dados relacionados a velocidade. Como padrão, a velocidade inicial (chamada de start) de todos os testes foi definida como zero e final possuiu alguns testes em 20 km/hora, e outros em 30 km/hora.

### 3.4 Definição do eixo representativo

Antes de prosseguir com as análises é necessário realizar a identificação de qual dos três eixos mensurados possui a maior representatividade para a classificação, que possui uma variação de acordo com o posicionamento do aparelho celular feito durante a medição. Ou seja, para as medições realizadas neste estudo, os eixos que possuem melhores descrições são o eixo Y e o eixo Z - aparelho na vertical e na horizontal, respectivamente.

Observando como os aparelhos celulares podem ficar posicionados no interior do veículo durante um trajeto, foram realizados testes nas duas posições destacadas como mais comuns: vertical e horizontal, tomando como ponto de referência a tela do aparelho, conforme ilustrado na Figura 5. Assim, caso os sensores do celular tenham feito os registros na posição vertical, o acelerômetro de estudo selecionado será o que representa o eixo Y, enquanto estando o aparelho deitado, o sensor que melhor representa as variações / vibrações do veículo perante o solo é o eixo Z.

### 3.5 Tratamento de dados

Como somente a conversão dos dados captados do domínio do tempo para o domínio da frequência - através da função FFT - não obteve resultados esperados para um tratamento completo, tal procedimento inicial foi incluído como uma etapa do desenvolvimento do filtro, e cada novo arquivo foi submetido diretamente para a conversão de sua extensão, transformando-os do tipo csv para json, possibilitando a leitura e processamento destes pelo classificador SVC.

Antes de passar pelo classificador, foram identificados cada um dos pontos anômalos nos sinais e exibidos juntamente como resposta na saída da classificação, contendo o instante inicial e final onde ocorre a anomalia dentro do espectro. Tal ação foi determinada visando contribuir para uma melhor identificação visual das anomalias nos sinais processados. Além disso, é registrada a velocidade média do veículo durante o percurso de análise, em quilômetros por hora (Km/h), partindo e encerrando sempre da imobilidade (0 Km/h).

Após leitura dos dados, foi implementada uma nova etapa de tratamento, com o intuito de adequar a frequência dos dados em relação ao tempo, usando o método FFT. Foi desenvolvida uma função para realizar a Transformada de Fourier dos sinais e executar a plotagem no domínio frequência. Este processo foi realizado visando uma padronização dos sinais amostrados - em relação à frequência, e comparar com os dados iniciais, além de obter uma melhor visualização gráfica. Porém, o uso do método

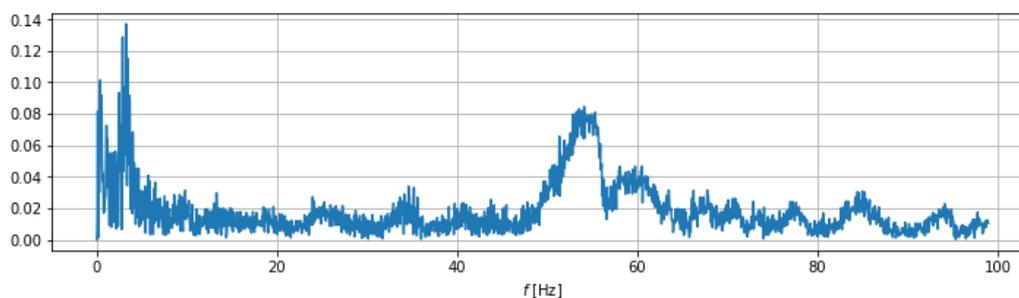
isoladamente não implicou em uma variabilidade desejada para avançar um melhor discernimento entre os resultados.

No trabalho desempenhado por M. R. Carlos, M. E. Aragón, L. C. González, H. J. Escalante e F. Martínez (CARLOS et al., 2018), foi desenvolvido um site gerador de dados de acelerômetros virtual com capacidade de distinção de três tipos de imperfeições que podem ser encontradas na região de estudo ([https://www.accelerometer.xyz/pothole<sub>lab</sub>/](https://www.accelerometer.xyz/pothole_lab/)), mas que pode servir de base para realização de testes neste trabalho em questão. Com isso, é possível realizar comparações entre o método utilizado por eles, além de desenvolvimento de viabilidade de aplicação do método criado por eles ser implementado nas vias locais em estudo.

Através do site Pothole Lab, foram gerados, para cada um dos três tipos de imperfeições (Speed bumps, Metal bumps e Potholes), arquivos de dados com duração de dez segundos e cinco pontos de imperfeições. Desse modo, foi possível verificar como o sistema se comporta de acordo com cada tipo de imperfeição selecionada.

Para análise, desenvolvimento e geração de testes da codificação dos arquivos do tipo Python (extensões “.py”), foi utilizada a extensão *Python extension for Visual Studio Code* do software Visual Studio Code. Com isso, foi possível realizar a primeira verificação do funcionamento projeto pelo autor com dados gerados através do software online.

Observando uma oportunidade de aprimorar ainda mais a classificação do algoritmo, foi avaliado o uso da técnica de filtragem do sinal amostrado, do qual compreende uma quantidade considerável de ruídos que acabam poluindo as oscilações das imperfeições visadas. A capacidade de amortecimento dos veículos (que possuem características distintas quanto à sistemas de suspensão, tamanho das rodas, tipo de pneu empregado) contribuem para criação de novas variáveis e ruídos que devem ser desconsiderados neste estudo. Além disso, as variações previstas durante às alterações do estado do veículo (acelerando ou retardando) geram desestabilização do sinal, podendo representar falsas anomalias ou ter dificuldade de distinguir o sinal anômalo de um trecho normal.

**Figura 14 – Exemplo de sinal em frequência [Hz]**

Fonte: Autoria própria.

As frequências de passagem selecionadas durante a configuração do filtro foram determinadas visando eliminar possíveis ruídos encontrados nas extremidades dos sinais. Com análise espectral da frequência global, foi notado que, a maioria dos sinais possuem uma quantidade variável de energia espalhadas acima dos 50 radianos por segundo (rad/s). Desse modo, foi determinada a frequência de rejeição. Isso possibilitou a melhor visualização das energias armazenadas próximas dos instantes onde foram registradas as anomalias de cada um dos sinais.

Assim, a atenuação da faixa de passagem escolhida foi de 0,1 decibéis (dB). Enquanto a atenuação definida para a faixa de rejeição possui valor de 50 dB. Além disso, a frequência angular de passagem foi dada por 20 radianos por segundo (rad/s), e a de rejeição equivalente à 50 rad/s. Tais definições foram tomadas levando em consideração às melhores respostas encontradas durante a fase de testes, com o objetivo de que a banda de passagem não fosse comprometida e desequilibrasse a saída desejada.

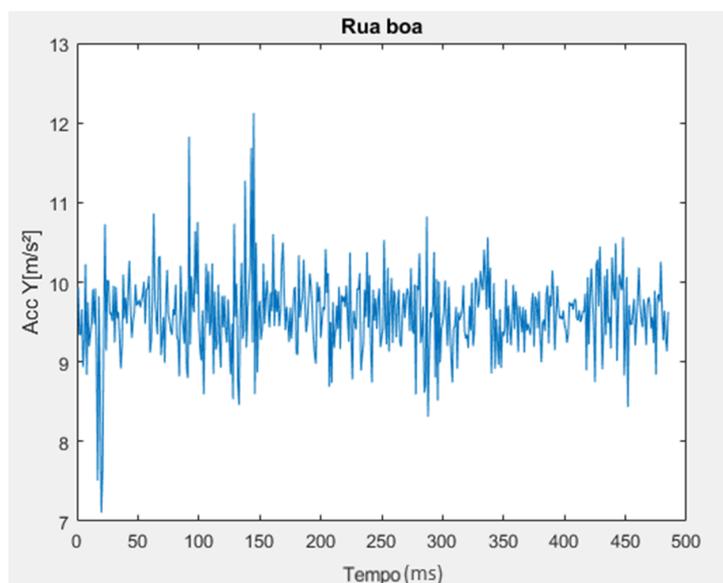
## 4 RESULTADOS

### 4.1 Análise e aplicação do SVM

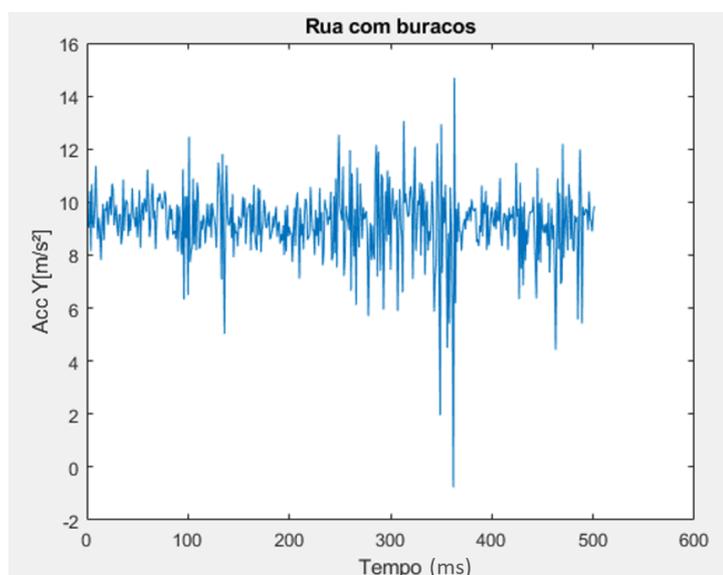
Já no início dos primeiros testes de leitura em aparelhos com sistema operacional Android, foi observado uma instabilidade do aplicativo Science Journal na leitura dos acelerômetros. Desse modo, a aplicação Physics Toolbox Accelerometer foi selecionada para uso neste objeto de estudo, visto que tem as mesmas funcionalidades do anterior, também é gratuita, com rápida geração de respostas e apresentou boa funcionalidade em ambos sistemas operacionais. Os dados mensurados neste aplicativo servirão de entrada para os softwares de simulação.

Durante a etapa de testes, pode-se distinguir o trabalho em duas etapas iniciais de análise: a primeira foram medições realizadas em dois trechos distintos de qualidade viária, enquanto na segunda foram os testes voltados para detecção e identificação de quais são as imperfeições nas vias, ou seja, trechos com boa qualidade asfáltica, porém com pontos específicos onde eram possíveis encontrar buracos e lombadas isolados. Na primeira etapa analisada, foram escolhidos dois trechos de ruas localizadas em determinada região da cidade de Betim, demarcadas onde eram perceptíveis visualmente as diferenças de qualidade entre si, uma rua com asfalto considerado de boa qualidade (rua-bona), e uma que continha diversas imperfeições em sua composição asfáltica (rua-ruim).

As medições foram realizadas utilizando os três eixos do acelerômetro (X, Y, e Z). O aplicativo de leitura utilizado armazena os dados lidos em um arquivo do tipo .csv. Levando em consideração que o celular se encontrava na posição vertical - o eixo X representa movimentação lateral / longitudinal, o eixo Y a movimentação vertical e o eixo Z a movimentação para frente ou para trás do veículo (ortogonal). Assim, as comparações realizadas foram baseadas nos resultados encontrados no eixo Y. As respostas dos acelerômetros em relação ao tempo podem ser observadas nas Figuras 15 e 16. Os dados são organizados em uma única coluna conforme a seguir: relative-time,AccX,AccY,AccZ. Dependendo do aplicativo utilizado, ainda podem trazer o sensor de aceleração linear (LinearAccelerometerSensor) e a força G (força-G).

**Figura 15 – Eixo Y de Rua boa**

Fonte: Autoria própria.

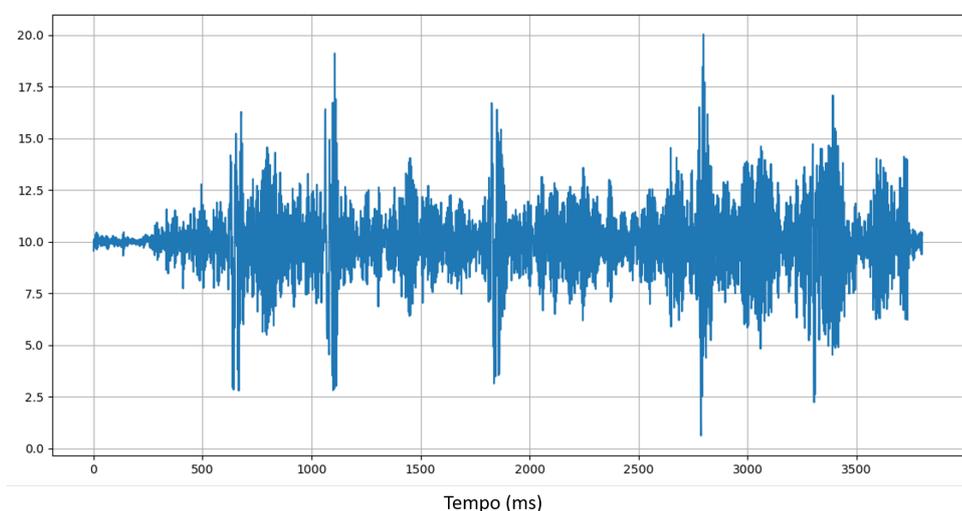
**Figura 16 – Eixo Y de Rua ruim**

Fonte: Autoria própria.

Em uma análise visual inicial, é possível determinar que, apesar da variação máxima (também chamado de range) de medição do acelerômetro na rua ruim ter sido maior (14m/s<sup>2</sup> contra aproximadamente 6m/s<sup>2</sup> da rua considerada de boa qualidade), não existem grandes diferenças visuais entre ambos sinais. As duas respostas apresentaram picos em trechos aleatórios (onde não havia nenhum registro de evento característico), com uma variabilidade desordenada. Este procedimento reforça a necessidade de buscar uma análise em diferentes domínios do sinal, além da eliminação de interferências.

Na segunda etapa inicial de análise, foram realizadas medições em trechos onde a maior parte do trecho estivesse em bom estado, mas que houvesse algumas anomalias isoladas localizadas durante o percurso (evento característico). Na Figura 17, o resultado inicial lido pelo eixo Y do acelerômetro.

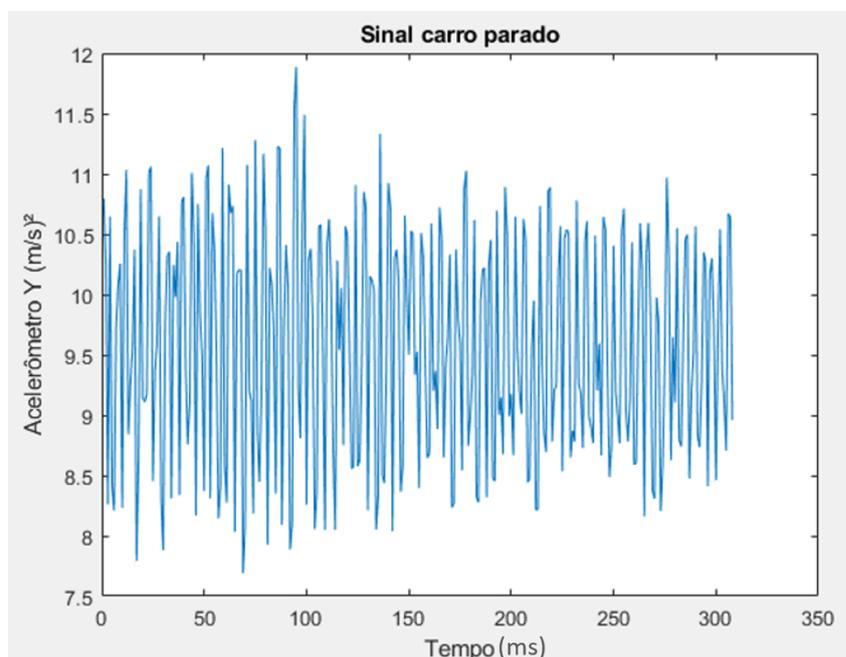
**Figura 17 – Eixo Y de Rua boa com buracos localizados**



Fonte: Autoria própria.

Considerando que, mesmo parados, os veículos automotores geram uma vibração na cabine devido ao funcionamento do motor a combustão, foi realizada também a medição dos acelerômetros na posição inicial em uma pista plana e o câmbio o veículo em posição neutra (comumente chamado de N ou ponto morto), observado na Figura 18. Além disso, não houve acionamento de nenhum dos pedais do veículo (acelerador, freio e embreagem) durante o processo de medição, que teve duração de aproximadamente dez segundos.

Figura 18 – Eixo Y com o veículo parado.



Fonte: Autoria própria.

## 4.2 Cálculo de Amostragem

Para a realização dos seus respectivos estudos e validações, M. R. Carlos, M. E. Aragón, L. C. González, H. J. Escalante e F. Martínez (CARLOS et al., 2018) fizeram uso de uma amostra de 30 trechos de rodovias geradas virtualmente pelo Pothole Patrol, e cada uma delas continha uma determinada quantidade de anomalias analisadas (no caso dividido entre buracos - *potholes*, valetas - *bumps*, e valetas metálicas - *metal bumps*). Considerando que este estudo possui interesse em identificar as imperfeições e anomalias presentes nas vias, indiferente de suas características específicas, não houve distinção entre tipos de imperfeição nos dados gerados artificialmente para o treinamento do classificador.

Desse modo, para chegar no valor de amostragem adequado para este trabalho, foi utilizada a equação que define o número de indivíduos na amostra ( $n$ ), onde:  $Z_{\alpha/2}$  é o valor crítico que corresponde ao grau de confiança desejado (neste caso, é esperado grau de 95% de confiança, ou seja, valor crítico  $Z_{\alpha/2} = 1,96$ ),  $dp$  é o desvio padrão populacional da variável estudada e  $E$  é a margem de erro ou erro máximo de estimativa, identificado pela diferença máxima entre a média amostral e a verdadeira

média populacional.

$$n = \left( \frac{Z_{\frac{\alpha}{2}} \cdot dp}{E} \right)^2 \quad (4)$$

Para encontrar o valor de amostragem mínimo necessário, a margem de erro considerada foi de 10 pontos percentuais, para mais ou para menos. Através do cálculo realizado, foi possível identificar que, para atingir os padrões determinados e semelhantes ao trabalho de (CARLOS et al., 2018), eram necessários pelo menos 13,34 amostras de trechos analisados. O cálculo sugere que, para resultados decimais, haja um arredondamento superior. O desvio padrão da variável F1 submetida foi de 0,1864. O resultado está presente na Tabela 2.

**Tabela 2 – Cálculo de Amostragem.**

n	Grau de Confiança	Margem de Erro	Desvio Padrão
14	95,0%	10,0%	0,1864

Fonte: Autoria própria.

Levando estes dados em consideração e visando reduzir possíveis erros de classificação, e também de medição, foram gerados 30 testes para representar os sinais virtuais, e outros 30 testes foram gerados usando o aplicativo Physics Toolbox Accelerometer. Estes últimos tratados como testes reais. As respostas descritas na Seção Resultados Encontrados foram baseadas nestas amostras encontradas.

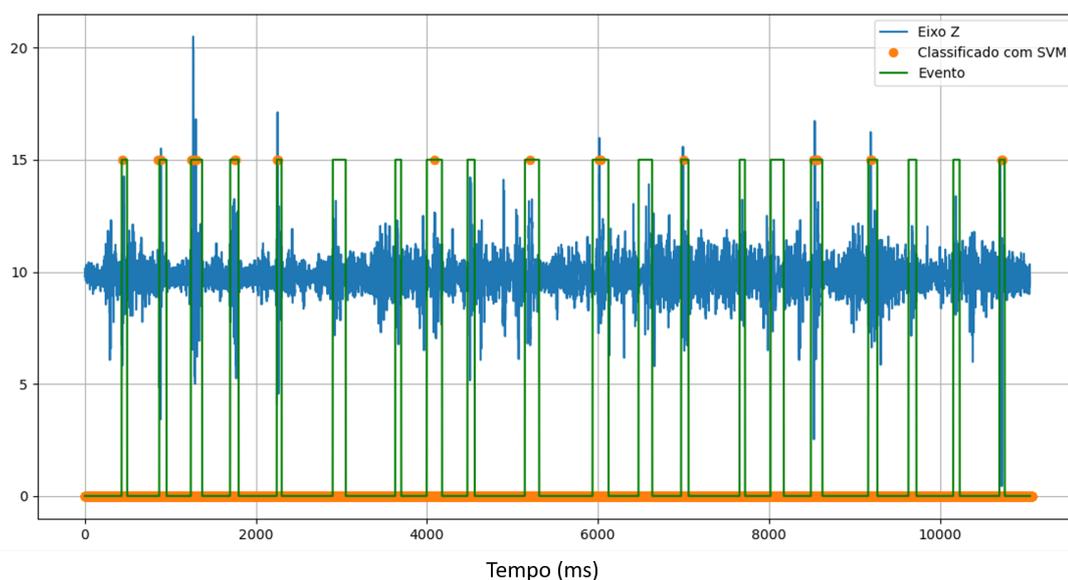
### 4.3 Resultados Encontrados e Validação

Nesta seção serão apresentados os resultados e suas respectivas avaliações quanto à aplicabilidade de acelerômetros na identificação de anomalias nas vias betinenses. É importante ressaltar que as apresentações a seguir foram comparadas sobre as mesmas condições do trabalho incorporado, e, portanto, é esperado que estes resultados sejam compatíveis e próximos aos encontrados anteriormente. Além disso, é desejado que as respostas encontradas possam ser avaliadas pelos autores referenciados e pela comunidade a fim de obter novos comentários e sugestões, sobre os limites e possíveis oportunidades das quais este trabalho possa aprofundar.

As apurações encontradas visam certificar o funcionamento da codificação adaptada. Para isso, foram utilizados os arquivos gerados via aplicação Pothole Lab, que já haviam sido previamente avaliados pelo trabalho anterior e disponibilizados para consulta, juntamente com o código classificador. Os resultados obtidos foram

satisfatórios, atendendo a expectativa de verossimilhança com os apresentados no trabalho referenciado. Apesar disso, os resultados mostrados revelam algumas falhas na identificação, que podem ter sido causadas pela grande quantidade de ruído presente nos sinais, apresentando, assim, espaço para melhorias durante a etapa de tratamento prévio.

**Figura 19 – Registro de sinal de acelerômetro sem tratamento.**



Fonte: Autoria própria.

Cada coluna representa o início e o fim da duração de uma imperfeição presente na via. Os pontos coloridos visíveis no topo das colunas definem a classificação do algoritmo. Ou seja, onde está presente um ponto representa que o classificador identificou uma anomalia naquele instante do sinal. A partir deste resultado, conforme descrito durante o período de captação e tratamento dos dados, foram gerados mais arquivos de dados na própria aplicação, visando ampliar e avaliar a adaptabilidade do sistema perante as anomalias simuladas no Pothole Lab.

Após os testes iniciais com os sinais simulados eletronicamente, foi a vez de atuar nas amostras reais adquiridas. Para a validação, a amostra de dados reais recolhida teve na conversão seu primeiro tratamento de dados, conforme descrito na etapa Tratamento e Formatação de dados. Após os arquivos convertidos, os mesmos foram alocados e executados pelo algoritmo classificador, que obteve como resposta a identificação de anomalias nos gráficos de saída, contendo representação semelhante aos sinais simulados, ou seja, por círculos coloridos. Para os dados coletados com os celulares na posição vertical, o eixo do acelerômetro utilizado para comparação e classificação é o Y, enquanto as medições registradas com o aparelho na posição horizontal tiveram esforços direcionados no eixo Z.

Os trechos mensurados com o aparelho celular posicionado horizontalmente dentro do veículo apresentaram resultados insatisfatórios, mesmo considerando o eixo Z como representativo. Os demais eixos também foram verificados e resultaram classificações igualmente indesejadas. Na Tabela 3, estão dispostos os valores de sensibilidade, precisão e nota F1 para estes sinais. Apesar de ter sensibilidade satisfatória, o classificador não foi preciso, detectando muito falsos positivos, contribuindo para um desempenho de acerto de apenas 41,1%. Dessa forma, optou-se por desconsiderar as mensurações realizadas de maneira horizontal. Os esforços de tratamentos e análises proferidos a seguir foram desempenhados levando em consideração somente os sinais registrados na posição vertical.

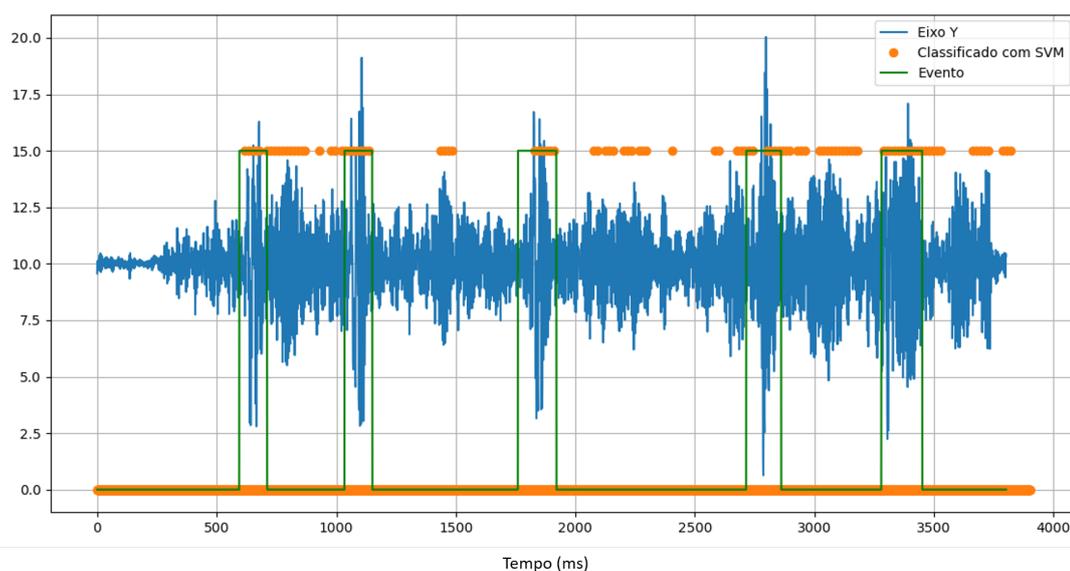
**Tabela 3 – Sensibilidade, Precisão e F1 - Medições Horizontais.**

	Sensibilidade	Precisão	F1
Sem Filtro	81,0%	30,0%	41,1%

Fonte: Aatoria própria.

A figura a seguir representa uma classificação obedecendo as premissas descritas anteriormente, neste caso, fazendo uso do eixo vertical (eixo Y relevante). É possível notar que, mesmo sem a utilização de filtros ou ajustes sinalizados, o classificador SVM obteve considerável êxito ao transmitir as imperfeições no sinal onde as imperfeições estão localizadas. Porém, neste exemplo foram encontradas muitas definições de falsos positivos, possivelmente gerados pelos ruídos indesejados.

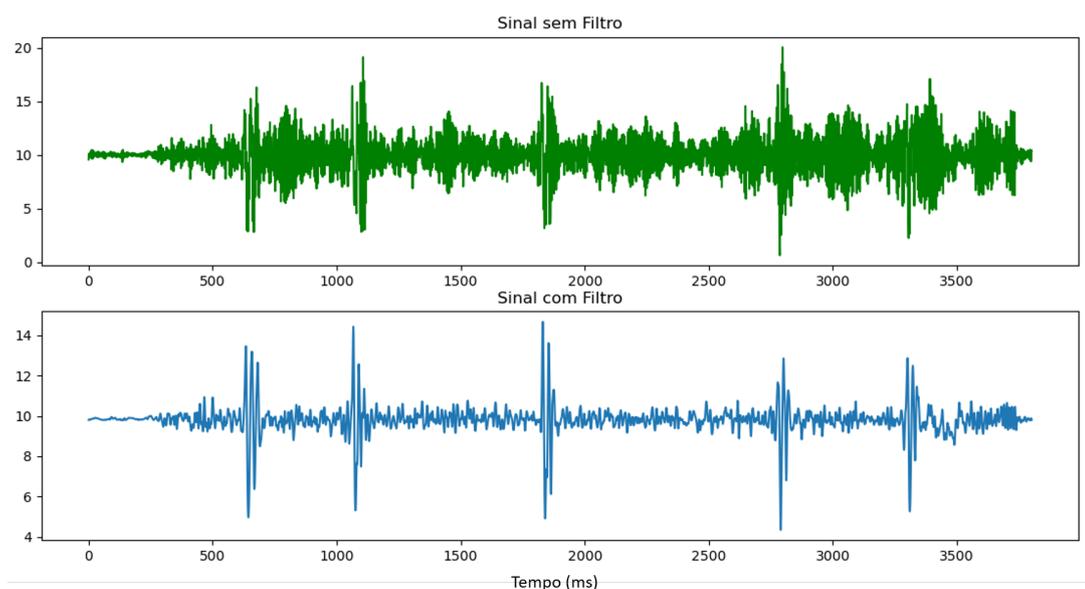
**Figura 20 – Registro de sinal classificado sem filtro.**



Fonte: Aatoria própria.

Após esta validação, o mesmo arquivo classificado anteriormente foi analisado após atravessar o filtro passa-baixa desenvolvido. Tal medição visa comparar a eficácia do filtro em reter os ruídos contidos nos sinais e melhorar a assertividade do algoritmo classificador, porém sem restringir os verdadeiros sinais anômalos. Abaixo, é possível verificar as diferenças encontradas do mesmo sinal antes e após a passagem pelo filtro.

**Figura 21 – Sinal original vs Sinal filtrado.**

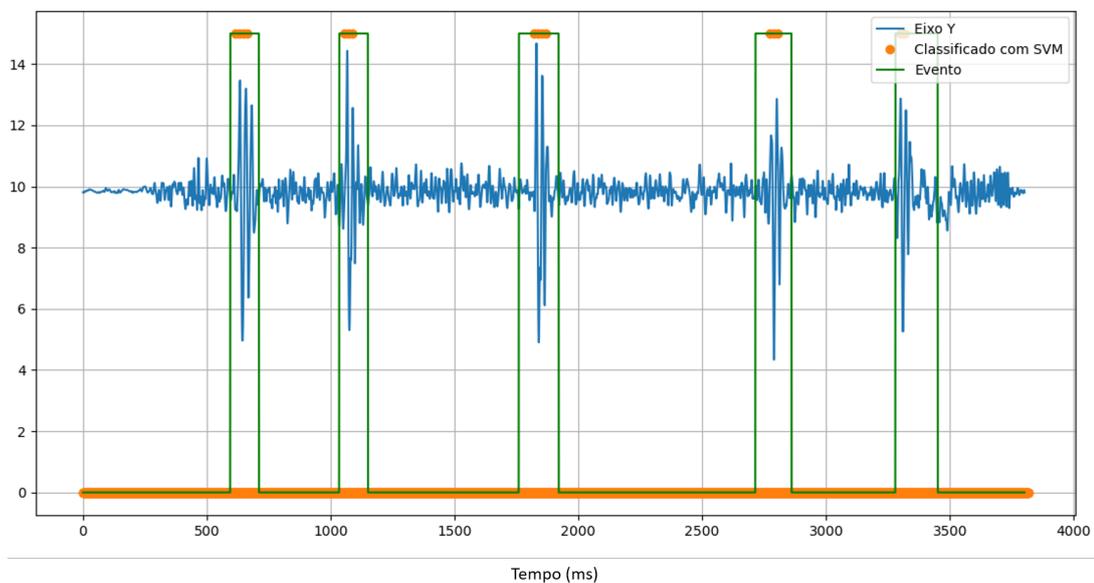


Fonte: Autoria própria.

O processo desempenhado pelo filtro trouxe algumas alterações visíveis quando se comparado com o sinal ainda virgem. É notável que, após o tratamento, o sinal estabelece uma variação (range) menor, reduzindo o valor médio global do sinal.

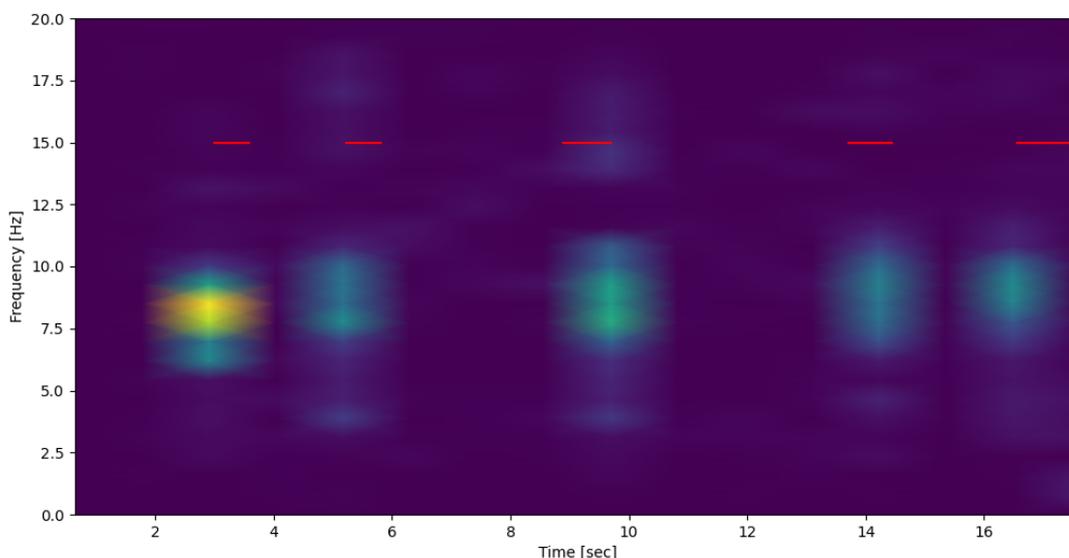
Em adjunto à passagem do sinal pelo filtro, foi realizada uma análise espectral das frequências ali presentes antes e depois do procedimento. Importante ressaltar como a análise espectral plotada contribui para melhores visualizações da energia armazenada nos sinais correspondentes às localizações das anomalias citadas, o que acaba por facilitar a visualização da região exata onde os trechos anômalos estão posicionados no espectro de frequência.

**Figura 22 – Registro de sinal classificado com filtro passa-baixa.**



Fonte: Autoria própria.

**Figura 23 – Espectrograma do sinal após filtro passa-baixa.**



Fonte: Autoria própria.

Apesar do espectro não conter uma visão de grande destaque, retendo as frequências destoantes, é possível notar a discrepâncias de energia armazenada durante os períodos anômalos e os certificados como padronizado (trecho bom). É evidenciado que cada anomalia possui uma determinada quantidade de energia armazenada, e dificilmente serão encontradas representações de sinais idênticos entre si, o que contribui para a alta variabilidade entre eles. Estes apontamentos representam a dificuldade de se estabelecer um único padrão de tratamento e classificação para regiões de características distintas. Na Tabela 4 abaixo, é possível observar as diferen-

ças de mensuração médias encontradas nos valores de sensibilidade, precisão e nota F1, antes e depois da utilização do filtro passa-baixa.

**Tabela 4 – Sensibilidade, Precisão e F1 - Medições Verticais.**

	Sensibilidade	Precisão	F1
Sem Filtro	83,89%	58,71%	60,80%
Com Filtro	60,00%	99,07%	71,70%
Filtro e Wavelet	63,52%	97,22%	74,13%

Fonte: Autoria própria.

A comparação entre os resultados das classificações com e sem o uso do filtro de atenuação das frequências valida a melhora percebida durante o cálculo de amostragem. Porém, é necessário destacar que a utilização de filtros que são configurados para realizar restrições de passagens de determinadas frequências, podem comprometer a estruturação do sinal analisado, impossibilitando a geração de respostas de classificação desejadas. Após a filtragem, em complemento aos tratamentos já processados, o uso da transformada de wavelet resultou em resultados mais satisfatórios, nas métricas Precisão e F1 analisadas. A perda demonstrada pela Sensibilidade já era esperada, visto que os falsos positivos foram reduzidos, e o algoritmo agora não define qualquer trecho como imperfeição.

Levando em consideração o tamanho dos sinais analisados, a wavelet realizou a decomposição do sinal em seis níveis, e sua reconstrução foi testada em cada um destes, visando identificar a melhor análise comportamental do algoritmo. Com o sinal totalmente reconstruído, obteve-se resultados similares aos encontrados nos testes sem o uso da wavelet. Com o sinal sendo reconstruído somente com um ou dois níveis, os resultados foram deficitários. Estes níveis de decomposição corroeram as imperfeições visadas. Assim, o SVM não correspondeu conforme esperado.

A reconstrução do sinal em três, quatro e cinco níveis obtiveram resultados parecidos e com pouca variabilidade entre si, constando valores bem próximos aos já encontrados com o sinal completo. A reconstrução usando cinco dos seis níveis disponíveis resultou em valores idênticos aos encontrados na reconstrução completa, inclusive. Na Tabela 5, é possível visualizar o desempenho de cada um.

**Tabela 5 – Sensibilidade, Precisão e F1 - Sinais Decompostos.**

Reconstrução	Sensibilidade	Precisão	F1
1 nível	13,70%	100%	18,60%
2 níveis	30,74%	100%	39,43%
3 níveis	61,67%	100%	73,89%
4 níveis	63,51%	97,84%	74,33%
5 níveis	63,52%	97,22%	74,13%
Completa	63,52%	97,22%	74,13%

Fonte: Autoria própria.

Também foram proferidos testes de decomposição do sinal, via wavelet, em porções diferentes de seis demonstrado acima: três, nove e doze decomposições. Foi observado que, para o tamanho espectral dos arquivos analisados, a decomposição acima de cinco níveis já não representa parte restante de sinal para composição dos níveis seguintes. Com isso, foi determinado o uso de seis níveis de decomposição como padrão, e os demais foram desconsiderados para este estudo.

Com intuito de evitar estas deficiências e estabelecer melhorias de sincronia entre os sinais de treino e os sinais classificados, alguns testes foram proferidos e direcionados aos sinais treinadores. Os resultados são descritos abaixo: na análise, os sinais de treino foram submetidos ao mesmo filtro dos quais os sinais reais foram inicialmente tratados. Nestes casos, foi observado que houveram baixas alterações de potência classificadora do algoritmo quando treinado pelos sinais previamente filtrados. É demonstrado que, mesmo que ambos sinais estejam recebendo tratamentos semelhantes, os efeitos observados estão dentro da margem de erro da amostra (abaixo de 5,0% de melhoria), mostrando que estes sinais não possuem grandes frequências perturbadoras a serem retiradas pelo filtro, não sofrendo grandes alterações no circuito de tratamento.

Para validar o ponto acima, outro teste realizado foi percorrer os arquivos treinadores pelo filtro, e após, o SVM realizou a classificação dos sinais originais (sem o uso do filtro) estabelecido para esses. Os resultados encontrados foram insatisfatórios, já que os sinais originais a serem classificados ainda possuíam os ruídos quando foram analisados.

Estes procedimentos mostram a importância e os impactos que um acertado tratamento de dados pode causar nos resultados do classificador. Como se pode observar, somente a utilização do filtro não trouxe todas as melhorias esperadas e descritas em referências citadas anteriormente. O uso e manipulação conjunta do sinal a ser analisado nos domínios da frequência e do tempo, por meio da análise de Fourier

e Wavelet, possibilitou encontrar o complemento esperado nas respostas do algoritmo classificador.

Toda codificação do algoritmo utilizada se encontra disponibilizada na plataforma aberta Github (GITHUB, 2021). O conversor de arquivos csv para json desenvolvido também está disponível. Os arquivos que compõem a base de dados utilizados durante o desenvolvimento e testes de validação acham-se em condições pré e pós tratamento, assim como o filtro passa baixa descrito anteriormente.

## 5 CONCLUSÕES

A utilização de aplicações para os aparelhos eletrônicos e, em especial para os celulares, que realizam captação e tratamento de dados cresce a cada dia, juntamente com a popularização das tecnologias. É notável também a necessidade de uma análise e identificação profunda das principais vias de locomoção dos grandes centros urbanos, no Brasil e no mundo, que apresente detalhes, deficiências e pontos de melhorias, visando facilitar o gerenciamento do trânsito como um todo.

Os estudos desenvolvidos neste trabalho possibilitam afirmar que a aplicabilidade da metodologia SVC na identificação de anomalias e imperfeições é viável nas condições analisadas até aqui, levando em consideração as adaptações para os determinados tipos de pavimentação, sinalização e veículos automotores existentes na região. Importante destacar que o método proposto não é universal, e sim um aperfeiçoamento que possui como visão melhorar a detecção na região de estudo, não sendo suficientemente robusto para diferentes tipos de estradas. Porém, a categorização ampla das estradas de acordo com a qualidade e o treinamento proposto, o modelo correspondente pode atender à outras aplicações da vida real.

A adaptabilidade deste trabalho não se limita apenas aos tipos de estradas. Outros fatores, como os aparelhos celulares e os tipos de veículos também merecem um estudo mais aprofundado. É importante salientar que os diversos métodos estudados e aplicados em trabalhos semelhantes contribuem para ampliação do campo de visão neste tema, que tem passado por uma franca expansão de novas oportunidades de implementação.

Foi identificado que a utilização do filtro passa-baixa, quando corretamente parametrizado para os sinais em destaque, contribuiu positivamente no grau de assertividade da classificação realizada. Essa melhoria demonstra a importância dos tratamentos de dados para se obter um desempenho satisfatório do algoritmo classificador, que pode desestabilizá-lo caso não esteja alinhado com os principais objetivos e metas de estudo.

A combinação de análise e decomposição dos sinais amostrados no domínio da frequência e do tempo, por meio da Transformada de Wavelet, possibilitou aprofundar nos sinais, com importância para visualizar e neutralizar os ruídos existentes. Apesar de não trazer grandes incrementos de desempenho nas respostas do algoritmo, viabilizou o trabalho focasse nas frequências desejadas.

Os métodos analisados confirmam a preferência do uso de acelerômetros portáteis e disponíveis nos aparelhos celulares como mais práticos, dos pontos de vista econômico (energético e financeiramente) e da facilidade de adaptabilidade às muitas ramificações de tratamento e classificação existentes.

## 6 TRABALHOS FUTUROS

Para trabalhos futuros, é observável a necessidade de uma análise mais profunda para aplicabilidade do classificador em demais regiões do país, que devido a sua extensão continental possui uma malha viária bastante diversa, e um leque de disponibilidade tecnológica desigual, assim como suas respectivas anomalias e imperfeições. Uma maior abrangência de recursos disponíveis, como tráfego de dados em velocidades superiores e constantes, possibilitando maior captação de dados, contribuirá para um mapeamento ampliado da região, com atualizações constantes e geração das informações, como anomalias e intervenções de determinada via. Essas respostas podem ser atribuídas ao usuário contendo informações em tempo real da via na qual o mesmo percorre.

Além disso, é viável um estudo mais aprofundado que interfira na melhora do desempenho das variáveis estudadas deste trabalho, para que sejam reduzidos os falsos negativos e melhore ainda mais a precisão de acerto das classificações em questão. Propostas de diferentes metodologias de aprendizado de máquina são visadas. É esperado que, com os conhecimentos adquiridos e descritos, possa haver uma contribuição para propostas de uso de sensores e detectores mais robustos, práticos e viáveis economicamente.

Outra proposta de continuidade inclusa é o desenvolvimento de um aplicativo *mobile* que integra todas as funcionalidades, captando os dados dos sensores, realizando o tratamento adequado e respondendo para o usuário com as informações encontradas. Dessa forma, existe a possibilidade de relatar o atual estado da via na qual o usuário se encontra em tempo real.

É factível estudos de implementação para o caso para que os próximos veículos desenvolvidos já possam realizar a captação de dados via acelerômetros próprios e enviar diretamente para tratamento e posterior análise do SVC, sem que haja a necessidade da interface com aparelhos celulares. Essa área de estudo certamente será beneficiada com o avanço de novas tecnologias, como a implementação de redes de internet 5G. Assim, o retorno da situação viária dos trechos percorridos poderá ser verificado e processado diretamente pelo veículo que trafega na região, possibilitando atualizar sua rota de maneira autônoma, desviando de interferências e anomalias presentes, como consequência das respostas e conclusões encontradas.

## REFERÊNCIAS

AL, P. et. Scikit-learn: Machine Learning in Python. 2011.

ALLOUCH, A.; KOUBAA, A.; ABBES, T.; AMMAR, A. RoadSense: Smartphone Application to Estimate Road Conditions Using Accelerometer and Gyroscope. **IEEE Sensors Journal**, v. 17, n. 13, jul 2017. ISSN 1530-437X.

Analytics Vidhya Content Team. **A Complete Tutorial on Tree Based Modeling from Scratch (in R Python)**. 2016. Disponível em: <[https://www.vooo.pro/insights/um-tutorial-completo-sobre-a-modelagem-baseada-em-tree-arvore-do-zero-em-r-python/#:\\\$sim\\$:text=Introdu{ç}{~{a}}o,estabilidadedeefacilidadeidei](https://www.vooo.pro/insights/um-tutorial-completo-sobre-a-modelagem-baseada-em-tree-arvore-do-zero-em-r-python/#:\$sim$:text=Introdu{ç}{~{a}}o,estabilidadedeefacilidadeidei)>.

ARIZA-LÓPEZ, F. J.; MOZAS-CALVACHE, A. T.; UREÑA-CÁMARA, M. A.; Gil de la Vega, P. Dataset of three-dimensional traces of roads. **Scientific Data**, v. 6, n. 1, dec 2019. ISSN 2052-4463.

BELLO-SALAU, I. H.; ONUMANYI, A. J.; AIBINU, A. M.; ONWUKA, E. N.; DUKIYA, J. J.; OHIZE, H. A Survey of Accelerometer Based Techniques for Road Anomalies Detection and Characterization. **International Journal of Engineering Science and Application**, v. 3, n. 1, 2019. Disponível em: <<https://www.semanticscholar.org/paper/A-Survey-of-AccelerometerBased-Techniques-for-Road-Bello-Salau/55d108b8a999d41335cff750d4d4ad0c80d57eba>>.

CÁRDENAS-MONTES, M. Sobreajuste - Overfitting. 2006. Disponível em: <[http://wwwae.ciemat.es/\\\$sim\\$cardenas/docs/lessons/sobreajuste.pdf](http://wwwae.ciemat.es/\$sim$cardenas/docs/lessons/sobreajuste.pdf)>.

CARLOS, M. R.; ARAGON, M. E.; GONZALEZ, L. C.; ESCALANTE, H. J.; MARTINEZ, F. Evaluation of detection approaches for road anomalies based on accelerometer readings-Addressing who's who. **IEEE Transactions on Intelligent Transportation Systems**, v. 19, n. 10, p. 3334–3343, 2018. ISSN 15249050.

CHENG, J.; CHENG, J.; ZHOU, M.; LIU, F.; GAO, S.; LIU, C. Routing in internet of vehicles: A review. **IEEE Transactions on Intelligent Transportation Systems**, v. 16, n. 5, p. 2339–2352, 2015. ISSN 15249050.

CLARK, P.; BOSWELL, R. Rule induction with CN2: Some recent improvements. In: **Machine Learning — EWSL-91**. Berlin/Heidelberg: Springer-Verlag, 1991.

CONSELHO NACIONAL DE TRÂNSITO - CONTRAN. **RESOLUÇÃO Nº 600, DE 24 DE MAIO DE 2016**. 2016. Disponível em: <[https://www.in.gov.br/materia/-/asset\\_publisher/Kujrw0TZC2Mb/content/id/22921408/do1-2016-05-27-resolucao-n-600-de-24-de-maio-de-2016-22921310](https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/22921408/do1-2016-05-27-resolucao-n-600-de-24-de-maio-de-2016-22921310)>.

CZAKO ZOLTAN. **SVM and Kernel SVM**. 2018. Disponível em: <<https://towardsdatascience.com/svm-and-kernel-svm-fed02bef1200#tuning>>.

DATA-FLAIR. **Kernel Functions-Introduction to SVM Kernel Examples**. 2021. Disponível em: <<https://data-flair.training/blogs/svm-kernel-functions/>>.

DEVMEDIA. **JSON Tutorial**. 2021. Disponível em: <<https://www.devmedia.com.br/json-tutorial/25275>>.

DOUANGPHACHANH, V.; ONEYAMA, H. A study on the use of smartphones under realistic settings to estimate road roughness condition. **Eurasip Journal on Wireless Communications and Networking**, v. 2014, n. 1, p. 1–11, 2014. ISSN 16871499.

GALAR, M.; FERNÁNDEZ, A.; BARRENECHEA, E.; BUSTINCE, H.; HERRERA, F. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. **Pattern Recognition**, v. 44, n. 8, p. 1761–1776, 2011. ISSN 00313203.

GHOSE, A.; BISWAS, P.; BHAUMIK, C.; SHARMA, M.; PAL, A.; JHA, A. Road condition monitoring and alert application: Using in-vehicle smartphone as internet-connected sensor. **2012 IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2012**, n. March, p. 489–491, 2012.

GITHUB. **Otimizacao-de-classificador-de-imperfeicoes-viarias-em-trechos-da-cidade-de-Betim**. 2021. Disponível em: <<https://github.com/rodolfoandlima/TCC-Rodolfo-Lima.git>>.

HAYKIN, S. **Redes Neurais Princípios e Prática**. 2.ed.. ed. Porto Alegre: [s.n.], 2001. ISBN 978-85-7307-718-6.

Honglak Lee; Roger Grosse; Rajesh Ranganath; Andrew Y. Ng. Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations. **Proceedings of the 26th Annual International Conference on Machine Learning (em inglês)**, 2009.

HORTA, G. F. **Acelerômetros e sua aplicação no transporte de componentes para uma usina eólica**. [S.l.], 2020.

JIA, N. Detecting human falls with a 3-axis digital accelerometer. **A forum for the exchange of circuits, systems, and software for real-world signal processing**, Volume 43, 2009.

KATTI FACELI; Ana Carolina Lorena; João Gama; Tiago Agostinho de Almeida; André C. P. L. F de Carvalho. **Inteligência Artificial - Uma abordagem de aprendizado de máquina**. 2|2021. ed. [S.l.: s.n.], 2021. v. 2. ISBN 9788521637349.

KNERR; STEFAN; Léon Personnaz; Gérard Dreyfus. **Single-layer learning revisited: a stepwise procedure for building and training a neural network**. [S.l.]: Neurocomputing, 1990. 41–50 p. ISBN 978-3-642-76153-9.

KONGYANG CHEN; Mingming Lu; Xiaopeng Fan; Mingming Wei; Jinwu Wu. Road condition monitoring using on-board Three-axis Accelerometer and GPS Sensor. In: **2011 6th International ICST Conference on Communications and Networking in China (CHINACOM)**. [S.l.]: IEEE, 2011. ISBN 978-1-4577-0101-6.

Mario Gerla, Eun-Kyu Lee, Giovanni Pau; LEE, U. Internet of Vehicles : From Intelligent Grid to Autonomous Cars and Vehicular Clouds The Genesis of IOT. p. 241–246, 2015.

MEDNIS, A.; STRAZDINS, G.; ZVIEDRIS, R. Real Time Pothole Detection using Android Smartphones with Accelerometers Research domain • Road infrastructure as blood vessels. 2011.

MILLER, J. S.; BELLINGER, W. Y. FHWA, Distress Identification manual for the Long-Term Pavement Performance Program. Report FHWA-HRT-13-092. n. May, p. 142, 2014.

National Institute for Phys Planning Constr Res, C. o. R. D. C. **CATALOGUE OF PAVEMENT DEFECTS (FLEXIBLE ROAD PAVEMENTS)**. 1985. Disponível em: <<https://trid.trb.org/view.aspx?id=279060>>.

NUNES, D. E.; FAGUNDES, A.; MOTA, V. F. S. **Classificação da Qualidade de Vias Urbanas baseado em Sensoriamento Participativo**. Porto Alegre, RS, Brasil, 2018. Disponível em: <[www.waze.comhttps://sol.sbc.org.br/index.php/courb/article/view/2334](http://www.waze.comhttps://sol.sbc.org.br/index.php/courb/article/view/2334)>.

PEDRO C T GOMES. **Introdução ao aprendizado de máquina**. 2019. Disponível em: <<https://www.datageeks.com.br/aprendizado-de-maquina/>>.

POPESCU, D.; STRUȚU, M. I. **Accelerometer based road defects identification system Cloud Architecture for an open Library of Complex re-Usable Logical function blocks for Optimized Systems View project Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement ACCELEROMETER BASED ROAD DEFECTS IDENTIFICATION SYSTEM**. [S.l.], 2014. v. 76, n. 3. Disponível em: <<https://www.researchgate.net/publication/289226869>>.

PYSCIENCE-BRASIL. **PyScience-Brasil**. 2021. Disponível em: <<http://pyscience-brasil.wikidot.com/module:numpy>>.

RICH, E.; KNIGHT, K. **Inteligência Artificial**. 2ª edição. ed. Rio de Janeiro: [s.n.], 1994. ISBN 85-346-0122-4.

RUSSELL, S. J.; NORVIG, P.; CANNY, J. F.; MALIK, J. M.; EDWARDS, D. D.; JONATHAN, S. J. S.; NORVIG, P. **A Modern Approach**. [S.l.: s.n.], 2010. ISBN 0131038052.

SEQUEIRA, C. Sensores para medições de Vibrações Mecânicas–Acelerómetros. **Revista Manutenção**, p. 4–6, 2013.

TASPINAR, A. A guide for using the wavelet transform in machine learning. 2018. Disponível em: <<https://ataspinar.com/2018/12/21/a-guide-for-using-the-wavelet-transform-in-machine-learning/>>.

TEDESCHI, A.; BENEDETTO, F. A real-time automatic pavement crack and pothole recognition system for mobile Android-based devices. **Advanced Engineering Informatics**, Elsevier Ltd, v. 32, p. 11–25, 2017. ISSN 14740346. Disponível em: <<http://dx.doi.org/10.1016/j.aei.2016.12.004>>.

VIEYRA, C.; VIEYRA, R. **Physics Toolbox Apps**. 2015. Disponível em: <[www.vieyrasoftware.net](http://www.vieyrasoftware.net)>.

VV VASCONCELOS; PP MARTINS-JÚNIOR. Protótipo de Sistema Especialista em Direito Ambiental para Auxílio à decisão em Situações de Desmatamento Rural. **NT-27. CETECMG**, 2004.

WAHLSTRÖM, J.; SKOG, I.; HÄNDEL, P.; MEMBER, S. Smartphone-Based Vehicle Telematics : A Ten-Year Anniversary. p. 1–24, 2017.

WEI, L.; FWA, T. F.; ZHE, Z. Wavelet Analysis and Interpretation of Road Roughness. **Journal of Transportation Engineering**, v. 131, n. 2, feb 2005. ISSN 0733-947X.

WU, C.; WANG, Z.; HU, S.; LEPINE, J.; NA, X.; AINALIS, D.; STETTLER, M. An Automated Machine-Learning Approach for Road Pothole Detection Using Smartphone Sensor Data. **Sensors**, v. 20, n. 19, sep 2020. ISSN 1424-8220.