

CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA “PAULA SOUZA”

FACULDADE DE TECNOLOGIA DE BEBEDOURO

TECNOLOGIA EM BIG DATA NO AGRONEGÓCIO

**ANÁLISE DE COEFICIENTE DE CORRELAÇÃO
ENTRE COMPONENTES AUTOMOTIVOS PARA
PREVISÃO DE MANUTENÇÃO**

AUTOR: JOEL CERQUINI JUNIOR

ORIENTADOR: PROF. DR. RENAN GUILERME NESPOLO

BEBEDOURO

2025

JOEL CERQUINI JUNIOR

**ANÁLISE DE COEFICIENTE DE CORRELAÇÃO
ENTRE COMPONENTES AUTOMOTIVOS PARA
PREVISÃO DE MANUTENÇÃO**

Monografia apresentada à Faculdade de Tecnologia
de Bebedouro, como parte dos requisitos para a
obtenção do título de Tecnólogo em Big Data no
Agronegócio

Orientador: **prof. Dr. Renan Guilherme Nespolo**

BEBEDOURO

2025

“Não espere as oportunidades acontecerem, crie-as”

George Bernard Shaw

Agradecimentos

Agradeço primeiramente com todo o meu carinho e gratidão a:

Deus pela minha saúde e por toda minha vida, também meus pais, Joel e Silvia, que sempre me apoiaram em todas minhas escolhas, e por último a minha esposa Williana e meu filho Joel Neto, que foram o combustível nessa jornada.

Meu orientador, Dr. Renan, por sua paciência e apoio a todo o projeto, com suas explicações e calma para tentar a me orientar tudo o que o precisava fazer para atingir o objetivo final.

CERQUINI, J. ANÁLISE DE COEFICIENTE DE CORRELAÇÃO ENTRE COMPONENTES AUTOMOTIVOS PARA PREVISÃO DE MANUTENÇÃO. Trabalho de Graduação (Monografia). Centro Estadual de Educação Tecnológica “Paula Souza”. Faculdade de Tecnologia de Bebedouro. n° p. 2025.

RESUMO

Nos dias de hoje é fundamental de implementar um sistema de manutenção preventiva em veículos pesados, especialmente considerando a complexidade e o custo elevado de reparos nesses equipamentos. Por meio de uma análise de dados operacionais e do diagnóstico de falhas, foi possível identificar padrões que indicam a necessidade de intervenção antes que problemas graves ocorram. Durante uma experiência prática realizada em uma oficina especializada, observou-se que muitas falhas frequentes em componentes críticos, como motores e câmbios, poderiam ser significativamente minimizadas com a adoção de técnicas de manutenção preditiva. Desse modo o presente estudo propõe a utilização de sensores de vibração e outros sensores de monitoramento veicular, em tempo real, para avaliar quais são os principais sensores que podem prever falhas antes que elas aconteçam. O estudo foi estruturado em etapas que incluíram levantamento bibliográfico, aquisição de dados e análise da correlação entre os sensores, comparando duas técnicas: O coeficiente de correlação de *Pearson* e o *Rank* de *Spearman*. Além das técnicas também foi abordado a estabilidade da correlação entre os sensores avaliando o desvio padrão dos resultados. Essa abordagem reforça a importância de uma gestão eficiente e preventiva, promovendo melhorias contínuas na operação e manutenção de frota.

Palavras-chave: Coeficiente de correlação de *Pearson*. *Rank* de *Spearman*. Análise de dados. Manutenção preditiva. Mapa de calor.

CERQUINI, J. ANÁLISE DE COEFICIENTE DE CORRELAÇÃO ENTRE COMPONENTES AUTOMOTIVOS PARA PREVISÃO DE MANUTENÇÃO. Trabalho de Graduação (Monografia). Centro Estadual de Educação Tecnológica “Paula Souza”. Faculdade de Tecnologia de Bebedouro. nº p. 2025.

ABSTRACT

Nowadays, it is essential to implement a preventive maintenance system in heavy vehicles, especially considering the complexity and high cost of repairs to this equipment. Through an analysis of operational data and fault diagnosis, it was possible to identify patterns that indicate the need for intervention before serious problems occur. During a practical experience carried out in a specialized workshop, it was observed that many frequent failures in critical components, such as engines and gearboxes, could be significantly minimized with the adoption of predictive maintenance techniques. Thus, this study proposes the use of vibration sensors and other vehicle monitoring sensors, in real time, to evaluate which are the main sensors that can predict failures before they occur. The study was structured in stages that included a bibliographic survey, data acquisition and analysis of the correlation between the sensors, comparing two techniques: Pearson's correlation coefficient and Spearman's Rank. In addition to the techniques, the stability of the correlation between the sensors was also addressed by evaluating the standard deviation of the results. This approach reinforces the importance of efficient and preventive management, promoting continuous improvements in fleet operation and maintenance.

Keywords: *Pearson correlation coefficient. Spearman rank. Data analysis. Predictive maintenance. Heat map.*

SUMÁRIO

1	INTRODUÇÃO	13
2	FUNDAMENTAÇÃO TEÓRICA.....	15
2.1	MANUTENÇÃO AUTOMOTIVA	15
2.2	MANUTENÇÃO PREDITIVA	15
2.3	SISTEMAS DE MONITORAMENTO POR SENSORES	15
2.4	ANÁLISE ESTATÍSTICA E INTELIGÊNCIA ARTIFICIAL	16
2.5	RECONHECIMENTO AUTOMÁTICO DE PADRÕES	16
2.6	IMPACTOS DA MANUTENÇÃO PREDITIVA	16
2.7	COEFICIENTE DE CORRELAÇÃO	17
2.7.1	COEFICIENTE DE CORRELAÇÃO DE <i>PEARSON</i>	18
2.7.2	COEFICIENTE DE CORRELAÇÃO POR POSTOS DE <i>SPEARMAN</i>	19
2.8	MATRIZ DE CONFUSÃO E MAPA DE CALOR.....	20
3	MÁTERIAS E MÉTODOS	23
3.1	CONJUNTO DE DADOS	23
3.2	DESENVOLVIMENTO DO PROJETO	24
3.3	MÉTRICAS DE AVALIAÇÃO.....	29
4	RESULTADOS E DISCUSSÕES.....	30
4.1	COEFICIENTE DE CORRELAÇÃO DE <i>PEARSON</i>	31
4.2	<i>RANK</i> DE <i>SPEARMAN</i>	33
4.3	DESVIO PADRÃO <i>PEARSON</i>	35
4.4	DESVIO PADRÃO <i>SPEARMAN</i>	37
4.5	DISCUSSÃO DOS RESULTADOS	38
4.6	INTERPRETAÇÃO PARA A MANUTENÇÃO	39
5.	CONCLUSÕES	41
	REFERÊNCIAS	43

LISTA DE FIGURAS

FIGURA 1 - EXEMPLO DE UM CICLO DE TESTES DO <i>DATASET</i> PS1 QUE CONTROLA . O SISTEMA DE REFRIGERAÇÃO DO ACUMULADOR HIDRÁULICO.....	26
FIGURA 2 - FLUXOGRAMA DO PRESENTE ESTUDO.....	28

Lista de Tabelas

TABELA 1 - INTERPRETAÇÃO DO COEFICIENTE DE CORRELAÇÃO DE <i>PEARSON</i>. .	17
TABELA 2 - PRINCIPAIS CARACTERÍSTICAS DOS <i>DATASETS</i> UTILIZADOS NO PRESENTE ESTUDO.	23
TABELA 3 - NORMALIZAÇÃO DOS DADOS.	25

Lista de Gráficos

GRÁFICO 1 - MAPA DE CALOR DA MATRIZ DE CONFUSÃO PARA DETECÇÃO DE FALHAS EM MOTORES.....	21
GRÁFICO 2 - MAPA DE CALOR DO COEFICIENTE DE CORRELAÇÃO DE <i>PEARSON</i> ENTRE OS SENSORES DE UM VEÍCULO.....	31
GRÁFICO 3 - MAPA DE CALOR DO RANK DE <i>SPEARMAN</i> ENTRE OS SENSORES DE UM..... VEÍCULO.	33
GRÁFICO 4 - MAPA DE CALOR DO DESVIO PADRÃO DE <i>PEARSON</i> ENTRE OS SENSORES DE ... UM VEÍCULO.	35
GRÁFICO 5 - MAPA DE CALOR DO DESVIO PADRÃO DE <i>SPEARMAN</i> ENTRE OS SENSORES DE UM VEÍCULO.	37

LISTA DE EQUAÇÕES

(1) COEFICIENTE DE CORRELAÇÃO DE PEARSON	18
(2) COEFICIENTE DE CORRELAÇÃO POR POSTOS DE SPEARMAN.....	19
(3) NORMALIZAÇÃO DE HERTZ.....	25
(4) NÚMERO DE PONTOS POR LINHA NO DATASET.....	25

Lista de Siglas

CE – *Cooling Efficiency*(Virtual)

CP – *Cooling Power*(Virtual)

EPS – Motor Power

FN - Falsos Negativos

FP - Falsos Positivos

FS - Frequência de Operação

Hz - Hertz

IA - Inteligência Artificial

kW - Quilowatt

PS - *Pressure*

SE – *Efficiency Factor*

TS - *Temperature*

VN - Verdadeiros Negativos

VP - Verdadeiros Positivos

VS - *Vibration*

W - Watt

1 INTRODUÇÃO

O setor de veículos pesados mudou drasticamente nas últimas décadas. Em meados do século XX, mecânicos experientes realizavam diagnósticos, identificando problemas complexos de forma empírica, observando ou ouvindo o funcionamento do maquinário, desenvolvendo conhecimento por meio da prática (HELWIG et al. 2015; MARQUES DE SÁ et al., 2001; KANO & NAKAGAWA et al., 2008). Atualmente, os caminhões embarcaram sistemas eletrônicos complexos, contando com sensores que registram fenômenos variados, tais como: temperatura, vibração, pressão e emissões de fluídos (HELWIG et al., 2015; WILCOX et al., 2010). Essas mudanças afetam especialmente regiões com forte atuação no agronegócio, nas quais falhas no transporte podem ocasionar prejuízos consideráveis durante períodos críticos como a colheita (HELWIG et al., 2015; SCHNEIDER et al., 2017; WOLD & ESBENSEN & GELADI, et al, 1987).

Com muitos dados sendo disponibilizados por esses sensores embarcados nos veículos, uma oportunidade interessante pode ser observada: usar os dados disponibilizados para prever falhas de funcionamento, gerando assim a manutenção preventiva (SCHNEIDER et al., 2017; OLSZEWSKI et al., 2001). Porém, a utilização de vários sensores pode não ser a estratégia mais eficiente, pois nem todos os sensores possuem a capacidade de indicar falhas de funcionamento em tempo real (HELWIG et al., 2015; HELWIG & SCHUTZE, 2015; MARQUEZ et al., 2012).

No campo estatístico existem técnicas que averiguam a associação entre duas variáveis, de forma numérica, denominada coeficiente de correlação (referencia). Em nosso contexto podemos verificar a correlação entre os principais sensores de um veículo: Pressão do sistema de circulação do ar (bar); temperatura do motor (C°); potência do motor (W); fluxo de volume (l/mim); vibração (mm/s); eficiência de resfriamento (%); energia de resfriamento (kW); fator de eficiência (%) . Existem mais de uma forma de avaliar o coeficiente de correlação, uma para dados que estão organizados de forma linear, utilizando a técnica de coeficiente de correlação de *Pearson* e outra técnica que quantifica o coeficiente de correlação para dados não organizados linearmente, chamada de *Rank de Spearman* (SCHNEIDER et al., 2017; BEYER & GOLDSTEIN & RAMAKRISHNAN & SHAFT et al.,1999).

Deste modo o presente trabalho abordou as duas técnicas junto, na atuação de vários ciclos dos sensores, com os seguintes objetivos:

1. Analisar o coeficiente de correlação entre os sensores de Pressão do sistema de circulação do ar (bar); temperatura do motor (C°); potência do motor (W); fluxo de volume (l/mim); vibração (mm/s); eficiência de resfriamento (%); energia de resfriamento (kW); fator de eficiência (%)
2. Avaliar a correlação entre os sensores comparando os métodos de *Pearson* e *Spearman*, para verificar se os dados estão organizados linearmente.
3. Gerar um mapa de calor em uma matriz de confusão para cada técnica escolhida, afim de gerar uma análise visual entre as técnicas.

O presente trabalho está segmentado em: 1 Introdução; 2 Fundamentação Teórica; 3 Materiais e Métodos; 4 Resultados e Discussões; 5 Conclusões; ao final as Referências.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 MANUTENÇÃO AUTOMOTIVA

A manutenção automotiva engloba diversas práticas essenciais ao funcionamento adequado dos veículos e à segurança dos usuários. Não se trata apenas de uma questão econômica - a manutenção apropriada é fundamental para que os sistemas de transporte sejam confiáveis e para que acidentes sejam prevenidos (HELWIG et al., 2015; HELWIG & SCHUTZE, 2015; LIU et al., 2012). Falhas mecânicas podem gerar consequências sérias, especialmente em sistemas críticos como freios, direção ou motor. A literatura da área mostra que a maior parte dessas falhas poderia ser evitada se houvesse programas de manutenção bem estruturados e executados adequadamente (HELWIG et al., 2015).

Em empresas, principalmente aquelas do setor de transporte e logística, a manutenção ganha importância estratégica. Quando veículos ficam parados por problemas mecânicos, isso impacta diretamente as operações e a receita, o que justifica desenvolver estratégias específicas de manutenção como parte da gestão operacional (HELWIG et al., 2015).

2.2 MANUTENÇÃO PREDITIVA

Representa um avanço tecnológico significativo na gestão da manutenção. Em vez de seguir cronogramas fixos ou esperar que falhas aconteçam, usa dados coletados continuamente para prever quando componentes realmente precisam de intervenção. Isso traz mais inteligência e eficiência para a gestão dos recursos de manutenção (HELWIG et al., 2015; SCHNEIDER et al., 2017; HIGGS & PARKIN & JACKSON & AL-HABAIBEH & ZORRIASSATINE & COY et al., 2004).

2.3 SISTEMAS DE MONITORAMENTO POR SENSORES

Veículos modernos têm sistemas eletrônicos avançados com vários sensores que monitoram constantemente parâmetros operacionais. Estes dispositivos capturam informações sobre temperatura, pressão, vibração, rotação e outros parâmetros críticos. Os dados ficam sendo registrados o tempo todo, possibilitando análises detalhadas para identificar padrões e

tendências (HELWIG et al., 2015; SCHNEIDER et al., 2017; BLEY & PIGNANELLI & SCHUTZE et al., 2012; LAOUTI & SHEIBAT-OTHMAN & OTHMAN et al., 2011).

2.4 ANÁLISE ESTATÍSTICA E INTELIGÊNCIA ARTIFICIAL

A grande quantidade de dados dos sensores permite usar técnicas estatísticas avançadas e algoritmos de inteligência artificial para identificar padrões que aparecem antes das falhas mecânicas (SCHNEIDER et al., 2017). Helwig, Pignanelli e Schütze (2015) mostraram que técnicas estatísticas multivariadas conseguem detectar problemas em equipamentos complexos bem antes deles se manifestarem fisicamente.

2.5 RECONHECIMENTO AUTOMÁTICO DE PADRÕES

No trabalho de Schneider, Helwig e Schütze (2017) foi evidenciado que algoritmos de aprendizado de máquina conseguem aprender sozinhos a identificar características importantes em dados cíclicos, como os que motores em funcionamento geram. Esta tecnologia faz com que sistemas computacionais desenvolvam capacidade de interpretar sinais de funcionamento anormal.

2.6 IMPACTOS DA MANUTENÇÃO PREDITIVA

Sistemas de manutenção preditiva têm mostrado resultados expressivos na redução de custos operacionais. Estudos de caso em frotas comerciais reportam reduções de até 30% nos custos de manutenção. Além da economia direta, há também maior disponibilidade dos veículos, já que intervenções podem ser programadas e otimizadas (HELWIG et al., 2015; SCHNEIDER et al., 2017; GUENEL & MESHRAM & BLEY & SCHÜTZE & KLUSCH et al., 2013).

Veículos mantidos adequadamente consomem menos combustível e emitem menos poluentes. Evitar descartes prematuros de peças por causa de falhas também reduz o desperdício de materiais. Assim, manutenção adequada contribui para sustentabilidade ambiental (HELWIG et al., 2015; International Standard ISO 13372 et al., 2004).

A digitalização dos veículos abriu novas possibilidades para prestadores de serviços de

manutenção. Hoje é possível saber exatamente quais componentes precisam ser substituídos antes mesmo do veículo chegar na oficina. Isso melhora o planejamento, otimiza estoques e aumenta eficiência operacional (HELWIG et al., 2015; SCHNEIDER et al., 2017).

Existem plataformas que monitoram frotas inteiras continuamente, mandando alertas em tempo real quando detectam problemas. Isso representa uma mudança fundamental na gestão de frotas, dando mais controle operacional e diminuindo riscos (HELWIG et al., 2015; HELWIG & SCHUTZE, 2015).

2.7 COEFICIENTE DE CORRELAÇÃO

Ao trabalhar com equipamentos industriais ou automotivos, muitas vezes precisamos entender as relações entre diferentes variáveis. Por exemplo, queremos saber se há conexão entre aumento da vibração de um motor e elevação do ruído, ou entre velocidade de rotação e aquecimento de uma máquina. Na engenharia, são usados dois métodos para medir correlação (HELWIG et al., 2015; SCHNEIDER et al., 2017; TCHAKOUA et al., 2014).

A Correlação de *Pearson* funciona bem para dados que seguem distribuição normal e têm relação linear, como por exemplo análises de desgaste de componentes de veículos pesados sob condições controladas. É mais precisa quando as condições estatísticas de aplicação são atendidas. Por outro lado, o coeficiente de *Spearman* é considerado flexível (comparado a técnica de *Pearson*), sendo útil quando os dados de manutenção apresentam comportamentos não lineares ou distribuições não normais, o que é comum em registros operacionais de frotas pesadas (MUKAKA, 2012), estabeleceu, na tabela 1, uma interpretação prática para esses valores:

Tabela 1 - Interpretação do Coeficiente de Correlação de *Pearson*.

Valores	Relação	Interpretação
0,7 a 0,9	Relação forte	Correlação significativa que vale investigar
0,5 a 0,7	Relação moderada	Possível influência, mas talvez indireta

Abaixo de 0,3	Relação fraca ou nula	Correlação provavelmente sem importância
---------------	-----------------------	--

Fonte: adaptado de ((MUKAKA, 2012).

É importante entender que correlação não significa causalidade. Duas variáveis podem ter correlação alta sem que uma cause a outra - ambas podem ser influenciadas por um terceiro fator. Em veículos em funcionamento, a correlação pode mudar ao longo do tempo por causa de condições operacionais, desgaste de componentes e fatores ambientais. Uma estratégia eficaz é analisar correlações em janelas temporais, o que permite detectar mudanças graduais que podem indicar problemas futuros (SCHNEIDER et al., 2017).

Em sistemas hidráulicos complexos, Helwig, Pignanelli e Schütze (2015) mostraram que analisar múltiplas variáveis simultaneamente permite detectar problemas que passariam despercebidos se analisássemos cada sensor individualmente. A correlação funciona como um filtro quando múltiplos sensores fornecem dados, permitindo identificar: Sinais redundantes que medem essencialmente o mesmo sensor e indícios que fornecem informações única.

Helwig e Schütze (2015) também trataram de uma questão prática importante: detectar falhas em sensores. Descobriram que a correlação entre sensores pode identificar quando um dispositivo está dando leituras inconsistentes, permitindo correção automática ou alertas para verificação manual.

Análise de correlação é base para técnicas mais avançadas como análise de componentes principais e regressão múltipla, servindo como ferramenta fundamental para entender comportamento de equipamentos e criar sistemas inteligentes de diagnóstico.

2.7.1 COEFICIENTE DE CORRELAÇÃO DE PEARSON

O coeficiente de correlação de *Pearson* é a medida mais conhecida e usada para avaliar intensidade e direção de relações lineares entre duas variáveis contínuas. Sua fórmula dada na equação 1:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1)$$

no qual x_i e y_i são os valores individuais das variáveis x e y ; $\bar{x}$ e $\bar{y}$ são as médias de x e y , respectivamente. O coeficiente de *Pearson* assume que os dados têm distribuição normal e que a relação entre as variáveis é linear. Em sistemas de monitoramento por sensores, o *Pearson* pode mostrar se aumentos em uma leitura acompanham aumentos em outra, sugerindo correlação direta.

O coeficiente de *Pearson* permite avaliar com precisão relações quantitativas em séries temporais, aspecto crucial quando sensores monitoram parâmetros ao longo do tempo. Ele dá uma medida objetiva para comparar conjuntos de dados mesmo quando não há relação visual clara.

Sua interpretação tem ampla aceitação na literatura científica (Helwig, Pignanelli e Schütze, 2015; Schneider et al., 2017). Valores próximos de +1 indica que as variáveis se movem juntas de maneira linear crescente, enquanto valores próximos de -1 mostram relação inversa. O coeficiente de correlação de *Pearson* virou ferramenta valiosa na identificação de anomalias, desvios e padrões operacionais esperados, sendo especialmente útil em programas de manutenção preditiva baseados em análise estatística.

2.7.2 COEFICIENTE DE CORRELAÇÃO POR POSTOS DE *SPEARMAN*

O coeficiente de *Spearman* é uma alternativa ao *Pearson*, especialmente útil quando os dados não seguem distribuição normal ou a relação entre eles não é linear. Ele avalia correlação monotônica entre duas variáveis, verificando se uma variável tende a aumentar quando a outra também aumenta, sem exigir proporcionalidade. Sua fórmula é apresentada na equação 2:

$$\rho = 1 - \frac{(6 \sum_{i=0}^n d_i^2)}{[n(n^2 - 1)]}, \quad (2)$$

no qual d_i é a diferença entre os *Ranks* das observações correspondentes e n é o número de pares de

dados. O *Rank* de *Spearman* converte dados em *Ranks* (posições relativas), ignorando valores absolutos. Por isso, resiste a valores extremos (outliers), sendo especialmente útil na análise de dados de sensores ruidosos ou onde a linearidade entre variáveis não é evidente (HELWIG, PIGNANELLI e SCHUTZE, 2015; SCHNEIDER et al., 2017).

O coeficiente de *Spearman* é particularmente adequado para dados que têm comportamento não-linear, situação comum em aplicações com sensores automotivos, onde oscilações, ruídos de leitura e interferências externas podem comprometer análise linear. Por converter valores em posições, ele prioriza a ordem dos dados sobre a magnitude, ajudando a identificar tendências e correlações mesmo quando os dados são assimétricos ou não seguem distribuição normal (HELWIG, PIGNANELLI e SCHUTZE, 2015; SCHNEIDER et al., 2017).

Usar o *Spearman* mostrou-se vantajoso na comparação entre diferentes tipos de sensores, permitindo detectar correlações sutis não identificadas apenas com o *Pearson*. Variações constantes em certos sensores podem ter padrões previsíveis que, embora não lineares, revelam comportamento correlacionado ao longo do tempo. A robustez estatística do *Spearman* fica ainda mais importante em contextos de manutenção preditiva, onde decisões devem ser tomadas baseadas em indícios precoces de falha ou degradação. Sua aplicação fortalece a confiabilidade do diagnóstico baseado em dados e complementa resultados obtidos por métodos lineares (HELWIG, PIGNANELLI e SCHUTZE, 2015; SCHNEIDER et al., 2017; EL-BETAR et al., 2006).

2.8 MATRIZ DE CONFUSÃO E MAPA DE CALOR

No desenvolvimento de sistemas de manutenção preditiva, uma questão fundamental é avaliar a eficácia real do modelo desenvolvido. A matriz de confusão é ferramenta essencial para esta avaliação. A matriz de confusão, conforme descrito por Fawcett (2006) e Powers (2011), é uma ferramenta de diagnóstico que permite avaliar desempenho de modelos de classificação. É essencialmente uma tabela que confronta previsões do modelo com resultados reais, dando visão detalhada sobre acertos e erros.

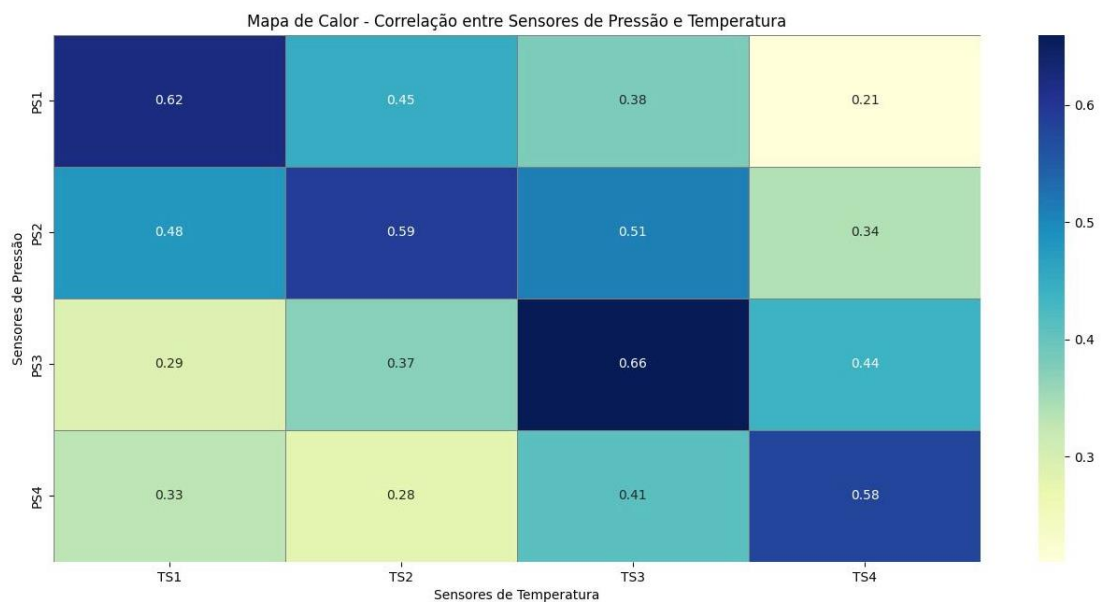
Sokolova e Lapalme (2009) descrevem esta ferramenta como uma matriz quadrada onde colunas representam previsões do modelo e linhas representam resultados reais. Apesar de simples, oferece informações valiosas sobre comportamento do sistema.

Usando como exemplo um sistema que detecta necessidade de manutenção em motores,

a matriz de confusão tem quatro componentes principais, conforme descrito por Provost (1998): Verdadeiros Positivos (VP), casos que o sistema identifica corretamente necessidade de manutenção, permitindo intervenção preventiva; Verdadeiros Negativos (VN), situações que o sistema identifica corretamente que não há necessidade de intervenção, evitando manutenção desnecessária; Falsos Positivos (FP), no qual o sistema indica necessidade de manutenção quando não há problema real, resultando em intervenção desnecessária; Falsos Negativos (FN), situações as quais o sistema falha em detectar problemas reais, podendo resultar em falhas inesperadas.

Para melhorar ainda mais a interpretação desses números, é comum usar o mapa de calor. Essa ferramenta visual aplica cores nas células da matriz, indicando a frequência de cada tipo de resultado. Quanto maior o número de ocorrências, mais forte a cor. A seguir, temos um exemplo de mapa de calor criado com dados simulados de um sistema de manutenção preditiva é apresentado no gráfico 1:

Gráfico 1 - Mapa de Calor da Matriz de Confusão para detecção de falhas em motores.



Fonte: elaborado pelo autor.

Ao olhar para o gráfico, percebe-se que o sistema acerta bem mais do que erra. As áreas mais escuras representam os maiores acertos. No entanto, um ponto que chama a atenção é a quantidade de Falsos Negativos (80 casos em mil análises). Isso pode ser um risco grave em operações onde a segurança vem em primeiro lugar. Chicco e Jurman (2020) reforçam que o uso de mapas de calor é uma boa prática para avaliar modelos de classificação, principalmente em projetos de manutenção preditiva, onde qualquer falha no diagnóstico pode gerar prejuízos ou acidentes. Estes resultados mostram que, embora o sistema tenha boa performance geral, ainda deixa passar 80 problemas reais a cada 1000 análises, o que pode ser crítico em aplicações onde segurança é prioritária (CHICCO, D.& JURMAN, G;2020).

3 MÁTERIAS E MÉTODOS

No trabalho comparou duas técnicas estatísticas diferentes: o *Rank* de *Spearman* e o coeficiente de correlação de *Pearson*. O objetivo foi analisar as relações entre diferentes tipos de dados coletados de sensores veiculares; usar ambas as técnicas juntas ajuda a validar os resultados; e permite identificar tanto relações diretas quanto padrões mais complexos que apenas um método poderia não detectar.

As seções seguintes apresentam cada técnica em detalhe: a Seção 3.1 aborda o conjunto de dados utilizados na pesquisa; a Seção 3.2 discute o desenvolvimento do projeto; a Seção 3.3 apresenta as métricas de avaliação.

3.1 CONJUNTO DE DADOS

Os dados utilizados nesta pesquisa vieram de oito sensores diferentes instalados em veículos comerciais durante seu uso normal. cada sensor recebeu uma sigla específica (ps, ts, vs, se, ce, cp, eps e fs) e monitora um aspecto particular do funcionamento do veículo.

Então a aquisição de dados aconteceu durante várias semanas, resultando em cerca de mil medições por sensor. Os veículos rodaram em diferentes situações: trânsito urbano, estradas, cargas variadas e horários distintos. Essa variedade foi importante para ter dados que realmente representem como os veículos funcionam no dia a dia. A tabela abaixo (tabela 2) mostra detalhes sobre cada arquivo de dados:

Tabela 2 - Principais características dos *Datasets* utilizados no presente estudo.

Sensor	Arquivo	O que mede	Unidade	Frequência original
PS	PS.csv	pressão do sistema	bar	100 hz
TS	TS.csv	temperatura	°c	1 hz
VS	VS.csv	velocidade/vibração	mm/s	1 hz
SE	SE.csv	eficiência do sistema	%	1 hz
CE	CE.csv	consumo energético	%	1 hz
CP	CP.csv	potência consumida	kw	1 hz
EPS	EPS.csv	potência elétrica	w	100 hz
FS	FS.csv	frequência de operação	1/min	10 hz

Fonte: Elaborado pelo autor.

3.2 DESENVOLVIMENTO DO PROJETO

Antes de começar as análises, foi preciso pré-processar os dados. Alguns arquivos estavam incompletos ou tinham valores estranhos, provavelmente por falhas momentâneas dos sensores. O maior problema encontrado no desenvolvimento do estudo foi constatado que os sensores coletavam dados em velocidades muito diferentes. os sensores PS e EPS são muito rápidos (100 medições por segundo), o FS têm frequência média (10 medições por segundo), o restante uma medição por segundo. Para comparar todos os sensores de forma única, foi necessário normalizar a frequência entre os ciclos de teste(linhas) presente nos *Datasets*. Normalizando todas as medições para o mesmo ritmo de uma medição por segundo. Isso significou selecionar cuidadosamente quais dados dos sensores mais rápidos usar, transformar os valores, sem perder informações importantes.

Foi necessário normalizar todos os sensores para trabalhar no mesmo ritmo: 1 medição por segundo. É como se fosse preciso que todos falassem no mesmo ritmo para conseguir entender a conversa completa. A técnica aplicada foi calcular a média de cada segundo dos sensores mais rápidos. Por exemplo, quando o sensor PS realizava 100 medições em 1 segundo essas 100 medições eram somadas e divididas por 100 - obtendo assim um valor representativo daquele segundo, como apresentado na tabela 3.

Tabela 3 - Normalização dos dados.

Tipo do Sensor	Frequência original	Dados por Ciclo	Dados por Ciclo normalizados
PS	100 Hz	6.000 valores	60 valores
TS	1 Hz	60 valores	60 valores
VS	1 Hz	60 valores	60 valores
SE	1 Hz	60 valores	60 valores
CE	1 Hz	60 valores	60 valores
CP	1 Hz	60 valores	60 valores
EPS	100 Hz	6.000 valores	60 valores
FS	10 Hz	600 valores	60 valores

Fonte: Elaborado pelo autor.

Como cada sensor mede sistemas diferentes (temperatura em graus, pressão em bar, potência em watts, etc.), todas as medidas tiveram que ser normalizadas para permitir comparações. A normalização foi realizada pela equação 3, apresentada como:

$$NormHz(CL_i) = \frac{CL_i}{Hz}, \quad (3)$$

no qual CL é o *Dataset*, CL_i é o ciclo (pontos dados em uma linha, ou o número de colunas por linha, correspondente a equação 4) e Hz são os valores apresentados na tabela 2.

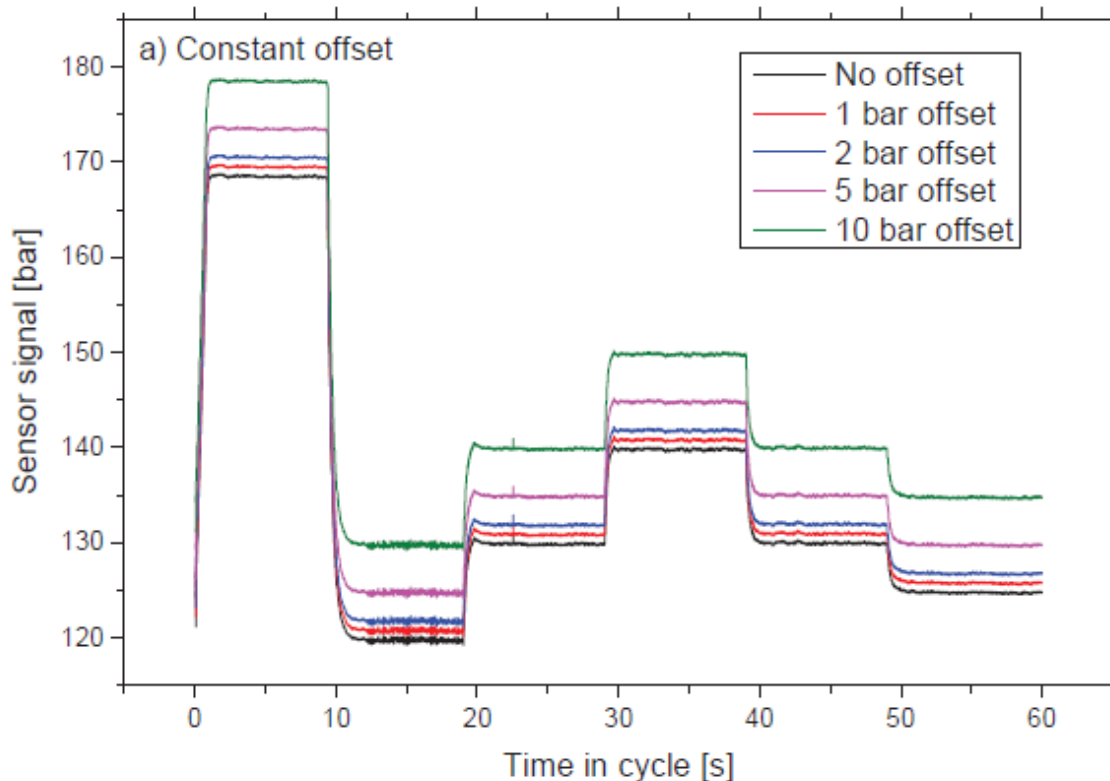
$$CL_i = 60 \times Hz. \quad (4)$$

Portanto necessário normalizar o tamanho das colunas para que foi necessário para normalizar as colunas para realizar uma comparação entre duas variáveis estatísticas (vetores).

O trabalho seguiu etapas bem definidas. Primeiro, foi necessário organizar os dados. Pois, caso os dados sejam mal preparados, a análise pode ser prejudicada (referência de viés bias). Foram criadas rotinas para encontrar e corrigir valores estranhos, além de sincronizar as medições dos diferentes sensores para a mesma frequência.

Um exemplo de um ciclo de testes do dataset PS1 é apresentada na figura 1.

Figura 1 - Exemplo de um ciclo de testes do dataset PS1 que controla o sistema de refrigeração do acumulador hidráulico.



Fonte: (HELWIG, N.; SCHÜTZE, A, 2015)

Uma abordagem complementar implementada durante o trabalho envolveu a análise não apenas dos valores diretos dos sensores, mas também de suas variações. A oscilação dos dados pode revelar padrões que permanecem ocultos quando observados apenas os valores absolutos.

Esta abordagem revelou aspectos importantes, já que alguns sensores sem correlação direta nos valores apresentaram padrões significativos na forma como variavam.

Para facilitar a interpretação dos dados numéricos, desenvolveram-se mapas de calor - representações gráficas coloridas que mostram visualmente as correlações mais fortes e mais fracas. Estes gráficos tornaram-se essenciais, considerando a dificuldade de identificar padrões

através da análise de tabelas numéricas.

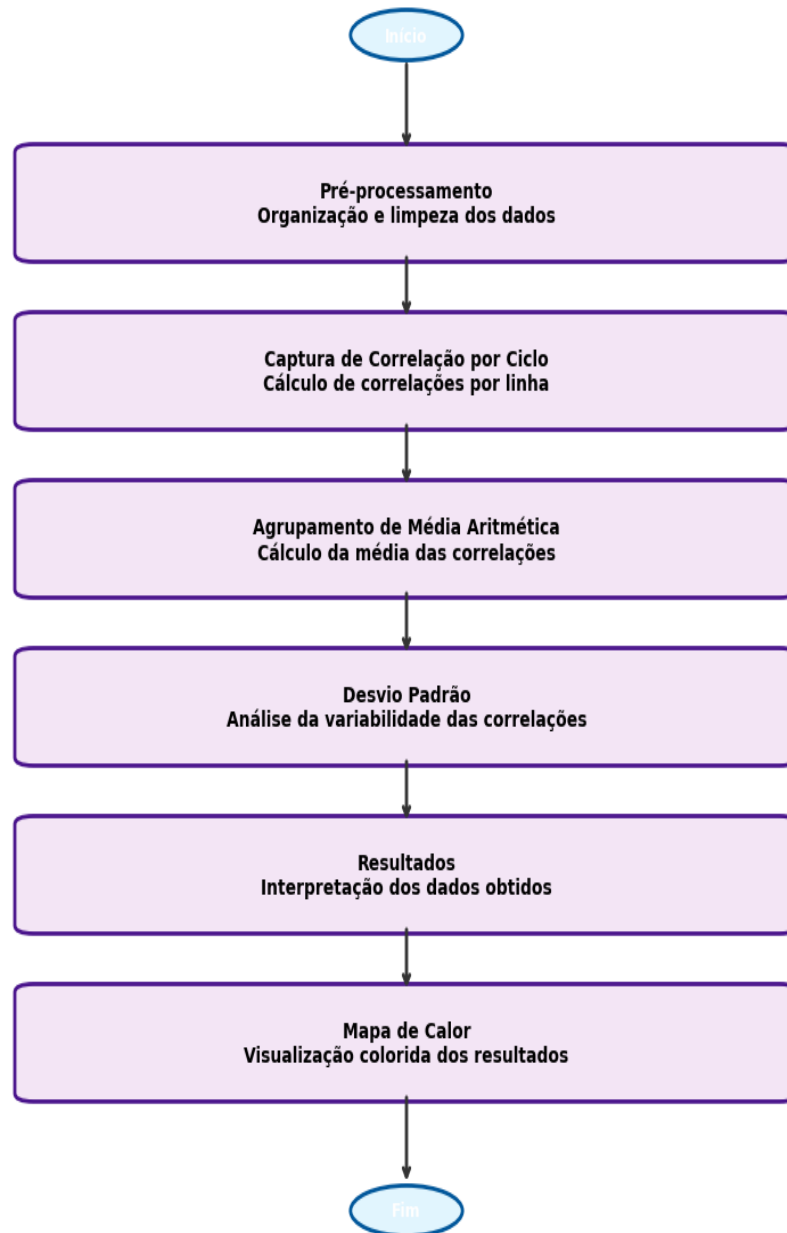
Os mapas de calor apresentaram tanto as correlações médias quanto suas variabilidades, oferecendo uma visão abrangente sobre quais relações entre sensores eram estáveis e quais apresentavam maior inconstância.

O processo permitiu identificar sensores com correlações fortes ou fracas entre si e compreender a consistência dessas relações. Observou-se que alguns sensores apresentavam correlações elevadas em média, mas com alta variação temporal, indicando relações instáveis. Outros sensores demonstraram correlações mais modestas, porém consistentes, sugerindo relações mais confiáveis e previsíveis.

A análise evidenciou que a estabilidade das correlações pode ser tão importante quanto sua magnitude para a compreensão do comportamento do sistema monitorado.

Abaixo um fluxograma é apresentado na figura 2 sobre os processos realizados por esse estudo.

Figura 2 - Diagrama de Atividade presente estudo.



Fonte: Elaborado pelo autor.

3.3 MÉTRICAS DE AVALIAÇÃO

Para interpretar os resultados, foram estabelecidos critérios baseados na literatura da área e nas orientações que recebi durante o desenvolvimento deste trabalho apresentados na seção 2.7.

As correlações como fortes quando o valor absoluto ficava acima de 0,7, moderadas entre 0,3 e 0,7, e fracas abaixo de 0,3. Esses intervalos são amplamente aceitos na comunidade acadêmica, embora eu tenha aprendido que o contexto também importa muito na interpretação (MUKAKA, 2012).

Para comparar os métodos de *Pearson* e *Spearman*, calculei as diferenças entre seus resultados. Quando essa diferença era pequena (menor que 0,1), indicava que os métodos apresentavam uma concordância, sugerindo relações lineares. Diferenças maiores alertavam para possíveis relações não-lineares ou presença de valores extremos.

Também foi estabelecido que correlações negativas mereciam atenção especial. Pois podem indicar mecanismos de compensação no sistema veicular quando um parâmetro aumenta, outro diminui para manter o equilíbrio (MUKAKA, 2012).

4 RESULTADOS E DISCUSSÕES

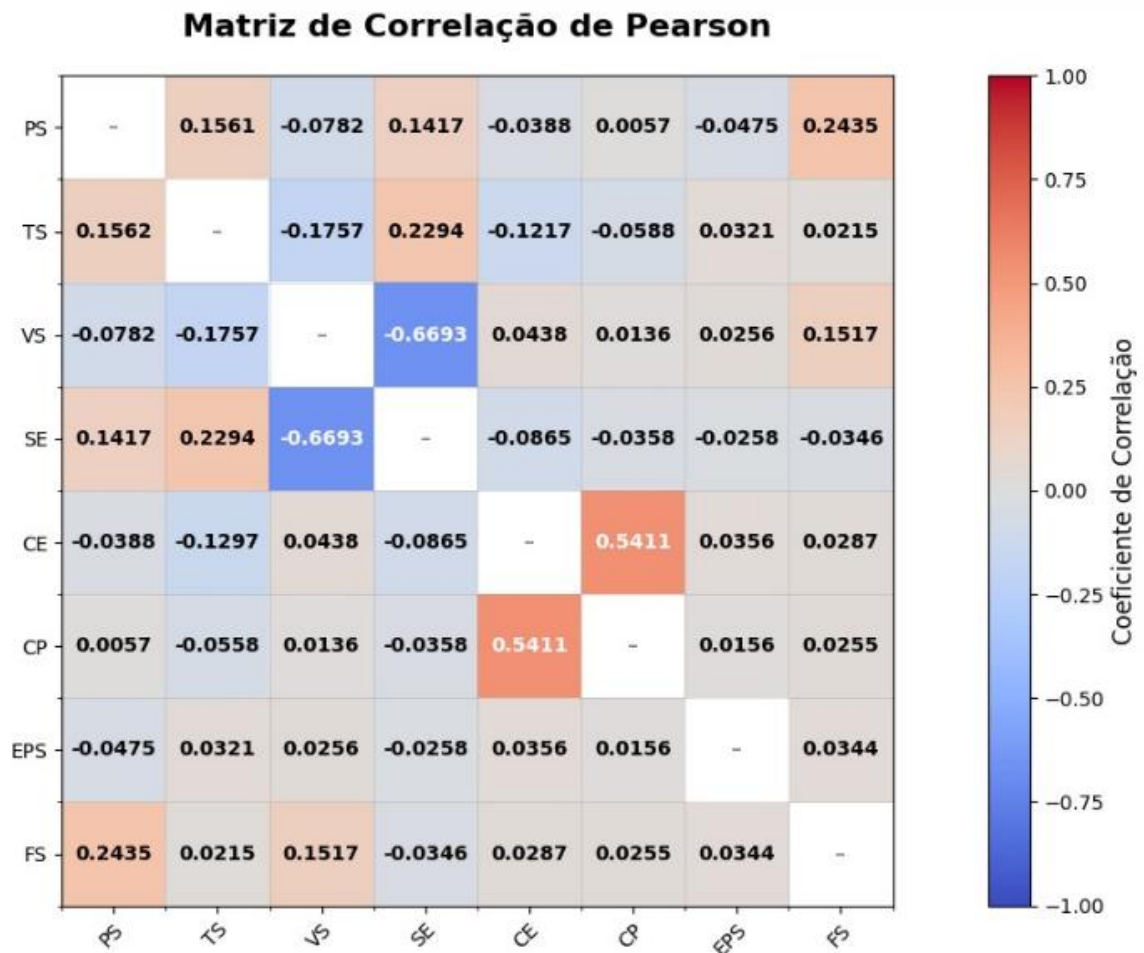
A análise realizada neste projeto teve como principal objetivo identificar possíveis correlações entre as médias dos valores disponibilizados por sensores distintos instalados em veículos. Para isso, foram conduzidas comparações entre os dados dos sensores TS1 a TS4 com os sensores FS1, FS2 e também com a referência EPS1, considerada um padrão confiável para análise.

Durante o processamento, cada ciclo de teste (linha de dados) foi analisado individualmente, com o cálculo dos coeficientes de correlação de *Pearson* e de *Spearman*. Em casos no qual a comparação entre os sensores não resultava em correlação válida, seja por divergência nos tamanhos dos vetores, dados constantes ou ruídos extremos. O sistema emitia alertas informando a exclusão desses pares da média final. Isso garantiu funcionamento do algoritmo, evitando distorções causadas por dados inválidos. Para que verificar a análise de forma visual a foi criado uma matriz de confusão com mapa de calor, como descrito na Seção 2.8.

4.1 COEFICIENTE DE CORRELAÇÃO DE PEARSON

O experimento utilizando o coeficiente de Correlação de *Pearson* é apresentado utilizando uma matriz de confusão com mapa de calor no gráfico 2:

Gráfico 2 - Mapa de Calor do Coeficiente de correlação de *Pearson* entre os sensores de um veículo.



Fonte: Elaborado pelo autor.

A matriz de correlação de *Pearson* apresenta as relações lineares entre oito sensores diferentes, utilizando uma escala de cores que varia do azul (correlações negativas) ao vermelho (correlações positivas).

A escala lateral indica que valores próximos a +1,00 (vermelho intenso) representam correlação positiva máxima, enquanto valores próximos a -1,00 (azul intenso) indicam correlação negativa mínima. Valores próximos a zero (cinza claro) indicam correlação linear nula.

O resultado mais importante do presente trabalho encontra-se na correlação VS-SE, o resultado mais valioso foi a forte correlação entre os sensores VS e SE, com valor de -0,6693. Esse resultado é destaca-se por: (1) sensorear componente diferentes no veículo; (2) Apresentar uma correlação forte negativa, caracterizando uma atividade inversamente proporcional.

Esse fato, faz deles indicadores excelentes para detectar problemas no sistema. Como fazem funções diferentes, quando ambos mostram mudanças ao mesmo tempo, é sinal de que há um problema maior no sistema. Ajuda a descobrir problemas mais cedo, porque se os dois sensores estão reagindo, é certeza de que algo não está funcionando bem.

O que torna essa correlação tão especial é justamente o fato de não serem sensores parecidos que naturalmente andariam juntos, mas sim componentes independentes que, quando se correlacionam, mostram que existe um problema mais amplo no sistema.

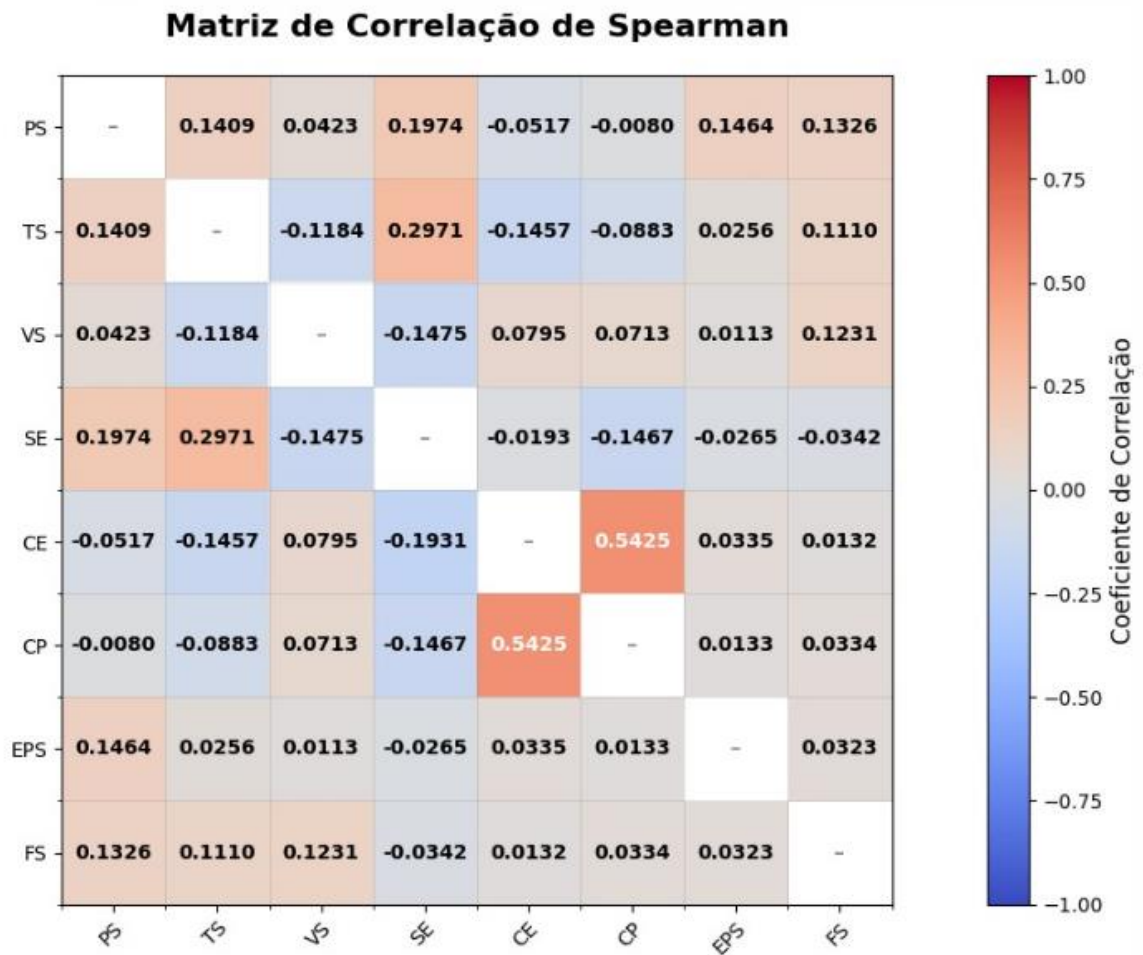
Outras correlações encontradas, os sensores CE e CP mostraram uma correlação positiva de 0,5411. Este resultado tem uma lógica por traz dessa correlação, já que esses dois sensores fazem trabalhos parecidos no sistema, confirmando que o sistema de monitoramento está funcionando de forma coerente.

As demais correlações entre os sensores foram consideradas nulas, por apresentarem valores próximas a zero, indicadas pelas cores cinza claro, o que sugere independência relativa entre suas medições.

4.2 RANK DE SPEARMAN

O experimento utilizando o *Rank de Spearman* é apresentado utilizando uma matriz de confusão com mapa de calor no gráfico 3:

Gráfico 3 - Mapa de Calor do *Rank de Spearman* entre os sensores de um veículo.



Fonte: Elaborado pelo autor.

A matriz de correlação de *Spearman* apresenta as correlações entre os oito sensores, utilizando o mesmo sistema de cores da análise anterior, na qual tons avermelhados indicam correlações positivas e tons azulados representam correlações negativas.

As diferenças mais relevantes em relação à correlação de *Pearson* são dadas pela única correlação mais forte, que agora ocorre entre os sensores CE e CP com um coeficiente de 0,5425, mantendo praticamente o mesmo coeficiente da análise anterior. Essa consistência reforça a relação genuína entre esses sensores que executam funções similares.

A diferença mais notável é apresentada na correlação VS-SE, que caiu drasticamente de

(-0,6693 no gráfico 2), apresenta agora um valor bem menor no gráfico 3, com o coeficiente apresentado -0,1475. Esta diferença indica como os dados estão organizados de forma linear. Já a correlação entre CE e CP se manteve constante (0,5425 com o método *Spearman* contra 0,5411 *Pearson*), mostrando que a relação entre esses sensores de funções correlatas.

Esta grande diferença entre os dois métodos nos mostra que, os dados dos sensores estão organizados de forma linear, o método *Pearson* funciona melhor para este tipo de dados, os padrões seguem relações lineares proporcionais, não apenas de ordem. O coeficiente de correlação de *Pearson*, funcionou melhor com dados organizados de forma linear, identificando relações diretas e proporcionais e medindo o quanto as variáveis se relacionam de forma linear, é um método que conhecido por ser mais sensível a padrões e dados não organizados linearmente. Por outro lado, o *Rank* de *Spearman*, funciona melhor com dados organizados de forma não-linear, identifica relações baseadas em ordem ou *Ranking*, não precisa que a relação seja linear, só que seja consistente na direção, é melhor para dados que seguem padrões de ordem, mesmo que não sejam proporcionais.

Os resultados apresentados indicam que o coeficiente de correlação de *Pearson* funcionou melhor para os sensores estudados. A correlação VS-SE praticamente desapareceu quando usamos o método *Spearman*, mostrando que, a relação entre VS e SE é organizada linearmente, não apenas uma questão de ordem, os dados estão organizados de forma proporcional, o método *Pearson* consegue capturar melhor as relações reais entre os sensores.

A constatação da correlação VS-SE usando o método *Pearson* permite avaliar que a observação dos dois sensores ajuda descobrir problemas mais cedo, antes que se tornem graves e oferecer um indicador confiável de como está a saúde geral do sistema, ajuda a planejar melhor quando fazer manutenções preventivas.

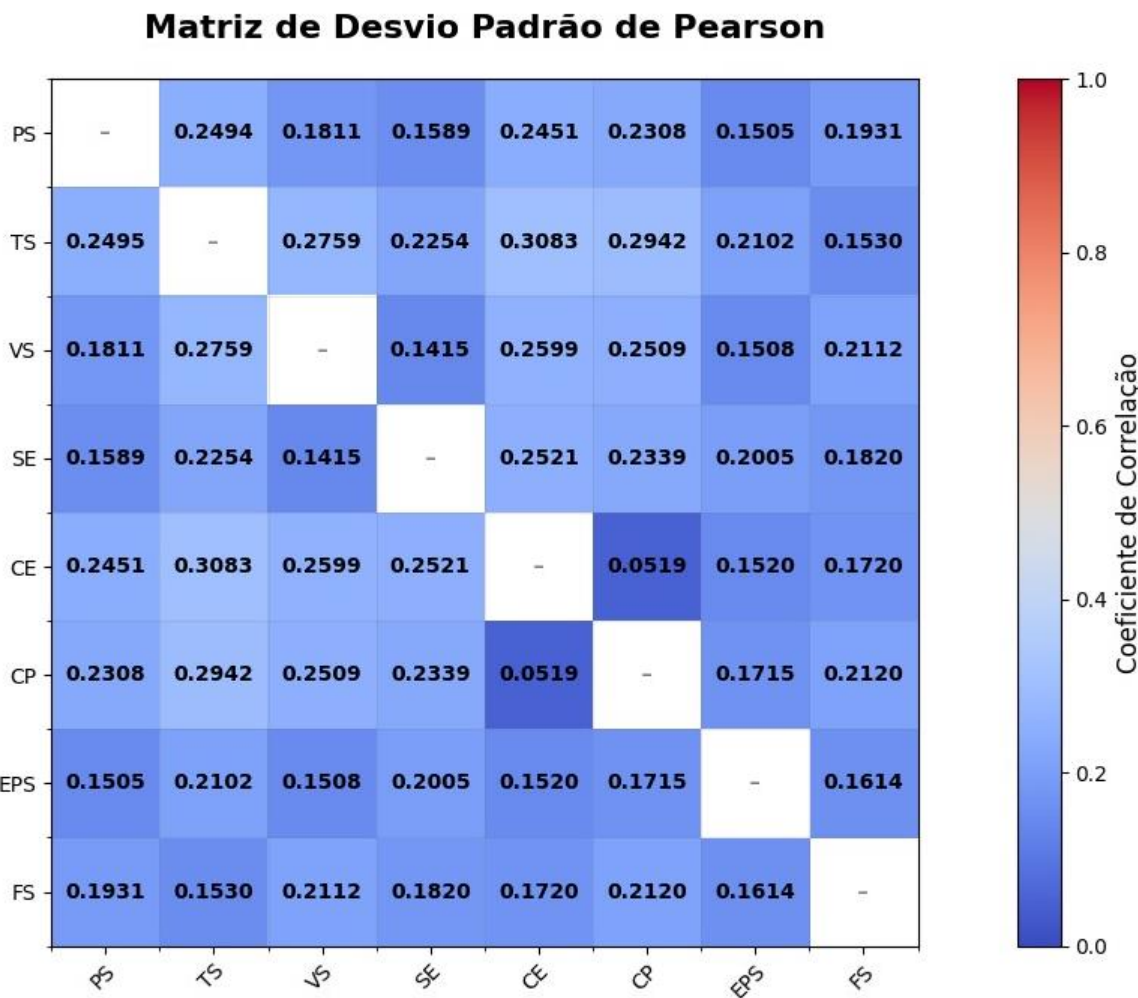
Estes resultados abrem portas para criar sistemas inteligentes de detecção de problemas baseados na correlação VS-SE, desenvolver alertas que avisem quando algo está começando a dar errado, melhorar as estratégias de manutenção preventiva, usar técnicas mais avançadas como inteligência artificial para prever problemas

A união entre análise estatística e manutenção de veículos pode ser muito mais eficaz quando entendemos adequadamente como sensores de funções diferentes se relacionam, conforme demonstrado pela análise de correlação de *Pearson*.

4.3 DESVIO PADRÃO PEARSON

Além de avaliar a correlação de *Pearson* e o *Rank* de *Spearman*, o presente estudo também buscou encontrar pistas sobre a estabilidade dos resultados providos pelas correlações utilizando o desvio padrão para isso. Dessa forma o desvio padrão para os resultados gerados pela técnica de *Pearson* é apresentado utilizando uma matriz de confusão com mapa de calor no gráfico 4:

Gráfico 4 - Mapa de Calor do Desvio padrão de *Pearson* entre os sensores de um veículo.



Fonte: Elaborado pelo autor.

Analisando esta matriz de desvio padrão das correlações de *Pearson*, pode-se observar

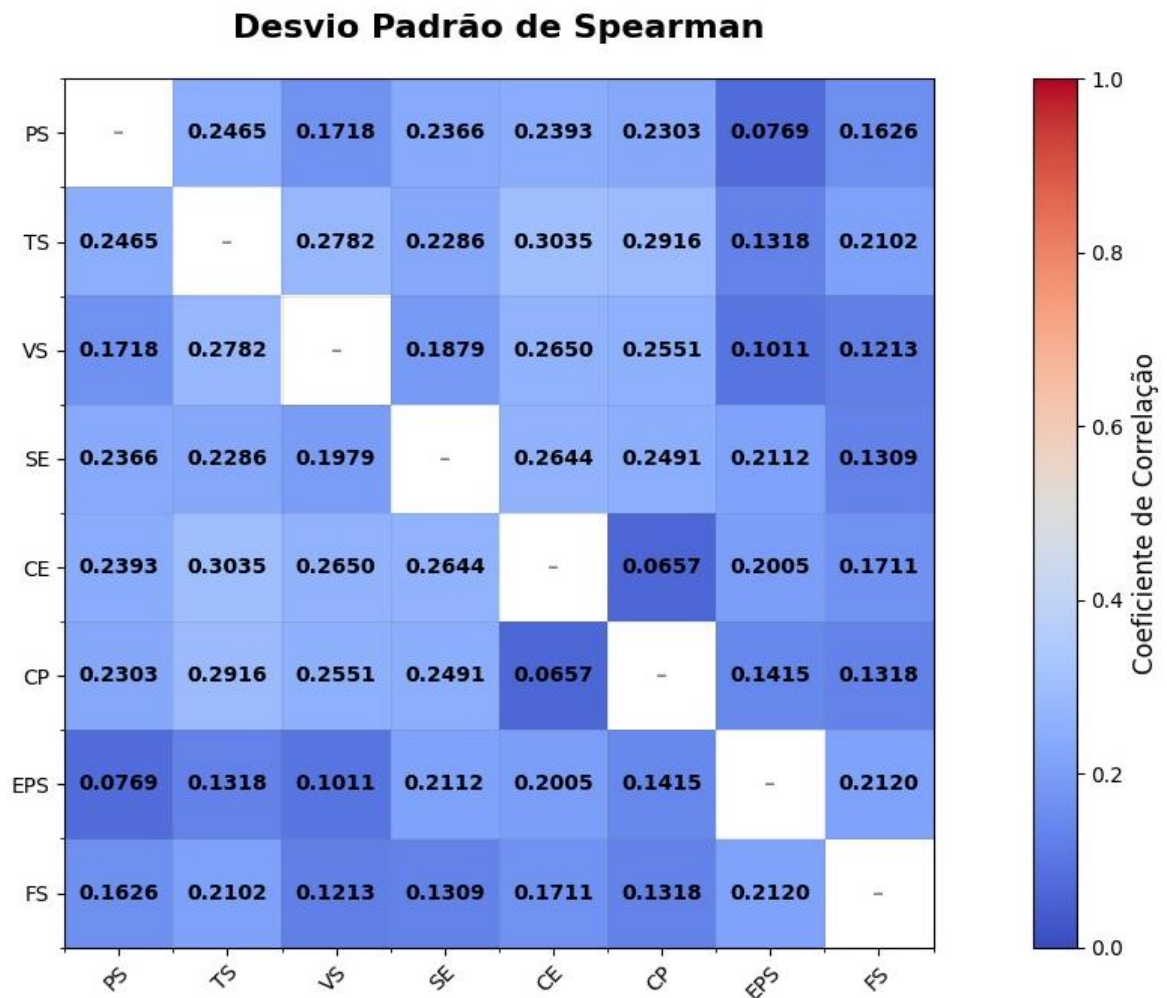
padrões que medem a variância das correlações entre as variáveis. Os valores menores indicam correlações mais consistentes e confiáveis ao longo do tempo ou entre diferentes amostras. As correlações mais estáveis, como apresentadas no Gráfico 4 são: CE e CP: 0,0519 VS e SE: 0,1415; PS e EPS: 0,1505; VS e EPS: 0.1508; e TS e FS: 0.1530. Os resultados com maior variabilidade foram: TS e CE: 0.308; TS e CP: 0.2942; VS e TS: 0.2759; VS e CE: 0.2599.

Todos os valores permanecem no intervalo adequado (0 a 1). A relação CE-CP continua destaca-se pela estabilidade, enquanto as interações envolvendo TS mostram maior imprevisibilidade. O resultado entre os sensores VS e SE confirmam que além de conter uma boa correlação no experimento utilizando o coeficiente de correlação de Pearson, os resultados também são bastantes estáveis.

4.4 DESVIO PADRÃO SPEARMAN

O experimento utilizando o desvio padrão de *Spearman* é apresentado utilizando uma matriz de confusão com mapa de calor no gráfico 5:

Gráfico 5 - Mapa de Calor do Desvio padrão de *Spearman* entre os sensores de um veículo.



Fonte: Elaborado pelo autor.

A lógica do desvio padrão dos resultados utilizando o *Rank* de *Spearman* seguem o mesmo modelo apresentado pelo desvio padrão utilizado para os resultados apresentados no coeficiente de correlação de *Pearson*, apresentados na Seção 4.3. As correlações mais estáveis, como apresentadas no Gráfico 5 são: CE e CP: 0,0657; PS e EPS: 0,0769; VS e EPS: 0,1011; VS e FS: 0,1213; TS e EPS: 0,1318. Já os resultados que apresentaram maior variabilidade foram: TS e CE: 0,3035; TS e CP: 0,2916; VS e TS: 0,2782 e; VS e CE: 0,2650.

Esses resultados indicaram que, o método de *Spearman* não é o melhor indicado para

os sensores estudados. Pois nenhuma correlação entre sensores de diferentes funções foram destacadas no modelo.

4.5 DISCUSSÃO DOS RESULTADOS

Durante o desenvolvimento do estudo, foi possível perceber que interpretar correlações vai muito além de simplesmente calcular coeficientes. Cada resultado conta uma história sobre como os sistemas do veículo interagem entre si.

A diferença marcante entre os resultados de desvio padrão e correlação direta representa um aspecto intrigante da análise. Após reflexão, conclui-se que estão sendo observados dois fenômenos diferentes: um relacionado ao comportamento temporal dos sensores e outro às relações instantâneas entre suas medições. Isso tem implicações importantes para quem trabalha com manutenção, pois sugere que problemas de longo prazo podem ter sinais diferentes dos problemas imediatos.

A discrepância entre Pearson e Spearman em alguns casos, como VS-SE, demonstra a importância de não confiar em apenas um método. Quando Pearson indica uma correlação forte (-0,67) e Spearman uma bem mais fraca (-0,15), isso confirma que existe uma relação linear genuína entre os sensores. A manutenção da correlação forte no método Pearson, combinada com o enfraquecimento no Spearman, indica que os dados seguem padrões lineares proporcionais, não apenas de ordem ou Ranking, validando ainda mais a robustez da correlação VS-SE para detecção de problemas sistêmicos.

Os sensores VS e SE emergiram como a combinação mais interessante para monitoramento veicular. Sua correlação de -0,6693 representa a relação mais forte identificada no estudo, sendo especialmente significativa porque esses sensores executam funções completamente diferentes no veículo. Esta independência funcional torna a correlação ainda mais relevante: quando sensores com propósitos distintos mostram comportamentos correlacionados, isso indica problemas sistêmicos mais amplos. A correlação VS-SE funciona como um sistema de alerta precoce, permitindo detectar falhas antes que se tornem críticas, oferecendo cobertura sistêmica superior ao monitoramento individual de cada sensor.

O caso CE-CP merece destaque especial por representar uma situação de controle ideal. Como ambos os sensores executam a mesma função no sistema, era esperado que apresentassem correlação positiva consistente. Os resultados confirmaram essa expectativa: *Pearson* mostrou 0,5411 e *Spearman* 0,5425, valores praticamente idênticos. Essa

concordância entre os métodos, aliada ao conhecimento técnico de que os sensores têm função equivalente, válida tanto a qualidade dos dados quanto a confiabilidade dos métodos de análise utilizados.

Essa validação com CE-CP fornece maior confiança para interpretar outras correlações consistentes no *dataset*. Quando ambos os métodos concordam em outros pares de sensores, provavelmente está-se diante de relações reais e estáveis no sistema, enquanto as discrepâncias sinalizam a necessidade de investigação mais aprofundada sobre comportamentos não-lineares ou presença de *outliers*.

Na análise do desvio padrão das correlações, todos os valores permaneceram positivos, variando entre 0,0519 e 0,3083 para Pearson, e entre 0,0657 e 0,3035 para Spearman. Os sensores CE e CP apresentaram a maior estabilidade com desvio padrão de apenas 0,0519, indicando que essa relação é altamente confiável ao longo do tempo. Por outro lado, as interações envolvendo TS mostraram maior variabilidade, com valores próximos a 0,31, sugerindo menor previsibilidade nessas correlações.

4.6 INTERPRETAÇÃO PARA A MANUTENÇÃO

A aplicação desses resultados na prática da manutenção automotiva representa o aspecto mais relevante desta pesquisa. Os números ganham significado real quando analisados sob a perspectiva de quem trabalha com frotas.

A correlação moderada entre CE e CP (0,54) representa um caso especial, pois esses sensores executam a mesma função no sistema. Esta correlação consistente entre *Pearson* e *Spearman* confirma que os sensores estão operando conforme esperado e fornece uma base confiável para estratégias de manutenção. Na prática, isso permite que uma oficina use um sensor para validar as leituras do outro, implementando redundância para aumentar a confiabilidade do diagnóstico. Quando essa correlação começar a se desviar significativamente de 0,54, será um indicador precoce de degradação ou falha em um dos sensores, permitindo manutenção preventiva antes da falha completa.

A discrepância entre VS e SE apresenta um cenário mais complexo. A forte correlação linear negativa detectada por *Pearson* (-0,67) versus a correlação fraca do *Spearman* (-0,15) confirma que existe uma relação linear genuína entre esses sensores. Isso indica que os dados seguem padrões lineares proporcionais, onde mudanças em VS provocam ajustes proporcionais

em SE.

Para a manutenção, isso significa que algoritmos de monitoramento linear podem capturar adequadamente essa relação, validando a eficácia de sistemas de detecção baseados na correlação VS-SE para identificar anomalias no sistema. Para sensores com correlações consistentemente fracas, como a maioria dos analisados, a interpretação é que fornecem informações independentes e igualmente importantes sobre diferentes aspectos do sistema veicular. Isso significa que cada um deve ser monitorado individualmente, sem assumir que o comportamento de um prediz o comportamento de outro.

Considerando sistemas de manutenção preditiva, esses resultados sugerem estratégias diferenciadas: sensores com correlações consistentes e moderadas (como CE-CP) podem ser agrupados para análise conjunta e validação cruzada, enquanto aqueles com correlações fracas necessitam de algoritmos independentes de monitoramento. Os casos de discrepância entre métodos requerem abordagens híbridas que considerem tanto relações lineares quanto não-lineares.

5. CONCLUSÕES

O estudo focou-se em entender como os sensores dos carros conversam entre si por meio de coeficientes de correlações. Foram observados dois métodos: *Pearson* e *Spearman*, para analisar oito sensores diferentes e buscar padrões que pudessem ajudar na manutenção preventiva dos veículos.

O resultado que mais destacou-se foi a forte correlação negativa entre o sensor de vibração (VS) e o de eficiência do sistema (SE) – apresentando um coeficiente de -0,6693 usando *Pearson*. Algo interessante foi que esses sensores trabalham de forma independente no carro, mas ao detectar problemas, o outro mensura na mesma proporção. Isso significa que se tem um indicador muito confiável de quando algo está começando a dar errado no veículo.

Também foi constatado algo lógico, mas igualmente importante: os sensores de consumo energético (CE) e potência consumida (CP) mostraram correlação positiva de 0,5411. Como eles monitoram aspectos parecidos do sistema, algo esperado que andassem juntos - e isso confirmou que os dados estavam consistentes.

A escolha do método faz considerável diferença. Para a relação vibração-eficiência, *Pearson* captou uma correlação forte (-0,6693), enquanto *Spearman* mostrou algo bem menos significativo (-0,1475). Essa observação indica que os dados seguem um padrão linear, tornando *Pearson* mais adequado para essa situação específica.

Já para consumo energético e potência, os dois métodos deram resultados muito parecidos (0,5411 e 0,5425), mostrando que essa relação é robusta independentemente da abordagem usada.

Uma característica que chamou à atenção foi verificar o quão estáveis essas correlações se mantiveram. A dupla CE-CP foi a que mais se destacou, com um desvio padrão de apenas 0,0519 - ou seja, é uma relação bastante confiável.

Por outro lado, qualquer correlação envolvendo o sensor de temperatura (TS) mostrou mais variabilidade, com desvios próximos a 0,31. Isso significa que o sensor apresenta alta instabilidade e um comportamento menos previsível.

A correlação entre vibração e eficiência pode ser a base para sistemas que detectam problemas antes que eles se tornem grandes dores de cabeça. Em vez de monitorar cada sensor isoladamente, pode-se criar um sistema que entende quando algo não está certo no conjunto.

O trabalho trouxe algumas contribuições práticas, tais como: identificação de

correlações específicas que podem ser usadas para detectar problemas de forma precoce; a validação de métodos estatísticos para trabalhar com dados de sensores automotivos, e criou-se diretrizes que podem ser implementadas em sistemas reais de manutenção preventiva.

O estudo ainda reconhece que trabalhar com um número limitado de veículos e situações específicas de uso. Em um cenário ideal, expandir para frotas maiores e condições mais variadas de operação. Além disso, implementar esse estudo na prática ainda necessitam de testes em ambiente real, com monitoramento contínuo.

Estudar como os sensores se relacionam é uma ferramenta valiosa para o campo de manutenção preditiva. A correlação vibração-eficiência como um indicador geral do sistema, junto com uma relação entre consumo energético e potência, oferece uma base sólida para sistemas mais inteligentes de monitoramento. Acredita-se que combinar análise estatística com conhecimento técnico automotivo pode revolucionar a manutenção preventiva, tornando o transporte mais confiável, reduzindo custos e aumentando a segurança.

Como trabalhos futuros o estudo busca avaliar modelos de aprendizado de máquinas para prever a manutenção de veículos agrícolas em dados reais.

REFERÊNCIAS

- BEYER, K.; GOLDSTEIN, J.; RAMAKRISHNAN, R.; SHAFT, U. *When Is ‘Nearest Neighbor’ Meaningful?*, in 7th International Conference on Database Theory – ICDT’99, 1999, pp. 217–235.
- BLEY, T.; PIGNANELLI, E.; SCHÜTZE, A. *COPS - Combined Oil quality and Particle measurement System*, ICST 2012, Sixth International Conference on Sensing Technology, Kolkata, India, December 18 - 21, 2012.
- CHICCO, D.; JURMAN, G. *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation*. BMC Genomics, v. 21, n. 1, p. 6, 2020.
- EL-BETAR, A. et al., *Fault Diagnosis of a Hydraulic Power System Using an Artificial Neural Network*, JKAU: Eng. Sci., Vol. 17 No. 1, pp: 117 – 137, 2006.
- FAWCETT, T. *An introduction to ROC analysis*. Pattern Recognition Letters, v. 27, n. 8, p. 861–874, 2006.
- F. MARQUEZ et al., *Condition monitoring of wind turbines: Techniques and methods*, Renewable Energy 46, 169-178, Elsevier, 2012.
- GUENEL, A.; MESHAM, A.; BLEY, T.; SCHÜTZE, A.; KLUSCH M., *Statistical and Semantic Multisensor Data Evaluation for Fluid Condition Monitoring in Wind Turbines*, Proc. 16th International Conference on Sensors and Measurement Technology (SENSOR), Nuremberg, Germany, 2013.
- HELWIG, N.; PIGNANELLI, E.; SCHÜTZE, A. *Condition Monitoring of a Complex Hydraulic System Using Multivariate Statistics*. In: IEEE INTERNATIONAL INSTRUMENTATION AND MEASUREMENT TECHNOLOGY CONFERENCE, 2015, Pisa. Proceedings... Pisa: IEEE, 2015. p. PPS1-39. DOI: 10.1109/I2MTC.2015.7151267.
- HELWIG, N.; SCHÜTZE, A. *Detecting and compensating sensor faults in a hydraulic condition monitoring system*. In: INTERNATIONAL CONFERENCE ON SENSORS AND MEASUREMENT TECHNOLOGY, 17., 2015, Nuremberg. Proceedings... Nuremberg: AMA Association for Sensor Technology, 2015. Oral presentation D8.1. DOI: 10.5162/sensor2015/D8.1.

HIGGS, P.; PARKIN, R.; JACKSON, M.; AL-HABAIBEH, A.; ZORRIASSATINE, F.; COY, J. **A Survey on Condition Monitoring Systems in Industry**, ASME 7th Biennial Conference on Engineering Systems Design and Analysis, No. ESDA2004-58216, pp. 163-178, 2004.

(International Standard) “ISO 13372 - **Condition monitoring and diagnosis of machines - Vocabulary**,” 2004.

KANO, M. NAKAGAWA, Y. **Data-Based Process Monitoring, Process Control, and Quality Improvement: Recent Developments and Applications in Steel Industry**, Computers & Chemical Engineering, Elsevier, Volume 32, Issues 1–2, January 2008, Pages 12–24.

LAOUTI, N.; SHEIBAT-OTHMAN, N., OTHMAN, S. **Support Vector Machines for Fault Detection in Wind Turbines**, Preprints of the 18th IFAC World Congress Milano (Italy) August 28 - September 2, 2011.

LIU, X. et al., **Study on Knowledge -based Intelligent Fault Diagnosis of Hydraulic System**, TELKOMNIKA, Vol.10, No.8, December 2012, pp. 2041~2046, e-ISSN: 2087-278X.

MARQUES DE SÁ, J.P. **Pattern Recognition – Concepts, Methods and Applications**, Springer Verlag, Berlin, 2001.

MUKAKA, M. M. **Statistics corner: A guide to appropriate use of correlation coefficient in medical research**. Malawi Medical Journal, v. 24, n. 3, p. 69–71, 2012.

OLSZEWSKI, R. T. **Generalized feature extraction for structural pattern recognition in time-series data**, PhD-thesis, Carnegie Mellon University, 2001.

POWERS, D. M. W. **Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation**. arXiv, 2011.

SCHNEIDER, T.; HELWIG, N.; SCHÜTZE, A. **Automatic feature extraction and selection for classification of cyclical time series data**. tm - Technisches Messen, Berlin, v. 84, n. 3, p. 198-206, 2017. DOI: 10.1515/teme-2016-0072.

SOKOLOVA, M.; LAPALME, G. **A systematic analysis of performance measures for classification tasks**. Information Processing & Management, v. 45, n. 4, p. 427–437, 2009.

TCHAKOUA, P. et al., **Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges**, Energies 2014, 7, 2595-2630, doi:10.3390/en7042595.

WASKOM, M. L. ***Seaborn: statistical data visualization***. [S. l.]: Python Software Foundation, 2014.

WILCOX, R. ***Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy***, Springer Science & Business Media, 2010, ISBN 1441955240.

WOLD, S.; ESBENSEN, K. H.; GELADI, P. ***Principal component analysis***, " *Chemom. Intell. Lab. Syst.*, vol. 2.1, pp. 37–52, 1987.