



UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO



Julio Cartier Maia Gomes

SMART SPR: Identificação e Análise das Respostas
Fornecidas por Sensores Baseados em Ressonância de
Plasmons de Superfície

Mossoró-RN

2020

Julio Cartier Maia Gomes

***SMART* SPR: Identificação e Análise das Respostas
Fornecidas por Sensores Baseados em Ressonância de
Plasmons de Superfície**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação - associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido, para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof^o Dr. Leiva Casemiro Oliveira

Coorientador: Prof^o Dr. Leandro Carlos de Souza

Mossoró-RN

2020

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

G633s Gomes, Julio Cartier Maia.
SMART SPR: Identificação e Análise das
Respostas Fornecidas por Sensores Baseados em
Ressonância de Plasmons de Superfície / Julio
Cartier Maia Gomes. - 2020.
119 f. : il.

Orientador: Leiva Casemiro Oliveira.
Coorientador: Leandro Carlos de Souza.
Dissertação (Mestrado) - Universidade Federal
Rural do Semi-árido, Programa de Pós-graduação em
Ciência da Computação, 2020.

1. Sensor Óptico. 2. Ressonância de Plasmons de
Superfície (RPS). 3. Aprendizagem de Máquina. 4.
Sensor Inteligente. I. Oliveira, Leiva Casemiro,
orient. II. Souza, Leandro Carlos de , co-orient.
III. Título.

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC's) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

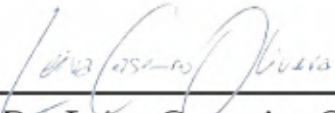
Julio Cartier Maia Gomes

***SMART* SPR: Identificação e Análise das Respostas
Fornecidas por Sensores Baseados em Ressonância de
Plasmons de Superfície**

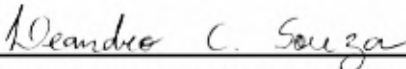
Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do título de Mestre em Ciência da Computação.

APROVADA EM: 01 / 06 / 2020.

BANCA EXAMINADORA



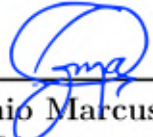
Prof^o Dr. Leiva Casemiro Oliveira
Orientador (Presidente) - UFERSA



Prof^o Dr. Leandro Carlos de Souza
Coorientador - UFPB



**Prof^o Dr. Marcus Vinicius Silverio
Costa**
Examinador Interno - UFERSA



**Prof^o Dr. Antonio Marcus Nogueira
Lima**
Examinador Externo - UFCG

Dedico esse trabalho aos meus pais Francisco Val-
derez e Margarida Soares pelo o exemplo.

Agradecimentos

Agradeço aos meus pais Francisco Valderez e Margarida Soares, pelo o amor, educação, ensino e por sempre acreditar no meu potencial para o meu crescimento pessoal e profissional. Agradeço aos meus irmãos Francisco Cristovão, José Valbercon, cunhadas Lidiane Silva, Gessica Alves, minhas sobrinhas Ana Lydia e Maria Clara, por todo o amor, conversas nos momentos mais difíceis que tenho passado e como também os exemplos que vocês tem ensinado. Também gostaria de agradecer aos meus avós paternos, Maria Nilce e José Gomes pelo o amor incondicional. Agradeço a minha prima Maria Beatriz, por sempre escutar minhas dificuldades e me apoiar com suas palavras.

Agradeço principalmente Profº Dr. Leiva Casemiro, por me apresentar a área de sensores baseado em SPR. E como também, sempre me incentivando a aprender da melhor forma possível e por acreditar no meu potencial, por desenvolver esse trabalho e demais trabalhos. O Senhor é um exemplo que levo para a vida toda. Agradeço também, ao Profº Dr. Leandro Carlos por me ensinar a parte de aprendizagem de máquina e incorporar ao trabalho, como também por acreditar no meu desenvolvimento e sempre retirar e solucionar as dúvidas em nossas reuniões.

Ao Profº Francisco de Assis (Xico) da UFERSA-Angicos por ter me proporcionado todo o ensino na graduação como também ter me ajudado nessa trajetória. Em especial, esse trabalho não teria sido desenvolvido sem o apoio, amor e carinho de Cynthia Moreira Maia, por sempre me escutar e tirar as minhas dúvidas. Desde a graduação me apoiando para o desenvolver o meu melhor, tanto no pessoal como no profissional.

Meus amigos Paulo Morais, Jedson Oriel, Bruna Morais e Vinicius Sales, Ariel, Thomas, Antonio Queiroz, Otilia Santos, Jackson Sousa, Josiran Barbosa, Thyago Alves, Elizeu Sandro e por todas as conversas como também incentivo. Agradeço também aos meus colegas de mestrados que fiz durante essa trajetória. Agradeço a CAPES e a UFERSA pelo apoio financeiro e bolsa de pesquisa.

”A cada geração, uma parte da humanidade se compromete em deixar o mundo um lugar melhor do que encontrou. Esse contingente populacional acredita que, para tanto, é preciso investir em inovações.”

Tiago Reis & Felipe Tadewald.

Resumo

Os sensores baseados em Ressonância de *Plasmons* de Superfície (RPS) permitem avaliar mudanças de índice de refração, com objetivo de detectar interações moleculares em segurança alimentar, proteção ambiental, detecção de diferentes compostos químicos, bioquímicos e orgânicos entre diversas outras aplicações. A confiabilidade nas informações repassadas por um sensor RPS deve ser garantida, diante que informações erradas podem ocasionar consequências, tais como, más interpretações e manipulações das ligações moleculares. Para melhorar e atestar a qualidade nas respostas fornecidas no sensor baseado na RPS, este trabalho investiga o uso de diferentes técnicas de aprendizagem de máquina e de descritores em sensorgramas adquiridos. De forma mais específica, este trabalho busca identificar padrões e anomalias no seu comportamento nas respostas RPS e analisar o impacto dessas técnicas de aprendizagem na qualidade do sensor. Os resultados obtidos por meio da análise estatística permitiu concluir que o descritor temporal de sensorgrama obteve um melhor desempenho com a classificação e que o uso da regressão linear possibilitou analisar as áreas de interesses no sensorgrama com resultados satisfatórios, criando assim, um sensor RPS com resposta inteligente. Este trabalho obteve como resultado também a criação de uma interface gráfica para a classificação de sensorgramas. Dessa forma, como contribuição científica tem-se aplicação de técnicas de aprendizagem de máquina que permitiram identificar, analisar e classificar as respostas do sensorgrama.

Palavras-chave: Sensor Óptico, Ressonância de Plasmons de Superfície (RPS), Aprendizagem de Máquina, Sensor Inteligente.

Abstract

Sensors based on Surface Plasmon Resonance (SPR) allow the evaluation of refractive index shifts, with the objective of detecting molecular interactions in food safety, environmental protection, detection of different chemical, biochemical and organic compounds, among several other applications. The reliability of the information transmitted by an SPR sensor must be guaranteed, given that wrong information can cause consequences, such as, misinterpretations and manipulations of molecular bonds. To improve and attest to the quality of the responses provided in the SPR-based sensor, this work investigates the use of different machine learning techniques and descriptors in acquired sensorgrams. More specifically, this work seeks to identify patterns and anomalies in their behavior in the SPR responses and to analyze the impact of these learning techniques on the quality of the sensor. The results obtained through the statistical analysis allowed to conclude that the sensorgram temporal descriptor obtained a better performance with the classification and that the use of linear regression made it possible to analyze the areas of interest in the sensorgram with satisfactory results, thus creating an SPR sensor with a intelligent response. This work also resulted in the creation of a graphical interface for the classification of sensorgrams. That way, the scientific contribution is the application of machine learning techniques that allowed to identify, analyze and classify the responses of the sensorgram.

Keywords: Optical Sensor, Surface Plasmon Resonance (SPR), Machine Learning (ML), Smart Sensor

Lista de ilustrações

Figura 1 – Estrutura Multicamada composta pelo prisma com espessura d_1 e permissividade elétrica ϵ_1 , Metal com espessura d_2 e permissividade elétrica ϵ_2 e os dielétricos (amostra) com espessura d_3 e d_4 , permissividade elétrica ϵ_3 e ϵ_4 . Vetores, ângulo e plano de incidência indicados na Figura.	23
Figura 2 – Excitação dos plasmons de superfície na geometria de Kretschmann e Raether (1968) no método de reflexão total atenuada. d_1 é a espessura, ϵ_1 a permissividade elétrica do prisma, ϵ_2 permissividade elétrica do metal, d_2 espessura do metal, d_3 é a espessura e ϵ_3 permissividade elétrica do dielétrico, k_{px} vetores de onda da luz incidente e k_{sp} vetores plasmons de superfície.	30
Figura 3 – Exemplo de curva SPR, dada pelo gráfico refletividade vs. ângulo (modo AIM) ou refletividade vs. comprimento de onda (modo WIM). O valor mínimo indica a posição de ressonância: λ_{res1} e λ_{res2} no modo WIM e θ_{res1} e θ_{res2} . Os valores C_R e C_L são parâmetros para calcular a largura e a assimetria da curva SPR.	31
Figura 4 – Exemplo de um ciclo experimental em um sensorgrama no modo WIM, em que foram inserido as substâncias $H_2O \rightarrow BSA \rightarrow H_2O \rightarrow Hypo. \rightarrow H_2O$. Destaque para as fases do sensorgrama: linha base, absorção, monolayer, dissociação, lavagem e regeneração.	34
Figura 5 – Estrutura da aprendizagem de máquina dividida em: Supervisionada e Não-Supervisionada.	35
Figura 6 – Exemplo do Algoritmo k -NN.	37
Figura 7 – Exemplo de um modelo de regressão linear múltipla.	42
Figura 8 – (a) Sinal de um relógio que está ligado e desligado no decorrer do tempo. (b) Obtenção dos coeficientes DC e AC com a DCT-4 do sinal do relógio.	46
Figura 9 – (a) Criação de um sinal com uma função $f(x) = 2*\text{rand}(x,1)$ em um domínio de tempo discreto. (b) Obtenção dos coeficientes DC e AC com a FChT do sinal da função $f(x)$.	47
Figura 10 – (a) Criação de um sinal com a função $f(x) = 10*\text{rand}(x,1)$ em um domínio de tempo discreto. (b) Obtenção dos coeficientes DC e AC com a FWHT do sinal da função $f(x)$.	48
Figura 11 – Diagrama de Blocos indicando as etapas para realização das tarefas de criar, descrever, identificar, classificar e tornar a resposta automática dos sensorgramas.	49
Figura 12 – Diagrama de Blocos do sensor SPR.	50

Figura 13 – Sensorgramas obtidos no modo AIM PC/Au, BK7/Au, PMMA/Au e TOPAS/Au utilizando o metal Au de 50 nm. O protocolo experimental inicializado com H_2O , depois Hipoclorito, retorno para H_2O , em seguida PBS, retorno H_2O ; depois inserido a solução BSA, remoção de excesso com H_2O , limpeza com Hipoclorito e finalizado com H_2O	51
Figura 14 – Sensorgramas obtidos no modo WIM com o vidro BK7 e polímeros PC, PMMA e TOPAS utilizando o metal Ag. O valor mínimo do comprimento de onda de ressonância foi utilizado o algoritmo minimum-hunt. O protocolo experimental para o vidro BK7 e polímero PC foi inicializado com H_2O , depois Etanol 25%, retornando para H_2O , em seguida Etanol 12,5%, retornou-se para H_2O ; depois BSA-monolayer, remoção com H_2O ; limpeza com Etanol 12,5% finalizado com H_2O . Para o polímero PMMA iniciou-se com H_2O ; inseriu-se Etanol 6,2%, retornou para H_2O ; depois Etanol 1,5%, retornou para H_2O ; depois BSA-Monolayer; retornou-se com H_2O , inserido o Etanol a 6,2% finalizado com H_2O . Com o polímero TOPAS inseriu H_2O ; depois Etanol 25%; retornou para H_2O , depois Etanol 6,25%, retorno para H_2O , em seguida BSA-Monolayer, remoção com H_2O ; inserido o Etanol a 6,25% e finalizado com H_2O	52
Figura 15 – Inserção de ruído gaussiano branco em sensorgramas com o vidro BK7 e o metal ouro. (a) Inserção com 20 dBW (b) 30 dBW. (c) 40 dBW e (d) 50 dBW.	54
Figura 16 – Erro médio quadrático da inserção de ruído gaussiano branco em sensorgramas com o vidro BK7 e o metal ouro.	54
Figura 17 – Aumento da posição de ressonância. (a) Sensorgrama original com o vidro BK7 e o metal ouro (BK7/Au). (b) Criado um novo sensorgrama com o aumento da posição de ressonância correspondente ao eixo y em $0.5(^{\circ})$	55
Figura 18 – Modificações temporais nas substâncias do sensorgrama. (a) Sensorgrama original com o vidro BK7 e o metal ouro (BK7/Au). (b) Criado um novo sensorgrama com a diminuição temporal das substâncias BSA-Monolayer e Hipoclorito.	55
Figura 19 – Inserção de bolha de ar na substância. (a) Sensorgrama original com o vidro BK7 e o metal ouro (BK7/Au). (b) Criado um novo sensorgrama com inserção de bolha de ar na substância BSA-Monolayer.	56

Figura 20 – Reconstrução do sensorgrama com o vidro BK7 e o metal ouro BK7/Au. (a), (b) e (c) apresenta a reconstrução utilizando os descritores FWHT, FChT e DCT-4 com 9, 100, 200 e 300 coeficientes AC. (d) reconstrução do sensorgrama com o descritor TSD com 9, 100, 200 e 300 coeficientes de números complexos.	57
Figura 21 – Erro Médio Quadrático utilizando a estrutura BK7/Au do grupo 1 com cada descritor	58
Figura 22 – Sensorgrama para o protocolo $H_2O \rightarrow \text{Etanol } 25\% \rightarrow H_2O \rightarrow \text{Etanol } 12.5\% \rightarrow H_2O \rightarrow \text{BSA-Monolayer} \rightarrow H_2O \rightarrow \text{Etanol } 25\% \rightarrow H_2O$ (azul). Valores estimados da regressão do regime permanente (vermelho). Valores estimados da regressão do regime de transição (preto).	61
Figura 23 – Protocolo experimental para detecção de Leishmaniose positivo e negativo. O protocolo (a) inicial é inserido a H_2O , em seguida inserido o PBS; depois com a inserção do peptídio, em seguida a solução BSA; retornando para o PBS, depois a inserção da solução leishmaniose positiva e lavagem para o PBS. (b) inicializou com H_2O , em seguida inserido o PBS; depois a inserção do peptídio; em seguida a solução BSA; retornando para o PBS, depois a inserção da solução leishmaniose negativa e lavagem com o PBS.	63
Figura 24 – Protocolo experimental para detecção de dengue negativo e positivo. O protocolo (a) inicial é inserido PBS, em seguida inserido a solução NA; depois com a inserção de PBS novamente; em seguida a solução DENV-2 AB; retornando para o PBS; depois a inserção de um Exemplo de Soro; retornando para PBS; em seguida inseriu o teste positivo; e ao final a inserção de PBS. (b) inicializado com PBS, em seguida inserido a solução NA; depois com a inserção de PBS novamente; em seguida a solução DENV-2 AB; retornando para o PBS; depois a inserção de um Exemplo de Soro; retornando para PBS; em seguida inseriu o teste negativo; e ao final a inserção de PBS.	64
Figura 25 – Aplicação do algoritmo TSD em que aplicou-se no protocolo experimental $H_2O \rightarrow \text{Hypo.} \rightarrow H_2O \rightarrow \text{PBS} \rightarrow H_2O \rightarrow \text{BSA} \rightarrow H_2O \rightarrow \text{Hypo.} \rightarrow H_2O$. A média de cada substância inserida no sensorgrama é representada no ponto azul. A tabela é composta pela média e o comprimento do protocolo experimental.	67
Figura 26 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o vidro BK7 com o metal ouro BK7/Au. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.	68

Figura 27 – Conjunto de dados de treinamentos em um gráfico 3D com os 99 sensorgramas dos 15 grupos para a (a) FWHT, (b) FChT e (c) DCT-4. (d) Conjunto de dados de treinamento com 3 regimes permanentes em um gráfico 2D do TSD.	69
Figura 28 – Resultados dos intervalos de confiança com o teste não paramétrico <i>bootstrap</i> para o $k = 3$, com o vidro BK7 e o metal ouro (BK7/Au). . .	70
Figura 29 – Resultados dos intervalos de confiança com o teste não paramétrico <i>bootstrap</i> para o $k = 5$, com o vidro BK7 e o metal ouro (BK7/Au). . .	71
Figura 30 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o polímero PC com o metal ouro PC/Au. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.	71
Figura 31 – (a) a (c) apresenta o conjunto de dados de treinamentos com os 99 sensorgramas dos 15 grupos para os descritores FWHT, FChT e DCT-4. (d) retrata o descritor TSD no qual apresenta o conjunto de dados de treinamento com 3 regimes permanentes.	72
Figura 32 – Resultados dos intervalos de confiança do teste não paramétrico <i>bootstrap</i> com o polímero PC e o metal ouro (PC/Au) para o $k = 3$	73
Figura 33 – Resultados dos intervalos de confiança do teste não paramétrico <i>bootstrap</i> com o polímero PC e o metal ouro (PC/Au) para o $k = 5$	74
Figura 34 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o polímero PMMA com o metal ouro PMMA/Au. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.	74
Figura 35 – Conjunto de dados de treinamentos com os 99 sensorgramas dos 15 grupos para a FWHT e FChT, DCT-4 e o TSD.	75
Figura 36 – Resultados dos intervalos de confiança do teste não paramétrico <i>bootstrap</i> para o $k = 3$, com o polímero PMMA e o metal ouro (PMMA/Au). . .	76
Figura 37 – Resultados dos intervalos de confiança do teste não paramétrico <i>bootstrap</i> para o $k = 5$, para o polímero PMMA e o metal ouro (PMMA/Au). . .	77
Figura 38 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o polímero TOPAS com o metal ouro TOPAS/Au. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.	77
Figura 39 – Conjunto de dados de treinamentos com os 99 sensorgramas dos 15 grupos para a FWHT e FChT, DCT-4 e o TSD.	78

Figura 40 – Resultados dos intervalos de confiança com o polímero TOPAS e o metal ouro (TOPAS/Au), em que utilizou o teste não paramétrico <i>bootstrap</i> para o $k = 3$	79
Figura 41 – Resultados dos intervalos de confiança com o polímero TOPAS e o metal ouro (TOPAS/Au), em que utilizou o teste não paramétrico <i>bootstrap</i> para o $k = 5$	80
Figura 42 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o vidro BK7 com o metal prata (BK7/Ag). De (a) até (g) o protocolo experimental consiste de 9 regimes permanentes. (h) até (l) 7 regimes permanentes. (m) até (o) 3 regimes permanentes.	81
Figura 43 – Conjunto de Dados de Treinamento em um gráfico com os 99 sensorgramas dos grupos para (a) FWHT, (b) FChT e (c) DCT-4. Conjunto de dados treinamento com 3 regimes permanentes em um gráfico 2D. . . .	82
Figura 44 – Resultados dos intervalos de confiança com o teste não paramétrico <i>bootstrap</i> para o $k = 3$, com o vidro BK7 e o metal prata (BK7/Ag). . .	83
Figura 45 – Resultados dos intervalos de confiança com o teste não paramétrico <i>bootstrap</i> para o $k = 5$, com o vidro BK7 e o metal prata (BK7/Ag). . .	83
Figura 46 – Grupos de sensorgramas representando 15 protocolos experimentais com quantidade de regimes permanentes diferentes utilizando o polímero e o metal PC/Ag. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.	84
Figura 47 – Conjunto de dados de treinamentos em um gráfico 3D com os 99 sensorgramas dos 15 grupos com os descritores (a) FWHT, (b) FChT e (c) DCT-4. (d) Conjunto de dados de treinamento com 3 regimes permanentes em um gráfico 2D.	85
Figura 48 – Resultados dos intervalos de confiança com com o polímero PC e o metal prata (PC/Ag) em que utilizou o teste não paramétrico <i>bootstrap</i> para o $k = 3$	86
Figura 49 – Resultados dos intervalos de confiança com com o polímero PC e o metal prata (PC/Ag) em que utilizou o teste não paramétrico <i>bootstrap</i> para o $k = 5$	86
Figura 50 – Grupos de sensorgramas no modo WIM representando 15 protocolos experimentais diferentes utilizando o polímero PMMA com o metal prata (PMMA/Ag). De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.	87

Figura 51 – Conjunto de Dados de Treinamento em um gráfico com os 99 sensorgramas dos grupos para (a) FWHT, (b) FChT e (c) DCT-4. (d) TSD conjunto de dados treinamento com 3 regimes permanentes em um gráfico 2D.	88
Figura 52 – Resultados dos intervalos de confiança do teste não paramétrico <i>bootstrap</i> para o $k = 3$, para o polímero PMMA e o metal prata (PMMA/Ag). . .	89
Figura 53 – Resultados dos intervalos de confiança do teste não paramétrico <i>bootstrap</i> para o $k = 5$, para o polímero PMMA e o metal prata (PMMA/Ag). . .	89
Figura 54 – Grupos de sensorgramas no modo WIM representando 15 protocolos experimentais diferentes utilizando o polímero TOPAS com o metal prata (TOPAS/Ag). De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.	90
Figura 55 – Conjunto de dados de treinamentos com os 99 sensorgramas dos 15 grupos para a FWHT, FChT, DCT-4 e TSD.	91
Figura 56 – Resultados dos intervalos de confiança do teste não paramétrico <i>bootstrap</i> para o $k = 3$, com o polímero TOPAS/Ag.	92
Figura 57 – Resultados dos intervalos de confiança do teste não paramétrico <i>bootstrap</i> para o $k = 5$, com o polímero TOPAS/Ag.	92
Figura 58 – Sensorgrama para o protocolo $H_2O \rightarrow \text{Etanol } 25\% \rightarrow H_2O \rightarrow \text{Etanol } 12.5\% \rightarrow H_2O \rightarrow \text{BSA-Monolayer} \rightarrow H_2O \rightarrow \text{Etanol } 25\% \rightarrow H_2O$ (azul). Resposta do coeficiente de regressão (preto). Destaque (seta vermelha) para a situação em que o coeficiente da regressão muda, indicando uma transição entre as substâncias analisadas.	93
Figura 59 – Teste de hipótese para verificar com 99.9% de confiança a presença de uma determinada substância. O nível de confiança define os limites superior e inferior do intervalo e o teste de hipótese verifica se na janela temporal estabelecida o coeficiente de regressão está contido. Em destaque tem-se que no intervalo 20 a 80 não foi possível confirmar com 99.9% de certeza que é uma substância.	94
Figura 60 – Histogramas da análise do resíduo do modelo de regressão. (a) Regime permanente. (b) Regime de transição.	95
Figura 61 – Identificação de padrão na resposta no modo WIM. (a) Identificação dos regimes permanentes com 99.9% de confiança, caracterizando as regiões que identificam as substâncias analisadas. (b) Identificação dos regimes de transição com 99.9% de confiança.	95
Figura 62 – Identificação dos regimes permanentes no protocolo experimental positivo e negativo da Leishmaniose com 99.9% de confiança, caracterizando as regiões que identificam as substâncias analisadas.	96

Figura 63 – Intervalo de confiança com duas amostras para o protocolo experimental da Leishmaniose. (a) Intervalo de confiança com o teste paramétrico <i>t-Student</i> . (b) Intervalo de confiança com o teste não-paramétrico <i>bootstrap</i> .	97
Figura 64 – Identificação dos regimes permanentes no protocolo experimental positivo e negativo da Dengue com 99.9% de confiança, caracterizando as regiões que identificam as substâncias analisadas.	98
Figura 65 – Intervalo de confiança com duas amostras para o protocolo experimental da Dengue. (a) Intervalo de confiança com o teste paramétrico <i>t-Student</i> . (b) Intervalo de confiança com o teste não-paramétrico <i>bootstrap</i>	99
Figura 66 – Diagrama de caso de uso para implementação da ferramenta SMART SPR.	114
Figura 67 – Diagrama de atividades para execução da ferramenta SMART SPR. . .	115
Figura 68 – Tela Inicial da ferramenta SMART SPR.	116
Figura 69 – Inserção do sensorgrama com a identificação dos regimes permanentes na ferramenta SMART SPR.	117
Figura 70 – Conjunto de dados de treinamento com o descritor TSD em que é separado os regimes permanentes para classificação do sensorgrama. . .	117
Figura 71 – Identificação dos regimes permanentes, como também a classificação das substâncias inserida e o intervalo de confiança para o vidro BK7 e o metal Ag.	118

Lista de tabelas

Tabela 1	– Algoritmos de inteligência artificial, aprendizagem de máquina e métodos numéricos aplicados ao uso de sensores e biossensores SPR.	27
Tabela 2	– Dados de um processo de fabricação de semicondutores para a construção de um modelo de regressão múltipla.	43
Tabela 3	– Protocolo experimental obtidos no modo AIM com o vidro BK7 e polímeros PMMA, TOPAS e PC utilizando o metal Au de 50 nm.	52
Tabela 4	– Protocolo experimental obtidos no modo WIM com o vidro BK7 e polímeros PMMA, TOPAS e PC com o metal Ag de 50 nm.	53
Tabela 5	– Resultados dos descritores em relação a quantidade de vizinhos com o algoritmo k -NN.	101

Lista de abreviaturas e siglas

3D-FDTD	<i>3D Finite-Difference Time-Domain</i>
AM	Aprendizagem de Máquina
AIM	<i>Angular Interrogation Mode</i>
AC	<i>Alternate Current</i>
ATR	<i>Attenuated Total Reflection</i>
BSA	<i>Bovine Serum Albumin</i>
DC	<i>Direct Current</i>
DCT	<i>Discrete Cosine Transform</i>
DChT	<i>Discrete Chebyshev Transform</i>
DT	<i>Decision Tree</i>
dWB	Decibel watt
FT	<i>Transform Fourier</i>
FWHT	<i>Fast Walsh-Hadamard Transform</i>
FEM	<i>Finite Element Method</i>
FWHM	<i>Full Width at Half Maximum</i>
HSA	<i>Human Serum Albumin</i>
IA	Inteligência Artificial
IC	Intervalo de Confiança
IDO	<i>Indoleamine 2,3-Dioxygenase</i>
IgG	- <i>Human Immunoglobulin G</i>
k-NN	<i>k-nearest neighbors</i>
LM	<i>Levenberg-Marquardt</i>
LSPR	<i>Localized Surface Plasmons Resonance</i>

LOO-CV	<i>Leave-One-Out-Cross Validation</i>
MLP	<i>Multilayer Perceptron</i>
NA	Neutravidin-Biotina
NB	<i>Naive Bayes</i>
NCs	Nanocristsais
NPs	Nanopartículas
PBS	<i>Phosphate Buffered Saline</i>
PPBIO	Prisma Polimérico para Aplicações BIOLógicas
PBS	<i>Phosphate Buffered Saline</i>
PSO	<i>Particle Swarm Optimization</i>
RP	<i>Recursive Partitioning</i>
RPS	Ressonância de Plasmons de Superfície
SOM	<i>Self-Organizing Maps</i>
SPR	<i>Surface Plasmon Resonance</i>
SP	<i>Surface Plasmons</i>
SVM	<i>Support Vector Machine</i>
TSD	<i>Temporal Sensorgram Descriptor</i>
WcBiM	<i>Waveguide-Coupled Bimetallic</i>
WHO	<i>World Health Organization</i>
WIM	<i>Wavelength Interrogation Mode</i>

Lista de símbolos

B	Número de replicações do <i>bootstrap</i>
d_i	Espessura do meio i da estrutura multicamada
ϵ_i	permissividade elétrica do meio i da estrutura multicamada
G	Largura da curva
H_0	Hipótese nula
H_A	Hipótese alternativa
k	Quantidade de vizinhos de mais próximos
k_{px}	Componente do vetor de onda da luz incidente
k_{sp}	Vetor de onda dos plasmons de superfície
λ	Ângulo
λ_R ou λ_{res}	Ângulo de ressonância
θ_R ou θ_{res}	Comprimento de onda de ressonância
$r(\lambda)$	Coefficiente de refletividade de Fresnel - WIM
$r(\theta)$	Coefficiente de refletividade de Fresnel - AIM
Γ	Assimetria
α	Alfa
σ	Valores de transição
ν	Valores de regimes permanentes
Φ	Distribuição empírica
μ	Média populacional
$\hat{\mu}$	Média da amostra

Sumário

1	INTRODUÇÃO	22
1.1	Objetivo Geral	24
1.2	Trabalhos Relacionados	25
1.3	Organização do Documento	28
2	REFERENCIAL TEÓRICO	29
2.1	Sensores de Ressonância de Plasmons de Superfície	29
2.2	Aspectos de Aprendizagem de Máquina	34
2.2.1	<i>k - Nearest Neighbor k-NN</i>	36
2.2.2	Teste de Hipótese	37
2.2.3	Validação <i>Bootstrap</i>	39
2.2.4	Regressão Linear Múltipla	41
2.3	Transformadas Discretas: Reconhecimento do Sinal	45
2.3.1	Transformada Discreta de Cosseno	45
2.3.2	Transformada Discreta de Chebyshev	46
2.3.3	Transformada Rápida de Walsh-Hadamard	47
3	METODOLOGIA	49
3.1	Etapa A - Definição do Ambiente e Dados para Inferência	50
3.2	Etapa B - Sensor em Ação	60
4	RESULTADOS	65
4.1	Descritor Temporal de Sensorgrama	65
4.2	Modo de Interrogação Angular (AIM)	67
4.3	Modo Interrogação Comprimento de Onda (WIM)	80
4.4	Identificação e Análise de Sensorgrama	93
4.5	Conclusão	99
5	CONCLUSÃO	102
	REFERÊNCIAS	104

APÊNDICES	112
APÊNDICE A – FERRAMENTA PARA GERAÇÃO DE RESPOSTAS BASEADO EM SENSORES SPR	113

1 Introdução

Os sensores são componentes específicos que tem como finalidade responder estímulos e convertê-los em sinais elétricos (LIU; BAIOCCHI, 2016). Sensores que detenham algum tipo de inteligência, podem ser denominados de sensores inteligentes. Esses sensores estão sendo cada mais usados, devido a sua praticidade e em relação aos custos baixos do seu uso. Os sensores inteligentes seguem a norma IEEE 1451 que tem como objetivo a definição de diretrizes de interfaces de comunicação comum para conectar transdutores a sistemas de microprocessadores, redes de controle e sistemas de instrumentação (LIU; BAIOCCHI, 2016) (SONG; LEE, 2010). Entre os sensores mais utilizados, estão os sensores magnéticos, capacitivos, indutivos e ópticos.

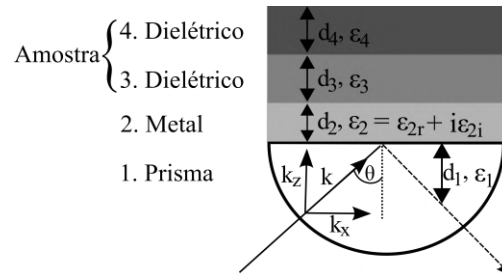
Os sensores ópticos são dispositivos que realizam a transformação de raios de luz em sinais elétricos. Seu funcionamento consiste na propagação da luz, por meio de um emissor e um receptor de luz (AHUJA; PARANDE, 2012). Esses sensores podem ser baseados em: elipsometria, espectroscópica e ressonância de plasmons de superfície (RPS, *Surface Plasmon Resonance* - SPR). A ressonância de plasmons de superfície é um fenômeno resultante da interação entre a luz e os elétrons. Esse fenômeno permite avaliar sensíveis mudanças de índice de refração ou na espessura em sua superfície com uma alta sensibilidade e em tempo real (JANG; CHAE; SHIN, 2015). Seu uso abrange diferentes campos, como no agronegócio, que é utilizado para detecção de pesticidas (THEPUDOM *et al.*, 2018) e *salmonella*, (LUKOSE *et al.*, 2018), já na indústria farmacêutica e alimentar é encontrado na adulteração de mel (ZAINUDDIN *et al.*, 2018), biomoléculas bianisotrópicas (WANG *et al.*, 2018), poluentes tóxicos na água (BORUAH; BISWAS, 2018) e glicose no sangue (LIDIYA *et al.*, 2019).

Nos sensores baseados em SPR é fundamental a especificação da estrutura e a configuração multicamada, em que a estrutura multicamada é construída a partir de finas camadas de materiais em escala nanométrica, formando uma interface metal/dielétrico. O filme metálico é depositado em um material opticamente transmissivo (substrato óptico) e o dielétrico são soluções gasosas, aquosas ou filmes (substância ou amostra) nos quais a interação molecular a ser detectada ocorre (OLIVEIRA, 2016).

Na Figura 1 é apresentada a estrutura multicamada constituída por 4 (quatro) camadas. A estrutura multicamada consiste no funcionamento do sensor SPR. Em incidir um feixe de luz, no qual, a oscilação dos elétrons de valência de um metal, os plasmons, possuem frequência e vetor de onda bem definidos. Essas oscilações são ondas que viajam ao longo da superfície metálica, no qual são denominadas de plasmons de superfície (SP - *Surface Plasmons*). Quando a superfície metálica é iluminada com uma luz em polarização tipo-p, ou seja, em um campo elétrico paralelo e campo magnético no plano-XZ, com

um ângulo de incidência (θ) bem definidos, ocorre a excitação dos SP. Essa excitação propaga-se paralelamente à interface metal/dielétrico e constituem ondas evanescentes na direção perpendicular à fronteira do metal e da substância em análise (amostra), sendo muito sensível a qualquer alteração nessa interface, tal como a adsorção de moléculas na superfície do metal (HOMOLA, 2008).

Figura 1 – Estrutura Multicamada composta pelo prisma com espessura d_1 e permissividade elétrica ϵ_1 , Metal com espessura d_2 e permissividade elétrica ϵ_2 e os dielétricos (amostra) com espessura d_3 e d_4 , permissividade elétrica ϵ_3 e ϵ_4 . Vetores, ângulo e plano de incidência indicados na Figura.



Fonte: Autoria Própria (2020).

Os sensores SPR tem sido amplamente utilizado nos últimos anos para projetar e desenvolver biossensores SPR, devido a oportunidade de se obter um biossensor sem a necessidade de marcadores, em que sua capacidade de monitorar a interação entre uma molécula imobilizada na superfície a aquisição de respostas rápidas, tornando-se uma ferramenta muito poderosa (PILIARIK; VAISOCHEROVÁ; HOMOLA, 2009). Os biossensores SPR exploram as características da estrutura multicamada de diferentes materiais para detectar interações moleculares. Sendo capazes de determinar o índice de refração de soluções analíticas e também detectar a presença e/ou evolução temporal de camadas de absorção.

Nos biossensores, a captura da luz refletida é possível por meio de fotodetectores. O sinal captado pelos fotodetectores passam por um processo de filtragem, tendo como objetivo minimizar ruídos e aumentar a faixa do sinal. Após esse processo, esse sinal pode ser então utilizado para gerar de forma gráfica a ressonância, denominada de curva SPR (OLIVEIRA *et al.*, 2015). A curva SPR apresenta um valor mínimo de refletividade e esse valor mínimo pode ser alterado a medida que diferentes analitos são inseridos. O monitoramento da alteração dos analitos permite alterar os parâmetros da curva SPR e o comportamento da curva. Esse monitoramento ao longo do tempo pode ser visualizado em um gráfico, denominado de sensorgrama.

O sensorgrama apresenta uma evolução temporal nos parâmetros que descrevem a curva, que permite visualizar alterações em eventos, como: ligações específicas e não específicas. O sensorgrama é importante para filtrar, manipular e entender essas ligações. Uma estratégia ainda pouco explorada para manipulação do sensorgrama é por meio de

técnicas de aprendizagem de máquina e descritores. A aprendizagem de máquina permite por meio de métodos de inferência realizar as manipulações de conhecimentos que possam gerar novas conclusões, e os descritores permite reconhecer e identificar de forma específica as características (MARSHLAND, 2015); (AMANATIADIS *et al.*, 2011).

Assim, tendo em vista a reprodutibilidade de um sensorgrama SPR para um protocolo experimental bem definido, é possível saber como a resposta de um biossensor SPR deverá ser, ou seja, qual resposta o sensor deverá fornecer para um ciclo experimental. Desse modo, sensores SPR inteligentes podem ser desenvolvidos para tomar decisões baseadas nas características particulares e comuns de um protocolo experimental, que possibilitem a classificação e a ordenação da resposta. Por meio de uma investigação do estado da arte de técnicas inteligentes em sensorgramas, foi encontrado apenas um trabalho que utiliza algoritmos de aprendizagem de máquina para classificação de respostas inteligentes, de forma mais específica, para classificação de inibidores. Não foram encontrados trabalhos que fazem uso de algoritmos de AM para descrever, classificar e identificar respostas SPR a partir de sensorgramas obtidos.

Nesse sentido, este trabalho visa utilizar diferentes técnicas de aprendizagem de máquina em sensores ópticos baseados na ressonância de plasmons de superfície para verificação das respostas do sensor, de forma mais específica, para descrever, classificar e identificar respostas SPR a partir de sensorgramas obtidos. Assim, com contribuições que permitam o avanço da área de estudo, como: identificação, classificação e análise de forma automática.

1.1 Objetivo Geral

O objetivo principal é investigar o uso algoritmos de aprendizagem de máquina para descrever, classificar, ordenar, identificar, atestar a qualidade em sensorgramas e realizar um teste automático adquirido em um sensor SPR.

Os objetivos específicos deste trabalho incluem:

1. Analisar o uso de inteligência artificial, aprendizagem de máquina e métodos numéricos em sensorgramas obtidos em sensores SPR;
2. Obter sensorgramas a partir do sensor SPR nos modos de interrogação angular e comprimento de onda;
3. Identificar as substâncias dos sensorgramas do sensor SPR com as transformadas discretas e aprendizagem de máquina;
4. Classificar e analisar o impacto dessas tarefas de aprendizagem de máquina na qualidade do sensor SPR;

5. Testar a qualidade dos descritores gerados a partir de intervalos de confiança não-paramétrico com *bootstrap*;
6. Teste automático nos protocolos experimentais da Dengue e Leishmaniose gerados a partir de intervalos de confiança paramétrico e não-paramétrico.
7. Desenvolver uma ferramenta baseada em algoritmos de aprendizagem de máquina para dar suporte aos estudos da área e como também fornecer de forma precisa as respostas sobre o fenômeno SPR.

1.2 Trabalhos Relacionados

Nesta seção são apresentados os trabalhos encontrados na literatura, que utilizam técnicas de inteligência artificial e métodos numéricos aplicados em sensorgramas obtidos com os sensores baseados ressonância de plasmons de superfície.

O trabalho de Yan *et al.* (2017) tem como objetivo detectar regiões de interesse e estimar parâmetros de forma precisa. Para as regiões de interesse foram utilizadas as técnicas de equalização de histograma, transformação afim, detecção de borda e a transformação de hough. A aplicação dessas técnicas permite identificar os pixels dentro das regiões de interesse para extração de dados. Já a estimação de parâmetros de associação e dissociação molecular, utilizou-se o Enxame de Partícula (PSO - *Particle Swarm Optimization*). Com as regiões identificadas o algoritmo enxame de partículas encontra os parâmetros de ligações moleculares. Desse modo, foi realizado uma comparação do algoritmo PSO e o algoritmo LM (*Levenberg-Marquardt*), no qual, foi observado que o algoritmo PSO tem um erro menor do que o algoritmo LM.

Tiwari *et al.* (2015) apresenta uma abordagem para identificar modelos de ligação bivalente em um sensor SPR. Para identificar as ligações bivalente, foi utilizado equações diferenciais não lineares acopladas. O modelo proposto ajusta diferentes concentrações de analitos, a fim de identificar e analisar inequivocamente o mecanismo bivalente. É apresentado também uma assinatura exclusiva para o modelo, em que a assinatura pode ser usada para distinguir o modelo de analito bivalente de outros modelos bifásicos. Dessa forma, o modelo proposto fornece um suporte para a identificação da análise bivalente.

No trabalho de Agrawal *et al.* (2017), é demonstrado o aprimoramento do campo próximo em torno de nanocristais (NCs), no qual utilizou-se a transformada de Fourier (FT - *Fourier Transform*) e o método dos elementos finitos (FEM - *Finite Element Method*) para observar e explicar a natureza dos fenômenos de acoplamento e para facilitar o acoplamento em matrizes auto-montadas, por meio da ressonância de plasmons de superfície localizado (LSPR - *Localized Surface Plasmon Resonance*). Com as abordagens teóricas e experimentais realizadas, os métodos descrevem a influência da força de acoplamento e da

relativa afinação entre as vibrações moleculares. O estudo apresentou resultados em vários fatores envolvidos na determinação do acoplamento de vibração molecular LSPR-LSPR e LSPR para materiais de óxido de metal e fornece uma base fundamental para o projeto de substratos sensores.

Lee *et al.* (2014) detecta amiloide- β 42, esse amilóide é um biomarcador da doença Alzheimer. Para a aquisição da resposta foi utilizado um chip de ressonância de plasmons de superfície bimetálica acoplado ao guia de ondas (WcBiM - *Waveguide-Coupled Bimetallic*) no modo de interrogação de intensidade. Dessa forma, foi aplicado a regressão linear para a curva de calibração, no qual é possível detectar o amilóide em uma concentração de 100 pg/ml a 2000 pg/ml, também foi possível distinguir o amiloide- β 42 do analito de controle (A- β 40) por meio de um receptor específico. Portanto, o resultado obtido do sensor apresenta que o A- β 42 pode ser detectado com sucesso para diagnósticos. Desse modo, o trabalho apresentou uma combinação do chip WcBiM com a modificação adequada da superfície no esquema de detecção de interrogação de intensidade, sendo capaz de fornecer a detecção de outros biomarcadores relacionados à doença.

O trabalho de Canoa *et al.* (2015) tem como objetivo comparar resultados de proteínas com o tipo de nanopartículas de óxido de metal, para medir quantitativamente a cinética e afinidades do fibrinogênio e da albumina sérica humana (HSA - *Human Serum Albumin*), em que utilizou-se o modelo de *Langmuir* para ajustar os dados. Por outro lado, o HSA é uma proteína de baixa afinidade e os dados se ajustam melhor ao modelo de mudança conformacional. O trabalho tem como destaque a imobilização direta de nanopartículas de óxido de metal para a avaliação quantitativa da ligação com proteínas injetadas. A caracterização dessas interações é importante devido as aplicações serem na área biomédica.

Zhang *et al.* (2018b) teve como objetivo utilizar tridimensionais de tempo-finito de diferença-finita (3D-FDTD - *3D Finite-Difference Time-Domain*) para simular os espectros de espalhamento da LSPR e calcular distribuições no campo elétrico para os nanotubos de ouro e prata durante o processo de hibridação. Os resultado do trabalho apresenta que o modelo fornece a base para o biosensoriamento inteligente e o controle genético. O trabalho também tem como contribuição um nanobiossensor para fornecer uma plataforma geral para a investigação quantitativa de vários problemas biológicos, como os do estudo de interações e cinética biomoleculares e computadores de DNA que funcionam tanto em *in vitro* como em *in vivo*.

No trabalho de Zhang *et al.* (2018a) teve como objetivo propor uma forma de identificar inibidores de indoleamina 2,3-dioxigenase (IDO - *Indoleamine 2,3-Dioxygenase*), com o uso de algoritmos de aprendizagem de máquina, em que utilizou os métodos de classificação Naive Bayes (NB - *Naive Bayes Classifier*) e particionamento recursivo (RP - *Recursive Partitioning*). Por meio de uma biblioteca de inibidores de IDO conhecidos

derivados de publicações recentes. Para identificar os inibidores de IDO, utilizou-se um sensor SPR para execução do experimento. O trabalho demonstrou que os métodos de aprendizagem de máquina são capazes de identificar inibidores de IDO, como também foi realizado simulações dinâmicas moleculares para demonstrar os modos de ligação dos inibidores IDO.

Chan (2009) apresenta uma patente para a redução de ruído no sensorgrama em sensores SPR. O filtro desenvolvido inclui um determinado comprimento ideal com base na inclinação do sinal do sensorgrama e em diferentes ocorrências de eventos. Dessa forma, é possível determinar o comprimento do filtro em tempo real durante a medição do sensorgrama para reduzir o ruído. Assim, o filtro dinâmico desenvolvido preserva o sinal, não apenas para reduzir ou eliminar ruídos, mas também para preservar os sinais SPR em processo ou evento bioquímico crítico, como o início de um evento de ligação entre analitos dentro de uma amostra. Dessa forma, o uso desse filtro também elimina a possibilidade de distorção de fase que outros filtros lineares ou não lineares podem introduzir. Na Tabela 1 são listados todos os trabalhos encontrados de forma resumida. Nessa tabela são apresentadas as referências, métodos aplicados, objetivos, sensor e a aplicação na ressonância de plasmons de superfície.

Tabela 1 – Algoritmos de inteligência artificial, aprendizagem de máquina e métodos numéricos aplicados ao uso de sensores e biossensores SPR.

Referências	Método	Objetivo	Sensor	Aplicação na Ressonância de Plasmons de Superfície
(YAN <i>et al.</i> , 2017)	Meta-heurística	O trabalho aborda o algoritmo PSO para detecção de parâmetros de ligação.	SPR	Obtenção de parâmetro
(TIWARI <i>et al.</i> , 2015)	Método-numérico	O artigo utilizou modelo de analito bivalente para identificar modelos de ligação,	SPR	Identificação e descrição
(AGRAWAL <i>et al.</i> , 2017)	Método-numérico	O trabalho apresenta nanocristais de In ₂ O ₃ (NCs) codificados por F e Sn cúbicos para acoplar à vibração C-H de ligantes de oleato ligados à superfície.	LSPR	Obtenção de parâmetros
(LEE <i>et al.</i> , 2014)	Aprendizagem de máquina supervisionada	O artigo apresenta a detecção de amiloide- β 42. Em que, foi realizado uma análise com a regressão linear para a curva de calibração.	SPR	Qualidade e calibração
(CANOA <i>et al.</i> , 2015)	Método numérico	O artigo aborda uma aplicação para medir quantitativamente a cinética e afinidades do fibrinogênio e da albumina sérica humana (HSA) para as nanopartículas (NPs) de TiO ₂ , CeO ₂ , Al ₂ O ₃ e ZnO imobilizados em um chip.	SPR	Modelagem numérica
(ZHANG <i>et al.</i> , 2018b)	Método numérico	O trabalho apresenta um nanobiosensor plasmons baseado em nanocubo de ouro e prata para detectar microRNA-21.	LSPR	Modelagem numérica
(ZHANG <i>et al.</i> , 2018a)	Aprendizagem de máquina supervisionada	O artigo propõe uma forma de identificar inibidores de indoleamina 2,3-dioxigenase (IDO).	SPR	Classificação de inibidores.
(CHAN, 2009)	Aprendizagem de máquina	Apresenta uma patente para a redução de ruído no sensorgrama em sensores SPR.	SPR	Filtro Inteligente

Fonte: Autoria Própria(2020).

Como é possível observar na Tabela 1, foram encontrados métodos de inteligência artificial, como meta-heurística, aprendizagem de máquina supervisionada e métodos numéricos. Em geral, os algoritmos de aprendizagem de máquina, meta-heurística e método numéricos foram utilizados para obtenção de parâmetros do designer do sensor e análise numérica para melhorar a qualidade de absorção dos analitos nos sensorgramas. Os métodos numéricos como o de Langmuir, método dos elementos finitos, equações diferenciais não lineares e tempo finito de diferença finita, torna-se um trabalho complexo e dependente de experimentos custosos e exigindo uma série de suposições. Além disso, tais métodos não são úteis para classificar sensorgramas, identificar substâncias e testar a qualidade da resposta de sensorgramas SPR.

Foi encontrado apenas um trabalho que utiliza algoritmos de aprendizagem de máquina para classificação de inibidores, esse trabalho utilizou os algoritmos de Naive Bayes e particionamento recursivo para identificar os inibidores. Dessa forma, o uso de algoritmos de aprendizagem de máquina e descritores pode auxiliar nas decisões baseadas em características particulares e comuns expressas pelo sensorgrama, para melhorar a precisão e eficiência, não exigir uma série de suposições e como também não depender de experimentos custosos. Dessa forma, neste trabalho é apresentada as direções futuras relacionadas a este contexto de pesquisa.

1.3 Organização do Documento

Este trabalho encontra-se organizado em capítulos. Na introdução foram apresentados a contextualização, objetivos e os trabalhos encontrados na literatura e os demais são descritos a seguir.

- Capítulo 2 é apresentados as definições utilizadas para este trabalho dos seguintes assuntos: ressonância de plasmons de superfície, descritores e aprendizagem de máquina;
- Capítulo 3 é apresentado a metodologia do trabalho, em que aborda os sensores SPR, protocolos experimentais, descritores e as técnicas de aprendizagem de máquina empregada para o trabalho;
- Capítulo 4 são apresentados os resultados obtidos com o uso dos algoritmos de aprendizagem de máquina e descritores para os modos de interrogação angular e comprimento de onda, como também são discutidos;
- Capítulo 5 são apresentados a conclusão, fornece uma visão sobre as limitações e os trabalhos futuros deste trabalho e potenciais melhorias.

2 Referencial Teórico

Este capítulo é destinado à apresentação do referencial teórico e está organizado da seguinte forma. Na Seção 2.1 são apresentados conceitos da ressonância de *plasmons* de superfície e a aquisição do sinal. Na Seção 2.2 são apresentados os conteúdos que fazem parte do escopo da aprendizagem de máquina, descrevendo cada algoritmo utilizado para esse trabalho. E ao final, na Seção 2.3 são apresentados os conceitos de descritores em detalhes extremos para estabelecer uma boa base para entender como é representado um descritor.

2.1 Sensores de Ressonância de Plasmons de Superfície

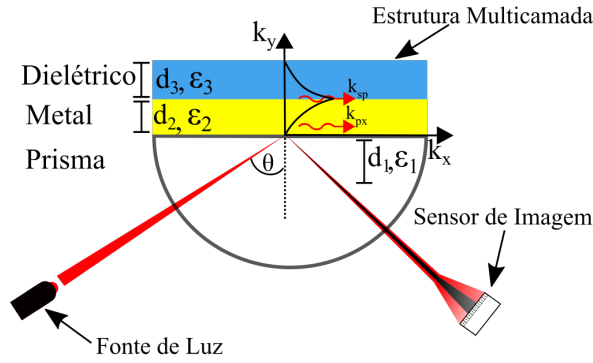
A ressonância de plasmons de superfície (SPR - *Surface Plasmon Resonance*) é resultante da interação entre uma onda eletromagnética e os plasmons de superfície, no qual consiste em uma oscilação de densidade de cargas (plasmons) presente em uma interface metal/dielétrico, sendo que no metal de espessura fina há a presença de elétrons livres distribuídos ao longo da superfície do metal. O dielétrico, por sua vez, faz oposição à passagem da corrente elétrica (SCHASFOORT, 2017).

As primeiras observações documentadas dos plasmons de superfície ocorreram em 1902, quando Wood iluminou com uma luz polarizada em uma grade de difração metálica e observou faixas escuras estreitas no espectro da luz difratada, em que se referiu como anomalias. Em 1941, Fano concluiu que as anomalias proposta por Wood, estavam associadas à excitação das ondas eletromagnéticas na superfície da rede de difração. Já em 1958, Turbadar ao iluminar com uma luz polarizado em uma camada fina de um metal/dielétrico, observou uma grande queda na refletividade, mas não associou este efeito a ressonância de plasmons de superfície. Em 1968, Kretschmann e Raether e Otto demonstraram o efeito SPR pelo processo da atenuação da reflexão interna total da luz (ATR - *Attenuated Total Reflection*), incidindo um feixe luminoso em um prisma acoplado a um metal, dessa forma, foi explicado os resultados obtidos por Turbadar (SILVA, 2017). Em 1983, foi apresentado o uso da ressonância de plasmons de superfície na monitoração de interação biomolecular por Liedberg, Nylander e Lundstro. O trabalho abordou a absorção da imunoglobulina humana G (IgG - *Human Immunoglobulin G*) em um metal, usando como ensaio para detectar IgG anti-humana em uma solução aquosa (PATIL *et al.*, 2019).

Uma das formas mais comuns para a excitação dos plasmons de superfície (PS) é com a configuração de acoplamento de prismas, proposta por Otto (1968) e Kretschmann e Raether (1968). A configuração de acoplamento de prismas, tem como resultado a transferência de energia dos fótons incidentes para os plasmons de superfície, reduzindo a

energia da onda refletida. Para demonstrar a existência da ressonância de plasmons de superfície é usado o processo de reflexão total atenuada da luz (ATR). A ATR consiste em incidir um feixe luminoso em um prisma acoplado a uma superfície metálica. Na geometria de Kretschmann e Raether (1968) no método ATR, a camada metálica é depositada diretamente sobre o prisma com espessura d_1 e permissividade elétrica ϵ_1 , em seguida o dielétrico é depositado acima da camada metálica com permissividade elétrica ϵ_2 (BOCKOVÁ *et al.*, 2019). Na Figura 2 é apresentada a configuração de Kretschmann e Raether (1968) com a reflexão total atenuada, com a indicação da onda evanescente gerada pela interação do vetor de onda dos plasmons e dos fótons, na direção perpendicular.

Figura 2 – Excitação dos plasmons de superfície na geometria de Kretschmann e Raether (1968) no método de reflexão total atenuada. d_1 é a espessura, ϵ_1 a permissividade elétrica do prisma, ϵ_2 permissividade elétrica do metal, d_2 espessura do metal, d_3 é a espessura e ϵ_3 permissividade elétrica do dielétrico, k_{px} vetores de onda da luz incidente e k_{sp} vetores plasmons de superfície.



Fonte: Autoria Própria (2020).

Na geometria de Kretschmann e Raether (1968), quando uma onda de luz se propaga através do prisma e no filme metálico, uma parte da luz é refletida de volta para o prisma e a outra parte se propaga no metal na forma de uma onda eletromagnética não homogênea (PILIARIK; VAISOCHEROVÁ; HOMOLA, 2009). Essa onda é fruto da interação da luz com os plasmons de superfície, ela decai exponencialmente na direção perpendicular à interface prisma/metal e consequentemente metal/dielétrico, não irradiando para fora dessas interfaces, assim caracteriza-se como uma onda evanescente.

A constante de propagação dos plasmons de superfície ao longo da superfície do metal denominada de k_{sp} , a qual é obtida através das equações de Maxwell levando em consideração as condições de contorno da estrutura multicamada (ADAM *et al.*, 2012). O valor de k_{sp} é uma aproximação para a oscilação dos plasmons de superfície, expresso pela Equação (2.1), em que, ϵ_m é a parte real da permissividade elétrica do metal, ϵ_d

permissividade elétrica do dielétrico e λ é o comprimento de onda da luz.

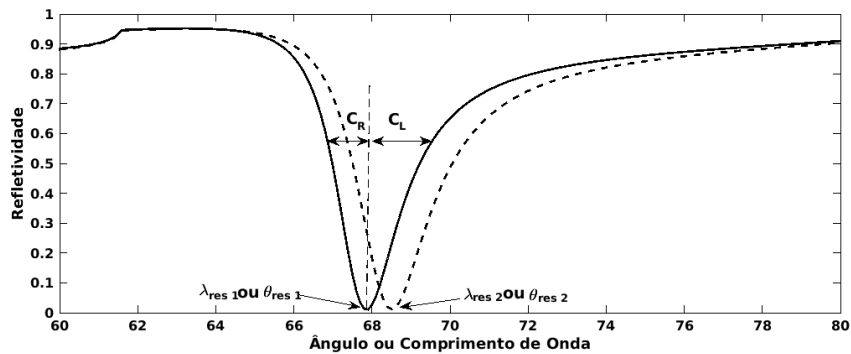
$$\underbrace{\frac{2\pi}{\lambda} \sqrt{\frac{\epsilon_m \epsilon_d}{\epsilon_m + \epsilon_d}}}_{k_{sp}} \cong \underbrace{\frac{2\pi}{\lambda} \sqrt{\epsilon_p} \sin(\theta_1)}_{k_{px}} \quad (2.1)$$

Para que ocorra a ressonância do PS é necessário o acoplamento do componente da luz incidente k_{px} na direção de k_{sp} , expresso pela Equação (2.1). Ocasionalmente uma diminuição na intensidade da luz refletida.

A ressonância é obtida variando as condições de acoplamento entre um feixe de luz e os plasmons de superfície. Dentre as formas para se obter a condição de ressonância é possível desenvolver sensores com dois modos de operação básicos, são eles: o modo de interrogação espectral (WIM *Wavelength Interrogation Mode*) constituído por um ângulo de incidência constante, dessa forma o espectro da fonte luminosa é empregado para medir a refletividade em função do comprimento de onda. E o modo de interrogação angular (AIM - *Angular Interrogation Mode*) constituído por um comprimento de onda fixo, variando o ângulo de incidência, até que a condição de ressonância seja alcançada.

Na Figura 3, é apresentado o valor da ressonância no modo AIM observado como θ e no modo WIM observado como λ . Com o valor mínimo da curva SPR, pode-se determinar o parâmetro FWHM (*Full Width at Half Maximum*), também conhecido como ponto médio entre o valor máximo e mínimo da curva, local onde os valores C_R e C_L são estimados. Os parâmetros C_R e C_L representam a distância em graus no modo AIM ou em nanômetros no modo WIM, entre o mínimo da curva e as bordas a esquerda e direita. A largura da curva é obtida através da distância entre a borda esquerda e direita, pela Equação $G = C_R + C_L$ e a assimetria da curva é obtida através da razão entre as duas metades da curva em relação ao ponto mínimo, pela Equação $\Gamma = C_R/C_L$.

Figura 3 – Exemplo de curva SPR, dada pelo gráfico refletividade vs. ângulo (modo AIM) ou refletividade vs. comprimento de onda (modo WIM). O valor mínimo indica a posição de ressonância: λ_{res1} e λ_{res2} no modo WIM e θ_{res1} e θ_{res2} . Os valores C_R e C_L são parâmetros para calcular a largura e a assimetria da curva SPR.



Fonte: Autoria Própria (2020).

Para a aquisição do valor mínimo de ressonância são necessárias duas medições para localizar o ângulo ou comprimento de onda, definidos pela Equação 2.2. A Equação 2.2 define a razão de intensidade, ou seja, o valor de $I_{dry}(\theta, \lambda)$ é o sinal de refletividade, obtido a partir da célula seca e $I_{ana}(\theta, \lambda)$ é registrado a presença do analito no sensor. Esse procedimento elimina as não uniformidades espaciais ou espectrais no feixe detectado, causadas pelas características da fonte de luz ou receptor óptico.

$$I_{ref} = \frac{I_{ana}(\theta, \lambda)}{I_{dry}(\theta, \lambda)} \quad (2.2)$$

O sensor de imagem é importante para aquisição da resposta. Esse sensor é composto por uma matriz fotodiodos unidimensional (1D) ou bidimensional (2D). A captura da posição de ressonância é adquirida por meio da imagem que detecta todas as regiões em uma matriz bidimensional e é processada logicamente para corresponder a matriz de fotodiodo e as áreas sensíveis. Dessa forma, cada matriz, contém sua própria curva SPR. A Equação 2.3 apresenta a imagem 2D com os dados brutos em uma matriz $I_{M \times N}$ (FERREIRA *et al.*, 2018).

$$I_{M \times N}(t) = \begin{bmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1j}(t) \\ x_{21}(t) & x_{22}(t) & \cdots & x_{2j}(t) \\ \vdots & \vdots & & \vdots \\ x_{i1}(t) & x_{i2}(t) & \cdots & x_{ij}(t) \end{bmatrix} \quad (2.3)$$

Uma técnica denominada de *multispot* voltada para imagem bidimensional, permite utilizar apenas um subconjunto de linhas para gerar a curva SPR $I_{M \times P}$. A Equação 2.4 apresenta a técnica *multispot*, no qual é composta por diferentes áreas sensíveis A^S , cada área corresponde uma porção física da superfície sensível do sensor SPR.

$$I_{M \times P}(t) = \begin{bmatrix} \hat{x}_{11}(t) & \hat{x}_{12}(t) & \cdots & \hat{x}_{1P}(t) \\ \hat{x}_{21}(t) & \hat{x}_{22}(t) & \cdots & \hat{x}_{2P}(t) \\ \hat{x}_{31}(t) & \hat{x}_{32}(t) & \cdots & \hat{x}_{3P}(t) \\ \hat{x}_{41}(t) & \hat{x}_{42}(t) & \cdots & \hat{x}_{4P}(t) \\ \hat{x}_{51}(t) & \hat{x}_{52}(t) & \cdots & \hat{x}_{5P}(t) \\ \hat{x}_{61}(t) & \hat{x}_{62}(t) & \cdots & \hat{x}_{6P}(t) \\ \vdots & \vdots & & \vdots \\ \hat{x}_{M-11}(t) & \hat{x}_{M-12}(t) & \cdots & \hat{x}_{M-1P}(t) \\ \hat{x}_{M1}(t) & \hat{x}_{M2}(t) & \cdots & \hat{x}_{MP}(t) \end{bmatrix} \begin{matrix} A^1(t) \\ A^2(t) \\ A^3(t) \\ \vdots \\ A^S(t) \end{matrix} \quad (2.4)$$

As áreas sensíveis formam os pontos sensíveis $[A^1(t), A^2(t), \dots, A^S(t)]$. Esses pontos indicam a quantidade de substâncias que podem ser monitoradas simultaneamente pelo sensor SPR. A maneira de agrupar as linhas para compor a área A^i pode ser expressa pela Equação 2.5. A Equação 2.5 representa a imagem média das linhas, onde N representa o

total de *pixels* no intervalo de $[a, b]$. O uso dessa técnica melhora a relação sinal-ruído e a resposta dinâmica do instrumento SPR (BOECKER *et al.*, 2008);(Oliveira *et al.*, 2015).

$$A^i(t) = \frac{1}{N} \sum \begin{bmatrix} x_{a1}(t) & x_{a2}(t) & \cdots & x_{aP}(t) \\ \vdots & \vdots & & \vdots \\ x_{b1}(t) & x_{b2}(t) & \cdots & x_{bP}(t) \end{bmatrix} \quad (2.5)$$

É necessário de algoritmos de rastreamento rápido para extrair o valor mínimo da curva SPR em tempo real. Alguns desses algoritmos utilizam os métodos de caça o mínimo (*Minimum Hunt*), centroide (NENNINGER; PILIARIK; HOMOLA, 2002) e interpolação polinomial (OWEGA; POITRAS, 2007). O algoritmo de caça o mínimo, consiste em um algoritmo simples para buscar o valor mínimo da curva SPR. Já a busca pelo método centroide, consiste em aplicar em um limite predefinido ou dinamicamente conhecido como linha de base. Para uma curva SPR com N pontos, o valor é calculado para os pontos com valores de intensidade abaixo da linha base (L_b), conforme a Equação 2.6, que p_k representa o valor do ponto k na curva SPR.

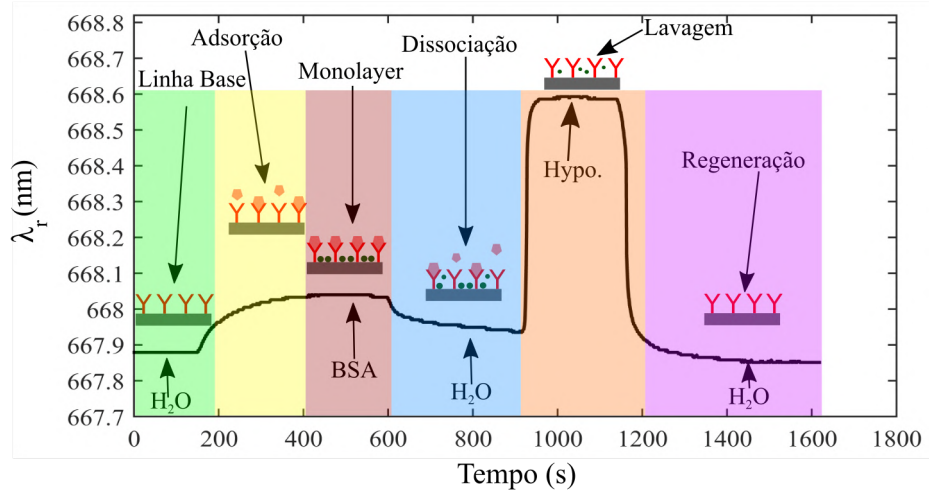
$$C(t) = \frac{\sum_{k=1}^N (p_k(t) - L_b)k}{\sum_{k=1}^N (p_k(t) - L_b)} \quad (2.6)$$

Para encontrar o valor mínimo, com o algoritmo de interpolação polinomial, é utilizado a Equação 2.7. A Equação 2.7 encontra o mínimo, em que consiste em ajustar toda a curva, ou uma porção próxima à região mínima, por um polinômio de grau n . Esse algoritmo é calculado através da primeira derivada.

$$\frac{d}{d(\theta, \lambda)} P(\theta, \lambda)_{\theta, \lambda = \theta_{ref}, \lambda_{ref}} = 0 \quad (2.7)$$

A ressonância de plamons de superfície é um excelente método para monitorar as alterações no índice de refração. Quando o índice de refração muda, o ângulo ou comprimento de onda mínimo também sofrerá uma mudança. Dessa forma, a ressonância de plamons de superfície pode monitorar a mudança do analito ao longo do tempo a partir da variação do ângulo ou comprimento de onda mínimo, em que pode ser denominado de sensorgrama. Na Figura 4 é apresentado um sensorgrama variando a condição de ressonância ao longo do tempo, em que, a cada substância inserida ocorre uma mudança na condição de ressonância. Dessa forma, as fases de um sensorgrama compreende-se em linha base, adsorção, *monolayer*, dissociação, lavagem e regeneração.

Figura 4 – Exemplo de um ciclo experimental em um sensorgrama no modo WIM, em que foram inserido as substâncias $H_2O \rightarrow BSA \rightarrow H_2O \rightarrow Hypo. \rightarrow H_2O$. Destaque para as fases do sensorgrama: linha base, absorção, monolayer, dissociação, lavagem e regeneração.



Fonte: Autoria Própria (2020).

Na fase de linha base é calibrado o sensor para uma substância conhecida. Na fase de adsorção ocorre o início da injeção da solução. Na fase de *monolayer* é um regime permanente, em que o valor de interesse é constante. Na fase de dissociação ocorre a remoção das moléculas em suspensão que não aderiram à superfície metálica. Na fase de lavagem é injetado uma substância para limpar a superfície e remover as moléculas fixadas na fase de adsorção. E a última fase, de regeneração, retorna a condição inicial do ciclo experimental (OLIVEIRA, 2016). Assim, o protocolo experimental de um sensorgrama contém informações sobre a interação entre as moléculas em termos de quão rápida ocorre as fases. Desse modo, ao longo das fases podem ser detectados padrões, verificado a qualidade do sensor, se pode ser reprodutivo ou não.

Para se ter uma melhor confiabilidade nas repostas fornecidas pelos sensorgramas, com um processo automático, podem ser utilizados algoritmos de aprendizado de máquina. Esses algoritmos visam aprender por meio de experiências passadas e assim podem prever padrões específicos. Na próxima subseção é apresentado alguns conceitos principais dessa área.

2.2 Aspectos de Aprendizagem de Máquina

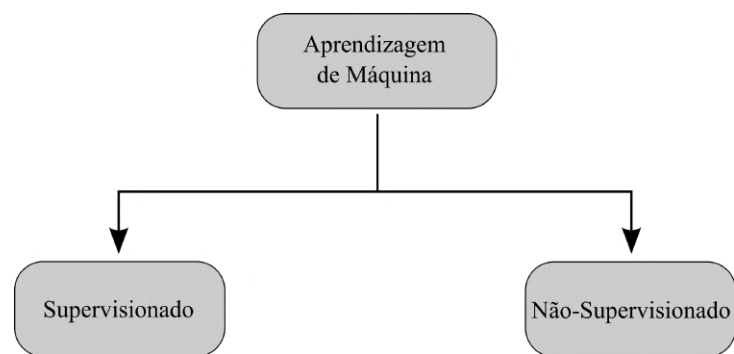
A Aprendizagem de Máquina (AM - *Machine Learning*) é uma área relacionada à Inteligência Artificial (IA - *Artificial Intelligence*), que fornece aos computadores a capacidade de aprender para descobrir tendências e padrões específicos (Basavaraju *et al.*, 2020). A AM permite estimar uma densidade de probabilidade sobre um subconjunto de

dados, denominado conjunto de dados de treinamento. O conjunto de dados de treinamento corresponde aos atributos que podem ser chamados de variáveis, que representam o campo da sua identificação, por exemplo, ângulo mínimo, comprimento de onda mínimo e tempo. Os modelos de AM são capazes de melhorar o desempenho na realização de determinadas tarefas por meio de experiências (SANTOS, 2012).

Um modelo da aprendizagem de máquina é o não-determinístico, esse modelo é capaz de melhorar o desempenho na realização de determinadas tarefas e apresentar comportamentos em diferentes execuções. Esse modelo pode ser dividido em: preditivos e descritivos. O modelo preditivo é uma forma de encontrar uma função que possa ser utilizada para prever um valor que caracterize um novo exemplo, com base nos seus valores de atributos de entrada. Já o modelo descritivo, explora um conjunto de dados, não fazendo o uso do atributo de saída (SUYKENS, 2014).

A AM utiliza o princípio de inferência denominado indução. A indução permite que conclusões genéricas sejam obtidas a partir de um conjunto de dados, ou seja, aprende a induzir uma função ou hipótese, sendo capaz de resolver um problema (SUYKENS, 2014). A aprendizagem pode ser dividida em duas categorias, como: supervisionada, não-supervisionada, como apresenta na Figura 5.

Figura 5 – Estrutura da aprendizagem de máquina dividida em: Supervisionada e Não-Supervisionada.



Fonte: Autoria Própria (2020).

A aprendizagem supervisionada, é um modelo preditivo que tem como objetivo aprender uma função com base no conjunto de dados de treinamento para uma saída desejada. O conjunto de dado de treinamento deve possuir um dado de destino. Esse dado de destino é a resposta em que o algoritmo deve produzir. Cada conjunto de dados de treinamento conhece a sua saída desejada, no qual, durante o treinamento, o algoritmo ajusta o indutor, como também a informação de sua saída associada (MARSHLAND, 2015). Existem dois tipos de problemas de aprendizagem supervisionado que é denominado de classificação e regressão. A classificação, tem como objetivo prever um exemplo com base em uma lista predefinida de possibilidades. Já a regressão o objetivo é prever valores com

o relacionamento entre uma variável dependente e uma ou mais variáveis independentes.

É possível utilizar a experiência adquirida anteriormente através do processo de treinamento para determinar a classificação de um possível rótulo. Existem diversos algoritmos para a aprendizagem supervisionada, é possível destacar alguns, como: k vizinhos mais próximos (k -NN - k - *Nearest Neighbor*) (CHOMBOON *et al.*, 2015), árvore de decisão (DT - *Decision Tree*) (MUSLIM; NURZAHPUTRA; PRASETIYO, 2018), máquina de vetores de suporte (SVM - *Support Vector Machine*) (CRISTIANINI; SHAWE-TAYLOR *et al.*, 2000) e *perceptron* multicamadas (MLP - *Multilayer Perceptron*) (HAYKIN; NETWORK, 2004).

A aprendizagem não-supervisionada contém um conjunto de dados de treinamento porém não está associada a uma saída, dessa modo é um processo de modelo descritivo, tendo como objetivo encontrar grupos de objetos semelhantes ao conjunto de dados de treinamento. Dessa maneira, o objetivo é agrupar exemplos através da identificação de padrões ou tendências, gerando assim os *clusters* distintos (SUYKENS, 2014);(SANTOS, 2012).

O processo de agrupamento é realizado através de duas abordagens, que são definidas como: distância ou correlação. No agrupamento por distância, os exemplos podem ser agrupados com outros mais próximos, formando assim um *cluster*. Já a abordagem de agrupamento por correlação, verifica se existe uma semelhança entre os *clusters*. É possível destacar a correlação de Pearson como uma medida de correlação usada na aprendizagem não-supervisionada. Dessa forma, é possível dizer que os exemplos formados no mesmo grupo podem apresentar uma maior semelhança entre si do que exemplos em grupos diferentes.

Existem diversos algoritmos para a aprendizagem não-supervisionada, podem ser destacados alguns, como: k -Médias (k -*Means*) (WANG; SU, 2011), agrupamento hierárquico (TOURETZKY; MOZER; HASSELMO, 1996) e redes neurais do tipo mapa auto-organizáveis (SOM - *Self-Organizing Maps*) (MILJKOVIC, 2017).

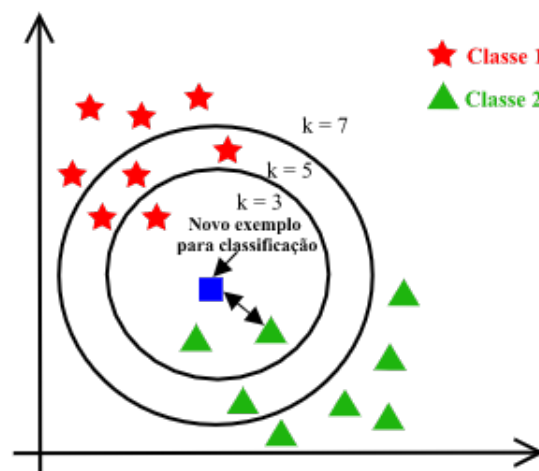
2.2.1 k - *Nearest Neighbor* k -NN

Para este trabalho será utilizado o algoritmo k -NN, por possuir algumas vantagens, como: simplicidade na sua implementação, flexibilidade em escolher a distância e contém um ótimo desempenho para classificação por simplesmente identificar o novo exemplo a partir do seu conjunto de dados de treinamento. Esse algoritmo é confiável para dados ruidosos de treinamento e é eficiente se os dados de treinamento forem grandes (IMANDOUST; BOLANDRAFTAR, 2013).

O algoritmo k -NN é uma extensão do 1-NN, ao contrário do 1-NN que faz uso de apenas um vizinho mais próximo faz uso de k vizinhos. O algoritmo k -NN classifica uma

amostra de teste não marcada com base na maioria das amostras semelhantes entre os vizinhos mais próximo a partir de uma distância. Essa distância entre a amostra de teste e cada uma das amostras de dados de treinamento são determinadas por uma medida específica. Na Figura 6 apresenta-se um exemplo do k -NN, em que contém um conjunto de dados de treinamento em duas classes, a primeira classe "estrela vermelha" e a segunda classe "triângulo verde". A amostra de teste é representada em quadrado azul. Essas amostras são colocadas em um espaço bidimensional (PRASATH *et al.*, 2017).

Figura 6 – Exemplo do Algoritmo k -NN.



Fonte: Autoria Própria (2020).

Para classificar a amostra de teste que pertence à classe "estrela vermelho" ou "triângulo verde", o k -NN adota uma função de distância para encontrar os k vizinhos mais próximos da amostra de teste. Para esse caso, quando o $k = 3$, o exemplo de teste é classificado na segunda classe como "triângulo verde", pois existe dois triângulos verdes e apenas uma estrela vermelha dentro do círculo interno, porém quando o $k = 5$ é classificado como classe 1 "estrela vermelha", pois existe três estrelas vermelhas e um triângulo verde. O algoritmo k -NN é simples, sendo um algoritmo altamente eficiente e eficaz para resolver vários problemas de classificação na vida real (PRASATH *et al.*, 2017).

2.2.2 Teste de Hipótese

Os métodos estatísticos são usados para verificar suposições, especulações ou afirmações referente à uma distribuição de probabilidade em populações. Com objetivo de demonstrar ou verificar uma suposição aceitável, pode ser utilizado hipóteses. As hipóteses são elaboradas sobre os parâmetros de uma certa distribuição com uma ou mais variáveis aleatórias com alternativas a ser testadas, em que essas alternativas incluem todos os valores possíveis que um teste estatístico pode assumir. Portanto, um teste para a hipótese será sempre com base em resultados amostrais, que pode ser aceita ou rejeitada. Desse modo, somente será rejeitada se o resultado da amostra for claramente improvável de

ocorrer quando a hipótese for verdadeira. As decisões a serem tomadas em relação aos resultados são chamadas de testes de hipótese (JUNIOR, 2017).

Os testes de hipótese são métodos de inferências que auxiliam na quantificação da tomada de decisão, baseado em uma amostra. O teste de hipótese têm como objetivo fornecer uma metodologia que permite verificar se os dados amostrais trazem evidências que apoiem ou não uma hipótese formulada. Os testes de hipótese podem apresentar duas hipóteses: A hipótese nula, H_0 e a alternativa, H_A . A hipótese nula considera que não existe relação entre dois fenômenos medidos, já a hipótese alternativa considera que exista correlação entre os fenômenos avaliados (JAMES DANIELA WITTEN, 2013).

Os testes de hipóteses com a média de uma população podem conter características importantes para os testes de hipótese unilateral e bilateral. O teste de hipótese unilateral contém uma desigualdade para a direita que envolve o sinal de maior ($>$) e a esquerda para o sinal de menor ($<$). Na equação 2.8 é apresentado as hipóteses unilaterais, em que define a média μ_2 para uma constante conhecida para os testes unilaterais, no qual a região de rejeição está em uma das caudas da distribuição.

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} \text{ ou } \begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases} \quad (2.8)$$

Já o teste de hipótese bilateral envolve uma não igualdade ($\neq$). A equação 2.9 são apresentadas as hipóteses para o teste bilateral, que define a média μ_2 para uma constância conhecida. A região de rejeição se distribui igualmente em ambas as caudas.

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \quad (2.9)$$

Desse modo, o teste de hipótese calcula a região crítica, definida pelo valor amostral mínimo ou máximo, no qual pode ser observado se será aceita ou refutada a hipótese nula, de acordo com o intervalo de confiança. Os intervalos de confiança expressam a precisão para estimar e estabelecer limites com certa probabilidade. Esses limites determinam um intervalo, que deverá estar o verdadeiro valor do parâmetro. Logo, a estimação por intervalo consiste na fixação de dois valores seja a probabilidade de que no intervalo, contenha o verdadeiro valor do parâmetro. Se a hipótese H_0 for rejeitada para valores da estatística de teste que caíam em uma das duas regiões críticas é denominado de teste bilateral. No entanto, se a hipótese H_0 for rejeitada para valores da estatística de teste que caíam em uma região crítica específica, esse teste é denominado de unilateral (JUNIOR, 2017).

Teste *t-Student* para Duas Distribuições Normais, Variâncias Desconhecidas e Diferença de Médias

Para este trabalho será utilizado o teste de hipótese para a diferença entre médias e variâncias desconhecidas, esse teste é utilizado com base na comparação de duas amostras extraídas. Dessa forma, o teste determina se as duas amostras são diferentes. Assim, as médias $\mu_1 - \mu_2$ são duas distribuições normais, variâncias σ_1 e σ_2 sejam desconhecidas e a hipótese $H_0 : \mu_1 - \mu_2 = \Delta_0$ for verdadeira, então é utilizado o teste na Equação 2.10, no qual, $\overline{X}_1$ e $\overline{X}_2$ são as médias conhecidas, S_1^2 e S_2^2 são as variâncias da amostras e n_1 e n_2 são os tamanhos das amostras.

$$T_0 = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (2.10)$$

São distribuídas normalmente o teste de hipótese com duas amostras com graus de liberdades, definida pela Equação 2.11. Caso v não for um número inteiro, é arredondado para o menor inteiro mais próximo.

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1/n_1)^2}{n_1-1} + \frac{(s_2/n_2)^2}{n_2-2}} \quad (2.11)$$

Em algumas situações não é razoável supor que as seguintes variâncias são iguais: $\sigma_1^2 = \sigma_2^2$. Quando essas situações não forem iguais, pode ser encontrado um intervalo de confiança de $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$. Na Equação 2.12 é apresentado o intervalo de confiança, composto pelas médias e variâncias. As médias e variâncias $\overline{x}_1, \overline{x}_2, s_1^2$ e s_2^2 de duas amostras aleatórias de tamanhos n_1 e n_2 , proveniente de duas populações normais independentes, com variâncias desconhecidas e desiguais. No qual v são os graus de liberdades em uma estatística t .

$$\overline{x}_1 - \overline{x}_2 - t_{\alpha/2,v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \overline{x}_1 - \overline{x}_2 + t_{\alpha/2,v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (2.12)$$

2.2.3 Validação *Bootstrap*

O *Bootstrap* é uma técnica estatística em que estima valores para parâmetros populacionais, em que será utilizada para esse trabalho. Essa técnica possibilita uma maior flexibilidade, rapidez e menos suposições na determinação de estimativas; além de poder ser aplicado a qualquer tipo de estimador, por mais que o estimador seja complicado possibilita calcular intervalos de confiança. A amostra é denominada de distribuição empírica Φ . Cada amostra denominada de x_i tem a mesma probabilidade de $1/n$ de ser selecionado quando uma nova amostra é gerada a partir de Φ , isto é, cada x_i tem a mesma

probabilidade de ser selecionado. No entanto, isso não significa que todos os elementos da amostra devem ou serão selecionados (MARTINEZ; MARTINEZ, 2016).

Dado uma estatística $s(\cdot)$, a validação *bootstrap* baseia-se na teoria da probabilidade para possibilitar inferências sobre Φ utilizando a estatística $s(\cdot)$ e o conjunto de amostras independentes gerados. O *bootstrap* fornece uma estimativa de distribuição de Φ . Para alcançar essa estimativa é necessário seguir algumas etapas, como segue:

1. Dada uma amostra de uma população $\Phi = \{x_1, x_2, \dots, x_n\}$
2. Substituição da amostra original $\Phi^{*b} = \{x_1^{*b}, x_2^{*b}, \dots, x_n^{*b}\}$
3. Calcula a mesma estatística usando a amostra *bootstrap* da etapa 2 para obter Φ^{*b} .
4. Repetir as etapas 2 e 3, B vezes.
5. Use esta estimativa da distribuição Φ^{*b} para obter o intervalo de confiança.

Para calcular a estimativa $\theta = s(\Phi)$ de um parâmetro θ do conjunto Φ . As réplicas do *bootstrap* $s(\cdot)$ são obtidas em cada amostra, desse modo é possível gerar o conjunto $\{\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*b}\}$, em que $\hat{\theta}^{*b} = s(\Phi^{*b})$, com $b = 1, 2, \dots, B$. A estimativa *bootstrap* do erro padronizado $\hat{s}e$, é equivalente ao desvio padronizado das replicações *bootstrap*, em que é dado pela equação 2.13.

$$\hat{s}e = \left[\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{*b} - \mu(\hat{\theta}))^2 \right]^{\frac{1}{2}}; \quad (2.13)$$

No qual, $\mu(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{*b}$. A distribuição de $\hat{\theta}$ se aproxima ainda mais de uma distribuição normal com média e a variância próxima de $\hat{\theta} \sim N(\theta, \hat{s}e^2)$. Para se calcular o intervalo de confiança é dada a equação 2.14, em que $z^{(1-\alpha/2)}$ indica o ponto em que equivale ao percentil $100 * (1 - \alpha/2)$ da distribuição normal com média 0 e variância 1, $N(0, 1)$. O valor de α é o nível de significância.

$$\left[\hat{\theta} - z^{(1-\alpha/2)} \hat{s}e, \hat{\theta} + z^{(1-\alpha/2)} \hat{s}e \right] \quad (2.14)$$

O intervalo de confiança é construído em um conjunto de valores por um determinado parâmetro, dado um erro fixado e controlado. Considere dois intervalos de confiança *bootstrap* c_1 e c_2 construídos para as médias de dois conjuntos Φ_1 e Φ_2 (OLIVEIRA, 2018). Testes de intervalo de hipóteses relacionados com estas médias podem ser analisados

através dos intervalos construídos. Caso o valor esteja presente na hipótese nula não esteja presente no intervalo, ela deve ser rejeitada, conforme a equação 2.15.

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \quad (2.15)$$

Se os intervalos de confiança c_1 e c_2 interceptem, então a hipótese H_0 não é rejeitada, a hipótese está dentro dos intervalos, pois existe valores satisfeito, caso ao contrário, essa hipótese é rejeita (SOUZA; SOUZA; AMARAL, 2019).

Os intervalos de confiança e testes de hipóteses, facilitam a comparação das médias quando vários métodos são analisados simultaneamente. Se esses valores se interceptam, as médias são consideradas iguais. Caso os valores do intervalo $c_1 > c_2$ conclui-se que a μ_1 é maior que μ_2 . Dessa forma, os resultados obtidos para esse trabalho na comparação dos descritores, são apresentados nesse trabalho através de intervalos de confiança construídos a partir dos percentis *bootstrap* provenientes de distribuições de medidas de desempenho.

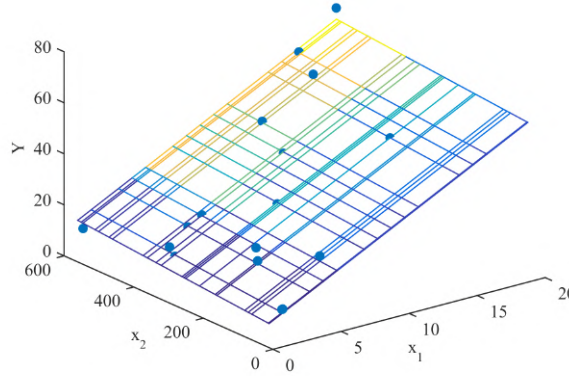
2.2.4 Regressão Linear Múltipla

A Regressão Linear Múltipla (MLR - *Multiple Linear Regression*) é uma técnica da estatística multivariada, usada para examinar a relação entre uma única variável dependente (Y_i) e duas ou mais variáveis independentes (x_{ik}). Uma vantagem da regressão linear múltipla é a capacidade de identificar discrepâncias ou anomalias. Dessa forma, a MLR usa as variáveis independentes cujos os valores são conhecidos para prever o valor dependente (HAIR *et al.*, 2013). O modelo da regressão linear múltipla com x_{ik} variáveis dependentes é definido pela Equação 2.16, em que, Y_i representa a variável dependente, $x_{i1}, x_{i2}, \dots, x_{ik}$, são os valores da i -ésima observação das k variáveis dependentes, os valores $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ são os parâmetros dos coeficientes de regressão e ϵ_i é o resíduo expresso por $y_i - \hat{Y}_i$.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i \quad (2.16)$$

Na Figura 7 é apresentado um modelo de regressão linear múltipla em uma superfície plana no espaço tridimensional, observa-se que x_1 e x_2 corresponde as variáveis independente e Y a variável dependentes.

Figura 7 – Exemplo de um modelo de regressão linear múltipla.



Fonte: Autoria Própria (2020).

Para determinar a equação linear, pode-se utilizar um método conhecido como o método dos mínimos quadrados. Esse método tem como vantagens a obtenção das melhores estimativas, possuir os desvios maiores como também permite realizar teste de significância (MORETTIN, 2010). Para encontrar os valores dos coeficientes da regressão β no qual $n > k$ observações estejam disponíveis e seja x_{ij} a i -ésima observação ou nível da variável x_j , denotado pela equação 2.17.

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i, \quad i = 1, 2, \dots, n \quad (2.17)$$

A função dos mínimos quadrados é dada pela equação 2.18.

$$L = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2 \quad (2.18)$$

Essa função deve ser minimizada com respeito à $\beta_0, \beta_1, \dots, \beta_j$. Esse método pode ser expressa em forma de matriz, em que a equação 2.19 representa a função dos mínimos quadrados.

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \cdots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{bmatrix} \quad (2.19)$$

Na Equação 2.19 é possível notar que há $p = k + 1$ equações, para cada coeficientes desconhecidos da regressão. Esse modelo é um sistema de n equações, que pode ser expresso

de forma reduzida pela equação 2.20.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n1} & \cdots & x_{nk} \end{bmatrix} \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ e } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (2.20)$$

Em que, Y é um vetor $(n \times 1)$ das variáveis dependentes, X é uma matriz $(n \times p)$ dos níveis das variáveis independentes (supondo que a interceptação seja sempre multiplicada por um valor constante e unidade), β é um vetor $(p \times 1)$ dos coeficientes da regressão e ϵ é um vetor $(n \times 1)$ de erros aleatórios. A matriz X é chamada de matriz modelo. Para encontrar os valores dos estimadores dos mínimos quadrados $\hat{\beta}$ pela equação ??.

O método dos mínimos quadrados de $\hat{\beta}$ é dado a equação 2.21.

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2.21)$$

A equação 2.20 possui $p = k + 1$ equações normais e incógnitas para $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$. A matriz $X'X$ é não singular, para observa a forma matricial foi expressa pela equação 2.19 em que resulta em uma equação normal. Dessa forma, $X'X$ é uma matriz simétrica $(p \times p)$ e $X'y$ é um vetor de coluna $(p \times 1)$. Os elementos da diagonal $X'X$ são as somas dos quadrados dos elementos da coluna de X e os elementos fora da diagonal são as somas dos produtos cruzados dos elementos das colunas de X . Desse modo, os elementos de $X'y$ são as somas dos produtos cruzados das colunas de X e das observações Y_i . Um exemplo de uma MLR seria para ajustar um modelo de regressão e identificar os coeficientes de regressão. Na Tabela 2 são apresentados os dados de um processo de fabricação de semicondutores, em que y é a resistência à tração de um fio colado, x_1 é o comprimento do fio e x_2 altura da garra. Fez uso da abordagem matricial para ajustar e encontrar os coeficientes.

Tabela 2 – Dados de um processo de fabricação de semicondutores para a construção de um modelo de regressão múltipla.

Número de Observações	Comprimento do fio x_1	Altura da Garra x_2	Resistência à tração y
1	2	50	9.95
2	8	110	24.45
3	11	120	31.75
4	10	550	35
5	8	295	25.02
6	4	200	16.86
7	2	375	14.38
8	2	52	9.6
9	9	100	24.35
10	8	300	27.5

Para encontrar os coeficientes, foi utilizado o método dos mínimos quadrados em forma matricial. Os valores X correspondem as variáveis independentes, no qual são obtidos os coeficientes de regressão e y corresponde aos valores das observações, ou seja, a variável dependente. Para encontrar os coeficientes de regressão, é calculado a matriz transposta de X pela matriz X , em que é definido pela Equação 2.22, no qual é encontrado o resultado.

$$X' * X \Rightarrow \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2 & 8 & \cdots & 8 \\ 50 & 110 & \cdots & 300 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 50 \\ 1 & 8 & 110 \\ \vdots & \vdots & \vdots \\ 1 & 8 & 300 \end{bmatrix} = \begin{bmatrix} 10 & 64 & 2152 \\ 64 & 522 & 15114 \\ 2152 & 15114 & 701854 \end{bmatrix} \quad (2.22)$$

A matriz transposta X' é calculada com os valores da variável dependente, desse modo $X'y$ é definido pela Equação 2.23.

$$X'y \Rightarrow \begin{bmatrix} 1 & 2 & 50 \\ 1 & 8 & 110 \\ \vdots & \vdots & \vdots \\ 1 & 8 & 300 \end{bmatrix} * \begin{bmatrix} 9.95 \\ 24.45 \\ \vdots \\ 27.50 \end{bmatrix} = \begin{bmatrix} 0.0219 \\ 0.1669 \\ 5.3577 \end{bmatrix} \quad (2.23)$$

As estimativas dos mínimos quadrados são encontradas a partir da Equação 2.24, no qual é calculado a inversa de $(X'X)^{-1}$ pelo resultado da Equação 2.23.

$$(X'X)^{-1} * X'y \Rightarrow \begin{bmatrix} 10 & 64 & 2152 \\ 64 & 522 & 15114 \\ 2152 & 15114 & 701854 \end{bmatrix}^{-1} * \begin{bmatrix} 0.0219 \\ 0.1669 \\ 5.3577 \end{bmatrix} = \begin{bmatrix} 4.5444 \\ 2.2158 \\ 0.0147 \end{bmatrix} \Leftarrow \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad (2.24)$$

O modelo de regressão com os coeficientes de regressão é $\hat{y} = 4.5444 + 2.2158x_1 + 0.0147x_2$. Esse modelo pode ser usado para prever valores da resistência à tração para valores do comprimento de fio x_1 e da altura da garra x_2 . A MLR é frequentemente útil para construir estimativas de intervalos de confiança para os coeficientes de regressão β_j , no qual fez uso para este trabalho. No desenvolvimento dos intervalos de confiança, os erros ϵ_i seguem uma distribuição normal e são independentemente distribuídos, com média e variância zero. A Equação 2.25 apresenta o intervalo de confiança para os coeficientes de regressão, em que $\hat{\beta}_j$ é normalmente distribuído com vetor médio β , $\hat{\sigma}^2 C_{jj}$ a matriz de covariância, o n é o tamanho do vetor e p a quantidade de parâmetros.

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \quad (2.25)$$

2.3 Transformadas Discretas: Reconhecimento do Sinal

Este trabalho utiliza algumas transformada discretas, tais como: Transformada Discreta de Cosseno (DCT - *Discrete Cosine Transform*), Transformada Discreta de Chebyshev (DChT - *Discrete Chebyshev Transform*) e Transformada Rápida de Walsh-Hadamard (FWHT - *Fast Walsh-Hadamard Transform*). As transformadas discretas desempenham um papel importante na área de processamento de sinais, consiste na conversão do sinal de tempo discreto para o domínio da frequência. O sinal é uma função de uma ou mais variáveis, no qual transporta informações sobre um fenômeno físico em observação. Esse sinal é classificado em relação ao tempo que pode ser comportado em tempo contínuo e discreto. O tempo contínuo é definido em uma faixa contínua de tempo, já o tempo discreto é em instantes isolados e a amplitude são uniformemente espaçados. Para os computadores, o sinal geralmente é discretizado por meio de *hardwares* específicos, sendo assim transformados em sinal de tempo discreto, desse modo, destaca-se o uso de técnicas matemáticas para representação e manipulação de sinais discretos (SHIRADO *et al.*, 2015).

2.3.1 Transformada Discreta de Cosseno

A DCT criada em 1974 por Ahmed, Natarajan e Rao (1974), é uma técnica para converter um sinal de tempo discreto para o domínio da frequência. Essa técnica é relacionada com a Transformada Rápida de *Fourier* (FFT - *Fast Fourier Transform*), no qual utiliza as funções dos senos e cossenos em um espaço real e imaginário (número complexo). Já a DCT faz uso da parte real a qual utiliza o somatório de cossenos. Com a transformação para o domínio da frequência obtém-se os coeficientes: DC (*Direct Current*) e AC (*Alternate Current*). No coeficiente DC é concentrada a maior parte da energia, sendo o primeiro coeficiente de transformação, o qual representa a média do sinal. O coeficiente AC indica a amplitude correspondente do componente da frequência. As DCTs mais utilizadas são do tipo: DCT-1, DCT-2, DCT-3 e a DCT-4, para esse trabalho será utilizada a DCT-4 (REIGOTA, 2007).

A DCT-4 é definida pela Equação 2.26, em que N é o tamanho do vetor, x o sinal e t o tempo.

$$c(x) = \alpha(x) \sum_{t=0}^{N-1} f(t) \cos \left(\frac{\pi}{2N} (2t-1)(x-1) \right) \quad (2.26)$$

A sua inversa é definida pela Equação 2.27.

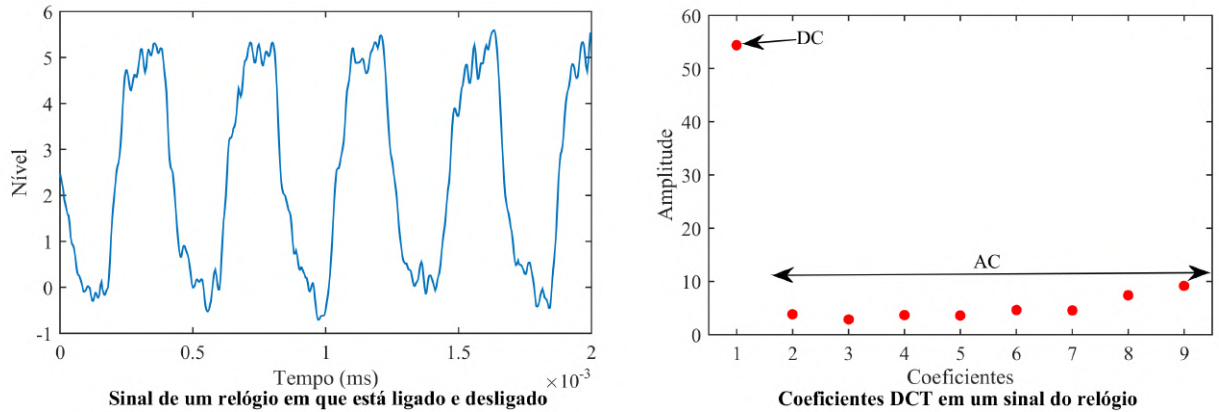
$$f(t) = \sum_{x=0}^{N-1} \alpha(x) c(x) \cos \left(\frac{\pi}{2N} (2t-1)(x-1) \right) \quad (2.27)$$

Para as Equações (2.26) e (2.27), é necessário encontrar o $\alpha(x)$ que pode ser definido pela Equação (2.28).

$$\alpha(x) = \begin{cases} \sqrt{\frac{1}{N}}, & \text{para } x = 0 \\ \sqrt{\frac{2}{N}}, & \text{para } x \geq 1 \end{cases} \quad (2.28)$$

Na Figura 8 (a) apresenta um sinal do relógio no tempo discreto, onde está ligado e desligado no decorrer do tempo. A Figura 8 (b) apresenta os coeficientes da DCT de forma compacta o sinal do relógio, o valor mais alto do gráfico corresponde ao coeficiente DC do sinal, representando a média do sinal, e os demais valores, representam o coeficiente AC, que indica a amplitude da frequência.

Figura 8 – (a) Sinal de um relógio que está ligado e desligado no decorrer do tempo. (b) Obtenção dos coeficientes DC e AC com a DCT-4 do sinal do relógio.



Fonte: Autoria Própria (2020).

2.3.2 Transformada Discreta de Chebyshev

Ahmed, Natarajan e Rao (1974), apontou que a DCT-2 está associada a uma série polinomial de Chebyshev, mas não foi descrito no artigo como isso ocorre. Desse modo, Corr *et al.* (2000) apresentou a transformada discreta de Chebyshev (DChT - *Discrete Chebyshev Transform*), no qual está associada a uma série polinomial de grau k . Com a DChT é obtido os coeficientes DC e AC para o domínio da frequência. O primeiro coeficiente DC, representando a média do sinal e o coeficiente AC a amplitude correspondente a frequência. A equação 2.29 permite obter a DChT, em que N é o tamanho do vetor, x o sinal e T_x o polinômio de *chebyshev*.

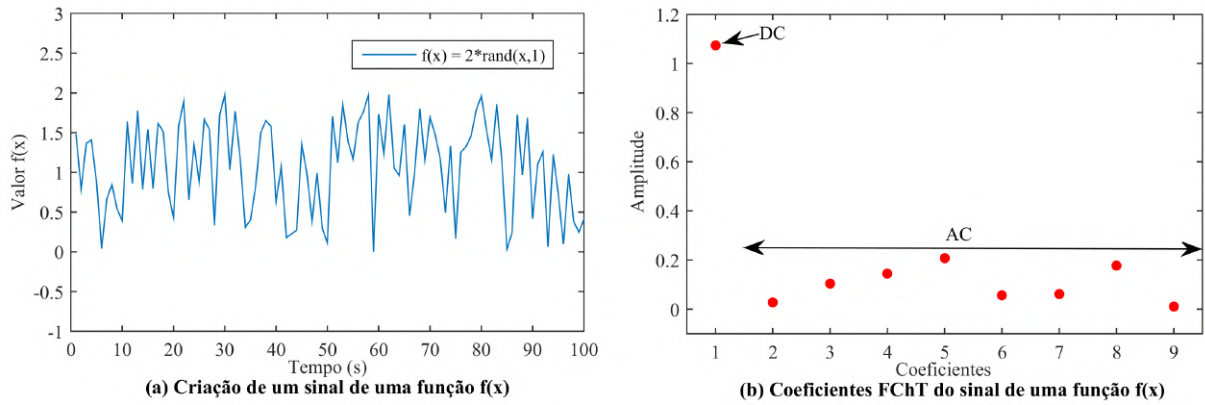
$$G(x) = \frac{2}{N} \sum_{j=1}^N g(x_j) T_x(x_j) \quad (2.29)$$

A sua inversa é obtida pela equação 2.30.

$$g(x) = \frac{1}{2}G(0) + \sum_{j=1}^{N-1} G(j)T_x(x_j) \quad (2.30)$$

O polinômio de *chebyshev* é útil devido as suas propriedades de aproximação para a interpolação polinomial (KOPRIVA, 2009). Na Figura 9 (a) é apresentado um sinal de tempo discreto, no qual foi criado com uma função $f(x) = 2 * \text{rand}(x, 1)$, x é um valor aleatório. A Figura 9 (b) apresenta os nove primeiros valores dos coeficientes, o coeficiente DC corresponde ao valor mais alto do gráfico, que representa a média do sinal, e os demais valores, representam o coeficiente AC, que indica a amplitude da frequência.

Figura 9 – (a) Criação de um sinal com uma função $f(x) = 2 * \text{rand}(x, 1)$ em um domínio de tempo discreto. (b) Obtenção dos coeficientes DC e AC com a FChT do sinal da função $f(x)$.



Fonte: Autoria Própria (2020).

2.3.3 Transformada Rápida de Walsh-Hadamard

A Transformada Rápida de Walsh-Hadamard (FWHT - *Fast Walsh-Hadamard Transform*) consiste em uma implementação simples, por conter um número de operações aritméticas reduzidas desenvolvida por Fino e Algazi (1976), com essa transformada obtém os coeficientes DC e AC. A FWHT pode ser calculada a partir da matriz de *Walsh-Hadamard*, no qual é formada pelo produto tensor. O produto tensor consiste em duas matrizes quadradas $A = [a_{mn}]$ e $B = [b_{lk}]$ formada pela matriz C de tamanho $mk * nl$, nesse modo, para se obter a matriz C_{mk*nl} , os elementos de a_{mn} corresponde a matriz A e B aos elementos da matriz b_{lk} definido pela equação 2.31

$$C_{mk*nl} = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} \quad (2.31)$$

Com o produtor tensor obtido, a matriz de Hadamard é definida como uma matriz quadrada de dimensão $n * m$, onde todas as entradas são 1 e quaisquer duas linhas distintas são ortogonais, definido pela equação 2.32.

$$H_{2^k} = \frac{1}{\sqrt{2^k}} \begin{bmatrix} H_{2^{k-1}} & H_{2^{k-1}} \\ H_{2^{k-1}} & -H_{2^{k-1}} \end{bmatrix}, \quad k = 0, 1, 2, \dots, n-1 \quad (2.32)$$

A equação 2.33 é definido o cálculo da FWHT, em que x é o sinal, N é o tamanho e $H(k, j)$ é a matriz de Hadamard.

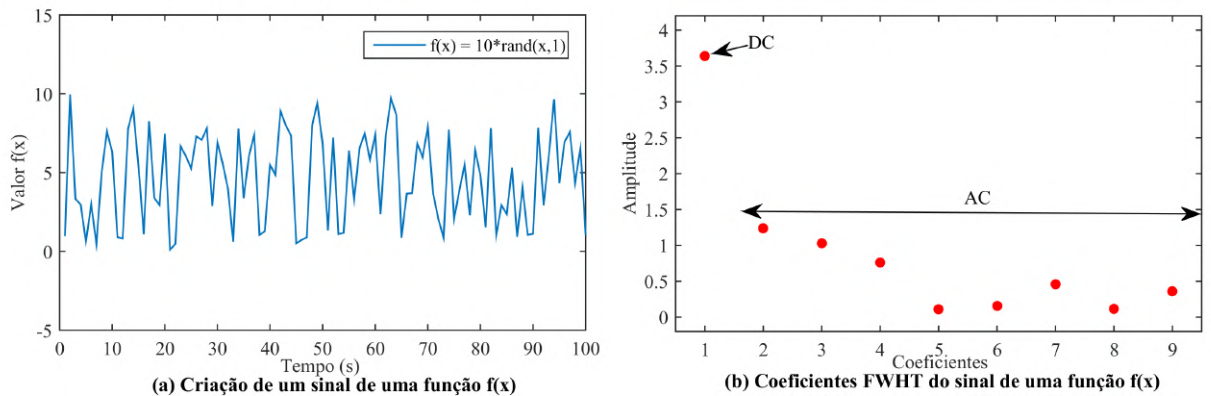
$$f(x) = \sum_{j=0}^{N-1} H_{2^k}(k, j)x(j) \quad (2.33)$$

A sua inversa pode ser calculada pela equação 2.34.

$$c(x) = \sum_{j=0}^{N-1} H_{2^k}(k, j)^{-1}x(j) \quad (2.34)$$

Para obter a inversa, é multiplicado a matriz inversa de Walsh-Hadamard, em que é a própria matriz (DONIAK, 2006). Na Figura 10 (a) ilustra um sinal de tempo discreto para uma função $f(x) = 10 * rand(x, 1)$, x é um valor aleatório. Na Figura 10 (b) apresenta os nove primeiros valores dos coeficientes, o coeficiente DC corresponde ao valor mais alto do gráfico, representa a maior parte da energia do produto tensor, e os demais valores, representa o coeficiente AC, que indica a amplitude da frequência.

Figura 10 – (a) Criação de um sinal com a função $f(x) = 10 * rand(x, 1)$ em um domínio de tempo discreto. (b) Obtenção dos coeficientes DC e AC com a FWHT do sinal da função $f(x)$.



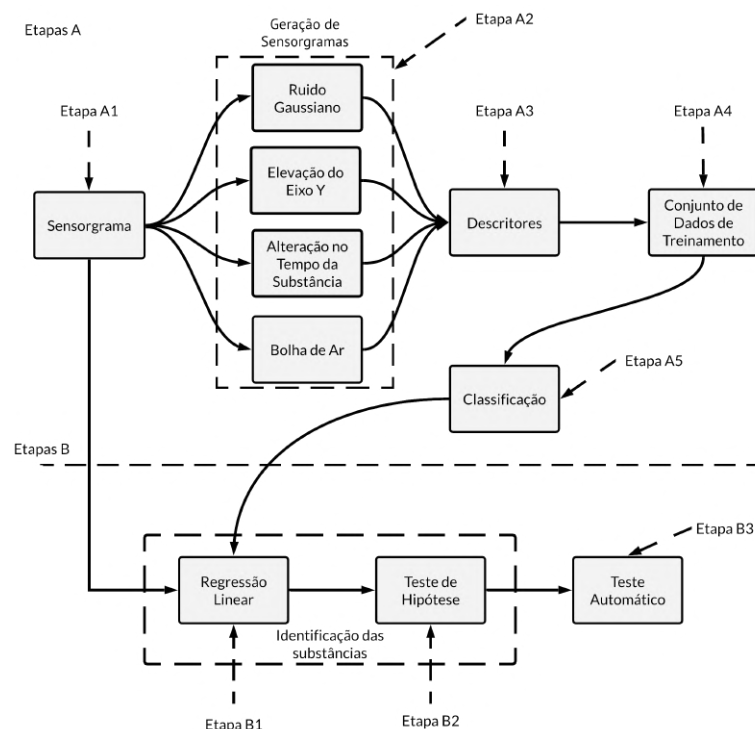
Fonte: Autoria Própria (2020).

3 Metodologia

Este trabalho aplica algoritmos de aprendizagem de máquina supervisionada, não-supervisionada e das transformadas discretas para a análise de sensorgramas, com o objetivo de analisar, descrever, identificar, classificar e como também tornar a resposta de protocolo experimental adquirido de forma automática. As abordagens e tecnologias de aprendizagem de máquina permitiram tornar o sensor SPR inteligente, para reconhecer partes de interesse da região do sensorgrama de tal forma que auxilie na tomada de decisão. Por meio do sensor SPR inteligente, será possível classificar e ordenar a resposta do sensorgrama para que possa fornecer uma resposta confiável.

A execução das tarefas de aprendizagem foram divididas em duas etapas. A primeira, denominada de etapa A, é destinada para descrever, classificar e ordenar os sensorgramas. A segunda etapa, denominada de etapa B, é para identificação das substâncias, averiguar a qualidade e realizar o teste automático. Na Figura 11 são apresentadas as etapas para a execução das atividades.

Figura 11 – Diagrama de Blocos indicando as etapas para realização das tarefas de criar, descrever, identificar, classificar e tornar a resposta automática dos sensorgramas.

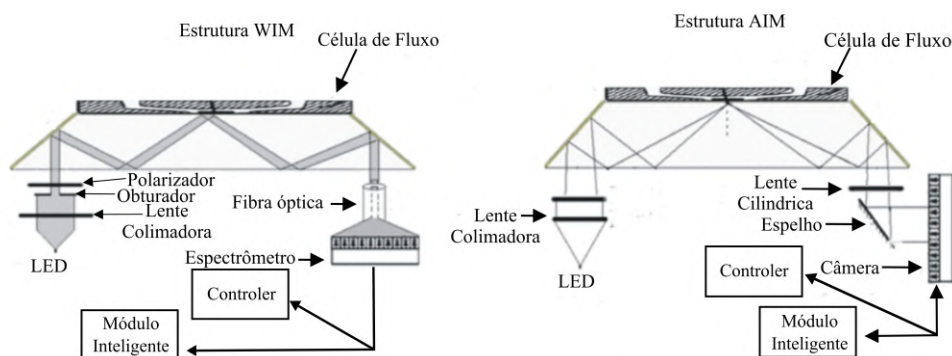


Fonte: Autoria Própria (2020).

3.1 Etapa A - Definição do Ambiente e Dados para Inferência

Os sensorgramas foram adquiridos da tese de doutorado de Oliveira (2016). Para a aquisição dos sensorgramas, foi utilizado um sensor SPR no modo AIM e WIM. O sensor SPR é baseado no PPBIO (Prisma Polimérico para Aplicações BIOLógicas). Na Figura 12 é apresentado a estrutura de um sensor SPR com o prisma PPBIO em ambos os modos AIM e WIM, em que diferem os tipos de componentes ópticos, mecânicos e eletrônicos utilizados (OLIVEIRA *et al.*, 2013). Na estrutura PPBIO é depositado o metal, com o uso da configuração de Kretschmann, esse prisma contém características bastantes peculiares que são a descartabilidade, o custo baixo e a flexibilidade, o que torna a instrumentação do sensor SPR menos dispendiosa, por evitar partes móveis e/ou rotacionais. As paredes laterais inclinadas do PPBIO são especulares para facilitar a reflexão da luz incidente. Esse prisma possui uma fina camada metálica depositada na base superior, que evita o uso de óleo/gel para acoplamento da superfície metálica (MOREIRA, 2010).

Figura 12 – Diagrama de Blocos do sensor SPR.



Fonte: Adaptado de (OLIVEIRA *et al.*, 2013).

Como é observado na Figura 12, nos modos de interrogação AIM e WIM, a sua forma de excitação com o uso do PPBIO muda. No modo WIM, os elementos envolvidos na excitação são: LED para a fonte de luz, lente esférica plano-convexa e um polarizador para colimar os feixes de luz, obturador para confinar a luz no PPBIO, uma célula de fluxo para inserir o analito e um espectrômetro para a captura da luz. No modo AIM, os elementos envolvidos na excitação SPR são: um LED de luz monocromática, duas lentes cilíndricas plano-convexas, um espelho para auxiliar na instrumentação do caminho óptico e uma câmera para captura da luz.

Em cada modo de interrogação as curvas SPR são exibidas por meio de um *software*, em que apresenta de forma gráfica na fase de controler. O *software* permite realizar a aquisição, processamento e exibição para cada modo de interrogação. Para aquisição das curvas SPR no modo AIM, foi utilizado o vidro BK7 e os polímeros PMMA, PC e TOPAS, cujo em cada um desses prismas na superfície superior foram depositado o metal ouro (Au). Já para o modo WIM também foi utilizado o vidro BK7 e os polímeros PMMA, PC

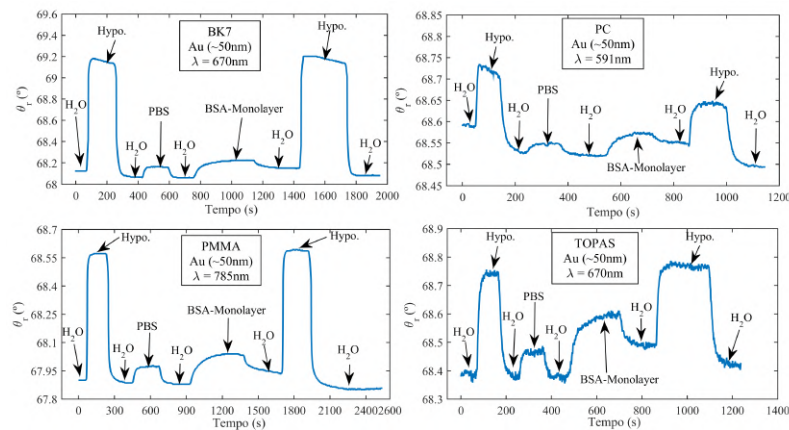
e TOPAS depositado na superfície em cada um desses prismas o metal prata (Ag). Nas curvas SPR pode ser capturado o valor mínimo e esse valor é apresentado na evolução temporal no sensorgrama.

Com a aquisição dos sensorgramas, é possível aplicar o módulo de inteligência. Esse módulo permite utilizar as técnicas de aprendizagem por meio do *software* MATLAB¹. Dessa forma, o sensorgrama é lido por meio do *software*, em que utiliza matrizes para expressar a matemática computacional, funções matemáticas, algoritmos de aprendizagem de máquina e ruídos (GUIDE, 2020).

A1 - Sensorgrama: Aspectos Experimentais

As substâncias dos sensorgramas diferem de acordo com o modo utilizado, seja AIM ou WIM. Para o modo AIM foi utilizado as substâncias: água deionizada e desgaseificada (H_2O), hipoclorito, soluções PBS (*Phosphate Buffered Saline*) e solução da proteína BSA-monolayer diluída em PBS. Na Figura 13, é apresentado os sensorgramas no modo AIM adquiridos por meio do sensor SPR com o uso do PPBIO e o filme de ouro, com espessura de 50 nm feito de polímeros PMMA, PC e TOPAS e vidro BK7. O vidro BK7 e o polímero TOPAS utilizou um comprimento de onda de 670 nm, já o polímero PC foi utilizado o comprimento de onda de 591 nm e o polímero PMMA foi inserido o comprimento de onda de 785 nm.

Figura 13 – Sensorgramas obtidos no modo AIM PC/Au, BK7/Au, PMMA/Au e TOPAS/Au utilizando o metal Au de 50 nm. O protocolo experimental inicializado com H_2O , depois Hipoclorito, retorno para H_2O , em seguida PBS, retorno H_2O ; depois inserido a solução BSA, remoção de excesso com H_2O , limpeza com Hipoclorito e finalizado com H_2O .



Fonte: Autoria Própria (2020).

O protocolo experimental para os polímeros e o vidro BK7 é apresentado na Tabela 3, em que as substâncias seguem a seguinte ordem: $H_2O \rightarrow \text{Hipoclorito} \rightarrow H_2O \rightarrow$

¹ Disponível em <https://www.mathworks.com/products/matlab.html>

$PBS \rightarrow H_2O \rightarrow BSA - \text{monolayer} \rightarrow H_2O \rightarrow \text{Hipoclorito} \rightarrow H_2O$.

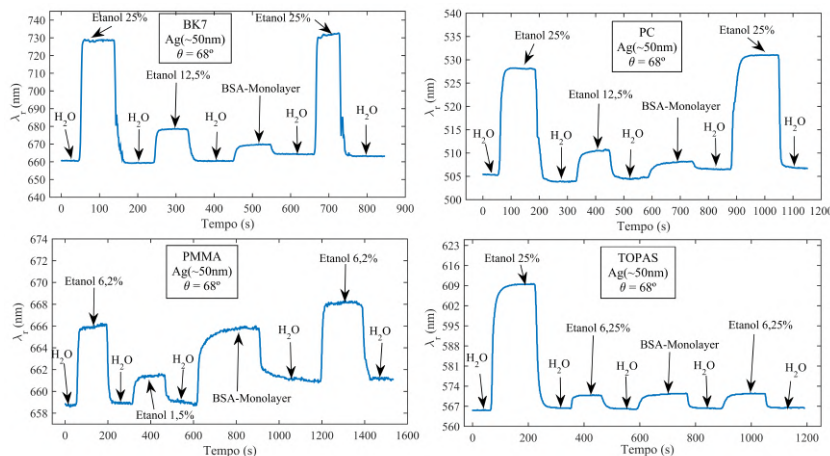
Tabela 3 – Protocolo experimental obtidos no modo AIM com o vidro BK7 e polímeros PMMA, TOPAS e PC utilizando o metal Au de 50 nm.

Material PPBIO	Metal	Protocolo Experimental
BK7/PC/PMMA/TOPAS	Au	$H_2O \rightarrow \text{Hipoclorito} \rightarrow H_2O \rightarrow$ $PBS \rightarrow H_2O \rightarrow BSA\text{-Monolayer}$ $\rightarrow H_2O \rightarrow \text{Hipoclorito} \rightarrow H_2O$

Fonte: Autoria Própria (2020).

Também foi adquirido sensorgramas no modo WIM com as substâncias: água deionizada e desgaseificada (H_2O), soluções de alta e baixa concentração de etanol e solução proteína albumina sérica bovina (BSA - *Bovine Serum Albumin*) diluída em PBS em uma concentração de 0,1mg/l, por meio do sensor SPR com o uso do PPBIO e o filme de prata com o vidro BK7 e polímeros PC, PMMA e TOPAS com o ângulo fixo de 68° . Na Figura 14 são apresentados os comprimentos de onda mínimo (nm) vs tempo de acordo com o protocolo experimental.

Figura 14 – Sensorgramas obtidos no modo WIM com o vidro BK7 e polímeros PC, PMMA e TOPAS utilizando o metal Ag. O valor mínimo do comprimento de onda de ressonância foi utilizado o algoritmo minimum-hunt. O protocolo experimental para o vidro BK7 e polímero PC foi inicializado com H_2O , depois Etanol 25%, retornando para H_2O , em seguida Etanol 12,5%, retornou-se para H_2O ; depois BSA-monolayer, remoção com H_2O ; limpeza com Etanol 12,5% finalizado com H_2O . Para o polímero PMMA iniciou-se com H_2O ; inseriu-se Etanol 6,2%, retornou para H_2O ; depois Etanol 1,5%, retornou para H_2O ; depois BSA-Monolayer; retornou-se com H_2O , inserido o Etanol a 6,2% finalizado com H_2O . Com o polímero TOPAS inseriu H_2O ; depois Etanol 25%; retornou para H_2O , depois Etanol 6,25%, retorno para H_2O , em seguida BSA-Monolayer, remoção com H_2O ; inserido o Etanol a 6,25% e finalizado com H_2O .



Fonte: Autoria Própria (2020).

O protocolo experimental para os polímeros e o vidro BK7 é apresentado na Tabela 4. O protocolo experimental com as substâncias segue a seguinte ordem para o vidro BK7 e o polímero PC: $H_2O \rightarrow$ Etanol 25% $\rightarrow H_2O \rightarrow$ Etanol 12,5% $H_2O \rightarrow$ BSA-monolayer $\rightarrow H_2O \rightarrow$ Etanol 12,5% $\rightarrow H_2O \rightarrow$ Etanol 25% $\rightarrow H_2O$.

Para o polímero PMMA o protocolo experimental, seguiu a ordem: $H_2O \rightarrow$ Etanol 6,2% $\rightarrow H_2O \rightarrow$ Etanol 1,5% $H_2O \rightarrow$ BSA-monolayer $\rightarrow H_2O \rightarrow$ Etanol 6,2% $\rightarrow H_2O \rightarrow$ Etanol 6,2% $\rightarrow H_2O$. E o polímero TOPAS o protocolo experimental seguiu: $H_2O \rightarrow$ Etanol 25% $\rightarrow H_2O \rightarrow$ Etanol 6,25% $H_2O \rightarrow$ BSA-monolayer $\rightarrow H_2O \rightarrow$ Etanol 6,25% $\rightarrow H_2O$.

Tabela 4 – Protocolo experimental obtidos no modo WIM com o vidro BK7 e polímeros PMMA, TOPAS e PC com o metal Ag de 50 nm.

Vidro ou Polímero	Metal	Protocolo Experimental
BK7/PC	Ag	$H_2O \rightarrow$ Etanol 25% $\rightarrow H_2O \rightarrow$ Etanol 12,5% $\rightarrow H_2O \rightarrow$ BSA Monolayer $\rightarrow H_2O \rightarrow$ Etanol 12,5% $\rightarrow H_2O$
PMMA	Ag	$H_2O \rightarrow$ Etanol 6,2% $\rightarrow H_2O \rightarrow$ Etanol 1,5% $\rightarrow H_2O \rightarrow$ BSA Monolayer $\rightarrow H_2O \rightarrow$ Etanol 6,2% $\rightarrow H_2O$
TOPAS	Ag	$H_2O \rightarrow$ Etanol 25% $\rightarrow H_2O \rightarrow$ Etanol 6,25% $\rightarrow H_2O \rightarrow$ BSA Monolayer $\rightarrow H_2O \rightarrow$ Etanol 6,25% $\rightarrow H_2O$

Fonte: Autoria Própria (2020).

A2 - Geração de Sensorgramas

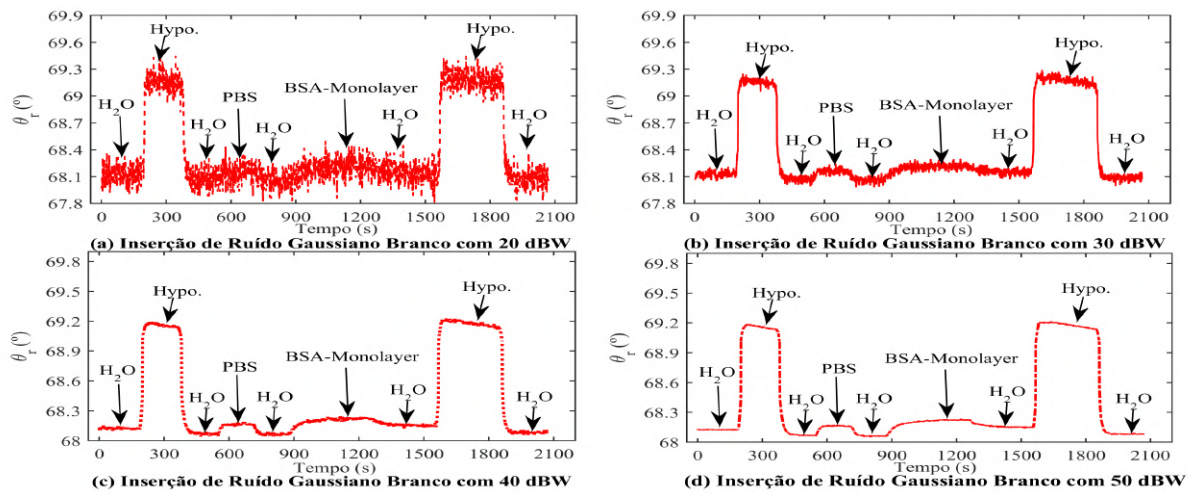
Foram criados novas amostras para os sensorgramas obtidos no modo AIM e WIM. Aplicou-se modificações nos sensorgramas como: introdução de ruído gaussiano branco, diminuição temporal em regiões de interesse, aumento ou diminuição no valor do eixo y (posição de ressonância) e inserção de bolhas de ar. Essas modificações podem representar o uso de diferentes algoritmos de detecção do valor mínimo da curva SPR e mudança no tempo de inserção de substâncias, no qual foi levado em consideração aspectos da cinética de reação das substâncias envolvidas, ou seja, foram tomados cuidados para a geração dos dados de forma significativa. As modificações geram economias de materiais, não sendo necessário a inserção de diferentes PPBIO para as novas amostras.

Na Figura 15 (a), (b), (c) e (d) que corresponde a estrutura multicamada BK7/Au são apresentados sensorgramas com a inserção de ruídos gaussianos branco sem alteração

do protocolo experimental. Foi utilizado o ruído gaussiano branco constante por selecionar uma frequência e como também contém uma dispersão no sinal. Esse ruído, produz modelos matemáticos simples e tratáveis, que são úteis para simular as informações sobre o comportamento do sinal (MCCLANING; VITO, 2001).

Observa-se na Figura 15 (a) a inserção de um ruído gaussiano com 20 dBW em que não há uma distinção entre as substâncias H_2O , PBS e BSA-Monolayer, ou seja, essas substâncias podem ser as mesmas. Já com o aumento do ruído gaussiano branco como é possível observar nas Figuras 15 (b), (c) e (d) já é notável a diferença entre as substâncias.

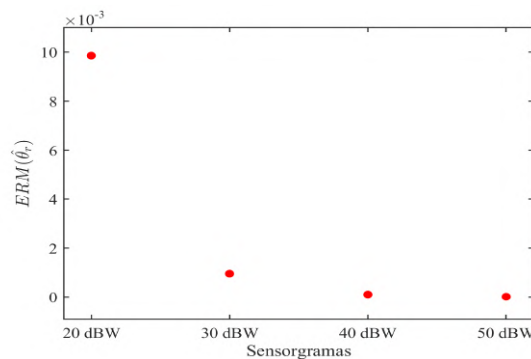
Figura 15 – Inserção de ruído gaussiano branco em sensorgramas com o vidro BK7 e o metal ouro. (a) Inserção com 20 dBW (b) 30 dBW. (c) 40 dBW e (d) 50 dBW.



Fonte: Autoria Própria (2020).

Na Figura 16 é apresentado o EMQ (Erro Médio Quadrático) da aplicação de ruídos gaussianos brancos entre 20 dBW a 50 dBW, a medida que aumenta o ruído gaussiano branco o EMQ diminui. Desse modo, foram criados novos sensorgramas com adição de ruído gaussiano branco de 40 dBW, pois se torna perceptível a adição do ruído como também a diferença entre as substâncias.

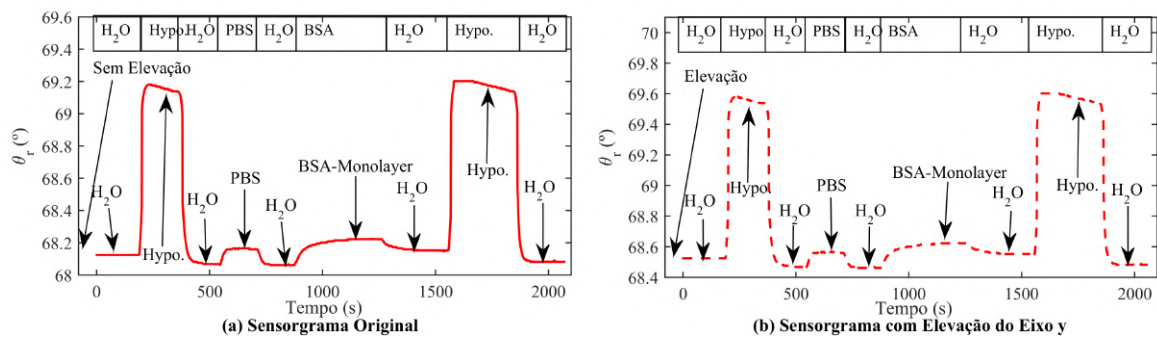
Figura 16 – Erro médio quadrático da inserção de ruído gaussiano branco em sensorgramas com o vidro BK7 e o metal ouro.



Fonte: Autoria Própria (2020).

Também nessa etapa foram criados sensorgramas em que foi diminuído ou aumentado o valor da posição de ressonância. A posição de ressonância corresponde ao valor do eixo Y, essa mudança representa um novo ponto de operação, em que pode ser a troca da luz, mudança do metal ou prisma. Para demonstrar como ocorreu essa mudança foi utilizado o sensorgrama com a estrutura multicamada BK7/Au, apresentada na Figura 17 (a). Na Figura 17 (b) é possível observar o valor das substâncias em que foi aumentado para 0.4° , criando assim novos sensorgramas.

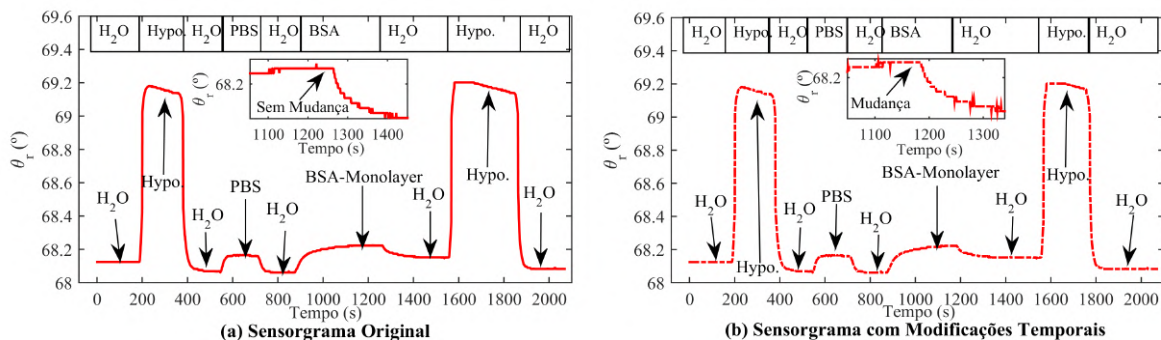
Figura 17 – Aumento da posição de ressonância. (a) Sensorgrama original com o vidro BK7 e o metal ouro (BK7/Au). (b) Criado um novo sensorgrama com o aumento da posição de ressonância correspondente ao eixo y em 0.5° .



Fonte: Autoria Própria (2020).

Foi diminuído os regimes permanentes (regiões de interesses) das substâncias inseridas no protocolo experimental nos sensorgramas. Essa situação representa a economia de materiais, tendo em vista o custo elevado que algumas substâncias possuem. Na Figura 18 (a), é apresentado o sensorgrama original com a estrutura multicamada BK7/Au. Como é possível observar na Figura 18 (b), ocorreu uma mudança no tempo na inserção das substâncias BSA-Monolayer e Hipoclorito. Para visualizar essa mudança apresenta-se um sub-gráfico da modificação no tempo da substância BSA-Monolayer.

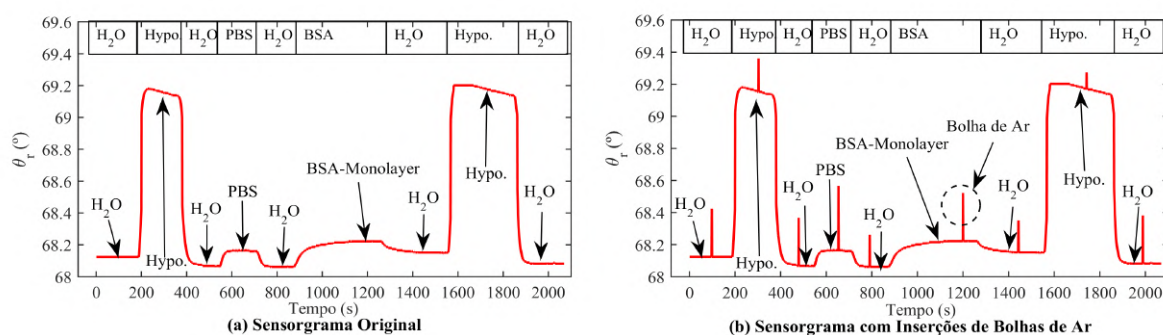
Figura 18 – Modificações temporais nas substâncias do sensorgrama. (a) Sensorgrama original com o vidro BK7 e o metal ouro (BK7/Au). (b) Criado um novo sensorgrama com a diminuição temporal das substâncias BSA-Monolayer e Hipoclorito.



Fonte: Autoria Própria (2020).

A última modificação foi a inserção de bolhas de ar em cada regime permanente do protocolo experimental. A bolha de ar degrada a resposta do sensorgrama, pois com essa bolha o sinal de calibração ou referências torna-se irrelevante. Na Figura 19 (a), é apresentado o sensorgrama original com a estrutura multicamada BK7/Au. Na Figura 19 (b), é criado um novo sensorgrama com inserções de bolhas de ar nos regimes permanentes.

Figura 19 – Inserção de bolha de ar na substância. (a) Sensorgrama original com o vidro BK7 e o metal ouro (BK7/Au). (b) Criado um novo sensorgrama com inserção de bolha de ar na substância BSA-Monolayer.



Fonte: Autoria Própria (2020).

A3 - Descritores

Consistiu em aplicar quatro descritores para fazer uma análise comparativa entre qual é o melhor descritor para classificar o protocolo experimental. Os descritores utilizados foram: transformada discreta de cosseno (DCT-4) (AHMED; NATARAJAN; RAO, 1974), transformada rápida de Walsh-Hadamard (FWTH) (SUJA *et al.*, 2016), transformada discreta de Chebyshev (DChT) (DEVILLE; LABROSSE, 1982) e o descritor de sensorgrama temporal (TSD - *Temporal Sensorgram Descriptor*).

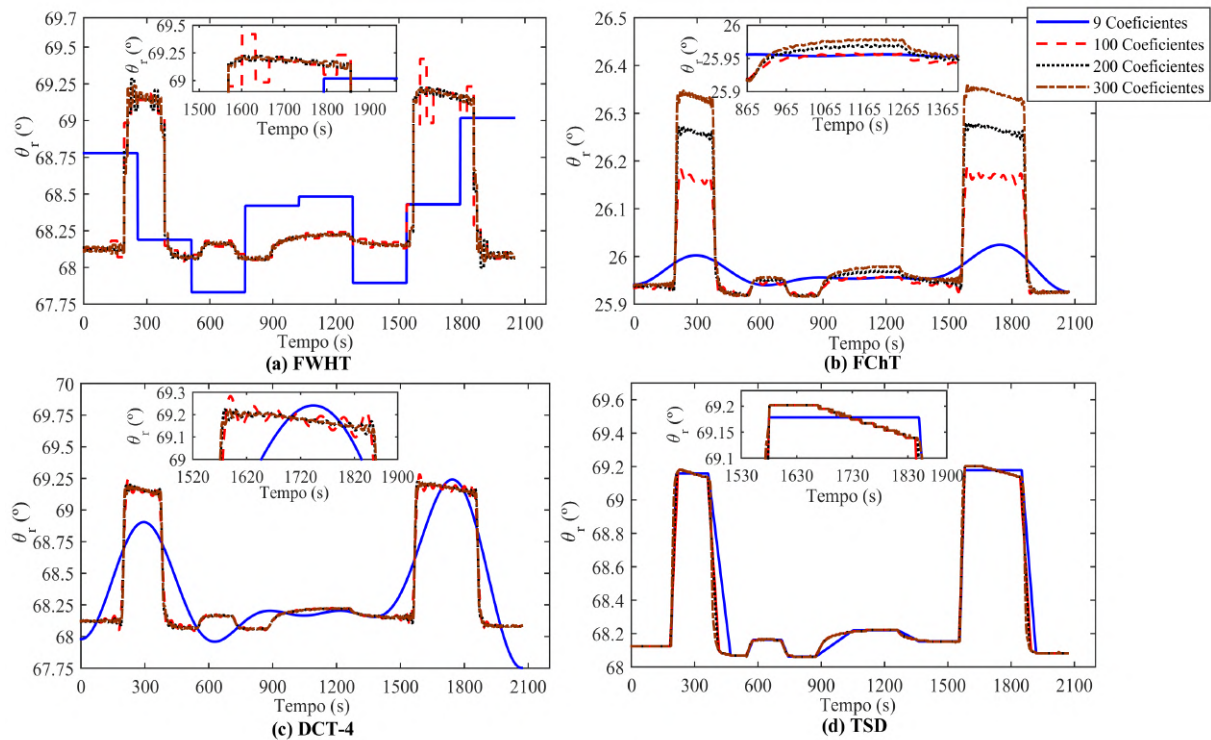
A escolha de um descritor é necessário para descrever um sensorgrama, pois exige a inclusão de aspectos bioquímicos, processos isotérmicos de absorção e dissociação das substâncias. Algumas formas possíveis de descrever um sensorgrama são por meio de modelos matemáticos, como o de *Langmuir*, método dos elementos finitos, equações diferenciais não lineares e tempo finito de diferença finita, porém a aplicação desses modelos torna-se um trabalho complexo, dependente de experimentos custosos e exigindo uma série de suposições (SOUSA *et al.*, 2018), (ZHANG *et al.*, 2018a); (AGRAWAL *et al.*, 2017); (TIWARI *et al.*, 2015).

Os descritores são excelentes para representar de forma compacta a informação dos sensorgramas. Foi escolhido os descritores FWHT, FChT e DCT-4 por não utilizar um número complexo, apenas o número real, tornando-se mais fácil de realizar o cálculo. Já o descritor TSD faz o uso de números complexos, porém a sua representação é composta pela média na parte real e o comprimento na parte imaginária do regime permanente.

Não foi utilizado a transformada rápida de *fourier*, por apresentar cálculos complexos, como também a sua representação em um gráfico se tornaria complicada (PARSONS; BOONMAN; OBRIST, 2000).

Na Figura 20 (a), (b) e (c) é apresentado as inversas dos descritores com 9, 100, 200 e 300 coeficientes de nível AC para FWTH, FChT, DCT-4 e o TSD que representa os seus coeficientes em forma de números complexos para a reconstrução dos sensorgramas do grupo 1 da estrutura multicamada BK7/Au. Na Figura 20 (d) o descritor TSD apresenta a reconstrução do sensorgrama semelhante ao um sensorgrama original, por possuir os regimes permanentes com a média e o seu comprimento, dessa forma, a representação do TSD necessita de uma quantidade menor de coeficientes para se assemelhar a um sensorgrama. Diferente dos descritores FWTH, FChT e DCT-4 que com o seu nível AC se faz necessário de muitos coeficientes para representar um sensorgrama de forma original.

Figura 20 – Reconstrução do sensorgrama com o vidro BK7 e o metal ouro BK7/Au. (a), (b) e (c) apresenta a reconstrução utilizando os descritores FWHT, FChT e DCT-4 com 9, 100, 200 e 300 coeficientes AC. (d) reconstrução do sensorgrama com o descritor TSD com 9, 100, 200 e 300 coeficientes de números complexos.

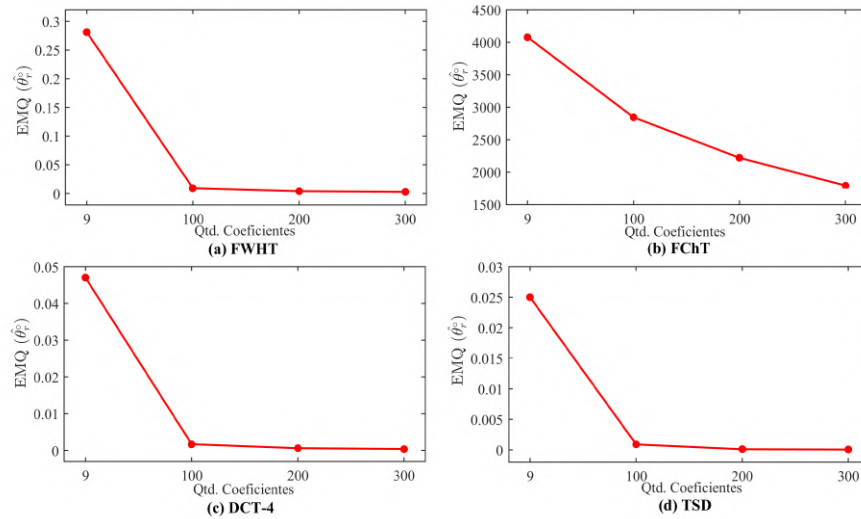


Fonte: Autoria Própria (2020).

A medida que vai adicionando mais coeficientes a representação do sensorgrama vai se tornando o sensorgrama original em relação aos outros descritores. Na Figura 21 tem-se a EMQ (Erro Médio Quadrático) para cada descritor utilizado. É possível observar na Figura 21, a medida que vai sendo adicionado mais coeficientes o EMQ diminui em cada descritor. O descritor que tem o maior erro é o FChT, ou seja, para se assemelhar ao

sensorgrama é necessário de mais coeficientes. Já o descritor TSD, apresenta o menor erro, sua representação com 9 coeficientes é bem semelhante ao sensorgrama. Essa representação da quantidade coeficiente do TSD, significa que em todo o sensorgrama contém espalhado pelo o sensorgrama as médias como também o comprimento.

Figura 21 – Erro Médio Quadrático utilizando a estrutura BK7/Au do grupo 1 com cada descritor .



Fonte: Autoria Própria (2020).

A4 - Conjunto de dados de treinamento

Consistiu em criar dois conjuntos de dados de treinamento para cada descritor. Com o conjunto de dados de treinamento, é atribuído um rótulo de destino em cada sensorgrama pertencente ao seu grupo. O conjunto de dados de treinamento é composto no total de 99 sensorgramas. O primeiro conjunto de dados de treinamento corresponde aos sensorgramas com 3, 5, 7 e 9 regimes permanentes. Para os descritores FWHT, FChT e DCT-4 foi utilizado os três primeiros coeficientes de nível AC. A escolha de nível AC é por corresponder a amplitude do componente da frequência. Já para o descritor TSD foi usado a média e o comprimento dos três primeiros regimes permanentes. O segundo conjunto de dados de treinamento utiliza os cinco primeiros coeficientes de nível AC dos descritores FWHT, FChT e DCT-4 aplicado aos sensorgramas com 5, 7 e 9 regimes permanentes e os cinco primeiros regimes permanentes para o descritor TSD.

A5 - Classificação

Para classificação, foi aplicado o algoritmo de aprendizagem de máquina k -NN em cada descritor com os seus coeficientes. Para o algoritmo k -NN é utilizado a distância, a distância descreve os elementos em um espaço por estarem próximos entre si, ou distantes um do outro. Essa distância pode ser descrita em dois valores x e y que define a distância

entre os valores como um número real não negativo. A função $d(x, y)$ deve satisfazer as propriedades a seguir (DEZA, 2016).

1. **Não Negatividade:** A função $d(x, y)$ é sempre um valor maior ou igual a zero.

$$d(x, y) \geq 0 \quad (3.1)$$

2. **Reflexidade:** A função $d(x, y)$ é igual a 0, se e somente se x for igual y .

$$d(x, y) = 0 \Leftrightarrow x = y \quad (3.2)$$

3. **Simetria:** A função de x e y é igual à distância entre y e x .

$$d(x, y) = d(y, x) \quad (3.3)$$

4. **Desigualdade Triangular:** Considerando a presença de um terceiro ponto na função, a função $d(x, y)$ entre x e y é sempre menor ou igual à soma da função $d(x, z)$ e a distância entre $d(y, z)$.

$$d(x, y) \leq d(x, z) + d(z, y) \quad (3.4)$$

Os trabalhos de Cha (2007), Chomboon *et al.* (2015), Mulak e Talhar (2015) e Alfeilat *et al.* (2019) apresentaram estudos sobre o desempenho do algoritmo k -NN com diferentes distâncias, em que depende principalmente do tipo da distância para classificar. Os resultados desses estudos apresentaram que as distâncias Euclidiana, *Cityblock*, *Chebyshev* e Cosseno apresentam um bom desempenho em relação as outras distâncias, dessa forma, foram escolhidas essas distâncias para o algoritmo.

A distância Euclidiana é uma extensão do teorema de Pitágoras. Essa distância representa a raiz da soma dos quadrado das diferenças entre os valores opostos, na equação 3.5 é representada a distância Euclidiana, em que: DE é o resultado da distância euclidiana, x e y correspondem aos pontos de entrada do vetor em que serão calculados.

$$DE(x, y) = \sqrt{\left(\sum_{i=1}^n |x_i - y_i|^2 \right)} \quad (3.5)$$

A distância *cityblock* também chamada de Manhattan, representa a soma das diferenças absolutas entre os valores opostos da função $d(x, y)$, na equação 3.6 é representada a distância *cityblock*, em que: DCB corresponde ao resultado da distância *cityblock*, x e y correspondem aos pontos de entrada do vetor em que serão calculados.

$$DCB(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (3.6)$$

A distância *Chebyshev* é apropriada em casos em que dois objetos devem ser definidos como diferentes em qualquer dimensão. Essa distância representa um espaço vetorial, em que a distância entre x e y é a maior de suas diferenças ao longo de qualquer dimensão na coordenada, na equação 3.7 é representada a distância *Chebyshev*, em que: DC corresponde ao resultado da distância *Chebyshev*, x e y correspondem aos pontos de entrada do vetor que serão calculados.

$$DC(c, y) = \max_i |x_i - y_i| \quad (3.7)$$

A distância cosseno é derivada da semelhança do cosseno em que mede o ângulo entre os vetores x e y , na qual a distância cosseno é obtida subtraindo a semelhança do cosseno de um. Na equação 3.8 é representada a distância do Cosseno, em que: $DCos$ corresponde ao resultado da distância cosseno, x e y correspondem aos pontos de entrada do vetor que serão calculados.

$$DCos(x, y) = 1 - \left(\frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2}} \right) \quad (3.8)$$

Para avaliar o desempenho do algoritmo k -NN, foi utilizado um método de validação cruzada, no qual consiste em avaliar o desempenho dividindo o conjunto de dados em treinamento e teste. A validação cruzada utilizada é denominada de *Leave-one-out* (LOO-CV - *Leave-one-out Cross-Validation*), nela o conjunto de dados consiste em $N - 1$ dados de treinamento e é deixado um de fora para teste, em que é repetido para um total de N vezes, cada vez é deixado de fora um exemplo diferente para teste. A validação *Leave-one-out* foi utilizada por conter um conjunto de dados de treinamento pequeno, dessa forma, essa validação valida o modelo para todas as combinações possíveis (Yang *et al.*, 2018).

A medida que o algoritmo fornece uma resposta, a média é calculada e um intervalo de confiança não-paramétrico para a média é construído. O intervalo de confiança utilizado foi o *bootstrap* com um erro de 5%, seu objetivo foi de provar uma comparação estatística com os descritores e a acurácia do algoritmo.

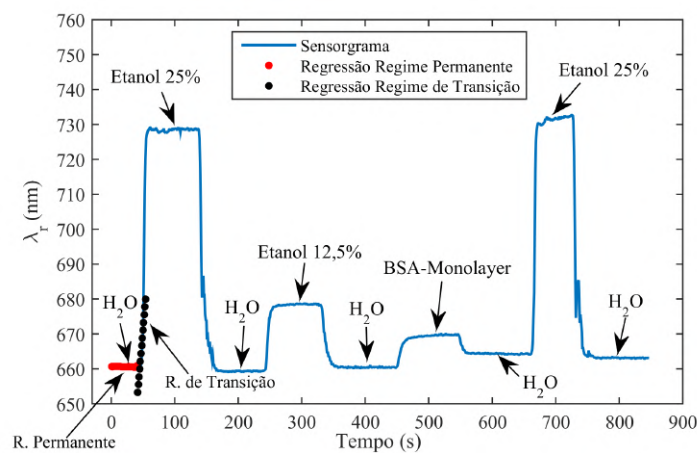
3.2 Etapa B - Sensor em Ação

Para identificar as regiões que representam as substâncias no sensorgrama, e consequentemente identifica as áreas de transição entre elas, foi aplicado a regressão linear para estimar as regiões de transição e os regimes permanentes no sensorgrama.

B1 - Regressão Linear

Com a regressão linear é possível observar a mudança no valor do coeficiente de regressão, com o coeficiente igual a 0 é dado como um regime permanente, caso for diferente é um regime de transição. Em seguida, é possível identificar a ordem temporal e a quantidade de substâncias que um sensorgrama apresenta. Na Figura 22 é apresentado os valores estimados com o uso da regressão linear, em que é possível observar os valores da área de regime permanente como também os de transição.

Figura 22 – Sensorgrama para o protocolo $H_2O \rightarrow$ Etanol 25% $\rightarrow H_2O \rightarrow$ Etanol 12,5% $\rightarrow H_2O \rightarrow$ BSA-Monolayer $\rightarrow H_2O \rightarrow$ Etanol 25% $\rightarrow H_2O$ (azul). Valores estimados da regressão do regime permanente (vermelho). Valores estimados da regressão do regime de transição (preto).



Fonte: Autoria Própria (2020).

B2 - Teste de Hipótese

Foi aplicado o teste de hipótese na resposta do sensor para aferir a qualidade do sensorgrama gerado e da identificação das substâncias. O teste de hipótese foi aplicado para o coeficiente de regressão. Dessa forma, o teste verifica se na janela temporal estabelecida o coeficiente de regressão está contido no intervalo de confiança aceitável para atestar a qualidade do sensor SPR. Assim, foi testado se o sensorgrama e suas regiões/substâncias refletem o protocolo experimental com 99.9% de confiança. O nível de confiança define os limites superior e inferior do intervalo de confiança.

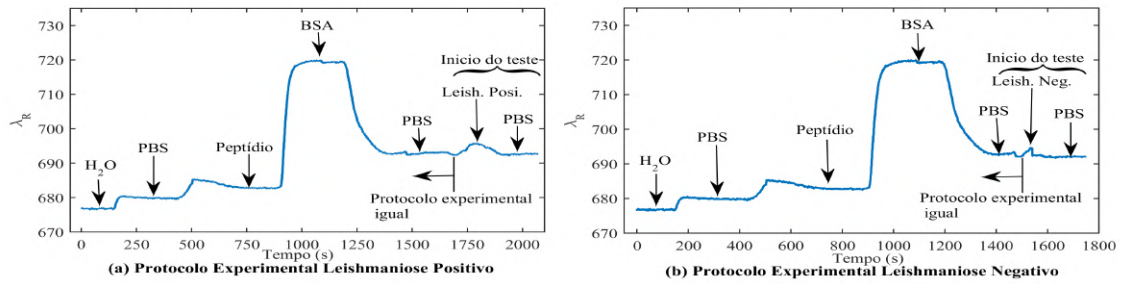
B3 - Teste Automático

Por fim, foi utilizado o teste de hipótese com duas amostras, em que foi utilizado o teste paramétrico t de *Student* e o não paramétrico *bootstrap* para um teste automático com os protocolos experimentais da Leishmaniose e Dengue, os sensorgramas para o diagnóstico dessas duas doenças se encontram na referência (FERREIRA *et al.*, 2017); (LOUREIRO *et*

al., 2017). Esses testes permitem adquirir uma maior confiabilidade nas respostas fornecidas pelo sensor, assim permite-se aferir se um teste é positivo ou não. A Leishmaniose representa uma doença causada por um protozoário do gênero *Leishmani*, em que afeta o homem e diferentes espécies animais, selvagens e domésticas, sendo transmitidas por diferentes espécies de mosquitos e com um amplo espectro de manifestações clínicas, no qual é causada pela espécie *Leishmania Donovan*i. A Leishmaniose visceral levantou preocupações pela organização mundial de saúde devido à uma grande disseminação desta doença nas últimas décadas (ORGANIZATION, 2020b). O vírus da Dengue é uma ameaça persistente em regiões tropicais e subtropicais, no qual os casos aumentam em quadros chuvosos. Esse vírus pode levar a alguns sintomas que podem ser perceptíveis aos profissionais de saúde e alguns que estão de certa forma imperceptíveis. Em 2019 a Organização Mundial de Saúde (WHO - *World Health Organization*) registrou um aumento de surtos, sendo denominado de uma ameaça potencial para essas regiões (ORGANIZATION, 2020a). Para diagnosticar o vírus da Dengue requer uma infraestrutura laboratorial bastante existente como um pessoal treinado, dessa forma o sensor SPR com o PPBIO é usado para monitorar em tempo-real um diagnóstico de dengue (LOUREIRO *et al.*, 2017).

Na Figura 23 é apresentado o protocolo experimental da Leishmaniose com o sensor SPR e PPBIO, foi utilizado um antígeno LC2 em que é um peptídeo sintético, altamente específico e sensível, derivado de um mimótipo de *Leishmania chagasi*. Para o protocolo experimental de detecção da leishmaniose positivo e negativo foi utilizado o 1000 μ L de peptídeo diluído em PBS e como também a mesma quantidade para as outras substâncias do protocolo experimental (BSA, Leishmaniose negativo e positivo diluído em PBS). No tempo 1700s se inicia a inserção da substância positivo da Leishmaniose, conforme apresenta na Figura 23 (a). Na Figura 23 (b) no tempo 1500s se inicia a inserção do teste negativo de leishmaniose, antes do início do teste é inserido o mesmo protocolo experimental.

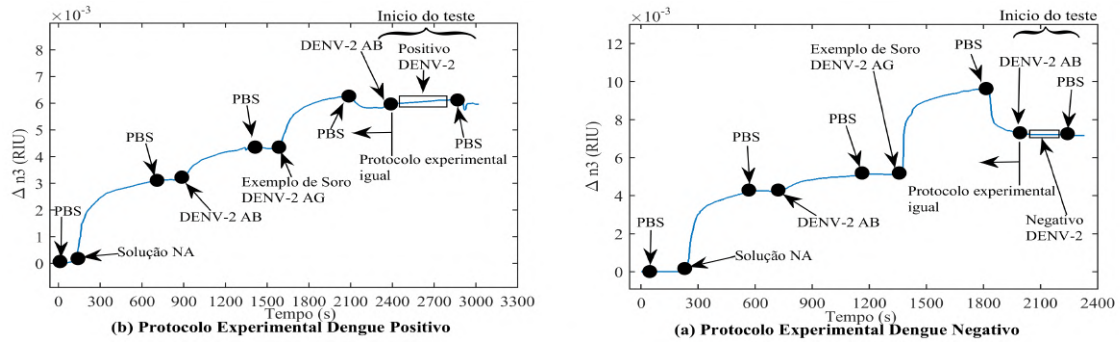
Figura 23 – Protocolo experimental para detecção de Leishmaniose positivo e negativo. O protocolo (a) inicial é inserido a H_2O , em seguida inserido o PBS; depois com a inserção do peptídio, em seguida a solução BSA; retornando para o PBS, depois a inserção da solução leishmaniose positiva e lavagem para o PBS. (b) inicializou com H_2O , em seguida inserido o PBS; depois a inserção do peptídio; em seguida a solução BSA; retornando para o PBS, depois a inserção da solução leishmaniose negativa e lavagem com o PBS.



Fonte: Autoria Própria (2020).

O protocolo experimental para detecção da dengue compreende a ligação do anticorpo monoclonal mediado por Neutravidin-Biotina (NA). A formação do complexo Dengue-Ag resulta em uma alteração acentuada da espessura, em que é gravada opticamente em tempo real. O protocolo experimental do teste negativo e positivo da Dengue é apresentado na Figura 24, em que é seguido por PBS, inserido depois a solução Neutravidin (NA), depois é inserido a solução DENV-2 AB para ser admitida e anexada à camada de NA, lavagem com PBS, depois é admitido soro diluído, mais uma vez para a lavagem de PBS, após começa o teste dengue negativo para o teste negativo, por último realizando a lavagem na superfície do metal. O protocolo da dengue para positivo é muito parecido com o anterior, porém o que muda é a ingestão da composição da dengue positiva em que ocorre uma alteração na detecção da dengue. Na Figura 24 (a) no tempo 2400s se inicia a inserção do teste positivo, já na Figura 24 (b) no tempo 2000s se inicia a inserção da substância dengue negativo, observa-se que há uma distinção entre as substâncias nos testes positivo e negativo da Dengue.

Figura 24 – Protocolo experimental para detecção de dengue negativo e positivo. O protocolo (a) inicial é inserido PBS, em seguida inserido a solução NA; depois com a inserção de PBS novamente; em seguida a solução DENV-2 AB; retornando para o PBS; depois a inserção de um Exemplo de Soro; retornando para PBS; em seguida inseriu o teste positivo; e ao final a inserção de PBS. (b) inicializado com PBS, em seguida inserido a solução NA; depois com a inserção de PBS novamente; em seguida a solução DENV-2 AB; retornando para o PBS; depois a inserção de um Exemplo de Soro; retornando para PBS; em seguida inseriu o teste negativo; e ao final a inserção de PBS.



Fonte: Autoria Própria (2020).

4 Resultados

Este capítulo apresenta os resultados deste trabalho, diante das etapas apresentada na metodologia da Seção 3, em que consiste em descrever, classificar e identificar as respostas do sensorgrama. Está organizado da seguinte forma, na Seção 4.1 é apresentado o Descritor Temporal de Sensorgrama. A seção 4.2, são apresentados os resultados no modo AIM. Na Seção 4.3, são apresentados os resultados no modo WIM e ao final na Seção 4.4 são apresentados os resultados para identificação e análise das respostas obtidas pelo sensorgrama.

4.1 Descritor Temporal de Sensorgrama

O Descritor Temporal de Sensorgrama (TSD - *Temporal Sensorgram Descriptor*) foi criado para representar o sensorgrama. Teve duas finalidades, a primeira foi voltada para encontrar os regimes permanentes e de transição. Para detectar os regimes, o algoritmo 1 apresenta o pseudocódigo em que a variável ϵ é um erro atribuído, θ são os valores do sensorgrama, α e β valor do sensorgrama com o erro em uma janela. Dessa forma, o algoritmo realiza uma comparação entre a variável do sensorgrama θ com as variáveis α e β . Caso o valor esteja dentro dessa janela com o erro, o valor é armazenado em um vetor de regimes permanentes, caso contrário é armazenado em um vetor de transição. Ao final, é retornado os vetores com os regimes permanentes e de transição.

Algoritmo 1: Pseudocódigo para encontrar os regimes permanentes e transição do sensorgrama.

Result: $R = \{ \nu, \sigma \}$

```

1  $\epsilon$  = valor de erro;
2 for  $i \leftarrow 1$  to  $\theta$  do
3    $\alpha = \theta_{i+4} + \epsilon$ ;
4    $\beta = \theta_{i-4} - \epsilon$ ;
5   if  $\theta \leq \alpha$  and  $\theta \geq \beta$  then
6      $\nu$  = armazena os valores de regime permanente;
7   else
8      $\sigma$  = armazena os valores de regime de transição;
9   end
10 end
```

Já a segunda finalidade é para calcular a média e o comprimento do regime permanente. O algoritmo 2 descreve a segunda parte do pseudocódigo do TSD. O pseudocódigo é inicializado com duas variáveis de incrementação. O parâmetro σ denota os valores de

transição e ν os valores dos regimes permanentes armazenado em uma auxiliar. Para percorrer os dados do sensorgrama, foi feita uma estrutura de repetição que permite comparar o valor do sensorgrama. Na comparação é verificado se o valor do sensorgrama é um valor de transição, caso o valor do sensorgrama seja de transição é feita uma contagem, em que nessa contagem o contador verifica se contém mais de 4 valores de transição. O valor de j corresponde a quantidade de dados do regime permanente. Sendo diferente de 0, os valores do regime permanente são armazenados em uma variável α . Caso o valor do valor do sensorgrama não seja de transição, os valores do regime permanente são armazenados em uma outra variável. Ao final, o valor da variável α a média e o comprimento do regime permanente são calculados e armazenados em um conjunto final R .

Algoritmo 2: Pseudocódigo do *Temporal Sensorgram Descript - TSD*

Result: $R = \{\alpha_1 + \alpha_1 i, \alpha_2 + \alpha_2 i, \dots, \alpha_n + \alpha_n i\}$

```

1  {Inicialização (i)}
2  {Inicialização (j)}
3   $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$ 
4   $V = \{\nu_1, \nu_2, \dots, \nu_n\}$ 
5  for  $k \leftarrow 1$  to  $\sigma$  do
6      if  $\sigma == V$  then
7           $\iota \leftarrow$  Contador;
8          if  $\iota > 4$  then
9              if  $j \neq 1$  then
10                  $\alpha(i) \leftarrow$  Armazena os valores do sensorgrama em uma variável
11                    auxiliar, e o vetor em que está os dados do sensorgrama é passado
12                    como vazio;
13                 end
14                 O algoritmo continua ser executado enquanto não encontra um valor de
15                 transição;
16             end
17         else
18              $V(j) \leftarrow$  Armazena os valores do sensorgrama em um vetor e realiza a
19                 contagem dos dados do regime permanente;
20         end
21     if  $\alpha > 1$  then
22          $R(i) \leftarrow$  Realiza o calculo da média e o seu comprimento armazenando em
23             um número complexo;
24     end
25 end

```

Para calcular a média de cada regime permanente, foi utilizada a Equação 4.1. A

Equação 4.1 apresenta a média, no qual n é o tamanho da amostra e x_i os valores do regime permanente. A média indica o centro de gravidade da distribuição de probabilidade.

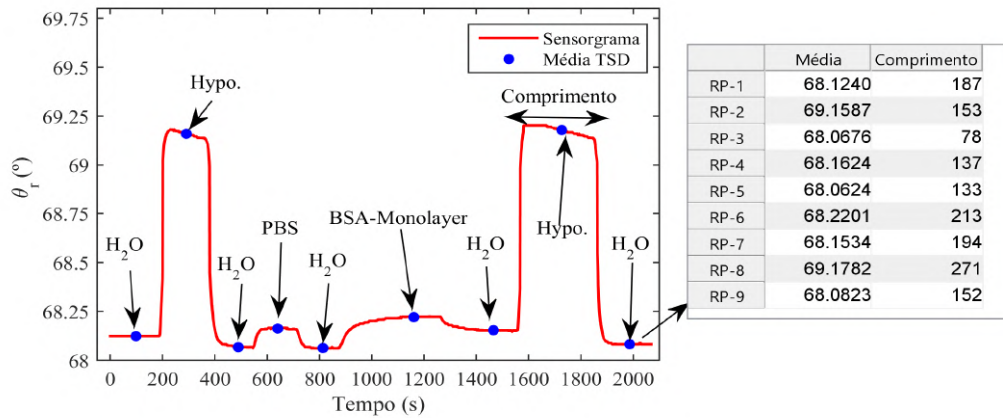
$$\bar{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.1)$$

O comprimento do regime permanente é calculado pela Equação 4.2, em que x_i corresponde a quantidade de observações daquele regime.

$$\bar{L} = x_i, \quad i = 0, 1, 2, \dots, n - 1 \quad (4.2)$$

Uma vez que um sensorgrama representa um protocolo experimental bem definido, isso é, a quantidade de substâncias (regimes permanentes) é conhecido. O algoritmo é composto por um número complexo, em que a parte real é a média do regime permanente e parte imaginária é o seu comprimento do regime. Na Figura 25 é ilustrado a aplicação do descritor TSD em um sensorgrama que corresponde a estrutura multicamada BK7/Au. É observado na Figura 25 uma tabela que representa os regimes permanentes do sensorgrama com a média e o comprimento desse regime.

Figura 25 – Aplicação do algoritmo TSD em que aplicou-se no protocolo experimental $H_2O \rightarrow Hypo. \rightarrow H_2O \rightarrow PBS \rightarrow H_2O \rightarrow BSA \rightarrow H_2O \rightarrow Hypo. \rightarrow H_2O$. A média de cada substância inserida no sensorgrama é representada no ponto azul. A tabela é composta pela média e o comprimento do protocolo experimental.



Fonte: Autoria Própria (2020).

4.2 Modo de Interrogação Angular (AIM)

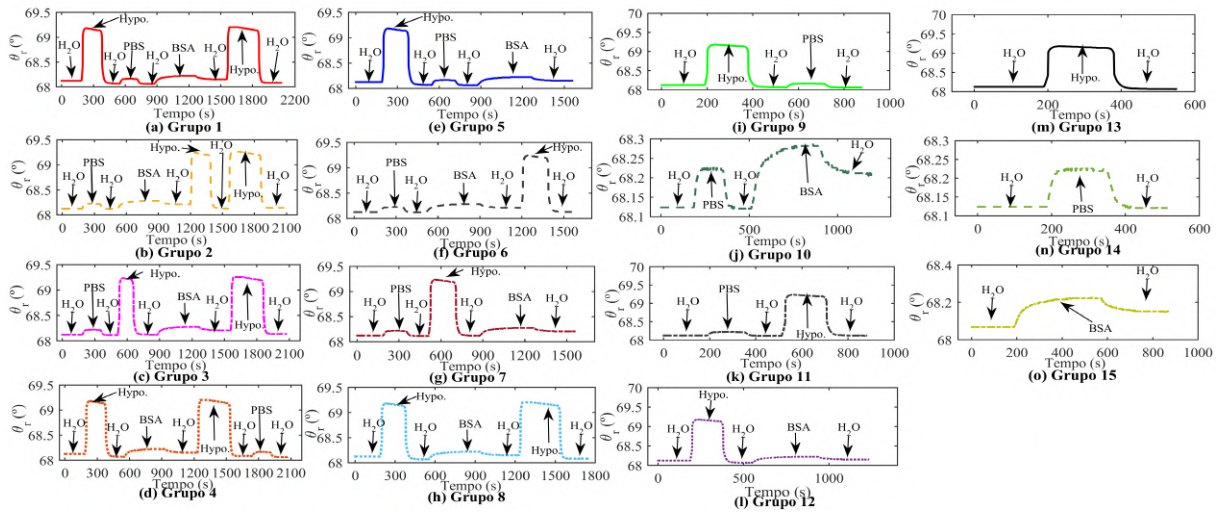
Após a aquisição dos sensorgramas, foi alterada a ordem do protocolo experimental para o vidro e cada polímero e também a inserção de quantidade de substâncias inseridas,

de tal forma que fosse criado 15 grupos de sensorgramas em cada vidro e polímero, com um total de 99 sensorgramas com 3, 5, 7 e 9 regimes permanentes.

Protocolo Experimental com o Vidro BK7 e o metal ouro (BK7/Au)

Com o protocolo experimental BK7/Au no modo AIM, foram criados 15 grupos de sensorgramas com 3, 5, 7 e 9 regimes permanentes. Em cada um dos sensorgramas, foi inserido uma quantidade e sequência temporal diferente no protocolo experimental. As substâncias para criação dos protocolos experimentais foram: H_2O , Hipoclorito (Hypo.), BSA-Monolayer e PBS (*Phosphate Buffered Saline*). Na Figura 26 são apresentados os 15 grupos de sensorgramas, é possível observar na Figura 26 as mudanças nas posições e na quantidade de substâncias inseridas, que permitiu criar sensorgramas com 3, 5, 7 e 9 regimes permanentes.

Figura 26 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o vidro BK7 com o metal ouro BK7/Au. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.

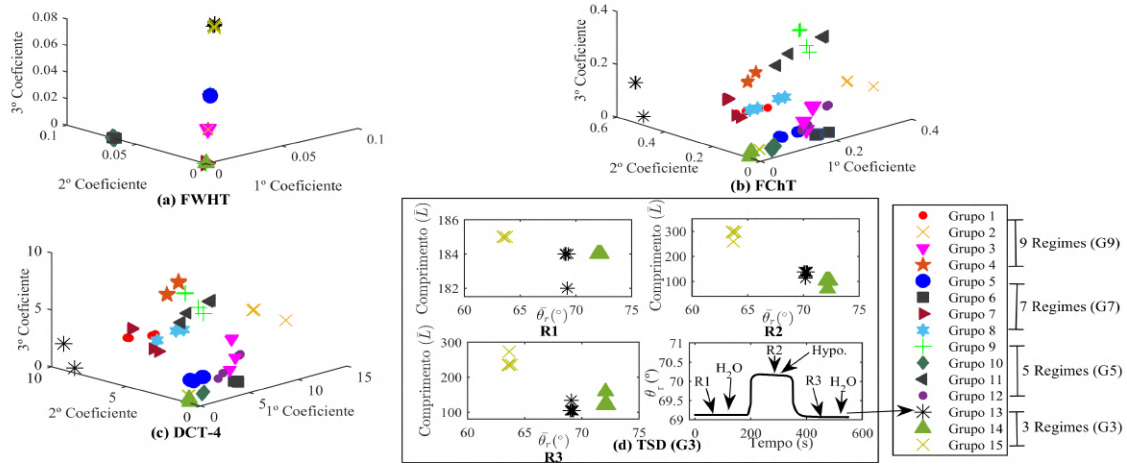


Fonte: Autoria Própria (2020).

Foi aplicado a geração de sensorgramas para criar os novos elementos para cada um dos grupos de sensorgramas, criando assim 99. Depois de realizar a geração de sensorgramas, foi aplicado quatro descritores em cada um dos sensorgramas pertencentes a cada grupo e foi extraído os seus coeficientes. A Figura 27 (a), (b) e (c) apresenta os três primeiros coeficientes de nível AC em que contém 99 sensorgramas com os descritores FWHT, FChT e DCT-4. É observado que os dados dos coeficientes de nível AC do descritor FWHT estão muito próximos entre si, já em relação aos descritores FChT e DCT-4 têm-se o espalhamento dos dados. O conjunto de dados de treinamento do descritor TSD com os três primeiros regimes permanentes, que corresponde aos grupos 13, 14 e 15 é apresentado

na Figura 27 (d) no gráfico do tipo 2D, cada sensorgrama é composto por um número complexo. No eixo x (parte real) é a média do regime permanente e no eixo y (parte imaginária) é o seu comprimento, é possível observar que os grupos de sensorgrama com 3 regimes permanentes estão um pouco distantes.

Figura 27 – Conjunto de dados de treinamentos em um gráfico 3D com os 99 sensorgramas dos 15 grupos para a (a) FWHT, (b) FChT e (c) DCT-4. (d) Conjunto de dados de treinamento com 3 regimes permanentes em um gráfico 2D do TSD.



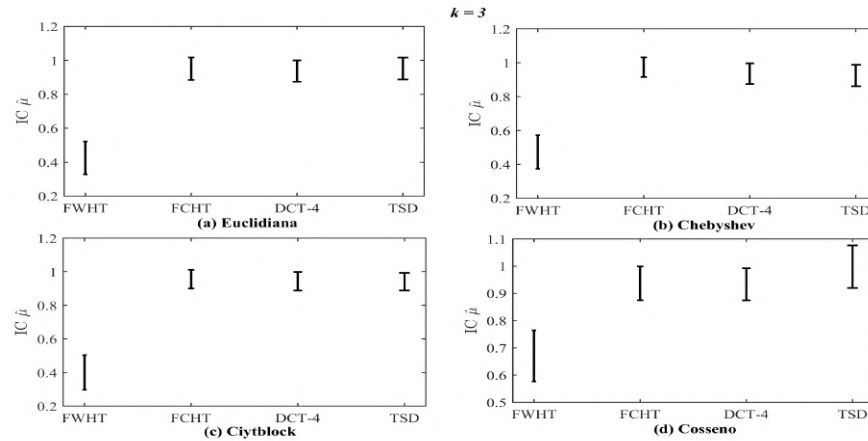
Fonte: Autoria Própria (2020).

Depois de criado os conjuntos de dados de treinamento com os coeficientes de nível AC dos descritores FWHT, FChT e a DCT-4, como também foi feito o conjunto com a média e o comprimento dos regimes permanentes do descritor TSD, foi atribuído o dado de destino para cada sensorgrama pertencente. Para classificar as respostas fornecidas pelo algoritmo k -NN, foi obtido a sua acurácia com a vizinhança $k = 3$, no qual, utilizou 99 sensorgramas com os regimes permanentes (3, 5, 7 e 9) e com o $k = 5$ vizinhos mais próximos foi utilizado 84 sensorgramas, com os regimes permanentes (5, 7 e 9). Com o $k = 3$ e $k = 5$ foi utilizada as distâncias euclidiana, *chebychev*, *cityblock*, e cosseno com a validação cruzada *Leave-one-out*. Para saber qual o melhor descritor, foi utilizado o teste estatístico não-paramétrico *bootstrap* na acurácia da validação cruzada. Para o teste *bootstrap* foi aplicado valores de replicações igual a 2000 (número de réplicas *bootstrap*) e $\alpha = 5\%$ (erro de confiança).

Na Figura 28 é apresentado os intervalos de confiança da acurácia obtidos da validação cruzada *Leave-one-out* com o $k = 3$ para cada descritor, e com as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno. Na Figura 28 é observado que o descritor FWHT apresentou o pior desempenho com as distâncias utilizadas, diante que o seu intervalo de confiança é menor do que os demais descritores, desse modo a sua classificação não é estável. Os descritores FChT, DCT-4 e TSD com as distâncias euclidiana, *chebyshev* e *cityblock* são estatisticamente iguais, porém o descritor TSD conduz a um resultado

melhor do que os descritores FChT e DCT-4, por possuir uma amplitude no seu intervalo menor, ou seja, a classificação obtida com o *Leave-one-out* para o descritor TSD obtém o seu resultado próximo a média. Com a distância cosseno, o descritor TSD obteve um resultado melhor que os descritores FChT e DCT-4, visto que o seu intervalo de confiança é superior, já os descritores FChT e DCT-4, são estatisticamente iguais.

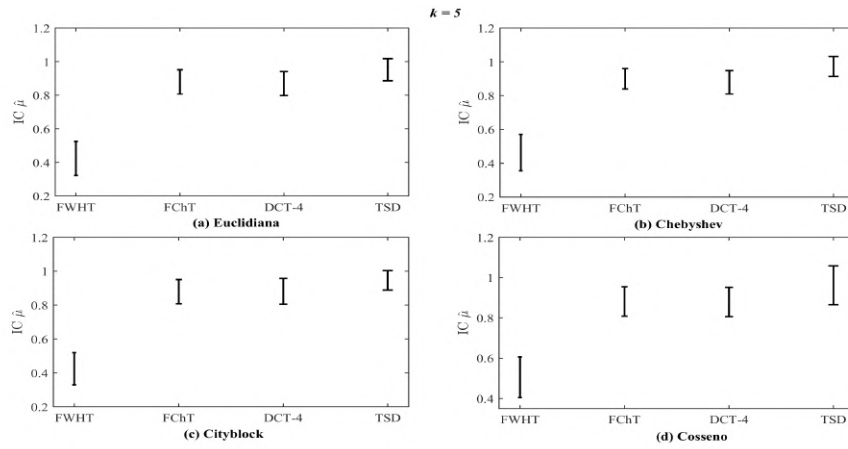
Figura 28 – Resultados dos intervalos de confiança com o teste não paramétrico *bootstrap* para o $k = 3$, com o vidro BK7 e o metal ouro (BK7/Au).



Fonte: Autoria Própria (2020).

Na Figura 29 é exposto os intervalos de confiança para o $k = 5$ com os descritores e as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno. Percebe-se na Figura 29 que o descritor FWHT com as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno, apresentou o pior desempenho, o seu intervalo é menor do que os demais descritores, assim não obtém um classificação estável. Os descritores FChT, DCT-4 e TSD são estatisticamente iguais com as distâncias euclidiana, *chebyshev*, *cityblock*, porém o descritor TSD com essas distâncias apresenta um resultado melhor, diante que a amplitude do intervalo de confiança é menor, ou seja, o resultado é próximo a média. Os intervalos de confiança com a distância cosseno, o descritor TSD obteve o melhor resultado entre os descritores FChT e DCT-4, visto que o seu intervalo de confiança é superior e os descritores FChT e DCT-4 são estatisticamente iguais.

Figura 29 – Resultados dos intervalos de confiança com o teste não paramétrico *bootstrap* para o $k = 5$, com o vidro BK7 e o metal ouro (BK7/Au).

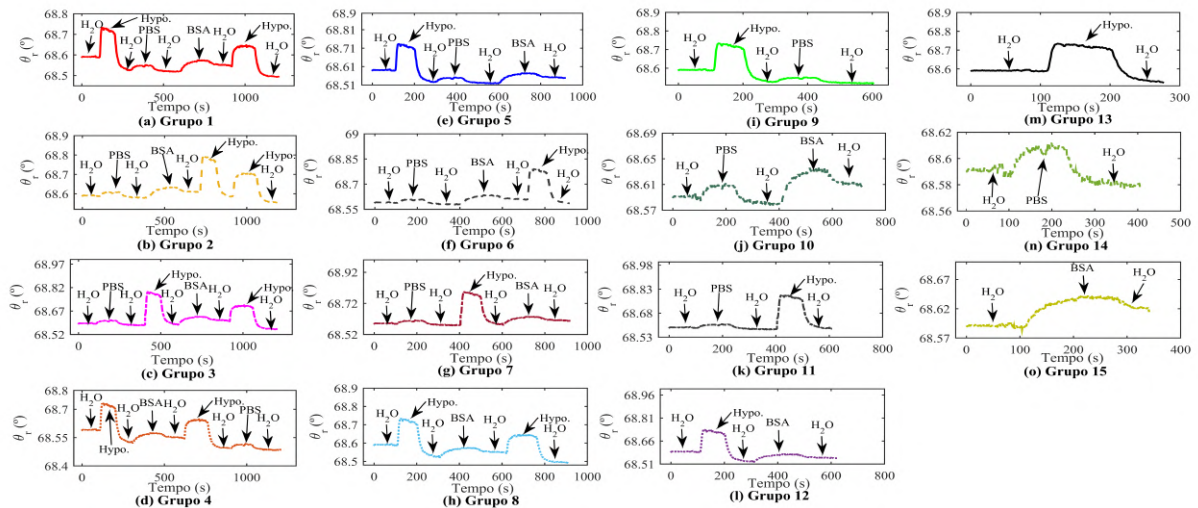


Fonte: Autoria Própria (2020).

Protocolo Experimental com o Polímero PC e o Metal Ouro PC/Au

Para o protocolo experimental PC/Au no modo AIM foram criados 15 grupos de sensorgramas, em que seguiu o mesmo protocolo experimental. Em cada um dos protocolos experimentais, foi inserido uma quantidade e sequência temporal diferente na inserção das substâncias. Na Figura 30 são apresentados os 15 grupos de sensorgramas com o polímero PC/Au com 3, 5, 7 e 9 regimes permanentes, percebe-se no protocolo experimental as mudanças nas posições e na quantidade de substâncias inseridas.

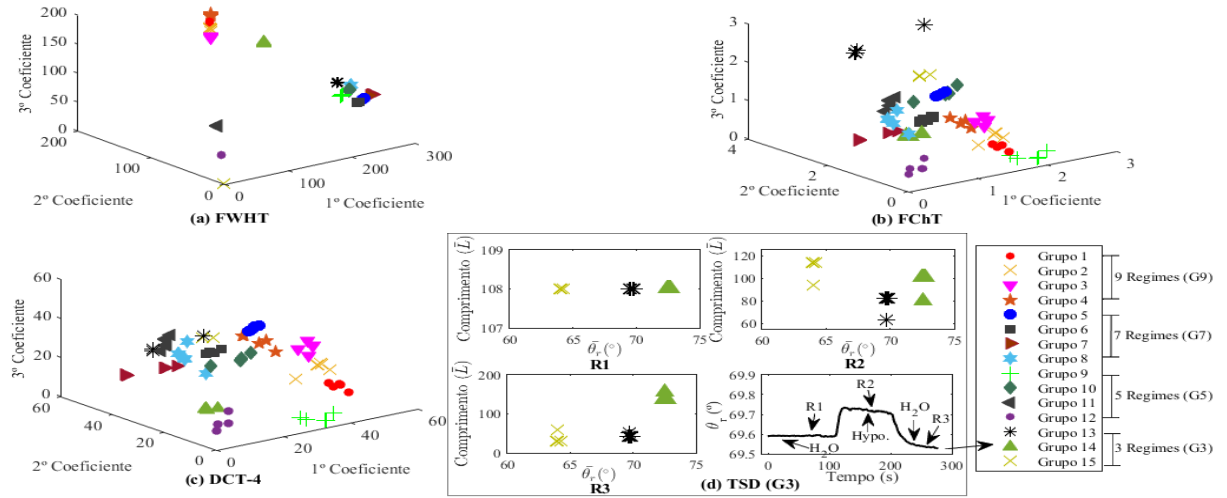
Figura 30 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o polímero PC com o metal ouro PC/Au. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.



Fonte: Autoria Própria (2020).

Também foram gerados novos sensorgramas para criar os novos elementos para cada um dos grupos, criando assim 99 sensorgramas. Depois de criado, foi aplicado os quatro descritores FWHT, FChT, DCT-4 e TSD em cada um dos sensorgramas pertencentes e foram criados dois conjuntos de dados de treinamento para classificação. O primeiro conjunto de dados corresponde aos sensorgramas com 3, 5, 7 e 9 regimes permanentes. Para os descritores FWHT, FChT e DCT-4, foi usado os três primeiros coeficientes de nível AC e para o descritor TSD foi utilizado os três primeiros regimes permanentes. Já o segundo conjunto de dados, corresponde aos cinco primeiros coeficientes de nível AC e o descritor TSD os cinco primeiros regimes permanentes. A Figura 31 (a), (b) e (c) apresenta os três primeiros coeficientes de nível AC em que contém 99 sensorgramas. Na Figura 31 (d) é apresentado o conjunto de dados de treinamento com os três primeiros regimes permanentes do descritor TSD, no qual corresponde aos grupos 13, 14 e 15.

Figura 31 – (a) a (c) apresenta o conjunto de dados de treinamentos com os 99 sensorgramas dos 15 grupos para os descritores FWHT, FChT e DCT-4. (d) retrata o descritor TSD no qual apresenta o conjunto de dados de treinamento com 3 regimes permanentes.

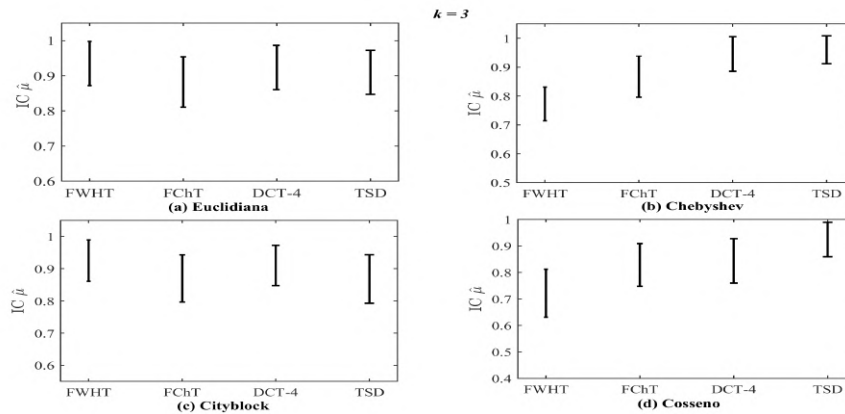


Fonte: Autoria Própria (2020).

Foi atribuído o dano de destino depois de criado o conjunto de dados de treinamento para os coeficientes de nível AC dos descritores FWHT, FChT e a DCT-4, como também para o conjunto com a média e o comprimento do regime permanente do descritor TSD. Para classificar as respostas do algoritmo k -NN, foi obtida a acurácia com o vizinho $k = 3$ para os 99 sensorgramas, com os regimes permanentes (3, 5, 7 e 9) e para o $k = 5$ foi utilizado 84 sensorgramas, com os regimes (5, 7 e 9). Para o $k = 3$ e $k = 5$ foi utilizada as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno, com a validação cruzada *Leave-one-out*. No fim, para saber qual o melhor descritor, foi utilizado o teste estatístico não paramétrico *bootstrap* na acurácia no método *Leave-one-out*. Para o *bootstrap* foi aplicado valores de replicações igual a 2000 (número de réplicas *bootstrap*) e $\alpha = 5\%$ (erro de confiança).

Os intervalos de confiança da acurácia com a validação cruzada *Leave-one-out* para o $k = 3$ é apresentado na Figura 32 com cada descritor. Na Figura 32 com as distâncias euclidiana e *cityblock* os descritores são estatisticamente iguais, porém os descritores FWHT e DCT-4 indicam a um desempenho melhor, diante que o seu intervalo de confiança é superior aos descritores FChT e TSD, como também os descritores FWHT e DCT-4 possui uma amplitude menor, ou seja, o resultado da classificação é próximo a média. Já os descritores FChT e TSD, são estatisticamente iguais para as distâncias euclidiana e *cityblock*. Na distância *chebyshev*, os descritores FWHT e FChT obtiveram o pior desempenho, diante que os intervalos de confiança apresentaram um resultado inferior aos descritores. Já os descritores DCT-4 e TSD, são estatisticamente iguais, porém o descritor TSD, obteve um intervalo de confiança menor, dessa forma, a sua classificação fica próxima a média. Com a distância cosseno, o descritor FWHT segue o mesmo para a distância *chebyshev*, e os descritores FChT e DCT-4 são estatisticamente iguais. Já o descritor TSD com a distância cosseno, teve um desempenho satisfatório, diante que o seu intervalo de confiança é superior aos demais.

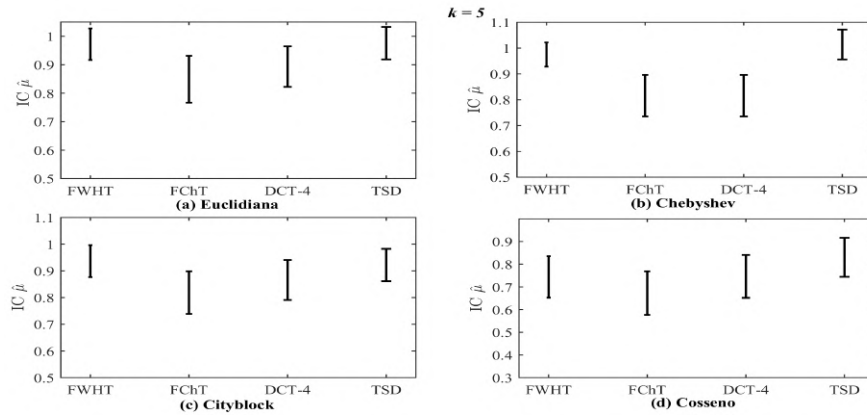
Figura 32 – Resultados dos intervalos de confiança do teste não paramétrico *bootstrap* com o polímero PC e o metal ouro (PC/Au) para o $k = 3$.



Fonte: Autoria Própria (2020).

Na Figura 33 apresentado os intervalos de confiança para o $k = 5$. Observa-se na Figura 33 que com as distâncias euclidiana, *cityblock* e cosseno, apresentaram os descritores FChT e DCT-4 os piores desempenho, visto que o resultado dos seus intervalos é menor em relação aos descritores FWHT e TSD, com as distância *euclidiana* e *chebyshev* são estatisticamente iguais. Para a distância *chebyshev* o descritor TSD é estatisticamente igual ao FWHT, porém o TSD indica um resultado melhor, diante que o seu intervalo de confiança superior e a amplitude é menor, desse modo, a sua classificação está próxima a média. Com a distância cosseno, os descritores são estatisticamente iguais, porém o descritor FChT obteve o pior desempenho, diante que o seu intervalo é inferior, já o TSD apresentou um intervalo de confiança menor e superior, ou seja, a classificação fica próxima a média.

Figura 33 – Resultados dos intervalos de confiança do teste não paramétrico *bootstrap* com o polímero PC e o metal ouro (PC/Au) para o $k = 5$.

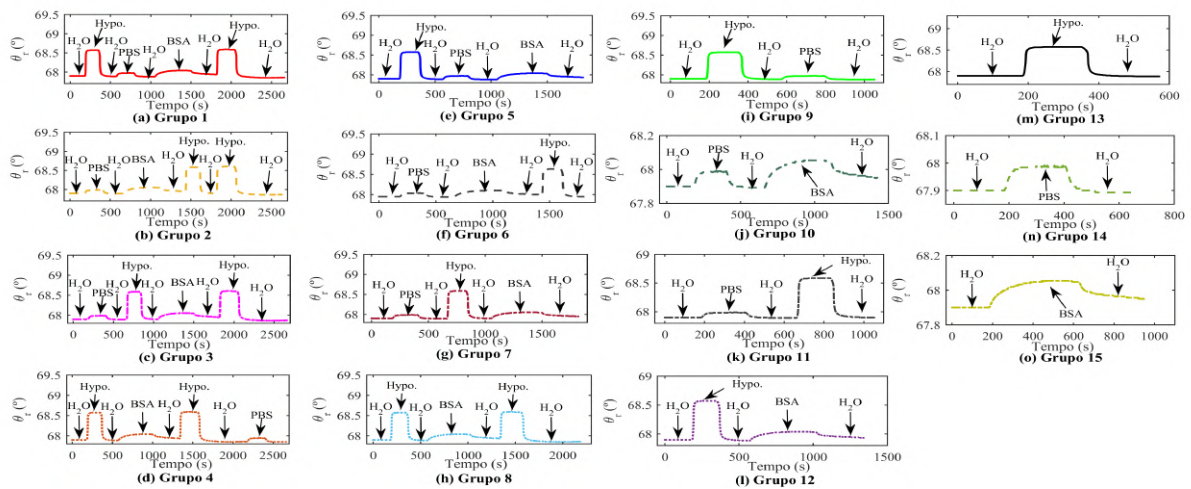


Fonte: Autoria Própria (2020).

Protocolo Experimental com o Polímero PMMA e o Metal Ouro PMMA/Au

Na Figura 34 são exibidos os grupos de sensorgramas do polímero PMMA/Au no modo AIM, com o mesmo protocolo experimental. Foi inserida uma quantidade e sequência temporal diferente na inserção das substâncias em cada protocolo experimental. Percebe-se na Figura 34 os grupos de sensorgramas com as mudanças nas posições e na quantidade de substâncias inseridas, em que se tem sensorgramas com 3, 5, 7 e 9 regimes permanentes.

Figura 34 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o polímero PMMA com o metal ouro PMMA/Au. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.



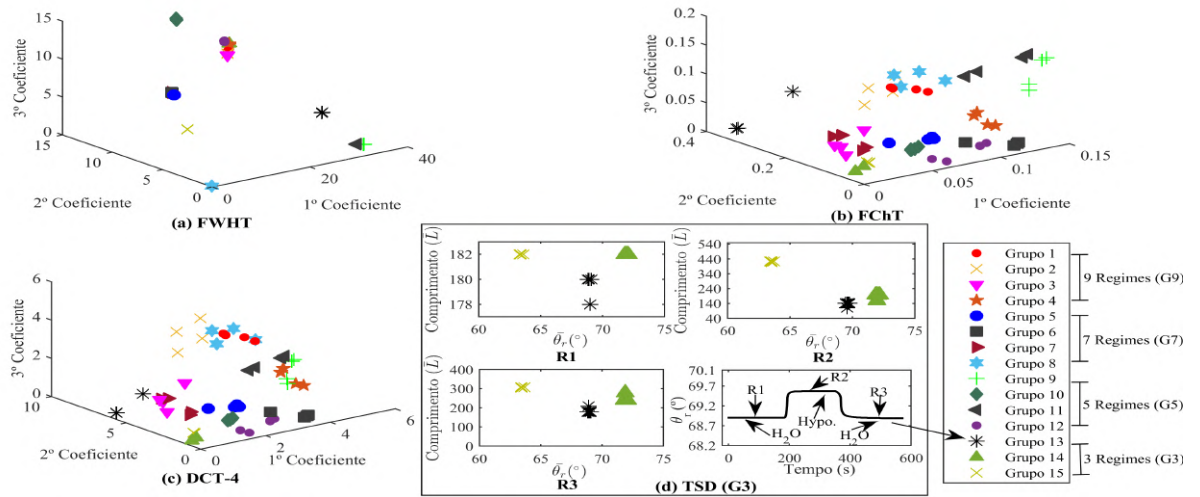
Fonte: Autoria Própria (2020).

Para criar novos elementos para cada um dos grupos, foi utilizada a geração de sensorgramas para criar 99. Depois de criado, foi aplicado quatro descritores FWHT,

FChT, DCT-4 e a TSD em cada um dos sensorgramas pertencentes a cada grupo. Com os coeficientes de nível AC para os descritores FWHT, FChT e DCT-4 e o TSD foi obtida a média e o comprimento do regime permanente. Foram criados dois conjuntos de dados, o primeiro corresponde aos sensorgramas com 3, 5, 7 e 9 regimes permanentes. Para esse conjunto, foi utilizado os descritores FWHT, FChT e DCT-4 e obtido os três primeiros coeficientes de nível AC e para o descritor TSD foi utilizado os três primeiros regimes permanentes para classificação. Já para o segundo conjunto, que corresponde os sensorgramas com 5, 7 e 9 regimes, foi utilizado os descritores FWHT, FChT e DCT-4 e obtido os cinco primeiros coeficientes de nível AC e para o descritor TSD foi utilizado os cinco primeiros regimes permanentes.

Na Figura 35 (a), (b) e (c) observa-se os três primeiros coeficientes do nível AC em um gráfico 3D com os 99 sensorgramas dos grupos 3, 5, 6 e 7 regimes permanentes. Percebe-se que os dados dos descritores FWHT, FChT e DCT-4, estão um pouco dispersos, porém próximos entre si. O conjunto de dados de treinamento do descritor TSD é apresentado na Figura 35 (d) em um gráfico 2D, é possível observar que os grupos com 3 regimes permanentes estão distantes.

Figura 35 – Conjunto de dados de treinamentos com os 99 sensorgramas dos 15 grupos para a FWHT e FChT, DCT-4 e o TSD.



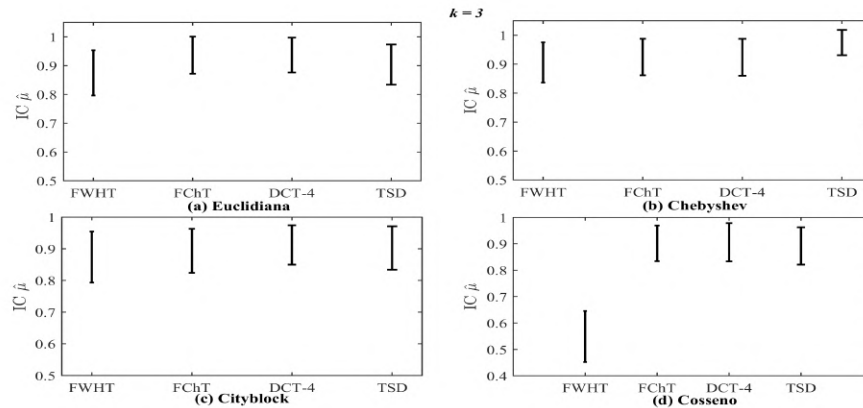
Fonte: Autoria Própria (2020).

Depois de criado os conjunto de dados de treinamento com os coeficientes dos descritores FWHT, FChT e DCT-4 e TSD, foi atribuído o dado de destino para cada sensorgramas pertencentes. Para verificar as respostas fornecidas pelo algoritmo k -NN, com o $k = 3$ foi utilizado 99 sensorgramas e para o $k = 5$ foi usado 84 sensorgramas, com as distâncias euclidiana, *cityblock*, *chebyshev* e cosseno e a validação cruzada *Leave-one-out*. Foi utilizado o teste estatístico não-paramétrico *bootstrap* na acurácia obtida pelo método *Leave-one-out*, para saber qual o melhor descritor. Para o teste *bootstrap* foi aplicado

valores de replicações igual a 2000 (número de réplicas *bootstrap*) e $\alpha = 5\%$ (erro de confiança).

Os intervalos de confiança para a acurácia com o $k = 3$ é apresentado na Figura 36, com cada descritor e distância. Observa-se na Figura 36, que os descritores nas distâncias euclidiana, *chebyshev* e *cityblock* são estatisticamente iguais. Na distância euclidiana os descritores FChT e DCT-4 apresentaram um resultado superior comprado aos descritores FWHT e TSD, devido que os seus intervalos de confiança são superiores e como também possui uma amplitude menor, sua classificação é estável. Para a distância *chebyshev* o descritor TSD teve um resultado superior em relação aos outros descritores, e a sua amplitude foi menor, ou seja, a resposta para a classificação é próxima a média. Na Figura 36 (d) com distância cosseno, o descritor FWHT apresentou o pior desempenho, diante que o intervalo de confiança é inferior aos demais descritores, a classificação ocasiona uma resposta errada. Já os descritores FChT, DCT-4 e TSD são estaticamente iguais, devido a sua amplitude ser igual.

Figura 36 – Resultados dos intervalos de confiança do teste não paramétrico *bootstrap* para o $k = 3$, com o polímero PMMA e o metal ouro (PMMA/Au).

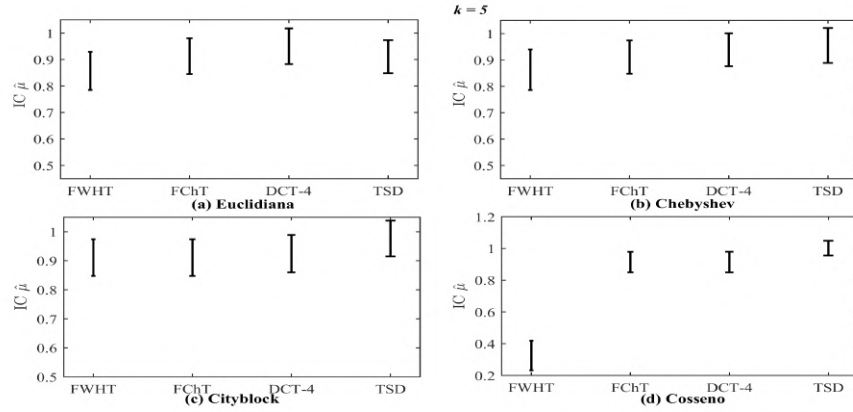


Fonte: Autoria Própria (2020).

Na Figura 37 é exposto os intervalos de confiança para o $k = 5$, obtido com a acurácia para cada descritor e distância. Percebe-se na Figura 37 com as distâncias euclidiana, *chebyshev* e *cityblock* os descritores são estaticamente iguais. Na distância euclidiana o descritor DCT-4 apresenta um desempenho melhor, diante que o seu intervalo de confiança é superior em relação aos outros descritores. Os descritores com o uso das distâncias *chebyshev* e *cityblock* são estatisticamente iguais. O descritor TSD indica um resultado melhor, devido que o seu resultado contém uma amplitude menor, então sua classificação está próxima a média. Já com a distância cosseno, o descritor FWHT apresentou um pior desempenho em relação aos demais descritores, visto que o seu resultado do intervalo de confiança foi inferior e os descritores FChT, DCT-4 e TSD são estaticamente iguais, no entanto o descritor TSD apresentou um desempenho superior, por possuir uma amplitude

menor, assim sua classificação fica próxima a média.

Figura 37 – Resultados dos intervalos de confiança do teste não paramétrico *bootstrap* para o $k = 5$, para o polímero PMMA e o metal ouro (PMMA/Au).

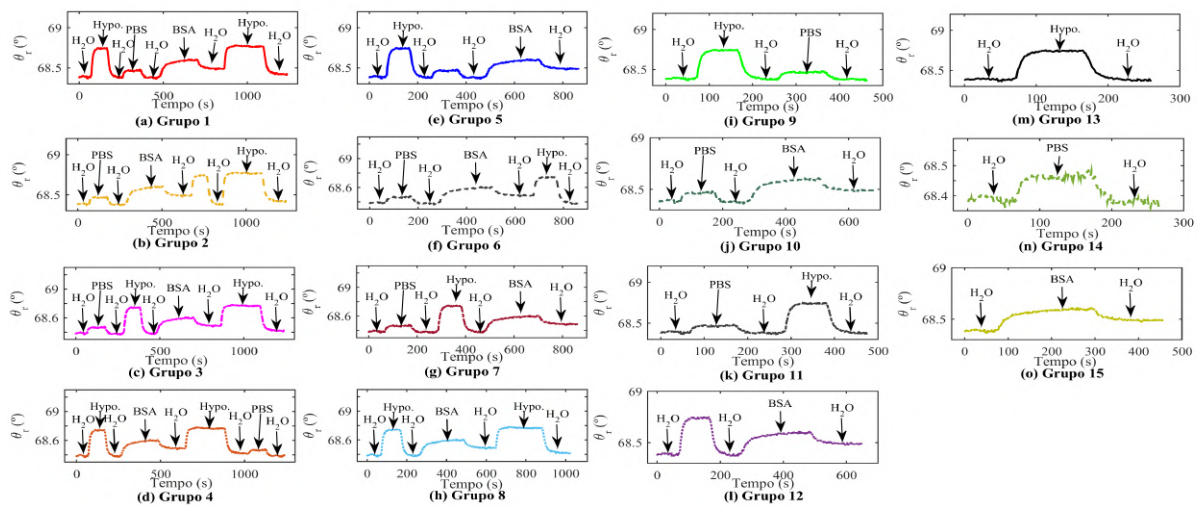


Fonte: Autoria Própria (2020).

Protocolo Experimental com o Polímero TOPAS e o Metal Ouro TOPAS/Au

No modo AIM para o protocolo experimental TOPAS/Au, foram criados 15 grupos de sensorgramas, em que seguiu o mesmo protocolo experimental. Em cada um dos protocolos experimentais, foi inserido uma quantidade e sequência temporal diferente na inserção das substâncias. É apresentado na Figura 38 o polímero TOPAS/Au com os protocolos experimentais 3, 5, 7 e 9 regimes permanentes. É possível observar nos protocolos experimentais as mudanças nas posições e na quantidade de substâncias inseridas.

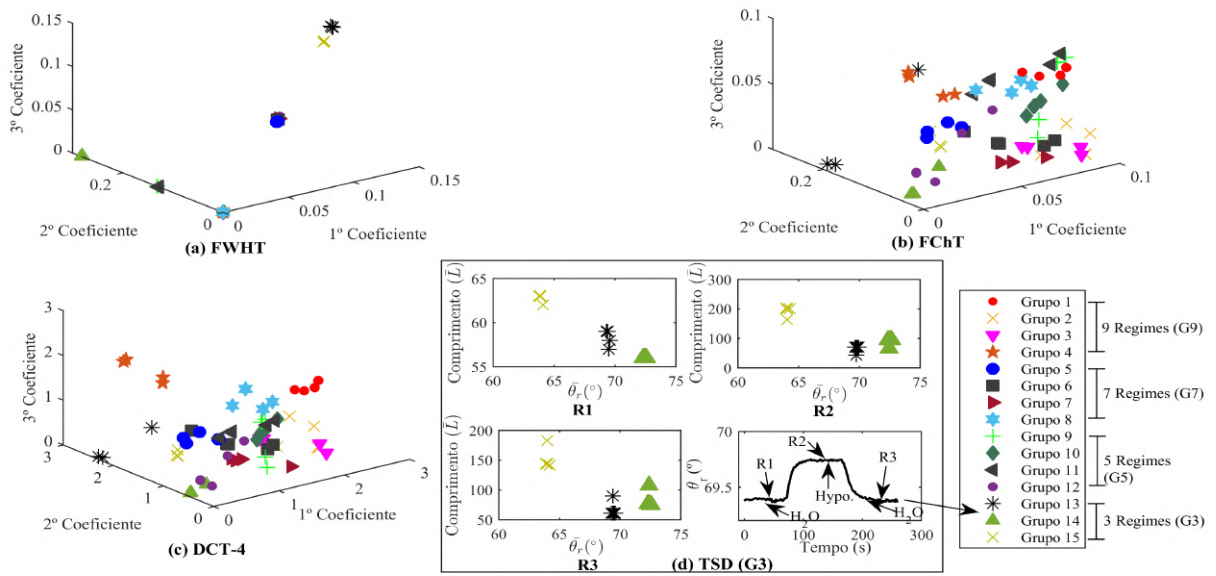
Figura 38 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o polímero TOPAS com o metal ouro TOPAS/Au. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.



Fonte: Autoria Própria (2020).

Para cada um dos grupos, foi utilizada a geração de sensorgramas para criar 99. Depois de criado, foi aplicado em cada um dos sensorgramas pertencentes a cada grupo, os descritores FWHT, FChT, DCT-4 e TSD. Com os coeficientes de nível AC para os descritores FWHT, FChT e DCT-4 e o TSD a média e o comprimento, foram criados dois conjuntos de dados de treinamento. Para o primeiro conjunto foi utilizado os três primeiros coeficientes de nível AC para os descritores FWHT, FChT e DCT-4 e o descritor TSD foi utilizado os três primeiros regimes permanentes, esse conjunto corresponde aos sensorgramas com 3, 5, 7 e 9 regimes permanentes. Já o segundo conjunto, corresponde os cinco primeiros coeficientes de nível AC dos descritores FWHT, FChT e DCT-4 e o TSD os cinco primeiros regimes permanentes. A Figura 39 (a), (b) e (c) apresenta os grupos de sensorgramas com os três primeiros coeficientes de nível AC. O conjunto de dados de treinamento com os grupos 13, 14 e 15 é apresentado na Figura 39 (d) com os três primeiros regimes permanentes.

Figura 39 – Conjunto de dados de treinamentos com os 99 sensorgramas dos 15 grupos para a FWHT e FChT, DCT-4 e o TSD.

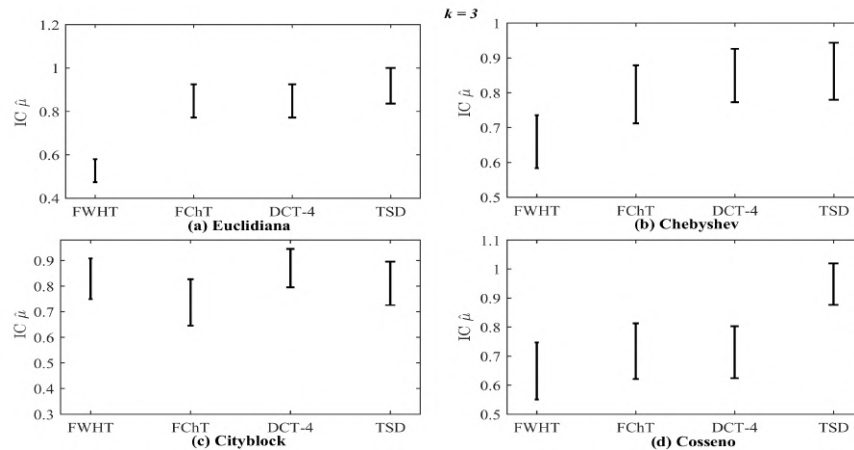


Fonte: Autoria Própria (2020).

Depois de criado os conjuntos de dados de treinamento para os descritores FWHT, FChT, DCT-4 e TSD, foi atribuído o seu dado de destino. Para classificar as respostas do algoritmo k -NN, foi obtido a acurácia com a vizinhança $k = 3$ para 99 sensorgramas e para o $k = 5$ foi utilizado 84 sensorgramas. Para o $k = 3$ e $k = 5$, foi utilizado as distâncias euclidiana, *cityblock*, *chebyshev* e cosseno com a validação cruzada *Leave-one-out*. Para saber qual o melhor descritor, foi utilizado teste estatístico não-paramétrico *bootstrap* em que utilizou replicações igual a 2000 (número de réplicas *bootstrap*) e $\alpha = 5\%$ (erro de confiança) na acurácia na validação cruzada.

Na Figura 40 apresenta os intervalos de confiança com as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno com o uso dos descritores obtidos a partir da validação cruzada *Leave-one-out*. O descritor FWHT apresentou o pior desempenho para as distâncias euclidiana e *chebyshev*, diante que os intervalos de confiança foi inferior, ou seja, o seu desvio padrão é menor em relação aos outros descritores. Os descritores FChT, DCT-4 e TSD com a distância euclidiana são estatisticamente iguais, porém o descritor TSD tem um desempenho satisfatório, diante que o seu intervalo de confiança é superior e a sua amplitude é menor, ou seja, sua classificação é próxima a média e o seu desvio padrão é menor. Os descritores FChT, DCT-4 e TSD com a distância *chebyshev*, são estatisticamente iguais. Para a distância *cityblock* todos os descritores são estatisticamente iguais, porém os descritores FWHT e DCT-4 apresentaram um desempenho superior aos descritores FChT e TSD. Os descritores FWHT, FChT e DCT-4 com a distância cosseno, são estatisticamente iguais, porém os resultados dos seus intervalos de confianças é inferior ao TSD, dessa forma, o TSD obtêm um resultado satisfatório.

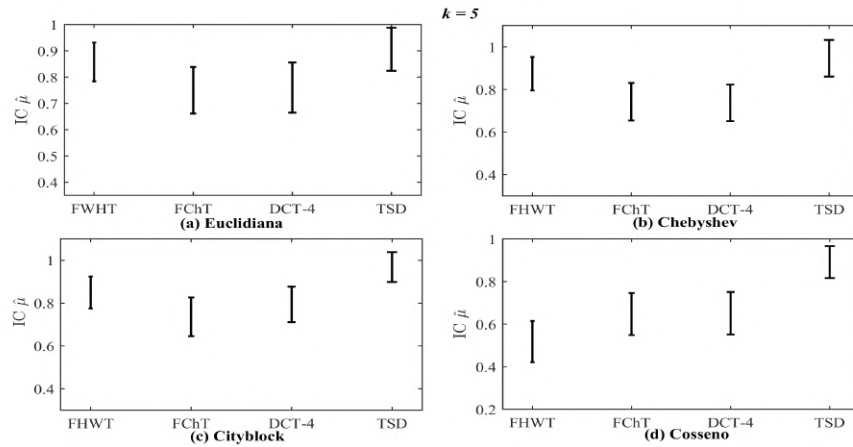
Figura 40 – Resultados dos intervalos de confiança com o polímero TOPAS e o metal ouro (TOPAS/Au), em que utilizou o teste não paramétrico *bootstrap* para o $k = 3$.



Fonte: Autoria Própria (2020).

Na Figura 41 é exposto os intervalos de confiança para o $k = 5$, obtidos da acurácia da validação cruzada LOO-CV com cada descritor e com as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno. É possível observar na Figura 41 com as distâncias euclidiana, *chebyshev* e *cityblock*, que os descritores são estatisticamente iguais, porém os descritores FWHT e TSD apresentam um desempenho melhor, devido que os seus intervalos de confiança serem superiores aos descritores FChT e DCT-4. Os descritores FWHT, FChT e DCT-4 com a distância cosseno, são estatisticamente iguais, porém obtiveram um pior desempenho em relação ao descritor TSD, ou seja, o descritor TSD teve um resultado superior, diante que o seu intervalo de confiança é maior.

Figura 41 – Resultados dos intervalos de confiança com o polímero TOPAS e o metal ouro (TOPAS/Au), em que utilizou o teste não paramétrico *bootstrap* para o $k = 5$.



Fonte: Autoria Própria (2020).

Com o uso do algoritmo k -NN e com a validação cruzada *Leave-one-out* para as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno para o vidro BK7 e os polímeros PC, PMMA e TOPAS, em que utilizou o metal ouro no modo AIM. O descritor TSD teve um bom desempenho para o $k = 3$ e $k = 5$ para o vidro BK7 e os polímeros PMMA e TOPAS com algumas distâncias. Já o descritor FWHT com o vidro BK7 e os polímeros, em alguns casos obteve o pior desempenho em relação aos descritores, devido que o seu conjunto de dados de treinamento está muito próximo o que pode ocasionar a sua classificação de forma errada. Porém o FWHT, obteve um bom resultado para o polímero PC e TOPAS com algumas distâncias, tanto com o uso do $k = 3$ e $k = 5$. Já os descritores FChT e DCT-4, são estatisticamente iguais com o uso do vidro BK7 e os polímeros PC, PMMA e TOPAS com o $k = 3$ e $k = 5$.

4.3 Modo Interrogação Comprimento de Onda (WIM)

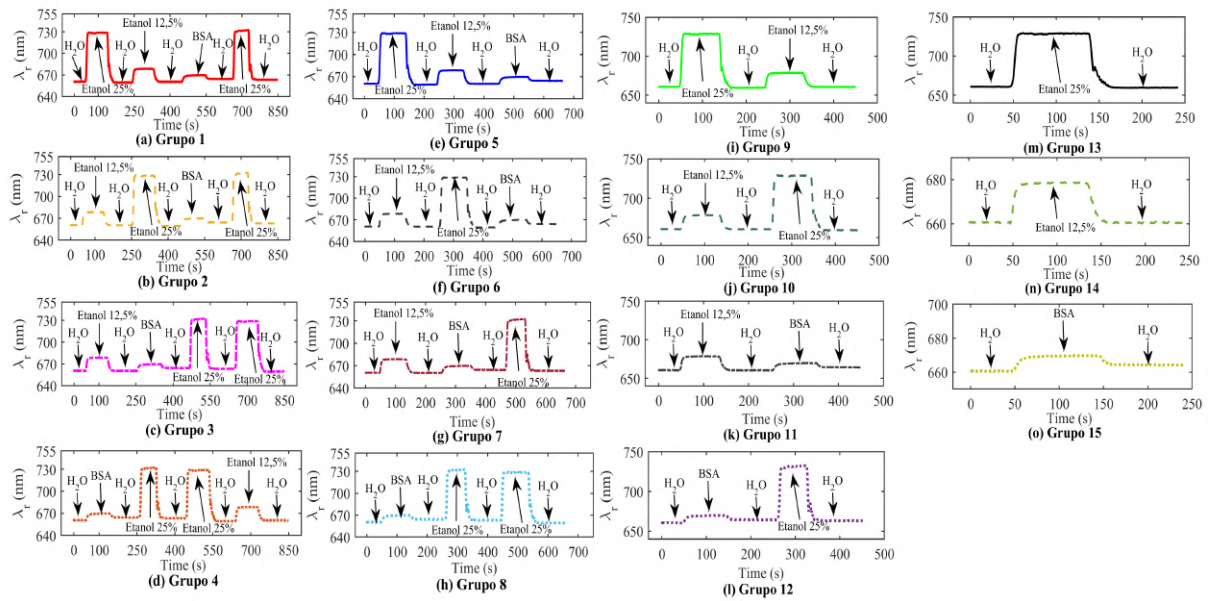
Posteriormente, no modo WIM após a aquisição dos sensorgramas, foi alterada a ordem do protocolo experimental para o vidro e cada polímero e também a inserção na quantidade de substâncias inseridas, de tal forma que fosse criado 15 grupos de sensorgramas em cada vidro e polímero, com um total de 99 sensorgramas com 3, 5, 7 e 9 regimes permanentes.

Protocolo Experimental com o Vidro BK7 e o metal prata (BK7/Ag)

Com o protocolo experimental BK7/Ag no modo WIM, foram criados 15 grupos de sensorgramas, com as substâncias H_2O , Etanol 25 %, Etanol 12,5% e BSA-Monolayer, com 3, 5, 7 e 9 regimes permanentes. Em cada um dos sensorgramas, foi inserido uma quantidade e sequência temporal diferente no protocolo experimental. Na Figura 42 são

apresentados os grupos de sensorgramas do vidro BK7/Ag, percebe-se que na Figura 42 houve mudanças nas posições do protocolo experimental e na quantidade de substâncias inseridas.

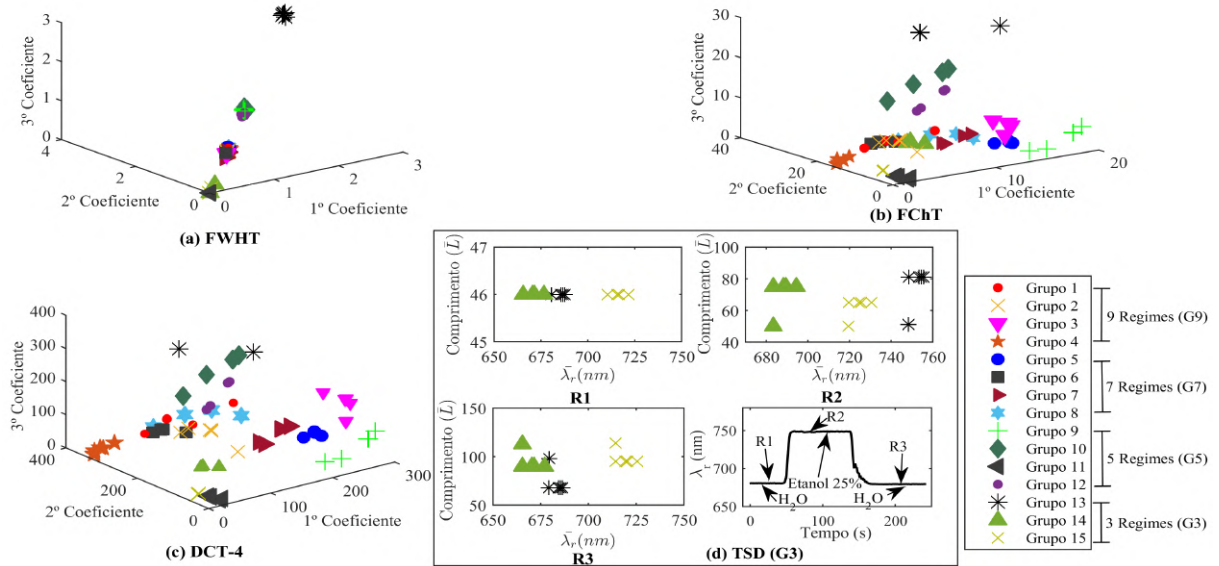
Figura 42 – Grupos de sensorgramas representando 15 protocolos experimentais diferentes utilizando o vidro BK7 com o metal prata (BK7/Ag). De (a) até (i) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.



Fonte: Autoria Própria (2020).

Foram criados novos elementos para cada um dos grupos de sensorgramas, assim foi criado 99. Depois de criado os sensorgramas, foi aplicado os descritores FWHT, FChT, DCT-4 e TSD em cada um dos sensorgramas pertencentes a cada grupo e foi extraído os seus coeficientes. Foram criado dois conjuntos de dados de treinamento para classificação, o primeiro conjunto foi extraído os valores dos três primeiros coeficientes de nível AC dos descritores FWHT, FChT e DCT-4 e para o descritor TSD foi utilizado os três primeiros regimes permanentes. Para esse conjunto de dados foram utilizado os sensorgramas com 3, 5, 7 e 9 regimes permanentes. O segundo conjunto de dados de treinamento corresponde aos cinco primeiros coeficientes de nível AC dos descritores FWHT, FChT e DCT-4 e os cinco primeiros regimes permanentes do descritor TSD. A Figura 43 (a), (b) e (c) apresenta os três primeiros coeficientes de nível AC. Na Figura 43 (d) é apresentado o conjunto de dados de treinamento com os três primeiros regimes permanentes dos grupos 13, 14 e 15.

Figura 43 – Conjunto de Dados de Treinamento em um gráfico com os 99 sensorgramas dos grupos para (a) FWHT, (b) FChT e (c) DCT-4. Conjunto de dados treinamento com 3 regimes permanentes em um gráfico 2D.

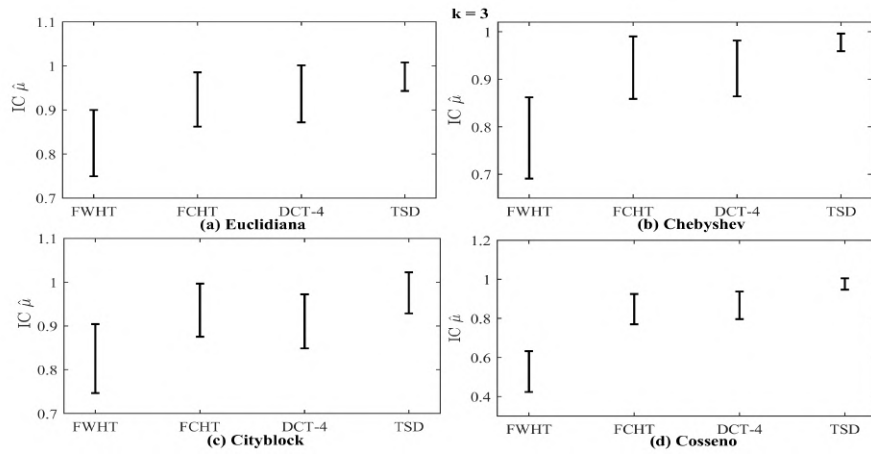


Fonte: Autoria Própria (2020).

Depois de criado os conjuntos de dados de treinamento com os coeficientes de nível AC dos descritores FWHT, FChT e a DCT-4, como também para os regimes permanentes do descritor TSD, foi atribuído o dado de destino. Para verificar as respostas fornecidas pelo o algoritmo k -NN para o $k = 3$, foi utilizado 99 sensorgramas e para o $k = 5$ 84 sensorgramas, com as distâncias euclidiana, *cityblock*, *chebyshev* e cosseno. Para cada distância foi utilizada a validação cruzada *Leave-one-out* e o teste estatístico não-paramétrico *bootstrap*, em que utilizou replicações igual a 2000 (número de réplicas *bootstrap*) e $\alpha = 5\%$ (erro de confiança) na acurácia na validação cruzada.

Os intervalos de confiança obtidos com a validação cruzada *Leave-one-out* e com a vizinhança $k = 3$ para cada descritor são apresentados na Figura 44. Na Figura 44 o descritor FWHT com as distâncias, apresentou o pior desempenho em relação ao descritores, o seu intervalo é inferior, ou seja, não obtém uma classificação estável. Já para os descritores FChT, DCT-4 e TSD com as distâncias euclidiana, *chebyshev*, *cityblock* são estatisticamente iguais, porém o descritor TSD apresentou um melhor desempenho, por possuir uma amplitude menor, ou seja, o resultado da classificação está próximo a média. Os descritores FChT e DCT-4 com a distância cosseno, são estatisticamente iguais, porém o TSD tem um resultado superior, devido que o seu intervalo de confiança é superior e também possui uma amplitude menor, dessa forma, o seu resultado está próximo a média.

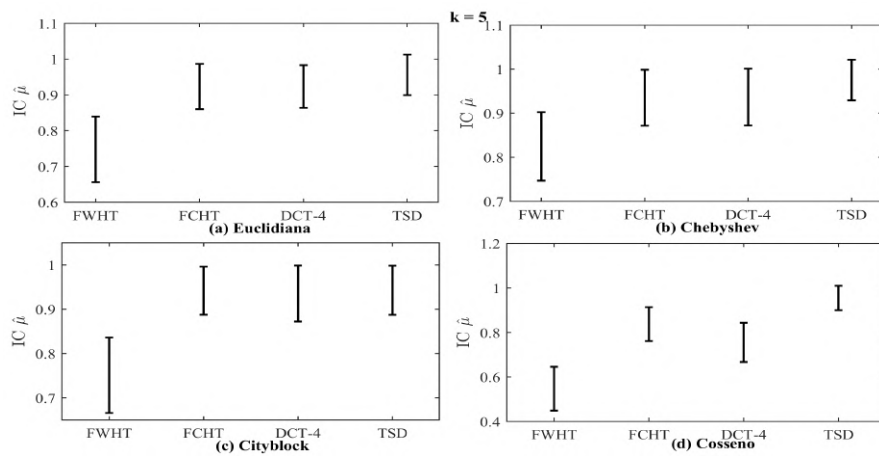
Figura 44 – Resultados dos intervalos de confiança com o teste não paramétrico *bootstrap* para o $k = 3$, com o vidro BK7 e o metal prata (BK7/Ag).



Fonte: Autoria Própria (2020).

Na Figura 45 é exposto os intervalos de confiança da acurácia obtida pelo o método de validação cruzada *LOO-CV* com a vizinhança $k = 5$. Observa-se na Figura 45, que o descritor FWHT com as distâncias obteve o pior desempenho em relação aos outros descritores, devido que o seu intervalo de confiança é inferior. Para as distâncias euclidiana, *chebyshev* e *cityblock* os descritores FChT, DCT-4 e TSD, são estatisticamente iguais, porém o descritor TSD obteve um melhor resultado. A amplitude do seu intervalo de confiança e o seu desvio padrão é menor, desse modo sua classificação tende a ser próxima da média. Os descritores FChT e DCT-4, também obtiveram um resultado inferior ao descritor TSD, devido que o seu intervalo de confiança não apresenta uma acurácia estável.

Figura 45 – Resultados dos intervalos de confiança com o teste não paramétrico *bootstrap* para o $k = 5$, com o vidro BK7 e o metal prata (BK7/Ag).

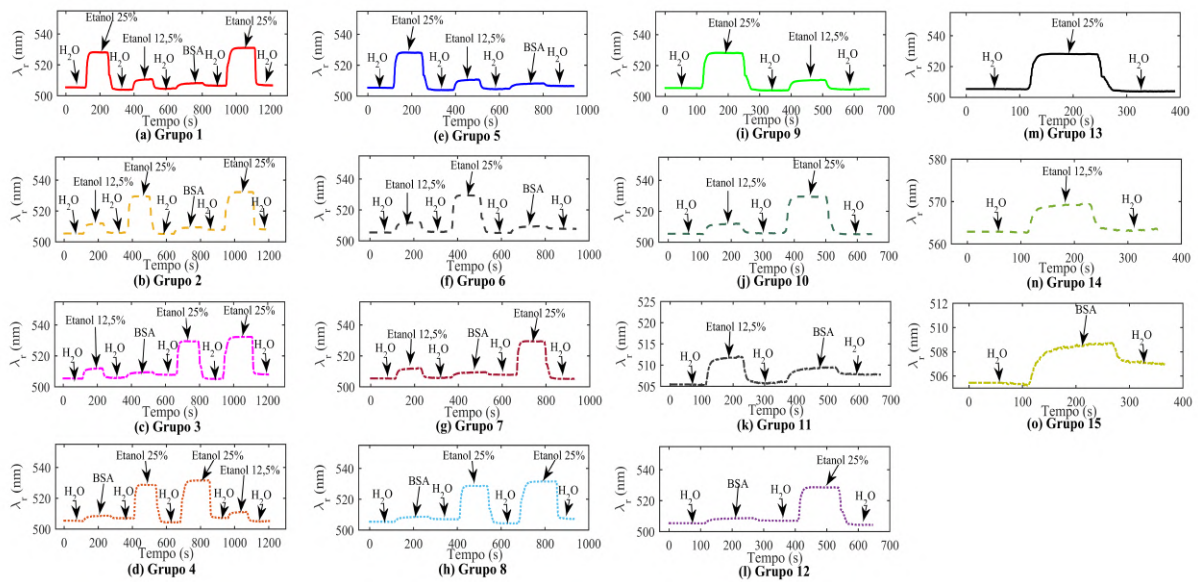


Fonte: Autoria Própria (2020).

Protocolo Experimental com o Polímero PC e o metal prata (PC/Ag)

No modo WIM para o protocolo experimental PC/Ag, em que seguiu o mesmo protocolo experimental, foram criados 15 grupos de sensorgramas. Em cada um dos protocolos, foi inserido uma quantidade e sequência temporal diferente na inserção das substâncias. Na Figura 46 são apresentados os grupos de sensorgramas do polímero PC/Ag, com a criação de sensorgramas com 3, 5, 7 e 9 regimes permanentes. Os protocolos experimentais contém mudanças nas posições e na quantidade de substâncias inseridas.

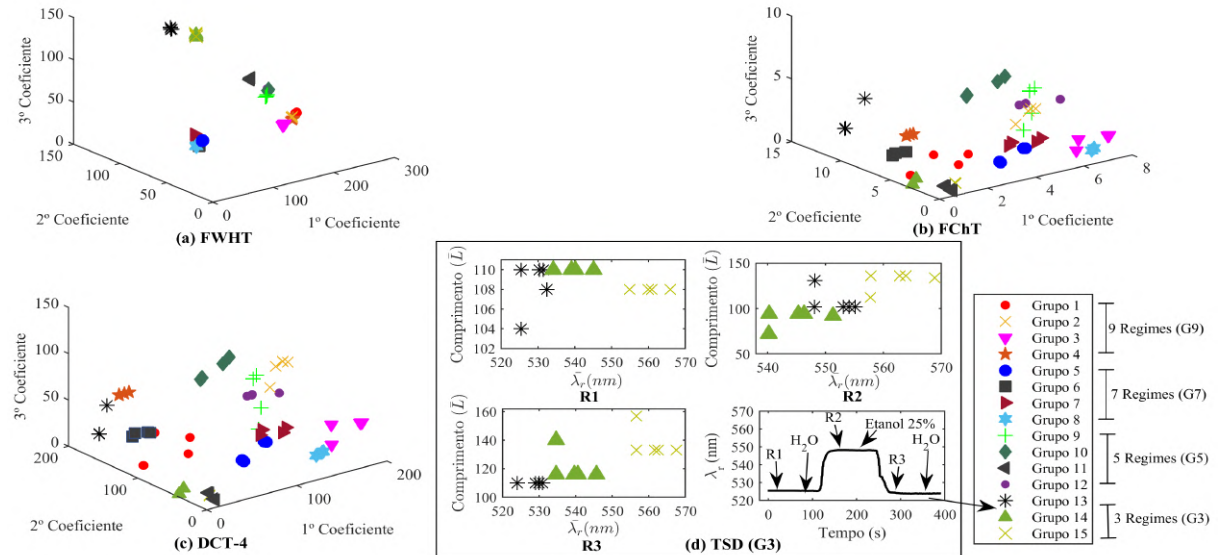
Figura 46 – Grupos de sensorgramas representando 15 protocolos experimentais com quantidade de regimes permanentes diferentes utilizando o polímero e o metal PC/Ag. De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.



Fonte: Autoria Própria (2020).

Para criar 99 sensorgramas, foi utilizada a geração de sensorgramas em seus grupos. Criado os sensorgramas, foi aplicado em cada um, os descritores FWHT, FChT, DCT-4 e TSD. Com os coeficientes de nível AC para os descritores e a média e o comprimento do descritor TSD foi criado dois conjuntos de dados de treinamento. Para o primeiro conjunto foi utilizado os três primeiros coeficientes de nível AC dos descritores FWHT, FChT e DCT-4 e os três primeiros regimes permanentes do descritor TSD. Esse conjunto de dados corresponde aos sensorgramas com 3, 5, 7 e 9 regimes permanentes. Já o segundo conjunto, corresponde aos cinco primeiros coeficientes de nível AC e aos cinco primeiros regimes permanentes do descritor TSD. Na Figura 47 (a), (b) e (c) é apresentado os grupos de sensorgramas com os três primeiros coeficientes de nível AC. Já na Figura 47 apresenta os três primeiros regimes permanentes com os grupos 13, 14 e 15.

Figura 47 – Conjunto de dados de treinamentos em um gráfico 3D com os 99 sensorgramas dos 15 grupos com os descritores (a) FWHT, (b) FChT e (c) DCT-4. (d) Conjunto de dados de treinamento com 3 regimes permanentes em um gráfico 2D.

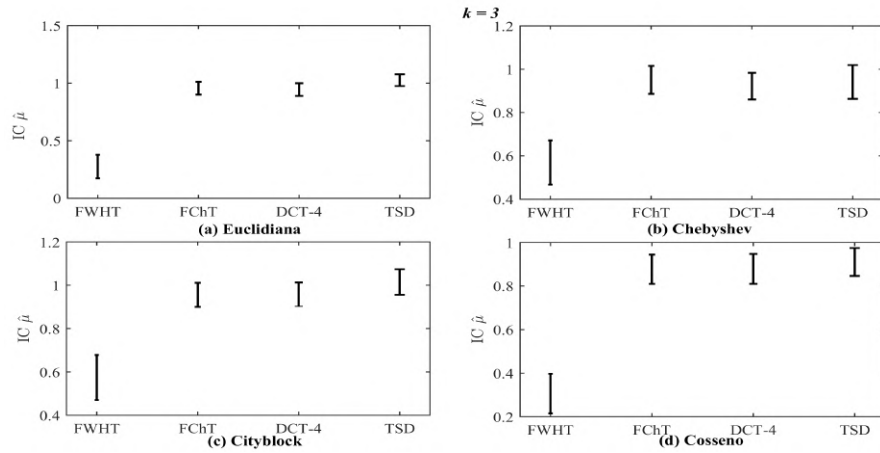


Fonte: Autoria Própria (2020).

Depois de criado os conjuntos de dados de treinamento, em cada conjunto de dados foi atribuído o dado de destino para cada sensorgrama pertencente. Para verificar as respostas fornecidas pelo algoritmo k -NN com o $k = 3$, foi obtido acurácia com 99 sensorgramas e o $k = 5$ para 84. Para as vizinhanças foi utilizada as distâncias euclidiana, *cityblock*, *chebyshev* e cosseno com a validação cruzada *Leave-one-out*. No fim, foi utilizado o teste estatístico não paramétrico *bootstrap* para a resposta fornecida pelo algoritmo. Para o teste estatístico foi aplicado valores de replicações igual a 2000 (número de réplicas *bootstrap*) e $\alpha = 5\%$ (erro de confiança).

Na Figura 48 é apresentado os intervalos de confiança obtidos da acurácia da validação cruzada *Leave-one-out* para o $k = 3$ com os descritores e as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno. O descritor FWHT com as distâncias, obteve o pior desempenho em relação aos outros descritores, e o seu intervalo de confiança foi inferior, assim ocasionou uma classificação errada. Os descritores FChT, DCT-4 e TSD com as distâncias euclidiana, *cityblock* e cosseno, são estatisticamente iguais, porém o descritor TSD obteve um resultado próximo a média, devido que a amplitude do seu intervalo é menor, dessa forma a sua classificação é estável. Já para a distância *chebyshev*, os descritores FChT, DCT-4 e TSD são estatisticamente iguais, porém o descritor DCT-4 possui amplitude menor em relação aos outros descritores, assim é possível uma classificação estável.

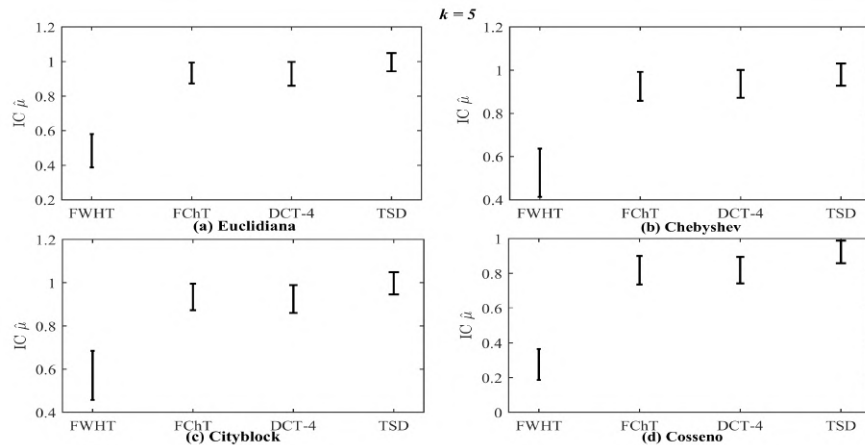
Figura 48 – Resultados dos intervalos de confiança com com o polímero PC e o metal prata (PC/Ag) em que utilizou o teste não paramétrico *bootstrap* para o $k = 3$.



Fonte: Autoria Própria (2020).

É exposto na Figura 48, os intervalos de confiança com a acurácia obtidos pela validação cruzada *LOO-CV* para o $k = 5$ com cada descritor e as distâncias. O descritor FWHT apresentou o pior desempenho em relação aos outros descritores, dessa forma, sua classificação não é estável, devido que o intervalo é inferior. Já para os descritores FChT, DCT-4 e TSD com as distâncias, são estatisticamente iguais, porém o descritor TSD, obteve um resultado próximo a média, por possuir uma amplitude menor, dessa forma a sua classificação é estável.

Figura 49 – Resultados dos intervalos de confiança com com o polímero PC e o metal prata (PC/Ag) em que utilizou o teste não paramétrico *bootstrap* para o $k = 5$.

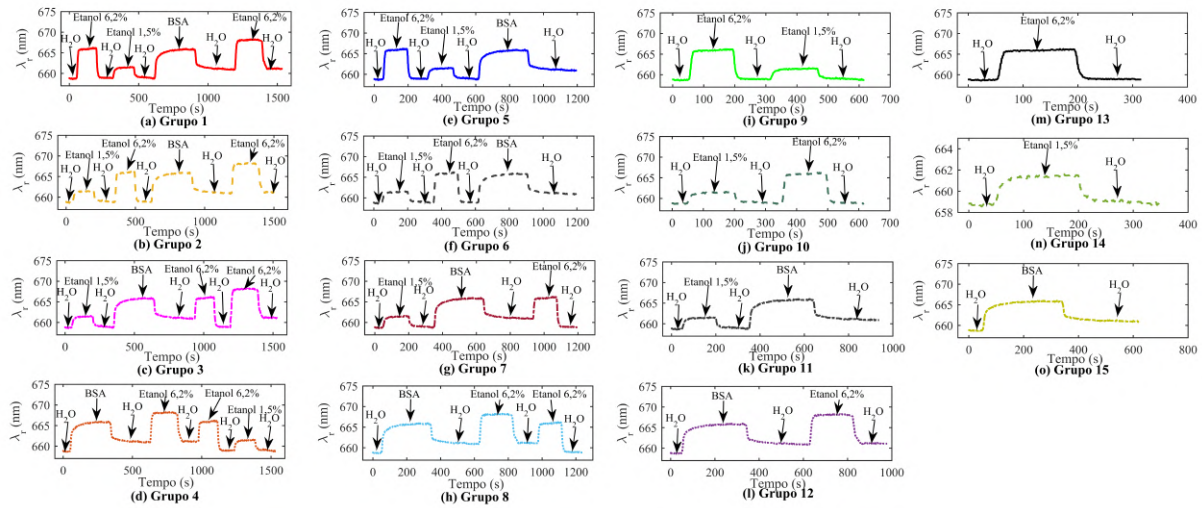


Fonte: Autoria Própria (2020).

Protocolo Experimental com o Polímero PMMA e o metal prata (PMMA/Ag)

Na Figura 50 é apresentado os grupos de sensorgramas do polímero PMMA/Ag no modo WIM. Para esses grupos, seguiu as substâncias H_2O , Etanol 6,2%, Etanol 1,5% e BSA-Monolayer. Para cada grupo, foi inserida uma quantidade e sequência temporal diferente na inserção das substâncias em cada protocolo experimental. É possível observar na Figura 50, os grupos de sensorgramas com as mudanças nas posições e na quantidade de substâncias inseridas, em que se tem os sensorgramas com 3, 5, 7 e 9 regimes permanentes.

Figura 50 – Grupos de sensorgramas no modo WIM representando 15 protocolos experimentais diferentes utilizando o polímero PMMA com o metal prata (PMMA/Ag). De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.



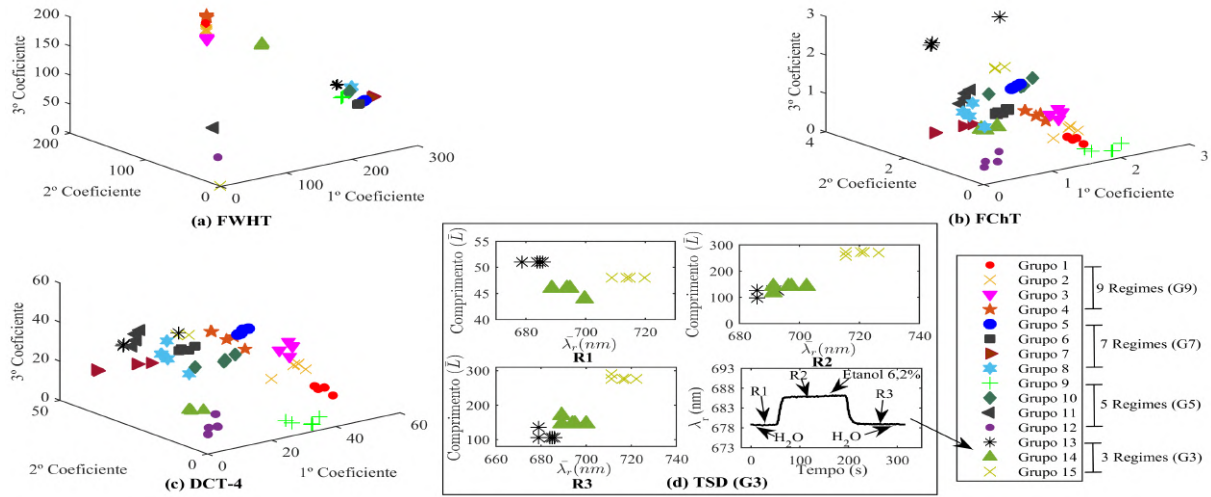
Fonte: Autoria Própria (2020).

Foi aplicado os descritores FWHT, FChT, DCT-4 e a TSD, em cada um dos sensorgramas pertencentes a cada grupo. Em cada grupo foi gerado novos sensorgramas, que consistiu em criar 99. Depois de criado, foi aplicado os descritores FWHT, FChT, DCT-4 e a TSD, em cada um dos sensorgramas pertencentes a cada grupo. Com os coeficientes de nível AC para os descritores FWHT, FChT e DCT-4 e a média e o comprimento do TSD, foram criados dois conjuntos de dados de treinamento. Para o primeiro conjunto de dados de treinamento, foi utilizado os três primeiros coeficientes de nível AC e os três primeiros regimes permanentes, que corresponde aos sensorgramas com 3, 5, 7 e 9 regimes permanentes. Já para o segundo o conjunto de dados de treinamento, foi utilizado os cinco primeiros coeficientes de nível AC e os cinco primeiros regimes permanentes.

A Figura 51 (a), (b) e (c) apresenta os três primeiros coeficientes de nível AC com os 99 sensorgramas. O conjunto de dados de treinamento com os três primeiros regimes permanentes do descritor TSD, corresponde aos grupos 13, 14 e 15, apresentado na Figura

51 (d).

Figura 51 – Conjunto de Dados de Treinamento em um gráfico com os 99 sensorgramas dos grupos para (a) FWHT, (b) FChT e (c) DCT-4. (d) TSD conjunto de dados treinamento com 3 regimes permanentes em um gráfico 2D.



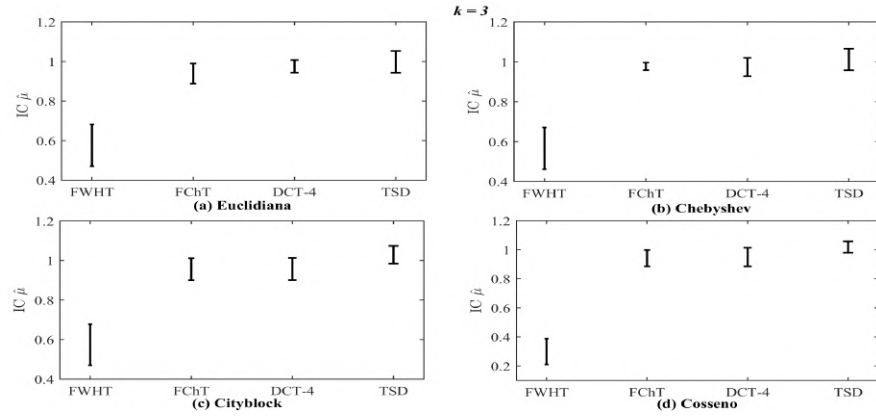
Fonte: Autoria Própria (2020).

Depois de criado os conjuntos de dados de treinamento dos descritores FWHT, FChT e DCT-4, como também foi criado os conjuntos com a média e o comprimento dos regimes permanentes do TSD, foi atribuído o dado de destino para cada sensorgrama pertencente. Para verificar as respostas fornecidas pelo algoritmo k -NN, foi usada o $k = 3$ com 99 sensorgramas e o $k = 5$ para 84 sensorgramas. Com as distâncias euclidiana, *cityblock*, *chebyshev* e cosseno para validação cruzada *Leave-one-out* obteve a acurácia. No fim, para saber qual o melhor descritor, foi utilizado o teste estatístico não paramétrico *bootstrap*. Para o teste *bootstrap* foi aplicado valores de replicações igual a 2000 (número de réplicas *bootstrap*) e $\alpha = 5\%$ (erro de confiança).

É apresentado na Figura 36 os intervalos de confiança da acurácia obtidos da validação cruzada *Leave-one-out* para o $k = 3$ com os descritores FWHT, FChT, DCT-4 e TSD e as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno. O descritor FWHT com as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno apresentou o pior desempenho em relação aos outros descritores, diante que o seu intervalo de confiança é inferior. Com a distância euclidiana os descritores FChT, DCT-4 e TSD são estatisticamente iguais, porém o descritor DCT-4 apresenta um resultado melhor, por possuir uma amplitude do seu intervalo menor. Já para a distância *chebyshev* os descritores FChT, DCT-4 e TSD são estatisticamente iguais, porém o descritor FChT conduz a um resultado melhor com a distância, diante que a sua classificação é estável e por conter uma amplitude menor. Para as distâncias *cityblock* e cosseno os descritores FChT, DCT-4 e TSD também são estatisticamente iguais, porém o descritor TSD apresenta uma amplitude e um desvio

padrão menor, dessa forma, a sua classificação está próxima a média.

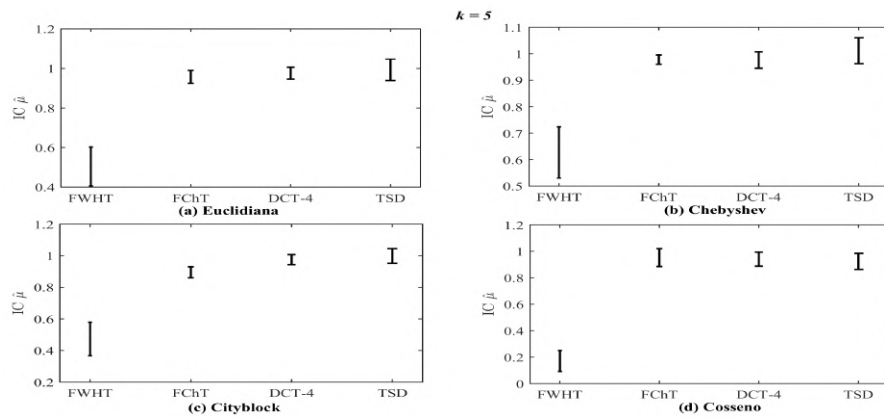
Figura 52 – Resultados dos intervalos de confiança do teste não paramétrico *bootstrap* para o $k = 3$, para o polímero PMMA e o metal prata (PMMA/Ag).



Fonte: Autoria Própria (2020).

Os intervalos de confiança obtidos da acurácia com a validação cruzada *Leave-one-out* é apresentado na Figura 53. O descritor FWHT apresentou o pior desempenho em relação aos descritores, diante que o seu intervalo de confiança é inferior. Os descritores FChT, DCT-4 e TSD com as distâncias euclidiana, *cityblock* e cosseno, são estatisticamente iguais, porém o descritor DCT-4 possui uma amplitude no seu intervalo de confiança menor em relação aos descritores FChT e TSD, ou seja, sua classificação é estável. Já com a distância *chebyshev* os descritores FChT, DCT-4 e TSD também são estatisticamente iguais, no entanto o descritor FChT apresenta um melhor resultado por possuir uma amplitude menor.

Figura 53 – Resultados dos intervalos de confiança do teste não paramétrico *bootstrap* para o $k = 5$, para o polímero PMMA e o metal prata (PMMA/Ag).

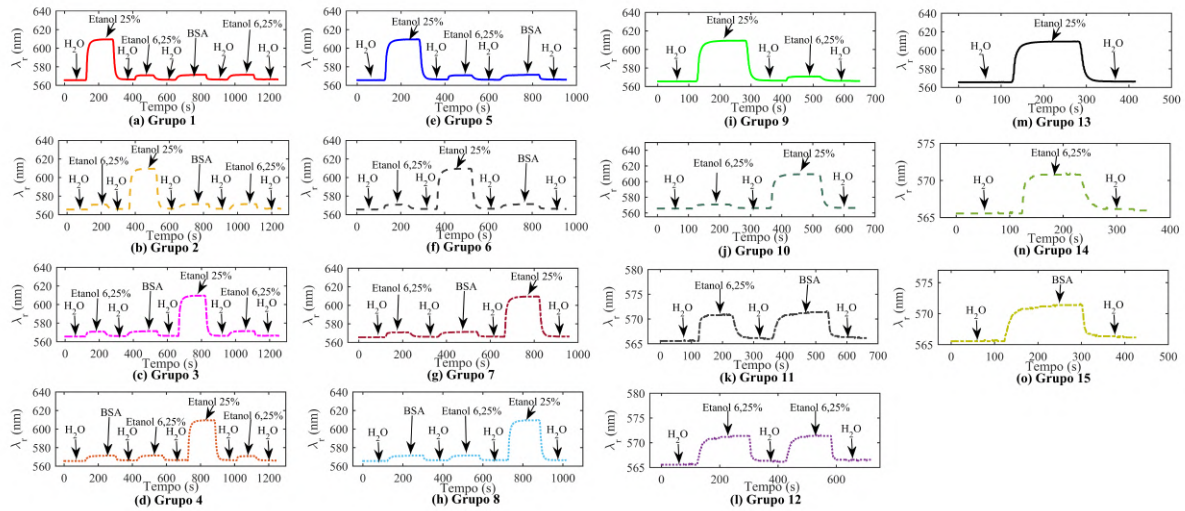


Fonte: Autoria Própria (2020).

Protocolo Experimental com o Polímero TOPAS e o metal prata (TOPAS/Ag)

Para o protocolo experimental TOPAS/Ag no modo WIM, foram criados 15 grupos de sensorgramas, em que seguiu as substâncias H_2O , Etanol 25%, Etanol 6,25% e BSA-Monolayer (*Phosphate Buffered Saline*). Em cada um dos protocolos experimentais, foi inserido uma quantidade e sequência temporal diferente na inserção das substâncias. Na Figura 54 são apresentados os grupos de sensorgramas do polímero TOPAS/Ag com os protocolos de 3, 5, 7 e 9 regimes permanentes. É possível observar nos protocolos experimentais as mudanças nas posições e na quantidade de substâncias inseridas.

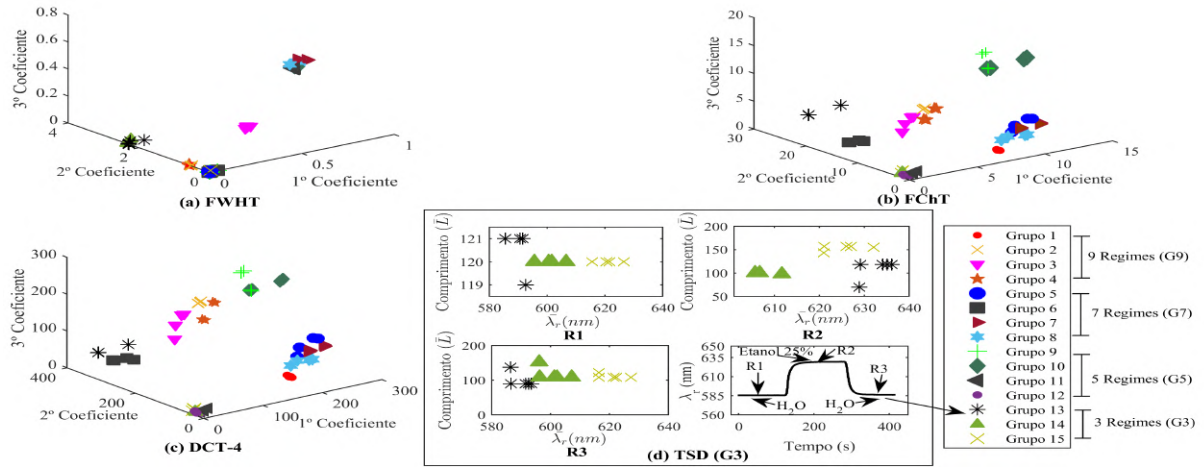
Figura 54 – Grupos de sensorgramas no modo WIM representando 15 protocolos experimentais diferentes utilizando o polímero TOPAS com o metal prata (TOPAS/Ag). De (a) até (d) o protocolo experimental consiste de 9 regimes permanentes. (e) até (h) 7 regimes permanentes. (i) até (l) 5 regimes permanentes. (m) até (o) 3 regimes permanentes.



Fonte: Autoria Própria (2020).

Para cada um dos grupos, foram gerados no total 99. Depois de criado, foi aplicado os quatro descritores FWHT, FChT, DCT-4 e o TSD em cada um dos sensorgramas. Com os coeficientes de nível AC dos descritores FWHT, FChT, DCT-4 e o TSD a média e o comprimento, foram criados dois conjuntos de dados de treinamento. Para o primeiro conjunto foi utilizado os três primeiros coeficientes de nível AC com os descritores FWHT, FChT e DCT-4. No descritor TSD foi utilizado os três primeiros regimes permanentes, esse primeiro conjunto corresponde aos sensorgramas com 3, 5, 7 e 9 regimes permanentes. Já o segundo conjunto, corresponde aos cinco primeiros coeficientes de nível AC e os cinco primeiros regimes permanentes. A Figura 55 (a), (b) e (c) apresenta os três primeiros coeficientes de nível AC em que contém 99 sensorgramas. Na Figura 55 (d) é apresentado o descritor TSD correspondente aos grupos 13, 14 e 15.

Figura 55 – Conjunto de dados de treinamentos com os 99 sensorgramas dos 15 grupos para a FWHT, FChT, DCT-4 e TSD.

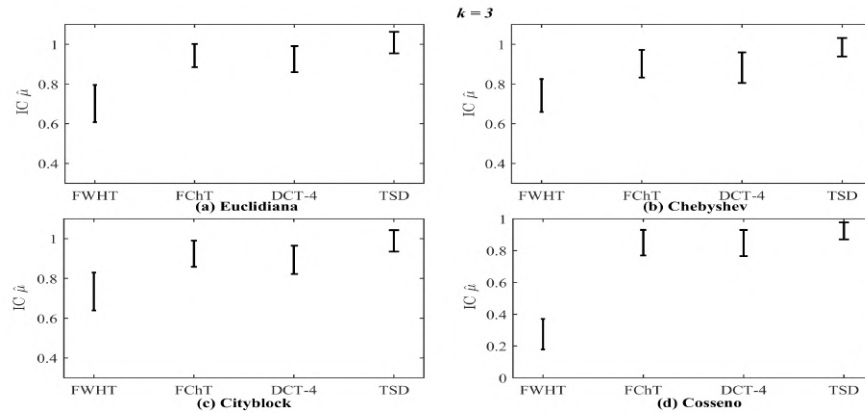


Fonte: Autoria Própria (2020).

Com a criação dos conjuntos de dados de treinamento, foi atribuído em cada sensorgrama o seu dado de destino com o uso dos descritores. Para classificar as respostas do algoritmo k -NN, foi obtido a acurácia com a vizinhança $k = 3$ para 99 sensorgramas e para o $k = 5$ foi 84. Desse modo, utilizou-se as distâncias euclidiana, *cityblock*, *chebyshev* e cosseno, com a validação cruzada *Leave-one-out* para o $k = 3$ e $k = 5$. Foi utilizado o teste estatístico não-paramétrico *bootstrap*, para o teste foi aplicado valores de replicações igual a 2000 (número de réplicas bootstrap) e $\alpha = 5\%$ (erro de confiança).

Na Figura 56 é apresentado os intervalos de confiança obtidos com os descritores FWHT, FChT, DCT-4 e TSD com as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno para o $k = 3$. O descritor FWHT contém um intervalo menor em relação aos descritores, ou seja, apresenta uma classificação errada para a sua vizinhança. Já os descritores FChT, DCT-4 e TSD, são estatisticamente iguais, porém o descritor TSD apresenta um melhor resultado, diante que a amplitude do seu intervalo e o seu desvio padrão é menor e o desempenho é maior, ou seja, sua classificação é mais estável.

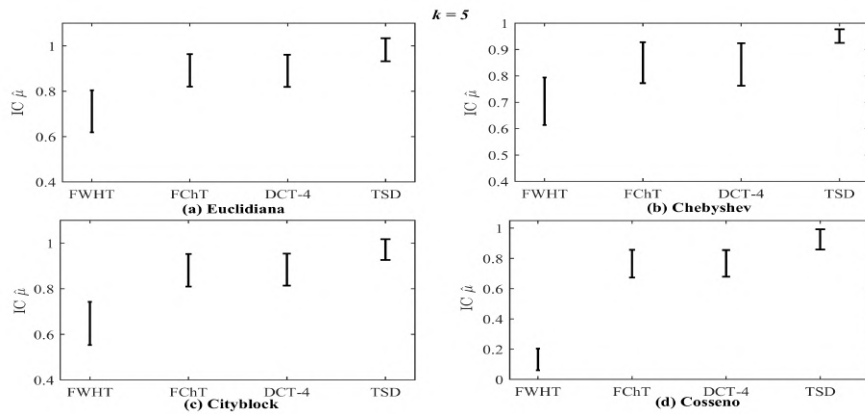
Figura 56 – Resultados dos intervalos de confiança do teste não paramétrico *bootstrap* para o $k = 3$, com o polímero TOPAS/Ag.



Fonte: Autoria Própria (2020).

É exposto os intervalos de confiança na Figura 53 para o $k = 5$. O descritor FWHT, apresentou o menor resultado em relação aos descritores FChT, DCT-4 e TSD, sua classificação não é estável e possui um desvio padrão maior. Já os descritores FChT, DCT-4 e TSD, são estatisticamente iguais, porém o descritor TSD conduz a um melhor resultado, diante que a amplitude do seu intervalo é menor e o desempenho é maior, ou seja, sua classificação é mais estável e possui um desvio padrão menor.

Figura 57 – Resultados dos intervalos de confiança do teste não paramétrico *bootstrap* para o $k = 5$, com o polímero TOPAS/Ag.



Fonte: Autoria Própria (2020).

Para o modo WIM foi utilizado o vidro BK7 e os polímeros PC, PMMA e TOPAS com o metal prata, foram utilizadas as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno com a validação cruzada *Leave-one-out*. O descritor TSD para o $k = 3$ e o $k = 5$ com as distâncias para o vidro e cada polímero, obteve um resultado superior em relação aos demais descritores. O descritor FWHT, obteve o pior desempenho com o vidro e polímeros com as distâncias utilizadas, devido que o seu conjunto de dados de treinamento estão

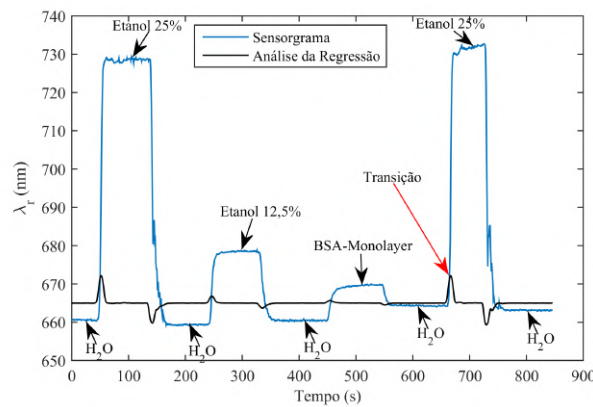
muito próximos. E os descritores FChT e DCT-4 são estatisticamente iguais, ou seja, os seus conjuntos de dados de treinamento são semelhantes.

4.4 Identificação e Análise de Sensorgrama

Para o modo WIM foi aplicado a regressão linear com a estrutura multicamada BK7/Ag, seguindo o protocolo: $H_2O \rightarrow \text{Etanol } 25\% \rightarrow H_2O \rightarrow \text{Etanol } 12.5\% \rightarrow H_2O \rightarrow \text{BSA-Monolayer} \rightarrow H_2O \rightarrow \text{Etanol } 25\% \rightarrow H_2O$. Com a análise de regressão é possível detectar as áreas de regimes permanentes e de transição do sensorgrama. A Figura 58 apresenta a análise de regressão, em que o sensor estima as regiões que representam as substâncias analisadas, e consequentemente, identifica as áreas de transição entre elas. Assim, observado a mudança no valor do coeficiente da regressão indica-se que há uma transição entre as substâncias.

Os valores para a inclinação da reta que descrevem o comportamento de um regime permanente são encontrados calculando os coeficientes β_0 e β_1 da reta $y = \beta_0 + \beta_1 x_i$, de modo a minimizar o erro (ϵ_i) da soma dos quadrados das distâncias de cada ponto à reta de aproximação, em que β_0 é a variável de interseção, β_1 a inclinação da e x_i são os dados do sensorgrama. O algoritmo criado contém uma janela de tempo e com ele é possível obter a resposta do coeficiente de inclinação dos pontos de acordo com essa janela. A janela deve ser ajustada de modo que ruídos presentes na instrumentação do sensor não caracterizem a ocorrência de transição entre as substâncias (MONTGOMERY; PECK; VINING, 2012).

Figura 58 – Sensorgrama para o protocolo $H_2O \rightarrow \text{Etanol } 25\% \rightarrow H_2O \rightarrow \text{Etanol } 12.5\% \rightarrow H_2O \rightarrow \text{BSA-Monolayer} \rightarrow H_2O \rightarrow \text{Etanol } 25\% \rightarrow H_2O$ (azul). Resposta do coeficiente de regressão (preto). Destaque (seta vermelha) para a situação em que o coeficiente da regressão muda, indicando uma transição entre as substâncias analisadas.

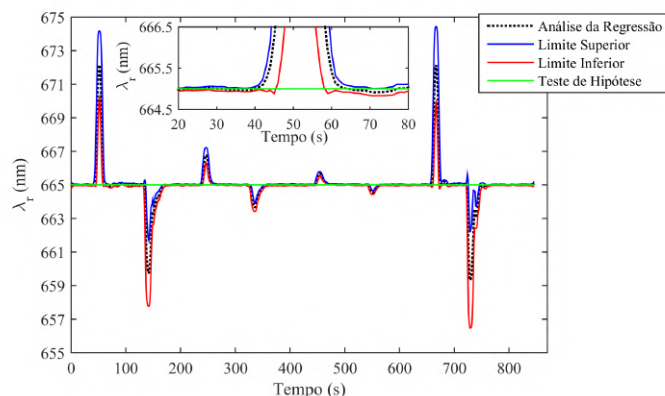


Fonte: Autoria Própria (2020).

Com o uso da regressão é possível proceder com testes de hipótese na resposta do sensor, com possibilidade de confiabilidade nos dados. Na Figura 59, é apresentado o teste de hipótese para verificar com 99.9% de confiança a presença de uma determinada substância no protocolo experimental da Figura 58. O nível de confiança define os limites superior e inferior do intervalo de confiança e o teste de hipótese verifica se na janela temporal estabelecida o coeficiente de regressão está contido nesses limites. Na Figura 59, é possível observar a análise de regressão, os intervalos de confiança superior e inferior e o teste de hipótese.

Quando o teste de hipótese (H_0) da inclinação da reta é nula ($H_0 \Rightarrow \beta_1 = 0, Verdadeiro$) significa que o valor de β_1 e seu teste encontra-se dentro dos limites estabelecidos. Na sub-figura tem-se no intervalo entre 20s a 80s onde não foi possível confirmar com 99.9% de certeza que é uma substância, em que os limites e o coeficiente da análise de regressão está fora do teste de hipótese (linha verde), ou seja o teste contém a hipótese ($H_0 \Rightarrow \beta_1 \neq 0, falso$), logo está região refere-se a uma transição entre substâncias. Com o teste de hipótese é possível verificar e auditar se um sensor com inteligência embarcada está se comportando de maneira correta, seguindo os padrões regulamentados pelas agências governamentais responsáveis. Por tanto, faz-se necessário a investigação das condições de operação e o impacto do uso de algoritmos de aprendizagem no desenvolvimento e na operação de um sensor SPR.

Figura 59 – Teste de hipótese para verificar com 99.9% de confiança a presença de uma determinada substância. O nível de confiança define os limites superior e inferior do intervalo e o teste de hipótese verifica se na janela temporal estabelecida o coeficiente de regressão está contido. Em destaque tem-se que no intervalo 20 a 80 não foi possível confirmar com 99.9% de certeza que é uma substância.

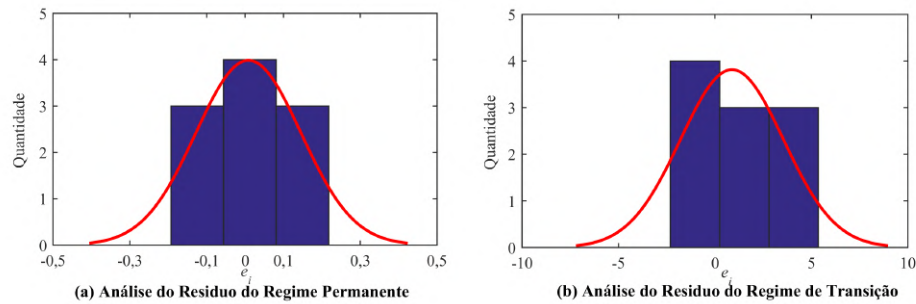


Fonte: Autoria Própria (2020).

Para verificar se os dados estão distribuídos na forma normal no modelo de regressão do regime permanente foi aplicado a análise dos resíduos. A análise dos resíduos é importante para investigar a adequabilidade do modelo de regressão. Para calcular a

análise é definido por um erro $e_i = y_i - \hat{y}_i$, em que, y_i é o valor observado e $\hat{y}_i$ o valor ajustado. Na Figura 60 (a) é possível observar o histograma da análise do resíduo do regime permanente. Sua distribuição está na forma normal, diante que segue a condição de ressonância estável. Já na Figura 60 (b) observa-se que no histograma do regime de transição os dados não estão na forma normal, ou seja, não segue a condição de ressonância estável.

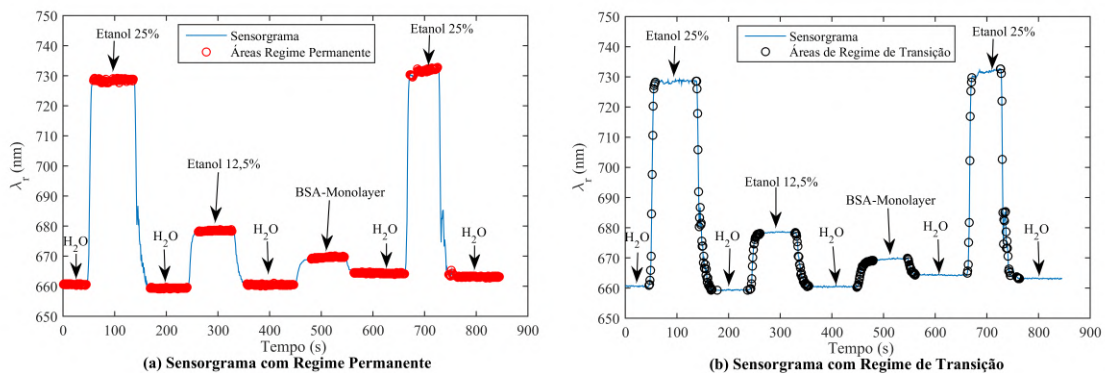
Figura 60 – Histogramas da análise do resíduo do modelo de regressão. (a) Regime permanente. (b) Regime de transição.



Fonte: Autoria Própria (2020).

Na Figura 61, tem-se a identificação do regime permanente e de transição no protocolo experimental. Os regimes permanentes identificados são expressos pela condição de ressonância estável quando se tem fração de uma única substância em contato com sensor, já os regimes de transição sua identificação é expresso quando sua condição de ressonância não está estável, ou seja, quando o regime está ascendendo e descendendo no valor de ressonância. Com o uso apropriado dos algoritmos inteligentes, é possível aferir a qualidade de um sensor e atestar em que condições de operação (valores estimados e seus erros) este funcionará com determinado nível de confiabilidade.

Figura 61 – Identificação de padrão na resposta no modo WIM. (a) Identificação dos regimes permanentes com 99.9% de confiança, caracterizando as regiões que identificam as substâncias analisadas. (b) Identificação dos regimes de transição com 99.9% de confiança.

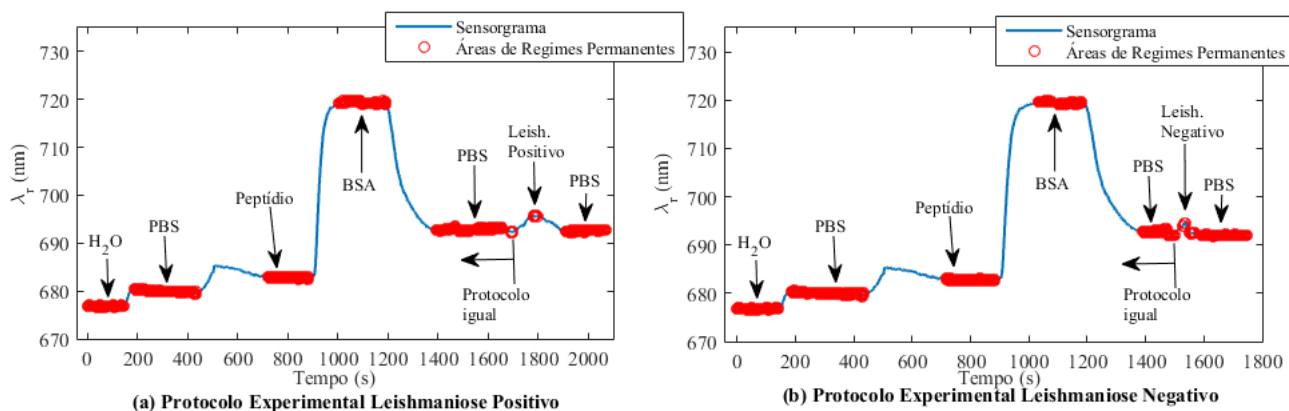


Fonte: Autoria Própria (2020).

Teste Automático

Estudos sobre o uso de técnicas de inteligência em sensores SPR permitem aprender um protocolo experimental, identificar padrões na resposta, bem como sinalizar anomalias no comportamento do sensor. Partes dessas tarefas de aprendizagem já foram demonstradas até aqui. Na Figura 62, tem-se a identificação de padrões no protocolo experimental no teste negativo e positivo da Leishmaniose. Em ambos os protocolos experimentais, a preparação para realizar seguiu as mesmas substâncias, demonstrados no chaveado das substâncias até o tempo 1700s (Leishmaniose Positivo) e 1500s (Leishmaniose Negativo), a partir dessas substâncias é realizado o teste. Uma vez estabelecido o teste positivo 62 (a), pode-se usá-lo como sensorgrama padrão e assim compará-lo com novos testes. Se um novo sensorgrama é estaticamente igual ao pré-teste positivo, pode-se afirmar que a amostra testada é positiva. Dessa forma, a automatização do teste é obtida comparando o novo sensorgrama já previamente armazenado ao sensorgrama positivo. Assim, o algoritmo k -NN com o $k = 3$ e distância Euclidana é usado para reconhecimento. Com o processo de validação cruzada LOO-CV, obteve-se a acurácia de 0.92.

Figura 62 – Identificação dos regimes permanentes no protocolo experimental positivo e negativo da Leishmaniose com 99.9% de confiança, caracterizando as regiões que identificam as substâncias analisadas.



Fonte: Autoria Própria (2020).

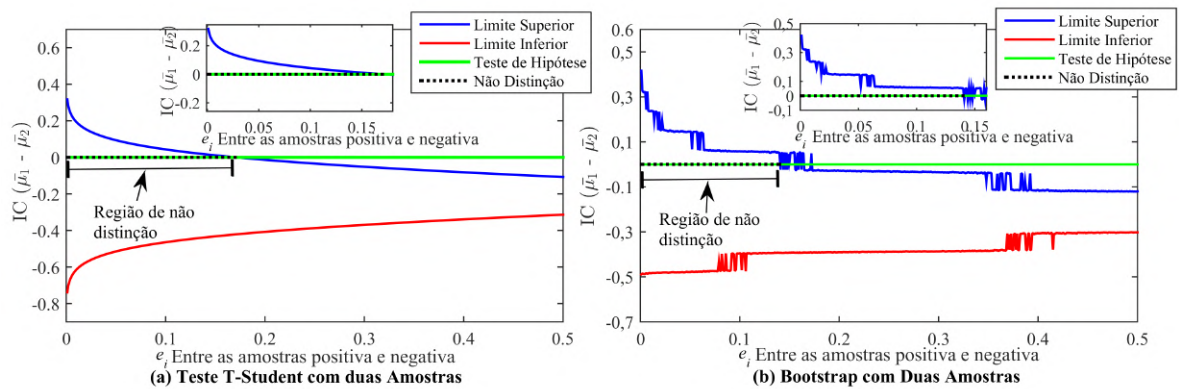
Com a análise da regressão foi possível identificar os regimes permanentes positivo e negativo no protocolo experimental da Leishmaniose. Com os padrões identificados, é possível fazer o teste de hipótese com as duas amostras. Para tanto foi realizado o teste paramétrico t -Student com diferenças entre as médias e variâncias desconhecidas e o teste não-paramétrico *bootstrap* (com 2000 replicações). Com esses testes é possível automatizar e saber se a substância da Leishmaniose inserida é positiva ou negativa.

Na Figura 63 (a) apresenta-se o intervalo de confiança com o teste t -Student com as duas amostras positivo e negativo da Leishmaniose em relação ao erro. Observa-se que foi obtido o intervalo de confiança com duas amostras (Positivo e Negativo). Nesses intervalos

existe uma fase marcada em preto em que não é possível distinguir a substância inserida, ou seja, se a substância é positiva ou negativa. A medida em que o erro vai aumentando é possível observar uma distinção nas substâncias inseridas. Assumindo o valor positivo de referência em que é apresentado na Figura 63 (b) torna-se possível verificar se um novo sensorgrama é estatisticamente igual ou diferente do sensorgrama positivo com o *bootstrap*. Um eventual novo teste-Leishmaniose é comparado com o sensorgrama-positivo já previamente armazenado através de teste não paramétrico *bootstrap* e contra-prova com teste paramétrico *t-Student*.

O teste de hipótese ($H_0 \Rightarrow IC < \epsilon_i$, Verdadeiro) significa que o novo teste (novo sensorgrama) é estatisticamente igual ao sensorgrama positivo, isto é, não há distinção entre as amostras. Assim, pode-se afirmar que a amostra testada é positiva. Observa-se também que é realizado o intervalo de confiança com as duas amostras na Figura 63, em que existe uma fase marcada em preto em que não é possível distinguir a substância inserida. Desse modo, com o teste estatístico *bootstrap* em relação ao teste *t-Student* tem um desempenho melhor, diante que na sub-figura 63 (b) a região de não distinção do intervalo de confiança é menor do que o teste *t-Student*.

Figura 63 – Intervalo de confiança com duas amostras para o protocolo experimental da Leishmaniose. (a) Intervalo de confiança com o teste paramétrico *t-Student*. (b) Intervalo de confiança com o teste não-paramétrico *bootstrap*.



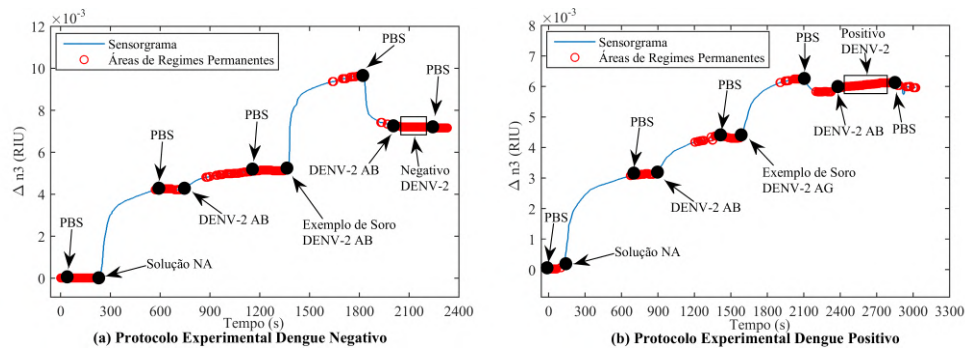
Fonte: Autoria Própria (2020).

Na Figura 64, tem-se a identificação de padrões no protocolo experimental no teste negativo e positivo da Dengue no modo AIM, observa-se que os protocolos experimentais seguem as mesmas substâncias. Com a análise da regressão também foi possível identificar, tem-se a identificação de padrões no protocolo experimental no testes da Dengue. Em ambos os protocolos experimentais, a preparação para realizar seguiu as mesmas substâncias, demonstrados no chaveado das substância até o tempo 1900s (Dengue Negativo) e 2400s (Dengue Positivo). Uma vez estabelecido o teste negativo 62 (a), pode-se usá-lo como sensorgrama padrão e assim compara-lo com novos testes. Se um novo sensorgrama é

estaticamente igual ao pré-teste negativo, pode-se afirmar que a amostra testada é negativa.

Com os padrões identificados, é possível fazer o teste de hipótese com duas amostras, para o teste foi realizado o teste paramétrico *t-Student* com diferenças entre as médias e variâncias desconhecidas. O teste não-paramétrico *bootstrap* (utilizou-se 2000 replicações), com esses testes é possível automatizar e saber se a amostra do paciente inserida é positiva ou negativa.

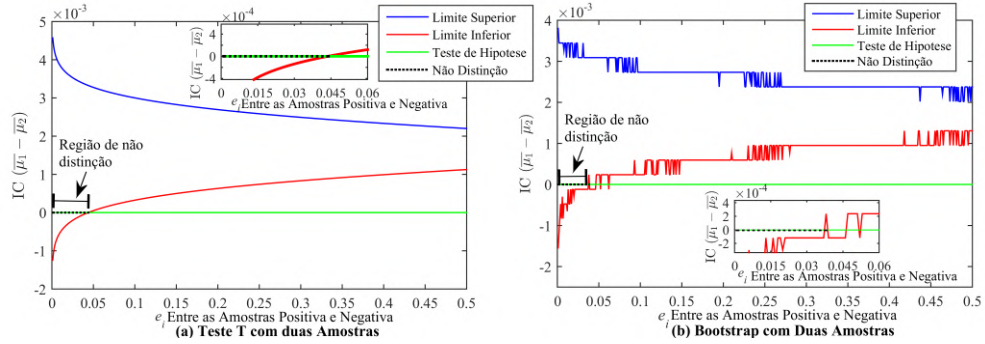
Figura 64 – Identificação dos regimes permanentes no protocolo experimental positivo e negativo da Dengue com 99.9% de confiança, caracterizando as regiões que identificam as substâncias analisadas.



Fonte: Autoria Própria (2020).

O intervalo de confiança com o teste *t-Student* com duas amostras positivo e negativo da Dengue em relação ao erro é apresentado na Figura 63. É observado que foi obtido o intervalo de confiança com duas amostras (Positivo e Negativo). No qual, existe uma fase marcada em preto em que não é possível distinguir a substância inserida, ou seja, se a substância é positiva ou negativa, a medida em que o erro vai aumentando é possível observar uma distinção nas substâncias inseridas. Na Figura 63 (b) apresenta-se o teste não-paramétrico com o *bootstrap*, observa-se também que é realizado o intervalo de confiança com as duas amostras, existe uma fase marcada em preto em que não é possível distinguir a substância inserida. Desse modo, com o teste estatístico *bootstrap* em relação ao teste *t-Student* tem um desempenho melhor, diante que na sub-figura 63 (b) a região de não distinção do intervalo de confiança é menor do que o teste *t-Student*.

Figura 65 – Intervalo de confiança com duas amostras para o protocolo experimental da Dengue. (a) Intervalo de confiança com o teste paramétrico *t-Student*. (b) Intervalo de confiança com o teste não-paramétrico *bootstrap*.



Fonte: Autoria Própria (2020).

4.5 Conclusão

Os resultados obtidos no modo AIM tiveram a mesma configuração em todos os descritores. Foi possível observar que o descritor TSD obteve o mesmo desempenho ou foi superior aos demais descritores para o vidro BK7 com todas as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno. O polímero PC com o uso do descritor TSD, também foi estatisticamente igual ou superior aos descritores com a distância *chebyshev* e cosseno para o $k = 3$ e o $k = 5$. Porém os descritores FWHT e DCT-4 com as distâncias euclidiana e *cityblock* para o $k = 3$ apresentaram melhores resultados do que os descritores FChT e TSD. Com o uso da vizinhança $k = 5$ para o PC e as distâncias euclidiana e *cityblock* os descritores FWHT e TSD são estatisticamente iguais, porém obtiveram resultados superiores aos descritores FChT e DCT-4.

Para o polímero PMMA com o uso do metal ouro, os descritores são estatisticamente iguais para as distâncias euclidiana, *chebyshev* e *cityblock* com o $k = 3$, porém os descritores FChT e DCT-4 apresentaram um intervalo de confiança superior, já que os seus dados do conjunto de dados de treinamento se encontram dispersos. O descritor FWHT com o uso da distância cosseno, obteve um pior desempenho, devido que os seus dados estão muito próximos entre si e os demais descritores são estatisticamente iguais. Com a vizinhança $k = 5$ para o polímero PMMA no modo AIM, o descritor DCT-4 apresentou um melhor desempenho, diante que o seu intervalo de confiança é superior, ou seja, seu conjunto de dados está bem disperso. Para o polímero PMMA com a distância cosseno e o $k = 5$ o descritor TSD apresentou uma classificação mais próxima a média, ou seja, os seus dados se encontram dispersos, assim é superior aos demais descritores.

No modo AIM também foram adquiridos os resultados dos intervalos de confiança com o uso do polímero TOPAS com o $k = 3$ e $k = 5$. O descritor FwHT apresentou o pior desempenho com as distâncias euclidiana, *chebyshev* devido que os seus dados estão muito

próximos entre si e o descritor TSD é estatisticamente igual aos descritores FChT e DCT-4 ou superior. Com a distância *cityblock* os descritores são estatisticamente iguais, porém os descritores FWHT e DCT-4 são superiores, com uma acurácia maior. E na distância cosseno o descritor TSD apresentou o melhor desempenho, sua classificação está mais próxima a média.

Também foi possível descobrir que o descritor TSD com o vidro e os polímeros utilizados para comparação com os outros descritores, obteve um desempenho melhor ou foi igual aos outros descritores com a distância cosseno. A reconstrução de um sensorgrama com o descritor TSD se apresentou de forma satisfatória, assemelhando-se a um sensorgrama original, como também o EMQ do descritor TSD apresentou um erro menor em relação aos outros descritores.

Em relação ao modo WIM, seguiu-se a mesma configuração do modo AIM, foi possível observar que o descritor TSD obteve o mesmo desempenho ou foi superior aos descritores FWHT, FChT e DCT-4 para o vidro BK7 com todas as distâncias euclidiana, *chebyshev*, *cityblock* e cosseno para o $k = 3$ e $k = 5$. Já o descritor FWHT apresentou o pior desempenho com a vizinhança $k = 3$ e $k = 5$, devido que o seu conjunto de dados de treinamento está muito próximo, assim proporciona uma classificação errada. Para os resultados do polímero PC, o descritor FWHT também apresentou o pior desempenho com a vizinhança $k = 3$ e $k = 5$ com as distâncias utilizadas e o descritor TSD obteve o mesmo desempenho ou foi superior aos descritores.

Para os resultados obtidos com polímero PMMA, o descritor FWHT apresentou um desempenho inferior com as vizinhanças $k = 3$ e $k = 5$ e as distâncias utilizadas, porém com o descritor DCT-4 com a distância euclidiana para as vizinhanças, apresentou uma amplitude menor em relação aos demais descritores, dessa forma, a sua classificação é estável. Para a distância *chebyshev*, o descritor FChT com o $k = 3$ e $k = 5$ possui um desempenho superior aos descritores DCT-4 e TSD. Com a distância cosseno e as vizinhanças, o descritor TSD apresenta uma classificação mais próxima a média, ou seja, os seus dados já se encontram dispersos.

O descritor TSD para o $k = 3$ e $k = 5$ é estatisticamente igual ou superior aos descritores utilizados para o polímero TOPAS e o metal prata, devido que o seu conjunto de dados de treinamento está bem espalhado, assim proporciona uma classificação correta e próxima a média. Já o descritor FWHT apresentou o pior desempenho para o polímero com as vizinhanças utilizadas, assim proporciona uma classificação errada.

O descritor TSD com os polímeros PC, TOPAS e o vidro BK7 com o metal prata, obteve um desempenho superior aos demais descritores com os vizinhos $k = 3$ e $k = 5$, devido que as distâncias do conjunto de dados de treinamento estão mais dispersas. Também descobriu-se que o descritor TSD com o vidro e os polímeros utilizados para comparação com os outros descritores obteve um desempenho melhor ou foi igual aos

outros descritores com o $k = 5$ no polímero TOPAS.

A Tabela 5 apresenta um resumo dos resultados dos descritores com a quantidade de vizinhos para os modos AIM e WIM, com as distâncias utilizadas. O sinal de positivo significa que os descritores obteve um desempenho igual ou superior. Já o sinal de negativo o desempenho é inferior. Observa-se na Tabela 5 que os descritores FChT, DCT-4 e TSD com o uso do $k = 3$ e as distâncias utilizadas, obteve um resultado igual ou superior, com a aplicação desses descritores os coeficientes de nível AC e os regimes permanentes estavam um pouco dispersos. Já o descritor FWHT obteve um desempenho inferior com as distâncias utilizadas, devido que o seu conjunto de dados de treinamento estavam próximos. Já em relação ao $k = 5$ os descritores FWHT, FChT e DCT-4, obteve um resultado inferior devido que os seus dados estavam próximos como também a distância utilizada para realizar a classificação. Já o descritor TSD obteve um desempenho superior aos demais descritores.

Tabela 5 – Resultados dos descritores em relação a quantidade de vizinhos com o algoritmo k -NN.

Qtd. de Vizinhos	FWHT	FChT	DCT-4	TSD
$k = 3$	-	+	+	+
$k = 5$	-	-	-	+

Fonte: Autoria Própria (2020).

Com a regressão linear foi obtido a análise e identificação das substâncias em um protocolo experimental de forma que é possível verificar e auditar de maneira correta as respostas do sensorgrama. Com as respostas obtidas torna-se possível saber se uma região é regime permanente ou de transição no sensorgrama. Com a análise de regressão foi possível aplicar intervalos de confiança com duas amostras para identificar se é possível distinguir as substâncias positiva e negativa nos protocolos experimentais da Dengue e Leishmaniose, para essa abordagem pode ser aplicado tanto no modo AIM como WIM. No Apêndice A é apresentado uma ferramenta que foi desenvolvida com o uso das técnicas de aprendizagem de máquina e o descritor TSD para identificar, classificar e descrever as substâncias inseridas no sensorgrama.

5 Conclusão

Este trabalho apresentou a utilização de algoritmos de aprendizagem de máquina com o objetivo de descrever, classificar, analisar, identificar e automatizar testes a partir de sensorgrama de um sensor baseado em ressonância de plasmons de superfície com o uso do PPBIO. Para descrever e classificar os sensorgramas foi feito uma análise comparativa entre diferentes descritores com o algoritmo de aprendizagem de máquina k -NN, em que foi utilizado teste estatísticos para obter qual o melhor descritor. Dessa forma, o descritor TSD desenvolvido um resultado satisfatório com o k -NN e com as distâncias para o $k = 3$ e $k = 5$. Para analisar, identificar e automatizar o protocolo experimental, foi utilizado testes de hipóteses, intervalos de confiança, regressão linear, em que apresentou-se de forma satisfatória a identificação e análise de padrões na resposta SPR. Também foi possível automatizar o teste da Dengue e Leishmaniose tendo como base sensorgramas de diagnóstico positivo e negativo.

Com o uso apropriado de algoritmos inteligentes é possível aferir a qualidade de um sensor SPR e atestar em que condições de operação este funcionará com determinado nível de confiabilidade. Assim, o sensor inteligente garante a confiabilidade nas respostas fornecidas, condição essencial para auditoria por órgãos fiscalizadores. Ademais, a análise sobre o uso de técnicas de inteligência em sensores SPR possibilita prever anomalias no comportamento do sensor, garantido melhores resultados nas medições.

Destaca-se as limitações do trabalho a construção da ferramenta em que utilizou-se o MATLAB, não sendo desenvolvido uma versão web ou *Desktop*. Apesar do *software* ser compatível com diversos sistemas operacionais é necessário da instalação da ferramenta para ser executando, não sendo possível sua portabilidade. Além disso, para a classificação dos sensorgramas se limitando-se ao uso do PPBIO, existindo diferentes materiais e configurações para aquisição de sensorgramas.

Como contribuição científica para área, os trabalhos apresentado na literatura utilizam métodos matemáticos para obtenção de parâmetro como análise numérica em que torna-se custoso em relação a obtenção de respostas. Dessa forma, a aplicação de técnicas de aprendizagem de máquina e o uso descritores para o desenvolvido de respostas inteligentes baseado em um sensor SPR, permitiu realizar análises estatísticas para o fornecimento de respostas satisfatória, com a criação de modelos de conjunto de dados de treinamento para classificação, como também um descritor em que reconhece os regimes permanentes de um sensorgrama e como também distinguir os testes da Dengue e Leishmaniose, tornando-se a forma de identificação, análise e classificação menos custosa e uma resposta rápida.

As perspectivas futuras, são:

- Apesar de ter alguns conjuntos de dados de treinamento para classificação dos sensorgramas, seria relevante um levantamento de diferentes tipos sensorgramas com diferentes materiais e configurações para a classificação;
- Criar uma base de dados com diferentes tipos de sensorgramas para classificar, analisar e identificar as regiões.
- Desenvolver um *software* web ou *Desktop* de domínio público, em que possa descrever, analisar, identificar e classificar os sensorgramas com diferentes materiais e configurações;
- Aprimorar os estudos sobre diferentes algoritmos de classificação de aprendizagem de máquina, para os sensorgramas, não limitando-se ao algoritmo k -NN.

Referências

- ADAM, P.; PILIARIK, M.; ŠÍPOVÁ, H.; ŠPRINGER, T.; VALA, M.; HOMOLA, J. Surface plasmons for biodetection. John Wiley Sons, Ltd, p. 1–58, 2012. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118310212.ch1>>. Citado na página 30.
- AGRAWAL, A.; SINGH, A.; YAZDI, S.; SINGH, A.; ONG, G. K.; BUSTILLO, K.; JOHNS, R. W.; RINGE, E.; MILLIRON, D. J. Resonant coupling between molecular vibrations and localized surface plasmon resonance of faceted metal oxide nanocrystals. *Nano Letters*, v. 17, n. 4, p. 2611–2620, 2017. PMID: 28337921. Disponível em: <<https://doi.org/10.1021/acs.nanolett.7b00404>>. Citado 3 vezes nas páginas 25, 27 e 56.
- AHMED, N.; NATARAJAN, T.; RAO, K. R. Discrete cosine transform. *IEEE Transactions on Computers*, C-23, n. 1, p. 90–93, Jan 1974. ISSN 0018-9340. Citado 3 vezes nas páginas 45, 46 e 56.
- AHUJA, D.; PARANDE, D. Optical sensors and their applications. *Journal of Scientific Research and Reviews*, v. 1, n. 5, p. 060–068, 2012. Citado na página 22.
- ALFEILAT, H. A. A.; HASSANAT, A. B.; LASASSMEH, O.; TARAWNEH, A. S.; ALHASANAT, M. B.; SALMAN, H. S. E.; PRASATH, V. S. Effects of distance measure choice on k-nearest neighbor classifier performance: A review. *Big Data*, 2019. PMID: 31411491. Disponível em: <<https://doi.org/10.1089/big.2018.0175>>. Citado na página 59.
- AMANATIADIS, A.; KABURLASOS, V.; GASTERATOS, A.; PAPADAKIS, S. Evaluation of shape descriptors for shape-based image retrieval. *IET Image Processing*, v. 5, p. 493–499, 09 2011. Disponível em: <<http://dx.doi.org/10.1049/iet-ipr.2009.0246>>. Citado na página 24.
- Basavaraju, A.; Du, J.; Zhou, F.; Ji, J. A machine learning approach to road surface anomaly assessment using smartphone sensors. *IEEE Sensors Journal*, v. 20, n. 5, p. 2635–2647, 2020. Citado na página 34.
- BOCKOVÁ, M.; SLABÝ, J.; ŠPRINGER, T.; HOMOLA, J. Advances in surface plasmon resonance imaging and microscopy and their biological applications. *Annual Review of Analytical Chemistry*, v. 12, n. 1, p. 151–176, 2019. PMID: 30822102. Disponível em: <<https://doi.org/10.1146/annurev-anchem-061318-115106>>. Citado na página 30.
- BOECKER, D.; ZYBIN, A.; NIEMAX, K.; GRUNWALD, C.; MIRSKY, V. Noise reduction by multiple referencing in surface plasmon resonance imaging. *The Review of scientific instruments*, v. 79, p. 023110, 03 2008. Citado na página 33.
- BORUAH, B. S.; BISWAS, R. An optical fiber based surface plasmon resonance technique for sensing of lead ions: A toxic water pollutant. *Optical Fiber Technology*, v. 46, p. 152 – 156, 2018. ISSN 1068-5200. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1068520018304243>>. Citado na página 22.

- CANOA, P.; SIMÓN-VÁZQUEZ, R.; POPPLEWELL, J.; GONZÁLEZ-FERNÁNDEZ África. A quantitative binding study of fibrinogen and human serum albumin to metal oxide nanoparticles by surface plasmon resonance. *Biosensors and Bioelectronics*, v. 74, p. 376 – 383, 2015. ISSN 0956-5663. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0956566315301688>>. Citado 2 vezes nas páginas 26 e 27.
- CHA, S.-H. Comprehensive survey on distance/similarity measures between probability density functions. *Int. J. Math. Model. Meth. Appl. Sci.*, v. 1, 01 2007. Citado na página 59.
- CHAN, D. Adaptive linear filter for real time noise reduction in surface plasmon resonance systems background. WO/2011/016769. 2009. Citado na página 27.
- CHOMBOON, K.; CHUJAI, P.; TEERARASSAMMEE, P.; KERDPRASOP, K.; KERDPRASOP, N. An empirical study of distance metrics for k-nearest neighbor algorithm. *Proceedings of the 3rd International Conference on Industrial Application Engineering*, p. 280–285, 2015. Citado 2 vezes nas páginas 36 e 59.
- CORR, P.; STEWART, D.; HANNA, P.; MING, J.; SMITH, F. J. Discrete chebyshev transform. a natural modification of the dct. v. 3, p. 1142–1145 vol.3, Sep. 2000. ISSN 1051-4651. Disponível em: <<https://doi.org/10.1109/ICPR.2000.903746>>. Citado na página 46.
- CRISTIANINI, N.; SHAW-TAYLOR, J. *et al.* An introduction to support vector machines and other kernel-based learning methods. Cambridge university press, 2000. Citado na página 36.
- DEVILLE, M.; LABROSSE, G. An algorithm for the evaluation of multidimensional (direct and inverse) discrete chebyshev transforms. *Journal of Computational and Applied Mathematics*, v. 8, n. 4, p. 293 – 304, 1982. ISSN 0377-0427. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0771050X82900559>>. Citado na página 56.
- DEZA, E. D. M. M. *Encyclopedia of Distances*. 4ed.. ed. Springer, 2016. ISBN 3662528436,978-3-662-52843-3,978-3-662-52844-0. Disponível em: <<https://www.springer.com/gp/book/9783642002342>>. Citado na página 59.
- DONIAK, M. H. *Estudo da Transformada de Walsh-Hadamard Aplicada à Transmissão OFDM*. Dissertação (Mestre em Engenharia Elétrica pelo Programa de Pós-Graduação em Engenharia Elétrica) — Universidade Federal de Santa Catarina, sep 2006. Citado na página 48.
- FERREIRA, E.; BALVEDI, R. P. A.; BONAN, P. R.; MEDEIROS, L. G. E.; LIMA, A. M. N.; NEFF, H. F.; OLIVEIRA, L. C.; CASTELLANO, L.; MOREIRA, C. D. S. Data processing influence in surface plasmon resonance biosensor for leishmania spp. detection. *Biosensors*, 2018. Citado na página 32.
- FERREIRA, E.; LIMA, J.; ALVES-BALVEDI, R. P.; BONAN, P. R.; MEDEIROS, E.; GOULART, L.; LIMA, A.; NEFF, H.; OLIVEIRA, L.; CASTELLANO, L.; ARAÚJO, A.; MOREIRA, C. Leishmania spp. detection using a surface plasmon resonance biosensor. *Proceedings*, v. 1, n. 4, 2017. ISSN 2504-3900. Disponível em: <<http://www.mdpi.com/2504-3900/1/4/536>>. Citado na página 61.

- FINO; ALGAZI. Unified matrix treatment of the fast walsh-hadamard transform. *IEEE Transactions on Computers*, C-25, n. 11, p. 1142–1146, Nov 1976. ISSN 2326-3814. Disponível em: <<https://doi.org/10.1109/TC.1976.1674569>>. Citado na página 47.
- GUIDE, M. U. The mathworks. *Inc., Natick, MA*, v. 5, p. 333, 2020. Citado 2 vezes nas páginas 51 e 113.
- HAIR, J.; BLACK, W.; BABIN, B.; ANDERSON, R. *Multivariate Data Analysis*. [S.l.]: Pearson Education Limited, 2013. (Always learning). ISBN 9781292021904. Citado na página 41.
- HAYKIN, S.; NETWORK, N. A comprehensive foundation. *Neural networks*, v. 2, n. 2004, p. 41, 2004. Citado na página 36.
- HOMOLA, J. Surface plasmon resonance sensors for detection of chemical and biological species. *Chemical Reviews*, v. 108, n. 2, p. 462–493, 2008. PMID: 18229953. Disponível em: <<https://doi.org/10.1021/cr068107d>>. Citado na página 23.
- IMANDOUST, S. B.; BOLANDRAFTAR, M. Application of k-nearest neighbor (knn) approach for predicting economic events theoretical background. *Int J Eng Res Appl*, v. 3, p. 605–610, 01 2013. Citado na página 36.
- JAMES DANIELA WITTEN, T. H. R. T. G. *An Introduction to Statistical Learning with Applications in R*. Springer, New York, NY, 2013. v. 103. (Springer Texts in Statistics, v. 103). ISBN 978-1-4614-7138-7. Disponível em: <<https://link.springer.com/book/10.1007/978-1-4614-7138-7>>. Citado na página 38.
- JANG, D.; CHAE, G.; SHIN, S. Analysis of surface plasmon resonance curves with a novel sigmoid-asymmetric fitting algorithm. *Sensors*, v. 15, n. 10, p. 25385–25398, 2015. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/15/10/25385>>. Citado na página 22.
- JUNIOR, G. R. de A. *Metodologia para subamostragem em grandes bancos de dados amostrais complexos para realização de testes de hipóteses*. Dissertação ((Mestrado em Estatística)) — Universidade de Brasília, oct 2017. Universidade de Brasília, Instituto de Ciências Exatas, Departamento de Estatística. Citado na página 38.
- KOPRIVA, D. A. Implementing spectral methods for partial differential equations. *Implementing Spectral Methods for Partial Differential Equations: Algorithms for Scientists and Engineers, Scientific Computation*, ISBN 978-90-481-2260-8. Springer Science+Business Media B.V., 01 2009. Disponível em: <<https://doi.org/10.1007/978-90-481-2261-5>>. Citado na página 47.
- KRETSCHMANN, E.; RAETHER, H. Radiative decay of nonradiative surface plasmons excited by light. *Z. Naturforsch. A*, v. 23, p. 2135, 1968. Citado 3 vezes nas páginas 9, 29 e 30.
- LEE, Y. K.; LEE, K.-S.; KIM, W. M.; SOHN, Y.-S. Detection of amyloid-42 using a waveguide-coupled bimetallic surface plasmon resonance sensor chip in the intensity measurement mode. *PLOS ONE*, Public Library of Science, v. 9, n. 6, p. 1–7, 06 2014. Disponível em: <<https://doi.org/10.1371/journal.pone.0098992>>. Citado 2 vezes nas páginas 26 e 27.

- LIDIYA, A. E.; RAJA, R. V. J.; PHAM, V. D.; NGO, Q. M.; VIGNESWARAN, D. Detecting hemoglobin content blood glucose using surface plasmon resonance in d-shaped photonic crystal fiber. *Optical Fiber Technology*, v. 50, p. 132 – 138, 2019. ISSN 1068-5200. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1068520019300483>>. Citado na página 22.
- LIU, X.; BAIOCCHI, O. A comparison of the definitions for smart sensors, smart objects and things in iot. p. 1–4, Oct 2016. Disponível em: <<https://doi.org/10.1109/IEMCON.2016.7746311>>. Citado na página 22.
- LOUREIRO, F. C.; NEFF, H.; MELCHER, E. U.; ROQUE, R. A.; FIGUEIREDO, R. M. de; THIRSTRUP, C.; BORRE, M. B.; LIMA, A. M. Simplified immunoassay for rapid dengue serotype diagnosis, revealing insensitivity to non-specific binding interference. *Sensing and Bio-Sensing Research*, v. 13, p. 96 – 103, 2017. ISSN 2214-1804. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2214180416301477>>. Citado na página 62.
- LUKOSE, J.; SHETTY, V.; BALLAL, M.; CHIDANGIL, S.; SINHA, R. K. Real-time and rapid detection of salmonella typhimurium using an inexpensive lab-built surface plasmon resonance setup. *Laser Physics Letters*, IOP Publishing, v. 15, n. 7, p. 075701, may 2018. Disponível em: <<https://doi.org/10.1088/1612-202x/aabed8>>. Citado na página 22.
- MARSHLAND, S. Machine learning : an algorithmic perspective. CRC Press, 2015. Citado 2 vezes nas páginas 24 e 35.
- MARTINEZ, W. L.; MARTINEZ, A. R. Computational statistics handbook with matlab. *Computational Statistics Handbook with Matlab, Third Edition*, 01 2016. Disponível em: <<https://10.1201/9781420035636>>. Citado na página 40.
- MCCLANING, K.; VITO, T. *Radio Receiver Design*. 1. ed. [S.l.]: Nobel Publishing, 2001. (1). ISBN 1-884932-07-X. Citado na página 54.
- MILJKOVIC, D. Brief review of self-organizing maps. p. 1061–1066, May 2017. Disponível em: <<https://10.23919/MIPRO.2017.7973581>>. Citado na página 36.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. Introduction to linear regression analysis. John Wiley & Sons, v. 821, 2012. Citado na página 93.
- MOREIRA, C. d. S. *Projeto e realização de um biochip óptico para aplicações biológicas baseado no princípio de ressonância de plásmos de superfície [thesis]*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2010. Citado na página 50.
- MORETTIN, W. d. O. B. P. A. Estatística básica. Saraiva, 2010. Citado na página 42.
- MULAK, P.; TALHAR, N. R. Analysis of distance measures using k-nearest neighbor algorithm on kdd dataset. *International Journal of Science and Research*, p. 2101–2104, 2015. ISSN 2319-7064. Citado na página 59.
- MUSLIM, M. A.; NURZAHPUTRA, A.; PRASETIYO, B. Improving accuracy of c4.5 algorithm using split feature reduction model and bagging ensemble for credit card risk prediction. *2018 International Conference on Information and Communications Technology (ICOIACT)*, p. 141–145, Mar 2018. Disponível em: <<https://10.1109/ICOIACT.2018.8350753>>. Citado na página 36.

- NENNINGER, G. G.; PILIARIK, M.; HOMOLA, J. Data analysis for optical sensors based on spectroscopy of surface plasmons. *Measurement Science and Technology*, IOP Publishing, v. 13, n. 12, p. 2038–2046, nov 2002. Disponível em: <<https://doi.org/10.1088-2F0957-0233-2F13-2F12-2F332>>. Citado na página 33.
- OLIVEIRA, L. C. *Construção e Caracterização de Sensores SPR: influências da camada metálica e do substrato dieletrico*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2016. Citado 3 vezes nas páginas 22, 34 e 50.
- OLIVEIRA, L. C.; LIMA, A. M. N.; THIRSTRUP, C.; NEFF, H. F. Surface plasmon resonance sensors: A materials guide to design and optimization. Springer, 2015. Citado na página 23.
- OLIVEIRA, L. C.; MOREIRA, C. d. S.; THIRSTRUP, C.; MELCHER, E. U. K.; LIMA, A. M. N.; NEFF, H. A surface plasmon resonance biochip that operates both in the angular and wavelength interrogation modes. *IEEE Transactions on Instrumentation and Measurement*, v. 62, n. 5, p. 1223–1232, May 2013. ISSN 1557-9662. Disponível em: <<https://doi.org/10.1109/TIM.2012.2232433>>. Citado na página 50.
- Oliveira, L. C.; Pereira, E. G.; Oliveira, R. C.; Morais, M. R. A.; Lima, A. M. N.; Neff, H. Spr sensor for tampering detection in biofuels. In: *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*. [S.l.: s.n.], 2015. p. 1471–1476. Citado na página 33.
- OLIVEIRA, N. A. P. de. *Single image super-resolution method based on linear regression and Box-Cox transformation*. Dissertação (Mestrado em Ciência da Computação pelo Programa de Pós-Graduação em Ciência da Computação) — Universidade Federal Rural do Semi-Arido, dez 2018. Citado na página 40.
- ORGANIZATION, W. H. *Dengue and severe dengue*. 2020. Disponível em: <https://www.who.int/health-topics/dengue-and-severe-dengue#tab=tab_1>. Citado na página 62.
- ORGANIZATION, W. H. *Epidemiological situation of Leishmaniasis*. 2020. Disponível em: <<https://www.who.int/leishmaniasis/burden/en/>>. Citado na página 62.
- OTTO, A. Excitation of nonradiative surface plasma waves in silver by the method of frustrated total reflection. *Zeitschrift für Physik A Hadrons and nuclei*, v. 216, n. 4, p. 398–410, Aug 1968. ISSN 0939-7922. Disponível em: <<https://doi.org/10.1007/BF01391532>>. Citado na página 29.
- OWEGA, S.; POITRAS, D. Local similarity matching algorithm for determining spr angle in surface plasmon resonance sensors. *Sensors and Actuators B: Chemical*, v. 123, n. 1, p. 35 – 41, 2007. ISSN 0925-4005. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925400506005077>>. Citado na página 33.
- PARSONS, S.; BOONMAN, A. M.; OBRIST, M. K. Advantages and Disadvantages of Techniques for Transforming and Analyzing Chiropteran Echolocation Calls. *Journal of Mammalogy*, v. 81, n. 4, p. 927–938, 11 2000. ISSN 0022-2372. Disponível em: <[https://doi.org/10.1644/1545-1542\(2000\)081<0927:AADOTF>2.0.CO;2](https://doi.org/10.1644/1545-1542(2000)081<0927:AADOTF>2.0.CO;2)>. Citado na página 57.

- PATIL, P. O.; PANDEY, G. R.; PATIL, A. G.; BORSE, V. B.; DESHMUKH, P. K.; PATIL, D. R.; TADE, R. S.; NANGARE, S. N.; KHAN, Z. G.; PATIL, A. M.; MORE, M. P.; VEERAPANDIAN, M.; BARI, S. B. Graphene-based nanocomposites for sensitivity enhancement of surface plasmon resonance sensor for biological and chemical sensing: A review. *Biosensors and Bioelectronics*, v. 139, p. 111324, 2019. ISSN 0956-5663. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0956566319303896>>. Citado na página 29.
- PILIARIK, M.; VAISOCHEROVÁ, H.; HOMOLA, J. Surface plasmon resonance biosensing. Humana Press, Totowa, NJ, p. 65–88, 2009. Disponível em: <https://doi.org/10.1007/978-1-60327-567-5_5>. Citado 2 vezes nas páginas 23 e 30.
- PRASATH, V. B. S.; ALFEILAT, H. A. A.; LASASSMEH, O.; HASSANAT, A. B. A. Distance and similarity measures effect on the performance of k-nearest neighbor classifier - A review. *CoRR*, abs/1708.04321, 2017. Disponível em: <<http://arxiv.org/abs/1708.04321>>. Citado na página 37.
- REIGOTA, N. d. S. *Comparação da Transformada Wavelet Discreta e da Transformada do Cosseno, para compressão de imagens de impressão digital*. Dissertação (Mestrado) — Universidade de São Paulo, 2007. Citado na página 45.
- ROY, R.; GOATMAN, M.; KHANGURA, K. User-centric design and kansei engineering. *CIRP Journal of Manufacturing Science and Technology*, v. 1, n. 3, p. 172 – 178, 2009. ISSN 1755-5817. Design Synthesis. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1755581708000588>>. Citado na página 113.
- SANTOS, A. M. *Investigando a Combinação de Técnicas de Aprendizado Semi-Supervisionado e classificação Hierárquica Multirrótulo*. Tese (Doutorado) — Programa de Pós Graduação em Sistemas e Computação - Universidade Federal Do Rio Grande do Norte, 2012. Citado 2 vezes nas páginas 35 e 36.
- SCHASFOORT, R. B. M. (Ed.). *Handbook of Surface Plasmon Resonance*. The Royal Society of Chemistry, 2017. P001-524 p. ISBN 978-1-78262-730-2. Disponível em: <<http://dx.doi.org/10.1039/9781788010283>>. Citado na página 29.
- SHIRADO, W.; DE, M.; MOREIRA, A.; PALMA, J.; BARBON, S. Estudo comparativo entre algoritmos das transformadas discretas de fourier e wavelet. *Revista Brasileira de Computação Aplicada*, v. 7, p. 97, 10 2015. Disponível em: <<https://doi.org/10.5335/rbca.2015.4880>>. Citado na página 45.
- SILVA, M. L. F. da. *Estudo e Desenvolvimento de Sensor em Fibra Óptica Polimérica Baseada na Ressonância de Plasmons de Superfície*. Dissertação (Mestrado) — Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, mar 2017. Programa de Pós-Graduação em Engenharia Elétrica. Citado na página 29.
- SONG, E. Y.; LEE, K. B. Ieee 1451.5 standard-based wireless sensor networks. Springer Berlin Heidelberg, Berlin, Heidelberg, p. 243–271, 2010. Disponível em: <https://doi.org/10.1007/978-3-642-12707-6_11>. Citado na página 22.
- SOUSA, T.; OLIVEIRA, L.; NEFF, H.; LABORDE, H.; LIMA, A. Numerical tool for estimating the dielectric constant, the thickness, and the coverage of immobilized

inhomogeneous protein films on gold in aqueous solution. *Appl. Opt.*, OSA, v. 57, n. 24, p. 6866–6875, Aug 2018. Citado na página 56.

SOUZA, L. C. de; SOUZA, R. M. C. R. de; AMARAL, G. J. A. do. Dynamic clustering of interval data based on hybrid lq l q distance. *Knowledge and Information Systems*, May 2019. ISSN 0219-3116. Disponível em: <<https://doi.org/10.1007/s10115-019-01367-w>>. Citado na página 41.

SUJA, K. P.; TARHUNI, N.; AHMAD, A.; SHIGINAH, F. B. Computationally efficient, robust watermarking using fast walsh hadamard transform. In: *Proceedings of the 7th International Conference on Computing Communication and Networking Technologies*. New York, NY, USA: Association for Computing Machinery, 2016. (ICCCNT '16). ISBN 9781450341790. Disponível em: <<https://doi.org/10.1145/2967878.2973276>>. Citado na página 56.

SUYKENS, J. A. Chapter 13 - introduction to machine learning. Elsevier, v. 1, p. 765 – 773, 2014. ISSN 2351-9819. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780123965028000139>>. Citado 2 vezes nas páginas 35 e 36.

THEPUDOM, T.; LERTVACHIRAPAIBOON, C.; SHINBO, K.; KATO, K.; KANEKO, F.; KERDCHAROEN, T.; BABA, A. Surface plasmon resonance-enhanced photoelectrochemical sensor for detection of an organophosphate pesticide chlorpyrifos. *MRS Communications*, Cambridge University Press, v. 8, n. 1, p. 107–112, 2018. Disponível em: <<https://doi.org/10.1557/mrc.2017.131>>. Citado na página 22.

TIWARI, P. B.; ÜREN, A.; HE, J.; DARICI, Y.; WANG, X. Note: Model identification and analysis of bivalent analyte surface plasmon resonance data. *Review of Scientific Instruments*, v. 86, n. 10, p. 106107, 2015. Disponível em: <<https://doi.org/10.1063/1.4933318>>. Citado 3 vezes nas páginas 25, 27 e 56.

TOURETZKY, D. S.; MOZER, M. C.; HASSELMO, M. E. Advances in neural information processing systems 8: Proceedings of the 1995 conference. Mit Press, v. 8, 1996. Citado na página 36.

WANG, J.; SU, X. An improved k-means clustering algorithm. p. 44–46, May 2011. Disponível em: <<https://10.1109/ICCSN.2011.6014384>>. Citado na página 36.

WANG, M.; LI, H.; XU, T.; YU, M.; LI, G.; ZHENG, H.; WU, J.; XU, J. Sensing and manipulation of bianisotropic biomolecules using a surface plasmon resonance based optical fiber sensor. *Journal of Lightwave Technology*, v. 36, n. 24, p. 5927–5934, 2018. Cited By 2. Disponível em: <<https://ieeexplore.ieee.org/document/8531746>>. Citado na página 22.

YAN, Y.; CHEN, L.; TAY, P. C.; TANAKA, M. L. Molecule binding parameter estimation from surface plasmon resonance measurements. p. 1–9, 2017. Citado 2 vezes nas páginas 25 e 27.

Yang, C.; Zhu, X.; Ahmad, Z.; Wang, L.; Qiao, J. Design of incremental echo state network using leave-one-out cross-validation. *IEEE Access*, v. 6, p. 74874–74884, 2018. Citado na página 60.

ZAINUDDIN, N. H.; FEN, Y. W.; ALWAHIB, A. A.; YAACOB, M. H.; BIDIN, N.; OMAR, N. A. S.; MAHDI, M. A. Detection of adulterated honey by surface plasmon resonance optical sensor. *Optik*, v. 168, p. 134 – 139, 2018. ISSN 0030-4026. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0030402618305412>>. Citado na página 22.

ZHANG, H.; LIU, W.; LIU, Z.; JU, Y.; XU, M.; ZHANG, Y.; WU, X.; GU, Q.; WANG, Z.; XU, J. Discovery of indoleamine 2,3-dioxygenase inhibitors using machine learning based virtual screening. *Med. Chem. Commun.*, The Royal Society of Chemistry, v. 9, p. 937–945, 2018. Disponível em: <<http://dx.doi.org/10.1039/C7MD00642J>>. Citado 3 vezes nas páginas 26, 27 e 56.

ZHANG, Y.; SHUAI, Z.; ZHOU, H.; LUO, Z.; LIU, B.; ZHANG, Y.; ZHANG, L.; CHEN, S.; CHAO, J.; WENG, L.; FAN, Q.; FAN, C.; HUANG, W.; WANG, L. Single-molecule analysis of microrna and logic operations using a smart plasmonic nanobiosensor. *Journal of the American Chemical Society*, v. 140, n. 11, p. 3988–3993, 2018. PMID: 29504757. Disponível em: <<https://doi.org/10.1021/jacs.7b12772>>. Citado 2 vezes nas páginas 26 e 27.

Apêndices

APÊNDICE A – Ferramenta para Geração de Respostas Baseado em Sensores SPR

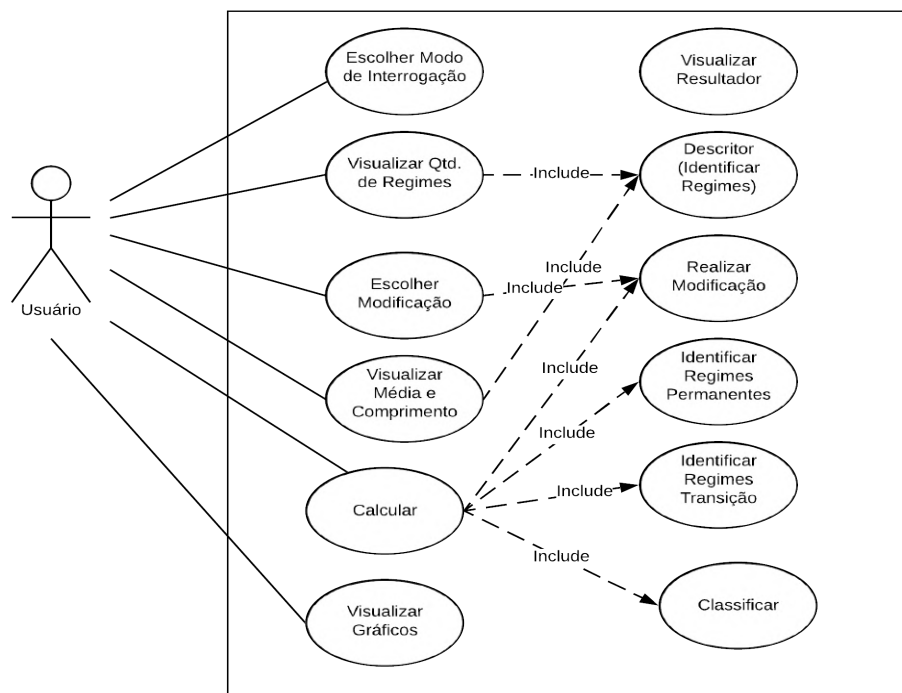
A construção de uma ferramenta computacional para classificar os sensorgramas de um sensor SPR faz-se necessário para dar um suporte aos estudos das técnicas e como também ajudar a classificar as respostas fornecidas pelo sensor. O desenvolvimento de ferramentas voltados para os sensorgramas necessitam de uma eficiência na implementação, que envolve a colaboração entre especialistas das áreas envolvidas para ajudar na concepção de ferramentas intuitivas para fornecer respostas precisas e necessidades específicas para o processo de classificação das substancias do sensorgrama.

Para conseguir atender a todos os especialistas, tais ferramentas tem o seu desenvolvimento centrado ao usuário (ROY; GOATMAN; KHANGURA, 2009). Nesse sentido, a ferramenta denominada de SMART SPR, foi desenvolvida para disponibilizar uma interface útil baseado nas necessidades e experiências vivenciadas pelos integrantes do laboratório de sensores da Universidade Federal Rural do Semi-Árido (UFERSA). A ferramenta desenvolvida foi utilizando o próprio MATLAB, em que é disponibilizado métodos para interface gráfica e a utilização das funções (GUIDE, 2020).

Com o desenvolvimento permitiu inserir arquivos para ler os sensorgramas, como também optar pelo modo de interrogação em que se deseja obter a classificação e as respostas. A implementação e utilização de funções na própria ferramenta, para permitir a geração de novos sensorgramas a partir das alterações de ruído, posição de ressonância e tempo e apresentar em Gráficos em 2D, para visualizar os sensorgramas inseridos e como também ver a classificação dos regimes permanentes estão presentes na ferramenta.

Para a implementação da ferramenta, foi criado um diagrama de caso de uso. Na Figura 66 apresenta-se o diagrama de caso de uso. O diagrama define as funcionalidades da ferramenta, o usuário pode escolher o modo de interrogação que são AIM e WIM, visualizar a média dos regimes permanentes como também o seu tamanho com o descritor TSD; pode reconhecer os regimes permanentes e de transição com a utilização da regressão linear e, como também o usuário pode classificar o sensorgrama com a utilização do algoritmo k -NN.

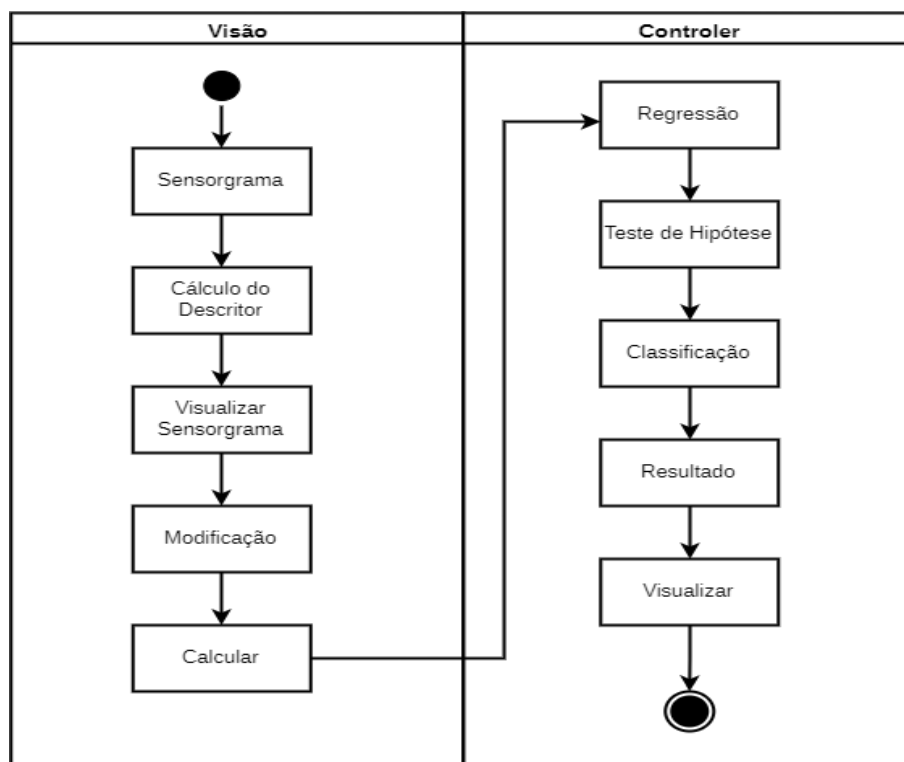
Figura 66 – Diagrama de caso de uso para implementação da ferramenta SMART SPR.



Fonte: Autoria Própria (2020).

O diagrama de atividades é apresentado na Figura 67 em que é apresenta-se uma série de passos a ser executado pela ferramenta. A ferramenta contém duas camadas, em que é possível observar na Figura 67, em que constitui-se de uma visão e um *controller*, em que na parte da visão o usuário visualiza todos os campos e é possível fazer as alterações, e tem a parte do *controller*, em que é realizado a série de cálculos para ser realizado, como também visualizar em formas de gráficos e o resultado final da classificação do sensorgrama.

Figura 67 – Diagrama de atividades para execução da ferramenta SMART SPR.

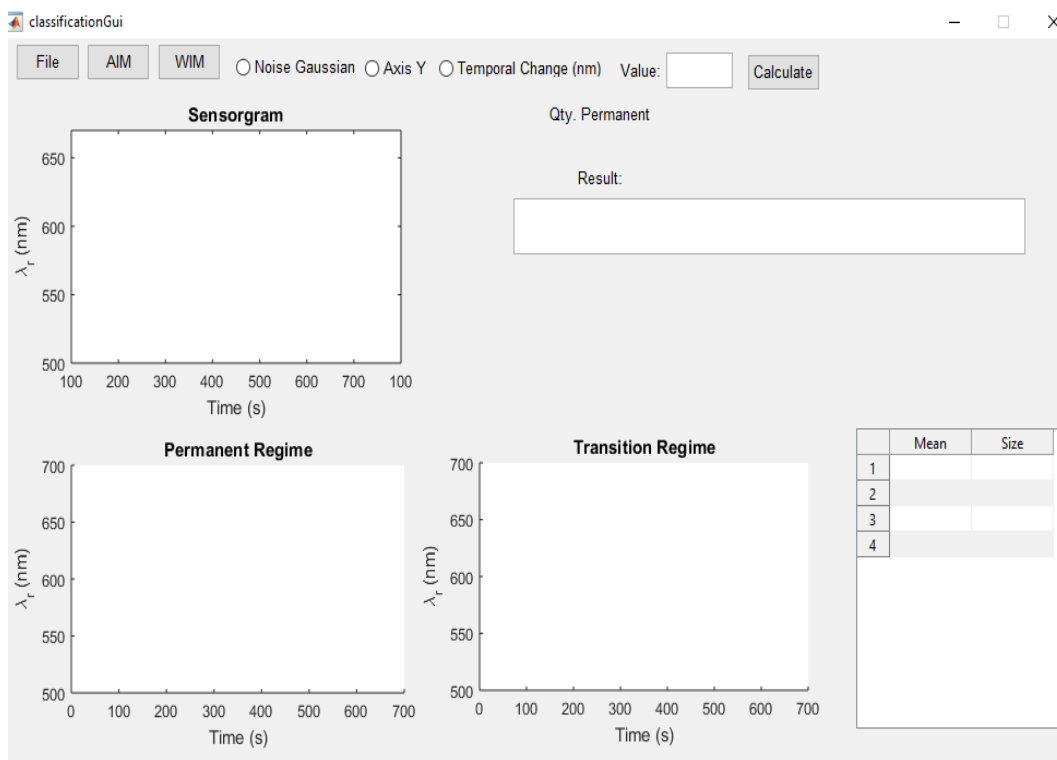


Fonte: Autoria Própria (2020).

Desse modo, caso o usuário queira fazer uma modificação no sensorgrama, é possível realizar essa modificação e visualizar o comportamento do mesmo em um gráfico.

Na Figura 68 é visualizado a tela inicial da ferramenta, observa-se que apresenta os dois modos de interrogação em que pode ser escolhido. A ferramenta também contém uma opção para ler arquivos de extensão do tipo ".txt", e nessa opção a visualização do sensorgrama é exibido em um gráfico.

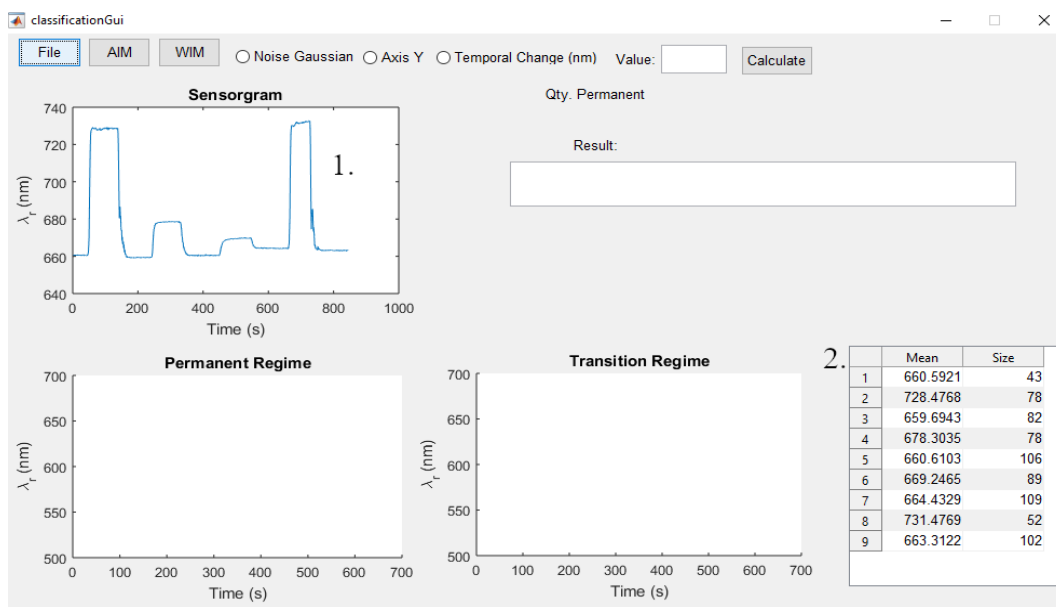
Figura 68 – Tela Inicial da ferramenta SMART SPR.



Fonte: Autoria Própria (2020).

A ferramenta apresenta a quantidade de regimes permanentes em um sensorgrama, como é possível visualizar em uma tabela ao lado, a média e o seu comprimento do regime permanente. Na Figura 68 é possível observar essas informações e como também gráficos em que identifica os regimes permanentes como os de transição. Nessa ferramenta é possível visualizar quais substâncias foram inseridas, em que identifica o PPBIO com os polímeros PC, TOPAS e PMMA, como também com o vidro BK7 e os metais por ouro (Au), prata (Ag), cobre (Cu) e alumínio (Al) que compõe o PPBIO.

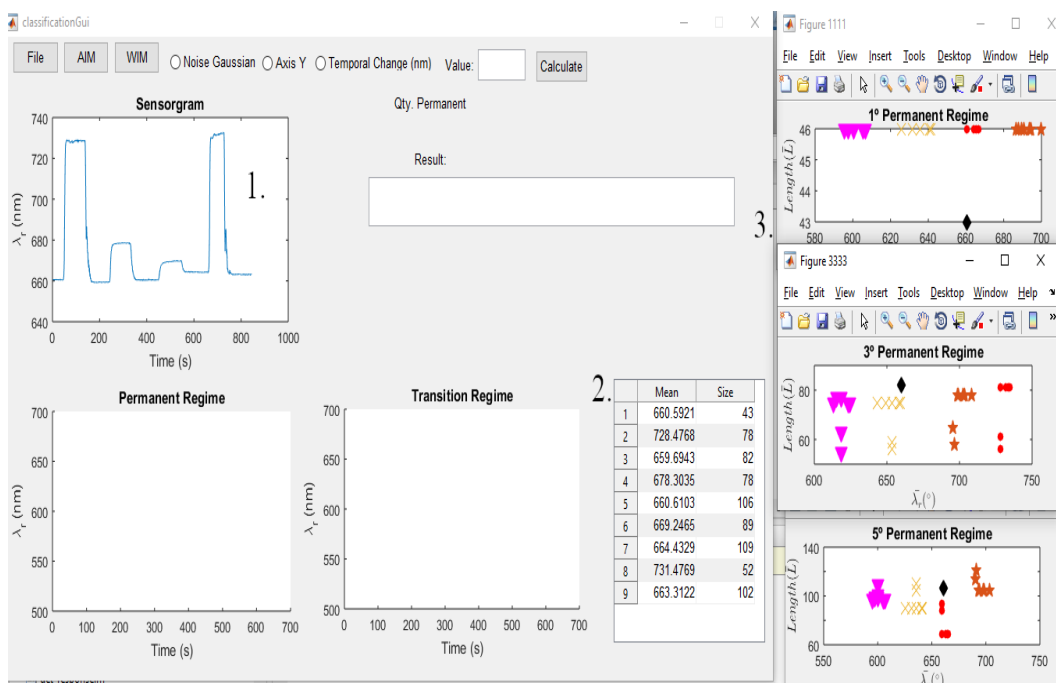
Figura 69 – Inserção do sensorgrama com a identificação dos regimes permanentes na ferramenta SMART SPR.



Fonte: Autoria Própria (2020).

Na Figura 69 é possível observar um sensorgrama no modo WIM com o vidro BK7 e o metal prata. Assim que é inserido o sensorgrama, é identificado quantos regimes permanentes tem ao lado e como também é possível visualizar a média e o tamanho do regime permanente.

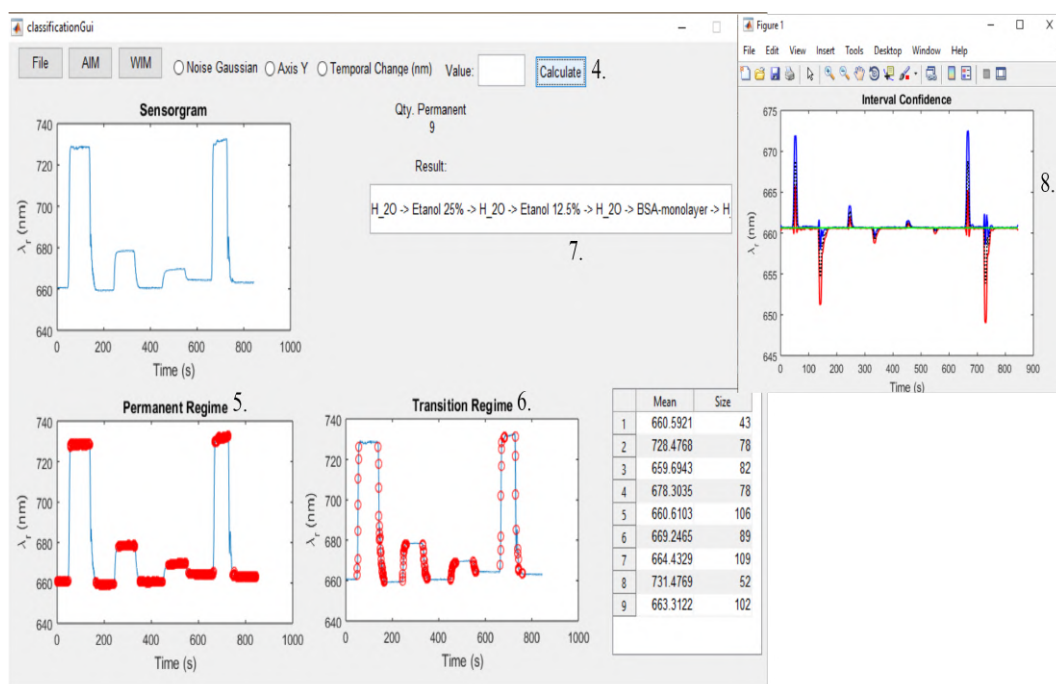
Figura 70 – Conjunto de dados de treinamento com o descritor TSD em que é separado os regimes permanentes para classificação do sensorgrama.



Fonte: Autoria Própria (2020).

Na Figura 70 é possível visualizar o conjunto de dados de treinamento e o para teste de acordo com a quantidade de regimes permanentes que se tem. Observa-se na Figura 70 o primeiro, terceiro e quinto regime permanente para classificação desse sensorgrama e identificação das substâncias inseridas no sensorgrama.

Figura 71 – Identificação dos regimes permanentes, como também a classificação das substâncias inserida e o intervalo de confiança para o vidro BK7 e o metal Ag.



Fonte: Autoria Própria (2020).

Quando é clicado no botão "calculate", é realizado todos os procedimentos para o reconhecimento do sensorgrama, como regimes permanentes e de transição. Na Figura 71 é possível identificar as substâncias inseridas, os regimes permanentes e de transição de acordo com o sensorgrama inserido e o intervalo de confiança com o teste de hipótese, é possível observar em que foi classificado corretamente as substâncias inseridas, diante disso, essa ferramenta é de grande importância para o desenvolvimento e reconhecimento das respostas dos sensores SPR.