



LUIZ PAULO DE OLIVEIRA SOUSA

**ANÁLISE DE DADOS DE LÍNGUA ELETRÔNICA
UTILIZANDO CURVAS PRINCIPAIS**

LAVRAS – MG

2019

LUIZ PAULO DE OLIVEIRA SOUSA

**ANÁLISE DE DADOS DE LÍNGUA ELETRÔNICA UTILIZANDO CURVAS
PRINCIPAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas e Automação, área de concentração em Engenharia de Sistemas e Automação, para a obtenção do título de Mestre.

Prof. Dr. Danton Diego Ferreira
Orientador

Prof. Dr. Juliano Elvis Oliveira
Coorientador

**LAVRAS – MG
2019**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Sousa, Luiz Paulo de Oliveira

Análise de dados de Língua Eletrônica utilizando Curvas Principais / Luiz Paulo de Oliveira Sousa. 2019.

71 p. : il.

Orientador(a): Prof. Dr. Danton Diego Ferreira.

Coorientador(a): Prof. Dr. Juliano Elvis Oliveira.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2019.

Bibliografia.

1. Língua eletrônica. 2. Curvas principais. 3. Reconhecimento de padrões. I. Ferreira, Danton Diego. II. Oliveira, Juliano Elvis. III. Título.

LUIZ PAULO DE OLIVEIRA SOUSA

**ANÁLISE DE DADOS DE LÍNGUA ELETRÔNICA UTILIZANDO CURVAS
PRINCIPAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas e Automação, área de concentração em Engenharia de Sistemas e Automação, para a obtenção do título de Mestre.

APROVADA em 06 de Setembro de 2019.

Prof. Dr. Danton Diego Ferreira UFLA
Prof. Dr. Juliano Elvis Oliveira UFLA
Prof. Dr. Luciano Manhães de A. Filho UFJF
Prof. Dr. Leandro Rodrigues M. Silva UFJF

Prof. Dr. Danton Diego Ferreira
Orientador

Prof. Dr. Juliano Elvis Oliveira
Co-Orientador

**LAVRAS – MG
2019**

AGRADECIMENTOS

Agradeço primeiramente a Deus que encaminhou o meu caminho durante esta jornada e por ter me dado saúde e a força necessária para percorrer e superar as dificuldades.

Ao meu pai, que significou segurança a certeza de que não estou sozinho nessa caminhada. A minha mãe, heroína que me deu apoio e incentivo nas horas difíceis. E, aos meus familiares, pelo carinho e apoio.

Ao meu orientador, pela dedicação, apoio e confiança. Agradeço também pela amizade e bom relacionamento construído durante o período de desenvolvimento do trabalho.

Sou imensamente grato à Universidade Federal de Lavras (UFLA), pelo ambiente criativo e amigável que proporciona. Em especial ao Programa de Pós-Graduação em Engenharia de Sistemas e Automação (PPGESISA) e ao departamento pela oportunidade de crescimento acadêmico.

Aos amigos de Caldas Novas - GO, por fazerem parte da minha formação pessoal. Aos colegas da república Bomgodo de Lavras pela convivência.

A todos os professores que, diretamente e/ou indiretamente, deram boas contribuições em meu processo de formação profissional. Em destaque o professor Dr. Cleiton Antônio Nunes por ter sido sempre solícito e atencioso quando sua ajuda se fez necessário.

E a todos os demais, que mesmo de maneira singela contribuíram de algum modo para o meu crescimento, sou grato.

RESUMO

Os sistemas de língua eletrônica são inspirados nos sistemas de reconhecimento biológico, onde as técnicas sensoriais e instrumentais são utilizadas para determinar os sabores ou as substâncias presentes nas amostras analisadas. Basicamente, as línguas eletrônicas são compostas por arranjos de sensores que atuam como coletores de informações das amostras utilizadas. Os dados coletados pelos arranjos de sensores correspondem a informações valiosas das amostras, úteis em reconhecimento, identificação ou quantificação dos diversos constituintes presentes, podendo ser obtidas por meio dos métodos sofisticados de reconhecimento de padrões e mineração de dados. O objetivo geral da pesquisa foi investigar dados de um sistema de língua eletrônica usando Curvas Principais via o algoritmo k-segmentos não suave com o intuito de projetar um modelo que seja capaz de classificar diferentes concentrações e também identificar as substâncias analisadas. Essas substâncias são realçadores de sabor, que são utilizados para provocar um sabor marcante ao produto. Os realçadores analisados foram o Glutamato Monossódico, o Inosinato Dissódico e o Guanilato Dissódico. Estes realçadores têm sido alvo de pesquisas que têm o objetivo de substituir ou diminuir a quantidade de cloreto de sódio por esses realçadores, visando produtos mais saudáveis. O método proposto de classificador usando Curvas Principais via k-segmentos se mostrou capaz de extrair informações importantes sobre as análises amostrais. Inicialmente, as análises foram realizadas para determinar alguns parâmetros do modelo, como a quantidade de segmentos que compõem as curvas e quais entradas (variáveis) teriam mais relevância nessas classificações. Em seguida foram examinadas questões sobre o sistema de língua eletrônica usado (como as arquiteturas dos eletrodos que constituem a língua eletrônica e os sinais de frequência que a mesma capta), com intuito de identificar quais seriam mais discriminantes ao modelo. E, por fim, para efeitos de comparação deste modelo, foram projetados outros modelos empregando diferentes tipos de redes neurais artificiais (Perceptron Multi-Camadas, Função de Base Radial e Mapas Auto Organizáveis). Como métrica de validação cruzada foi utilizado o *hold-out* com vinte subconjuntos mutuamente excludentes, além da análise das matrizes de confusão dos melhores resultados alcançados. Os modelos com Curvas Principais atingiram em seus melhores resultados entre 88,02% e 91,39% de precisão em classificações de novos eventos (tanto nas análises das concentrações, quanto das substâncias) e a faixa de desvio padrão desses resultados, ficaram de 2,27% a 1,28%. Os modelos com Perceptron Multi-Camadas ficaram com percentuais de 74,50% a 97,03%, com desvio padrão entre 7,65% e 1,39%. Usando a Função de Base Radial, os modelos tiveram de 85,70% a 96,61% de precisão e o desvio padrão entre 2,80% e 0,83%. Com redes do tipo Mapas Auto Organizáveis, a precisão ficou de 80,78% a 87,03% com desvio entre 2,85% e 1,04%. Os modelos implementados usando redes neurais do tipo Perceptron Multi-Camadas (com 2 camadas ocultas) e Função de Base Radial tiveram percentuais de precisão superiores ao método proposto que usa Curvas Principais, porém com o custo computacional mais elevado. Em termos de operações matemáticas, o Perceptron Multi-Camadas realiza aproximadamente 6 vezes mais operações do que o modelo baseado em Curvas Principais, enquanto a Função de Base Radial realiza aproximadamente 158 vezes mais operações.

Palavras-chave: Língua eletrônica. Reconhecimento de padrões. Classificação. Curvas principais. Redes neurais artificiais.

ABSTRACT

Electronic tongue systems are inspired by biological recognition systems, where sensory and instrumental techniques are used to determine the flavors or substances in the analyzed samples. Basically, electronic tongue are composed of sensor arrays that act as information collectors of the samples used. The data collected by the sensor arrays correspond to valuable sample information, useful in recognizing, identifying or quantifying the various constituents present, and can be obtained through sophisticated pattern recognition and data mining methods. The core objective of the is Dissertation was to analyse data from an electronic tongue system using Principal Curves, through the k-segment algorithm. Based on Principal Curves, a model capable of classifying different concentrations and identifying the substances is developed. These substances are flavor enhancers, which are used to bring about a marked flavor to the product. The enhancers analyzed were Monosodium Glutamate, Disodium Inosinate and Disodium Guanylate. These enhancers have been the subject of research aimed at replacing or decreasing the amount of sodium chloride for healthier products. The proposed classifier method using Principal Curves proved to be able to extract important information from the samples analyzed. Initially, the analyzes were performed to determine some parameters of the model, such as the number of segments used to build up the principal curves and which inputs (variables) are more relevant in these classifications. The structure of the electronic tongue system used, such as the electrode architectures and the frequency were investigated, in order to identify which are more discriminating to the model. Finally, for comparison purposes, other models were designed employing different types of artificial neural networks (Multi-Layer Perceptron, Radial Base Function and Self-Organizing Maps). As a cross-validation metric we used hold-out with twenty mutually exclusive subsets, as well as the analysis of the confusion matrices of the best results achieved. The models based on Principal Curves achieved accuracy results between 88,02% and 91,39% in classifying new events (both concentration and substance analysis) and the standard deviation range of from 2,27% to 1,28%. Perceptron Multi-Layer models achieved percentages of 74,50% to 97,03%, with standard deviation between 7,65% and 1,39%. Using the Radial Base Function, the models had 85,70% to 96,61% accuracy and the standard deviation between 2,80% and 0,83%. With Self-Organizing Maps, the accuracy was 80,78% to 87,03% with a deviation between 2,85% and 1,04%. The models implemented using Perceptron Multi-Layer (with 2 hidden layers) and Radial Base Function neural networks had higher percentages of accuracy than the proposed method based on Principal Curves, but with the highest computational cost. The Perceptron Multi-Layer model performs 6 times operations more than the proposed method while the Radial Base Function requires about 158 times more.

Keywords: Electronic tongue. Pattern Recognition. Classification. Principal curves. Artificial neural networks.

LISTA DE FIGURAS

Figura 2.1 – Comparação entre o sistema biológico e o sistema eletrônico.	14
Figura 2.2 – Elementos de uma LE.	15
Figura 2.3 – Dados de uma Curva Principal.	21
Figura 2.4 – Regiões de Voroni.	23
Figura 2.5 – Fluxograma do algoritmo k-segmentos não suave.	24
Figura 2.6 – Representação do sistema nervoso em diagrama de blocos.	24
Figura 2.7 – Características de uma RNA.	25
Figura 2.8 – Modelo de um neurônio do tipo RBF.	27
Figura 2.9 – Características de uma rede SOM.	29
Figura 3.1 – Fluxograma do método proposto.	35
Figura 3.2 – Espaço de características em duas dimensões do eletrodo E1.	36
Figura 3.3 – Fluxograma do classificador que usa CPs.	36
Figura 3.4 – Fluxograma do classificador que usa MLP ou RBF.	37
Figura 3.5 – Fluxograma do classificador que usa SOM.	37
Figura 4.1 – Precisão dos subconjuntos de validação atingida conforme o número de segmentos.	41
Figura 4.2 – Teste ANOVA usando as médias de validação de GD/ID.	42
Figura 4.3 – Sinais captados por cada eletrodo em 2 dimensões para a substância GM.	43
Figura 4.4 – Curvas Principais extraídas com o modo 1.	51
Figura 4.5 – Curvas Principais extraídas de ID com E6 e F3.	53
Figura 1 – Curvas Principais de GM com E6 que teve $88,02 \pm 2,06\%$	63
Figura 2 – Curvas Principais de GD/ID com E1 que teve $91,15 \pm 1,34\%$	63
Figura 3 – Curvas Principais de ID com E1 que teve $88,75 \pm 2,27\%$	64
Figura 4 – Curvas Principais de GM com E4 e E6 que teve $87,46 \pm 1,64\%$	65
Figura 5 – Curvas Principais de GD/ID com E1 e E6 que teve $85,50 \pm 1,49\%$	65
Figura 6 – Curvas Principais de ID com E1 e E4 que teve $85,84 \pm 1,19\%$	66
Figura 7 – Curvas Principais de E1 que teve $91,39 \pm 1,28\%$	67
Figura 8 – Curvas Principais de E5 que teve $89,33 \pm 1,32\%$	67

LISTA DE TABELAS

Tabela 3.1 – Arquiteturas dos eletrodos da LE.	31
Tabela 3.2 – Concentrações dos realçadores de sabor (mol L^{-1}).	32
Tabela 3.3 – Segmentos de Frequências que serão analisadas (Hz).	34
Tabela 4.1 – Percentual médio e desvio padrão da precisão dos subconjuntos de validação obtidos com GM (%).	40
Tabela 4.2 – Percentual médio e desvio padrão da precisão dos subconjuntos de validação alcançados com GD/ID (%).	40
Tabela 4.3 – Percentual médio e desvio padrão da precisão dos subconjuntos de validação atingidos com ID (%).	41
Tabela 4.4 – Percentual médio e desvio padrão da precisão dos subconjuntos de validação conforme a divisão dos dados (%).	42
Tabela 4.5 – Percentual médio e desvio padrão da precisão dos subconjunto de validação de acordo com as arquiteturas dos eletrodos (%).	44
Tabela 4.6 – Matriz de confusão dos subconjuntos de validação de GM com E6 (%). . .	45
Tabela 4.7 – Matriz de confusão dos subconjuntos de validação de GD/ID com E1 (%). .	45
Tabela 4.8 – Matriz de confusão dos subconjuntos de validação de ID com E1 (%). . . .	46
Tabela 4.9 – Percentual médio e desvio padrão da precisão dos subconjunto de validação combinando os eletrodos de GM (%).	47
Tabela 4.10 – Percentual médio e desvio padrão da precisão dos subconjunto de validação combinando os eletrodos de GD/ID (%).	48
Tabela 4.11 – Percentual médio e desvio padrão da precisão dos subconjunto de validação combinando os eletrodos de ID (%).	49
Tabela 4.12 – Percentual médio e desvio padrão da precisão dos subconjunto de validação conforme o número de eletrodos (%).	49
Tabela 4.13 – Matriz de confusão dos subconjuntos de validação de GM com E4 e E6 (%). .	50
Tabela 4.14 – Matriz de confusão dos subconjuntos de validação de GD/ID com E1 e E6 (%).	50
Tabela 4.15 – Matriz de confusão dos subconjuntos de validação de ID com E1 e E4 (%). .	51
Tabela 4.16 – Percentual médio e desvio padrão da precisão dos subconjunto de validação de combinando as Frequências (%).	53
Tabela 4.17 – Resultados de treino das classificações das substâncias (%).	54

Tabela 4.18 – Matriz de confusão E1 (%).	55
Tabela 4.19 – Matriz de confusão E5 (%).	55
Tabela 4.20 – Percentual médio e desvio padrão da precisão dos subconjunto de validação dos classificadores (%).	57
Tabela 4.21 – Equações para obtenção do custo computacional dos modelos.	58
Tabela 4.22 – Custo computacional dos modelos propostos.	59

SUMÁRIO

1	Introdução	9
1.1	Objetivos	12
1.2	Estrutura do Trabalho	13
2	Referencial Teórico	14
2.1	Língua Eletrônica	14
2.2	Estado da Arte	16
2.3	Reconhecimento de padrões	18
2.4	Curvas Principais	20
2.4.1	K-segmentos não suave	21
2.5	Redes Neurais Artificiais	24
2.5.1	Função de Base Radial	26
2.5.2	Mapas Auto Organizáveis	28
3	Metodologia	31
3.1	Língua Eletrônica	31
3.2	Amostras	32
3.3	Base de Dados	33
3.4	Método Proposto	34
4	Resultados	39
4.1	Identificando os parâmetros	39
4.2	Eletrodo mais discriminante	43
4.2.1	Combinando os eletrodos	46
4.3	Combinando as frequências	52
4.4	Classificando as substâncias	54
4.5	Comparação de métodos	55
5	Conclusões e Discussões	60
	APENDICE A – Curvas Principais dos melhores resultados	63
	REFERÊNCIAS	68

1 INTRODUÇÃO

Os primeiros sistemas inspirados em órgãos olfativos e gustativos foram propostos na década de 1980, e desde então tem-se observado um grande esforço da comunidade acadêmica para desenvolver ferramentas de modelamento e melhora das respostas sensoriais destes sistemas. O primeiro sistema de língua eletrônica (LE) foi proposto por Otto e Thomas (1985). O funcionamento desse sistema é baseado no princípio de reconhecimento biológico, onde técnicas sensoriais e instrumentais são utilizadas para determinar sabores ou substâncias presentes nas amostras analisadas (NERY; KUBOTA, 2016).

O sistema de LE é considerado um instrumento analítico, sendo capaz de caracterizar amostras aquosas com composição complexa de forma mais rápida. Isso, através de uma fusão de conhecimentos de vários ramos da ciência, como métodos de reconhecimento de padrões, tecnologias sensoriais, ferramentas quimiométricas e inteligência artificial (KARKRA et al., 2017). Em consequência disso, o interesse no conceito LEs aumentou consideravelmente nos últimos anos (WEI; WANG, 2014).

As LEs são compostas por arranjos de sensores, que são responsáveis pela coleta das informações contidas nas amostras, e o processamento dos dados gerados é feito usando técnicas matemáticas. Essas amostras são soluções aquosas e suas informações são consideradas como a “impressão digital” da solução, o que representa o conjunto de diferentes constituintes e suas informações, como reconhecimento, identificação ou quantificação de substâncias (NERY; KUBOTA, 2016; WEI; WANG, 2014).

Avanços importantes relacionados à melhoria da resposta das LEs vêm sendo relatados na literatura, normalmente direcionados para a implementação de novos conceitos e configurações baseados no desenvolvimento de novas plataformas sensoriais (GONZÁLEZ-CALABUIG; CETÓ; VALLE, 2016; JR et al., 2010). Em oposição às abordagens convencionais, os arranjos de sensores nas LEs são empregados visando à combinação de sensores de baixa seletividade que possam responder a diferentes componentes químicos e que forneça informações globais sobre a solução analisada (WEI; WANG; ZHANG, 2013; JR et al., 2010). Assim, dada a abrangência de respostas provenientes da análise por meio de LEs faz-se necessário o uso de ferramentas de processamento de dados capazes de revelar as informações específicas referentes as amostras.

A LE é uma tendência recente do campo sensorial, mas já existe uma grande quantidade de aplicações e em diversos segmentos. Na indústria alimentícia seu uso pode ser aplicado em

questões de controle de qualidade, monitoramento de processos, controle de envelhecimento e controle automatizado do sabor. Na área industrial, pode ser aplicada na detecção de explosivos e na avaliação de pureza de compostos. Na área medicinal, tem uso no monitoramento clínico, diagnóstico não invasivo, avaliação de produtos farmacêuticos e crescimento de culturas celulares (NERY; KUBOTA, 2016; WEI; WANG; ZHANG, 2013).

Nesse estudo, a aplicação da LE se enquadra na área alimentícia, onde as amostras analisadas são compostas por realçador de sabor. O incentivo dessa pesquisa vem do grande desafio para a indústria de alimentos em reduzir a concentração de sódio em seus produtos, devido sua influência direta nas propriedades sensoriais dos alimentos, como: sabor, cor, textura, e conservação. Por isso, a utilização de substitutos do cloreto de sódio vem sendo estudada como uma estratégia para a redução da concentração, entre elas, a utilização de realçador de sabor (AL-BARRACÍN et al., 2011; PEREIRA et al., 2015). Os realçadores analisados nessa dissertação de mestrado foram o Glutamato Monossódico, o Inosinato Dissódico e Guanilato Dissódico.

O foco das pesquisas realizadas nesse campo tem sido em três aspectos principais: o desenvolvimento da matriz de sensores, a otimização e comparação de métodos de reconhecimento de padrões e a aplicação em várias tarefas analíticas. Na análise sensorial é comum os dados terem uma grande variabilidade, dessa maneira métodos estatísticos mais sofisticados devem ser usados para estabelecer associações, fazer inferências ou caracterizar tais amostras (LIU et al., 2013; ZIELINSKI et al., 2014).

As contribuições registradas nessa dissertação de mestrado estão relacionadas com a comparação de desempenhos obtidos com diferentes métodos de reconhecimento de padrões com foco em Curvas Principais, aplicados em uma investigação analítica qualitativa e quantitativa. O reconhecimento de padrões tem como objetivo principal a classificação de padrões em um número de categorias ou classes. As técnicas de reconhecimento de padrões são usadas para classificar e/ou descrever padrões por meio do conjunto de propriedades ou características extraídas dos dados obtidos. Nesse estudo, essa investigação tem como objetivo criar um modelo que seja capaz de identificar automaticamente qual é a substância e sua concentração.

Dentre as principais técnicas de reconhecimento de padrões aplicadas envolvendo LEs, destacam-se os classificadores baseados na teoria de Bayes, os classificadores lineares e os não lineares e o “*template matching*”. Os trabalhos de Peris e Escuder-Gilabert (2013) , Zielinski et al. (2014) e Jain, Duin e Mao (2000) apresentam uma revisão apontando vantagens e desvantagens de algumas das principais técnicas de classificação empregadas nessa área. Portanto,

a principal colaboração desta com o estudo de LEs, é a metodologia proposta baseada na aplicação da técnica de Curvas Principais, haja vista que na revisão bibliográfica realizada não foi encontrado estudos que usam essa técnica.

Muitos trabalhos encontrados na literatura empregam *Principal Component Analysis* (PCA) em suas análises. Entretanto, essa não é a ferramenta mais adequada para ser usada em problemas de classificação de objetos, pois usa estatística de segunda ordem realizando uma transformação linear e ortogonal nos dados, em que no novo espaço gerado pelo PCA os dados são descorrelacionados. Em algumas situações essa descorrelação não é suficiente para agrupar as classes adequadamente. Além de depender da análise do projetista e interpretação dos agrupamentos formados em gráficos 2D.

As Curvas Principais consistem numa generalização não linear do PCA e independe de análises qualitativas. As mesmas são como curvas suaves, unidimensionais, que passam no “meio” de um conjunto multidimensional de dados, fornecendo um bom resumo unidimensional das mesmas (SHLENS, 2005). Dentre os algoritmos propostos para a extração dessas curvas, destaca-se o k-segmentos não suave que, devido a sua robustez na estimação das curvas, apresenta menor susceptibilidade a mínimos locais e convergência garantida (VERBEEK; VLASSIS; KROSE, 2002). A motivação para o uso dessa técnica está na sua boa capacidade de extrair modelos compactos dos dados e baixa complexidade computacional na fase de operação (FERNANDEZ et al., 2005).

Para comparar o modelo classificador com Curvas Principais, outros modelos foram projetados usando diferentes tipos de Redes Neurais Artificiais, como Perceptron Multi-Camadas (MLP), Função de Base Radial (RBF) e Mapas Auto Organizáveis (SOM-1D). As Redes Neurais Artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência.

A LE usada nessa pesquisa é baseada na espectroscopia de impedância, o que significa que esse sistema mede as propriedades elétricas das amostras em função da frequência. Observando os trabalhos disponíveis na literatura nessa área, nota-se que cada autor utiliza um segmento de frequência diferente na coleta dos dados. O trabalho feito por Breijo et al. (2013a) utiliza a faixa de 1 a 2×10^5 Hertz (Hz) e a pesquisa realizada por Ghosh, Das e RoyChaudhuri (2016) usa de 30 a 10^5 Hz. Já os trabalhos elaborados por Wei e Wang (2014), Lu et al. (2016) e

Karkra et al. (2017) aplicam suas metodologias usando apenas os seguintes segmentos: 1, 100, 1000, 10^4 e 10^5 Hz.

No entanto, não há um direcionamento ou estudo que aponte a frequência ou conjunto de frequência com maior capacidade de discriminação para o estudo. Em consequência disso, outra contribuição desse estudo é investigar a caracterização das soluções e reproduzir as análises considerando todas as frequências e depois os intervalos definidos separadamente, de forma a dizer qual é a mais indicada para o problema. Em seguida, os objetos determinados para atingir tais contribuições são apresentados.

1.1 Objetivos

O objetivo geral dessa pesquisa é investigar os resultados de impedância de um sistema de LE usando técnicas de reconhecimento de padrões. Pretende-se criar um modelo que seja capaz de identificar automaticamente qual é a substância e sua concentração, através da técnica de Curvas Principais.

Com o intuito de alcançar esse objetivo geral, foram determinados os seguintes objetivos específicos:

1. Projetar um modelo capaz de identificar automaticamente as amostras compostas por Glutamato monossódico, Inosinato dissódico e Guanilato dissódico, e suas diferentes concentrações.
2. Por meio do modelo projetado, determinar quais parâmetros propiciam a melhor distinção de classificação.
3. Analisar as arquiteturas dos eletrodos que compõem a LE individualmente, com intuito de identificar quais terão melhores precisões nas classificações. Ao identificá-los, verificar se a combinação dos mesmos gera melhores resultados.
4. Determinar em qual faixa de frequência o problema é mais discriminante.
5. Projetar o mesmo modelo usando técnicas diferentes e compará-los conforme o desempenho de classificação.

1.2 Estrutura do Trabalho

De acordo com o desenvolvimento da pesquisa e a execução dos objetivos estabelecidos, os resultados alcançados em cada etapa estão registrados no decorrer deste documento que está organizado em cinco capítulos.

O primeiro capítulo traz uma breve introdução ao assunto, apresentando a descrição de forma sucinta do uso de LEs. Além disso, esse capítulo apresenta os objetivos almejados nesse estudo.

O Capítulo 2 utiliza os dados extraídos de artigos para responder de forma coesa as questões de pesquisa e ainda apresenta as aplicações encontradas na literatura acerca do tema discutido. Esse capítulo tem o intuito de apresentar e fundamentar o leitor sobre as técnicas que serão empregadas nessa pesquisa.

No Capítulo 3, é onde contém a apresentação das características do experimento e do banco de dados obtido. Além disto, apresenta a metodologia desenvolvida para alcançar os objetivos definidos. Em sequência, o Capítulo 4 traz os resultados obtidos e as análises pertinentes.

As conclusões gerais deste trabalho são apresentadas no Capítulo 5. E por fim, o apêndice traz imagens referentes aos melhores resultados obtidos de acordo com as análises realizadas.

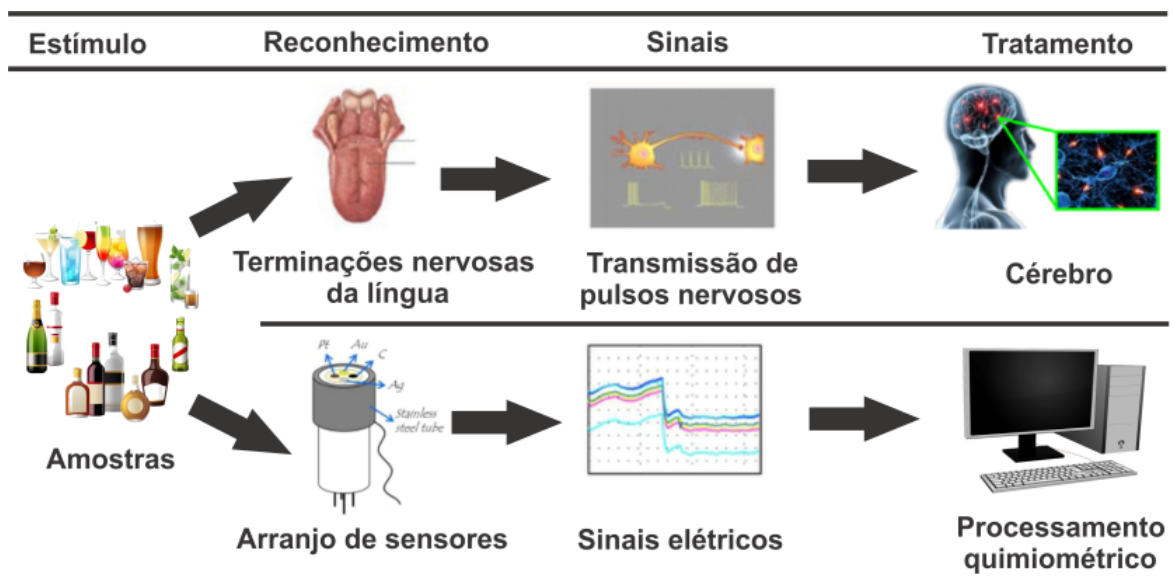
2 REFERENCIAL TEÓRICO

Com intuito de fundamentar o leitor, esse capítulo traz seções com alguns conceitos sobre língua eletrônica e reconhecimento de padrões. Também apresenta definições à respeito das técnicas que serão empregadas, como diferentes tipos de redes neurais e, em destaque Curvas Principais. Além disso, mostra exemplos de trabalhos que têm sido realizados acerca do tema abordado.

2.1 Língua Eletrônica

O conceito de LE teve sua origem e é baseado nos mecanismos do sistema gustativo do ser humano. A Figura 2.1 mostra a relação entre esses sistemas. A língua humana é o órgão responsável por fornecer informações sobre o gosto, textura e temperatura dos alimentos. Através do conceito de seletividade global, a língua é capaz de identificar sabores básicos, como: azedo, doce, salgado, amargo e umami (saboroso). Esta percepção é possível devido os receptores gustativos presentes na língua, que por meio de estímulos, os receptores geram sinais elétricos que são transmitidos ao cérebro e assim, interpretados pelo mesmo (DEISINGH; THOMPSON, 2004).

Figura 2.1 – Comparação entre o sistema biológico e o sistema eletrônico.



Fonte: Adaptado pelo autor.

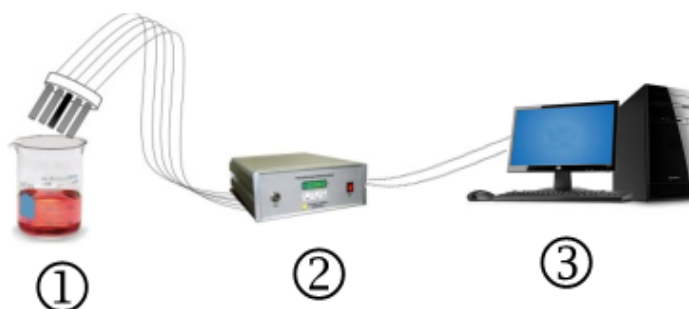
Assim, inspirado nesse processo, a LE foi desenvolvida. Esse sistema é considerado um instrumento analítico baseado em uma série de sensores químicos de baixa seletividade e com o auxílio de processamentos matemáticos são capazes de caracterizar amostras líquidas

complexas. Se for configurado, treinado e/ou calibrado corretamente, a LE é capaz de reconhecer a composição qualitativa e quantitativa de soluções de diferentes naturezas (WEI; WANG; ZHANG, 2013; PERIS; ESCUDER-GILABERT, 2013).

Segundo a definição da *International Union of Pure and Applied Chemistry* (IUPAC), uma LE é “um sistema multissensorial, que consiste de um arranjo de sensores pouco seletivos e utiliza procedimentos matemáticos avançados para processamento de sinais baseados em reconhecimento de padrões e/ou análise multivariada de dados” (VLASOV et al., 2005).

Devido à sua alta sensibilidade, versatilidade, simplicidade e robustez, o sistema de LE já provou ser valioso em diversas aplicações (WEI; WANG; ZHANG, 2013). De modo geral, esse sistema é composto por três elementos: (1) uma matriz de sensores químicos com diversos seletivos; (2) um amplificador ou gerador de sinal; e (3) um software com o algoritmo apropriado para processar os sinais e analisar os resultados, conforme a Figura 2.2 (PERIS; ESCUDER-GILABERT, 2013).

Figura 2.2 – Elementos de uma LE.



Fonte: Adaptado pelo autor.

A LE desempenha um papel importante na determinação rápida de substâncias químicas e suas propriedades sensoriais. Nos últimos anos, as LEs desenvolvidas podem ser classificadas em (WEI; WANG, 2014; LIU et al., 2013):

- LE potenciométrica, que é composta de vários tipos de membranas lipídicas/poliméricas, limitando seu uso à detecção de espécies iônicas carregadas. Esse sistema é amplamente usado para monitorar o tempo de armazenamento de alimentos e bebidas (WEI; WANG; ZHANG, 2013).
- LE voltamétrica, que deriva a informação do componente químico a partir da medição da corrente versus o potencial aplicado sob condições que ajudam a polarização do eletrodo de trabalho. Esses eletrodos de trabalho são interessantes devido à sua sensibilidade

cruzada, onde cada eletrodo apresenta informações diferentes em relação às amostras (BREIJO et al., 2013b; WEI; WANG; ZHANG, 2013).

- LE baseada na espectroscopia de impedância, aqui usada, permite a análise das propriedades de materiais e sistemas, aplicando-lhes sinais elétricos alternados de diferentes frequências e medindo os sinais elétricos de saída. Esse sistema tem sido amplamente aplicado na análise de alimentos e medicinal (BREIJO et al., 2013a).

A principal vantagem do sistema de LE em comparação aos métodos tradicionais é que eles permitem a implementação de sistemas de medição rápida e de baixo custo, evitando a necessidade de interpretação humana para realizar as análises. Esse sistema requer pequeno volume de amostra e pouco espaço, devido a robustez de seus sensores. Além disso, a LE proporciona maior objetividade em comparação ao painel sensorial humano (BREIJO et al., 2013b; SAHA et al., 2014).

A seção a seguir apresenta uma revisão da literatura a respeito das aplicações que têm sido realizadas utilizando sistemas de LEs, com destaque para a contextualização e o estado da arte.

2.2 Estado da Arte

Devido á grande quantidade de aplicações possíveis acerca deste tema, podem ser encontrados diversos estudos realizados na literatura. O estudo apresentado por Cetó et al. (2015) que relata a aplicação de uma LE como ferramenta para discriminação de diferentes origens de vinhos. O autor usou *Fast Fourier Transform* (FFT) para pré-processar os dados e *Linear Discriminant Analysis* (LDA) e Redes Neurais Artificiais (RNA) para análises qualitativa e quantitativa. O desempenho geral alcançado foi de 92,9% na identificação correta para a aplicação qualitativa e um coeficiente de correlação de 0,830 para o quantitativo.

Nery e Kubota (2016) mostram as etapas da preparação de uma LE que foi usada para analisar 34 amostras de cerveja e 11 de vinho. O sistema proposto foi capaz de discriminar cervejas de diferentes marcas e tipos utilizando *Principal Components Regression* (PCR) e *Principal Component Analysis* (PCA). Além disso, também foi possível a identificação da presença de estabilizadores e antioxidantes por meio de *K Nearest Neighbors* (KNN) com apenas 3 erros de classificação. Além disso, o pH de 24 das 34 amostras testadas foi previsto com erro menor que 3% e o teor alcoólico com o erro de previsão menor que 5% em 18 das 34 amostras. A

análise das amostras de vinho foi discriminada de acordo com a variedade da uva, utilizando PCA, *Partial Least Squares* (PLS) e KNN.

Ainda relacionado a vinhos, Ouyang, Zhao e Chen (2013) trazem uma pesquisa com objetivo de classificar amostras de vinho de acordo com seu tempo de envelhecimento (3, 5, 8 e 10 anos). Para isso, foram desenvolvidos modelos de classificação utilizando PLS, KNN, RNA e *Support Vector Machine* (SVM). Os resultados dos modelos na identificação das amostras foram 80%, 87%, 100% e 95%, respectivamente.

Há diversos estudos relacionados a detecção e quantificação de substâncias em diversos tipos de bebidas. Note que os trabalhos apresentados anteriormente foram todos utilizando amostras de vinho, porém cada pesquisa desenvolve aplicações e análises diferentes usando diversas técnicas, o que mostra a importância e a gama de aplicações dessa área.

A pesquisa de Wei e Wang (2014) aplica um sistema de LE como instrumento para classificar e prever amostras de mel de diferentes origens florais e geográficas. As origens florais das amostras analisadas foram: acácia, trigo sarraceno e *motherwort* e as amostras referentes ao mel de acácia foram selecionadas de diferentes origens geográficas. Para classificar as amostras foram usadas as técnicas PCA e *Discriminant Function Analysis* (DFA) e as previsões foram realizadas com PCR, PLS e SVM. Os modelos implementados com PCR e PLS atingiram resultados com coeficientes de correlação iguais a 0,96. O melhor resultado encontrado foi com SVM, o qual teve coeficiente igual a 0,99.

O trabalho realizado por Ghosh, Das e RoyChaudhuri (2016) propõe um sistema de LE para detectar várias toxinas alimentares em misturas complexas, utilizando a espectroscopia de impedância. Os parâmetros extraídos foram processados usando PCA, FFT e PLS. Os resultados demonstraram que é possível quantificar essas substâncias com até 0,1 fg/ml nas amostras. Dando continuidade à sua pesquisa, Ghosh, Das e RoyChaudhuri (2017) apresentaram uma melhora na precisão da detecção de toxinas múltiplas, com um erro entre 10% a 20%, o que segundo o autor é um feito notável no domínio de detecção de toxinas em alimentos.

O artigo de Lu et al. (2016) aborda uma análise atributiva para a caracterização e quantificação das particularidades do sabor de amostras de arroz, como a suavidade, a viscosidade, a doçura e o aroma, empregando um sistema de LE. As informações obtidas foram pré-processadas com PCA e FFT, as quais foram usadas como variáveis de entrada para um modelo projetado usando RNA. Os coeficientes de correlação para todos os atributos foram superiores a 0,9 demonstrando uma boa capacidade preditiva do modelo de análise desenvolvido.

O estudo feito por Liu et al. (2014) traz a análise de dados de uma LE com intuito de prever a intensidade da amargura de um determinado medicamento. O modelo de droga utilizado para estabelecer o modelo de predição de amarguras foi o cloridrato de berberina, por meio dos dados fornecidos pela LE e com auxílio de um algoritmo genético combinado com RNA. Para comparar com esse modelo, outros foram projetados usando *Multiple Linear Regression* (MLR) e PLS. Os índices de validação cruzada foram de 99,59%. Percentual este superior aos métodos comparados, que tiveram 89,57% e 95,21% de precisão.

Os artigos (BREIJO et al., 2013a; BREIJO et al., 2013b) empregaram sistemas de LE que são capazes de obter medidas de voltametria ou espectroscopia de impedância com diferentes configurações dos eletrodos. Esses sistemas foram aplicados na detecção e quantificação de Trinitrotolueno (TNT) com diferentes níveis de concentração. O primeiro artigo utiliza RNA e *Self-Organizing Map* (SOM) para classificar a concentração de TNT e o segundo usa PCA e PLS em seus modelos de previsão. Os resultados obtidos expressam que as LEs foram capazes de detectar essa substância em suas respectivas amostras.

Outra pesquisa que utiliza LE na determinação de compostos explosivos é relatada por González-Calabuig, Cetó e Valle (2016), onde as informações de partida foram os dados voltamétricos obtidos. Estes dados foram, primeiramente, analisados com PCA para discriminar os diferentes compostos individuais e depois, RNA para a quantificação de algumas misturas. O índice de correlação entre as concentrações foi de 0,929 e o erro médio quadrático para o conjunto de teste foi de 0,108.

De acordo com os artigos que fazem parte da revisão bibliográfica dessa dissertação, as técnicas mais usuais envolvendo sistemas de LEs são: PCA, PLS e RNA. Há inúmeros estudos que relatam diversas metodologias e aplicações que utilizam esses métodos (BERRÀS et al., 2015; PERIS; ESCUDER-GILABERT, 2013). Por isso, a importância deste trabalho é referente a utilização de um método sofisticado, com boa capacidade de representação de dados não lineares, para discriminação de dados, com a finalidade de avaliar e analisar o desempenho dos mesmos.

2.3 Reconhecimento de padrões

O Reconhecimento de Padrões (RP) tem uma longa história, mas antes da década de 1960 era destinado principalmente a produção de pesquisa teórica na área de estatística. Em consequência da necessidade de manuseio e recuperação de informações, o RP foi impulsio-

nado nas pesquisas de engenharia de hoje. Sendo parte integrante da maioria dos sistemas de inteligência de máquina criados para a tomada de decisões (THEODORIDIS; KOUTROUMBAS, 2009).

O RP é o estudo de como as máquinas podem observar o ambiente, aprender a distinguir padrões de interesse de acordo com seu histórico e tomar decisões sensatas e razoáveis sobre as categorias dos padrões (JAIN; DUIN; MAO, 2000). Existem diversas aplicações relacionadas a essa área, como o reconhecimento de caracteres (letras ou números), com grandes implicações na automação e no tratamento de informações. Outra área de aplicação é a mineração de dados e a descoberta de conhecimento em bancos de dados, usado na medicina, biologia, análise de mercado financeiro e recuperação de imagens. Também é aplicado na identificação pessoal com base em vários atributos físicos, tais como o rosto e impressões digitais (biometria) (THEODORIDIS; KOUTROUMBAS, 2009).

O escopo deste trabalho foca em problemas de classificação, e diante do exposto, é necessária uma arquitetura elaborada para cumprir tal tarefa. Em geral, inicia-se com a aquisição das informações, nesse caso o sistema de LE, que transforma os fenômenos físicos em sinais elétricos e por intermédio de conversores analógicos ou digitais disponibilizam uma informação numérica.

Após a coleta desses sinais, recomenda-se a etapa de pré-processamento. Nessa etapa é onde ocorre a adequação dos sinais, através de técnicas como a normalização dos sinais, a filtragem para remoção de ruído, entre outras. Essa transformação dos dados também é necessária para dar a cada resposta variável o mesmo peso e a mesma importância prévia na análise, evitando, portanto, polarização dos dados (ZIELINSKI et al., 2014; DUDA; HART; STORK, 2001). Normalmente, os dados obtidos por um sistema de LE exigem algum tipo de pré-processamento nos dados antes de realizar a análise estatística ou aplicar algum método de classificação.

Seguindo, a etapa de classificação atribui aos eventos uma categoria. Se esta associação é pautada em um conjunto de regras lógicas e claras, o classificador tem como base métodos sintáticos. Caso contrário, a associação possui como base uma natureza probabilística e os métodos empregados são os estatísticos. Neste último caso, uma atribuição é realizada através da probabilidade de pertinência a cada possível categoria (BEZERRA, 2010).

Por fim, uma etapa de pós-processamento realiza o trabalho de adequação das saídas do classificador através da escolha de uma ação, tendo como base não apenas os valores

provenientes do classificador, mas podendo levar em conta os custos de uma decisão errada, o contexto em que se dá a classificação, entre outros parâmetros (BEZERRA, 2010). A seguir, as técnicas que serão empregadas para realizar a classificação do conjunto de dados.

2.4 Curvas Principais

As Curvas Principais (CPs) foram apresentadas pela primeira vez por Hastie e Stuetzle (1989) como uma ferramenta capaz de fornecer uma descrição não linear de um conjunto de dados. Essa técnica tem sido aplicada em diversos tipos de dados com bastante sucesso na classificação de padrões, devido sua capacidade de extrair modelos compactos dos dados.

As CPs são curvas unidimensionais que passam no “meio” de um conjunto de dados no espaço multidimensional, fornecendo uma descrição não linear dos dados, ou seja, são curvas não paramétricas e a sua forma é sugerida pelo conjunto de dados (HASTIE; STUETZLE, 1989). Matematicamente, essas curvas são definidas a partir do conceito de auto consistência. Isso significa que os pontos que compõem a CP são as médias dos dados que nela se projetam. Para entender este conceito é necessário definir o índice de projeção de um ponto x_i em uma curva \mathbf{f} .

Uma curva unidimensional no espaço d dimensional é um vetor $\mathbf{f}(t)$ de d funções contínuas e uma única variável t , ou seja, $\mathbf{f}(t) = [f_1(t), f_2(t), \dots, f_d(t)]^T$. Essas funções são denominadas funções de coordenadas e o parâmetro t está relacionado ao ordenamento ao longo da curva. Seja \mathbf{f} uma curva suave no intervalo fechado $I \subseteq \mathfrak{R}^1$ que não intercepta a si própria, isto é, $t_1 \neq t_2 \rightarrow f(t_1) \neq f(t_2)$ e com seu comprimento finito dentro de uma esfera de dimensões finitas em \mathfrak{R}^d .

Tomando \mathbf{x} como um vetor aleatório em \mathfrak{R}^d , com densidade de probabilidade h , e momentos de segunda ordem finitos, o índice de projeção $t_f : \mathfrak{R}^d \rightarrow \mathfrak{R}^1$ é definido como:

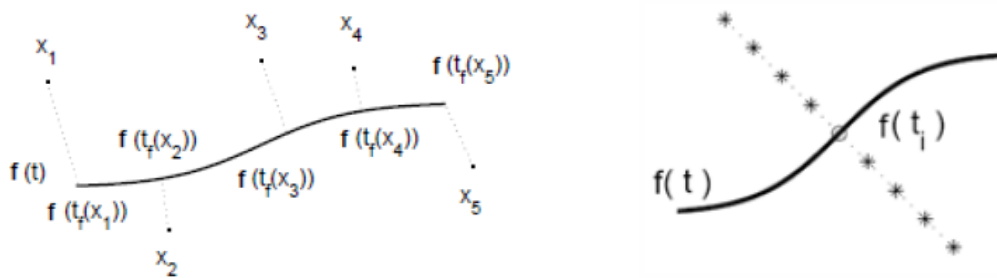
$$t_f(x) = \sup_t \{t : \|x - \mathbf{f}(t)\| = \inf_{\mu} \|x - \mathbf{f}(\mu)\|\}, \quad (2.1)$$

onde μ é uma variável auxiliar definida em \mathfrak{R}^1 . O índice de projeção $t_f(x)$ é o valor de t para o qual a curva principal $\mathbf{f}(t)$ está mais próxima de x . Caso haja mais de um valor possível, o maior deles é selecionado. A Figura 2.3(a) mostra o índice de projeção relativo a uma curva principal. A Figura 2.3(b) mostra um ponto auto consistente (t_i) de uma curva para uma distribuição de

dados em duas dimensões, representada parcialmente. Assim, de acordo com a definição de Hastie e Stuetzle (1989) a curva principal é auto consistente conforme a equação:

$$\mathbf{f}(t) = E[x | t_f(x) = t], \forall t. \quad (2.2)$$

Figura 2.3 – Dados de uma Curva Principal.



(a) Projeção dos dados na curva.

(b) Ponto auto consistente de uma curva.

Fonte: Adaptados de Kégl et al. (2000), Hastie e Stuetzle (1989).

2.4.1 K-segmentos não suave

Diversos algoritmos foram desenvolvidos para extrair as CPs. Em Hastie e Stuetzle (1989), as curvas foram extraídas baseando-se no conceito de auto consistência, entretanto o algoritmo posposto apresentou problemas como: tendências em trechos de curvatura, convergência e a existência de uma curva principal que não ocorria para qualquer distribuição.

Estes fatores deram margem para novos trabalhos, como a proposta de correção da tendência de estimação em Banfield e Raftery (1992). Duchamp e Stuetzle (1996) determinaram as curvas através de um sistema de equações diferenciais. Kégl et al. (2000) propuseram um algoritmo baseado em linhas poligonais. Delicado e Huerta (2002) relatam um algoritmo baseado na definição em que se enfatizam a definição de curvas principais a partir de pontos orientados auto consistentes.

Neste estudo, o algoritmo utilizado foi apresentado por Verbeek, Vlassis e Krose (2002), conhecido como k-segmentos não suave (k-seg), que além de ser robusto e ter convergência garantida, é menos susceptível a mínimos locais. Esse algoritmo, propõe a construção passo a passo da curva principal criando-se primeiramente um único segmento e assim, o número de

segmentos é aumentado progressivamente de acordo com os parâmetros do algoritmo. Todos os segmentos são interligados por segmentos independentes da curva principal.

Em seu trabalho, Verbeek, Vlassis e Krose (2002) utilizam um modelo probabilístico para encontrar as CPs baseado no trabalho de Tibshirani (1992), através da maximização da verossimilhança logarítmica, onde os dados são modelados por:

$$p(\mathbf{x}) = \int_0^1 p(\mathbf{x} | t) p(t) dt, \quad (2.3)$$

em que t é a variável latente distribuída no comprimento do arco da curva parametrizada de comprimento l , $p(\mathbf{x} | t)$ é um modelo gaussiano esférico modelando o ruído localizado no ponto t da curva com variância sobre todo t . A variância é um parâmetro de suavização, que pode ser configurado pelo usuário.

O algoritmo inicia-se na inserção do primeiro segmento s , sendo que todos os dados do conjunto são levados em conta, tendo assim um conjunto de dados em que o centro corresponde ao valor médio dos dados. O primeiro segmento s tem a direção da primeira componente principal dos dados, com comprimento equivalente a $3\sigma/2$. A linha s e sua distância ao ponto x são definidos pelas respectivas equações:

$$\mathbf{s} = \{\mathbf{s}(t) = \mathbf{c} + \mathbf{u}t \mid t \in \mathbb{R}\}, \quad (2.4)$$

$$d(x, s) = \min_{t \in \mathbb{R}} \|s(t) - x\|. \quad (2.5)$$

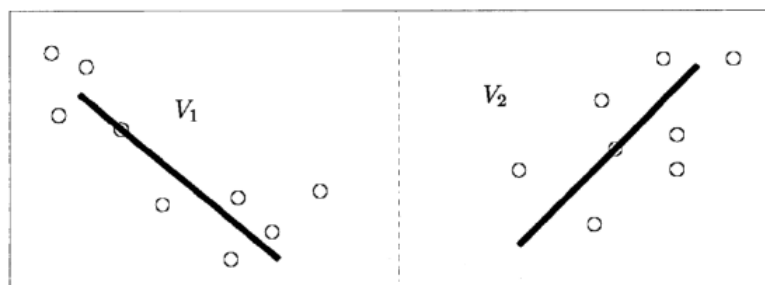
Seja \mathbf{X}_n um conjunto de dados em \mathfrak{R}^d . As regiões de Voronoi V_1, \dots, V_k é definido como (VERBEEK; VLASSIS; KROSE, 2002):

$$V_i = \{\mathbf{x} \in X_n \mid i = \arg \min_j d(\mathbf{x}, s_j)\}, \quad (2.6)$$

onde, V_i contém todos os pontos de maneira em que a i -ésima linha esteja mais próxima dos pontos, como ilustrado na Figura 2.4 que mostra as regiões de Voronoi V_1 e V_2 com seus respectivos segmentos. Dessa forma, o algoritmo busca minimizar a distância total quadrática de todos os pontos à linha k de acordo com:

$$\sum_{i=1}^k \sum_{x \in V_i} D(\mathbf{x}, s_i)^2. \quad (2.7)$$

Figura 2.4 – Regiões de Voroni.



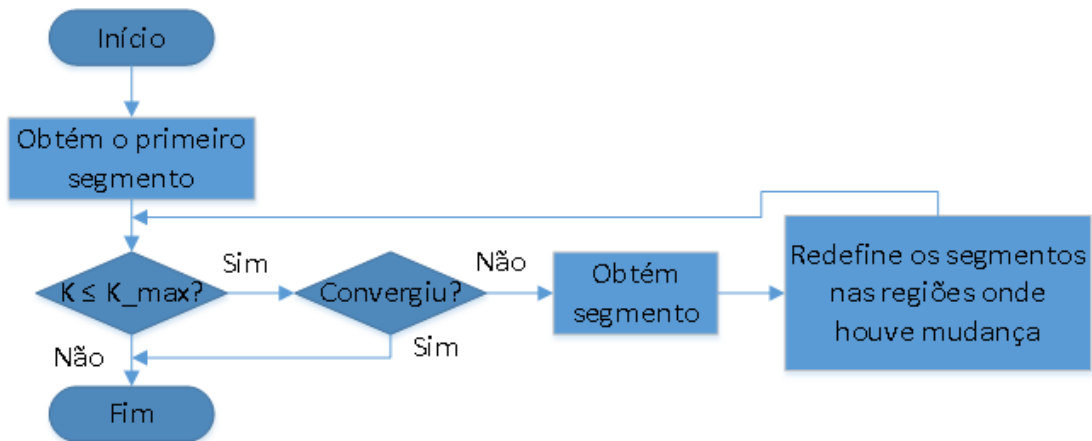
Fonte: Adaptado de Verbeek, Vlassis e Krose (2002).

Verbeek, Vlassis e Krose (2002) notaram que o algoritmo apresentava melhores resultados removendo a restrição de que os segmentos devem incluir todas as projeções dos pontos da região de Voroni na primeira componente principal. Desse modo, foi encontrado empiricamente, que o algoritmo apresenta melhor desempenho se forem usados segmentos da primeira componente principal, limitados ao comprimento de $3\sigma/2$ no centro da região de Voroni. Para garantir a convergência, verifica-se para cada região de Voroni se a Equação 2.7 decresce. Caso a Equação 2.7 não decresça, usa-se o segmento que inclui todas as projeções dos pontos na primeira componente principal ao invés de usar um segmento de comprimento $3\sigma/2$, o que garanti o decrescimento da Equação 2.7. Em qualquer uma dessas situações, garante-se que todo segmento que compõe a curva principal atravesse os dados na direção da primeira componente principal de cada região de Voroni, e a união destes segmentos resulta na curva principal poligonal que melhor se ajusta aos dados.

Para a inserção do segundo segmento é realizado um teste com todo o conjunto de dados, cujo objetivo de encontrar qual é o melhor ponto a ser definido como centro do agrupamento correspondente a este segundo segmento. Dessa forma, para definir quais dados pertencem ao novo agrupamento é usado o algoritmo *k-means* (DUDA; HART; STORK, 2001; THEODORIDIS; KOUTROUMBAS, 2009). Após a inserção do segundo segmento é realizado um cálculo das distâncias dos dados aos segmentos mais próximos, caso um evento anterior pertencia ao primeiro segmento, é repetido o processo de agrupamento.

Após este passo, estes dois segmentos são unidos por uma linha. A inserção de novos segmentos segue os passos acima até que o algoritmo convirja. Existem dois requisitos para que ocorra essa convergência: (i) quando o número máximo de seguimentos é atingido ($k = k_{max}$) e (ii) quando a região de Voroni tem menos de três eventos. O requisito que ocorrer primeiro leva o algoritmo à convergência. A Figura 2.5 mostra o fluxograma do processo de extração das CPs com o algoritmo k-seg.

Figura 2.5 – Fluxograma do algoritmo k-segmentos não suave.

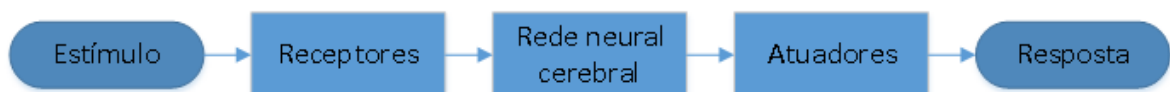


Fonte: Adaptado de Verbeek, Vlassis e Krose (2002).

2.5 Redes Neurais Artificiais

O sistema nervoso humano pode ser visto como um sistema de três estágios, ilustrados na Figura 2.6, onde se inicia com os receptores convertendo o estímulo vindo do corpo humano ou do ambiente externo em impulsos elétricos que conduzem a informação para a rede neural, o cérebro. O cérebro recebe informações continuamente, as interpreta e toma decisões apropriadas. Os atuadores convertem esses impulsos elétricos em respostas discerníveis como saídas do sistema (HAYKIN, 1999).

Figura 2.6 – Representação do sistema nervoso em diagrama de blocos.



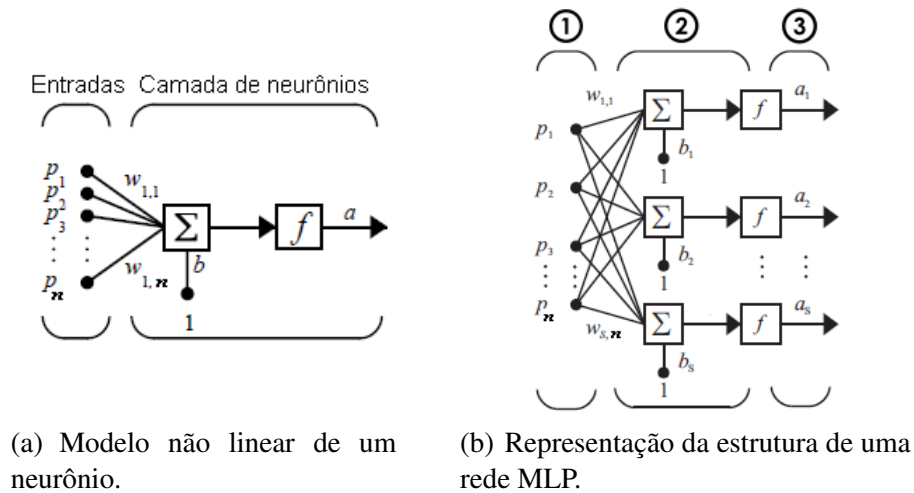
Fonte: Adaptado de Haykin (1999).

Conforme as “redes neurais naturais”, as Redes Neurais Artificiais (RNA) também consistem da interconexão de unidades de processamento, os neurônios. As conexões entre os neurônios são chamadas sinapses ou pesos sinápticos. A Figura 2.7(a) representa o modelo básico de um neurônio utilizado no projeto de RNA (HAYKIN, 1999).

Observe que o modelo de neurônio apresentado é composto por sinais de entrada (\mathbf{p}_n) ponderadas pelos pesos sinápticos (\mathbf{W}_n) somados a uma polarização externa, o *bias* (b), o qual tem o efeito de aumentar ou diminuir o argumento da função de ativação. Tal função ($f(\cdot)$), serve para limitar a amplitude da saída do neurônio (a). Em termos matemáticos, um neurônio k pode ser descrito pela equação:

$$a = f(\sum W_n p_n + b). \quad (2.8)$$

Figura 2.7 – Características de uma RNA.



Fonte: Adaptado de Beale, Hagan e Demuth (2012).

A maneira em que os neurônios da rede são estruturados, está intimamente relacionada ao algoritmo de aprendizagem usado para treinar a rede. Uma RNA é organizada em forma de camadas, onde a forma mais simples é composta por uma camada de entrada, que se projeta sobre uma camada de saída de neurônios, conectando a rede ao ambiente externo. Essa rede é chamada de rede de camada única (*Single Layer Perceptron*).

Outro tipo de arquitetura é a rede de múltiplas camadas (MLP - *Multiple Layer Perceptron*), onde a diferença é que essa rede possui uma ou mais camadas ocultas. A função dos neurônios dessa camada é intervir entre a entrada e a saída da rede, tornando a rede capaz de extrair estatísticas de ordem elevada (HAYKIN, 1999). A Figura 2.7(b) mostra a estrutura de uma rede MLP com uma camada oculta.

A propriedade mais significativa de uma RNA é certamente a habilidade de aprender a partir do seu ambiente e melhorar seu desempenho através desse aprendizado, de acordo com critérios específicos adotados para atingir tal propósito. Em outras palavras, através da informação de uma resposta desejada, a rede tenta aproximar um sinal alvo durante seu processo de aprendizagem. Essa aproximação é obtida através do ajuste, de forma sistemática, de um conjunto de parâmetros livres, característico de cada rede neural. Existem vários algoritmos de aprendizado que são utilizados no projeto de RNA, cada um deles possuindo características e vantagens específicas (HAYKIN, 1999).

Em resumo, existem dois paradigmas de aprendizado: o aprendizado supervisionado e o não-supervisionado. O aprendizado supervisionado é o aprendizado obtido por meio de um tutor, o qual detém o conhecimento do ambiente. No processo de aprendizado, a RNA é exposta a um vetor de treino extraído do ambiente e o mesmo provê à rede uma resposta desejada para este específico vetor de treino. Assim, a resposta desejada é o resultado ótimo que a rede deveria apresentar para aquele determinado vetor do conjunto de treino (HAYKIN, 1999).

O aprendizado não-supervisionado não há um tutor externo para supervisionar o processo de aprendizado. Durante o treinamento, somente os padrões de entrada são apresentados à RNA até que a rede se torne “sintonizada” às regularidades estatísticas dos dados de entrada. A partir desta condição, a rede desenvolve a habilidade de formar representações internas para codificar características da entrada e, por meio disto, agrupar os padrões de entrada em grupos com características similares ou criar novos grupos automaticamente.

O processo de ajuste dos pesos da rede é chamado de treinamento e tem o intuito de otimizar o desempenho da rede. Basicamente, há dois estilos diferentes de treinamento: o incremental e o em lote (*batching*). No treinamento incremental, os pesos e o *bias* da rede são atualizados cada vez que uma entrada é apresentada à rede. No treinamento em lote, os pesos e o *bias* só são atualizados depois que todas as entradas são apresentadas. Para a maioria dos problemas, o treinamento em lote é significativamente mais rápido e produz erros menores do que o treinamento incremental (BEALE; HAGAN; DEMUTH, 2012).

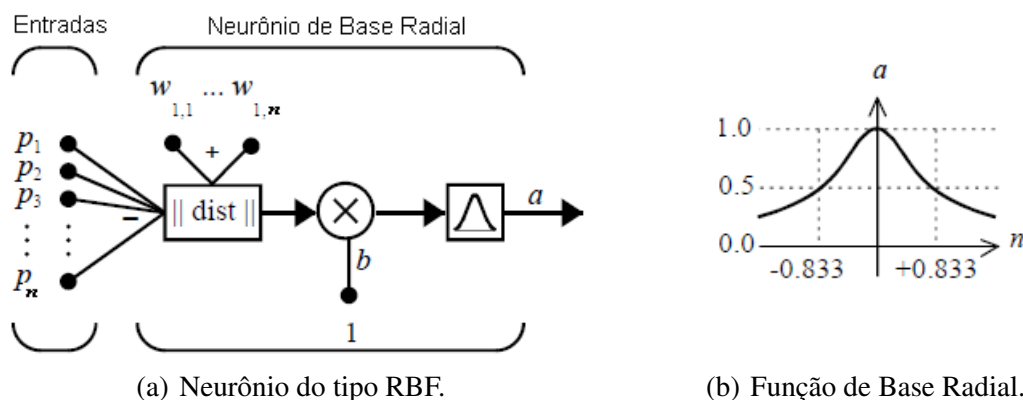
Geralmente, os algoritmos de treinamento baseados na otimização de Levenberg-Marquardt e o que utiliza semi-Newton são os mais rápidos. Ambos os métodos tendem a ser menos eficazes para redes com milhares de pesos, uma vez que exigem mais memória e mais tempo de computação para estes casos. Ao treinar grandes redes, os algoritmos baseados em gradiente conjugado dimensionado e *resilient backpropagation* são boas escolhas. Seus requisitos de memória são relativamente pequenos e eles ainda podem ser mais rápidos que os algoritmos de gradiente de descida padrão. As definições e procedimentos de cada um desses algoritmos podem ser encontrados em Beale, Hagan e Demuth (2012).

2.5.1 Função de Base Radial

As RNAs de Função de Base Radial (RBF - *Radial Basis Function*) são unidades ocultas que fornecem um conjunto de funções que constituem uma base arbitraria para os padrões de entrada, quando eles são expandidos sobre o espaço oculto. O projeto de uma rede neural RBF

pode ser visto como um problema de ajuste de curvas (ou, dito de outra forma, um problema de aproximação de funções) em um espaço de alta dimensionalidade. A Figura 2.8 representa o modelo de um neurônio utilizado no projeto de uma rede RBF (HAYKIN, 1999).

Figura 2.8 – Modelo de um neurônio do tipo RBF.



Fonte: Adaptado de (BEALE; HAGAN; DEMUTH, 2012).

A construção de uma RBF, usualmente, também envolve três camadas: (1) a camada de entrada, que é constituída por sinais da fonte (unidades sensoriais) que conectam a rede ao seu ambiente; (2) a camada oculta da rede, onde aplica-se uma transformação não linear do espaço de entrada para o espaço oculto, na maioria das aplicações o espaço oculto de alta dimensionalidade; e (3) a camada de saída, que fornece a resposta da rede ao padrão de ativação aplicado à camada de entrada.

Uma RNA do tipo RBF possui em sua arquitetura uma camada escondida definida por um conjunto de funções de base radial, das quais a rede deriva seu nome. O aprendizado de uma rede RBF é equivalente a ajustar uma superfície não linear ao conjunto de dados, em um espaço multi dimensional, de acordo com algum critério estatístico. O processo de treinamento equivale a usar esta superfície multi dimensional para interpolar outros pontos que não pertençam ao conjunto de treino, mas estejam em sua vizinhança (HAYKIN, 1996).

Quando uma RBF é usada para executar uma tarefa complexa de classificação de padrão, o problema é basicamente resolvido transformando-o em um espaço de alta densidade de maneira não linear. De acordo com Haykin (1999), um problema de classificação de padrões lançado em um espaço de alta dimensão é mais provável de ser linearmente separável do que em um espaço de baixa dimensão.

Considerando uma família de superfícies em que cada uma divide naturalmente um espaço de entrada em duas regiões. Seja ε um conjunto de N eventos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, onde

cada um deles é atribuído a uma das duas classes: ε_1 e ε_2 . Esta dicotomia dos pontos é dito ser separável caso exista uma superfície na família que separa os pontos na classe ε_1 daqueles da classe ε_2 . Para cada padrão $\mathbf{x} \in \mathcal{E}$, define-se um vetor composto de um conjunto de funções reais como $\varphi(\mathbf{x}) = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots, \varphi_{m_1}(\mathbf{x})]^T$ (HAYKIN, 1999).

Suponha que o evento \mathbf{x} seja um vetor em um espaço de entrada m_0 -dimensional. O vetor $\varphi(\mathbf{x})$ então mapeia pontos nesse espaço em pontos correspondentes em um novo espaço de dimensão m_1 . Uma dicotomia $\{\varepsilon_1, \varepsilon_2\}$ de \mathcal{E} é dita ser separável se existe um vetor \mathbf{w} m_1 -dimensional que possa ser escrito como:

$$\begin{aligned} \mathbf{w}^T \varphi(\mathbf{x}) &> 0, & x \in \varepsilon_1, \\ \mathbf{w}^T \varphi(\mathbf{x}) &< 0, & x \in \varepsilon_2. \end{aligned} \quad (2.9)$$

O hiperplano descreve a superfície de separação no espaço φ , ou seja, no espaço oculto. A imagem inversa deste hiperplano define a superfície de separação no espaço de entrada. A saída y_i da rede RBF é, portanto, a soma das saídas de cada função de base radial $\varphi(\cdot)$, ponderadas pelos respectivos pesos sinápticos w_i , de tal forma que a combinação linear seja expressa pela equação:

$$y_i = \sum_{i=1}^n w_i \varphi(\|u - u_i\|^2), \quad (2.10)$$

onde u_i são referidos como centros no contexto de redes neurais RBF e o operador $\|\cdot\|$ é usualmente a norma Euclidiana e mede o módulo do vetor argumento. Existe um conjunto de funções que são adequadas para empregar em φ , por resultarem em um conjunto de equações lineares para as incógnitas w_i para as quais existe uma única solução.

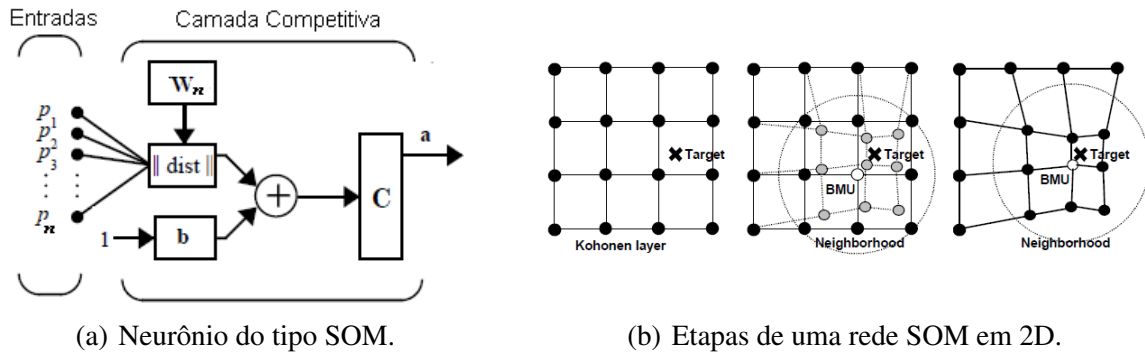
2.5.2 Mapas Auto Organizáveis

O modelo de rede de mapas auto organizáveis (SOM - *Self-Organization Maps*), também conhecido como rede de Kohonen, é um método de rede neural competitiva que aprende a classificar os vetores de entrada de acordo com a forma como eles são agrupados no espaço de entrada. As localizações espaciais (ou coordenadas) dos neurônios no espaço de saída são indicativos de características estatísticas inerentes contidos no espaço de entrada (BEALE; HAGAN; DEMUTH, 2012; ASAN; ERCAN, 2012).

Ao contrário das redes RBF, o objetivo principal da rede SOM é transformar um espaço complexo de alta dimensionalidade em um espaço de baixa dimensionalidade, mas com suas

relações preservadas. A rede aprende tanto a distribuição quanto a topologia dos vetores de entrada em que são treinados. A Figura 2.9(a) mostra o modelo de um neurônio do tipo SOM.

Figura 2.9 – Características de uma rede SOM.



Fonte: Adaptados de Beale, Hagan e Demuth (2012), Vesanto et al. (2000).

A formação dessa rede envolve três processos característicos (BEALE; HAGAN; DEMUTH, 2012; ASAN; ERCAN, 2012). A concorrência, onde os neurônios de saída competem uns com os outros para melhor representar a amostra de entrada particular, onde o neurônio específico com seus pesos de conexão mais semelhantes à amostra de entrada é declarado vencedor da competição, nominado como *Best-Matching Unit* (BMU).

A cooperação, pois como a rede é uma organização topográfica em que os locais próximos no espaço de saída representam entradas com propriedades semelhantes, o BMU determina a localização espacial de uma vizinhança de neurônios cooperantes. Assim, compartilhando recursos comuns e estimulando uns aos outros para aprender algo a partir da mesma entrada. E por último, a adaptação, que com o processo de aprendizagem, os vetores de peso do BMU e suas unidades vizinhas no mapa são ajustados e os neurônios relevantes tornam-se mais semelhantes à amostra de entrada. Desse modo, os neurônios que têm uma resposta forte a uma determinada peça de dados de entrada terão uma chance maior de responder a dados de entrada semelhantes no futuro.

Esses processos descritos podem ser melhores compreendidos por meio da Figura 2.9(b). Matematicamente, o primeiro passo é inicializar os vetores de pesos $\mathbf{w}_i = (\mathbf{w}_{i1}, \mathbf{w}_{i2}, \dots, \mathbf{w}_{im})^T$, onde $i = 1, 2, \dots, n$ denota o número de neurônios de saída na rede. Esses pesos são links que conectam os neurônios de entrada com os da saída e são atualizados através do processo de aprendizagem. Para encontrar o BMU, as distâncias entre um conjunto de dados de entrada (\mathbf{x}) e todos os vetores de peso (\mathbf{w}_i) da rede são comparadas. O vencedor é o que satisfaz:

$$c = \operatorname{argmin}_i \| \mathbf{x} - \mathbf{w}_i \|, \quad (2.11)$$

onde $\| \cdot \|$ é a norma de distância adotada, como distância de Manhattan, de Chebyshev, de Mahalanobis ou a Euclidiana (ASAN; ERCAN, 2012).

Depois de encontrar o BMU, os vetores de peso são atualizados para se tornarem mais representativos em relação as características do espaço de entrada. Uma determinada vizinhança do BMU também são atualizados para preservar a estrutura topológica do mapa (conforme a Figura 2.9(b)). Essa atualização requer a consideração de dois parâmetros: o tamanho da vizinhança e a taxa de aprendizagem $\alpha(t)$ que controla a taxa de variação dos vetores de peso. A regra de atualização do vetor de peso $\mathbf{w}_i(t+1)$ na iteração $t+1$ é definido por (ASAN; ERCAN, 2012; VESANTO et al., 2000):

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \alpha(t)h_{ci}(t)[\mathbf{x}(t) - \mathbf{w}_i(t)]. \quad (2.12)$$

Todos os processos descritos são repetidos para os dados de treinamento restantes até os pesos convergirem e não haja alterações perceptíveis na camada de saída de baixa dimensão. Os nós na camada de saída da rede são dispostos de acordo com uma função topológica, onde essa estrutura é geralmente bidimensional e organizada em *grids* retangulares ou hexagonais, como na Figura 2.9(b). Essas topologias têm propriedades diferentes, como por exemplo, a quantidade de vizinhos dos nós internos (BEALE; HAGAN; DEMUTH, 2012; ASAN; ERCAN, 2012).

Os processos descritos anteriormente valem tanto para redes SOM 1D ou 2D, com mudanças apenas na estrutura topológica da rede. As redes SOM 1D podem ser relacionadas às CPs, pois em certo sentido, o algoritmo SOM pode ser considerado como uma generalização do algoritmo de CPs. No entanto, há diferenças nos processos de projeção e suavização dos algoritmos. Com SOM 1D, os dados são projetados nos nós, e não na curva. Além disso, nesse caso, o algoritmo SOM visa otimizar um conjunto de vértices equipado com uma topologia unidimensional. Já as CPs, executam a suavização inteiramente no espaço de dados e seu objetivo é encontrar uma curva ótima conforme os dados (YIN, 2008).

Nota-se que existem diversas especificações entre os tipos de RNA, como sua arquitetura e os algoritmos de aprendizado e treinamento. Nesse estudo será empregado redes do tipo MLP, RBF e SOM 1D, com intuito de averiguar o desempenho de diferentes tipos de redes.

3 METODOLOGIA

Esse capítulo traz as informações sobre como o experimento foi realizado e o banco de dados utilizado para validar o método proposto. Além disso, apresenta a metodologia empregada no trabalho para alcançar os objetivos propostos.

3.1 Língua Eletrônica

A LE utilizada nesse estudo, teve sua caracterização, elaboração e aplicação do conjunto de unidades sensoriais feitas no Laboratório Nacional de Nanotecnologia para o Agronegócio (LNNA), Embrapa Instrumentação (CNPDI). O arranjo de sensores utilizado é composto por eletrodos inter digitados de ouro, fabricados pela técnica convencional de fotolitografia no Laboratório de micro fabricação do Laboratório Nacional de Nanotecnologia (LNNano-LNLS) em Campinas-SP.

Os eletrodos inter digitados são compostos por 50 pares de trilhas com comprimento de 4 mm e apresentando inter digitados com $10\mu\text{m}$ de largura e de separação entre eles. As modificações na superfície do eletrodo foram feitas através de filmes nano estruturados utilizando a técnica de automontagem (*Layer-by-Layer self assembly technique*), que permite ter o conhecimento em nível molecular de mecanismos de adsorção e a estabilidade de filmes ultra-finos. O processo é feito a partir de interações eletrostáticas entre poli cátions e poli ânions, os quais são responsáveis pela formação de pares iônicos, resultando na formação de um filme (FUKUSHIMA, 2017).

Desse modo, foram construídas seis arquiteturas diferentes, conforme apresenta a Tabela 3.1 que traz a quantidade de camadas e as substâncias¹ que as compõe.

Tabela 3.1 – Arquiteturas dos eletrodos da LE.

Eletrodo	Camadas	Arquitetura
E1	2 tetra camadas	PAH/rGO/PANI/CuTsPc
E2	2 tetra camadas	PAH/rGO/PAH/CuTsPc
E3	2 tetra camadas	PANI/AgNPs/PANI/CuTsPc
E4	2 tetra camadas	PAH/AgNPs/PAH/CuTsPc
E5	4 bi camadas	PANI/CuTsPc
E6	4 bi camadas	PAH/CuTsPc

¹ Ftalocianin tetrassulfonada de cobre (CuTsPc), poli(anilina) (PANI) (em seu estado condutor), óxido de grafeno reduzido (rGO), poli(alilamina hidrocloreada) (PAH) e nanopartículas de prata (AgNP).

Como o intuito desta dissertação é analisar os resultados obtidos pela LE com diferentes ferramentas, os detalhes sobre o preparo e a caracterização das soluções e dos filmes utilizados não serão discutidos, mas podem ser encontradas em (FUKUSHIMA, 2017).

3.2 Amostras

As amostras utilizadas no experimento são compostas por Glutamato Monossódico (GM), Inosinato Dissódico (ID), e uma mistura 1:1 (50% m/m) de Guanilato Dissódico (GD) com ID, os quais são realçadores de sabor muito utilizados em indústrias alimentícias. Os realçadores de sabor são aditivos alimentares usados para causar um efeito de gosto e sabor mais marcante nos alimentos ao paladar do consumidor (FUKUSHIMA, 2017; MILANI, 2013).

O GM é um aminoácido não essencial encontrado naturalmente em alimentos e em organismos vivos, como peixes, carnes vermelhas, leite, diversos vegetais e a partir da fermentação bacteriana do melão de cana. O ID e o GD são nucleotídeos obtidos a partir da fermentação do amido e podem ser encontrados em sardinhas secas, atum, carnes suína, bovina e de frango ou em alguns cogumelos (MILANI, 2013).

Tais aditivos são comumente adicionados aos produtos cárneos, sopas, snacks, temperos, molhos, condimentos preparados e alimentos à base de peixe. Esses realçadores proporcionam o quinto sabor básico, o “Umami”, que em japonês significa saboroso. Além disso, ele também é utilizado pelas indústrias alimentícias para o desenvolvimento de produtos com a redução da concentração de sódio (MILANI, 2013; FUKUSHIMA, 2017).

As concentrações a seguir foram determinadas segundo as instruções fornecidas em (FUKUSHIMA, 2017), considerando a utilização de uma porção de caldo em cubo diluído em 250 ml de água. A Tabela 3.2 mostra tais concentrações.

Tabela 3.2 – Concentrações dos realçadores de sabor (mol L^{-1}).

	GM	GD/ID	ID
C1	$3,54 \times 10^{-4}$	$5,00 \times 10^{-4}$	$1,01 \times 10^{-4}$
C2	$3,54 \times 10^{-10}$	$5,00 \times 10^{-10}$	$1,01 \times 10^{-10}$
C3	$1,00 \times 10^{-3}$	$1,00 \times 10^{-3}$	$1,00 \times 10^{-3}$
C4	$1,00 \times 10^{-6}$	$1,00 \times 10^{-6}$	$1,00 \times 10^{-6}$
C5	$1,00 \times 10^{-9}$	$1,00 \times 10^{-9}$	$1,00 \times 10^{-9}$
C6	$1,00 \times 10^{-12}$	$1,00 \times 10^{-12}$	$1,00 \times 10^{-12}$

3.3 Base de Dados

Os dados medidos pela LE foram obtidos usando seis unidades sensoriais conectadas nos canais para as leituras, as quais foram feitas em triplicata utilizando os *softwares Zplot 2.6* e *Zview 2.6*. As medidas de espectroscopia de impedância foram realizadas usando um analisador de impedância Solartron (1260 A).

Os dados foram coletados a partir das medidas de resistência em função da frequência de 10 Hz a 1 MHz utilizando uma tensão aplicada de 50 mV e potencial fixo (uma vez que a LE baseada na espectroscopia de impedância realiza suas medições sem a necessidade de usar um eletrodo de referência, geralmente empregada em medições eletroquímicas) (FUKUSHIMA, 2017).

Um passo importante e decisivo para realizar a análise quimiométrica para um determinado conjunto de dados está relacionado à organização dos resultados experimentais, ou seja, a matriz de resultados deve ser definida e estruturada (ZIELINSKI et al., 2014). Assim, com relação a organização dos dados, cada observação consiste em n variáveis medidas, agrupadas em um vetor de linhas n dimensional $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kn}]^T$, $\mathbf{x}_k \in R^n$. Um conjunto de N observações é denotado por $\mathbf{X} = \mathbf{x}_k, k = 1, 2, \dots, N$ e é representado como uma matriz $N \times n$:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nn} \end{bmatrix} \quad (3.1)$$

Na terminologia de RPs, as linhas de \mathbf{X} são chamadas de padrões ou objetos, as colunas são chamadas de recursos ou atributos e \mathbf{X} é chamado de matriz de padrões ou simplesmente matriz de dados. As variáveis obtidas pela LE utilizada foram: Frequência em Hertz (F), a parte real e imaginária da Impedância (Z_r e Z_i), Capacitância (C) e Resistência (R). Ou seja, nesse caso, o vetor \mathbf{X} é composto por quatro variáveis medidas com 1224 observações por classe ($\mathbf{X}_{4 \times 1224}$).

Os seis eletrodos que foram usados no experimento capturam os sinais em uma gama de frequência de 10 Hz a 1 MHz em 50 passos, em resposta à mistura das amostras. Um dos objetivos desse estudo é determinar em qual segmento ou faixa de frequência se obtém a melhor discriminancia. Mas, em detrimento da baixa quantidade de observações (eventos) disponíveis a cada passo de Frequência coletado, o que interfere diretamente na aplicação da metodolo-

gia proposta, as frequências serão analisadas em conjunto. Cada conjunto de frequências foi dividida contendo dez passos da aferição feita pela LE, conforme a Tabela 3.3 mostra.

Tabela 3.3 – Segmentos de Frequências que serão analisadas (Hz).

F₁	F₂	F₃	F₄	F₅
12,589	125,89	1258,9	12589	125890
15,849	158,49	1584,9	15849	158490
19,953	199,53	1995,3	19953	199530
25,119	251,19	2511,9	25119	251190
31,623	316,23	3162,3	31623	316230
39,811	398,11	3981,1	39811	398110
50,119	501,19	5011,9	50119	501190
63,096	630,96	6309,6	63096	630960
79,433	794,33	7943,3	79433	794330
100	1000	10000	100000	1000000

3.4 Método Proposto

O trabalho foi desenvolvido através da linguagem de programação MatLab, que é uma linguagem de alto nível orientada a diversos tipos de análises. O MatLab é um *software* científico para computação numérica que fornece um poderoso ambiente computacional para aplicações científicas. Essa linguagem provê um ambiente para interpretação, com diversas ferramentas numéricas. O uso dessa ferramenta facilita a codificação de algoritmos complexos em comparação com outras linguagens. Outra grande vantagem do MatLab é a quantidade de bibliotecas com funcionalidades específicas.

O problema de reconhecimento aqui está sendo colocado como uma tarefa de classificação ou categorização, onde as classes ou são definidas pelo projetista do sistema ou são aprendidas com base na similaridade dos padrões. A escolha da métrica é crucial para o sucesso dessa abordagem.

A metodologia proposta nessa pesquisa seguiu alguns procedimentos recomendados em uma aplicação que envolve reconhecimento de padrões, e neste caso específico, um problema de classificação. A Figura 3.1 apresenta um fluxograma dos procedimentos que foram implementados nessa pesquisa. A primeira etapa inicia-se com a aquisição dos dados que foi realizada por meio de um sistema de língua eletrônica (Seção 3.1).

Seguindo na Figura 3.1, o próximo passo é o pré-processamento do conjunto de dados obtidos. Nessa etapa, foi realizado uma exploração dos dados para observar possíveis falhas,

como a falta de dados. Além disso, essa exploração é importante e fundamental para o projetista adquirir conhecimento da aplicação e dos dados que serão analisados. Outra técnica empregada foi a normalização dos dados realizada via transformada *z-scores*.

Figura 3.1 – Fluxograma do método proposto.



Essa técnica inicia-se com a centralização dos resultados removendo o valor médio de cada recurso (\bar{x}) e, em seguida, dimensionando os resultados dividindo os recursos não constantes pelo seu desvio padrão (s), como mostra a equação abaixo:

$$\mathbf{x} = \frac{\mathbf{x}_{old} - \bar{x}}{s}, \quad (3.2)$$

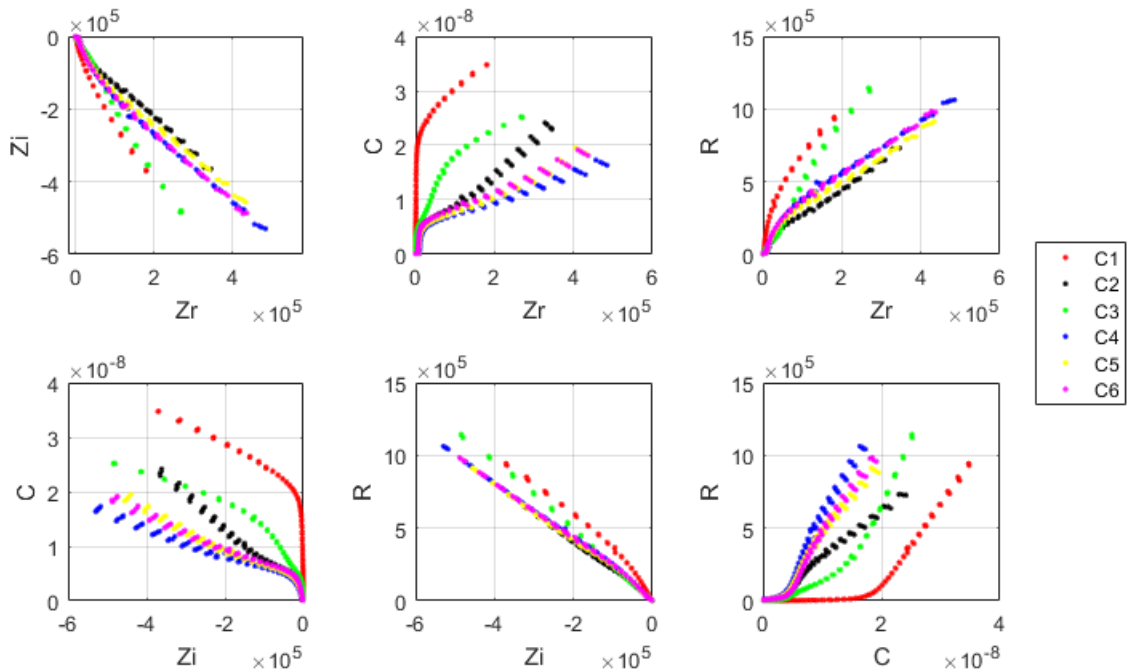
onde \mathbf{x}_{old} é o vetor de observações original e \mathbf{x} é o vetor normalizado (ZIELINSKI et al., 2014).

Em seguida, inicia-se a fase de seleção de características, selecionando-se as características (variáveis ou parâmetros) que são mais significativos para estabelecer os padrões de cada classe. Esta etapa também tem como objetivo eliminar redundância nos dados. Devido à pequena quantidade de variáveis disponíveis no banco de dados, o método escolhido para determinar as características (variáveis) mais relevantes nesse problema foi explorar todas as combinações das mesmas e, de acordo com a precisão obtida na classificação das observações, identificar quais combinações tiveram os melhores resultados.

Após definir as variáveis mais relevantes, aplica-se o classificador. A Figura 3.2 mostra o espaço de características formado pela combinação das variáveis Z_r , Z_i , C e R , em duas dimensões, para a substância GM. Os dados apresentam uma disposição no espaço de características de forma alongada e não linear. Essa disposição característica dos dados é a principal

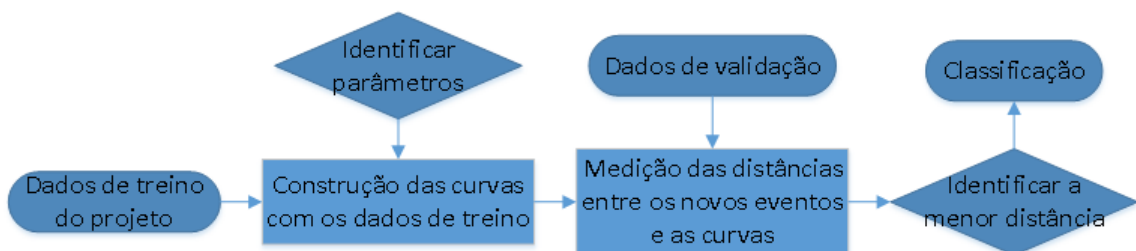
motivação para o uso da técnica de curvas principais, compreendendo que com essa técnica é possível gerar uma boa representação destes dados (MORAES et al., 2019).

Figura 3.2 – Espaço de características em duas dimensões do eletrodo E1.



A Figura 3.3 apresenta as etapas do classificador projetado que usa Curvas Principais. Em sua fase de projeto, o modelo inicia com a divisão dos dados em conjuntos de treino e de validação. Com esses conjuntos definidos, o próximo passo é a extração das CPs, feita com o algoritmo k-seg (Seção 2.4.1). Em seguida, com intuito de identificar quais parâmetros proporcionam a melhor classificação, as curvas extraídas são testadas e avaliadas de acordo com seu desempenho de classificação. Assim, estabelecendo as curvas que melhor modelam os dados.

Figura 3.3 – Fluxograma do classificador que usa CPs.



Após determinar as CPs, começa a fase operacional do modelo. Com o conjunto de dados de validação, que são novos eventos, é feito o cálculo da distância entre esses eventos e as CPs definidas. A métrica adotada foi a norma Euclidiana. A pertinência de uma nova

observação é definida a CP que atingi o menor valor dessa distância. Cada concentração (classe) é representada por uma ou mais curvas, assim bastando identificar a qual concentração a CP representa.

Para fins de comparação, outros modelos de classificador foram projetados fazendo o uso de das técnicas apresentadas na Seção 2.5, que são RNAs dos tipos MLP, RBF e SOM 1D. A Figura 3.4 traz um fluxograma com os procedimentos realizados nos modelos que usam MLP ou RBF. Os passos iniciais são análogos aos apresentados no modelo que usa CP, com a diferença que são criadas redes neurais, ao invés de curvas. Com as redes treinadas usando os melhores parâmetros definidos, os dados de validação são apresentados as mesmas. Assim, com as simulações dessas redes, são determinadas as pertinências das observações as suas respectivas concentrações (classes).

Figura 3.4 – Fluxograma do classificador que usa MLP ou RBF.

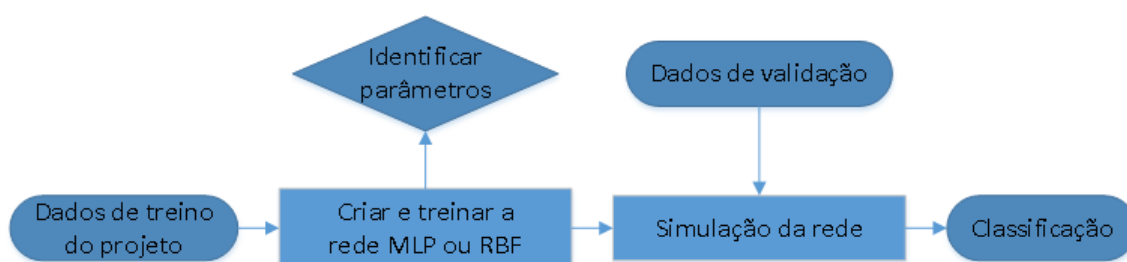
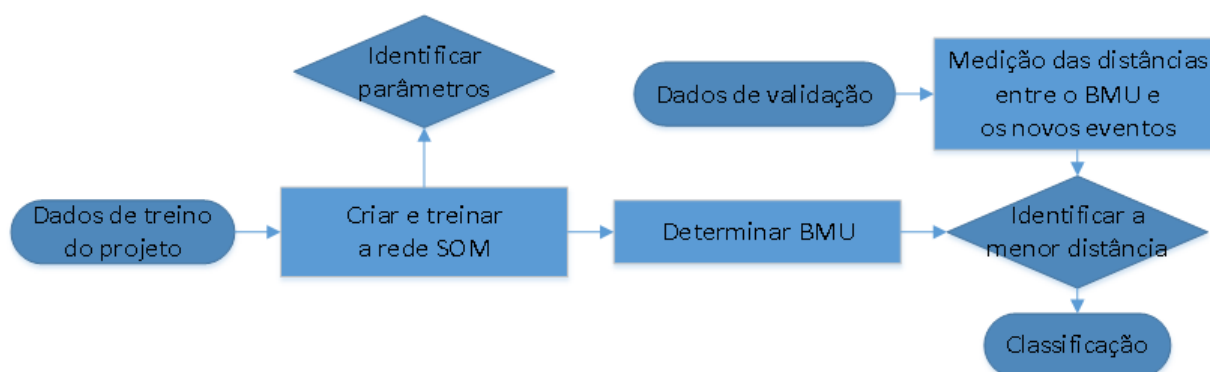


Figura 3.5 – Fluxograma do classificador que usa SOM.



Conforme traz a Figura 3.5, no caso das curvas principais via rede SOM 1D ao apresentar os dados de validação à rede, a classe do evento é definida de acordo com o neurônio vencedor (dito BMU), ou seja, aquele que apresenta a menor distância do evento processado. Como cada concentração é representada por uma certa quantidade de BMUs, basta identificá-los.

A validação cruzada é frequentemente usada para estimar a capacidade de generalização de um classificador, ou seja, o desempenho em dados previamente não vistos (CAWLEY; TALBOT, 2003). Nesse sentido, o método escolhido foi o *hold-out*. Esse método reserva um conjunto para teste e seu complemento para treinamento e, neste estudo, foram testados conjuntos com 10, 20, 30 e 40 % dos dados para teste. Uma deficiência que esse método pode apresentar é a não representatividade dos dados, em consequência disso, essa técnica foi aplicada usando estratificação, o que assegura que cada classe seja representada com a mesma distribuição em ambos conjuntos de treino e teste.

As estimativas podem ser mais confiáveis se o processo de *hold-out* for repetido com diferentes amostras, onde uma certa proporção do conjunto é aleatoriamente selecionada para treinamento. Ficou definido que esse processo será aplicado vinte vezes, assim gerando uma boa estimativa dos resultados. Os percentuais de precisão equivalem à média desses diferentes conjuntos. Em consequência de resultados similares ou próximos, a ferramenta ANOVA *One-way* (unidirecional) foi empregada, por ser um método usado para testar hipóteses nulas sobre médias iguais em várias populações (SHEIKHAN; BEJANI; GHARAVIAN, 2013).

4 RESULTADOS

Este capítulo apresenta os resultados alcançados através dos testes realizados. Os testes foram definidos de acordo com os objetivos almejados e alinhado com a metodologia traçada na seção anterior.

4.1 Identificando os parâmetros

Para identificar quais variáveis são mais relevantes na classificação, foram testadas todas as combinações possíveis entre as variáveis que compõem o banco de dados, o que resultou nas seguintes combinações:

- Zr e Zi
- Zi e R
- Zr e R
- Zr e C
- C e R
- Zi e C
- Zi, C e R
- Zr, Zi e C
- Zr, Zi e R
- Zr, Zi, C e R

Com isso e para identificar quais parâmetros devem ser usados no algoritmo k-seg para ter a melhor classificação das amostras, foram determinadas as seguintes métricas:

- Cada modelo projetado faz a classificação de vinte subconjuntos de dados mutuamente exclusivos, selecionados de acordo com o método de validação *hold-out*;
- As arquiteturas dos eletrodos foram consideradas de maneira individual, ou seja, um modelo para cada eletrodo. Desse modo, as médias das Tabelas 4.1, 4.2 e 4.3 são referentes a classificação de 120 subconjuntos de dados (6 eletrodos x 20 subconjuntos);
- Todas as frequências foram usadas, de 10 Hz a 1 MHz;
- Nesse momento, a divisão dos dados de treinamento e validação, foram definidos sendo 60% e 40%, respectivamente, o que resulta em 738 e 486 eventos, nessa ordem;
- O número máximo de segmentos que compõe as curvas foram testados usando 10, 15, 20 e 25 segmentos;

Vale lembrar que o método de curvas principais aplicado é determinístico e, portanto, as classificações do modelo sofrem alterações apenas quando o conjunto de entradas é alterado. A seguir, a Tabela 4.1 apresenta os resultados obtidos conforme as especificações anteriores com o realçador GM para o conjunto de validação.

Tabela 4.1 – Percentual médio e desvio padrão da precisão dos subconjuntos de validação obtidos com GM (%).

Entradas	10 segmentos	15 segmentos	20 segmentos	25 segmentos
ZrZi	71,11 ± 3,33	72,28 ± 2,96	72,27 ± 3,11	72,43 ± 3,34
ZrC	67,05 ± 5,97	68,51 ± 5,88	68,66 ± 5,53	68,64 ± 5,92
ZrR	66,91 ± 5,91	68,00 ± 5,76	68,53 ± 5,85	68,30 ± 5,85
ZiC	70,06 ± 3,40	71,50 ± 3,47	72,55 ± 3,69	71,60 ± 3,39
ZiR	63,95 ± 6,62	65,32 ± 6,09	65,34 ± 6,47	65,26 ± 6,61
CR	71,11 ± 3,33	72,28 ± 2,96	72,27 ± 3,11	72,43 ± 3,34
ZrZiC	76,30 ± 4,31	77,49 ± 4,26	78,13 ± 3,99	77,76 ± 3,94
ZrZiR	73,60 ± 5,18	75,37 ± 5,39	75,66 ± 5,33	75,60 ± 5,37
ZiCR	77,46 ± 4,32	79,01 ± 3,75	78,89 ± 3,65	78,83 ± 3,69
ZrZiCR	79,46 ± 4,32	80,43 ± 3,84	80,97 ± 3,94	81,02 ± 3,88

De acordo com os dados contidos na Tabela 4.1, independente da quantidade de segmentos escolhidos, os melhores resultados foram alcançados utilizando todas entradas (ZrZiCR). Os resultados obtidos por essa combinação de entrada com 15 e 20 segmentos foram bons, com 80,43% e 80,97% de precisão, respectivamente. Porém, o melhor índice de classificação foi com 25 segmentos, atingindo 81,02% de acurácia. O desvio padrão para este caso ficou abaixo de 4,00%, o que mostra a capacidade do modelo.

A Tabela 4.2 apresenta os índices de classificação dos realçadores GD/ID. Observa-se que os valores obtidos analisando esse realçador foram menores, ao compará-los com os do GM. O melhor resultado alcançou 73,76% de precisão e foram obtidos com ZrZiCR e 25 segmentos. Note que com esse realçador, o desvio padrão do melhor resultado de classificação aumentou para entorno de 9,00%.

Tabela 4.2 – Percentual médio e desvio padrão da precisão dos subconjuntos de validação alcançados com GD/ID (%).

Entradas	10 segmentos	15 segmentos	20 segmentos	25 segmentos
ZrZi	62,08 ± 7,03	64,36 ± 6,60	64,39 ± 6,71	64,42 ± 6,32
ZrC	60,99 ± 5,35	63,58 ± 5,69	64,21 ± 5,40	64,18 ± 5,79
ZrR	58,47 ± 7,15	60,27 ± 6,94	59,89 ± 7,15	60,01 ± 7,18
ZiC	59,66 ± 5,83	62,59 ± 6,07	62,82 ± 6,71	63,05 ± 6,29
ZiR	51,31 ± 8,89	53,03 ± 7,90	53,78 ± 7,68	53,17 ± 7,80
CR	62,08 ± 7,03	64,36 ± 6,60	64,39 ± 6,71	64,42 ± 6,32
ZrZiC	67,13 ± 10,67	69,75 ± 9,71	70,67 ± 9,51	70,02 ± 9,77
ZrZiR	63,75 ± 6,97	66,15 ± 7,60	66,16 ± 7,13	66,81 ± 6,91
ZiCR	65,62 ± 9,66	68,71 ± 9,14	69,29 ± 8,70	69,55 ± 8,76
ZrZiCR	70,06 ± 9,15	72,81 ± 8,94	73,59 ± 8,48	73,76 ± 8,46

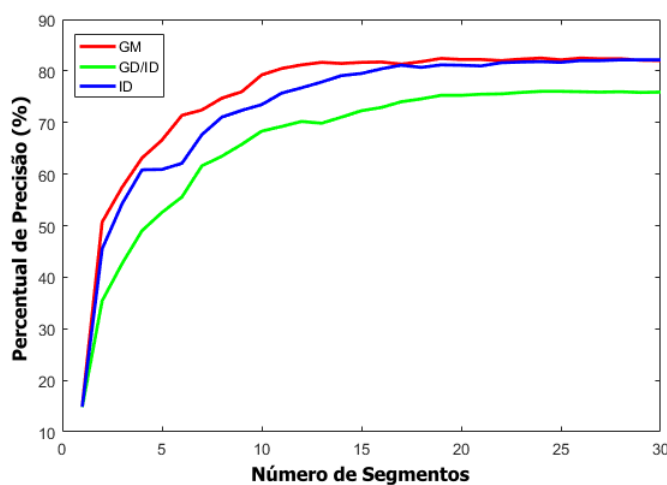
A Tabela 4.3 traz os índices de classificação dos realçadores GD/ID e ID, nessa ordem. De forma análoga às substâncias anteriores, as melhores respostas ao analisar o realçador ID foram obtidas com ZrZiCR e 25 segmentos, atingindo 79,03% de acurácia. Os resultados obtidos por essa entrada com 15 e 20 segmentos foram de 78,20% e 78,73% de precisão, nessa ordem. O desvio padrão para este caso ficou abaixo de 7,00%.

Tabela 4.3 – Percentual médio e desvio padrão da precisão dos subconjuntos de validação atingidos com ID (%).

Entradas	10 segmentos	15 segmentos	20 segmentos	25 segmentos
ZrZi	66,83 ± 7,22	68,46 ± 6,82	68,84 ± 6,73	69,57 ± 6,52
ZrC	64,68 ± 6,49	66,91 ± 6,04	67,74 ± 5,34	67,52 ± 5,67
ZrR	59,37 ± 6,69	61,67 ± 5,45	62,28 ± 4,91	62,51 ± 5,14
ZiC	65,31 ± 6,83	67,90 ± 5,89	68,37 ± 5,45	68,67 ± 6,07
ZiR	58,98 ± 6,18	60,90 ± 6,03	61,45 ± 5,28	61,18 ± 5,33
CR	66,83 ± 7,22	68,46 ± 6,82	68,84 ± 6,73	69,57 ± 6,52
ZrZiC	71,74 ± 8,15	73,91 ± 7,11	75,08 ± 6,86	75,40 ± 6,86
ZrZiR	69,61 ± 4,82	71,85 ± 5,25	73,12 ± 4,64	72,89 ± 4,79
ZiCR	70,50 ± 7,60	73,08 ± 6,93	74,04 ± 6,90	73,92 ± 6,70
ZrZiCR	75,15 ± 7,48	78,20 ± 6,71	78,73 ± 6,05	79,03 ± 5,89

Usando a mesma métrica já estabelecida, a quantidade máxima de segmentos que compõem a curva foi testada usando de 1 a 30 segmentos, para averiguar o percentual de precisão conforme o número de segmentos aumenta. A Figura 4.1 traz os resultados atingidos.

Figura 4.1 – Precisão dos subconjuntos de validação atingida conforme o número de segmentos.



Observa-se que a precisão média tende a estabilizar a partir de 20 segmentos. Ao tentar usar valores superiores a 30, o algoritmo não consegue alocar os novos seguimentos acrescentados, em detrimento das condições de convergência do algoritmo k-seg (Seção 2.4). Assim, com

base nas Tabelas 4.1, 4.2 e 4.3 e na Figura 4.1, entende-se que os parâmetros responsáveis pelas melhores médias de classificação foram: (i) ZrZiCR, combinação que usa todas as entradas e (ii) o número máximo de segmentos igual a 25.

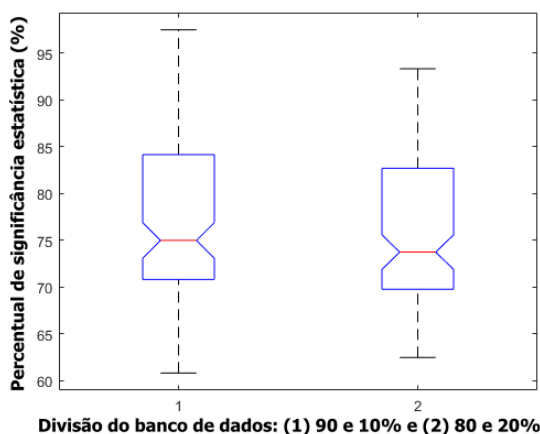
Outro parâmetro que foi explorado com esses testes foi o percentual de divisão do banco de dados. De forma que, os conjuntos de treinamento assumiram 90%, 80%, 70% ou 60% dos dados e seus respectivos complementos para os conjuntos de validação. Assim, usando os parâmetros já definidos (todas as entradas e 25 segmentos), a Tabela 4.4 traz os resultados obtidos conforme a divisão dos conjuntos de treino e validação. Lembrando que essas médias também equivalem a 120 subconjuntos de treino e validação mutuamente exclusivos.

Tabela 4.4 – Percentual médio e desvio padrão da precisão dos subconjuntos de validação conforme a divisão dos dados (%).

Treino	Validação	GM	GD/ID	ID
90%	10%	84,55 ± 4,86	77,40 ± 8,71	82,46 ± 6,70
80%	20%	83,40 ± 4,50	76,24 ± 8,52	81,58 ± 6,91
70%	30%	81,88 ± 4,03	74,95 ± 8,22	80,41 ± 6,15
60%	40%	81,02 ± 3,88	73,76 ± 8,46	79,03 ± 5,89

Os resultados foram similares, com destaque para os conjuntos de 90% e 10%. A diferença de precisão desse par com o par de conjuntos 80% e 20% foi pequena, haja vista que não há uma superioridade estatística em definitivo baseado no uso da técnica de Análise de variância (ANOVA). A Figura 4.2 traz o resultado do ANOVA analisando as médias de GD/ID. Além do teste realizado, recomendam-se uma quantidade maior que 10% de observações para validação, para ter maior representatividade dos eventos. Logo, a partir daqui, os modelos projetados destinarão 80% dos dados para treino e 20% para validação.

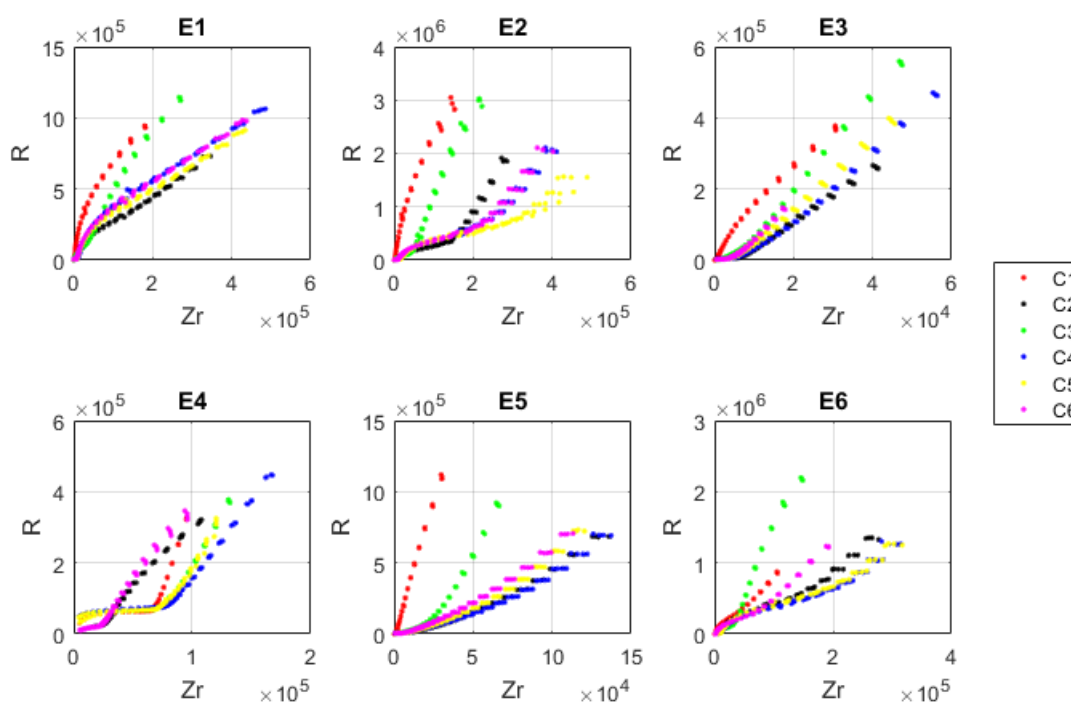
Figura 4.2 – Teste ANOVA usando as médias de validação de GD/ID.



4.2 Eletrodo mais discriminante

Esta seção tem o objetivo de identificar qual arquitetura de eletrodo que compõe a LE é mais discriminante nas classificações das concentrações analisadas. A Figura 4.3 mostra os sinais capturados por cada eletrodo em duas dimensões (Impedância real e Resistência), para a substância GM, com intuito de apresentar a diferença dos mesmos.

Figura 4.3 – Sinais capturados por cada eletrodo em 2 dimensões para a substância GM.



Conforme o objetivo definido, os classificadores foram treinados e testados com dados de seus respectivos eletrodos. A Tabela 4.5 traz os resultados obtidos por dados de cada arquitetura de eletrodo. Esses percentuais foram obtidos de acordo com os seguintes parâmetros:

- Modelo projetado usando todas entradas (Zr, Zi, C e R);
- Todas as frequências foram usadas, de 10 Hz a 1 MHz.
- O número máximo de segmentos sendo 25;
- A divisão dos dados de treinamento e validação, com 80% e 20%, respectivamente, o que resulta em 984 e 240 eventos, em ordem;
- A validação cruzada *hold-out*, com vinte subconjuntos;

Tabela 4.5 – Percentual médio e desvio padrão da precisão dos subconjunto de validação de acordo com as arquiteturas dos eletrodos (%).

Eletrodo	GM	GD/ID	ID
E1	84,33 ± 2,54	91,15 ± 1,34	88,75 ± 2,27
E2	81,81 ± 1,57	73,46 ± 2,39	68,38 ± 3,17
E3	83,85 ± 1,56	72,54 ± 2,91	84,17 ± 1,69
E4	86,33 ± 3,45	71,27 ± 3,15	84,85 ± 2,49
E5	76,02 ± 2,63	66,73 ± 2,35	79,73 ± 2,37
E6	88,02 ± 2,06	82,27 ± 2,71	83,60 ± 2,39

Analisando as amostras de GM na Tabela 4.5, observe que o eletrodo E5 foi o que teve o pior percentual, com 76,02%. Já o E6 teve o melhor desempenho, atingindo 88,02%. De modo geral, ambos eletrodos tiveram boas médias de classificação e o percentual de desvio padrão em todos os casos ficaram abaixo de 3,50%. Já com as amostras de GD/ID, apenas os eletrodos E1 e E6 tiveram boas classificações, com 91,15% e 82,27% de precisão, respectivamente. Porém, os eletrodos E2, E3 e E4 ficaram abaixo de 73,50% e E5 teve o pior desempenho, novamente. O desvio padrão do modelo ficou abaixo de 3,50%.

Nas amostras de ID, apenas o eletrodo E2 teve uma precisão baixa, ficando com 68,38%. O melhor resultado foi obtido com E1, mas os eletrodos E3, E4, E5 e E6 ficaram com percentuais superiores a 79,00% de precisão. O desvio padrão obtido foi análogo às outras substâncias analisadas.

Os eletrodos tiveram desempenhos diferentes conforme a substância analisada. Para GM, o eletrodo E6 foi o mais discriminante. Já as amostras compostas por GD/ID e ID foram melhor distinguidas pelo eletrodo E1. O eletrodo E5 foi o menos discriminativo para todas as substâncias analisadas.

Um diagnóstico importante que pode ser realizado, é investigar quais foram as concentrações que estão sendo confundidas nas classificações, sendo possível mediante a análise da matriz de confusão. As matrizes que serão apresentadas correspondem aos melhores resultados de classificação apresentados na Tabela 4.5. O Apêndice A mostra as curvas extraídas com essas especificações.

É importante evidenciar que os índices apresentados equivalem ao valor médio e seu desvio padrão, utilizando os vinte bancos de dados de treino. A Tabela 4.6 expõe a matriz de confusão das análises feitas com GM, quando o foco é a identificação da concentração da substância.

Tabela 4.6 – Matriz de confusão dos subconjuntos de validação de GM com E6 (%).

	C1	C2	C3	C4	C5	C6
C1	99,0 ± 2,2	0,6 ± 1,4	0,0 ± 0,0	0,3 ± 0,8	1,4 ± 1,5	0,0 ± 0,0
C2	0,4 ± 1,2	79,3 ± 7,0	0,0 ± 0,0	11,6 ± 7,0	9,8 ± 4,4	1,6 ± 2,6
C3	0,1 ± 0,6	0,0 ± 0,0	100,0 ± 0,0	0,0 ± 0,0	0,0 ± 0,0	0,0 ± 0,0
C4	0,0 ± 0,0	9,8 ± 4,5	0,0 ± 0,0	74,0 ± 9,9	10,8 ± 5,3	0,1 ± 0,6
C5	0,5 ± 1,5	8,5 ± 4,8	0,0 ± 0,0	13,0 ± 6,5	77,6 ± 5,5	0,0 ± 0,0
C6	0,0 ± 0,0	1,9 ± 2,3	0,0 ± 0,0	1,1 ± 1,5	0,5 ± 1,0	98,3 ± 2,6

A respeito das amostras de GM, vale destacar que a concentração C3 obteve 100,0% de classificações corretas e as concentrações C1 e C6 também tiveram ótimos índices e com percentual de confusão menor que 2,0%. As concentrações C2, C4 e C5 também tiveram bons resultados, ficando com valores entre 74,0% e 80,0%. Além disso, as mesmas foram bastante confundidas entre si, mantendo-se o valor de confusão em até 13,0%. Este resultado está de acordo com a Tabela 4.6, em que observa-se que para o Eletrodo E6 as concentrações C2, C4 e C5 estão mais confundidas no espaço de características, o que dificulta o processo de classificação.

Já as amostras de GD/ID (Tabela 4.7), teve a maior média de classificações corretas, onde as concentrações C1, C2 e C3 tiveram médias superiores a 95,0% de precisão, onde as mesmas foram pouco confundida com as demais, ficando com valores abaixo de 3,5%. As concentrações C4, C5 e C6 ficaram com médias de 83,0% a 88,0% de acurácia. Essas concentrações também foram confundidas entre si, com índices chegando a 12,0% de confusão.

Tabela 4.7 – Matriz de confusão dos subconjuntos de validação de GD/ID com E1 (%).

	C1	C2	C3	C4	C5	C6
C1	95,9 ± 4,4	0,0 ± 0,0	1,0 ± 1,7	0,0 ± 0,0	0,5 ± 1,0	0,0 ± 0,0
C2	0,0 ± 0,0	97,4 ± 3,1	0,0 ± 0,0	0,4 ± 0,9	2,6 ± 2,6	0,6 ± 1,4
C3	3,3 ± 3,7	0,1 ± 0,6	98,9 ± 1,9	0,1 ± 0,6	2,4 ± 2,4	0,0 ± 0,0
C4	0,0 ± 0,0	1,1 ± 2,1	0,0 ± 0,0	87,6 ± 6,3	5,4 ± 4,8	11,5 ± 5,3
C5	0,9 ± 2,5	0,9 ± 1,5	0,1 ± 0,6	4,5 ± 2,6	83,3 ± 5,3	4,0 ± 3,3
C6	0,0 ± 0,0	0,5 ± 1,0	0,0 ± 0,0	7,4 ± 5,0	5,9 ± 3,6	83,9 ± 5,5

Analisando amostras de ID (Tabela 4.8), os resultados foram similares aos anteriores. As três primeiras concentrações (C1, C2 e C3) ficaram com médias de 90,5% a 99,0% de precisão.

Dentre essas classes, a concentração C2 foi a mais confundida, tendo percentual de 3,4% e 5,5% de confusão com as concentrações C4 e C5, nessa ordem. As concentrações C5 e C6 tiveram boas classificações com 89,3% e 81,5%, e foram bastante confundidas com a concentração C4, ficando com 5,8% e 10,9% de confusão. A concentração C4 teve a pior classificação, com 74,5% de precisão e sendo muito confundida com as concentrações C2, C5 e C6.

Tabela 4.8 – Matriz de confusão dos subconjuntos de validação de ID com E1 (%).

	C1	C2	C3	C4	C5	C6
C1	97,9 ± 2,0	0,0 ± 0,0	0,9 ± 2,0	0,0 ± 0,0	0,0 ± 0,0	1,5 ± 1,7
C2	0,0 ± 0,0	90,5 ± 5,6	0,1 ± 0,6	10,1 ± 5,8	3,4 ± 3,0	3,1 ± 2,0
C3	0,0 ± 0,0	0,0 ± 0,0	98,9 ± 2,1	0,0 ± 0,0	0,0 ± 0,0	0,1 ± 0,6
C4	0,1 ± 0,6	5,5 ± 3,8	0,0 ± 0,0	74,5 ± 8,5	5,8 ± 4,4	10,9 ± 6,7
C5	0,6 ± 1,4	3,4 ± 3,9	0,0 ± 0,0	8,1 ± 5,0	89,3 ± 5,7	2,9 ± 3,0
C6	1,4 ± 1,5	0,6 ± 1,4	0,1 ± 0,6	7,3 ± 4,4	1,6 ± 2,3	81,5 ± 8,0

4.2.1 Combinando os eletrodos

Dando continuidade à proposta deste estudo e conforme os testes já realizados, foram estabelecidos os seguintes parâmetros:

- A combinação de entradas que teve as melhores classificações em todas as especificações utilizadas, independente do realçador, foi usando todas as variáveis disponíveis Zr, Zi, C e R;
- Todas as frequências foram usadas, de 10 Hz a 1 MHz;
- O número máximo de segmentos, com 25 segmentos;
- A divisão dos dados de treinamento e validação, com 80% e 20%, respectivamente;
- A validação cruzada *hold-out*, com vinte subconjuntos;
- Em relação aos eletrodos houve divergências nos resultados de acordo com as substâncias analisadas. Lidando com os realçadores GM ou ID, os eletrodos que se destacaram foram: E1, E3, E4 e E6. Já com as amostras de GD/ID, foram os eletrodos E1, E2, E3 e E6;

Assim, essas serão as especificações que serão empregadas nesta seção de testes. Alinhando com os objetivos estabelecidos na Seção 1.1, essa seção tem o intuito de investigar se os modelos projetados são mais eficientes em suas classificações caso os eletrodos sejam analisados em conjunto. Há duas formas de implementar os modelos com os eletrodos em conjuntos, onde a diferença entre elas está na quantidade de curvas usada para representar cada concentração.

No primeiro modo (modo 1), o modelo extrai uma curva para cada concentração usando os dados de todos os eletrodos, ou seja, apenas seis CPs. Nesse caso, quando a combinação tiver 2 eletrodos, os subconjuntos de dados contam com 1956 e 492 observações de treino e validação, na ordem. Com 3 e 4 eletrodos, os subconjuntos contam com 2940 e 3918 observações de treino e 732 e 978 observações de validação, respectivamente.

Já no segundo modo (modo 2), o modelo vai gerar seis curvas por eletrodo. Assim, ao combinar dois ou quatro eletrodos, o modelo vai contar com 12 ou 24 curvas, respectivamente. Assim, cada curva usa 984 e 240 observações de treino e validação, em ordem. Os eletrodos foram escolhidos conforme os resultados da seção anterior. Sendo assim, os resultados obtidos aplicando essas implementações estão expostos nas Tabelas 4.9, 4.10 e 4.11.

Tabela 4.9 – Percentual médio e desvio padrão da precisão dos subconjunto de validação combinando os eletrodos de GM (%).

Eletrodos	Modo 1	Modo 2
E1 e E3	80,36 ± 1,88	83,58 ± 1,10
E1 e E4	80,24 ± 2,23	85,00 ± 0,99
E1 e E6	78,62 ± 1,99	85,72 ± 1,32
E3 e E4	76,30 ± 2,24	84,45 ± 1,59
E3 e E6	81,86 ± 2,24	85,39 ± 1,44
E4 e E6	81,94 ± 1,99	87,46 ± 1,64
E1, E3 e E4	69,59 ± 1,94	84,97 ± 1,20
E1, E3 e E6	67,00 ± 2,42	85,01 ± 0,96
E1, E4 e E6	68,51 ± 2,28	86,40 ± 1,57
E3, E4 e E6	70,53 ± 1,24	86,02 ± 1,14
E1, E3, E4 e E6	62,25 ± 1,47	85,10 ± 0,92

Analisando as substâncias separadamente, as amostras de GM (Tabela 4.9), combinando dois eletrodos, tiveram os resultados variando entre 76,00% e 82,00% no modo 1 e 83,00% a 88,00% no modo 2. O melhor resultado obtido foi usando E3 e E4, que atingiu 87,46%. Com três eletrodos, o modo 1 ficou com a precisão na faixa de 67,00% a 71,00% e o modo 2, entre 84,00% e 87,00%. Note que a combinação de três eletrodos com o modo 2 tiveram

bons resultados e o melhor foi obtido usando as arquiteturas E1, E4 e E6, com 86,40% de precisão. Usando os eletrodos E1, E2, E3 e E6, as classificações tiveram 85,10% de precisão. Os melhores resultados citados tiveram sua variância abaixo de 1,7%.

Examinando as amostras compostas por GD/ID (Tabela 4.10) com dois eletrodos, os resultados de precisão ficaram entre 65,00% e 79,00% no modo 1 e 72,00% a 86,00% no modo 2. O melhor resultado foi obtido combinando dois eletrodos e com os modelos implementados pelo modo 2, por meio dos eletrodos E1 e E5, tendo 85,50% de acurácia. Nota-se que usando três eletrodos, os índices alcançados com o modo 1 ficaram na faixa de 59,00% e 68,00% e com o modo 2, entre 76,00% a 82,00%. Os eletrodos E1, E2 e E6 tiveram a melhor precisão combinando três eletrodos, com 81,96%. Ao utilizar os quatro (E1, E2, E3 e E6) a precisão foi inferior, alcançando apenas 78,78% de acurácia. Os desvios padrões dos melhores resultados mencionados foram menores que 1,5%.

Tabela 4.10 – Percentual médio e desvio padrão da precisão dos subconjunto de validação combinando os eletrodos de GD/ID (%).

Eletrodos	Modo 1	Modo 2
E1 e E2	72,44 ± 1,97	81,55 ± 1,68
E1 e E3	72,46 ± 2,42	80,62 ± 1,66
E1 e E6	78,15 ± 2,05	85,50 ± 1,49
E2 e E3	65,96 ± 2,91	72,58 ± 1,77
E2 e E6	70,16 ± 1,82	78,04 ± 1,51
E3 e E6	70,65 ± 3,39	77,56 ± 2,04
E1, E2 e E3	60,09 ± 3,40	78,33 ± 1,13
E1, E2 e E6	67,49 ± 1,95	81,96 ± 1,20
E1, E3 e E6	62,02 ± 2,15	81,52 ± 1,96
E2, E3 e E6	59,21 ± 2,16	76,22 ± 1,27
E1, E2, E3 e E6	58,55 ± 1,85	78,78 ± 1,38

Considerando as amostras de ID (Tabela 4.11), os modelos implementados com dois eletrodos e o modo 1, ficaram com percentuais de precisão entre 72,00% e 80,00%. Já com o modo 2, entre 81,00% e 86,00%. Análogo aos resultados anteriores, o melhor foi obtido combinando dois eletrodos (E1 e E3) e o modo 2, atingindo 85,77%. Ao usar três eletrodos, as médias do modo 1 ficaram entre 63,00% e 69,00% de precisão e com o modo 2, entre 82,00% e 86,00%. Onde, a combinação de E1, E3 e E4 teve o melhor desempenho, atingindo 85,04% de acurácia. Utilizando os quatro eletrodos (E1, E3, E4 e E6), a precisão foi de 83,13%. A variância das classificações dos melhores resultados apontados ficaram abaixo de 1,7%.

Tabela 4.11 – Percentual médio e desvio padrão da precisão dos subconjunto de validação combinando os eletrodos de ID (%).

Eletrodos	Modo 1	Modo 2
E1 e E3	79,35 ± 2,07	85,77 ± 1,57
E1 e E4	79,50 ± 1,91	85,84 ± 1,19
E1 e E6	72,45 ± 2,54	81,82 ± 1,25
E3 e E4	75,10 ± 2,40	83,86 ± 0,83
E3 e E6	73,41 ± 2,55	84,13 ± 1,68
E4 e E6	73,83 ± 1,59	84,49 ± 1,32
E1, E3 e E4	68,49 ± 1,85	85,04 ± 1,68
E1, E3 e E6	63,52 ± 2,63	82,97 ± 1,34
E1, E4 e E6	66,09 ± 2,13	83,38 ± 1,14
E3, E4 e E6	63,74 ± 1,95	83,95 ± 1,11
E1, E3, E4 e E6	59,53 ± 1,32	83,13 ± 1,38

A Tabela 4.12 traz os melhores resultados em detrimento da quantidade de eletrodos usados no classificador. Observe que em apenas um caso a média foi inferior a 80% de precisão. Apesar da proximidade dos resultados, nota-se que os modelos são mais precisos quando são implementados com um eletrodo. Além disso, a complexidade e o custo computacional usando um eletrodo, é menor.

Tabela 4.12 – Percentual médio e desvio padrão da precisão dos subconjunto de validação conforme o número de eletrodos (%).

	1 eletrodo		2 eletrodos		3 eletrodos		4 eletrodos	
GM	E6	88,02±2,06	E4 e E6	87,46±1,64	E1, E4 e E6	86,40±1,57	E1, E3, E4 e E6	85,10±0,92
GD/ID	E1	91,15±1,34	E1 e E6	85,50±1,49	E1, E2 e E6	81,96±1,20	E1, E2, E3 e E6	78,78±1,38
ID	E1	88,75±2,27	E1 e E4	85,84±1,19	E1, E3 e E4	85,04±1,68	E1, E3, E4 e E6	83,13±1,38

A seguir, a análise da matriz de confusão dos melhores resultados alcançados em cada substância ao combinar os eletrodos. A Tabela 4.13 traz a matriz de confusão das classificações feitas com GM. Note que as concentrações C1, C3 e C6 tiveram bons percentuais de classificações, ficando acima de 89,0%, e com relação as médias de confusão, apenas concentração C6 foi confundida com C2, a concentração C6 foi bastante confundida com C2, atingindo 10,2% de confusão, o restante ficou abaixo de 1,0%. Isso se repete ao analisar a concentração C2, onde houve 13,9% de confusão com C6. Essa classe teve a precisão mais baixa (com 76,6%) e além de C6, também foi confundida com C4 e C5, com índices de 4,1% e 4,9%, respectivamente. As concentrações C4 e C5 tiveram bons resultados, mas foram confundidas entre si e com

C2 e C3, com valores de 4,0% a 7,5% de confusão.

Tabela 4.13 – Matriz de confusão dos subconjuntos de validação de GM com E4 e E6 (%).

	C1	C2	C3	C4	C5	C6
C1	99,2 ± 1,3	0,4 ± 0,6	0,2 ± 0,5	0,4 ± 0,6	3,1 ± 2,5	0,0 ± 0,0
C2	0,0 ± 0,0	76,6 ± 3,5	0,0 ± 0,0	4,7 ± 2,9	5,4 ± 2,4	10,2 ± 4,2
C3	0,1 ± 0,3	0,1 ± 0,3	98,8 ± 1,9	6,3 ± 2,4	5,7 ± 2,3	0,0 ± 0,0
C4	0,0 ± 0,0	4,9 ± 2,8	0,8 ± 1,7	80,9 ± 5,7	5,2 ± 2,4	0,3 ± 0,7
C5	0,7 ± 1,3	4,1 ± 1,8	0,2 ± 0,6	7,3 ± 4,0	80,2 ± 4,6	0,3 ± 0,9
C6	0,0 ± 0,0	13,9 ± 3,8	0,0 ± 0,0	0,5 ± 0,6	0,4 ± 0,6	89,1 ± 4,6

Com as amostras compostas por GD/ID, Tabela 4.14, as concentrações C1, C2 e C3 ficaram com médias acima de 95,0% de precisão e apenas um caso que essas concentrações foi confundida com percentual acima de 1,5%, que foi com as concentrações C1 e C3. As concentrações C4, C5 e C6 também tiveram boas classificações, ficando entre 83,0 e 88,0% de precisão. Apesar disso, as mesmas foram bastante confundidas entre si, como por exemplo, as concentrações C4 e C6 que tiveram percentuais de 7,4% e 11,5% de confusão entre elas.

Tabela 4.14 – Matriz de confusão dos subconjuntos de validação de GD/ID com E1 e E6 (%).

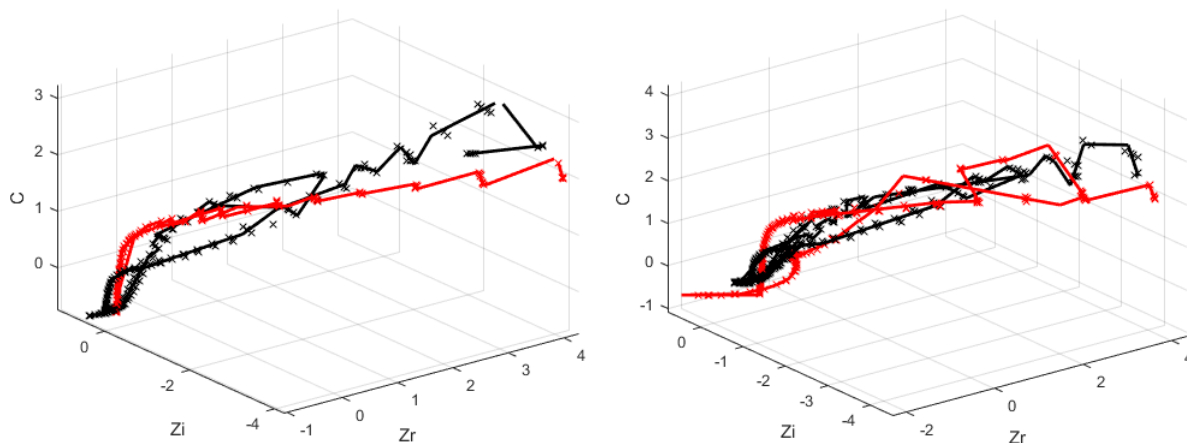
	C1	C2	C3	C4	C5	C6
C1	95,9 ± 4,4	0,0 ± 0,0	1,0 ± 1,7	0,0 ± 0,0	0,5 ± 1,0	0,0 ± 0,0
C2	0,0 ± 0,0	97,4 ± 3,1	0,0 ± 0,0	0,4 ± 0,9	2,6 ± 2,6	0,6 ± 1,4
C3	3,3 ± 3,7	0,1 ± 0,6	98,9 ± 1,9	0,1 ± 0,6	2,4 ± 2,4	0,0 ± 0,0
C4	0,0 ± 0,0	1,1 ± 2,1	0,0 ± 0,0	87,6 ± 6,3	5,4 ± 4,8	11,5 ± 5,3
C5	0,9 ± 2,5	0,9 ± 1,5	0,1 ± 0,6	4,5 ± 2,6	83,3 ± 5,3	4,0 ± 3,3
C6	0,0 ± 0,0	0,5 ± 1,0	0,0 ± 0,0	7,4 ± 5,0	5,9 ± 3,6	83,9 ± 5,5

E por fim, analisando amostras de ID por meio da Tabela 4.15, as concentrações C1, C2, C3 e C5 atingiram precisão acima de 89,0%. Referente aos percentuais de confusão, os que se destacaram negativamente foram C2 e C4, com 5,5% e C5 com C2 e C4, atingindo 3,4% e 5,8%, as demais foram abaixo desses valores. A concentração C6 teve 81,5% de precisão e foi mais confundida com C4, em torno de 10,9%. A concentração C4 teve o pior percentual de precisão, isso por que ela foi bastante confundida com C2, C5 e C6, com percentuais de 10,1%, 8,1% e 7,3%.

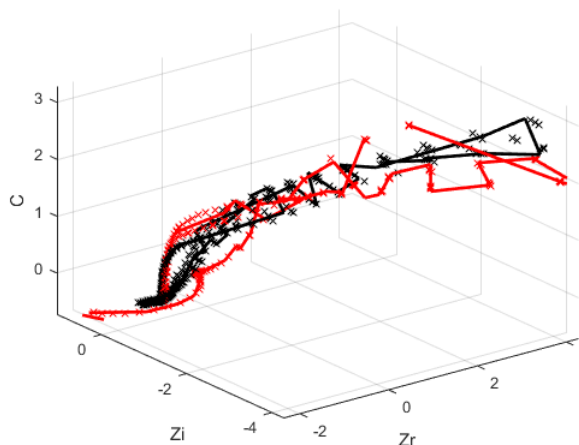
Tabela 4.15 – Matriz de confusão dos subconjuntos de validação de ID com E1 e E4 (%).

	C1	C2	C3	C4	C5	C6
C1	97,9 ± 2,0	0,0 ± 0,0	0,9 ± 2,0	0,0 ± 0,0	0,0 ± 0,0	1,5 ± 1,7
C2	0,0 ± 0,0	90,5 ± 5,6	0,1 ± 0,6	10,1 ± 5,8	3,4 ± 3,0	3,1 ± 2,0
C3	0,0 ± 0,0	0,0 ± 0,0	98,9 ± 2,1	0,0 ± 0,0	0,0 ± 0,0	0,1 ± 0,6
C4	0,1 ± 0,6	5,5 ± 3,8	0,0 ± 0,0	74,5 ± 8,5	5,8 ± 4,4	10,9 ± 6,7
C5	0,6 ± 1,4	3,4 ± 3,9	0,0 ± 0,0	8,1 ± 5,0	89,3 ± 5,7	2,9 ± 3,0
C6	1,4 ± 1,5	0,6 ± 1,4	0,1 ± 0,6	7,3 ± 4,4	1,6 ± 2,3	81,5 ± 8,0

Figura 4.4 – Curvas Principais extraídas com o modo 1.



(a) Curvas extraídas com modo 1 e os eletrodos E1 e E3. (b) Curvas extraídas com modo 1 e os eletrodos E1, E3 e E4.



(c) Curvas extraídas com modo 1 e os eletrodos E1, E3, E4 e E6.

De maneira geral, empregando o modo 1, nota-se que ao adicionar os eletrodos no modelo, a eficiência do mesmo diminui, isso ocorre devido o aumento da complexidade dos dados. Os baixos índices de classificação apresentados com esse modo podem ser justificados devido a

dificuldade em definir de maneira clara a qual classe uma nova observação pertence, haja vista a proximidade, e até mesmo a interceptação, das curvas de cada classe. A Figura 4.4 corrobora com esses aspectos, e apresenta as curvas extraídas de amostras de GM com o modo 1 usando dois, três e quatro eletrodos.

Em consequência disso e conforme os resultados das Tabelas 4.9, 4.10 e 4.11 fica evidente que o modo 2 é mais eficaz que o modo 1. Com relação a quantidade de eletrodos combinados, também fica claro que com dois eletrodos os resultados são melhores, mesmo que seja com uma pequena diferença em alguns casos. O Apêndice A traz as curvas extraídas com essas especificações.

4.3 Combinando as frequências

De acordo os objetivos definidos, nessa seção pretende-se determinar algum intervalo específico de frequência em que o percentual de precisão aumente de forma considerável. No primeiro momento, o intuito da pesquisa seria descobrir em qual sinal de frequência resulta as melhores classificações. Dessa forma, os sinais seriam analisadas ponto a ponto, de acordo com os valores captados pelo sistema de LE implementado (apresentados na Tabela 3.2).

Porém, conforme já mencionado anteriormente, em problemas de classificação com alto nível de complexidade, a quantidade de observações disponíveis na etapa de treinamento influencia na capacidade de generalização das classes. Para realizar a análise descrita acima, seria mais adequado que o banco de dados tivesse um número maior de observações. Assim, uma solução encontrada foi agrupar essas ondas. Os sinais pertinentes a cada conjunto foram evidenciadas na Tabela 3.2 e os parâmetros usados nesse modelo foram:

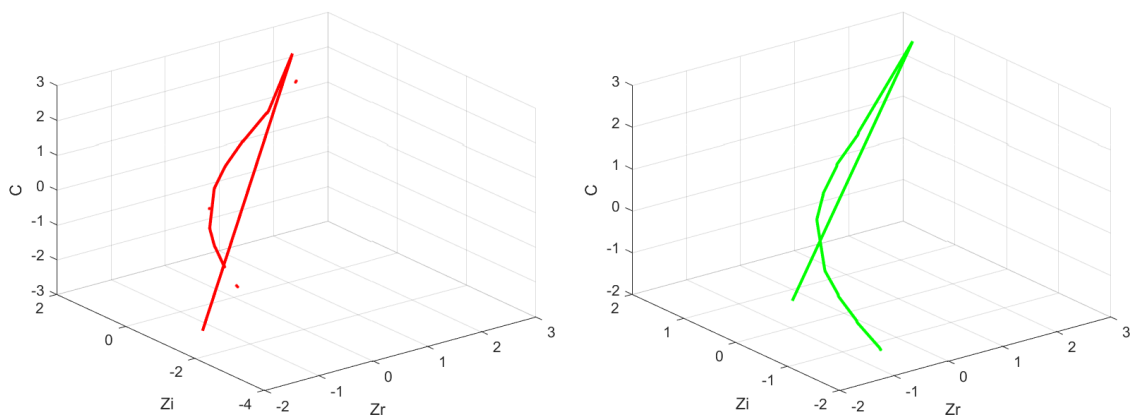
- Todas as entradas disponíveis foram utilizadas, Zr, Zi, C e R;
- O número máximo de segmentos, com 25 segmentos;
- A divisão dos dados de treinamento e validação, com 80% e 20%, respectivamente;
- A validação cruzada *hold-out*, com vinte subconjuntos;
- Os modelos foram projetados usando os eletrodos separadamente, devido os resultados apresentados na Tabela 4.12. Foram considerados apenas as melhores arquiteturas, conforme a Tabela 4.5;

A Tabela 4.16 traz os resultados obtidos com essas especificações. Observa-se que os resultados apresentados na Tabela 4.16 foram ruins, onde o melhor resultado teve apenas 69,27% de precisão (com amostras de ID, E6 e F3), fora um alto índice de variação, com 7,99%.

Tabela 4.16 – Percentual médio e desvio padrão da precisão dos subconjunto de validação de combinando as Frequências (%).

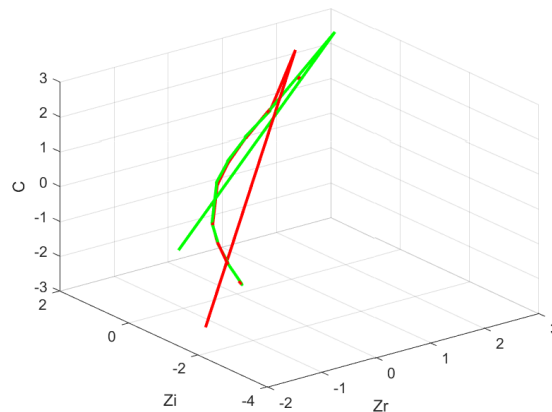
Eletrodo		F1	F2	F3	F4	F5
GM	E1	55,52±5,75	42,60±6,03	36,88±4,93	51,56±4,77	51,98±9,38
	E3	49,17±6,77	43,02±6,77	57,71±5,90	52,92±4,89	38,33±6,06
	E4	52,92±7,95	46,98±8,07	43,23±7,67	48,33±9,06	52,50±10,31
	E6	59,48±5,67	63,85±4,24	58,85±4,63	62,60±4,19	62,71±5,45
GD/ID	E1	61,25 ± 8,07	44,38 ± 7,22	63,23 ± 5,29	62,60 ± 7,07	54,06 ± 6,91
	E2	41,15 ± 3,88	47,19 ± 3,84	50,52 ± 5,65	50,52 ± 5,40	53,02 ± 7,66
	E3	43,65 ± 5,67	41,56 ± 6,60	55,73 ± 5,27	43,02 ± 4,50	41,88 ± 7,58
	E6	51,88 ± 7,25	54,58 ± 8,18	63,13 ± 7,71	64,27 ± 6,56	61,04 ± 7,03
ID	E1	45,63 ± 5,49	59,06 ± 3,46	41,98 ± 6,67	57,60 ± 4,93	62,19 ± 8,15
	E3	42,29 ± 6,56	37,71 ± 6,92	53,02 ± 6,81	48,23 ± 6,90	47,71 ± 7,34
	E4	30,94 ± 5,86	51,56 ± 7,31	33,13 ± 8,84	40,83 ± 6,74	41,25 ± 7,36
	E6	54,27 ± 6,03	62,92 ± 8,67	69,27 ± 7,99	54,27 ± 7,10	56,35 ± 8,64

Figura 4.5 – Curvas Principais extraídas de ID com E6 e F3.



(a) Curvas extraídas da concentração C1.

(b) Curvas extraídas da concentração C3.



(c) Curvas extraídas: (a) e (b).

Além da quantidade de dados usadas na fase de treinamento do modelo, esses resultados podem ser justificados devido a similaridade das classes (concentrações), como mostra a Figura 4.5. Em consequência disso, não foi possível definir ou identificar um segmento ou intervalo de frequência em que os modelos projetados tenham melhores classificações.

4.4 Classificando as substâncias

Além de classificar as diferentes concentrações das substâncias analisadas, outro objetivo dessa pesquisa é projetar um modelo que seja capaz de identificar qual é a substância em análise. Esses modelos foram projetados com as seguintes diretrizes:

- O número máximo de seguimentos do k-seg sendo 25;
- O modelo conta com todas as entradas disponíveis (Zr, Zi, C e R);
- As arquiteturas dos eletrodos foram consideradas individualmente;
- Os conjuntos de treino e validação têm 80% e 20% dos dados, resultando em 2937 e 735 observações, respectivamente;
- Os sinais obtidos por todos os pontos de Frequência foram considerados;
- E, o método de validação cruzada continua o mesmo, *hold-out*, com vinte banco de dados mutuamente exclusivos;

Dessa forma, seguindo essas especificações, a Tabela 4.17 traz os resultados obtidos.

Tabela 4.17 – Resultados de treino das classificações das substâncias (%).

Eletrodo	Precisão
E1	91,39 ± 1,28
E2	75,27 ± 2,24
E3	83,25 ± 2,34
E4	83,08 ± 2,31
E5	89,33 ± 1,32
E6	75,56 ± 2,56

Os eletrodos E1 e E5 propiciaram os melhores índices de precisão, obtendo 91,38% e 89,33%. A variação desses resultados também foram os melhores, ficando abaixo de 1,5%. Os eletrodos E3 e E4 também tiveram boas classificações, porém com os percentuais de precisão e

variação piores que os já referidos. As piores classificações foram obtidas usando os eletrodos E2 e E6, atingindo apenas 75,27% e 75,56% de precisão.

As Tabelas 4.18 e 4.19 mostram as matrizes de confusão dos eletrodos E1 e E5. Nota-se que as matrizes são parecidas, onde em ambas as amostras de ID tiveram os melhores percentuais de precisão, com mais de 91,0%. Com o eletrodo E1, os valores de confusão dessas amostras foram abaixo de 1,0% e com E5, entre 2,5% e 6,0%. As amostras de GM e GD/ID tiveram precisão superior à 85,0%, entretanto as mesmas tiveram valores de confusão entre 7,0% e 11,0% entre si. Com amostras de ID, esse percentual ficou abaixo de 3,5%. O Apêndice A apresenta as curvas extraídas com essas especificações.

Tabela 4.18 – Matriz de confusão E1 (%).

	GM	GD/ID	ID
GM	89,6 ± 2,4	10,7 ± 2,7	0,4 ± 0,4
GD/ID	9,1 ± 2,3	85,9 ± 3,3	0,9 ± 0,9
ID	1,3 ± 0,6	3,4 ± 1,2	98,7 ± 1,0

Tabela 4.19 – Matriz de confusão E5 (%).

	GM	GD/ID	ID
GM	88,4 ± 3,2	7,0 ± 3,1	2,8 ± 1,4
GD/ID	8,3 ± 2,3	88,4 ± 2,9	6,0 ± 2,0
ID	3,3 ± 1,4	4,6 ± 1,5	91,2 ± 2,3

4.5 Comparação de métodos

Para identificar o nível de eficácia do modelo que usa Curvas Principais, essa seção tem a finalidade de compará-lo com modelos implementados usando redes neurais do tipo MLP, RBF e SOM 1D. Esses modelos foram projetados conforme os fluxogramas apresentados nas Figuras 3.4 e 3.5 e os testes foram realizados levando em conta os melhores resultados de precisão apresentados nas Tabelas 4.12 e 4.17.

Iniciando as análises com modelos que usa redes MLP, existem diversos parâmetros que podem ser investigados usando essa ferramenta. Nessa pesquisa, foram considerados:

- o número de épocas de treinamento, assumindo 25, 50, 75 e 100 épocas.

- a função de treinamento, onde a mesma deve ser escolhida conforme as características do problema, segundo Beale, Hagan e Demuth (2012). Mesmo assim foram realizados testes usando as funções 'trainscg', 'trainbr' e 'trainlm'. E, os melhores resultados foram usando 'trainlm', que usa a otimização de Levenberg Marquardt;
- a quantidade de neurônios na camada oculta, variando em 5, 10, 15, 20 e 25 neurônios;
- o número de camadas ocultas, o que aumenta consideravelmente o custo computacional do modelo. Foram testados com uma e duas camadas;
- os demais, como por exemplo, o gradiente de desempenho mínimo, foram usados seus respectivos padrões.

Com relação ao número de neurônios da camada oculta e de épocas de treinamento, os melhores resultados foram obtidos com 25 neurônios e 100 épocas. Já, os classificadores projetados usando redes RBF têm os seguintes parâmetros a serem considerados:

- a meta de erro quadrático médio (*mse - mean squared error*), valor padrão 0,0. Esses valores são descritos com detalhes em (BEALE; HAGAN; DEMUTH, 2012);
- a taxa de propagação das funções de base radial, igual a 1,0;
- o número máximo de neurônios, alterando em 10, 25, 50, 75 e 100.

Os modelos com RBF tiveram melhores resultados usando no máximo 100 neurônios. Os modelos projetados com SOM 1D têm os seguintes parâmetros:

- o número de épocas de treinamento, definido igual ao utilizado com redes MLP, 100 épocas;
- a quantidade de neurônios, variando em 10, 25, 50 e 100 neurônios;
- o tamanho inicial da vizinhança, com valores 3, 10 e 25;
- a função de distância, testando os métodos 'dist', 'linkdist', 'mandist' e 'boxdist'. Detalhes dos mesmos são relatados em (VESANTO et al., 2000);
- o número de etapas de treinamento para cobertura inicial do espaço de entrada e a função de topologia foram usados os valores padrões.

Para os modelos que usaram SOM 1D, os melhores resultados também foram obtidos com 100 neurônios. Em relação ao tamanho inicial da vizinhança e a função de distância, os resultados não tiveram mudanças consideráveis entre os parâmetros testados. Assim, foram adotados os valores padrões, 3 e 'linkdist', respectivamente.

Como já mencionado, os resultados seguem as diretrizes da validação cruzada definida (método *hold-out*), ou seja, o conjunto de dados é particionado em vinte subconjuntos mutuamente exclusivos. Isto posto, é importante evidenciar que os resultados apresentados nessa tabela, foram obtidos usando os mesmos subconjuntos de dados definidos pela validação, o que propicia uma melhor comparação das ferramentas.

Além disso, os parâmetros foram testados com o máximo de similaridades possíveis para garantir que os métodos estejam sujeitos às mesmas condições que poderiam influenciar o resultado. A seguir, a Tabela 4.20 mostra todos resultados concatenados para comparar os percentuais de precisão de cada modelo e sua respectiva técnica.

Tabela 4.20 – Percentual médio e desvio padrão da precisão dos subconjunto de validação dos classificadores (%).

Classificação	Eletrodo	CP	MLP1	MLP2	RBF	SOM 1D
Concentrações	GM e E6	88,02±2,06	78,06±4,44	93,63±2,38	91,71±2,20	84,54±2,33
	GD/ID e E1	91,15±1,34	77,08±4,99	87,35±3,92	89,00±2,10	86,13±2,85
	ID e E1	88,75± 2,27	74,50±7,65	87,63±5,62	89,00±2,10	85,60±1,95
Substâncias	E1	91,39± 1,28	87,06±7,13	97,03±1,39	96,61±0,83	87,03±1,04
	E5	89,33±1,32	85,98±3,85	96,29±1,69	85,70±2,80	80,78±1,19

Com relação as redes neurais do tipo MLP1 e MLP2, os resultados mostraram que as redes MLP2 e 25 neurônios em cada tiveram os melhores resultados, destacando que todos resultados foram superiores à 87,00% de precisão. Usando redes MLP1, as classificações das concentrações ficaram entre 74,00% e 79,00% e classificando as substâncias, E1 teve 87,06% e E5 atingiu 85,98% de precisão.

A respeito das redes RBF e SOM-1D, ainda na Tabela 4.20, verifica-se que as redes RBF tiveram melhores precisões que as redes SOM 1D. Entretanto, mesmo assim as redes SOM 1D tiveram médias entre 80,00% e 88,00% de precisão, que são considerados bons resultados. Usando as redes RBF, houveram dois resultados com destaque: classificando as concentrações, amostras de GM com E6 atingiu 91,71% e classificando as substâncias, E1 teve 96,61%.

Realizando uma comparação geral de todos métodos, sobre as classificações das concentrações das amostras, os resultados mostram que conforme as amostras analisadas os métodos

tiveram diferentes precisões, como por exemplo, as amostras de GM, onde os melhores resultados foram usando MLP2 e RBF, atingindo 93,63% e 91,71% de precisão, respectivamente. Já analisando as amostras compostas por GD/ID e ID, foram usando CP e RBF com médias entre 88,75 a 91,15% de precisão.

Analisando as classificações das substâncias das amostras, de modo geral os resultados foram melhores, onde o pior foi obtido usando SOM 1D, com 80,78% de precisão. Novamente, os métodos MLP2, RBF e CP tiveram destaque (com os melhores resultados usando E1). Um fato que pode ter influenciado essas classificações terem melhores resultados é a quantidade de classes e de dados, o que diminui ou aumenta consideravelmente a complexidade do problema e a capacidade do modelo definir suas classificações.

Vale ressaltar uma característica importante dos modelos que usaram as redes neurais artificiais (MLP1, MLP2, RBF e SOM) são técnicas estocásticas, ou seja, os resultados dependem da inicialização dos pesos sinápticos da rede, assim gerando um resultado diferente cada vez que a rede é treinada e simulada. Os modelos com CPs, acontece o contrário, já que esse é um método determinístico, o que significa que os resultados só mudam caso os dados de treinamento sejam diferentes. Esse fator justifica a diferença entre o desvio padrão dos resultados de precisão dos modelos.

A fim de mensurar a complexidade computacional de cada método proposto, o número de operações matemáticas necessárias para classificar um novo evento na fase operacional (após o método já ter sido projetado) foi calculado. A Tabela 4.21 apresenta as equações que levam ao custo computacional das técnicas utilizadas nos modelos projetados, em que N_c representa a quantidade de curvas principais extraídas, N_d é a dimensão da amostra processada e nos modelos com redes neurais, N_n e N_s são os números de neurônios da(s) camada(s) oculta(s) e de saída, respectivamente.

Tabela 4.21 – Equações para obtenção do custo computacional dos modelos.

Técnicas	Adições	Multiplicações	Raiz Quadrada	Exponencial
CP	$N_c(2N_d - 1)$	$N_c N_d$	N_c	-
RBF	$N_n(2N_d - 1)$	$N_n(N_d + N_n) + N_n$	N_n	N_n
MLP2	$N_d(2N_n + N_s)$	$N_d(2N_n + N_s)$	-	-
SOM 1D	$N_n(2N_d - 1) + N_n$	$N_d N_n$	N_n	-
MLP1	$N_d(N_n + N_s)$	$N_d(N_n + N_s)$	-	-

Com auxílio da Tabela 4.21 e o objeto de análise sendo os melhores resultados (apresentados na Tabela 4.20), a seguir a Tabela 4.22 traz o custo computacional dos modelos projetados em função do número de operações necessárias para processar uma amostra do banco de dados. Neste caso, N_d é igual a 4 e é fixo para todos os modelos, já que todos foram treinados com os mesmos sub-conjuntos de dados. No caso dos modelos que contam com CP, MLP1 e MLP2, as variáveis N_c e N_s assumem 6, caso a classificação seja relacionada às concentrações e 3 se for relacionado às substâncias. Independente do tipo de classificação (concentração ou substância), N_n vale 25 para MLP1 e MLP2, é igual a 100 quando forem redes RBF ou SOM 1D.

Tabela 4.22 – Custo computacional dos modelos propostos.

Técnicas	Tipo	Adições	Multiplicações	Raiz Quadrada	Exponencial
CP	concentrações	42	24	6	0
	substâncias	21	12	3	0
RBF	ambas	700	10500	100	100
MLP2	concentrações	224	224	0	0
	substâncias	212	212	0	0
SOM 1D	ambas	800	400	100	0
MLP1	concentrações	124	124	0	0
	substâncias	112	112	0	0

Observa-se na Tabela 4.22 que os modelos que utilizam Curvas Principais realizam menos operações que as demais técnicas. Dentre os demais métodos, as redes SOM e RBF mostraram-se as mais complexas. As redes do tipo MLP1 e MLP2 só utilizam adições e multiplicações, porém com uma quantidade elevada de vezes. Apesar de ter apresentado boas médias de classificação, a rede RBF é a que realiza mais operações, ultrapassando 11 mil.

5 CONCLUSÕES E DISCUSSÕES

Neste presente trabalho usou-se um sistema de língua eletrônica com o intuito de determinar informações importantes acerca das amostras analisadas. Esse sistema é considerado um instrumento analítico capaz de caracterizar amostras aquosas com composição complexa de forma mais rápida, por meio de uma fusão de conhecimentos químicos, sensoriais e matemáticos. Devido às suas diversas vantagens, como a robustez de seus sensores e a necessidade de um pequeno volume de amostra para análise, a língua eletrônica pode ser aplicada em diversas áreas, conforme apresentado na Seção 2.2.

As amostras analisadas são compostas por Glutamato Monossódico (GM), Inosinato Dissódico (ID), e uma mistura de Guanilato Dissódico (GD) com ID. Essas substâncias, são realçadores de sabor muito utilizados em indústrias alimentícias. Tais aditivos alimentares são usados para causar um efeito de gosto e sabor mais marcante nos alimentos ao paladar do consumidor. Normalmente, adicionados a produtos cárneos, sopas, *snacks*, temperos, molhos, condimentos preparados e alimentos à base de peixe.

Geralmente, na análise sensorial os dados têm uma grande variabilidade, dessa maneira métodos estatísticos mais sofisticados devem ser utilizados para estabelecer associações, fazer inferências ou caracterizar tais amostras. A principal contribuição dessa dissertação é a metodologia proposta (Capítulo 3) e a aplicação da técnica de Curvas Principais. A justificativa dessa escolha, está na boa capacidade dessa técnica em extrair modelos compactos dos dados e a baixa complexidade computacional na fase de operação. O algoritmo usado para extrair essas curvas foi o k-segmentos não suave, devido à sua robustez na estimação das curvas, menor susceptibilidade a mínimos locais e convergência garantida.

A metodologia proposta, foi baseada em alguns procedimentos recomendados em um problema de classificação que envolve reconhecimento de padrões. Aplicando etapas de pré-processamento, seleção de características, classificador e pós-processamento. Uma das métricas definidas para avaliar o desempenho dos modelos projetados foi a validação cruzada *hold-out*, aplicada usando estratificação dos sub-conjuntos. Além disso, esse processo foi repetido vinte vezes, para ter estimativas mais confiáveis. Nas análises dos melhores resultados, fez-se o uso do ANOVA unidirecional e avaliação de suas matrizes de confusão.

Alinhando com um dos objetivos definidos (Seção 1.1), todos os modelos projetados foram implementados variando seus principais parâmetros, com intuito de encontrar os que gerassem melhores percentuais de classificações. E, conforme outra meta estabelecida, a língua

eletrônica teve as arquiteturas de seus eletrodos investigadas individualmente e combinando os mesmos para verificar com qual configuração os modelos seriam mais discriminantes.

Assim, a Tabela 4.12 mostrou que conforme o número de eletrodos eram adicionados os resultados diminuía sua precisão. Os melhores percentuais com 1 eletrodo ficaram entre 88,02% e 91,15%, enquanto com 4 eletrodos as médias alcançadas variaram de 78,78% a 85,10%. Ficou claro que ao adicionar os eletrodos nos modelos, a complexidade dos dados e consequentemente das classes aumentavam consideravelmente.

Assim, com relação a classificação das concentrações das amostras (Tabela 4.5), as arquiteturas E1 e E6 foram as mais discriminantes, ficando com médias de precisão de 84,33% a 91,15% e 82,27% a 88,02%, nesta ordem. A respeito da classificação das substâncias (Tabela 4.17), os eletrodos E1 e E5 propiciaram os melhores índices de precisão, obtendo 91,39% e 89,33%.

Por não haver um direcionamento ou estudo que aponte a frequência ou conjunto de frequência com maior capacidade de discriminação para o estudo, esta investigação foi realizada aqui. Lembrando que os sinais capturados com a língua eletrônica foram em uma gama de frequência de 1 Hz a 100 kHz em 50 passos, em resposta à mistura das amostras. Os resultados foram apresentados na Tabela 4.16 e os mesmos não foram satisfatórios, pois o melhor resultado teve apenas 69,27% de precisão e com um alto valor de desvio padrão. A Seção 4.3 discute as possíveis justificativas.

Por fim, para comparar o nível de eficácia do modelo com Curvas Principais proposto, foram implementados outros modelos usando redes neurais artificiais do tipo MLP, RBF e SOM 1D. Esses modelos foram projetados conforme mostra a Seção 3.4. Vale enfatizar que todos os modelos foram treinados e testados utilizando-se os mesmos sub-conjuntos de dados, para não haver interferência dos dados nas classificações dos modelos.

Sobre as classificações das concentrações das amostras (Tabela 4.20), os resultados mostram que conforme as amostras analisadas os métodos tiveram diferentes precisões, como por exemplo, as amostras de GM, os melhores resultados foram usando MLP2 e RBF, atingindo 93,63% e 91,71% de precisão, respectivamente. Já analisando as amostras compostas por GD/ID e ID, os melhores resultados foram usando CP e RBF com médias entre 88,75% a 91,15% de precisão. Analisando as classificações das substâncias das amostras, de modo geral os resultados foram melhores, onde o pior foi obtido usando SOM 1D, com 80,78% de preci-

são. E, novamente, os métodos MLP2, RBF e CP tiveram destaque (com os melhores resultados usando E1).

Uma vantagem dos modelos com CP, é apresentado na Tabela 4.22 que traz o custo computacional em termos de operações matemáticas das técnicas utilizadas. E, os resultados evidenciam que os modelos com redes neurais tem o custo bem mais elevado, o que torna o modelo com Curvas Principais mais atraente que os demais.

Além disso, ficou evidente a capacidade dos modelos projetados com CPs de representar muito bem o problema apresentado, levando a resultados promissores devido à sua característica não linear e boa capacidade de representação. Os resultados apresentados no Apêndice A enfatiza isso.

APÊNDICE A – Curvas Principais dos melhores resultados

As figuras a seguir mostram as Curvas Principais extraídas em suas respectivas classes (concentrações). Essas curvas seferentes aos melhores índices apresentados na Tabela 4.5.

Figura 1 – Curvas Principais de GM com E6 que teve $88,02 \pm 2,06\%$.

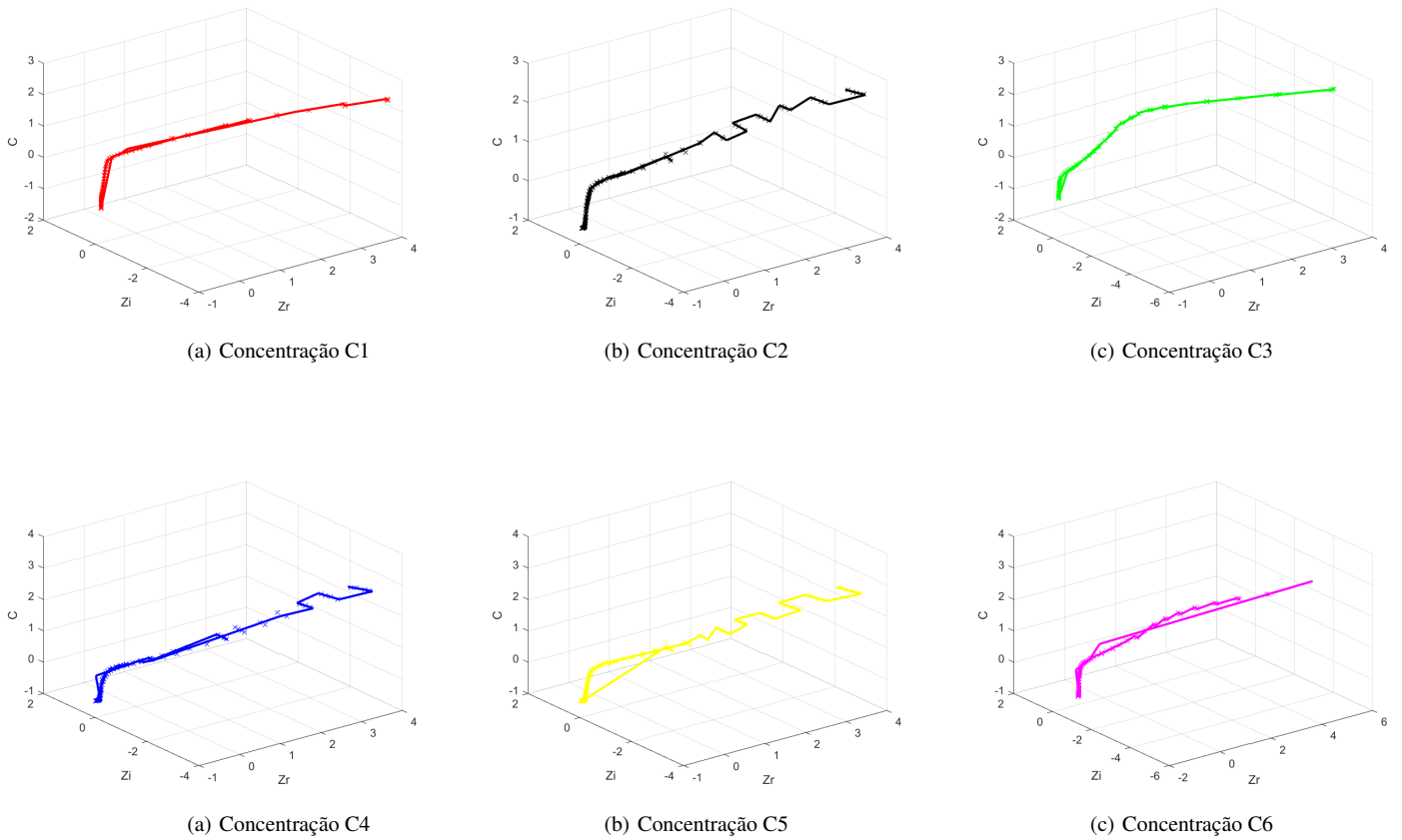
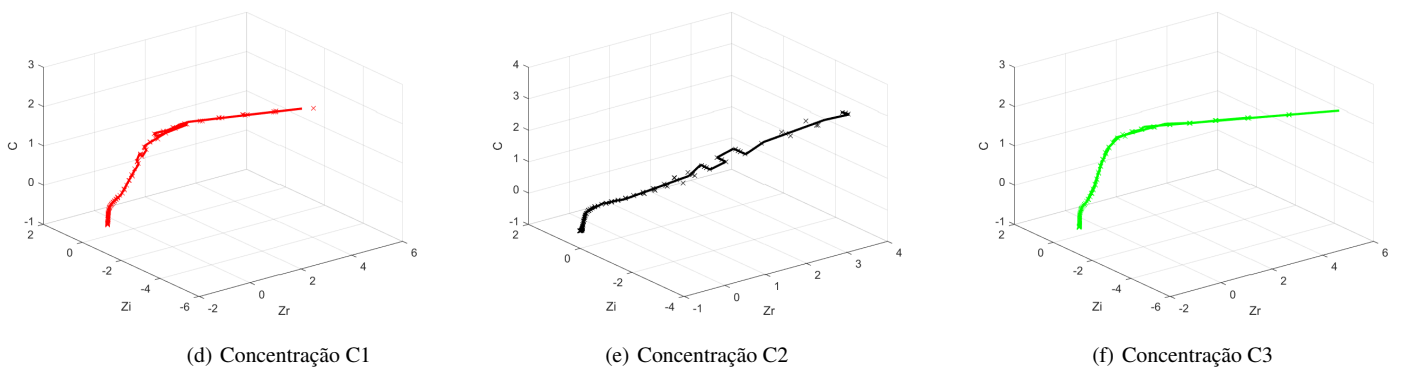
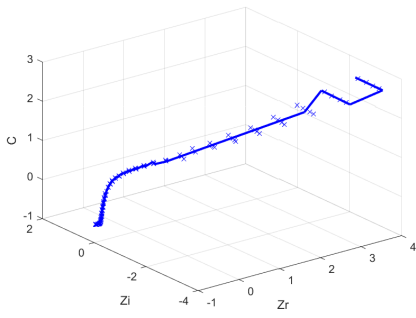
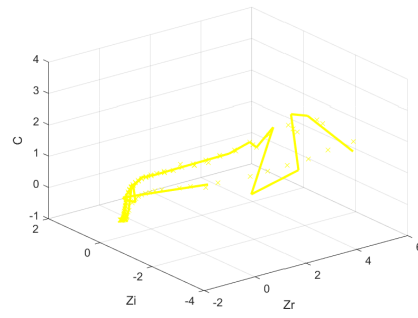


Figura 2 – Curvas Principais de GD/ID com E1 que teve $91,15 \pm 1,34\%$.

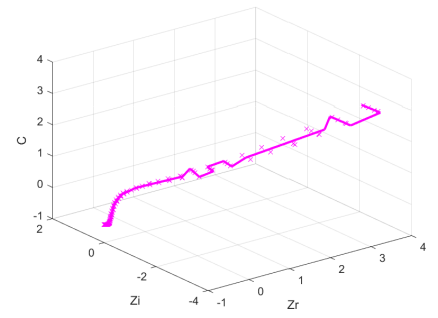




(a) Concentração C4

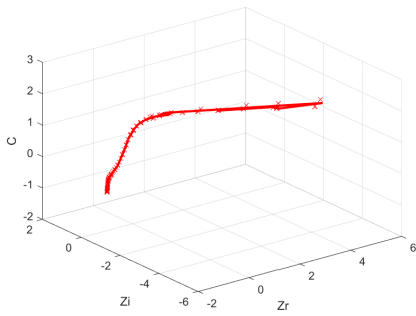


(b) Concentração C5

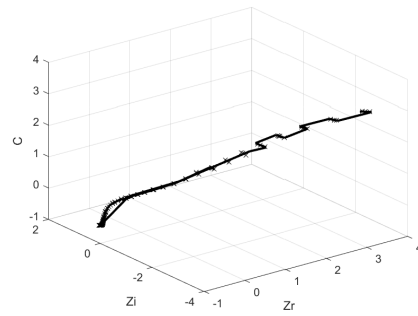


(c) Concentração C6

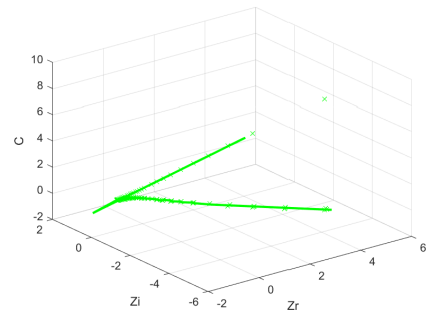
Figura 3 – Curvas Principais de ID com E1 que teve $88,75 \pm 2,27\%$.



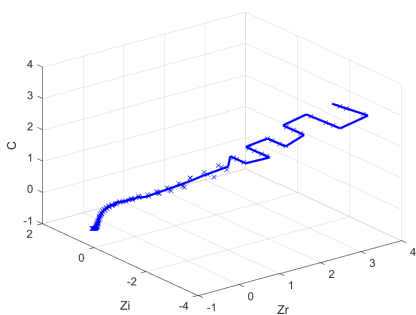
(d) Concentração C1



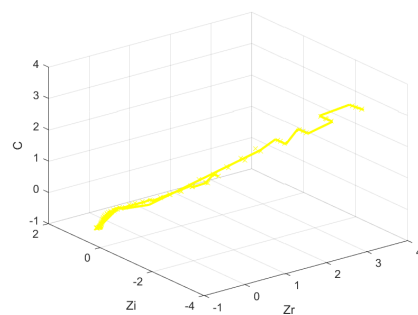
(e) Concentração C2



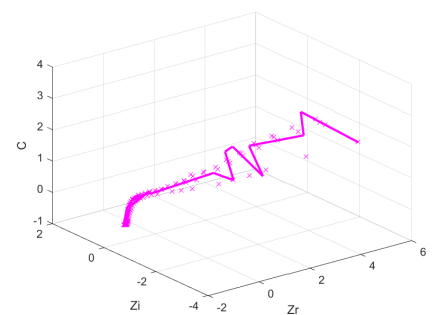
(f) Concentração C3



(a) Concentração C4



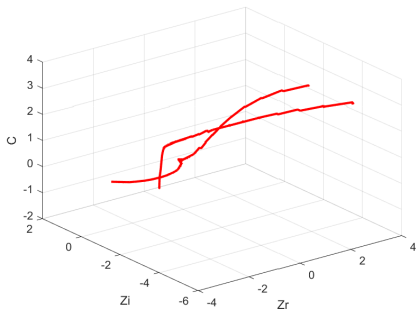
(b) Concentração C5



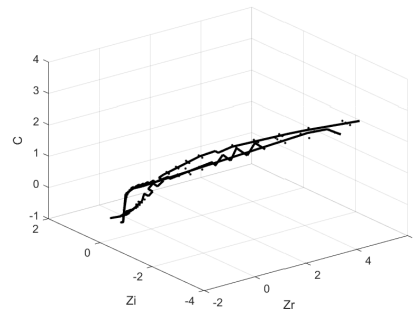
(c) Concentração C6

As próximas figuras apresentam as Curvas Principais obtidas pelos melhores índices apresentados nas Tabelas 4.9, 4.10 e 4.11.

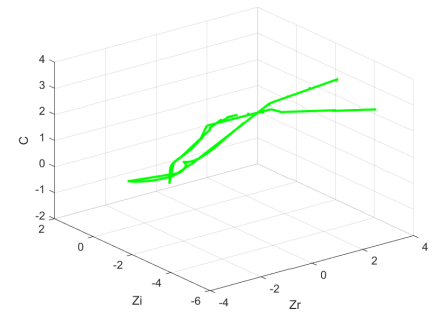
Figura 4 – Curvas Principais de GM com E4 e E6 que teve $87,46 \pm 1,64\%$.



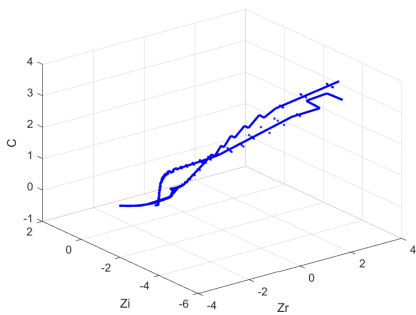
(d) Concentração C1



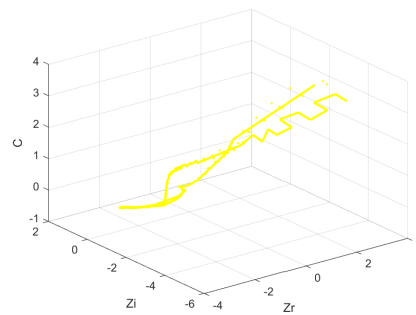
(e) Concentração C2



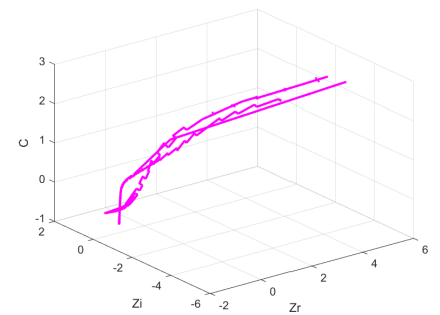
(f) Concentração C3



(a) Concentração C4

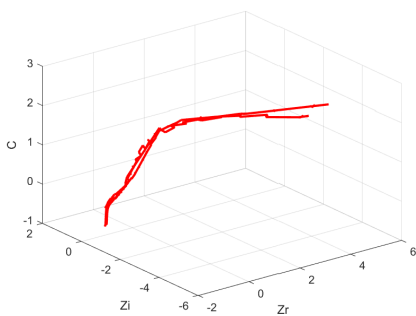


(b) Concentração C5

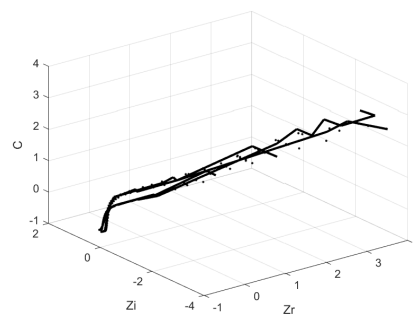


(c) Concentração C6

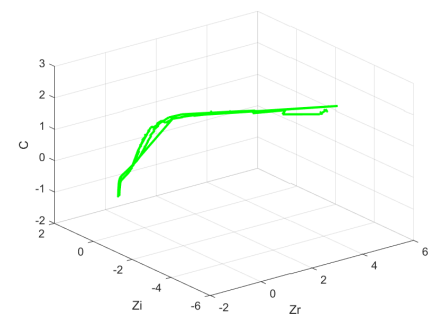
Figura 5 – Curvas Principais de GD/ID com E1 e E6 que teve $85,50 \pm 1,49\%$.



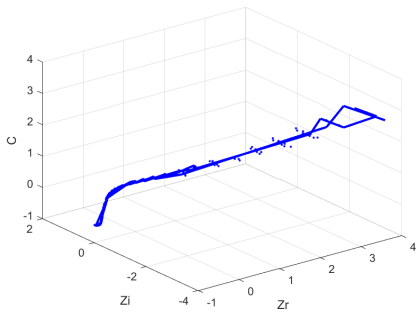
(d) Concentração C1



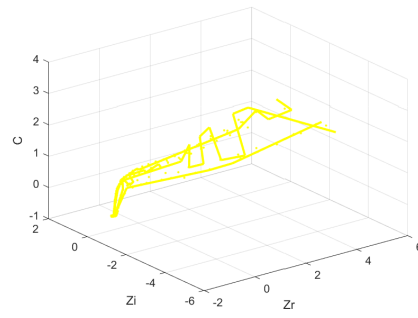
(e) Concentração C2



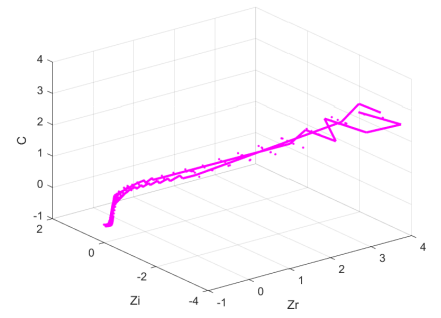
(f) Concentração C3



(a) Concentração C4

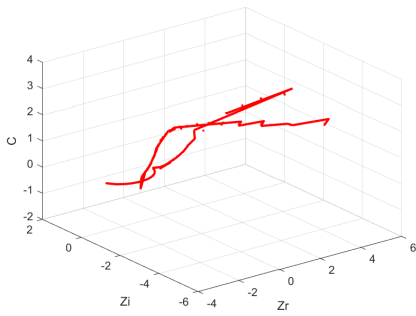


(b) Concentração C5

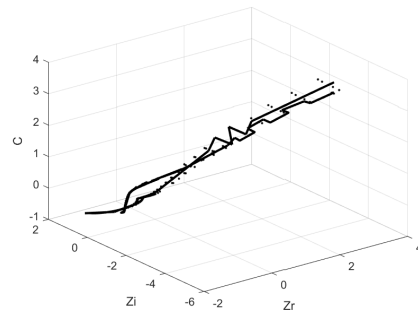


(c) Concentração C6

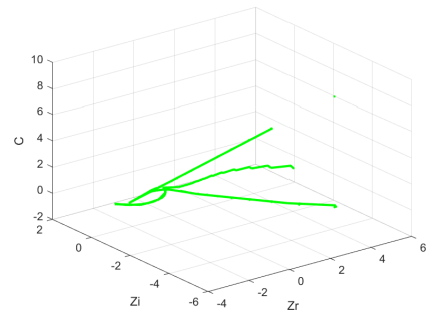
Figura 6 – Curvas Principais de ID com E1 e E4 que teve $85,84 \pm 1,19\%$.



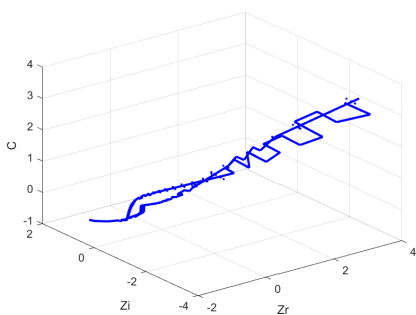
(d) Concentração C1



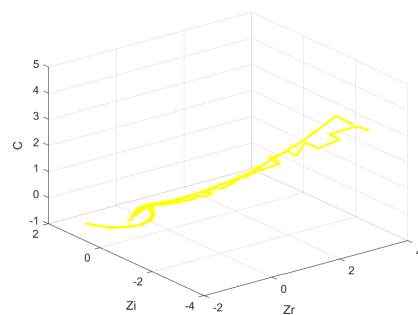
(e) Concentração C2



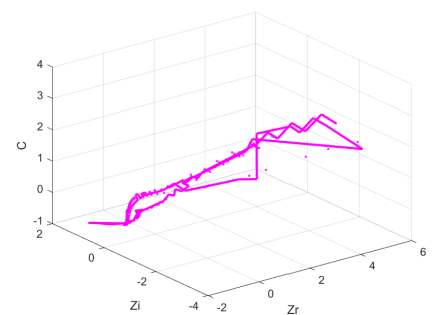
(f) Concentração C3



(a) Concentração C4

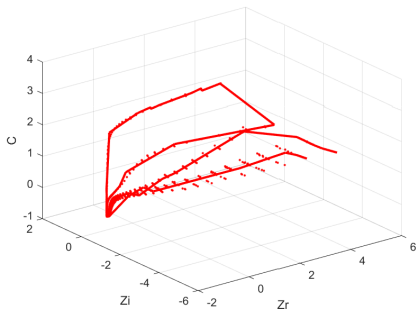


(b) Concentração C5

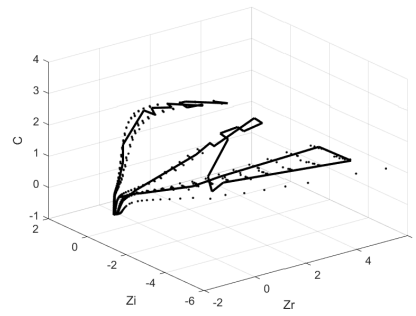


(c) Concentração C6

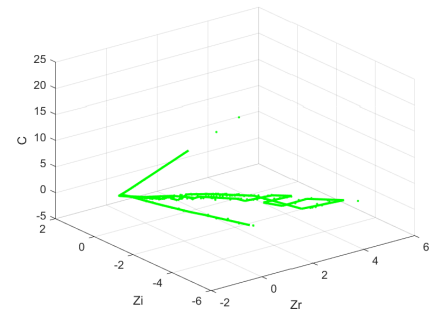
Em sequência, as figuras expõem as Curvas Principais geradas em suas respectivas classes (substâncias). Essas curvas seferentes aos melhores índices apresentados na Tabela 4.17.

Figura 7 – Curvas Principais de E1 que teve $91,39 \pm 1,28\%$.

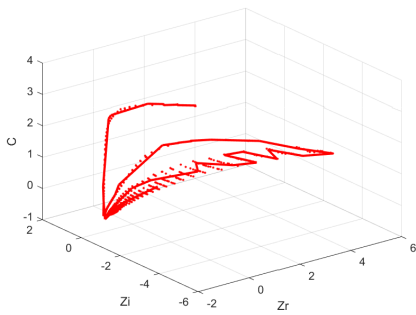
(d) Substância GM



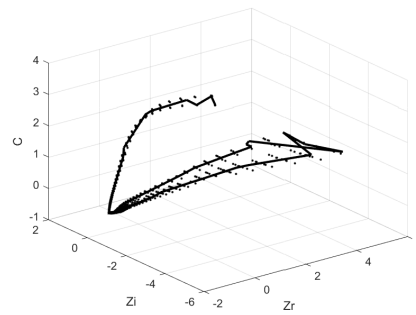
(e) Substância GD/ID



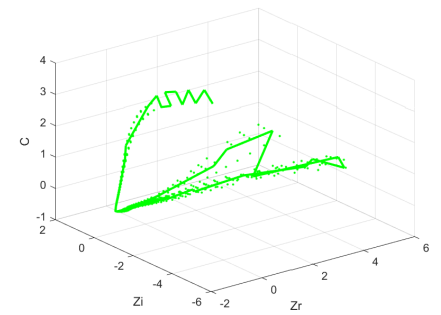
(f) Substância ID

Figura 8 – Curvas Principais de E5 que teve $89,33 \pm 1,32\%$.

(a) Substância GM



(b) Substância GD/ID



(c) Substância ID

REFERÊNCIAS

- ALBARRACÍN, W. et al. Salt in food processing; usage and reduction: A review. **International Journal of Food Science Technology**, v. 46, p. 1329 – 1336, 2011.
- ASAN, U.; ERCAN, S. **Computational Intelligence Systems in Industrial Engineering: With Recent Theory and Applications**. [S.l.]: Atlantis Press, 2012. v. 6.
- BANFIELD, J. D.; RAFTERY, A. E. Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. **Journal of the American Statistical Association**, v. 87, n. 417, p. 7 – 16, 1992.
- BEALE, M. H.; HAGAN, M. T.; DEMUTH, H. B. Neural network toolbox 7 user's guide. In: **R2012a, The MathWorks Inc.** [S.l.: s.n.], 2012.
- BEZERRA, M. A. D. **Aplicação de redes neurais artificiais no reconhecimento de padrões de cartas dinâmométricas de fundo em sistemas de bombeio mecânico de petróleo**. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Mecatrônica, Universidade Federal da Bahia, 2010.
- BORRÀS, E. et al. Data fusion methodologies for food and beverage authentication and quality assessment—a review. In: **Analytica Chimica Acta**. [S.l.]: Elsevier, 2015. v. 891, p. 1 – 14.
- BREIJO, E. G. et al. Low-cost electronic tongue system and its application to explosive detection. In: **IEEE Transactions on Instrumentation and Measurement**. [S.l.]: IEEE, 2013. v. 62, n. 2, p. 424 – 431.
- BREIJO, E. G. et al. Tnt detection using a voltammetric electronic tongue based on neural networks. **Sensors and Actuators A: Physical**, v. 192, n. Supplement C, p. 1 – 8, 2013.
- CAWLEY, G. C.; TALBOT, N. L. Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. **Pattern Recognition**, v. 36, n. 11, p. 2585 – 2592, 2003.
- CETÓ, X. et al. Instrumental measurement of wine sensory descriptors using a voltammetric electronic tongue. **Sensors and Actuators B: Chemical**, v. 207, n. Part B, p. 1053 – 1059, 2015.
- DEISINGH, A. K.; THOMPSON, D. C. S. and M. Applications of electronic noses and tongues in food analysis. **International Journal of Food Science and Technology**, v. 39, p. 587 – 604, 2004.
- DELICADO, P.; HUERTA, M. Principal curves of oriented points: theoretical and computational improvements. **Computational Statistics**, v. 18, n. 2, p. 293 – 315, 2002.
- DUCHAMP, T.; STUETZLE, W. Robust statistics, data analysis and computer intensive methods. In: _____. [S.l.]: Springer, 1996. v. 109, cap. Geometric Properties of Principal Curves in the plane.
- DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2. ed. [S.l.]: Wiley-Interscience, 2001.
- FERNANDEZ, H. L. et al. Combining morphological mapping and principal curves for ship classification. In: **International Symposium on Signals, Circuits and Systems**. [S.l.: s.n.], 2005. v. 2, p. 605 – 608.

FUKUSHIMA, K. L. **Desenvolvimento de sensores nanoestruturados para análise de realçadores de sabor em água**. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia de Biomateriais, Universidade Federal de Lavras, 2017.

GHOSH, H.; DAS, R.; ROYCHAUDHURI, C. Impedance based nanocrystalline silicon oxide immunosensor electronic tongue for ultrasensitive and low cost multiple food toxin detection. In: **IEEE International Instrumentation and Measurement Technology Conference Proceedings**. [S.l.]: IEEE, 2016.

GHOSH, H.; DAS, R.; ROYCHAUDHURI, C. Optimized nanocrystalline silicon oxide impedance immunosensor electronic tongue for subfemtomolar estimation of multiple food toxins. In: **IEEE Transactions on Instrumentation and Measurement**. [S.l.]: IEEE, 2017. v. 66, n. 5, p. 964 – 973.

GONZÁLEZ-CALABUIG, A.; CETÓ, X.; VALLE, M. del. Electronic tongue for nitro and peroxide explosive sensing. In: **Talanta**. [S.l.]: Elsevier, 2016. v. 153, p. 340 – 346.

HASTIE, T.; STUETZLE, W. Principal curves. **Journal of the American Statistical Association**, v. 84, n. 406, p. 502 – 516, 1989.

HAYKIN, S. **Adaptive Filter Theory**. 3. ed. [S.l.]: Prentice Hall, 1996.

HAYKIN, S. **Neural Networks - A Comprehensive Foundation**. 2. ed. [S.l.]: Prentice Hall, 1999.

JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. **IEEE Transactions on pattern analysis and machine intelligence**, v. 22, p. 4 – 37, 2000.

JR, A. R. et al. Recent advances in electronic tongues. **The Analyst**, v. 135, p. 2481–95, 10 2010.

KARKRA, R. et al. Classification of heavy metal ions present in multi-frequency multi-electrode potable water data using evolutionary algorithm. **Applied Water Science**, v. 7, p. 3679–3689, 2017.

KÉGL, B. et al. Learning and design of principal curves. pattern analysis and machine intelligence. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 3, p. 281 – 297, 2000.

LIU, M. et al. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and chinese vinegar. **Sensors and Actuators B: Chemical**, v. 177, n. Supplement C, p. 970 – 980, 2013.

LIU, R. et al. Bitterness intensity prediction of berberine hydrochloride using an electronic tongue and a ga-bp neural network. **Experimental and Therapeutic Medicine**, v. 7, p. 1696 – 1702, 2014.

LU, L. et al. Visualized attribute analysis approach for characterization and quantification of rice taste flavor using electronic tongue. In: **Analytica Chimica Acta**. [S.l.]: Elsevier, 2016. v. 919, p. 11 – 19.

MILANI, T. M. G. **Estudo da aromatização pré-extrusão e das condições de extrusão sobre as características físicas e retenção de aroma em proteína isolada de soja**. Dissertação (Mestrado) — Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista Julio de Mesquita Filho, 2013.

MORAES, E. C. C. et al. Data clustering based on principal curves. In: **Advances in Data Analysis and Classification**. [S.l.]: Springer Berlin Heidelberg, 2019. p. 1–20.

NERY, E. W.; KUBOTA, L. T. Integrated, paper-based potentiometric electronic tongue for the analysis of beer and wine. In: **Analytica Chimica Acta**. [S.l.]: Elsevier, 2016. v. 918, p. 60 – 68.

OTTO, M.; THOMAS, J. D. R. Model studies on multiple channel analysis of free magnesium, calcium, sodium, and potassium at physiological concentration levels with ion-selective electrodes. **Analytical Chemistry**, v. 57, n. 13, p. 2647 – 2651, 1985.

OUYANG, Q.; ZHAO, J.; CHEN, Q. Classification of rice wine according to different marked ages using a portable multi-electrode electronic tongue coupled with multivariate analysis. **Food Research International**, v. 51, n. 2, p. 633 – 640, 2013.

PEREIRA, H. C. et al. Optimization of low sodium salts mix for shoestring potatoes. **Journal of Food Science**, v. 80, p. 1 – 5, 2015.

PERIS, M.; ESCUDER-GILABERT, L. On-line monitoring of food fermentation processes using electronic noses and electronic tongues: a review. In: **Analytica chimica acta**. [S.l.]: Elsevier, 2013. v. 804, p. 29 – 36.

SAHA, P. et al. A novel technique of black tea quality prediction using electronic tongue signals. In: **IEEE Transactions on Instrumentation and Measurement**. [S.l.]: IEEE, 2014. v. 63, n. 10, p. 2472 – 2479.

SHEIKHAN, M.; BEJANI, M.; GHARAVIAN, D. Modular neural-svm scheme for speech emotion recognition using anova feature selection method. **Neural Computing and Applications**, v. 23, n. 1, p. 215 – 227, 2013.

SHLENS, J. A tutorial on principal component analysis. In: **Salk Institute for Biological Studies**. [S.l.: s.n.], 2005. v. 2, p. 1 – 13.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 4. ed. [S.l.]: Elsevier, 2009.

TIBSHIRANI, R. Principal curves revisited. **Statistics and Computing**, v. 2, n. 4, p. 183 – 190, 1992.

VERBEEK, J. J.; VLASSIS, N.; KROSE, B. A k-segments algorithm for finding principal curves. **Pattern Recognition Letters**, v. 23, n. 8, p. 1009 – 1017, 2002.

VESANTO, J. et al. Som toolbox for matlab 5. **Helsinki University of Technology**, 2000.

VLASOV, Y. et al. Nonspecific sensorarrays (“lectronic tongue”) for chemical analysis of liquids (iupac technicalreport). In: **Pure and Applied Chemistry**. [S.l.]: Elsevier, 2005. v. 77, p. 1965 – 1983.

WEI, Z.; WANG, J. Tracing floral and geographical origins of honeys by potentiometric and voltammetric electronic tongue. In: **Computers and Electronics in Agriculture**. [S.l.]: Elsevier, 2014. v. 108, p. 112 – 122.

WEI, Z.; WANG, J.; ZHANG, X. Monitoring of quality and storage time of unsealed pasteurized milk by voltammetric electronic tongue. In: **Electrochimica Acta**. [S.l.]: Elsevier, 2013. v. 88, p. 231 – 239.

YIN, H. Learning nonlinear principal manifolds by self-organising maps. In: GORBAN, A. N. et al. (Ed.). **Principal Manifolds for Data Visualization and Dimension Reduction**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 68–95.

ZIELINSKI, A. A. F. et al. Chemical composition, sensory properties, provenance, and bioactivity of fruit juices as assessed by chemometrics: a critical review and guideline. **Comprehensive reviews in food science and food safety**, Wiley Online Library, v. 13, n. 3, p. 300 – 316, 2014.