CEFET/RJ
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO

DISPOSITIVO ELETRÔNICO, UTILIZANDO APRENDIZADO DE MÁQUINA, PARA IDENTIFICAÇÃO DE MATERIAL ORGÂNICO ALIMENTAR

Almir Silva da Silveira

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de mestre em Ciência da Computação

Orientador: Prof. DSc. João Roberto de Toledo Quadros

Rio de Janeiro,

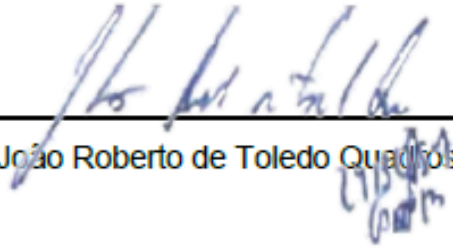
Maio de 2025

DISPOSITIVO ELETRÔNICO, UTILIZANDO APRENDIZADO DE MÁQUINA, PARA IDENTIFICAÇÃO DE MATERIAL ORGÂNICO ALIMENTAR

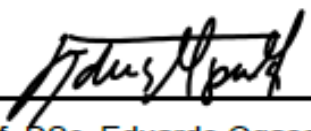
Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título em Mestre em Ciência da Computação.

Almir Silva da Silveira

Banca Examinadora:



Presidente, Prof. DSc. João Roberto de Toledo Quadros (orientador)



Prof. DSc. Eduardo Ogasawara



Prof. DSc. Carlos Otávio Schocair Mendes

Rio de Janeiro,

Maio de 2025

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

S587 Silveira, Almir Silva da
Dispositivo eletrônico, utilizando aprendizado de máquina, para identificação de material orgânico alimentar / Almir Silva da Silveira. — 2025.
81f. : il. color. , enc.

Dissertação (Mestrado) Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, 2024.
Bibliografia : f. 75-81
Orientador: João Roberto de Toledo Quadros

1. Robótica. 2. Aprendizado do computador. 3. Susentabilidade. 4. Resíduos orgânicos. 5. Algoritmos. 6. Reaproveitamento (Sobras, refugos, etc.). I. Quadros, João Roberto de Toledo. (Orient.). II. Título.

CDD 629.892

Elaborada pela bibliotecária Tania Mello – CRB/7 nº 5507/04

DEDICATÓRIA

Dedico essa dissertação a minha esposa querida, que sempre me ajudou em todas ocasiões; a Deus e toda minha família.

Rio de Janeiro,

Maio de 2025

AGRADECIMENTOS

Agradeço à minha família, a meu orientador, a todo corpo docente do PPCIC, a meus colegas de classe e, por fim, agradeço ao órgão de fomento do CNPq e CAPES, por fornecer fomento a vários trabalhos desse curso.

Rio de Janeiro,

Maio de 2025

RESUMO

DISPOSITIVO ELETRÔNICO, UTILIZANDO APRENDIZADO DE MÁQUINA, PARA IDENTIFICAÇÃO DE MATERIAL ORGÂNICO ALIMENTAR

Atualmente, há uma necessidade de ferramentas automatizadas que usem algoritmos complexos de aprendizado de máquina para ajudar na sustentabilidade. Nesse campo, é essencial criar mecanismos para identificar e separar a matéria orgânica da não orgânica para ajudar a separar e reutilizar o lixo. Este trabalho apresenta uma plataforma formada por um microcontrolador e sensores de gás, visando identificar material orgânico animal ou vegetal, mas que esteja sendo usado para fins alimentares, dentro de um padrão de certeza entre 80%-95%, que seria considerado, a princípio, aceitável. Para atingir essa meta, faz-se uso de algoritmos baseados em aprendizado de máquina. Esses algoritmos vão ser aplicados de modo a treinar a plataforma na identificação da emissão de gases de amônia, carbônico e metano, de modo que o material orgânico possa ser identificado com o grau de precisão indicado. Esse dispositivo visa a contribuição no modelo de sustentabilidade indicado pelas ODS nº 11 e 12 da ONU, podendo ser usado em latas de lixo, para facilitar o descarte correto de alimentos orgânicos, auxiliando na correta separação e posterior reutilização dos mesmos.

Palavras-chave: Robótica; aprendizado de máquina; sustentabilidade; material orgânico.

Rio de Janeiro,

Maio de 2025

ABSTRACT

ELECTRONIC DEVICES, USING MACHINE LEARNING, TO IDENTIFYING ORGANIC FOOD MATERIAL

Currently, there is a need for automated tools that use complex machine learning algorithms to help with sustainability. In this field, it is essential to create mechanisms to identify and separate organic and non-organic matter to help separate and reuse waste. This work presents a platform made up of a microcontroller and gas sensors, aimed at identifying organic animal or plant material, but which is being used for food purposes, within a standard of 80%-95% certainty, which would be considered, in principle, acceptable. To achieve this goal, algorithms based on machine learning will be used. These algorithms will be applied to train the platform to identify the emission of ammonia, carbon dioxide and methane gases, so that organic material can be identified with the indicated degree of accuracy. This device aims to contribute to the sustainability model indicated by the 11 e 12 UN's SDGs, and can be used in garbage cans to facilitate the correct disposal of organic food, helping with its correct separation and subsequent reuse.

Keywords: Robotics; Machine learning; Sustainability; Organic material.

Rio de Janeiro,

Maio de 2025

SUMÁRIO

Introdução. 15

1 - Referencial Teórico 22

1.1 - Caracterização e identificação de matéria orgânica **22**

1.2 - Algoritmos de reconhecimento de padrões usados para identificar material orgânico **25**

1.3 - Dispositivos de detecção utilizados em sistemas de captação de gases e similares **28**

1.3.1 - Sensores usados em sistemas de captação de gases 29

1.3.2 - Outros componentes dos dispositivos de identificação de matéria orgânica 31

1.4 - Algumas outras aplicações de E-Nose **33**

1.4.1 - Aplicações na indústria 33

1.4.2 - Aplicações para sustentabilidade 34

1.4.3 - Aplicação utilizada em sistemas de controle atmosférico 36

2 - Métodos e Materiais **38**

2.1 - Base de conhecimento **38**

2.2 - Modelagem do hardware e repositório **40**

2.3 - Modelagem do Software. **45**

2.3.1 - Modelagem do Aplicativo. 49

3 - Construção da base de conhecimento e metodologia de aplicação **52**

3.1 - Construção da base de conhecimento da pesquisa **52**

3.2 - Metodologia e análise dos testes **58**

4 - Discussão dos Resultados **66**

5 - Conclusões **71**

Referências **74**

Rio de Janeiro,

Maio de 2025

Lista de Figuras

Figura 1- Um exemplo de E-Nose usado em indústrias automobilísticas, com seu diagrama visto em https://phillipsind.com/ e FABER, *et al*, (2013) 34

Figura 2- Dispositivo E-nose para identificação de fase de decomposição de um alimento(a) e seu diagrama eletrônico(b) Fonte:(VASCONCELOS, *et al.*,2022) 35

Figura 3- Exemplo de código para o modulo de treinamento do Sistema E-Nose visto em VASCONCELOS, *et al.* (2022) 36

Figura 4- Dispersão de CO2 pela pressão atmosférica, obtida de um UAV a 30 m de altitude em um ambiente aberto próximo a fábricas. Fonte: (CIDADE, 2022) 37

Figura 5- Involucro de polietileno utilizado para realizar as medições dessa pesquisa. Fonte: Próprio autor 42

Figura 6- Modelo da disposição dos componentes de hardware da plataforma, no qual se visualiza a posição mais adequada desses componentes no suporte de fixação do circuito. Fonte: Próprio autor 43

Figura 7- Alocação dos componentes no dispositivo de polietileno, com o Bluetooth no lado inverso da tampa, sendo (a) alocação básica e (b) com fios de ligação. Fonte: Próprio autor 44

Figura 8- Esquema elétrico/eletrônico dos dispositivos de hardware da plataforma, mostrando suas ligações com a interface Arduino. Fonte: Próprio autor 44

Figura 9- Arquitetura do sistema, com a caixa com sensores de gás, o aplicativo móvel, os microsserviços de qualidade (dados integrados garantidos); o estimador (que lida com estatísticas de dados); os serviços para treinamento de ML e classificação de dados e o serviço de conexão. onte: Adaptação do próprio autor da figura baseada em VASCONCELOS, *et al*.(2022) 46

Figura 10- Trecho de código associado aos microsserviços utilizados, apresentando a configuração e construção das rotas usadas pela API Gateway. Fonte: Próprio autor 47

Figura 11- Parte do código do microsserviço *Treino* , para tratar da geração de tuplas JSON para serem inseridas na base de conhecimento. Fonte: Próprio autor 49

Figura 12 - Parte da programação no MIT App Inventor do módulo aplicativo, para dispositivos móvel, referente a ação de conexão com dispositivo Bluetooth da Plataforma. Fonte: Próprio autor 50

Figura 13 - Tela do aplicativo, com o comando de conexão Bluetooth, o botão para iniciar medição do material e um texto indicando a classificação da amostra, no caso, classificada como não alimento. 51

Figura 14- Exemplo de técnica de Imputação Múltipla (MI) Fonte: (STOCHERO, JACOBI e LÚCIO, 2020) 53

Figura 15- Gráfico de consolidação da base de conhecimento, sendo X o tempo (de dez em dez segundos, durante 20 minutos) e Y ppm (representando a emissão de gás). Esse gráfico apresenta os dados das emissões de gases representativos dos alimentos do tipo climatéricos, mostrando as emissões de CO2 entre 120 e 130 ppm, CH4 entre 4 e 9 ppm e NH3 entre 5 e 9 ppm. Fonte: Próprio autor 56

Figura 16- Gráfico de consolidação da base de conhecimento, sendo X o tempo (de dez em dez segundos, durante 20 minutos) e Y ppm (representando a emissão de gás). Esse gráfico apresenta os dados das emissões de gases representativos dos alimentos do tipo climatéricos, mostrando as emissões de CO2 entre 10 e 400 ppm, CH4 entre 1 e 5 ppm e NH3 entre 1 e 6 ppm. Fonte: Próprio autor. 57

Figura 17- Gráfico de consolidação da base de conhecimento, sendo X o tempo (de dez em dez segundos, durante 20 minutos) e Y ppm (representando a emissão de gás). Esse gráfico apresenta os dados das emissões de gases representativos dos alimentos do tipo carne, mostrando as emissões de CO2 entre 100 e 120 ppm, CH4 entre 4 e 12 ppm e NH3 entre 1 e 5 ppm. Fonte: Próprio autor 57

Figura 18- Gráfico de consolidação da base de conhecimento, sendo X o tempo (de dez em dez segundos, durante 20 minutos) e Y ppm (representando a emissão de gás). Esse gráfico apresenta os dados das emissões de gases representativos dos alimentos do tipo misturas mostrando as emissões de CO2 entre 120 e 141 ppm, CH4 entre 5 e 25 ppm e NH3 entre 2 e 10 ppm. Fonte: Próprio autor 58

Figura 19- Gráficos da emissão de CO2, CH4 e NH3 de uma amostra de hambúrguer e de uma amostra de lata de alumínio vazia obtidos pela plataforma. Percebe-se que a taxa de emissão de CO2, CH4 e NH3 da lata de alumínio é bem menor, em relação a amostra de hambúrguer, o que indicou o hambúrguer como orgânico alimentar e a lata como não alimento. Fonte: Próprio autor 61

Figura 20- Tela com a apresentação da resposta da plataforma em relação a amostra de hambúrguer colocada dentro da caixa de verificação da plataforma. Após clicar no botão “verificar amostra”, após três minutos, a plataforma envia os dados consolidados e o aplicativo responde com a informação de ser “orgânico alimentar” (a) ou “não alimento” (b). Fonte: Próprio autor 62

Figura 21- Gráficos de dados de emissões de CO2, CH4 e NH3 obtidos pela plataforma de uma amostra com mistura de cenoura e uva (a) e um invólucro de vidro(b).O gráfico(a) foi identificado como Orgânico Alimentar, com emissão de CO2 entre10 e 145 ppm, de NH3 entre 1 e 4 ppm e CH4 entre 1 e 5 ppm, já o gráfico(b) foi identificado como não alimento, com uma emissão de CO2 entre 0 e 10 ppm, CH4 entre 0 e 1 ppm e NH4 0 ppm Fonte: Próprio autor 63

Figura 22- Gráficos de emissões uma amostra de lata com liquido (tipo refrigerante) (a),um vidro com restos de comida (b) e um involucro de papelão(b)Nesses casos, a lata com liquido e o vidro com resto foram identificados pelo analisador com orgânicos alimentares e o papelão como não alimento, apesar de papel ser de origem orgânica, mas não ser um alimento. Fonte: Próprio autor 64

Figura 23- Arranjo do depósito de polietileno, com os componentes de hardware da plataforma, com o uso de saco plástico para isolar a amostra e permitir sua medição sem contaminação da amostra anterior. Fonte: Próprio autor 65

Lista de Tabelas

Tabela 1- Algoritmos utilizados em sistemas de reconhecimento de padrões tipo E-Nose, ou para reconhecimento e classificação de tipos de gases emitidos 28

Tabela 2- Sensores utilizados em sistemas do tipo E-Nose 30

Tabela 3- Divisão das amostras de cada tipo de alimento utilizado nessa fase da pesquisa. 40

Tabela 4- Resultados finais da 1º medição de 20 amostras de testes 68

Tabela 5- Resultados finais da 2º medição de 20 amostras de testes 69

Lista de Equações

Equação – 1 Equação de Vignoli (2021) para tratar de outliers nas planilhas das amostras de cada objeto 55

Lista de Siglas

ANN - Rede Neural Artificial

API – Application interface

ARM – Advanced RISC Machine

CO2 – Gás carbônico

CH4 – Metano

CSV – Comma Separeted Values

E-Nose - Nariz Eletrônico

GPIO – Pinos gerais de entrada e saída

HDMI - Interface Multimídia de Alta Definição

ISP - Image Signal Processor

KNN – K-vizinhos mais próximos

LDA - Análise Discriminante Linear

MAR – Missing at Random

MIT – Massachusetts Institute Technology

ML – Machine Learning

MOS – Semicondutor óxido metálico

NH3 - Amônia

ONU - Organização das Nações Unidas

ODS – Objetivos de Desenvolvimento Sustentável

PCA - Análise de Componentes Principais

RAM - Radom Access Memory

SD – Cartão de Memória Digital

SVM - Máquina de Vetores de Suporte

RISC – Reduced Instructions Set Computer

UAV – Veiculo aéreo não-tripulado

USB – Porta Serial Universal

USART - Transmissor/receptor assíncrono universal

Introdução

A quantidade de material orgânico alimentar desperdiçado no mundo, seja por descarte ou identificação incorreta, seja pela manipulação não tratável desse tipo de lixo, ou pela mistura de itens orgânicos com inorgânicos, é muito alta, ainda nos tempos atuais (RIBEIRO, 2020).

Cerca de 30% a 40% do material alimentar jogado fora torna-se inútil para reaproveitamento, gerando um aumento de 10 a 20% de custo no tratamento do lixo urbano, o que faz com que haja a necessidade de mecanismos mais elaborados para separar o lixo orgânico alimentar, de modo a evitar a contaminação e consequente inutilidade do mesmo, pois uma contaminação, por menor que seja, impede seu reaproveitamento (RIBEIRO, 2020).

Entre os problemas que existem sobre esse assunto existem, o tratamento e manuseio correto dos resíduos orgânicos alimentares; a identificação desse tipo de material, que depende da escolha da matriz mais precisa para caraterização dos mesmos e a inexistência de um sensor ou uma plataforma específica, que auxilie na classificação desse material imediatamente após seu descarte, e também muito tempo após o descarte (MELO, SILVA e DIAS, 2008) (DANTAS, et al, 2021) (ULLOA-MURILLO, *et al*, 2022).

Entre esses, o principal problema é o desenvolvimento de um ambiente de hardware e software que possa tratar da identificação da matéria orgânica de origem alimentar, auxiliando no processo de separação desses resíduos antes ou depois da ida para as unidades de tratamento de lixo. Quando o material não é identificado de forma correta, o resíduo é descartado de forma inadequada, sem a devida consideração do seu tipo, facilitando a contaminação do mesmo com resíduos inorgânicos ou venenos e sem levantar o custo em *FootPrint* associado ao descarte incorreto (AYAMBA, *et al*, 2023).

Nessa prática de separação de lixo, há aplicações com dispositivos que não são configurados especificamente para identificar lixo orgânico alimentar, e, quando instalados para tentar auxiliar no descarte de resíduos orgânicos, os sinais e avisos emitidos por esses dispositivos, se mostram inadequados e tendem a ser ignorados por aqueles que os manipulam (MARTIN, LUNARDI e

DOLCI, 2020). Quando não são corretamente identificados, os resíduos orgânicos podem ser descartados incorretamente, prejudicando as etapas de reciclagem, ou seja, o tratamento de lixo orgânico, e a capacidade de reutilização para outros fins, por sofrerem contaminações com resíduos prejudicais ao reuso desse material (PERA, *et al*., 2023) (MARTIN, LUNARDI e DOLCI, 2020).

No ato da separação do lixo, as fases que podem ser observadas são: o descarte em depósitos comuns de lixo (lixeiras em locais públicos, por exemplo), a separação em unidades de tratamento, que é muitas vezes manual e exigem várias sub-etapas para esse propósito e finalmente a alocação de lixo orgânico alimentar em local específico para seu reuso (PERA, *et al*., 2023).

O descarte de resíduos não orgânicos (por exemplo, embalagens plásticas, baterias ou material não perecível) em contêineres, nos quais deveriam estar somente resíduos orgânicos, prejudica a cadeia de reciclagem, inserindo mais uma sub-etapa de separação do material descartado incorretamente e aumentado os custos no tratamento do lixo (CONCEIÇÃO, *et al*., 2021) (BARROZO, *et al.*, 2019).

Caso essa nova sub-etapa de separação não ocorra, o resíduo orgânico pode ser contaminado e não poderá ser reciclado, principalmente se houver um resíduo não orgânico do tipo tóxico ou venenoso que se misture ou entre em difusão com o lixo alimentar, tomando-o imprestável para a sustentabilidade (CONCEIÇÃO, *et al*., 2021) (BARROZO, *et al*., 2019). Nesse caso, contribuir para desenvolver métodos e dispositivos para identificar lixo orgânico alimentar, melhorando a etapa de separação e o reaproveitamento desse tipo de resíduo, é essencial para ajudar nos critérios de sustentabilidade e reduzir o desperdício de alimentos, pois lixos orgânicos podem ser usados inclusive para gerar combustíveis não-fósseis (MARTIN, LUNARDI e DOLCI, 2020).

Uma proposta de solução, já que não existem dispositivos de baixo custo para serem usados na separação de lixo orgânico alimentar, é desenvolver um dispositivo de baixo e acessível custo, que possa ser utilizado tanto em latas de lixo comuns (como antecedendo a etapa de separação), quanto em ambientes da pós-etapa do recolhimento do lixo, para melhorar a separação e descarte de lixo dos materiais que não foram classificados antes.

Considerando esse desenvolvimento, que é a proposta dessa pesquisa, de gerar um dispositivo, com hardware e software, que seja capaz de ser inserido tanto na pré-etapa de separação (como em latas de lixo, ou depósitos comuns e públicos), quanto no auxílio da etapa pôs-recolhimento, melhorando acuricidade na identificação de material orgânico alimentar, faz-se importante escolher a matriz correta do que vai poder ser usado para identificar material orgânico alimentar.

Para atingir um grau de clareza de classificação dos resíduos, é importante a escolha da matriz mais adequada ao ambiente no qual a plataforma desenvolvida irá se inserir, o que implica na preferência sobre os parâmetros específicos, tais como, gases emitidos, umidade e textura, ou outros que servem para classificá-los (RIBEIRO, 2020).

Além disto, para melhorar a precisão dessa identificação, vai-se verificar se há contaminação orgânica em um material não-orgânico, tais como, líquidos de sucos em copos de plástico, ou se o material orgânico é de origem alimentar ou não (por exemplo, caixa de papel são material orgânico, mas não são alimentares) (PERA, *et al*., 2023) (RIBEIRO, 2020).

A solução proposta para atender a essa demanda de identificação de resíduos orgânicos, tem como objetivo apresentar uma plataforma simples e portátil, adaptável à maioria dos contêineres pequenos de descarte de resíduos existentes (tipos latas de lixo), de modo que se possa identificar ou detectar um objeto descartado como resíduo orgânico alimentar, dentro de um grau de acuricidade razoável, ou seja, que possa ajudar na identificação, e avisar ao usuário do tipo de material descartado (se orgânico alimentar ou não alimento) (DANTAS, *et al*, 2021).

Destaca-se que não foi encontrado, nas buscas das referências teóricas, um dispositivo com todas essas características, que servisse para detecção de material orgânico alimentar, concentrados em um único tipo de circuito integrado, justamente pela ausência no desenvolvimento de plataformas de identificação para esse fim. Isso pode estar ocorrendo, em parte, por não haver ainda um interesse maior nessa área e também pela hipótese de existirem circuitos similares, mas com custo elevado, utilizando-se de sensores bem mais caros

(MOLTENE E ORSATO, 2021), voltados mais para o uso no mundo industrial e não para o consumidor comum (BARROZO, *et al*., 2019).

Além do hardware, composto por sensores de alguns tipos de gases, a pesquisa também se concentrou no desenvolvimento de um modelo de software de reconhecimento de padrões, de modo que pudesse ser possível estabelecer um padrão do material orgânico alimentar com o grau de certeza entre 80% a 95%, que foi visto como o mais aceitável em outros estudos (DANTAS, *et al*, 2021), (BARROZO, *et al.*, 2019). Para essa tarefa, optou-se pelo uso de classificação de material orgânico alimentar em quatro categorias básicas:

- Climatérico: que são os alimentos que, após a colheita, continuam a amadurecer e a produzir gás carbônico e etileno, que influencia no amadurecimento e na deterioração final (abacate, cenoura, banana etc) (VASCONCELOS *et al.*, 2022).

- Não-Climatéricos: são os que não madurecem após a colheita, ou seja, devem ser colhidos apenas quando estiverem em pleno amadurecimento, produzindo menos etileno, mas ainda produzindo uma quantidade perceptível de gás carbônico (uva, limão, laranja, abacaxi etc) (VASCONCELOS *et al.*, 2022).

- Carnes: Podendo ser brancas (frangos e peixe) ou vermelhas 9bovina suína), são alimentos de origem animal que são as que fornecem proteínas indispensáveis ao com trato biológico, além de disporem de alta gama de vitaminas, minerais e gorduras (VASCONCELOS *et al*., 2022).

- Misturas: que é uma classificação introduzida neste trabalho, particularizada por não ser identificada como nenhuma das anteriores especificamente, mas sim como uma mistura de vários tipos de material orgânicos alimentares, e que existem em abundância em ambientes de praças de alimentação de shopping, por exemplo

Esse pesquisa se torna justificável pela proposta de se desenvolver uma plataforma móvel, de custo aceitável (menos de R$ 1.000,00 por unidade), capaz de identificar material orgânico alimentar com uma precisão entre 80% e 95%, conforme visto em alguns trabalhos (CAMPUZANO e GONZÁLEZ-MARTÍNEZ, 2016), sendo que essa plataforma servirá para auxiliar nos processos de separação do lixo orgânico alimentar, permitindo redução de custo e tempo nas

etapas de classificação, melhorando as ações de reaproveitamento e reuso desse tipo de lixo.

A pesquisa, e a plataforma desenvolvida, também vem de encontro aos propósitos de sustentabilidade identificados na ODS da ONU nº 11, que trata de cidades e comunidades sustentáveis, uma vez que esse dispositivo desenvolvido pode ser utilizado em lixeiras de escola, shoppings e similares; e na ODS nº 12, que trata de assegurar padrões de produção e consumo sustentável, que inclui métodos de descartes que auxiliem o reaproveitamento correto do que não foi consumido (ONU, 2024)

Como dito, além de um hardware, composto por microcontrolador de fácil programação, baseado em plataformas de desenvolvimento, do tipo Arduino (com AVR ou ARM), sensores e mecanismos de memória, fez-se importante trabalhar com algoritmos de reconhecimento de padrões, que possibilitem configurar uma etapa de treinamento e análise, usada para formar e identificar as características do padrão alvo, que é de material orgânico alimentar (WANG, *et al.*, 2022).

O tipo de algoritmo para essa situação, que melhor se adequou aos propósitos da pesquisa, foram os algoritmos de Machine Learning ou Deep Learning (ML) (WANG, *et al.*, 2022), pois, com eles pode-se construir o mecanismo para criação e formação da base de conhecimento de padrões e também o módulo que analisa os dados obtidos, o módulo analisador, no qual vai ocorrer a classificação do tipo de material (GHARAVI, *et al.*, 2022).

Não foram encontrados sensores específicos para a identificação de material orgânico alimentar (LEE, *et al.*, 2018), de modo que, para que pudesse haver o reconhecimento deste tipo de material, focou-se no uso combinado de sensores de gases sensíveis emitidos pelos processos de decomposição, ligados a matéria orgânica em geral, tais como gás carbônico, metano ou amônia (BAO, *et al.*, 2021).

Outros mecanismos de classificação, tais como verificação de textura ou mudanças de cor (CAHYA, *et al.*, 2023), não foram usados no desenvolvimento da plataforma dessa pesquisa, por encarecerem em muito o produto e por aumentar o tempo de identificação, contudo podem vir a serem usados em uma continuidade dessa pesquisa (CAHYA, *et al.*, 2023). Definiu-se, então, os

componentes essenciais, utilizados no desenvolvimento do hardware da plataforma, como os sensores para verificação das quantidades emitidas de gás carbônico, metano e amônia de cada amostra e uma interface de controle de baixo custo e processamento rápido (ZI-JUN, *et al*., 2024).

Para a pesquisa também foi utilizado um repositório simples de polietileno, com base nos dados de VASCONCELOS *et al*. (2024), no qual foram inseridas diversas amostras de materiais. Nesse repositório foram instalados o circuito, com os sensores e microcontrolador, usado para detecção da matriz escolhida para a identificação, que foi, nesse caso, o gás metano, amônia e gás carbônico. Junto a esse circuito, foi acrescentado um dispositivo de comunicação do tipo Bluetooth/WiFi, para poder enviar os dados obtidos para o construtor da base e também para enviar os dados da emissão de gases da amostra para o analisador.

Para a configuração da base de conhecimento, o sinal emitido pelo dispositivo de comunicação envia os dados dos gases identificados para um módulo construtor. As amostras para esse fim foram inseridas no repositório de polietileno, de modo a traçar as características de emissão de gases associados a tipos de lixo diferentes, para ocorrer a classificação adequado no momento dos testes.

Os detalhes desses procedimentos são descritos ao longo dessa dissertação, e consistem em deixar as amostras por um tempo específico no repositório, de modo a captar os dados de emissão de gás carbônico (identificado como CO2), amônia (identificado como NH3) e metano (identificado como CH4), em processos de decomposição bem inicial (VASCONCELOS, *et al*., 2024), sendo que essa exposição e captação serviu tanta para construção da base de conhecimento, quanto, posteriormente, na ação de classificação de amostras em geral.

Na construção da base de conhecimento, esses dados das amostras enviados foram formatados como planilhas, para que pudesse ocorrer a “limpeza” e acertos dos dados, como, por exemplo a retirada de *outliers*, ou dados muito discrepantes, e, posteriormente, a consolidação das informações. A limpeza realizada para que não fossem consideradas as informações que poderiam causar problemas na classificação, no caso, os *outliers*, por exemplo,

foram percebidos em situações de instabilidade dos sensores, ou mesmo por problema na emissão dos dados via comunicação sem fio, para o módulo de construção da base de conhecimento.

Depois de formado a base de conhecimento, iniciou-se os testes e exames para aferição da plataforma. Diversos tipos de amostras, foram colocados no repositório, por um tempo, e os dados obtidos foram usados para a classificação pelo componente analisador, resultando em uma identificação como material orgânico alimentar, orgânico não alimentar ou inorgânico. Como visto, uma parte importante da pesquisa, além do desenvolvimento da própria plataforma em si, cujo construção apresentou diferenças do visto na literatura, se concentrou também no desenvolvimento do software, baseado na manipulação dos dados obtidos pelos sensores de gás, para a detecção acurada do tipo de material orgânico alimentar.

Para melhor entendimento sobre os métodos e resultados dessa pesquisa, ela é contém, além da introdução, a fundamentação teórica (com a respectiva revisão bibliográfica sobre o assunto, inclusive com verificação do uso de sistemas similares em outras aplicações), metodologia utilizada, com os materiais, os resultados já obtidos, os testes executados, discussão sobre eles e a conclusão.

1- Referencial Teórico

Visando dar maior segurança nos dados e conclusões observadas nesse trabalho, foi realizado uma pesquisa bibliográfica, visando fornecer a devida sustentação teórica para o desenvolvimento da plataforma, seja para o hardware ou o software. A busca de trabalhos acadêmicos e profissionais se deu nas plataformas Scopus, Google Scholar e Elsevier, procurando trabalhos mais recentes, ou que se constituíssem como estado da arte sobre o tema.

Os assuntos abordados para essa pesquisa bibliográfica foram: modelos e dispositivos de identificação de matéria orgânica comum, associando a métodos de reconhecimento e treinamento de padrões, procurando-se algoritmos e metodologias baseadas em ML; dispositivos de baixo custo, a serem utilizados na construção da plataforma, e em sistemas de nariz eletrônico (E-Nose) de outras aplicações já existentes.

1.1- Caracterização e identificação de matéria orgânica

Conforme visto em SAILER, *et al.* (2021), um material é considerado orgânico quando definido como matéria de origem animal ou vegetal, na forma biológica, ou também como compostos geológicos que tiveram origem orgânica, tais como os encontrados abaixo da superfície do solo. Os parâmetros que melhor identificam esse tipo de material incluem a sua textura, cor, as emissões de gases específicos (metano, gás carbônico, oxigênio, nitrogênio, etileno e amônia), evaporação de água/tempo e sua aparência geral (SAILER, *et al.*, 2021).

Esses parâmetros podem ser capturados, em ambientes próprios, através de técnicas de processamento e classificação configuradas para atuar sobre esses aspectos (MARATOS, *et al.*, 2023) (VASCONCELOS, *et al*., 2022)

(CAMPUZANO e GONZÁLEZ-MARTÍNEZ, 2016). Entre os diversos métodos de identificação de material orgânico, incluem-se os que identificam:

- A textura, que tratam de verificar os tipos de materiais que têm alta viscosidade, que seriam representativos de matéria orgânica (STRIBIŢCAIA, *et al.*, 2020) (CAMPUZANO e GONZÁLEZ-MARTÍNEZ, 2016);
- Os que trabalham com identificação de cor e umidade, que prescindem a utilização de algoritmos de reconhecimentos de padrões de alta precisão de classificação (MARATOS, *et al.*, 2023) (VIEJO e FUENTES, 2020), e
- Os que trabalham através da identificação da emissão de gases específicos dos materiais (SANJAYA, *et al*.,2023) (VASCONCELOS, *et al*., 2022).

Essa diversidade de métodos faz com que a identificação de matéria orgânica se torne bem mais complexa do que de outros tipos de materiais, sendo que a identificação desse tipo de material requer uma combinação de hardware (sensores) e software (algoritmos de reconhecimento) específicos (VIEJO e FUENTES, 2020).

A partir desses métodos, verifica-se que os parâmetros, que podem ser usados para identificar e classificar material orgânico alimentar são: cor, textura, umidade e emissão de gases. Para se trabalhar com cada um desses parâmetros são necessários algoritmos de treinamentos bem sofisticados, para que seja realizado uma detecção e classificação mais acurada dos objetos examinados (VIEJO e FUENTES, 2020) (WILSON, 2012).

Os métodos que trabalham com aparência (cor ou textura) requerem ajustes e hardwares muito complexos e, com uso de ML, etapas que demandam muito tempo para construir a base de conhecimento (WANG, *et al.*, 2022). Isso se deve ao fato de ter que se construir uma base de conhecimento que tenha que trabalhar com os diversos modos diferentes de preparação do material, que influenciam no quesito textura (TCHUINKOU e BOBDA, 2018).

Por exemplo, uma carne pode adquirir textura e cores diferentes, de acordo com o processo de preparação, pois uma mesma carne se for cozida, ou assada, ou frita, vai apresentar diferentes tipos de textura, umidade ou cor do material, o que implica em uma etapa de preparação de base de conhecimento que vai precisar de numerosos tipos de amostras diferentes, em situações distintas de preparação, requerendo especificidades complexas de tempo (BRACCINI, *et al*, 2021),

Caso fosse utilizado essa abordagem (parâmetros de aparência), seria necessário trabalhar com diversas nuances e características bem distintas, aplicados aos mesmo parâmetro (cor e textura), para a identificação e classificação da matéria orgânica, necessitando-se de algoritmos mais complexos, principalmente no que tange ao aprendizado sobre as variações desses padrões (BRACCINI, *et al.,* 2021).

Em VASCONCELOS, *et al.* (2022) identifica-se uma forma possível de ser aplicada para atuar em classificação de alguns tipos de material orgânico, mais especificamente, os que podem ser usados como alimentos. Esse modo de identificação se utiliza da percepção dos gases emitidos pelo objeto examinado, durante seu processo de decomposição, procurando obter as difusões do metano (CH4), gás carbônico (CO2) e amônia (NH3) ao longo do processo, pois esses são os principais gases emitidos por matérias orgânicas de origem alimentar (BAJWA, *et al*., 2021).

No método de Vasconcelos, *et al.* (2022), há uma desvantagem relacionada a alguns materiais orgânicos com altas doses de conservantes, que prejudicariam a percepção sobre a emissões dos gases trabalhados no processo de decomposição, podendo gerar confusões sobre a detecção e classificação do mesmo, necessitando outras abordagens mais complexas e que fugiriam do escopo principal desse trabalho.

Para se utilizar esse método nessa pesquisa, houve necessidade de adaptações voltadas par que ele tratasse de matéria orgânica alimentar, pois o original foi concebido para trabalhar com a ideia de examinar a decomposição da matéria. Nessa pesquisa há etapas de menos tempo de duração para

identificação do tipo de material a ser verificado, pois o objetivo é classificar o objeto como matéria orgânica alimentar ou não alimentar.

Observando os métodos pesquisados e o objetivo desse trabalho, verificou-se que o método de obtenção de dados por reconhecimento da porção de gases emitidos foi o que melhor pode ser adaptado para se adequar na tarefa de classificação de material orgânico alimentar. Desse modo foi possível ajustar a captação de gases para que os materiais alimentares de origem orgânica pudessem ser distinguidos, evitando-se que outros materiais orgânicos, como papel, madeira etc, pudessem ser classificados como alimento. Observou-se também que esse método de classificação e reconhecimento de padrões foi verificado como o mais simples e econômico a ser aplicado, inclusive podendo usar dispositivos eletrônicos de baixo custo (BAJWA, *et al*., 2021).

1.2- Algoritmos de reconhecimento de padrões usados para identificar material orgânico

De acordo com KARAKAY, ULUCAN e TURKAN (2020), o reconhecimento de padrões é uma técnica que abrange a análise de um conjunto de dados por meio da verificação de tendências ou atributos distintos, permitindo a classificação e a categorização desses dados.

Em geral, o reconhecimento padrões realizado sobre uma base em dados obtidos de emissão de gases de objetos específicos são conhecidos como tipos de sistemas E-Nose (KARAKAYA, ULUCAN e TURKAN, 2020) (TAN e XU, 2020). Esses sistemas usam abordagens de reconhecimento de padrões baseadas em algoritmos de ML específicos (STOCHERO, JACOBI e LÚCIO, 2020).

Entre esses algoritmos usado em sistemas E-Nose, os vistos na Tabela 1 são os principais utilizados para esse fim e nesses diferentes tipos de aplicação

(VASCONCELOS, *et al*., 2022) (YE, LIU e LI, 2021). Esses algoritmos também podem, inclusive, serem usados para identificação de padrões de gases em outras situações, como, por exemplo em padrões de poluição atmosférica (CIDADE, 2020) (WILSON, 2012).

Esses tipos de algoritmos analisam os dados obtidos de uma matriz de sensores, visando identificar as regularidades ou semelhanças contidas nos dados, arranjando-os em classes distintas, de acordo com critérios pré-determinados (HASAN, SARNO e SABILA, 2020) (VIEJO e FUENTES, 2020) (EAMSA-ARD, *et al*., 2018). Entre os algoritmos examinados estão:

- PCA (Análise de Componentes Principais), que funciona como um algoritmo de aprendizado não supervisionado, sendo usado para examinar as inter-relações entre um conjunto de variáveis, reduzindo a dimensionalidade dos dados e, ao mesmo tempo, preservando suas variações mais importantes (PLUMBLEY e OJA, 2004);
- LDA (Análise Discriminante Linear), que é uma abordagem usada no aprendizado de máquina supervisionado para resolver problemas de classificação, separando as várias classes possíveis, visando a redução da dimensionalidade dos dados. (GHOSH. *et al*, 2020);
- SVM (Máquina de Vetores de Suporte) que é um algoritmo de aprendizado de máquina que usa o aprendizado supervisionado para classificar ou regredir dados, no qual se encontram hiperplanos, separando os pontos de dados em classes diferentes, maximizando a distância entre os pontos de dados mais próximos de cada classe. (YANG, *et al*, 2021); e
- ANN (Rede Neural Artificial), que se refere a um conjunto de algoritmos projetados para ajudar no reconhecimento padrões, sem serem explicitamente programados, consistindo em um grupo de nós interconectados, que representam os neurônios do cérebro

biológico, e auxiliam na classificação de padrões. (STOCHERO, JACOBI e LÚCIO, 2020).

- KNN (Vizinhos mais próximos), é um classificador de aprendizado supervisionado não paramétrico, que usa a proximidade para fazer classificações ou previsões sobre o agrupamento de um determinado ponto de dados. É comumente utilizado para sistemas de recomendação simples, reconhecimento de padrões, mineração de dados, previsões de mercados financeiros, detecção de intrusões e muito mais. (HALDER, *et al.*, 2024).

Cada um deles apresenta suas vantagens e desvantagens, conforme visto na Tabela 1 e a escolha do mais adequado para uma determinada aplicação baseia-se no tipo de categorização necessária ou em qual limite de classificação é viável para os objetivos do sistema a ser desenvolvido. No caso dessa pesquisa, verificou-se que o modelo deveria possuir boa precisão, ótima generalização e eficiência computacional, para que pudesse ter uma classificação dentro dos parâmetros esperados, possibilitando o nível de acertos entre 80% e 95% estabelecidos como meta.

Na busca do modelo ideal, baseando-se na pesquisa de VASCONCELOS, *et al*. (2022), optou-se pela construção de um modelo combinado as técnicas PCA e LDA para tratar do pré-processamento, e KNN e SVM para atuar na questão e organização da árvore de decisão, já a técnica ANN foi considerada muito complexa, e poderia elevar em muito o tempo para investigar os dados e fornecer os resultados.

Tabela 1 **-** Algoritmos utilizados em sistemas de reconhecimento de padrões tipo E-Nose, ou para reconhecimento e classificação de tipos de gases emitidos.

Tipo de Algoritmo	Vantagens	Desvantagens
PCA	- Eliminação de correlações. - Mede estimativas de probabilidade de dados de alta dimensão.	- Alta sensibilidade a valores discrepantes. - Suposição de linearidade. - Dependência da escala dos dados.
LDA	- Trabalha com dimensionalidades menores. - Melhora a separação de classes. - Eficiência computacional mais acurada. - Trabalha melhor com limites de decisão linear.	- Dependência de classes rotuladas. - Sensibilidade a valores discrepantes. - Restrição linear. - Requer um tempo computacional elevado.
SVM	- Eficaz em espaços grandes e com alta dimensionalidade. - Possui uma gerência mais eficiente da memória, - Bom desempenho em conjuntos de dados pequenos. - Alta capacidade de generalização	- Alta sensibilidade aos parâmetros de ajuste. - Alto tempo computacional ao lidar com grandes quantidades de dados. - Dificuldade para interpretar os resultados.
ANN	-Favorece uma classificação mais pormenorizada. - Pode operar com conhecimento incompleto. - Alta tolerância a falhas. - Depois de treinadas, as previsões são rápidas. - Muito utilizado em processamento paralelo.	- É difícil encontrar a estrutura de rede neural ideal. - A interoperalidade com outros algoritmos é muito limitada. - Resultados precisam de complexa estrutura de interpretação. - Pode ocorrer sobreposição de classes. - Eleva o tempo de classificação.
KNN	- É supervisionado e não paramétrico utilizando-se de proximidade nas classificações e nas previsões sobre o um determinado ponto de dados. - Aplicado para solucionar problemas de classificação;.	- Ele apenas armazena um conjunto de dados de treinamento, em vez de passar por uma fase de treinamento. - Depende bastante da memória para armazenar todos os seus dados de treinamento,

Fontes: (VASCONCELOS, *et al*., 2022), (BAJWA, *et al*., 2021) (TAN E XU, 2020) (HASAN, SARNO E SABILA, 2020) e (WILSON, 2012) (HALDER, *et al*., 2024).

1.3- Dispositivos de detecção utilizados em sistemas de captação de gases e similares

Entre os tipos de dispositivos utilizados em sistema de captação de cheiro ou gás, os de tipo E-Nose são os mais comuns. Neles encontram-se diversos tipos

de sensores e captadores de gás, tais como, sensores e transmissores (SANJAYA, *et al*., 2023) (MOHAMED, *et al.,* 2022).

1.3.1- Sensores usados em sistemas de captação de gases

Entre os tipos de sensores usados em sistema E-Nose, há os sensores que se utilizam da propriedade piezoelétrica do cristal de quartzo, nos quais o sinal é obtido através da variação de frequência gerada pela oscilação do cristal de quartzo em resposta a mudança de massa na superfície do sensor (JONCA, *et al*., 2022).

Nesse caso, a medida da frequência de ressonância se altera com a percepção de massa, induzida pela adsorção de gás na superfície do eletrodo modificado. Esses sensores são do tipo analógicos e são mais usados para medir as variações de compostos que emitem gás do tipo oxigênio e nitrogênio (JONCA *et al*., 2022).

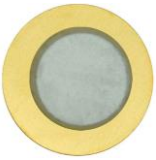
Há também os sensores químicos, que costuma se apresentar em três camadas diferentes, sendo uma para tratar de reconhecimento dos padrões, que contém o elemento químico que percebe o sinal; a de transdução, que transforma a informação química em sinais elétricos e a de propagação e transmissão, que amplifica o sinal e auxilia a outros captadores de processamento de sinais. Geralmente são utilizados para identificação de acetonas e similares (BHOWMIK, 2023).

Um outro tipo de sensores de detecção de gás, também utilizado em E-Noses, são os baseados no princípio de semicondutância ou MOS, no qual o material sensível a gás é um semicondutor MOS com baixa condutividade. Para esses tipos de sensores, quando há detecção de algum gás, a condutividade aumenta há alteração na condutividade, que é transformada em sinal de saída, correspondente à concentração de gás (VA, SUBRAMONIAM e MATHEW, 2021) (ABBAS, *et al*, 2020).

Os sensores de gás baseados em semicondutância são os de mais baixo

custo, mas aprestam a desvantagem de serem muito sensíveis a variações de tensão, podendo sofrer danos, caso as configurações elétricas/eletrônicas tenham muitas alterações (ABBAS, *et al*, 2020). Apesar dessa desvantagem, como foi dito, esses sensores são os mais baratos e são usados em diversos tipos de aplicações E-Nose, com um grau de segurança de uso contínuo aceitável, isto é, com um tempo médio entre falhas de cerca de 1488 horas (RAHMAN, *et al.*, 2024) (MOHAMED, *et al.*, 2022). A Tabela 2 apresenta os tipos de sensores mais sutilizados e suas características gerais.

Tabela 2- Sensores utilizados em sistemas do tipo E-Nose.

Sensores	**Modelo**	**Características**
Piezoelétricos		- Usa conversão de energia, na qual uma força é aplicada para gerar uma tensão associada. - É usado no setor de motores para detectar óxidos de nitrogênio, monóxido de carbono e hidrocarbonetos.
Químicos		- São os mais precisos e sensíveis. - São usados nos setores médicos, farmacêuticos e de cosméticos, no monitoramento ambiental e para detectar gases inflamáveis, como sulfeto de hidrogênio, cloro, dióxido de enxofre e etanol.
MOS		- Foram desenvolvidos para monitorar a qualidade do ar e gases tóxicos, como CO, NO_2 e NH_3. - São usados para sistemas de identificação associados a matéria orgânica - São os de menor custo.

Fonte: Próprio autor

1.3.2 - Outros componentes dos dispositivos de identificação de matéria orgânica

Um outro componente existente em sistemas E-Nose são os microcontroladores, que são utilizados para atuar como a gerência sobre os sensores, dispositivos de comunicação e, dependendo do potencial de memória de programação, também sobre parte da análise dos dados (BIN HAFEEZ, *et al*, 2019).

Entre os microcontroladores de baixo custo, existem os AVRs. Os AVRs são, atualmente, desenvolvidos pela empresa Microchip Technology, possuindo uma arquitetura de do tipo RISC (Reduced Instruction Set Computer), associada a uma arquitetura Harvard, no qual pode-se separar barramentos de dados e de instruções em duas linhas de comunicações distintas, desta forma, permitindo que o microprocessador execute instruções em todo ciclo de clock (BARRET e PACK. 2019). As principais características da arquitetura AVR são :

- Usam memória flash para armazenar a programação.
- Têm uma estrutura RISC, com um pouco mais de uma centena de instruções.
- Trabalham com operandos de oito bits.
- Possuem interface com diversos tipos de dispositivos (ISP, CAN, USART entre outros)

Contudo, por conta da limitação na sua memória de programa (cerca de um a 2KB), e por usar um pipeline simples de cinco estágios, o AVR possui a desvantagem de não poder processar certos algoritmos que demandem um processador de mais um núcleo básico, tendo, contudo, uma velocidade de processamento de 12MIPS (KNOWN, *et al.*, 2021). A principal plataforma que trabalha com esse tipo de microcontrolador é a família Arduino R3, que é uma interface de desenvolvimento para aplicações de controle em tempo real, que demandem um clock entre oito e vinte MHz (SINGH e GAUR, 2023).

Outra família de microcontroladores utilizadas em sistemas E-Nose são os ARMs. Os microcontroladores dessa família são mais robustos que os da

AVRs, sendo que alguns deles podem trabalhar como processadores comuns, ao invés de somente ser um microcontrolador (LI. *et al.*, 2023). Os microcontroladores comuns costuma trabalha com um processo por vez, usando mecanismo de interrupção, quando sensores ou dispositivos conectados precisam se comunicar. Os ARMs, além desse tipo de funcionamento, possuem multi-núcleos, permitindo, dependendo da montagem de sua interface, que mais de um processo possa funcionar ao mesmo tempo, guardando características de multiprocessamento existentes em microprocessadores modernos. (PANDURANGI, *et al*., 2021),

Entre as caraterísticas dos microcontroladores ARMs estão (ORĂȘAN, SEICULESCU e CĂLEANU, 2022):

- Baixo consumo de energia.
- São opções consistentes para sistemas que exigem muitas operações recorrentes executadas por longos períodos.
- Trabalham com instruções RISC, consumindo menos por tarefa executada, sendo utilizados em ambientes que demandem tempo de processo rápido, com baixo consumo de tensão.
- São aplicados em smartphones, tablets e dispositivos IoT.

A interface mais comum para uso de ARMs é a da família xPi, tendo como exemplo a placa de interface com ARM RaspberryPi (PALAPARTHY, *et al*., 2020). Esse tipo de placa possui memória RAM embutida, processador ARM do tipo Cortex M4, interface para monitores e GPIO (pinos de entrada e saída) de dez pinos, assim como conectores para USB, HDMI e micro SD (PALAPARTHY, *et al*., 2020).

As placas ARMs são mais caras que as AVRs, contudo, atualmente a família Arduino possui uma nova versão interfaces, que são as Arduino R4, que juntam a facilidade e o baixo custo proporcional das placas Arduinos R3 com a capacidade de maior processamento dos microcontroladores ARMs (ARDUINO-CC, 2024). Com o aparecimento dessa nova plataforma com ARM, que é o Arduino R4, os custos de acesso a esse tipo de microcontrolador reduziram-se

substancialmente, possibilitando o uso dessa interface como a principal do sistema a ser desenvolvido.

1.4- Algumas outras aplicações de Sistemas E-Nose

1.4.1- Aplicações na indústria

As primeiras aplicações dos sistemas E-Nose foram nos setores de cosméticos e perfumes (PERSAUD e DODD, 1982). Entretanto, outras aplicações críticas para esses sistemas surgiram com novos hardwares e softwares, atuando em diversas outras aplicações industriais (JOÑCA, *et al.,* 2022) (LI, *et al*., 2021). As aplicações industriais usam sensores piezoeléctricos adaptados para trabalhar com sistemas E-Nose, como, por exemplo, no setor automotivo.

Nesse tipo de aplicação, eles servem para testar as emissões de poluentes dos motores de combustíveis fósseis. Nesse caso, os sistemas E-Nose são usados para medir a quantidade de poluente emitida por um motor para que ele possa ser melhorado, caso exceda o limite aceito para veículos móveis. A Figura 1 mostra um exemplo de um sistema E-Nose para o setor automotivo (TAN e XU, 2022). Esse tipo de sistema é complexo, exigindo conexões própria para comunicação com outros dispositivos, e utiliza algoritmos do tipo ANN. A segunda geração desse modelo também usa sensores químicos, que são mais caros, mas muito mais precisos (TAN e XU, 2022).

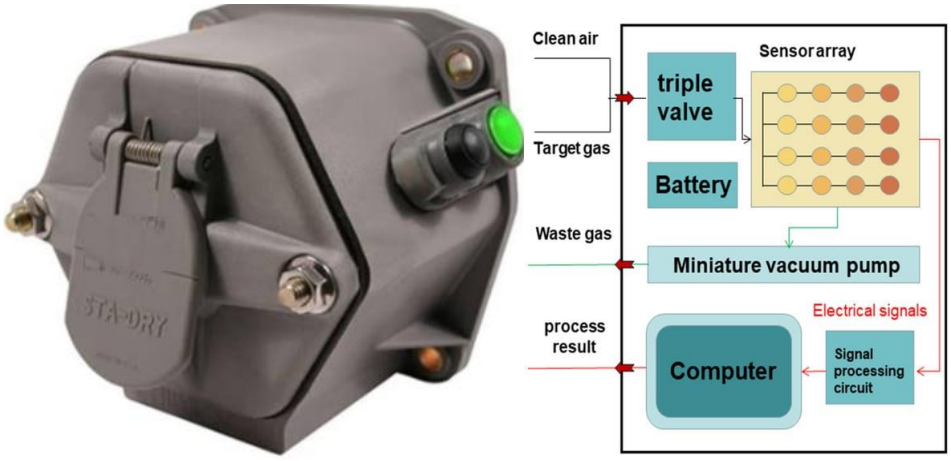
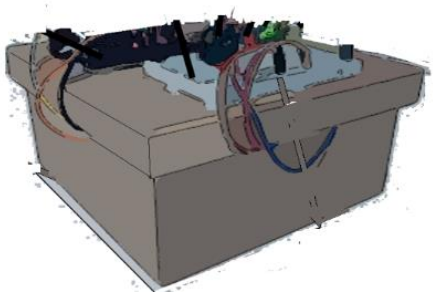


Figura 1 - Um exemplo de E-Nose usado em indústrias automobilísticas, com seu diagrama visto em https://phillipsind.com/ e FABER, *et al*. (2013).

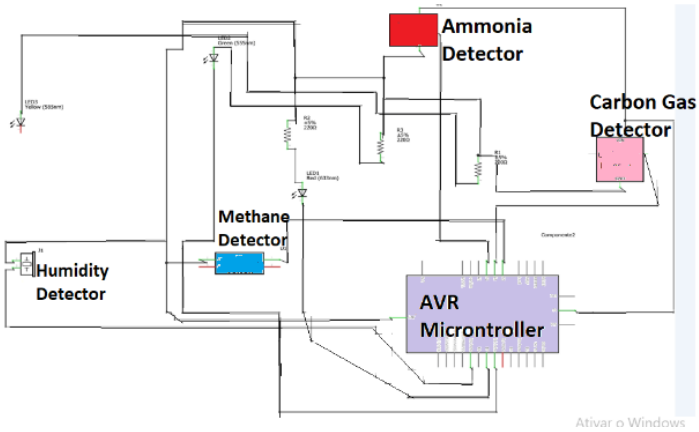
1.4.2 Aplicações para sustentabilidade

Em VASCONCELOS, *et al*. (2022) é apresentado uma aplicação de sistemas E-Nose, de baixo custo, com sensores MOS, associados a um modelo de software que utiliza uma combinação de algoritmos PCA e LDA, para identificar se um produto alimentício é próprio para consumo ou não, sendo isso uma aplicação que se utiliza das normas da ODS nº 11 e 12 da ONU.

Esse sistema é visto como capaz de ser de fácil transporte e útil para ajudar pessoas idosas e deficientes visuais a comprar legumes e frutas. Por meio desse sistema E-Nose, é possível elas saberem, se o produto é próprio para consumo. A Figura 2 mostra esse sistema E-nose e seu esquema eletrônico. Essa é uma das aplicações possíveis para E-Nose no item de sustentabilidade, sendo uma aplicação importante para que não haja desperdício alimentar ou carga excessiva de lixo orgânico.



a)



b)

Figura 2- Dispositivo E-nose para identificação de fase de decomposição de um alimento (a) e seu diagrama eletrônico (b). Fonte: VASCONCELOS, *et al.*, 2022.

A Figura 3 apresenta um exemplo de módulo de software dessa aplicação, que se faz uso de uma mescla de algoritmo LDA com PCA visando trabalhar na construção do módulo de treinamento e identificação usado para verificar material do tipo orgânico em sistema E-Nose (VASCONCELOS ,*et al.*, 2022).

```python
@app.route('/treinar/<string:metodo>', methods=['POST'])
def treinar(metodo):
    try:
        dados = request.get_json()
        df = pd.DataFrame(dados) # Converter JSON para DataFrame

        # Especifica as features (X) e o target (y)
        X = df.drop('target', axis=1)
        y = df['target']

        if metodo == 'combinado':
            resultado = treinamento_combinado(X, y)
        else:
            resultado = treinamento(X, y, metodo)

        return jsonify(resultado)
```

Figura 3- Exemplo de código para o modulo de treinamento do Sistema E-Nose visto em VASCONCELOS, *et al.* (2022).

1.4.3 Aplicação utilizada em sistemas de controle atmosférico

Outra aplicação na qual sistemas E-Nose têm sido usados é na área de controle ambiental. Na década de 1990, os sistemas E-Nose foram criados para controlar a poluição atmosférica. Esses sistemas geralmente usavam sensores piezoeléctricos, que se mostraram mais adequados para medir a poluição (JOÑCA, *et al.*, 2022). O uso do E-Nose para medir a qualidade do ar pode ser visto em KUMAR, *et al.* (2023), no qual um sistema baseado em microcontrolador ARM com sensores químicos é usado para avaliar a qualidade do ar em um ambiente de condomínio, verificando a concentração de dióxido de carbono e dióxido de nitrogênio.

Outro sistema para medir as condições atmosféricas, usando um dispositivo ARM IoT e sensores MOS, foi usado para medir a concentração de gases nocivos em um ambiente interno e, ao mesmo tempo, a parte que escaparia para a atmosfera livre (CIDADE, 2022). Ambos os sistemas utilizaram algoritmos ANN para seus módulos de treinamento e classificação. Um exemplo de um

gráfico de dispersão obtido de um sistema E-Nose instalado em um UAV pode ser visto na Figura 4.

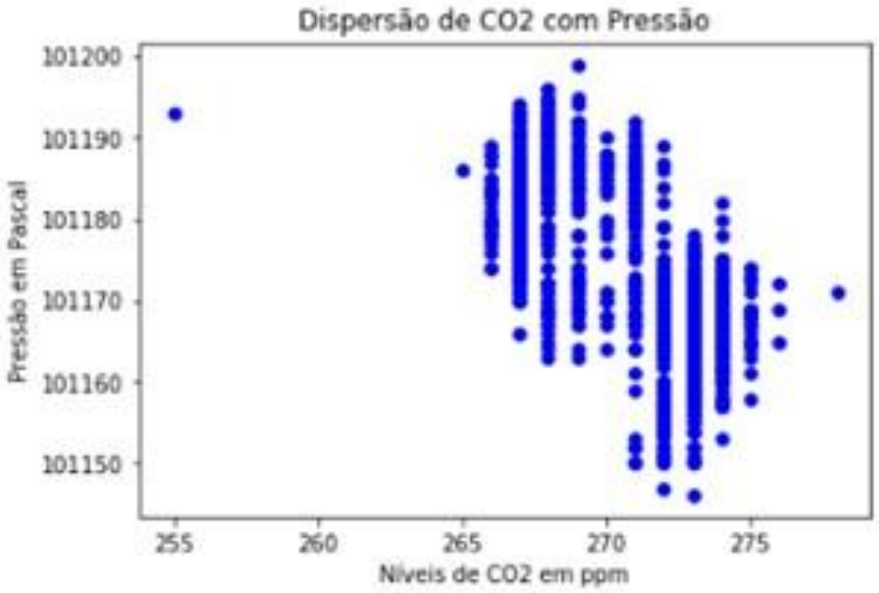


Figura 4- Dispersão de CO2 pela pressão atmosférica, obtida de um UAV a 30 m de altitude em um ambiente aberto próximo a fábricas.
Fonte: CIDADE, 2022.

2- Métodos e Materiais

2.1 Base de conhecimento

Para haver o reconhecimento de padrões que identificam a matéria orgânica alimentar, precisou-se montar a construção de uma base de conhecimento que pudesse dar captar valores indicativos das características dos materiais, visando a maior precisão possível na identificação de matéria orgânica alimentar. Neste trabalho, essa base de conhecimento foi confeccionada a partir das diversas amostragens obtidas sobre diversos tipos de matérias orgânicas alimentar, com vista a poder estabelecer o padrão básico relativo a emissão de gases desse tipo de material.

O foco foi na obtenção de valores sobre os gases mais comuns emitidos em materiais orgânicos alimentares, que são, em geral, o CO_2 (dióxido de carbono), CH_4 (metano) e NH_3 (amônia), conforme visto em CONCEIÇÃO, *et al*. (2021) e em STRIBIŢCAIA, *et al.* (2020). Nos experimentos de obtenção de dados, foram utilizadas quarenta amostras. Essas amostras se compunham de material relativo a dois tipos de frutas, um tipo de vegetal comestível, um tipo de carne e amostras de material alimentar misturado.

A meta foi verificar os sinais indicativos de emissão de gases de tecidos orgânicos, além de indicações da decomposição, putrefação ou dissociação do alimento, associando-os a características similares de materiais do tipo orgânicos (VASCONCELOS, *et al,* 2022). Os dados obtidos sobre emissão de gases de CO2, CH4 e NH3 de cada uma das amostras foram planilhados para compor a base de dados a ser utilizado na comparação com outros materiais, formando assim a base de conhecimento inteligente, usada para identificação de um objeto como material orgânico alimentar ou não.

Com base nos experimentos vistos em VASCONCELOS, *et al.* (2022), e MUDGULKAR (2024), o tempo médio para a captação de dados para criação da base de conhecimento, utilizando as quarenta amostras para os experimentos de construção da base de conhecimento, foi de no máximo 20 minutos por amostra, que é o tempo médio no qual uma pequena porção de matéria orgânica começa a emitir sinais de putrefação.

Nesse caso, foi feita uma obtenção de dados, para cada amostra, sendo as captações obtidas de dez em fez segundos, totalizando seis captações/min, por material orgânico alimentar, durante 20 minutos, totalizando 120 medições. Esse procedimento, para cada objeto de amostra, foi repetido cinco vezes, cada dessas medições também com 20 minutos, gerando um total de 600 medições por objeto de amostra.

Sendo gerado seis captações por minuto, com 20 minutos de exposição dos objetos das amostras, repetindo esse procedimento cinco vezes para cada objeto da amostra, e usando os 40 objetos de amostras da Tabela 2, gerou-se cerca de 24.000 tuplas de dados para formação da base de conhecimento.

O tempo de 20 minutos foi observado como sendo um tempo limite de pequenas amostras entrarem em estado total de putrefação, sinalizando o término de vida útil do material, fazendo com que os dados comecem a estar fora da escala pretendida para o sistema desse trabalho, já que o dispositivo se propõe a identificar matéria orgânica alimentar ainda dentro da vida útil da mesma. Na Tabela 3 é possível observar a quantidade detalhada de cada tipo de amostra, com cada tipo de alimento, que foram as usadas para a trabalhar e construir a base de conhecimento que foi aplicada no trabalho.

A divisão de alimentos foi baseada na classificação dos mesmos entre climatérios e não climatérios, carnes e amostras misturadas. Os climatérios, que são aqueles que depois de colhidos, apresentam uma elevação na taxa respiratória, e uma produção de um gás, chamado etileno. Devido à produção desse gás, conseguem amadurecer transformando, por exemplo, o sabor adstringente da banana verde em doce (BRON e JACOMINO, 2007).

Os alimentos não-climatérios são aqueles que param de amadurecer após a colheita, e são identificados em material orgânico associado ao limão, laranja, morango, uva, framboesa, amora-preta e demais tipos de frutas consideradas cítricas ou vermelhas (FUKANO e TACHIKI, 2021).

As carnes são tecidos de origem animal, que podem ser utilizados como alimentos, em sua maioria, que possuem um tempo de putrefação mais rápido, devido a atuação de bactérias típicas (EDITA, *et al*, 2018). Além disso, foram inseridas cinco amostras de comida pré-confeccionada, com 100 gramas de peso, com mistura de alimentos, sendo estas, duas amostras de arroz com feijão, duas de purê de batata com um molho de tomate e uma de espinafre cozido com carne moída.

Tabela 3- Divisão das amostras de cada tipo de alimento utilizado nessa fase da pesquisa.

Material orgânico para alimentos	**Tipo de alimento**	**Número de amostras**
Banana	Climatério	10
Limão	Não climático	9
Frango	Carne	4
Cenoura	Climatério	12
Amostras de comida	Mistura	5

Fonte: Próprio autor

2.2- Modelagem do hardware e repositório

Para poder obter os sinais relativos aos gases de CH4, NH3 e CO2, foram utilizados sensores MOS nos modelos MQ137, MQ135 e MQ4, respectivamente (BANZI E SHILOH, 2015). Esses sensores, conforme vistos no item I.3.1 desse trabalho, são os mais fáceis de manipular e que tem menor custo de acesso. Além disso, em VASCONCELOS, *et al.* (2022), observa-se uma interface entre esses sensores e um dispositivo de microcontrole baseado na interface Arduino,

que foi utilizada nessa pesquisa, com adaptações, facilitando o uso dessa nova plataforma.

O controle da captação dos gases foi realizado por uma interface Arduino, que, no caso dessa pesquisa, foi um Arduino R4, que possui um microcontrolador da família ARM, modelo ARM Cortex-M4, que opera a uma velocidade de até 48 MHz, proporcionando desempenho superior em comparação com as versões anteriores do Arduino Uno, contendo também 256 KB de memória flash, 32 KB de SRAM e 8 KB de EEPROM, (MAKERHERO, 2024).

A escolha de uma interface Arduino com ARM pode ser justificada com os argumentos vistos no item I.3.2 dessa pesquisa. No caso, para que o custo do projeto não ficasse muito expansivo, optou-se por uma placa Arduíno R4 com configuração mínima (sem Wi-Fi), pois a placa com Wi-Fi ainda se apresenta com um custo mais elevado (MAKERHERO,2024). Para comunicação entre a plataforma de controle (com os sensores acoplados) com um dispositivo móvel qualquer, optou-se por um dispositivo Bluetooth HC-05, que tem baixo custo e permite comunicação full-duplex (BANZI E SHILOH, 2015).

Com base no trabalho de VASCONCELOS, *et al*. (2022), verificou-se qual o mais adequado tipo de material para confeccionar o invólucro ou repositório, que estivesse mais associado ao objetivo desse trabalho, que foi deduzido como sendo um repositório de polietileno, que é flexível, leve e resistente a impactos e a produtos químicos, sendo fácil de limpar (FERREIRA, SILVA e MADEIRA, 2019). Um modelo de involucro usado pode ser visto na Figura 5.



Figura 5- Involucro de polietileno utilizado para realizar as medições dessa pesquisa.
Fonte: Próprio autor.

A escolha deste material também é por ele compor a maioria dos repositórios de lixo de escolas e lugares público, que são o destino principal da plataforma, para identificação de material orgânico alimentar. Conforme visto em VASCONCELOS, *et al.* (2022), há algumas restrições para o uso continuo desse material. Entre essas limitações são que o uso constante pode “contaminar” as paredes internas, causado erros de leitura, sendo que, para evitar essa situação, acrescenta-se sacos de lixo, de plástico, para colocar as amostras.

Outra limitação é que, independente de não haver contaminação com as amostras anteriores, o repositório deve ser constantemente limpo, o que pode causar problemas quanto ao manuseio dos circuitos inseridos na tampa do repositório. Mesmo assim, esse foi o repositório considerado mais adequado para essa plataforma. Com relação a medição, que faz uso da base de conhecimento gerada, para determinar se o material é orgânico alimentar, ou não, percebeu-se que o tempo de amostragem não deve ir além de três minutos, sendo feita com medições de dez em dez segundos. Esse tempo de três minutos, é devido ao objetivo de uso real da plataforma, pois em ambientes de escolas e shopping, até três minutos é o suficiente para que nova amostra seja colocada na mesmo dispositivo.

Mais de três minutos poderia gerar dados compatíveis com decomposição, que podem altera-se como alimentar orgânico, principalmente bebidas baseadas em carbono (refrigerantes), que tendem a perder o carbono quanto maior for o tempo de exposição ao ar. Os dispositivos eletrônicos foram acoplados ao invólucro em no lado interno da tampa do repositório, conforme visto no esquema da Figura 6 e na colocação real na Figura 7.

Como é visto, foi utilizado um protoboard de 100 pontos para as conexões e para alocação dos dispositivos de gás e do Bluetooth, o Arduino foi inserido ao lado do protoboard, assim como a bateria de 9V, tanto para a construção da base, quanto nos testes de medição. Um protoboard de 400 pontos pode ser utilizado, de forma ocasional, para alocar os componentes eletrônicos, de modo que possam ficar mais firmes na caixa.

Na Figura 8 é apresentado o esquema de ligação elétrica dos componentes eletrônicos com o Arduino, mostrando as ligações entre esses componentes, que foram utilizados para a construção da parte de hardware da plataforma, incluindo-se também a ligação de fonte de força com a plataforma.

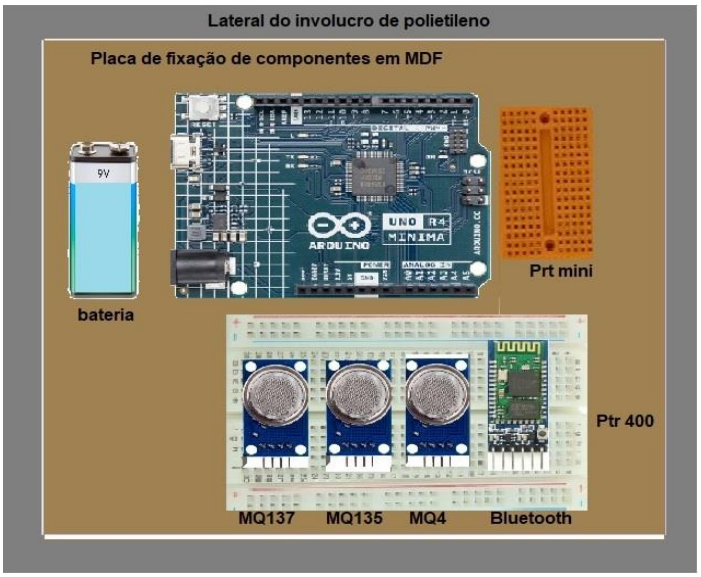


Figura 6- Modelo da disposição dos componentes de hardware da plataforma, no qual se visualiza a posição mais adequada desses componentes no suporte de fixação do circuito. Fonte: Próprio autor.

a) 

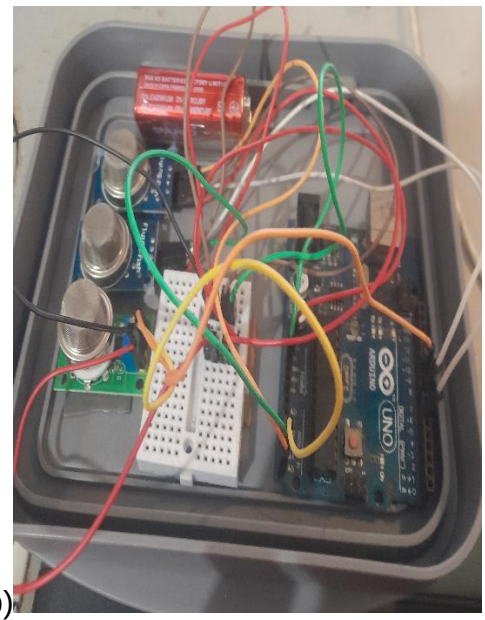
b) 

Figura 7- Alocação dos componentes no dispositivo de polietileno, com o Bluetooth no lado inverso da tampa, sendo (a) alocação básica e (b) com fios de ligação.
Fonte: Próprio autor.

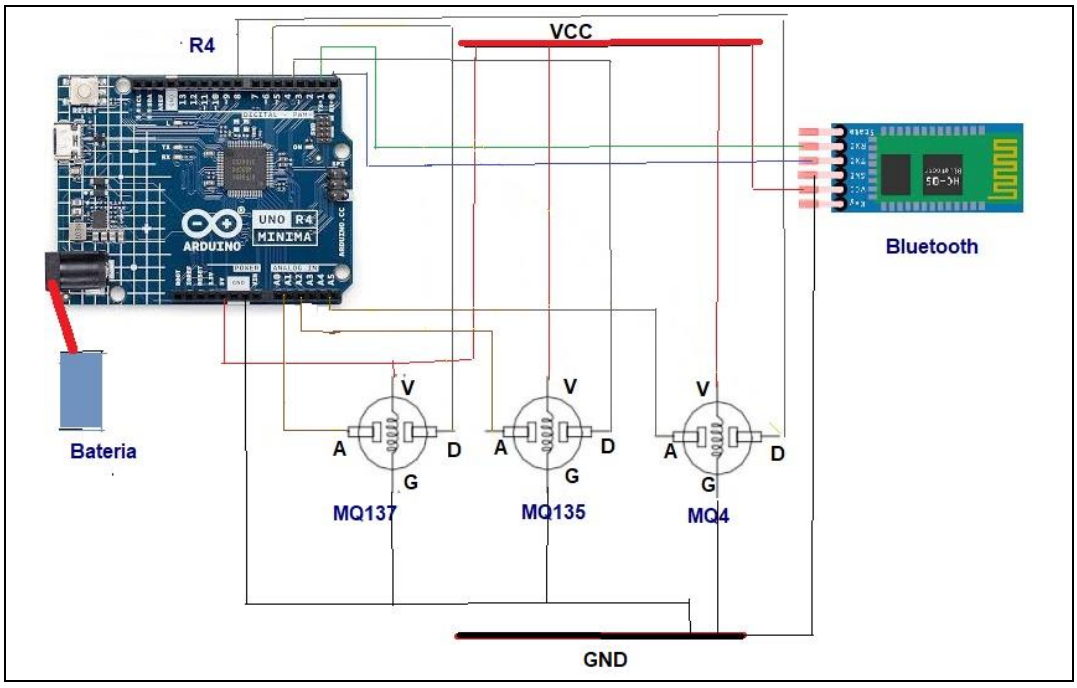


Figura 8- Esquema elétrico/eletrônico dos dispositivos de hardware da plataforma, mostrando suas ligações com a interface Arduino.
Fonte: Próprio autor.

2.3- Modelagem do Software

O software da plataforma possui dois módulos básicos, o principal, que é a parte de ML, no qual a análise dos dados, utilizando a base de conhecimento, é realizada, e o módulo de comunicação, que é o aplicativo a ser instalado em dispositivos móveis, com o objetivo de interfacear com o usuário final e fornecer informações sobre se o objeto no invólucro é um alimento orgânico alimentar ou não.

A principal função do modelo de ML é aprender com os dados e usar o conhecimento adquirido para fazer previsões ou tomar decisões. Após analisar várias técnicas, constatou-se que o modelo deveria ter alta precisão, ótima generalização e eficiência computacional. Para tanto o módulo principal da plataforma foi concebido dentro de uma combinação de técnicas de três algoritmos vistos no item 1,2 dessa pesquisa, o PCA (Análise de Componentes Principais), o LDA (Análise Discriminante Linear) e o SVM (Máquina de Vetores de Suporte).

Com a análise das técnicas vista no item I.2, foi aplicado um modelo para se obter uma precisão acurada, com ótima generalização e eficiência computacional. Para isso, esse modelo final foi baseado em uma combinação das técnicas PCA, LDA e SVM. Com o PCA é possível reduzir a dimensionalidade dos dados, preservando a maior parte da variabilidade. No LDA, pode-se maximizar a separação entre classes nos dados, obtendo a máxima distância entre as médias das classes e a mínima dispersão intraclasse. O SVM foi usado para a classificação, usando-se uma ideia de hiperplano, maximizando a margem de reconhecimento entre elas.

O design do aplicativo móvel foi realizado usando-se uma arquitetura de microsserviços (WASEEM, LIANG e SHAHIN, 2020). Esse aplicativo é a interface gráfica do usuário, sendo seus eventos de controle:

- permitir que o usuário se conecte ou desconecte do dispositivo eletrônico e

- realizar uma ou mais verificações do material colocado na caixa após a conexão com o dispositivo.

A Figura 9 mostra a arquitetura usada pelo sistema de software como um todo, tanto a parte usada para criar a base de conhecimento e testar a amostra, quanto a parte de conexão e identificação, confeccionada para o dispositivo móvel, que irá avisar ao usuário se o alimento é orgânico alimentar ou não.

Cada um desses microsserviços foi implementado aplicando a tecnologia mais aderente ao caso. No caso dessa pesquisa, os microsserviços foram confeccionados com o *Java Spring Boot,* associado ao Maven (BOAGLIO, 2018). Maven é uma ferramenta usada para gerenciar a construção e as dependências em projetos Java O *Java Spring Boot*, por sua vez, é uma extensão do *Spring Framework,* que é específico para a criação de aplicações Java de microsserviços (BOAGLIO, 2018).

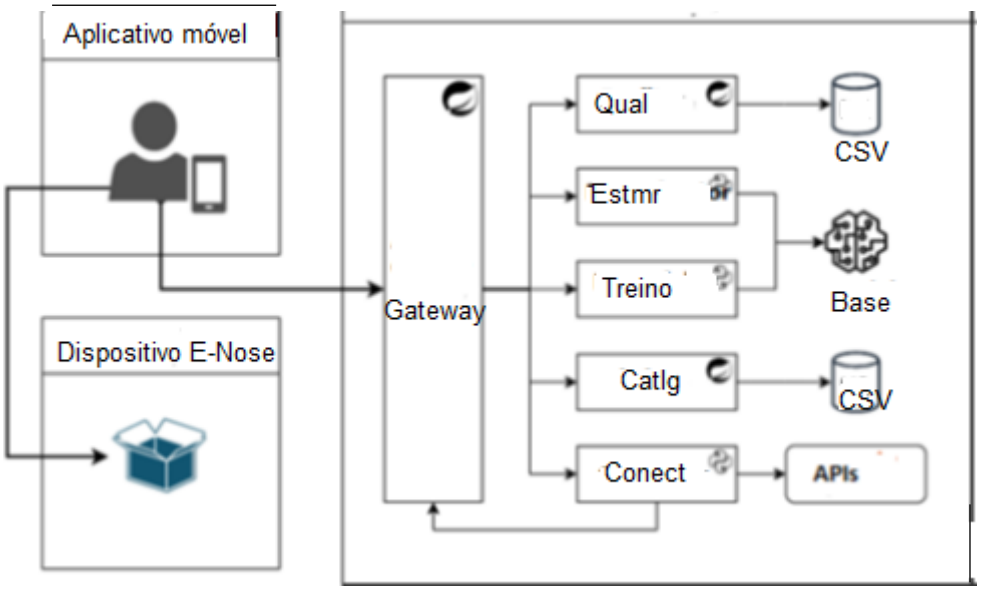
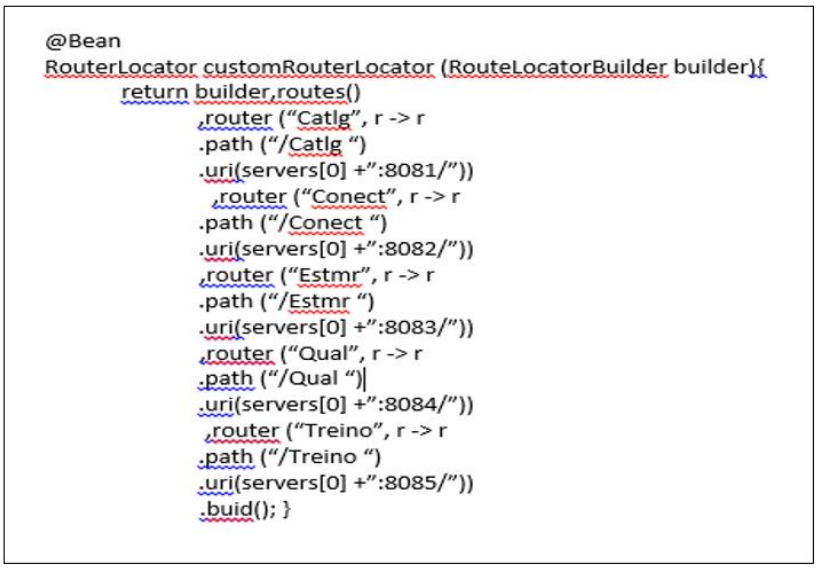


Figura 9- Arquitetura do sistema, com a caixa com sensores de gás, o aplicativo móvel, os microsserviços de qualidade (com os dados CSv integrados garantidos); o estimador (que lida com estatísticas de dados); os serviços para treinamento de ML e classificação da base de dados (formato CSV) e o serviço de conexão. Fonte: Adaptação do próprio autor da figura baseada em VASCONCELOS, *et al*. (2022).

Esse modelo de construção dos microsserviço foi porque ele consegue tratar do paralelismo das operações dos microsserviços, também facilitando o desenvolvimento ágil, com entrega de novos recursos rápidos. Além disso, com essa metodologia, pode-se focar mais no tratamento de falhas e independência dos microsserviços.

A comunicação entre os microsserviços foi através de *Application Interfaces* (APIs) do modelo *Gateway*, que servem para simplificar a comunicação da aplicação principal com os microsserviços e entre eles distribuídos, de modo a evitar solicitações individualizadas de serviços, o que demandaria mais tempo para executá-las (HECKLER, 2021). A Figura 10 apresenta um trecho de código Java de implementação de um dos microsserviços, que opera através da porta 8080 em diante, e estabelece as rotas dos demais serviços, associando os microsserviços de:

- catálogo (*Catlg*) à porta 8081;
- conexão (*Conect*) à porta 8082;
- para estimar valores (*Estmr*) à porta 8083;
- para tratar da qualidade (*Qual*) à porta 8084; e
- para treinamento (*Treino*) à porta 8085.



```
@Bean
RouterLocator customRouterLocator (RouteLocatorBuilder builder){
        return builder.routes()
                .router ("Catlg", r -> r
                .path ("/Catlg ")
                .uri(servers[0] +":8081/"))
                 .router ("Conect", r -> r
                .path ("/Conect ")
                .uri(servers[0] +":8082/"))
                .router ("Estmr", r -> r
                .path ("/Estmr ")
                .uri(servers[0] +":8083/"))
                .router ("Qual", r -> r
                .path ("/Qual ")
                .uri(servers[0] +":8084/"))
                 .router ("Treino", r -> r
                .path ("/Treino ")
                .uri(servers[0] +":8085/"))
                .buid(); }
```

Figura 10- Trecho de código associado aos microsserviços utilizados, que trata da configuração das portas de trabalho e da construção dos caminhos dos mirosserserviços *Catlg, Conect, Estmr, Qual* e *Treino*, usados pelo aplicativo.Fonte: Próprio autor

Todos os microsserviços (*Catlg, Conect, Estmr, Qual* e *Treino*) foram implementados em Java com o *Spring Framework,* sendo que o:

- *Catlg* possui a responsabilidade de gerenciar as sugestões ou dicas disponibilizadas no aplicativo mobile, que são aqueles que informam que o material é orgânico alimentar ou não.
- *Qual* realiza os tratamentos e validações nos dados obtidos, através da matriz de sensores, visando tratar de inconsistências e problemas com os mesmos.
- *Conect* é responsável por iniciar o processamento dos dados, através da interface *Apache Spark*, que gera uma base de dados não-SQL, ou no estilo CSV, podendo trabalhar com até 100.000 tuplas distintas, gerando a base de dados obtida pelos sensores.
- *Treino* recebe os dados do *Conect* e é responsável pelos modelos de treinamento e ML utilizados. Esse serviço vai retornar as métricas de avaliação e treinamento, apresentando erros anotados. A execução do *Treino*, para geração da base de conhecimento, foi configurada com o parâmetro tempo reduzido, e serviu para processar 8.000 registros no formato JSON de forma rápida (WASEEM, LIANG e SHAHIN, 2020). A Figura 11 apresenta uma parte do código relativo ao microsserviço *Treino* para gerar a tabela com registro JSON.
- *Estmr* recebe os dados, aplica sobre eles um modelo previamente salvo e retornar com a predição. O retorno deste serviço será “0” para alimentos orgânicos e “1”para outros tipos de matéria, que não se encaixem na base de conhecimento, e, por isso, seriam vistos como não alimento.

Cada um dos microsserviços tem sua função específica, não existindo acoplamento entre eles, sendo que cada um deles possui sua própria conexão com o Gateway, A conexão deles com a plataforma é realizada por meio de APIs. Sendo que os microsserviços *Catlg* e *Etmr* possuem processamento síncrono e dedicado, pois são microsserviços de mais urgência para a plataforma, e que precisam ser executados de forma imediata.

```
strcpy ("/Treino/dados",str1);
jsonify treino(str1, dados, tipo)
{
  rqst = request.get_json();
  dados=rqst;
  search(str1, dados);
  df =make.dataFrame(dados);
  x = df.drop('alvo', 1);
  y = df['alvo'];
 if (tipo == 'post)
    {tupla = combinar-Treino (x,y);}
 return jsonify(tupla);
}
```

Figura 11- Parte do código do microsserviço *Treino* , para tratar da geração de tuplas CSV-JSON, que serão inseridas na grande planilha, que constituirá parte da base de conhecimento.
Fonte: Próprio autor.

2.3.1- Modelagem do aplicativo

O módulo de aplicativo, que visa ser instalado em dispositivos móveis, foi construído com base em uma linguagem visual, para dispositivos com Android, visto que esses são os dispositivos ainda mais numerosos no mercado, usando a linguagem APP Inventor do Massachusetts Institute of Tecnology (MIT) (FAKHRURRIANA, 2023). Esse módulo possui um comando para conexão com dispositivos Bluetooth, de modo que o aplicativo possa utilizar-se da tecnologia Bluetooth já existente em todos os dispositivos móveis atuais (). Um exemplo dessa codificação pode ser visto na Figura 12, no qual se observa algumas das particularidades desse tipo de programação.

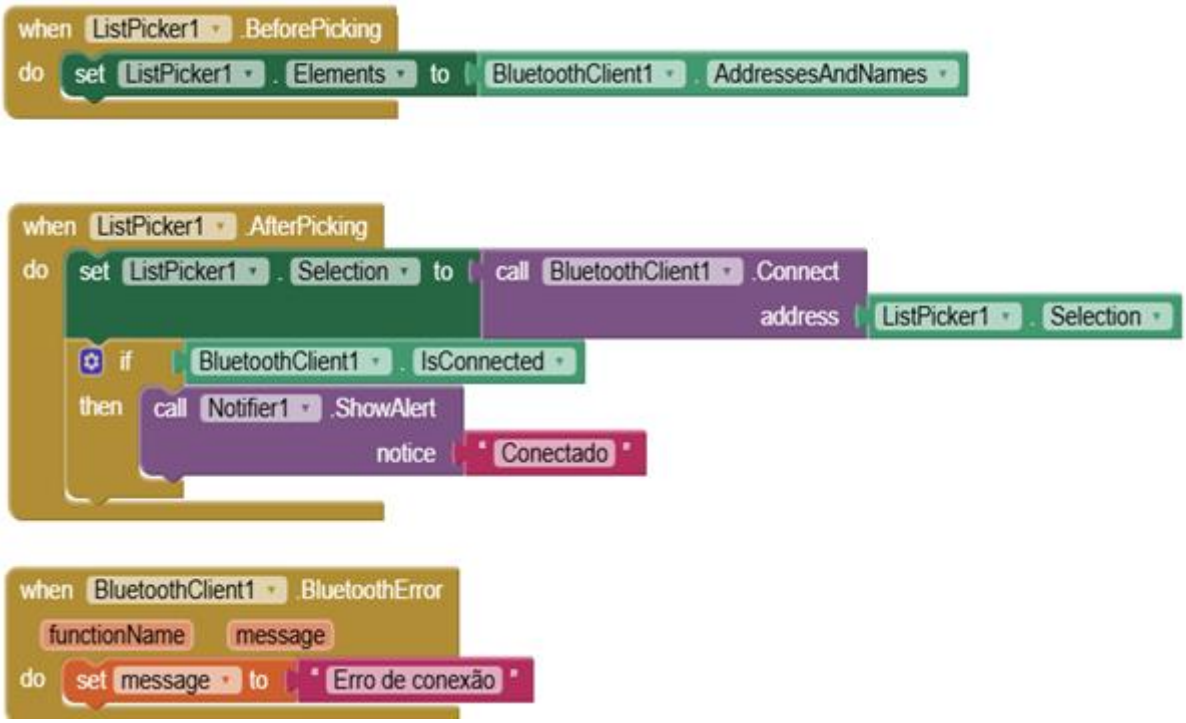


Figura 12- Módulo da programação no MIT App Inventor, para o dispositivo móvel, sendo que esse módulo trata das ações para implementar a conexão do dispositivo móvel com a plataforma, através do protocolo Bluetooth. Fonte: Próprio Autor

Nesse tipo de programação é possível acessar os módulos do *Estmr* e *Catlg*, feitos em Java, pois esse tipo de programação permite chamadas à aplicações (tipo APIs) feitas em outras linguagens. Como ambos microsserviços são sincronizados, as chamadas as funções de ambos podem ser feitas ao longo do aplicativo configurado no MIT App Inventor.

Na programação desse módulo, o *Conect* providencia o acesso ao arquivo CSV, que contém, os dados da base de conhecimento, separados por tipo de material orgânico, no qual, ao apertar o botão, os dados recebidos pelo Bluetooth são comparados dentro desse arquivo, resultando na classificação possível. Se os dados estiverem contidos nas pré-classificações de tipo climatérico, não climatérico, carnes ou mistura, a amostra é classificada como orgânico alimentar, se não for possível classificá-la dentro desses tipos, ele será classificado como não alimento.

Nos microsserviços no qual o aplicativo se apoia, E*stmr e Catlg,*, os dados recebidos dos sensores) são aplicados à base de conhecimento em CSV,

administrada pelo microsserviço *Conect,* que vai retornar com o resultado da classificação. O retorno deste serviço será “0” para amostra orgânico alimentar e “1” para amostras ditas como não alimentos (mesmo se for orgânica). Um exemplo da tela básica apresentada pode ser visto na Figura 13.

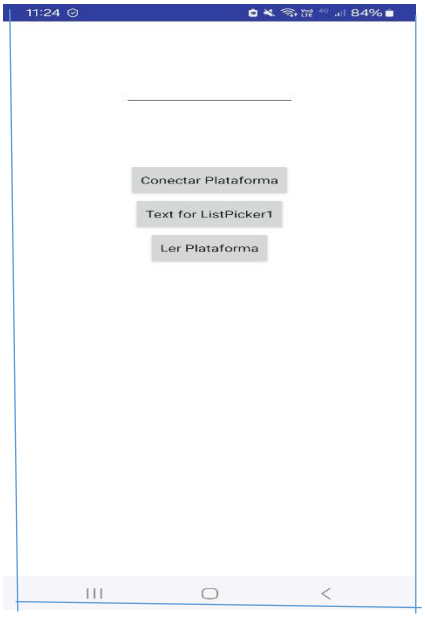


Figura 13- Tela do aplicativo, com o comando de conexão Bluetooth, o botão para iniciar medição do material e um texto indicando a classificação da amostra, no caso, classificada como não alimento. Fonte: Próprio autor

3- Construção da base de conhecimento e metodologia da aplicação

3.1 Construção da base de conhecimento da pesquisa

Conforme explicado no item 2.1, para a construção da base de conhecimento que é utilizada para identificação de material orgânico alimentar, foram realizadas medições, para as amostras vistas na Tabela 2. Como visto, foram submetidas amostras de frutas, carnes, vegetais e alimentos misturados. Para cada uma das 40 amostras, realizou-se dez medições. Para cada objeto, foram realizadas, durante 20 minutos, seis medições por minuto (ou seja, uma medição de dez em dez segundos), totalizando 120 medições por objeto.

Ainda, considerando as medições, esse método de medição foi executado cinco vezes (por objeto de amostra), gerando-se uma tabela de 600 tuplas, contendo todas essas medições por objeto de amostra. Esses dados foram transferidos para arquivos do tipo CSV (gerados com JSON), que foram utilizados na construção final da base de dados de conhecimento geral da pesquisa e dos testes subsequentes. Desse modo, gerou-se arquivos de planilhas para cada objeto de amostra, que foram utilizados para a construção de base de conhecimento geral, aplicados no tratamento dos dados nos testes.

No final, cada conjunto de dados de cada amostra foi colocado em sua própria planilha e depois todas foram combinadas em uma planilha única, de forma consolidada. Desse modo, a base de conhecimento foi separada em quatros tipos de tabelas, uma para os alimentos climatéricos, outras para os não climatéricos, carnes e dos misturados. Houve uma fase para execução de correspondência de dados, de modo que, pudessem ser tratados os dados ausentes ou com “defeito”, assumindo-os como perdas aleatórias.

Para construção da base, de modo que ela pudesse ser utilizada em qualquer ocasião, o mais genérico possível, com uso dos algoritmos de ML vistos no Item I.2, foi efetuado um controle sobre situações de *underfitting* e o *overfitting*, para melhor treinar o modelo proposto. (SANJAYA, *et al*., 2023). Conforme visto no trabalho de HASTIE, *et al.* (2009), o *underfitting* ocorre quando o modelo apresenta variância baixa e um viés alto nas predições e não possui capacidade de generalizar os dados de forma eficiente. Neste caso, acaba sendo um modelo que "decora" os dados de treinamento, ao invés de aprender os padrões.

No caso de *overfitting*, tal situação ocorre quando o modelo se ajusta excessivamente aos dados de treinamento (BISHOP, 2006). Dessa forma, o modelo permanece muito complexo e voltado só para o universo dos dados usados no treinamento. Nesse caso, o modelo não pode ser utilizado em novos dados (HASTIE, *et al*.,2009). Para tratar dessa situação e também dos *inputs* inconsistentes, utilizou-se da combinação das técnicas de PCA, LDA e SVM, de modo que, a base de conhecimento fosse gerada sem dados prejudiciais a análise dos testes de verificação. Também visando tratar essa situação, utilizou-se da técnica de Imputação Múltipla (MI), proposta por RUBIN (1987), para recuperar parte dos dados inconsistentes, sendo que um modelo dessa aplicação pode ser observado na Figura 14.

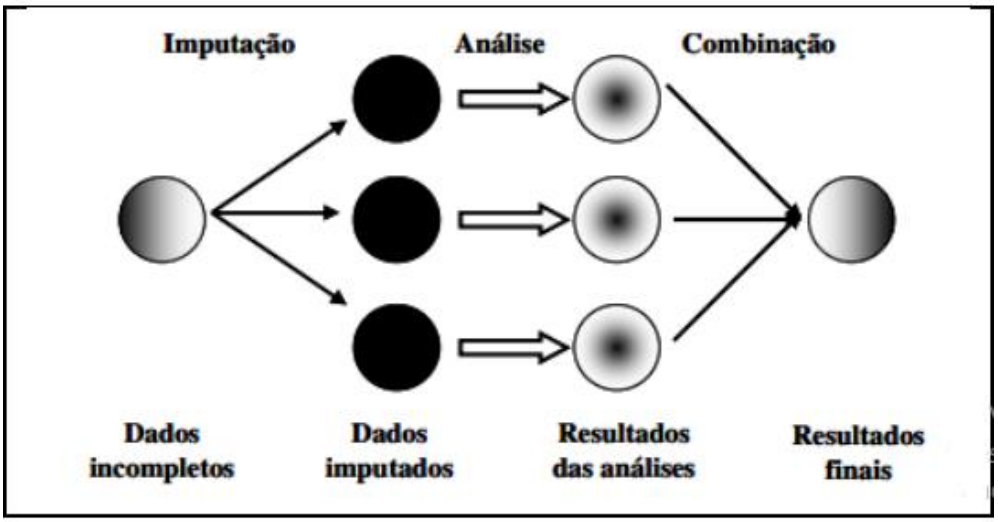


Figura 14- Exemplo de técnica de Imputação Múltipla (MI)
Fonte: STOCHERO, JACOBI e LÚCIO, 2020

Nesta técnica, a primeira fase consistiu em assumir que os dados faltantes devem ser considerados como tipo perdas ao acaso (ou do tipo MAR - *Missing at Random*) (RUBIN, 1987). No caso da pesquisa, essa técnica foi usada para não haver aferições incorretas por algum problema na captação dos sensores, como, por exemplo, um sensor com problema de captação, ou de bateria fraca.

Como foram captados 600 dados finais por objeto de amostra (contando-se todas as cinco medições), houve necessidade de eles serem suavizados e agrupados com base em uma série temporal. O processo de suavização serviu para reduzir a variabilidade da série temporal gerada, de modo ela permanece mais simplificada, facilitando sua manipulação. Uma parte do método para suavização de dados foi usar a média móvel, sendo que o cálculo da média considerou os pontos adjacentes, usando-se a técnica de suavização para identificar as tendências de valores ao longo do tempo das coletas.

A fim de mitigar o problema de perdas aleatórias, também foi estabelecido uma padronização das bases, de modo a ter-se um agrupamento de dados para sincronizar a frequência de medições das amostras, para que nenhuma corra o risco de ter muito menos valores úteis do que uma outra. No caso de dados ausentes, foi realizado a imputação através de um algoritmo KNN condicional, que é uma variante do KNN tradicional. A execução desse algoritmo permitiu ter uma estimativa probabilística mais de acordo com os valores presentes, de modo que os valores ausentes não se tornassem discrepantes (VIGNOLI, 2021).

Para isso, o algoritmo selecionou os 20 valores vizinhos mais próximos dos que possuíam valores não ausentes, e aplicação de uma predição baseada nestes vizinhos. Desse modo, foi preservado as relações entre as variáveis relacionadas aos dados ausentes com os dados presentes, sendo que, ao final, gerou-se a planilha por cada objeto de amostra.

Contudo, como dito, as planilhas por objeto de amostra foram mescladas por classificação, ou seja, as planilhas da base de conhecimento ficaram sendo: uma planilha para alimentos climatéricos (medições das dez amostras de banana e das 12 de cenoura), uma planilha para alimentos não climatéricos (medições das nove amostras de limão), uma planilha para alimentos tipo carnes (medições das quatro amostras de frango) e uma planilha para alimentos do tipo

misturas (medições das cinco amostras de misturas).

Os valores finais inseridos nessas planilhas de classificação também foram calculados através da aplicação de um algoritmo de média truncada, buscando eliminar a influência de *outliers* no cálculo da média. Essa técnica foi aplicada para evitar que valores anormalmente altos ou baixos que pudessem distorcer os valores da planilha unificada, para tal, aplicou-se nas planilhas a Equação 1:

$$TM_p = \frac{1}{n - np} \sum_{i=(np/2)+1}^{n-(np/2)} x_{(i)}$$

Equação 1- Equação de Vignoli (2021) para tratar de outliers nas planilhas das amostras de cada objeto.

Nessa equação, os parâmetros descritos são:

- TM_p é a média truncada,
- n é o número total de observações,
- p é a porcentagem de observações a ser truncada,
- $x_{(i)}$ são os valores das observações ordenadas em ordem crescente,
- A soma é calculada sobre as observações restantes após o truncamento

Para aplicação dessa técnica, ajustou-se o percentual de observação como sendo de 2,5% , indicando (no total de todas as amostras e medições) cerca de 600 dados muito fora da média normal calculada, sendo 330 dados das amostras de banana e cenoura; 135 dados de amostras de limão; 60 dados das amostras de frango e 75 dados das amostras de misturas. Assim os dados de maior e menor valor foram descartados e a média inserida em cada linha da tabela aproveitou os dados restantes para o cálculo.

Com os dados consolidados, pode-se gerar os gráficos indicados nas Figuras 15, 16, 17 e 18, que apresentam mostram os resultados do

processamento de dados confeccionados, que vão ser utilizados para identificar material orgânico alimentar, considerando as situações de material orgânico do tipo climatério, não climatério, tipo carne de misturas. Nos gráficos, foram consideradas as médias de cada tupla, que representam as medições durante um tempo de 20 minutos.

Os índices de CH4, CO2 e amônia NH3 desses alimentos podem ser observados (após a consolidação de todos os dados), de modo que as características de emissão desses gases são as que foram usadas para identificar um material orgânico alimentar, nos testes, em qualquer estágio de decomposição (ainda não decomposto, inicial ou final). Depois que a base de conhecimento foi construída e os sinais de metano, dióxido de carbono e amônia emitidos pelos materiais orgânicos alimentares foram tagueados, a plataforma foi considerada pronta os testes de identificar da situação de um material, se ele seria um material orgânico alimentar ou não.

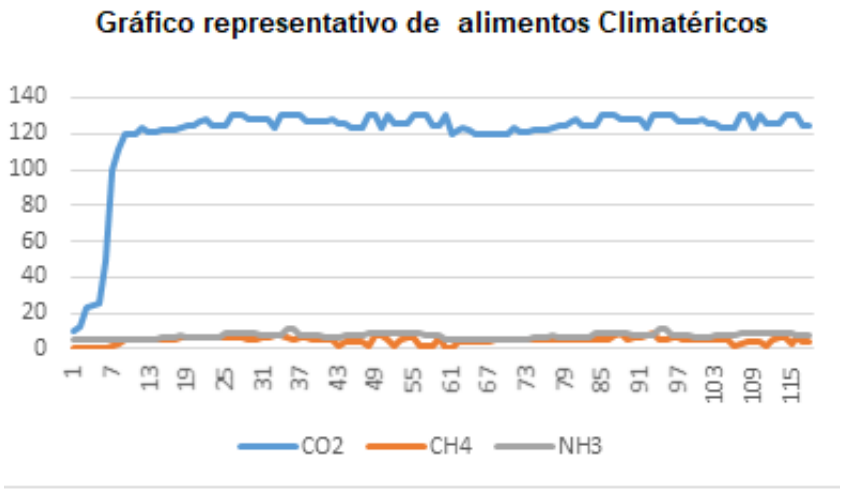


Figura 15- Gráfico de consolidação da base de conhecimento, sendo X o tempo (de dez em dez segundos, durante 20 minutos) e Y ppm (representando a emissão de gás). Esse gráfico apresenta os dados das emissões de gases representativos dos alimentos do tipo climatéricos, mostrando as emissões de CO2 entre 120 e 130 ppm, CH4 entre 4 e 9 ppm e NH3 entre 5 e 9 ppm.

Fonte: Próprio autor.

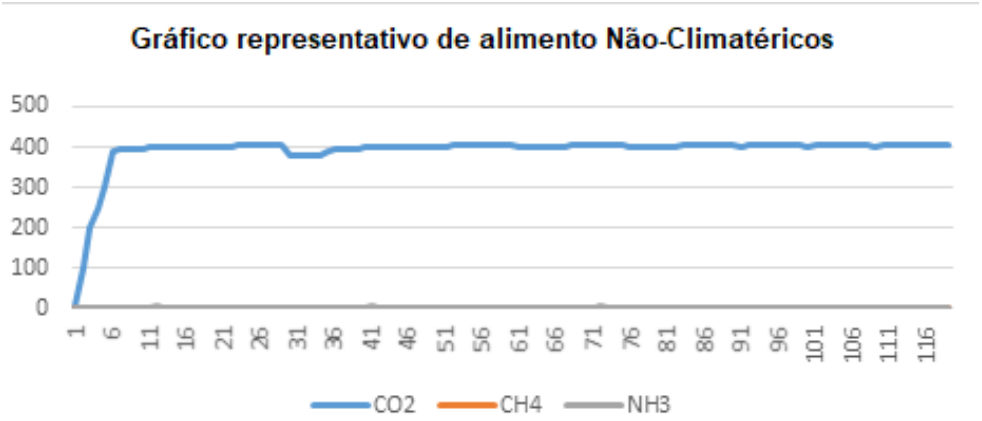


Figura 16- Gráfico de consolidação da base de conhecimento, sendo X o tempo (de dez em dez segundos, durante 20 minutos) e Y ppm (representando a emissão de gás). Esse gráfico apresenta os dados das emissões de gases representativos dos alimentos do tipo não-climatéricos, mostrando as emissões de CO2 entre 10 e 400 ppm, CH4 entre 1 e 5 ppm e NH3 entre 1 e 6 ppm.

Fonte: Próprio autor.

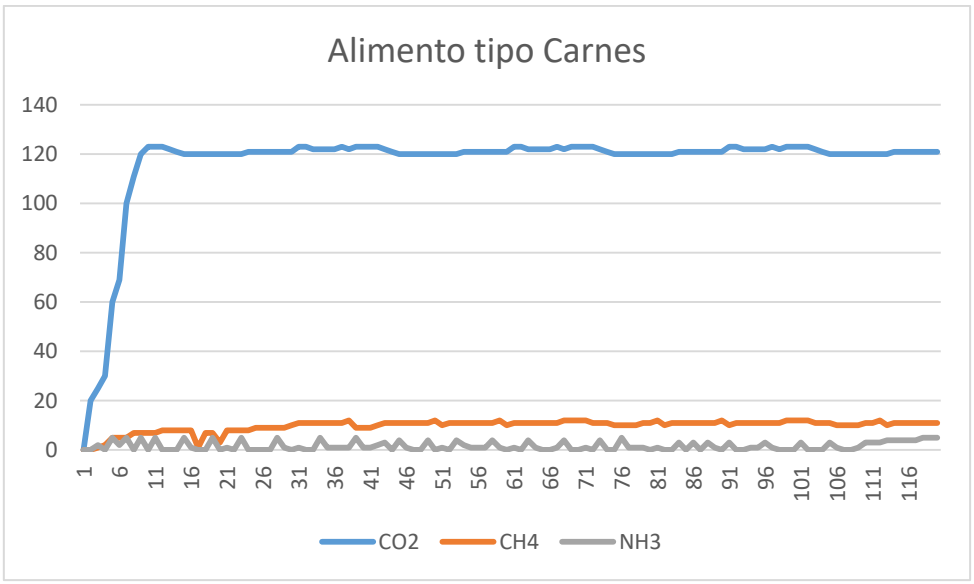


Figura 17- Gráfico de consolidação da base de conhecimento, sendo X o tempo (de dez em dez segundos, durante 20 minutos) e Y ppm (representando a emissão de gás). Esse gráfico apresenta os dados das emissões de gases representativos dos alimentos do tipo carne, mostrando as emissões de CO2 entre 100 e 120 ppm, CH4 entre 4 e 12 ppm e NH3 entre 1 e 5 ppm.

Fonte: Próprio autor.

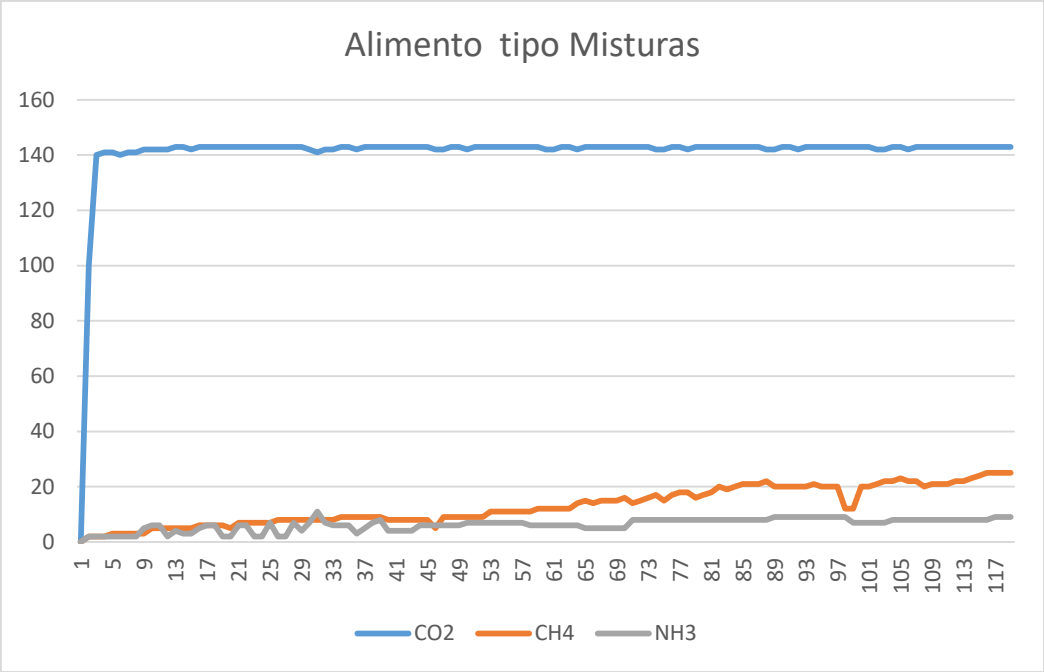


Figura 18- Gráfico de consolidação da base de conhecimento, sendo X o tempo (de dez em dez segundos, durante 20 minutos) e Y ppm (representando a emissão de gás). Esse gráfico apresenta os dados das emissões de gases representativos dos alimentos do tipo misturas mostrando as emissões de CO2 entre 120 e 141 ppm, CH4 entre 5 e 25 ppm e NH3 entre 2 e 10 ppm. Fonte: Próprio autor.

Com essas tabelas e gráficos, percebeu-se que o principal gás de emissão de matéria orgânica alimentar é o CO2 (entre 10 e 140 ppm), com emissões entre 2 e 25 ppm de CH4 e entre 1 e 10 ppm de NH3. Essas características foram usadas para classificar os objetos de amostras dos testes, verificando se, após um tempo, e emissão de gases se aproximava da identificação entre as classes de matéria orgânica alimentar, cujos dados foram obtidos na fase de treinamento e classificação da plataforma.

Ressalta-se que, nos testes preliminares, feitos antes da consolidação final, nas amostras de carnes e misturas, percebeu-se uma taxa de emissão de NH3 e CH4 maior do que nos testes definitivos (entre 5 e 70 ppm). Isso deveu-se ao preparo equivocado das amostras, que estavam em estado de putrefação adiantado, causando essa distorção. om as amostras mais bem preparadas, os testes definitivos mostraram índices de CH4 e NH3 de matéria orgânica alimentar

recém-feita bem menores do que os que foram vistos nas amostras preliminares, pois estavam em estado de decomposição bem inicial.

3.2- Metodologia e análise dos testes

Para testar a plataforma, foram escolhidas amostras diversas, tais como, invólucros de latas (alumínio), de vidro, peças de plástico e alimentos comuns em lanchonetes de shopping, tipo restos de hambúrgueres, restos de salada, pedaços de cenoura, pedaços de maça, pedaços de banana e restos de morango e uva.

Para efeito de medição dessas amostras, e de acordo com o visto no trabalho de VASCONCELOS, *et al.* (2022), cada amostra de teste, de forma aleatória, foi colocada na caixa e as medições ocorreram entre 10 segundo até, no máximo, três minutos, com todos os dados captados nesse interim. A maioria das amostras emitiu resultados da classificação já a partir dos 10 segundos, algumas demoraram quase um minuto, mas manteve-se, assim mesmo o tempo máximo de três minutos

A partir dos 10 segundos, a plataforma iniciou a emissão dos sinais de CO2, NH3 e CH4, que foram capturados e enviados ao módulo analisador da plataforma. Conforme visto no capítulo II dessa dissertação, a parte de análise ocorreu dentro do módulo instalado no dispositivo móvel, que contém a parte para classificar o tipo de material como orgânico alimentar ou não.

A plataforma gerou os dados em formato CSV, que podem ser visualizados como gráficos através de um software de planilha. Esses dados no formato CSV de cada amostra foram submetidos a análise, via microsserviços *Estmr* e *Catlg*, comparando-os com os dados da base de conhecimento (que também usa o microsserviço *Conect*), obtidos pelo treinamento, para gerar a classificação da amostra para o usuário final, podendo essa amostra ser classificada como “alimento orgânico” ou “não alimento”.

A base de conhecimento foi estruturada de modo a que fosse observado os dados como alimento climatério, não climatério, carnes e misturas, assim, se os dados de uma das amostras de testes não se encaixavam, nessas classificações, a amostra, como já dito, foi classificada como “não alimento”. Ou seja, as amostras que não puderam ser classificados como orgânico alimentar, são classificados como material não alimentar (independentemente de serem de origem orgânica ou não), uma vez que dados gerados não correspondam a nenhum outro contido na base de conhecimento. Para qualquer dos casos (“orgânica alimentar” ou “não alimento”), vai ser sempre emitido um aviso em texto para o usuário do resultado final.

A Figura 19 apresenta gráficos dos resultados da emissão de gases de CO2, NH3 e CH4 de duas amostras testadas, envolvendo restos de hambúrguer e uma lata de alumínio vazia (como a de refrigerantes). Nessa Figura, se visualiza a taxa de emissão dos gases ao longo de, no máximo 3 minutos, mas observa-se que a partir dos 10 segundo já é possível, para a plataforma, determinar se o material é orgânico alimentar ou não.

Outra observação obtida da Figura 17, é que a amostra de hambúrguer apresentou uma ascendência de emissão de CO2, que vai de 10 a 100 ppm. Essa ascendência foi percebida nos primeiros 10 segundos, no qual a taxa já chegou a 20 ppm de medição. No caso do CH4 e NH3, a taxa de subida do CH4 ficou entre 0 e 6 ppm, e a do NH3 entre 0 a 3 ppm de NH3 durante todos os três minutos, fazendo com que a plataforma identificasse essa amostra como orgânica alimentar

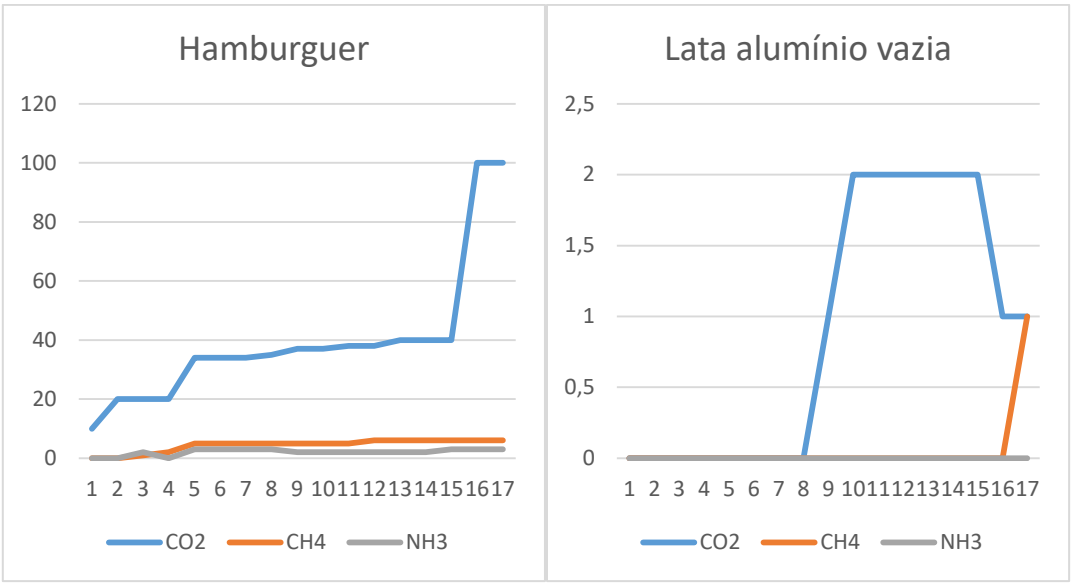


Figura 19- Gráficos da emissão de CO2, CH4 e NH3 de uma amostra de hambúrguer e de uma amostra de lata de alumínio vazia, obtidos pela plataforma. Nesse caso, o gráfico mostra uma taxa mais alta de emissão de CO2, NH3 e CH4 do hambúrguer, o que o coloca classificado como orgânico alimentar, já as latas de alumínio, com taxas de emissão dos mesmos gases bem menor, fizeram a plataforma identifica-la como não alimento.
Fonte: Próprio autor.

Já na amostra da lata de alumínio vazia, a emissão de CO2 foi muito abaixo do que foi observado na amostra de hambúrguer (entre 0 e 2 ppm), sendo que nos primeiros 10 segundos ficou em 0 ppm. Isso também aconteceu com as detecções de NH3 e CH4, que variaram de 0 a 1 ppm durante os três minutos. Com esses dados obtidos, a plataforma classificou essa amostra como “não alimento”.

Nas Figuras 20a e 20b são apresentados exemplos de telas para um dispositivo móvel, após o pedido de verificação da amostra pela plataforma, com as respostas relativas à amostra de hambúrguer e da lata de alumínio vazia. Os dados foram classificados pelo módulo de classificação que está no aplicativo instalado no dispositivo móvel.

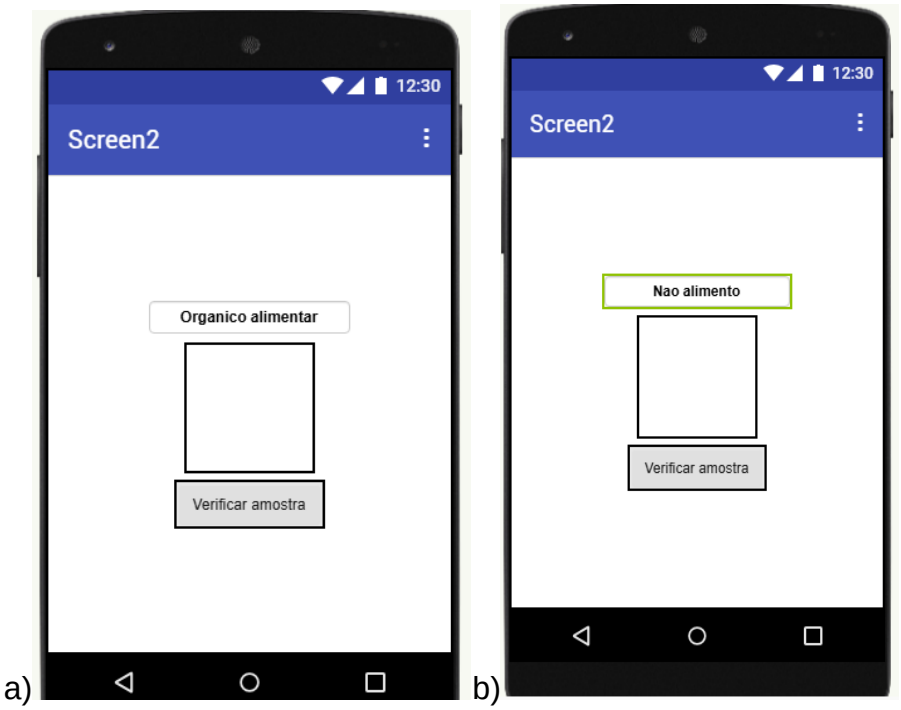
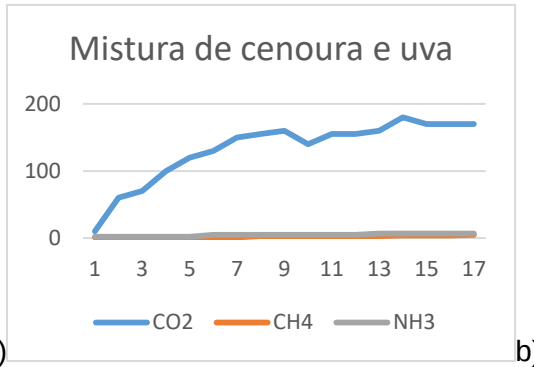


Figura 20- Tela com a apresentação da resposta da plataforma em relação a amostra de hambúrguer colocada dentro da caixa de verificação da plataforma. Após clicar no botão “verificar amostra”, após três minutos, a plataforma envia os dados consolidados e o aplicativo responde com a informação de ser “orgânico alimentar” (a) ou “não alimento” (b).

Fonte: Próprio autor.

Outros resultados dos testes, em formato de gráfico, também são vistos nas Figuras 21 e 22, sendo que na Figura 21a é observado um gráfico da mistura de restos de cenoura (que é um alimento climatério) com restos de uva (um alimento não-climatério), mostrando as emissões de CO2, CH4 e NH3. Já na Figura 21b, é visto as emissões de CO2, CH4 e NH3 de um invólucro de plástico vazio.

Na Figura 22 aparecem resultados que indicam algumas discrepâncias na execução do algoritmo usado nos exames durante os testes. No caso da lata com líquido (Figura 22a), o analisador indicou que havia uma amostra orgânica alimentar, mesmo o índice de CO2 estando abaixo do indicado pelo treinamento (1 a 20 ppm). Ao perscrutar a origem dessa classificação, viu-se que o algoritmo, que usa de probabilidade, identificou a subida do CO2 como significativa para identificar uma possível amostra alimentar.

a) 

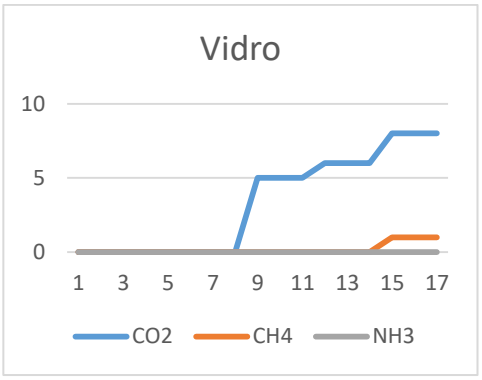
b) 

Figura 21- Gráficos de dados de emissões de CO2, CH4 e NH3 obtidos pela plataforma de uma amostra com mistura de cenoura e uva (a) e um invólucro de vidro(b). O gráfico (a) foi identificado como Orgânico Alimentar (com emissão de CO2 entre10 e 145 ppm, de NH3 entre 1 e 4 ppm e CH4 entre 1 e 5 ppm), já o gráfico (b) foi identificado como não alimento, com uma emissão de CO2 entre 0 e 10 ppm, CH4 entre 0 e 1 ppm e NH4 0 ppm
Fonte: Próprio autor.

O mesmo ocorreu com o vidro com resto de comida (Figura 22b), que teve uma emissão de CO2, CH4 e NH3 mais significativa (2 a 50 ppm, 1 a 3 ppm e 1 a 5 ppm, respectivamente), fazendo com que o analisador identificasse a amostra como orgânica alimentar. Outras amostras de vidro com resto também foram submetidas, mas a emissão de CO2 captada, e o indicio de ascendência dos valores da emissão foram menores do que nessa amostra, fazendo com que elas fossem classificadas como não alimento.

O tempo para o analisador classificar essas amostras foi maior do que as amostras anteriores, pois o analisador demorou de 20 a 50 segundos para verificar uma taxa de subida de CO2, CH4 e NH3 nessas amostras e, por fim, classifica-las como orgânicas alimentares. Já na amostra de invólucro de papelão (Figura 22c), observou-se que o sinal ascendente de CO2 só ocorreu a partir dos 90 segundos, e, mesmo assim, o valor emitido de CO2 estava bem abaixo dos 10 ppm. Esse valor baixo, ao ser submetido aos resultados de classificação de amostras da base de conhecimento (climatérico, não-climatérico, carnes e misturas) indicou que o papelão não se encaixou em nenhum desses tipos, sendo, então, classificado como não alimentar.

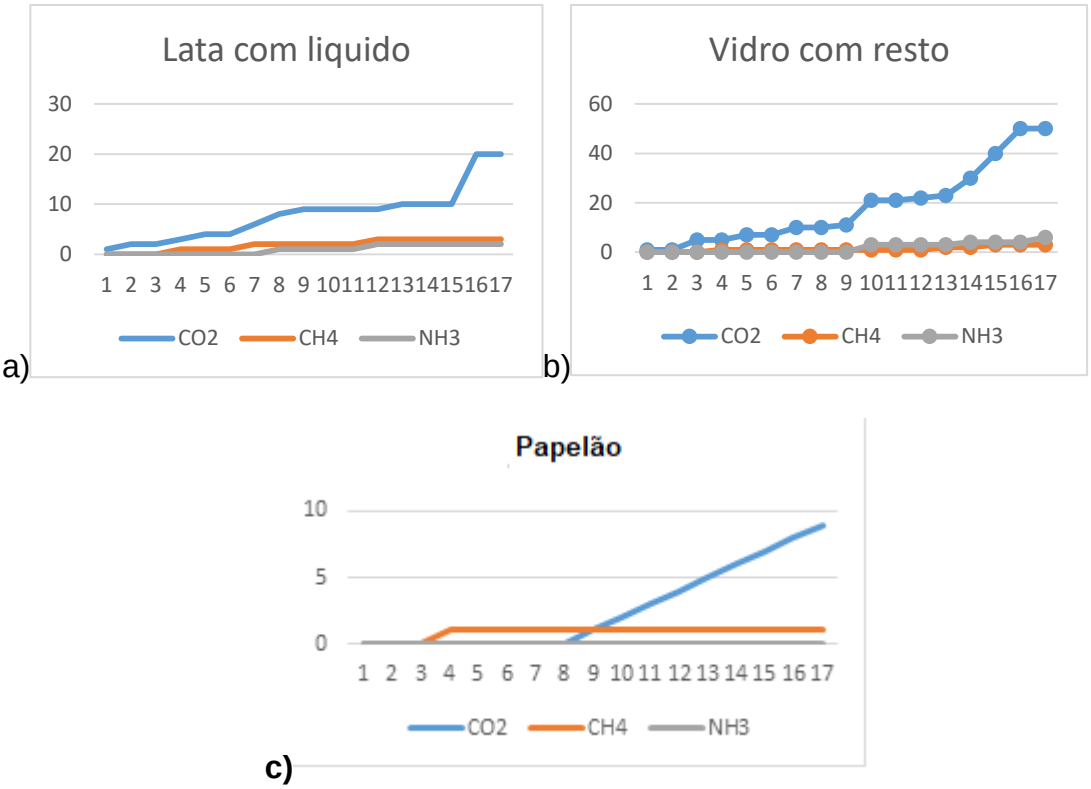


Figura 22- Gráficos de emissões uma amostra de lata com liquido (tipo refrigerante) (a), um vidro com restos de comida (b) e um involucro de papelão (b). Nesses casos, a lata com liquido e o vidro com resto foram identificados pelo analisador com orgânicos alimentares e o papelão como não alimento, apesar de papel ser de origem orgânica, mas não ser um alimento.
Fonte: Próprio autor.

No total, vinte amostras, de diversos tipos, foram usadas nos testes e todas geraram informações de ser uma amostra orgânica alimentar ou não alimento conforme o analisador as computava, ou seja, não houve uma situação em que o analisador não mostrasse um resultado. Outros detalhes importantes referentes a metodologia, aplicação e análise inicial dos testes foram:

- Cada amostra foi envolta em um plástico para latas de lixo, para que a inserção dela não contaminasse o invólucro de polietileno (conforme visto na Figura 23);

- O tempo de inserção entre uma medição de amostra e outra foi de cinco minutos;
- Para cada uma das 20 amostras usadas, foram feitas duas medições, com intervalos de tempo considerável, resultando no que é visto nas Tabelas 4 e 5 (vistas e discutidas no capítulo a seguir).
- Os sensores do tipo MOS necessitam ser aquecidos antes do uso correto. No caso dessa pesquisa, para usa-los eles foram aquecidos por 24 horas, e somente utilizados, tanto para construção da base, quanto nos testes, de forma continua, após esse tempo.



Figura 23- Arranjo do depósito de polietileno, com os componentes de hardware da plataforma, com o uso de saco plástico para isolar a amostra e permitir sua medição sem contaminação da amostra anterior.
Fonte: Próprio autor.

4- Discussão dos resultados

Como visto no Capítulo III desse trabalho, a plataforma classifica os objetos nela inseridos, a partir da obtenção dos dados emitidos pelos sensores de CO2, CH4 e NH3, classificando-os como orgânico alimentar ou não alimento.

A base de dados foi criada a partir de um conjunto de medições de diversos tipos de amostras, apresentando alguns problemas de perda de dados ocasionais, que forma resolvidos com aplicação do algoritmo KNN, que se mostrou satisfatório no preenchimento de informações disjuntivas.

Outra situação, baseada em situações que podem ocorrer no mundo real, certas amostras são invólucros (de lata, vidro ou plástico) contendo algum elemento orgânico, em algumas análises de amostras a plataforma classificou o material inorgânico (os invólucros) junto com material orgânico e pode os indicaram como material orgânico alimentar, mesmo apenas parte sendo alimento de fato.

Como visto no capítulo anterior e de acordo com o visto nas Figuras 22a e 22b, algumas medições indicaram situações que poderiam ser consideradas controversas, pois com a mistura de material inorgânico com orgânico, a execução do algoritmo analisador indicou uma elevação dos gases CO2, NH3 e CH4, o que fez interpretar (em certas situações) como material orgânico alimentar, mesmo a amostra apresentando junto material inorgânico.

Observou-se que isso aconteceu apenas quando a medição de CO2 (que foi o gás que mais influenciou nessas medições controversas) se comportava em ascendência, atingindo em menos de minuto uma medida maior ou igual que 20 ppm. Mas nem todas as amostras de invólucros inorgânicos, com conteúdo orgânico, foram classificadas deste modo.

Foram testadas na plataforma cerca de 20 amostras, com intervalos, entre cada deposição, de até cinco minutos. Para o repositório de polietileno não ser “contaminado” com restos, e influenciar as medidas, uma pequena sacola de

lixo (de plástico) foi inserida antes de se colocar cada amostra e depois retirada, conforme visto na Figura 23. Quando outra amostra era colocada, uma nova sacola no repositório também foi posta.

Na Tabela 4 apresentam-se os dados de medições das amostras de testes, relativo à primeira rodada de medição das 20 amostras submetidas e inseridas na plataforma, com suas classificações dadas pelo módulo analisador. Para confirmar essas medições, foi realizada uma segunda rodada de medição das mesmas amostras, cujos resultados podem também ser observados na Tabela 4.

Na primeira medição, viu-se que o tempo máximo de três minutos se mostrou mais que satisfatório para a obtenção dos resultados. A identificação observada no dispositivo móvel levou uma média de 30s para ser apresentada. O maior tempo de apresentação de resultados foi de latas de alumínio com restos (30 a 50s). A casca de banana, apesar de poder ser considerada orgânico alimentar, foi o material orgânico que levou mais tempo para ser identificado (40s).

As identificações controversas foram as que envolviam invólucros não orgânicos, com algum material orgânico inserido. Se o invólucro (fosse de alumínio, plástico ou vidro) estivesse fechado de forma hermética, o analisador não podia captar nenhuma emissão de gases, classificando essa amostra como não alimento.

Como essas amostras de invólucros não possuem um modo “automático” de separação em latas de lixo comum, precisa-se de haver uma fase, no qual, manualmente, a pessoa separasse o invólucro do material. Quando o invólucro está vazio, não corre identificação controversa e o material, seja plástico, alumínio, vidro ou qualquer outro (inclusive de papel) é identificado corretamente como não alimento.

Já na segunda medição (Tabela 5), o tempo médio de identificação foi menor, uma vez que já havia se passado mais de cinco minutos, fazendo que a emissão de gases estivesse mais acelerada, por conta da fase de decomposição, que acelera a emissão de CO2, NH3 e CH4. Os tempos dos

invólucros vazios ou fechados não foram alterados (não houve captação de decomposição nessas amostras).

Tabela 4- Resultados finais da 1º medição de 20 amostras de testes.

Número da amostra	Tipo de material	Tempo que se demorou para classificar	Classificação resultante
1	Restos de hambúrguer	20s	Orgânico alimentar
2	Restos de salada	20s	Orgânico alimentar
3	Restos de espaguete	20s	Orgânico alimentar
4	Mistura de cenoura com uva	20s	Orgânico alimentar
5	Mistura de arroz e feijão	20s	Orgânico alimentar
6	Casca de banana	40s	Orgânico alimentar
7	Restos de salada de frutas	30s	Orgânico alimentar
8	Mistura de arroz com frango	20s	Orgânico alimentar
9	Lata de alumínio vazia	20s	Não alimento
10	Lata de alumínio com restos de refrigerante	40s	Orgânico alimentar
11	Lata de alumínio com restos de suco	40s	Não alimento
12	Vidro aberto com frango	30s	Orgânico alimentar
13	Mistura de cenoura com uva	20s	Orgânico alimentar
14	Vidro sem nada (puro)	20s	Não alimento
15	Plástico com suco	30s	Não alimento
16	Plástico puro	20s	Não Alimento
17	Papelão	20s	Não alimento
18	Vidro fechado com carne	20s	Não alimento
19	Lata de alumínio cheia, mas fechada	20s	Não Alimento
20	Mistura de couve e milho	20s	Orgânico alimentar

Fonte: Próprio Autor

Tabela 5- Resultados finais da 2º medição de 20 amostras de testes.

Número da amostra	Tipo de material	Tempo que a plataforma demorou para classificar	Classificação resultante
1	Restos de hambúrguer	10s	Orgânico alimentar
2	Restos de salada	10s	Orgânico alimentar
3	Restos de espaguete	10s	Orgânico alimentar
4	Mistura de cenoura com uva	10s	Orgânico alimentar
5	Mistura de arroz e feijão	10s	Orgânico alimentar
6	Casca de banana	20s	Orgânico alimentar
7	Restos de salada de frutas	10s	Orgânico alimentar
8	Mistura de arroz com frango	10s	Orgânico alimentar
9	Lata de alumínio vazia	20s	Não alimento
10	Lata de alumínio com restos de refrigerante	40s	Não alimento
11	Lata de alumínio com restos de suco	30s	Orgânico alimentar
12	Vidro aberto com frango	30s	Não alimento
13	Mistura de cenoura com uva	10s	Orgânico alimentar
14	Vidro sem nada (puro)	20s	Não alimento
15	Plástico com suco	10s	Orgânico alimentar
16	Plástico puro	20s	Não Alimento
17	Papelão	20s	Não alimento
18	Vidro fechado com carne	20s	Não alimento
19	Lata de alumínio cheia, mas fechada	20s	Não alimento
20	Mistura de couve e milho	10s	Orgânico alimentar

Fonte: Próprio Autor

Em algumas amostras de invólucros abertos, com restos orgânicos, as medições, além de demorarem mais, indicaram resultados diferentes. No que antes, na primeira medição, foi classificado como orgânico alimentar, na segunda a classificação alterou-se para não alimento. Isso ocorreu nas amostras de número 10 e 12. Já na amostra 15 (vidro aberto com suco), houve o contrário, na primeira medição foi identificado como não alimento, mas na segunda medição foi classificado como orgânico alimentar.

Isso representou uma discrepância de classificação de 15% entre as amostras. No caso da amostra de lata de alumínio com líquido de refrigerante (amostra 10), uma justificativa dessa alteração pode ser devido ao fato de a maioria dos refrigerantes são líquidos carbonados, ou seja, com muito carbono, como demorou-se um tempo entre a primeira e segunda medição, a emissão de carbono reduziu-se, o que justificaria a nova classificação de não alimento, na segunda medição.

Nas amostras 12 e 15 , precisar-se-ia de mais de um conjuntos de exames para poder encontrar uma provável causa dessas discrepâncias entre as duas medições, mas não foi o tema da pesquisa, podendo ser um tema futuro de outras pesquisas

Assim, a plataforma identificou 85% das amostras em ambas medições, com a mesma classificação. Apesar desse índice estar dentro da expectativa vista na Introdução (que era entre 80% a 95% de certeza de identificação), não significa que a plataforma consiga distinguir de forma clara situações em que haja um invólucro aberto, com material orgânico.

Mesmo com a inserção de outros métodos de identificação (textura, cor etc), a plataforma não possui, nem foi planejada, capacidade de retirar e separar material orgânico de inorgânico nesses casos. Mas a classificação indica uma situação predominante. Por exemplo, invólucros fechados, mesmo com material orgânico, foram identificados como não alimento pelo fato de não emitirem os gases escolhidos.

No caso de material separado, vidro vazio, lata vazia, mistura de material orgânico etc, não houve discrepâncias de classificação. O material inorgânico foi

classificado como não alimento (inclusive o papelão, que é uma matéria orgânica, mas não alimentar) e o material orgânico foi classificado (exceto o papelão) como orgânico alimentar, nas duas medições.

5- Conclusões

A plataforma proposta visou ser um dispositivo para instalação em latas de lixo de lugares públicos cobertos (shoppings e escolas), com a intenção de facilitar a separação de material orgânico alimentar de material inorgânico ou não alimentar.

O desenvolvimento incluiu a utilização de sensores de gás tipo MOS, para identificar CO2, NH3 e CH4 em restos alimentares e, assim, poder classificar a amostra como orgânica alimentar ou não; e, também. o uso de algoritmos de ML para tratar os dados de treinamento, criar um base de conhecimento e para poder , através de dados de amostras de testes, verificar o grau de acuracidade dessa classificação.

Os algoritmos de classificação se basearam nas medições e na ação de treinamento. A classificação em 4 tipos , climatério, não-climatério, carnes e misturas foi, para essa plataforma, adequada para as situações apresentadas, consolidando mais ainda o uso desses algoritmos para sistemas de ML básicos.

A plataforma consegui, entre 20 amostras de testes, confirmar, em duas medições, os dados classificados de 85% dessas amostras, ou seja, 17 amostras de testes foram classificadas iguais na primeira e segunda medição. Contudo, ocorreram problemas na identificação de três amostras de invólucros abertos, com material orgânico em seu interior, que foram classificadas na primeira medição como orgânico alimentar e na segunda como não alimenta, ou vice-versa.

Nesses casos, a plataforma apenas verificou a emissão de gases ao longo de um tempo específico, sendo que, em certas amostras, principalmente as com invólucros não orgânicos, ela verificou uma situação de “confundir” a classificação, vendo uma amostra como orgânico alimentar ou não alimento na primeira medida e invertendo essa classificação na segunda medida.

Ainda que possa haver razões químicas e físicas para essa ocorrência percebe-se que a plataforma tende a considerar qualquer amostra que contenha restos de comida como orgânica alimentar, não podendo, nesses casos, separar esse tipo de lixo, prescindido de uma etapa manual de separação para concluir essa etapa.

Ainda assim, o índice de 85% de confirmação entre duas medidas, feitas com intervalos de tempo razoável (mais de cinco minutos), esteve dentro da expectativa inicial do uso da plataforma (que foi entre 80% e 90%). Mas isso significa que, ainda se deve buscar alguma forma mais eficiente para ajudar na separação do lixo, sendo que a plataforma, mesmo podendo identificar material orgânico alimentar, ainda está incompleta para toda as situações possíveis.

Contudo, confirma-se que o modelo estabelecido por ela, de poder usar sensores mais baratos, ainda que precisem de cuidados (como pré-aquecimento), mostrou-se um caminho possível para implantar soluções dessa natureza em ambiente escolar e shoppings, sendo uma plataforma barata e de fácil manipulação, além de ser capaz de funcionar por longo tempo com uma abateria simples, evitando-se trocas constantes.

Para poder aumentar a precisão de classificações, poder-se-ia utilizar sensores mais sensíveis, tipo pzioelétricos ou químicos, mas que tem um custo bem maior, encarecendo a montagem da plataforma como um todo. Um dos objetivos secundário deste trabalho visou desenvolver um dispositivo de baixo custo para classificar materiais, o que se mostrou possível.

Uma outra ideia de melhoramento seria atualizar a base de conhecimento, com mais dados de outros tipos de amostras, misturando mais material inorgânico com orgânico, para ver até que ponto não ocorreria um desvio total de classificação. Isso pode ser uma visão de trabalho futuro sobre a plataforma, assim como experimentar outros tipos de algoritmos de ML

Não foi objeto dessa pesquisa a identificação com uso de sensores de textura ou cor, sendo esse um campo de pesquisa que pode ser explorado em trabalhos posteriores e pode indicar uma continuidade no desenvolvimento proposto, inclusive para detectar e diferenciar material orgânico não alimentar, como subprodutos de papel ou da flora (flores não-comestíveis).

Referências

ABBAS, F N., *et al* (2020) Capable of Gas Sensor MQ-135 to Monitor the Air Quality with Arduino uno. **International Journal of Engineering Research and Technology**. v 13, n 10, pp. 2955-2959 https://dx.doi.org/10.37624/IJERT/13.10.2020.2955-2959 2955

ARDUINO-CC (2024). Sobre Arduino R4. Visto em https://docs.arduino.cc/hardware/uno-r4-minima/. Acessado em outubro de 2024.

AYAMBA, B. E., *et al* (2023). Mechanisms for nutrient interactions from organic amendments and mineral fertilizer inputs under cropping systems: a review, **PeerJ**, Doi: 10.7717/peerj.15135.

BAJWA, I. S., *et a*l. (2021). Smart eNose Food Waste Management System. **Hindawi Journal of Sensors**, v 2, p. 13.

BAO Y. *et al.* (2021). Organic matter- and temperature-driven deterministic assembly processes govern bacterial community composition and functionality during manure composting. **Waste Management, Elsevier,** v 131, pp 31-40, Doi: https://doi.org/10.1016/j.wasman.2021.05.033.

BANZI, M. e SHILOH, M. (2015). Getting started with Arduino: the open-source electronics prototyping platform. **Maker Media**, 3ª. Ed, Inc., 262 p. ISBN 9781449363338.

BARRET, S. F. e PACK, D. J. (2019). Microchip AVR. Microcontroller Primer: Programming and Interfacing. **Morgan & Claypool Publisher**, 3rd Edit, USA

BARROZO, V. P. *et al*. (2019). Desperdício de Alimentos: O peso das perdas para os Recursos Naturais. **Revista Agroecossistemas**, v. 11, n. 1, p. 75-96.

BHOWMIK, B. (2023).Chemical sensors for e-nose: an effective route for disease diagnosis. **Woodhead Publishing Series: Electronic and Optical Materials**, Woodhead Publishing, pp 395-413, https://doi.org/10.1016/B978-0-323-91157-3.00011-8.

BIN HAFEEZ, S. H. *et al* (2019). Electronic Nose (E-Nose). **2019 International Conference on Information Science and Communication Technology (ICISCT***)*, Karachi, Pakistan, pp. 1-6, Doi: 10.1109/CISCT.2019.8777425.

BRACCINI, V. P. *et a*l (2021). Types of muscle fibers, identification, characteristics, and quality of meat. **Brazilian Journal of Development,** v 7, n 3, https://doi.org/10.34117/bjdv7n3-025

BRON I. U. e JACOMINO A. P. (2007). Classificação de frutos por "climatério" é um conceito em extinção? **Visão Agrícola**, pp. 8-10, São Paulo, Brasil.

CAHYA, M. *et al*, (2023). The Effect of Fermentation on Increasing the Physico-Chemical Quality of Fish Feed Based on Food Waste, **Applied Mechanics and Materials**, v 916, p. 63, DOI: https://doi.org/10.4028/p-VX5Nyu

CAMPUZANO, R. e GONZÁLEZ-MARTÍNEZ, S. (2016). Characteristics of the organic fraction of municipal solid waste and methane production: A review, **Waste Management Elsevier**, v 54. pp 3-12,. DOI: https://doi.org/10.1016/j.wasman.2016.05.016.

CIDADE, I. A. (2022) Um Sistema para Sondagem Ambiental e Verificação da Qualidade do Ar. Dissertação de mestrado. PPCIC, CEFET-RJ.

CONCEIÇÃO, J. T. P. *et al.* (2021). Os desafios da gestão de resíduos sólidos orgânicos urbanos e sua contribuição para a arborização urbana. **UNG Geosciences Journal**, v. 20, pp. 67-81.

DANTAS, R., *et al* (2021). Identificação de compostos orgânicos em lixiviado de aterro sanitário durante tratamento por processo fenton. **Revista Ibero-Americana de Ciências Ambientais**, v 12, n 6. https://doi.org/10.6008/CBPC2179-6858.2021.006.0023

EAMSA-ARD, T. *et al*. (2018). Development of Electronic Nose for evaluation of Fragrance and human body Odor in the cosmetic industry. **2018 IEEE 7th global Conference on consumer Electronics (GCCE)**., pp. 363-364, Nara, Japan.

EDITA, R. *et al.* (2018). Rapid evaluation of fresh chicken meat quality by electronic nose. **Czech Journal of Food Sciences**, v 36, pp 420-426.

FAKHRURRIANA, R. (2023). Developing Android-Based Teaching Modules Using MIT App Inventor **BATARA DIDI: English Language Journal,** v 2, n 1. doi: https://doi.org/10.56209/badi.v2i1.56

FERREIRA, D. SILVA, P. e MADEIRA, T. F. (2019). Embalagens verdes: Conceitos e aplicações. Revista Americana de Empreendedorismo e Inovação, v 1, n 2, DOI: https://doi.org/10.33871/26747170.2019.1.2.2780

FABER,J. *et al* (2013) Air Pollution in New Vehicles as a Result of VOC Emissions from Interior Materials. **Pol. J. Environ. Stud**., v 22, pp 1701–1709.

FUKANO, Y. e TACHIKI, Y (2021) Evolutionary ecology of climacteric and non-climacteric fruits**. Biology Letters, UK.** Doi: https://doi.org/10.1098/rsbl.2021.0352

GHARAVI, A, *et al*. (2022) Application of machine learning techniques for identifying productive zones in unconventional reservoir. **International Journal of Intelligent Networks**, v 3, n 3, pp 87-101. DOI: https://doi.org/10.1016/j.ijin.2022.08.00

GHOSH, P. *et al.* (2020). Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm. **15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP***)*, pp. 1-6; Bangkok, Thailand, , doi: 10.1109/iSAI-NLP51646.2020.9376787.

HALDER, R.K., *et al.* (2024) Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. **Journal of Big Data**, v 11, n 113. https://doi.org/10.1186/s40537-024-00973-y

HASAN, M. A.; SARNO, R. e SABILA, S. I. (2020). Optimizing Machine Learning Parameters for Classifying the Sweetness of Pineapple Aroma Using Electronic Nose. **International Journal of Engineering and Intelligent Systems**, v. 13, n. 5, p. 122-132.

HASTIE, T. et al. (2009) **The elements of statistical learning: data mining, inference, and prediction**. Springer, v. 2. New York, USA ISBN 978-0-387-84857-0.

HECKLER, M. (2021) **Spring Boot: Up and running**. O'Reilly, 1st Edition, UK.

JOŃCA, J. *et al*. (2022) Electronic noses and their applications for sensory and analytical measurements in the waste management plant: A review. **Sensors**, v 22, n. 4. Pp 1510.

KARAKAYA, D., ULUCAN O. e TURKAN. M. (2020). Electronic Nose and Its Applications: A Survey. **International Journal of Autom. Comput.,** v 17, pp 179-209. https://doi.org/10.1007/s11633-019-1212-9

KNOWN, H., *et al.* (2021) High-Speed Implementation of PRESENT on AVR Microcontroller. **Mathematics**, v *9*, n 4,pp 374; https://doi.org/10.3390/math9040374

KUMAR, K. *et al.* (2023) An IoT-Enabled E-Nose for Remote Detection and Monitoring of Airborne Pollution Hazards Using LoRa Network Protocol. **Sensors.** v 23, n 10. https://doi.org/10.3390/s23104885

LEE, Y. H *et al.* (2018). Recent advances in organic sensors for health self-monitoring systems**. Journal of Material Chemistry C,** v 32, pp 8569-8612. Doi: https://doi.org/10.1039/C8TC02230E

LI, Z. *et al*. (2023) E-nose based on a high-integrated and low-power metal oxide gas sensor array. In: **Sensors and Actuators B: Chemical,** v 380, DOI:https://doi.org/10.1016/j.snb.2023.133289.

MAKERHERO (2024). Arduino R4 mínima. Visto em https://www.makerhero.com/produto/arduino-uno-r4-minima/. Acessado em julho de 2024.

MARATOS, F.A. *et al.* (2023). Exploring pattern recognition: what is the relationship between the recognition of words, faces, and other objects?. **Cogn Process** , v 24, pp 59-70, Doi: https://doi.org/10.1007/s10339-022- 01111-3

MARTIN, A.S.S.; LUNARDI, G. L. e DOLCI, D. B. (2020). "Razões que influenciam as organizações a adotarem práticas sustentáveis na área de Tecnologia da Informação". **Journal of Applied Technology**, v 9, n 3, pp. 3-19.

MELO, L. C. A.; SILVA, C. A. e DIAS, B. O. (2008). Characterization of the organic matrix of residues from different origins. In: Química e mineralogia do solo. **Revista Brasileira de Ciências de Solo**, v 32, n 1. Brasil. DOI: https://doi.org/10.1590/S0100-06832008000100010

MOHAMED, R.R *et al.* (2022). Fuzzy Classification In E-NOSE For Food Freshness Determination.. **Journal of Pharmaceutical Negative Results,** pp 9177-9188. https://doi.org/10.47750/pnr.2022.13.S09.1075

MOLTENE, L. e ORSATO, R. J. (2021). La economía compartida en la práctica: Un estudio exploratorio sobre la aceptación y el uso de plataformas digitales para reducir el desperdicio de alimentos. **Revista de la Adm. Empresarial,** v 61, n 5. Doi: https://doi.org/10.1590/S0034-759020210508x

PALAPARTHY, V. S. *et al*. (2020). E-Nose: Multichannel Analog Signal Conditioning Circuit With Pattern Recognition for Explosive Sensing, **IEEE Sensors Journal**, v. 20, n. 3, pp. 1373-1382, doi: 10.1109/JSEN.2019.2946253.

PANDURANGI, V. J. *et al.* (2021). Detection & Classification of Electronic Nose System. **5th International Conference on Intelligent Computing and Control Systems (ICICCS),** pp. 1-4, Madurai, India. doi: 10.1109/ICICCS51141.2021.9432248.

PERA, A. *et al.* (2023). Environmental and economic impacts of improper materials in the recycling of separated collected food waste through anaerobic digestion and composting. **Science in Total Environment**, v 880, Elsevier, https://doi.org/10.1016/j.scitotenv.2023.163240.

PERSAUD, K. e DODD, G. (1982). Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. **Nature**, v 299, pp 352–355, UK.

PLUMBLEY, M. D. e OJA, E. (2004). A "nonnegative PCA" algorithm for independent component analysis. **IEEE Transactions on Neural Networks**, v. 15, n. 1, pp. 66-76. Doi: 10.1109/TNN.2003.820672.

ONU (2024). Objetivos de desenvolvimento sustentável no Brasil. In https://brasil.un.org/pt-br/sdgs. Acessado em dezembro de 2024.

ORĂȘAN,L. I.; SEICULESCU, C. e CĂLEANU, C.D (2022) A Brief Review of Deep Neural Network Implementations for ARM Cortex-M Processor. **Electronics,** v 11*,* pp 2545. https://doi.org/10.3390/electronics11162545

RAHMAN, K.S. *et al.* (2024). Prediction of mango quality during ripening stage using MQ-based electronic nose and multiple linear regression. **Smart Agricultural Technology**, v 9, Elsevier. https://doi.org/10.1016/j.atech.2024.100558.

RIBEIRO, J. S. (2020). Food Waste Indicators. **Commercial Restaurants (Brazil]: Rosa dos Ventos**, 2020, v. 12, n. 2, ISSN: 2178-9061. DOI: https://doi.org/10.18226/21789061.v12i2p350

RUBIN, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. **Hoboken: John Wiley & Sons**. ISBN 9780470316696.

SAILER, G. et al. (2021). Characterization of the separately collected organic fraction of municipal solid waste from rural and urban districts for a one-year period in Germany. **Waste Management**, Elsevier, v. 131 pp 471-482. Doi: https://doi.org/10.1016/j.wasman.2021.07.004.

SANJAYA, W. S. M. *et al*. (2023). Electronic Nose for Tea Identification Detection based on Machine Learning K-Nearest Neighbors Method and Raspberry Pi 4. **2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)**, pp. 55-60, Banda Aceh, Indonesia, doi: 10.1109/COSITE60233.2023.10250144.

SINGH, S. e GAUR, S. (2023). Development of rapid and non-destructive electric nose (E-nose) system for shelf life evaluation of different edible seeds. **Food Chemistry**, v 426, Doi: https://doi.org/10.1016/j.foodchem.2023.136562.

STOCHERO, E. L. M., JACOBI, L e LÚCIO, A. D. (2020). Imputation of data in analysis of variance in experiments in the Completely Randomized Design. Ciência e Natura, v 17, UFSM. DOI: 10.5902/2179460X40446

STRIBIŢCAIA, E. *et al.* (2020). Food texture influences on satiety: systematic review and meta-analysis. **Science Rep.**, v 10, Doi: https://doi.org/10.1038/s41598-020-69504-y

TAN, J. e XU, J. (2020). Applications of electronic nose (e-nose) and electronic tongue (e-tongue) in determining properties related to food quality: a review. **Journal of Artificial Intelligence in Agriculture,** v 4, pp 104-115.

TAN, J. e XU. J. (2022). Development Trend of Electronic Nose Technology in Closed Cabins Gas Detection: a Review. **Applied Sciences**, v *12,* n 18, pp 9326 https://doi.org/10.3390/app12189326

TCHUINKOU, D. e BOBDA, C. (2018). R-covnet: Recurrent neural convolution network for 3d object recognition. **25ª Conferência Internacional do IEEE sobre Processamento de Imagens (ICIP)***,* pp. 331-335.

ULLOA-MURILLO, L.M., *et al*. (2022). Management of the Organic Fraction of Municipal Solid Waste in the Context of a Sustainable and Circular Model: Analysis of Trends in Latin America and the Caribbean. **International Journal Environ. Res. Public Health**, v 19, pp 6041. https://doi.org/10.3390/ijerph19106041

VA, B.; SUBRAMONIAM, M., e MATHEW, L. (2021). Noninvasive detection of COPD and Lung Cancer through breath analysis using MOS Sensor array based e-nose. **Expert Review of Molecular Diagnostics**, v *21*, n 11, pp 1223–1233. DOI: https://doi.org/10.1080/14737159.2021.1971079

VASCONCELOS, A. C. T. *et al.* (2022). Aplicando a abordagem Gestalt como um método para ensinar a prática da ciência da computação em sala de aula: Um estudo de caso em escolas de ensino fundamental no Brasil. **Educação e Tecnologias da Informação**, v. 28, pp. 200-220.

VASCONCELOS. A. C. T. *et al.* (2024) . Utilizing Educational Robotics for learning about Sustainability issues. **19o IEEE Latin American Robotics Symposium (LARS)**, São Bernardo do Campo. Brasil.

VIEJO, C. G. e FUENTES S. (2020). Low-cost methods to assess beer quality using artificial intelligence involving robotics, an electronic nose, and machine learning. **Fermentation Journal**, v 6, n. 4, pp 104.

WANG, H. *et al.* (2022). "Comparison and evaluation of advanced machine learning methods for performance and emissions prediction of a gasoline Wankel rotary engine". **Energy**, Elservier Pub, Vol 248. https://doi.org/10.1016/j.energy.2022.123611.

WILSON A. D. (2012). Review of electronic-nose technologies and algorithms to detect hazardous chemicals in the environment. **Procedia Technology**, v 1, pp 453-463.

Yang, Y. et al. (2021). “Application of E-nose combined with ANN modelling for qualitative and quantitative analysis of benzoic acid in cola-type beverages”. **Journal of Food Measurement and Characterization**, 15, pp 5131-5138.

YE, Z.; LIU, Y. e LI, Q. (2021) Recent Progress in Smart Electronic Nose Technologies Enabled with Machine Learning Methods. **Sensors**, v 21, n 22, pp 7620. DOI: https://doi.org/10.3390/s21227620

WASEEM, M.; LIANG, P. e SHAHIN, M. (2020). A Systematic Mapping Study on Microservices Architecture in DevOps. **Journal of Systems and Software**, v 170, DOI: https://doi.org/10.1016/j.jss.2020.110798.

ZI-JUN, W. *et al*. (2024). Effects of organic material properties on microbial carbon fractions in black soil farmland. **Journal of Plant Nutrition and Fertilizers**, v 30, n 5, pp 980-995. DOI: 10.11674/zwyf.2023526