UNIVERSIDADE FEDERAL DE ITAJUBÁ – UNIFEI

Instituto de Engenharia de Sistemas e Tecnologias da Informação – IESTI

TINYML - APRENDIZADO DE MÁQUINA APLICADO PARA DISPOSITIVOS IOT EMBARCADOS

Reconhecimento de Vogais Através do Uso de Aprendizado de Máquina com Arduino

Bruna Longuim – 2025000668

Gabriel Batista de Oliveira Vilas Boas – 2024000467

Leonardo Sousa Ferreira Fadul – 2021001101

Professor: Marcelo José Rovai

ITAJUBÁ

2025

Introdução

Aprendizado de máquina aplicado para dispositivos embarcados, também chamado de *TinyML*, consiste na criação, modelagem e treinamento de um algoritmo capaz de receber dados (inferências, ou *inputs*) e gerar respostas com base no conteúdo recebido, sendo geralmente usado para uma tarefa específica que exija processamento rápido, privado e seguro, com a capacidade de tomar decisões sem a necessidade de conexão com a nuvem. Neste relatório, é detalhado o método utilizado para a criação de um sistema capaz de reconhecer e classificar vogais, através dos dados obtidos por meio de sensores integrados à placa Arduino Nano 33 BLE Sense. O resultado é um modelo que, após uma inferência, é capaz de classificar o gesto realizado entre 6 classes, além do reconhecimento de anomalias para movimentos que fujam do padrão para qual foi treinado.

Objetivo

O seguinte trabalho teve como objetivo desenvolver um modelo de aprendizado de máquina capaz de reconhecer letras (nesse caso, vogais) por meio de dados capturados pelos sensores acelerômetro e giroscópio, integrados ao kit de *TinyML* disponibilizado para o curso. O seguinte sistema, apesar de sua natureza simples para o trabalho final da disciplina, possui uma vasta gama de funcionalidades para dispositivos presentes no mercado como, por exemplo, a função vista em celulares da linha Samsung Galaxy que, através da *S–Pen*, permitem controlar o dispositivo à distância com o uso de gestos e movimentos que são enviados da caneta ao smartphone.

Metodologia

O modelo foi desenvolvido com o uso da plataforma Edge Impulse, onde foram realizados a coleta de dados, construção da rede neural, treinamento e implantação (*deployment*) do modelo no microcontrolador. Após o *deploy*, as configurações adicionais como o funcionamento dos LEDs RGB e tempo de intervalo entre cada inferência foram realizadas através do software Arduino IDE.

Todos os processos foram realizados em um computador usando o sistema operacional Debian 12, fazendo necessário o uso da ferramenta de linha de comando da Edge Impulse para a comunicação da placa com a plataforma.

Desenvolvimento

Inicialmente foi utilizado o software Edge Impulse CLI para realizar a conexão entre o kit e a plataforma Edge Impulse. A seguir, tendo sido feitas as configurações necessárias para utilizar o módulo IMU (sensor de 9 eixos, usado para medir dados de orientação), procedeu-se com a coleta de dados, tendo sido coletadas amostras de 5 segundos para cada letra. Ao total, foram coletados cerca de 22 minutos de dados, com 29% sendo designados para o teste do modelo.

A criação do *"impulse"* (processo de treinamento) consistiu nos seguintes blocos:

- Leitura dos dados coletados em 9 eixos;
- Bloco de análise espectral que utiliza apenas os dados do acelerômetro e giroscópio;
- Bloco de aprendizado usando um modelo de classificação, para receber os *inputs* e designá-los entre as 6 vogais;
- Características (*features*) de saída contendo as classes: letra A, letra E, letra I, letra O, letra U e parado.

Também foi ajustado o tempo de cada amostra a ser analisado pelo modelo, neste caso escolhido como 3 segundos, para garantir que todo o movimento realizado pelo usuário possa ser devidamente processado.

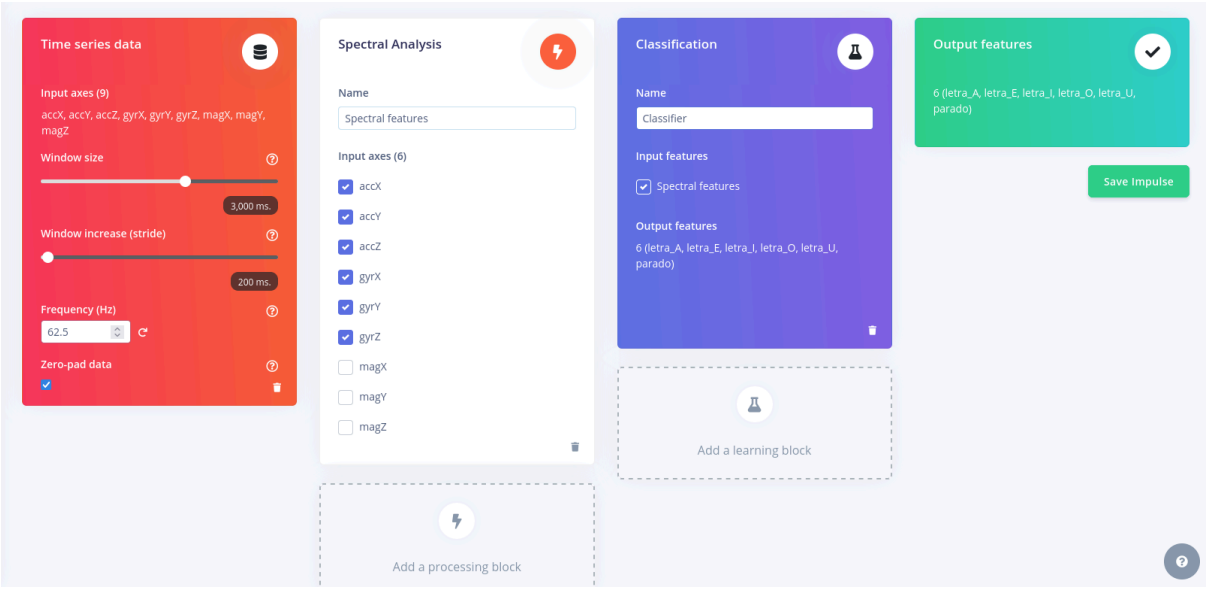

Figura 1: Modelagem do *impulse*.

Em seguida foi realizada a análise dos parâmetros e dos dados "puros", para então gerar os *features* necessários antes do treinamento. Aqui foi possível ver que os "*clusters*" (conjuntos de dados), apesar de muito próximos uns dos outros, ainda se separaram entre classes iguais, o que indica que o modelo poderá reconhecê-los corretamente nos testes.

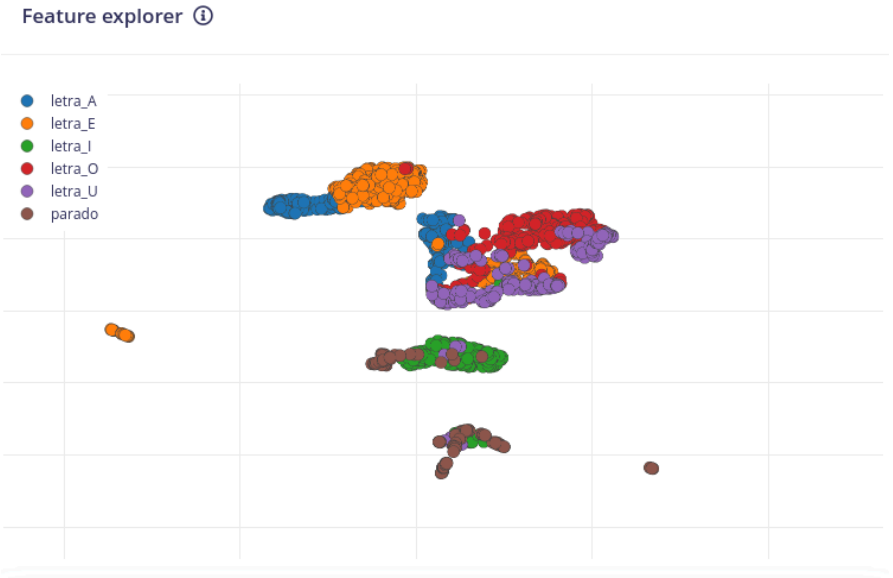

Figura 2: Plotagem dos dados.

O treinamento foi realizado com 50 “épocas” (iterações), uma taxa de aprendizagem de 0.001 e com o uso de uma CPU. A arquitetura da rede neural consistiu em uma camada de dados de entrada com 78 *features*, duas camadas do tipo *dense*, cada uma contendo 20 e 10 neurônios, respectivamente, e uma camada de saída contendo as 6 classes. O resultado foi um modelo com 95,7% de precisão, e uma matriz de confusão com poucos falsos negativos.

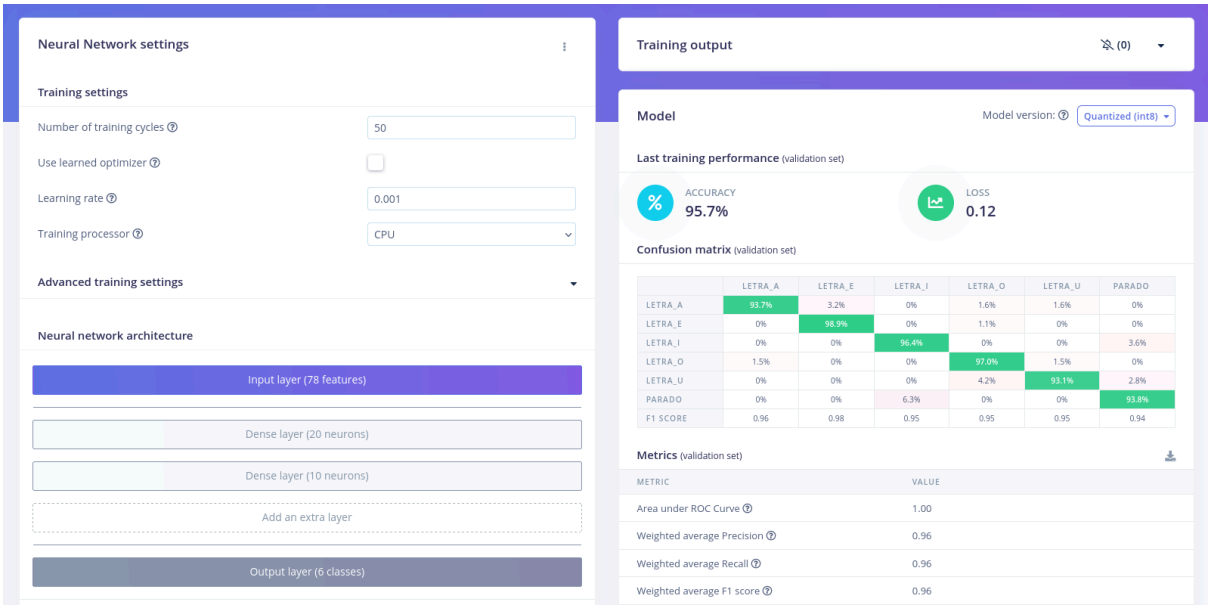

Figura 3: Configuração da rede neural e matriz de confusão.

Figura 4: Plotagem dos resultados pós-treinamento..

Resultados

Por fim, o desempenho em dispositivo, utilizando o compilador EON (proprietário da Edge Impulse) consistiu em um tempo de inferência de 1 milissegundo, uso máximo de memória RAM de 1,4 kilobytes e aproximadamente 15,8 kilobytes de armazenamento em memória flash. Utilizando a ferramenta de testes, antes de fazer o *deploy* para a placa, o modelo acertou 8 das 10 inferências realizadas, obtendo um desempenho considerado satisfatório pela equipe.

Enfim, para realizar o *deployment* e testar de forma prática com o kit, foi feito o download de todos os códigos necessários convertidos em bibliotecas para a linguagem Arduino. Vale ressaltar que, ao fazer o download da biblioteca, o modelo passa pelo processo de quantização, cujo objetivo é reduzir o tamanho e custo computacional de modelos de aprendizado de máquina. Neste caso, o formato passou de utilizar valores com ponto flutuante (*float32*) para valores inteiros (*int8*), garantindo desempenho no uso real.

Conclusão

Em conclusão, o projeto pôde demonstrar a viabilidade e eficácia da aplicação de modelos *TinyML*, destacando seu desempenho e abordagem livre de conexão com a nuvem para obter resultados de forma prática e confiável. Ademais, evidenciou-se que, mesmo com recursos computacionais limitados, é possível realizar tarefas complexas como o reconhecimento de padrões em tempo real, com baixo consumo de memória e energia.

O uso da plataforma Edge Impulse facilitou todas as etapas de desenvolvimento, desde a coleta de dados até a implantação do modelo, permitindo uma experiência prática de integração entre hardware e inteligência artificial. O modelo desenvolvido, com alta taxa de precisão e baixo tempo de inferência, demonstrou um ótimo desempenho no reconhecimento de vogais com base em gestos, apontando para possíveis aplicações em interfaces gestuais, acessibilidade e controle de dispositivos por movimentos.

Além disso, os resultados obtidos ao longo da execução reforçaram o bom desempenho do modelo em condições controladas e práticas. A acurácia de 80% nas inferências prévias ao deploy e o tempo de resposta de apenas 1 milissegundo no dispositivo validam a eficiência do sistema em tempo real. O baixo consumo de recursos — 1,4 KB de RAM e cerca de 15,8 KB de flash — evidencia a adequação do modelo ao ambiente embarcado, confirmando a viabilidade técnica da proposta e a capacidade do sistema em operar de forma autônoma, leve e responsiva.

Por fim, o projeto cumpriu com êxito os objetivos propostos, reforçando o potencial do aprendizado de máquina embarcado em soluções inovadoras e acessíveis para o cotidiano.

Referências

REDDI, Vijay Janapa (org.). **Machine Learning Systems**: Principles and Practices of Engineering Artificially Intelligent Systems. Harvard University, 10 jun. 2025. Disponível em: https://mlsysbook.ai/.