UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO (EMC)
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO (PPGEEC)

CAROLINA SOUZA FLORIANO

Sistema de Comunicação Alternativa para Pessoas com Distúrbios Neuromotores Severos usando Redes Neurais Artificiais

GOIÂNIA
2023

UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[X] Dissertação [ ] Tese [ ] Outro*:_____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Carolina de Souza Floriano

3. Título do trabalho

Sistema de Comunicação Alternativa para Pessoas com Distúrbios Neuromotores Severos usando Redes Neurais Artificiais

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;

- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Leonardo Da Cunha Brito**, **Professor do Magistério Superior**, em 12/01/2024, às 11:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Carolina Souza Floriano**, **Discente**, em 15/01/2024, às 10:58, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4311918** e o código CRC **2C781B07**.

Referência: Processo nº 23070.066748/2023-41 SEI nº 4311918

Carolina Souza Floriano

Sistema de Comunicação Alternativa para Pessoas com Distúrbios Neuromotores Severos usando Redes Neurais Artificiais

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação (PPGEEC), da Escola de Engenharia Elétrica, Mecânica e de Computação (EMC), da Universidade Federal de Goiás (UFG), como requisito parcial para a obtenção do Título de Mestre em Engenharia Elétrica e de Computação.
Área de Concentração: Engenharia de Computação.
Linha de pesquisa: Sistemas Inteligentes e Computação Aplicada

Orientador: Professor Dr. Leonardo da Cunha Brito
Coorientador: Professor Dr. Adson Rocha Silva

Goiânia
2023

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Floriano, Carolina Souza
Sistema de Comunicação Alternativa para Pessoas com Distúrbios Neuromotores Severos usando Redes Neurais Artificiais [manuscrito] / Carolina Souza Floriano. - 2023.
98 f.: il.

Orientador: Prof. Dr. Leonardo da Cunha Brito; co-orientador Dr. Adson Silva Rocha.
Dissertação (Mestrado) - Universidade Federal de Goiás, Escola de Engenharia Elétrica, Mecânica e de Computação (EMC), Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Goiânia, 2023.
Bibliografia.
Inclui siglas, fotografias, abreviaturas, símbolos, gráfico, tabelas, lista de figuras, lista de tabelas.

1. Redes neurais artificiais. 2. Comunicação alternativa. 3. Comunicação aumentativa. 4. Distúrbios neuromotores. 5. Tecnologia Assistiva. I. Brito, Leonardo da Cunha, orient. II. Título.

CDU 62+004+005

UNIVERSIDADE FEDERAL DE GOIÁS

ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº **11** da sessão de Defesa de Dissertação de **Carolina de Souza Floriano**, que confere o título de Mestra em **Engenharia Elétrica e de Computação**, na área de concentração em **Engenharia de Computação**.

Aos **quinze dias do mês de dezembro de dois mil e vinte e três**, a partir das **14h00min.**, realizou-se a sessão pública de Defesa de Dissertação intitulada **"Sistema de Comunicação Alternativa para Pessoas Portadoras de Distúrbios Neuromotores Severos usando Redes Neurais Artificiais"**. Os trabalhos foram instalados pelo Orientador, Professor Doutor **Leonardo da Cunha Brito - (EMC/UFG)**, com a participação dos demais membros da Banca Examinadora: Professor Doutor **Adson Silva Rocha - (IFGoiano)** membro titular externo e Professor Doutor **Renato de Sousa Gomide - (IFGoiano)** membro titular externo; **cujas participações ocorreram através de videoconferência** pelo link da videochamada: https://meet.google.com/ayy-ewgi-hzm. Durante a arguição os membros da banca **fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sida a candidata **aprovada** pelos seus membros. Proclamados os resultados pelo Professor Doutor **Leonardo da Cunha Brito**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos **quinze dias do mês de dezembro de dois mil e vinte e três**.

TÍTULO SUGERIDO PELA BANCA

Sistema de Comunicação Alternativa para Pessoas com Distúrbios Neuromotores Severos usando Redes Neurais Artificiais

Documento assinado eletronicamente por **Leonardo Da Cunha Brito**, **Professor do Magistério Superior**, em 15/12/2023, às 15:37, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Renato de Sousa Gomide**, **Usuário Externo**, em 15/12/2023, às 15:37, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Adson Silva Rocha**, **Usuário Externo**, em 15/12/2023, às 15:38, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Carolina Souza Floriano**, **Discente**, em 15/12/2023, às 15:45, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Referência: Processo nº 23070.066748/2023-41 SEI nº 4256169

Dedico este trabalho ao meu amado pai, Cícero Antônio Floriano, cujas lições de vida continuam a guiar meu caminho, mesmo em sua ausência física.

Pai, suas palavras de encorajamento ecoam nos corredores da minha mente, e é com gratidão profunda que dedico este trabalho a você. Seus ensinamentos sobre a importância do conhecimento e da perseverança são a essência desta jornada. Tenho a certeza de que, onde quer que esteja, você está orgulhoso, como sempre esteve, do que me tornei e das conquistas que alcançamos juntos.

Agradeço por ser meu guia, meu maior apoiador e a voz que sempre sussurrará: "Você é capaz". Este mestrado é uma homenagem ao seu legado, à sua crença inabalável em meu potencial e à influência duradoura que você teve em minha vida.

Agradecimentos

Gostaria de expressar minha sincera gratidão a todas as pessoas que tornaram possível a realização deste trabalho e me acompanharam ao longo dessa jornada desafiadora.

Em primeiro lugar, dedico este agradecimento à minha família, cujo apoio incondicional e compreensão foram fundamentais para que eu pudesse me dedicar inteiramente à minha pesquisa. Meu marido, Christian Douglas, e meu filho, Aaron, foram fontes constantes de incentivo e paciência, tornando cada desafio mais fácil de superar.

À minha mãe, Elizabeth, por seu apoio incansável e por ser a minha maior fonte de incentivo, mesmo nos momentos em que duvidava de mim mesma. Seu amor e confiança foram a luz que guiou cada passo desta jornada acadêmica.

Ao meu orientador, Leonardo, e ao meu coorientador, Adson, gostaria de expressar minha profunda gratidão. Suas orientações, insights e contribuições foram inestimáveis para o desenvolvimento deste trabalho. Obrigada por compartilharem seu conhecimento e por dedicarem seu tempo e energia para me guiar ao longo deste percurso acadêmico.

Ao corpo docente, colegas de mestrado e a todos os envolvidos no processo educacional, agradeço por proporcionarem um ambiente estimulante e enriquecedor. Cada interação e discussão contribuíram significativamente para o meu aprendizado e crescimento profissional.

Por fim, quero expressar minha profunda gratidão a todos que, de alguma forma, contribuíram para o meu caminho durante o mestrado. Este trabalho não seria possível sem o apoio e a colaboração de cada um de vocês.

Obrigada por fazerem parte desta jornada.

Resumo

Dificuldades de comunicação são frequentes para muitas pessoas com deficiências motoras graves, tornando difícil para elas interagir com suas famílias, cuidadores e a sociedade em geral. A Comunicação Aumentativa e Alternativa (CAA) tem como objetivo compensar o déficit de comunicação dessas pessoas, proporcionando uma melhor qualidade de vida ao indivíduo. No entanto, esses indivíduos com distúrbios neuromotores graves e restrições severas de movimento enfrentam grandes desafios no uso de várias tecnologias assistivas atuais. Nesse contexto, o objetivo deste trabalho é apresentar um Sistema de Comunicação Alternativa baseado em Redes Neurais Artificiais com uma abordagem centrada no usuário e suas necessidades para uso por esse público. A entrada e o processamento dos sinais são realizados pela leitura dos pontos de referência facial, utilizando a biblioteca MediaPipe FaceMesh, e o desenvolvimento do classificador de gestos/expressões faciais é realizado por meio da implementação e comparação dois modelos diferentes, um Modelo de Redes Neurais Convolucionais (CNN) e um Modelo de Redes Neurais Recorrentes, usando unidades de memória de longo prazo (LSTM) e camadas densas. Foram implementados desafios dinâmicos para realizar uma análise mais aprofundada do desempenho dos modelos em diversos contextos, variando parâmetros como a quantidade de amostras e a inserção de gestos semelhantes. Os resultados globais em tempo real apontam para um desempenho consistente do sistema proposto, sugerindo que em ambas as abordagens, a Rede Neural Convolucional (CNN) destaca-se significativamente em relação à Rede Neural Recorrente de Longa Memória (LSTM) no reconhecimento de gestos.

Palavras-chave: redes neurais artificiais, comunicação alternativa, comunicação aumentativa, distúrbios neuromotores, tecnologia assistiva.

Abstract

Communication difficulties are frequent for many people with severe motor disabilities, making it difficult for them to interact with their families, caregivers and society in general. Augmentative and Alternative Communication (AAC) then aims to compensate for the communication deficit of these people, providing the individual with a better quality of life. However, these individuals with severe neuromotor disorders who have severe movement restrictions find great challenges in the use of several current assistive technologies. In this context, the objective of this research is to present an Alternative Communication System based on Artificial Neural Networks with a user-centered approach and their needs, for use by this public. The input and signal processing are carried out by reading facial landmark points, using the MediaPipe FaceMesh library. The development of the gesture/facial expression classifier is performed through the implementation and comparison of two different models: a Convolutional Neural Network (CNN) model and a Recurrent Neural Network model using Long Short-Term Memory (LSTM) units and dense layers. Dynamic challenges were implemented to conduct a more in-depth analysis of the models' performance in various contexts, varying parameters such as the quantity of samples and the inclusion of similar gestures. Real-time overall results indicate a consistent performance of the proposed system, suggesting that, in both approaches, the Convolutional Neural Network (CNN) stands out significantly compared to the Long Short-Term Memory Recurrent Neural Network (LSTM) in gesture recognition.

Keywords: artificial neural networks, alternative communication, augmentative communication, neuromotor disorders, assistive technology.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

TA	*Tecnologia Assistiva*
ADA	*American with Disabilities Act*
CAA	*Comunicação Aumentativa e Alternativa*
ASHA	*Associação Americana de Fala, Linguagem e Audição*
ELA	*Esclerose Lateral Amiotrófica*
EOG	*Eletrooculografia*
BCI	*Brain Computer Interface*
SUS	*System Usability Scale*
SVM	*Máquina de Vetores de Suporte*
RNA	*Rede Neural Artificial*
CNN	*Rede Neural Convolucional*
ReLU	*Rectified Linear Unit*
RNN	*Rede Neural Recorrente*
LSTM	*Long Short-Term Memory*
VP	*Verdadeiros Positivos*
VN	*Verdadeiros Negativos*
FP	*Falsos Positivos*
FN	*Falsos Negativos*
ROC	*Curva Característica de Operação do Receptor*

Sumário

1 Introdução

Dificuldades de comunicação são frequentes para diversas pessoas com deficiência motora severa, dificultando a sua interação com os seus familiares, cuidadores e sociedade em geral.

O desenvolvimento e a inovação da tecnologia são responsáveis por dispor a sociedade uma série de benefícios, como melhorias na qualidade de vida, a facilidade e agilidade no acesso a informações e conhecimentos, a simplificação de mecanismos de comunicação, a otimização e automatização de processos, e inclusive uma contribuição significativa para a para a área da saúde.

Além disso, a tecnologia tem papel essencial na promoção da inclusão social, apresentando o conceito de tecnologia assistiva, cujo objetivo é auxiliar na resolução de problemas funcionais encontrados por pessoas com alguma deficiência, ou agilizar e promover habilidades essenciais do cotidiano dessas pessoas (BERSCH, 2009).

Para atender a necessidade de interação de alguns indivíduos com restrições de movimentos e atividades motoras com dispositivos de comunicação, é cada vez mais comum o uso de novas modalidades de interação como entrada por gestos ou objetos tangíveis, comandos de voz e diversos outros tipos de sensores (SILVA; VEIGA, 2021).

Atualmente existem um grande número e variedade de interfaces assistivas, e estudos recentes relacionados à Comunicação Aumentativa e Alternativa (CAA) buscam otimizar dispositivos já existentes, procurando diminuir a fadiga visual, aumentar a velocidade de digitação e encontrar formas mais eficientes de entrada de informação através do movimentos dos olhos ou direção do olhar (SILVA; VEIGA, 2021).

A estimativa da prevalência de utilizadores de CAA é uma tarefa difícil à grande variedade de pessoas presentes nessa população. Para tal, normalmene são utilizados fatores como diagnóstico, idade, localização, modalidade de comunicação e extensão do uso de CAA. Em geral, quanto maior for o défice de comunicação de um indivíduo, maior será a probabilidade de ele se beneficiar do apoio da CAA. Estudos indicam que aproximadamente 97 milhões de pessoas em todo o mundo podem se beneficiar da CAA ((ASHA), 2022).

1.1 Motivação

A comunicação é um fator básico na vida de todo e qualquer indivíduo, sendo um separador de águas na forma como a pessoa se desenvolve mentalmente desde seus primeiros anos de vida, também ditando a forma de como nós, enquanto seres sociaveis, somos aceitos e/ou nos inserimos na sociedade.

A pesquisa nacional conduzida nos Estados Unidos pelos autores (ANDZIK et al., 2018) entre educadores especiais revelou que 18,2% dos alunos que necessitam de apoio para se comunicar fazem uso de Comunicação Assistida por Computador (CAA). Dentro desse grupo, 6,9% recorrem a modos gestuais, 6,5% fazem uso de suportes pictóricos, e 4,8% utilizam dispositivos geradores de fala. Além disso, a Pesquisa Nacional de Crianças com Necessidades Especiais de Saúde, realizada nos anos de 2009 e 2010, indicou que 4% das crianças com deficiências de desenvolvimento e 10,5% das crianças com necessidades especiais de cuidados de saúde nos EUA não tiveram suas necessidades de comunicação atendidas por meio de tecnologia assistiva (LIN; GOLD, 2018).

No Reino Unido, estima-se que 0,5% da população necessite do uso de CAA com base em condições associadas a essa necessidade. As maiores populações que poderiam se beneficiar da AAC têm diagnósticos de Alzheimer/demência, doença de Parkinson, transtorno do espectro do autismo, dificuldades de aprendizagem e acidente vascular cerebral (CREER et al., 2016). Além disso, uma pesquisa com prestadores de serviços de dispositivos de geração de fala no Reino Unido constatou que cerca de 0,0155% dos indivíduos utilizam recursos de comunicação motorizados (JUDGE et al., 2017).

Em populações pediátricas específicas, (IACONO; TREMBATH; ERICKSON, 2016) estimaram que 25% a 30% das crianças australianas com autismo tenham habilidades de fala limitadas e se beneficiariam da CAA, enquanto (KRISTOFFERSSON; SANDBERG; HOLCK, 2020) observou que 44,4% das crianças suecas com paralisia cerebral utilizam alguma forma de CAA. Além disso, (BROWN; GRAMES; SKOLNICK, 2021) indentificou que 6,9% das crianças norte-americanas com fissura de palato ou anomalias craniofaciais utilizam CAA.

Em relação a pacientes hospitalizados, estudos mostraram que uma porcentagem significativa de pacientes na unidade de terapia intensiva (UTI) atende aos critérios para o uso de AAC, assim como um número considerável de pacientes fora da UTI (ZUBOW; HURTIG, 2013). No entanto, mesmo em situações em que a comunicação é difícil, os pacientes utilizam modos alternativos de comunicação com pouca frequência (FREEMAN-SANDERSON; MORRIS; ELKINS, 2019).

No caso de adultos com esclerose lateral amiotrófica (ELA), uma pesquisa identificou que 17,3% desses pacientes adquirem equipamentos de CAA para melhorar a fala, substituir a fala ou fornecer suporte para a comunicação escrita (ELLIOTT et al., 2020). Na Alemanha, 46% dos pacientes com ELA demonstraram necessidade de CAA, mas 39% não conseguiram acesso a dispositivos adequados (FUNKE et al., 2018).

1.2 Justificativa

Desde os primórdios da existência humana, nós fomos caracterizados pela nossa habilidade em desenvolver novas tecnologias e, junto a elas, novas ferramentas que tornassem as tarefas cotidianas mais fáceis. Durante muito tempo, a maioria dessas ferramentas eram simples e diretas, como martelos ou copos, o que garantia uma interação simplificada entre o usuário e a ferramenta (VOIGT et al., 2018).

Com o avanço tecnológico, surgiram e se popularizaram os microcomputadores, ferramentas formidáveis capazes de nos auxiliar em diversas tarefas com finalidades distintas. No entanto, uma ferramenta tão complexa pode exigir uma interação usuário-ferramenta igualmente complexa.

Inicialmente, os comandos para os computadores eram enviados por meio de cartões perfurados. Mas a introdução de novos dispositivos, como teclados, mouses e telas sensíveis ao toque, facilitou a interação com essas máquinas, tornando-a mais simples, precisas e acessíveis.

Atualmente, os comandos de voz já são uma realidade em todos os smartphones. Embora os comandos visuais, baseados na detecção de características faciais, ainda não sejam tão acessíveis e utilizadas, há pesquisas e trabalhos na área buscando expandir suas aplicações.

Assim como os comandos de voz permitem a interação com os computadores enquanto nos concentramos em outras tarefas, uma interação visual, permitiria um controle rápido e preciso sobre várias atividades. Além disso, proporcionaria o acesso a essas facilidades para usuários com deficiência auditiva ou de fala, abrindo a possibilidade de comunicação entre eles e outras pessoas.

Um grande desafio nesse sentido é a detecção precisa dessa interação. Nos últimos anos, diversas novas tecnologias foram desenvolvidas e popularizadas, principalmente na captura dos movimentos, como o Microsoft Kinect, o Intel RealSense 3D ou o Leap Motion (VOIGT et al., 2018; BILESAN et al., 2019; JÚNIOR; KNOP, 2015; NARCIZO; CÂMARA, 2013; HAUSAMANN et al., 2021; VASCONCELOS, 2017; MELO, 2015; SOUSA et al., 2017). No entanto, mesmo com essas tecnologias, o reconhecimento de movimentos específicos é complexo, pois é necessário tolerar pequenas variações em cada caso, uma vez que duas pessoas diferentes inevitavelmente executarão o mesmo movimento de forma ligeiramente diferente, e mesmo duas execuções do mesmo indivíduo raramente serão idênticas.

Além disso, esses dispositivos de captura ainda apresentam suas próprias dificuldades, como a dificuldade em rastrear corretamente as partes do corpo humano quando estão obstruídas por objetos, incluindo outras partes do corpo que estão sendo rastreadas.

Enquanto isso, outras tecnologias também avançaram significativamente nos últimos anos. Especialmente, o aumento da capacidade computacional tornou o uso de Redes Neurais Profundas uma realidade, mesmo em modelos mais complexos. As Redes Neurais Convolucionais são capazes de categorizar imagens com uma precisão altíssima, diferenciando objetos ou cenários distintos com muita confiabilidade. Além disso, Redes Neurais Recorrentes tornaram possível a avaliação de sequências de dados com tamanhos arbitrários.

1.3 Objetivos do Trabalho

1.3.1 Objetivo Geral

O objetivo desta pesquisa consiste em desenvolver um sistema de comunicação alternativa direcionado a indivíduos com distúrbios neuromotores severos. Para isso, será utilizada a tecnologia de redes neurais, que irá permitir a coleta de gestos personalizados identificados através de pontos faciais dos usuários. A partir desses dados, um modelo classificador eficiente será treinado, tendo a capacidade de estimar de forma precisa os gestos realizados pelos usuários. Essa solução possibilitará que as pessoas interajam com o sistema utilizando esses gestos, proporcionando uma forma de comunicação mais acessível e efetiva para aqueles que enfrentam distúrbios neuromotores severos.

1.3.2 Objetivos Específicos

Para atingir o objetivo geral, foram definidos alguns pontos necessários. Esses pontos foram destacados como objetivos ou tarefas específicas para alcançar resultados satisfatórios. Mais detalhadamente, esses objetivos podem ser delineados como:

- Pesquisar e analisar os problemas enfrentados pelas pessoas com distúrbios neuromotores severos no uso das tecnologias de comunicação assistivas atuais;
- Elaborar a arquitetura do sistema proposto, visando proporcionar um meio de comunicação alternativa para o público que não consegue usar as tecnologias atuais;
- Pesquisar e definir as melhores arquiteturas para cada módulo do sistema proposto, bem como suas conexões;
- Implementar o sistema de captura de amostras que constituem o conjunto de dados de treinamento;
- Treinar os modelos classificadores em diferentes arquiteturas e técnicas;
- Verificar a eficácia dos modelos treinados no reconhecimento dos gestos faciais previamente cadastrados.

1.4 Metodologia da Pesquisa

De acordo com os objetivos estabelecidos neste trabalho, a metodologia a ser seguida é composta por quatro etapas. Na primeira etapa, foi realizado um levantamento dos principais sistemas de comunicação alternativa de alta tecnologia atuais. Em seguida, a literatura é analisada através de uma revisão sistemática a fim de detectar e entender quais públicos não são atendidos por estes sistemas e sua motivação.

Na segunda etapa, as ideias para o desenvolvimento do sistema são elaboradas utilizando como referência as pesquisas realizadas na etapa anterior. Em seguida, é elaborada uma proposta de arquitetura do sistema, e um estudo exploratório é realizado a fim de validar os principais conceitos utilizados.

Na terceira etapa, a proposta do sistema é implementada, inicialmente por meio do sistema que fará a captura de gestos que comporão o conjunto de dados, que servirá como base para o treinamento da rede neural. Em seguida, o treinamento da rede neural é realizado através de diferentes arquiteturas gerando os modelos classificadores que servirão como base para os experimentos e validação do sistema proposto.

Na quarta e última etapa, foram realizados experimentos para análise da eficácia e comparação dos modelos classificadores gerados, com o objetivo de validar o sistema proposto.

1.5 Estrutura deste Trabalho

Com base nos objetivos estabelecidos, a dissertação está organizada em seis capítulos. O Capítulo 1 apresenta a problemática e a justificativa da pesquisa, seus objetivos e uma descrição da metodologia adotada no decorrer do trabalho.

O Capítulo 2 apresenta os conceitos relacionados à tecnologia assistiva, comunicação aumentativa e alternativa, software assistivo, usabilidade e os principais distúrbios neuromotores que podem levar à perda severa de movimentos, dificultando ou até impossibilitando a comunicação deste público. Além disso, apresenta uma revisão sistemática da literatura dos principais estudos relacionados à proposta desse trabalho.

O Capítulo 3 apresenta os conceitos relacionados à inteligência artificial, aprendizado de máquina, redes neurais artificiais e arquiteturas de aprendizado profundo no contexto de estimação de classificação de dados a partir de dados sequenciais.

O Capítulo 4 apresenta a proposta desse trabalho, que consiste na metodologia de desenvolvimento do sistema proposto para a identificação de gestos faciais personalizados através de uma câmera, transformando esses gestos em ações para que o usuário realize interações no sistema e possa se comunicar.

O Capítulo 5 apresenta uma análise dos experimentos que servirão como validação da proposta. Por fim, o Capítulo 6 apresenta a conclusão da dissertação.

2 Tecnologia Assistiva

2.1 Introdução

O termo Tecnologia Assistiva (Assistive Technology) foi criado oficialmente em 1988, como importante elemento jurídico dentro da legislação norte-americana, conhecida por Public Law 100-407, que compõe, com outras leis, o ADA - American with Disabilities Act. Este conjunto de leis regula os direitos dos cidadãos com deficiência nos EUA, além de prover a base legal dos fundos públicos para compra dos recursos que estes necessitam. Houve a necessidade de regulamentação legal deste tipo de tecnologia (TA) e, a partir desta definição e do suporte legal, a população norte-americana, de pessoas com deficiência, passa a ter garantido pelo seu governo o benefício de serviços especializados; bem como o acesso a todo o arsenal de recursos que necessitam e que venham favorecer uma vida mais independente, produtiva e incluída no contexto social geral (GARCIA; FILHO, 2012).

Em (XIII, 1999), o termo tecnologia assistiva foi definido de acordo com o conceito do American with Disabilities Act (ADA) como sendo "uma ampla gama de equipamentos, serviços, estratégias e práticas concebidas e aplicadas para minorar os problemas funcionais encontrados pelos indivíduos com deficiências".

A tecnologia assistiva tem se tornado cada vez mais relevante e necessária na sociedade contemporânea. Através da utilização de tecnologias inovadoras, como dispositivos eletrônicos, softwares e equipamentos especializados, a tecnologia assistiva busca oferecer suporte e auxílio para aqueles que enfrentam desafios na realização de atividades diárias. É um campo de estudo que se dedica ao desenvolvimento e aplicação de produtos, dispositivos, equipamentos e sistemas voltados para pessoas com deficiência ou mobilidade reduzida. Seu principal objetivo é promover a inclusão e a autonomia dessas pessoas, proporcionando-lhes maior independência e qualidade de vida (NISBET, 2019).

Essa área da tecnologia busca soluções e recursos que possam auxiliar indivíduos com diferentes tipos de limitações a se comunicarem, se locomoverem, obterem informações e realizarem diversas tarefas. É um campo multidisciplinar que envolve conhecimentos técnicos, de engenharia, design, psicologia, pedagogia, entre outros.

A tecnologia assistiva tem se mostrado uma área de pesquisa em rápido crescimento, com o potencial de melhorar significativamente a qualidade de vida de pessoas com deficiência. Sobre o tema podemos destacar alguns pontos.

O primeiro ponto é a definição de tecnologia assistiva: qualquer dispositivo ou equipamento que ajude uma pessoa com deficiência a realizar atividades que seriam desafiadoras sem auxílio. Isso pode incluir desde simples adaptações de utensílios domésticos

até dispositivos altamente avançados, como próteses robotizadas.

Outro aspecto importante é a variedade de deficiências para as quais a tecnologia assistiva pode ser aplicada. Ela pode beneficiar pessoas com deficiências físicas, sensoriais, cognitivas e até mesmo pessoas com transtornos do espectro autista. Essa diversidade de aplicações é fundamental para atender as necessidades individuais de cada pessoa.

Uma das principais vantagens da tecnologia assistiva é a capacidade de promover a independência das pessoas com deficiência. Por exemplo, dispositivos de mobilidade, como cadeiras de rodas motorizadas, permitem que as pessoas se locomovam com mais facilidade e autonomia. Além disso, a tecnologia assistiva pode melhorar a comunicação e proporcionar acesso a informações e recursos, reduzindo a exclusão social. Cada pessoa tem necessidades e habilidades diferentes, e é essencial que os dispositivos sejam adaptados de acordo com suas especificidades. Isso pode envolver a personalização de interfaces, ajustes de parâmetros ou até mesmo a criação de dispositivos sob medida (BERSCH, 2018).

Um dos desafios enfrentados pela tecnologia assistiva é o alto custo dos dispositivos. Muitas inovações são muito caras, o que dificulta o acesso de pessoas com menos recursos. No entanto, o artigo relata o surgimento de iniciativas e projetos de código aberto, que visam tornar a tecnologia assistiva mais acessível e inclusiva.

Alguns exemplos de tecnologia assistiva incluem próteses, órteses e dispositivos de comunicação alternativa, como as pranchas de comunicação visual. Também são utilizados sistemas de controle de ambiente, como o acionamento de luzes, portas automáticas e outros equipamentos por meio de comandos de voz ou sensores de movimento.

Além disso, existem softwares e aplicativos que apoiam pessoas com dificuldades de aprendizado, como leitores de texto e programas de reconhecimento de voz. Há ainda dispositivos que facilitam a mobilidade, como cadeiras de rodas motorizadas e equipamentos de locomoção assistida.

Os principais objetivos da tecnologia assistiva são:

- Promover a inclusão social: ao oferecer recursos e dispositivos que permitem às pessoas com deficiência participarem plenamente da sociedade, a tecnologia assistiva busca reduzir as barreiras e discriminações existentes.
- Aumentar a autonomia: ao fornecer meios de superar as limitações e realizar atividades do dia a dia de forma autônoma, a tecnologia assistiva busca proporcionar mais independência às pessoas com deficiência.
- Melhorar a qualidade de vida: oferecendo soluções que facilitem o acesso a serviços, comunicação, educação e trabalho, a tecnologia assistiva contribui para a melhoria da qualidade de vida das pessoas com deficiência.

- Potencializar habilidades: ao contar com recursos que permitem o desenvolvimento de habilidades específicas, a tecnologia assistiva possibilita que as pessoas com deficiência possam explorar seu potencial máximo.

Considerando como objetivo principal das Tecnologias de Apoio o uso de tecnologias que ajudem a ultrapassar as limitações funcionais dos seres humanos num contexto social, é de extrema importância identificar não só os aspectos puramente tecnológicos, mas também os aspectos relacionados com os fatores humanos e socioeconômicos. Um modelo de formação e treino em tecnologias de apoio deve ser baseado num modelo de desenvolvimento humano que tenha em consideração os problemas que as pessoas com deficiência apresentam quando tentam adaptar-se a um ambiente adverso (XIII, 1999).

A Tecnologia Assistiva e seus recursos deve ser distinguida de outras tecnologias que são aplicadas nas áreas médicas e na reabilitação, sendo entendida como "recurso do usuário" em contraste ao "recurso do profissional", uma vez que a Tecnologia Assistiva tem a finalidade de promover a autonomia e a eficiência na execução das tarefas cotidianas dos usuários (BERSCH, 2018).

A tecnologia assistiva desempenha um papel fundamental na inclusão de pessoas com deficiência, contribuindo para a igualdade de oportunidades e a garantia dos direitos humanos. Ela está em constante evolução, buscando desenvolver soluções cada vez mais eficientes e acessíveis, visando o bem-estar e a inclusão de todos.

O desenvolvimento de novas soluções e aprimoramento de tecnologias existentes são cruciais para atender às necessidades em constante evolução das pessoas com deficiência. Em resumo, a tecnologia assistiva é uma área em expansão que visa melhorar a qualidade de vida das pessoas com deficiência. Ela pode promover a independência, facilitar a comunicação e proporcionar acesso a recursos importantes. Apesar dos desafios, a tecnologia assistiva tem o potencial de transformar a vida de muitas pessoas, oferecendo soluções personalizadas e acessíveis.

Portanto, conclui-se que o objetivo de uma tecnologia assistiva é proporcionar a uma pessoa com deficiência uma maior independência, a promoção de melhorias em sua qualidade de vida e inclusão social, mediante o favorecimento ou ampliação de sua comunicação, mobilidade, independência e desenvolvimento de habilidades.

2.2 Comunicação Aumentativa e Alternativa (CAA)

A Comunicação Aumentativa e Alternativa (CAA) é uma subárea da Tecnologia Assistiva que representa um amplo conjunto de estratégias, métodos e ferramentas cujo propósito é aumentar a capacidade comunicativa de pessoas que não conseguem utilizar a comunicação verbal de forma funcional e eficaz.

A CAA foi descrita pela Associação Americana de Fala, Linguagem e Audição (ASHA) como o "esforço de estudar e quando necessário compensar deficiências temporárias ou permanentes, limitações de atividades e restrições de participação de pessoas com distúrbios graves na produção e/ou compreensão da linguagem falada ou escrita"((ASHA), 2022).

A comunicação aumentativa e alternativa (CAA) é um campo da fonoaudiologia que busca auxiliar pessoas que têm dificuldades na comunicação verbal ((ASHA), 2022). É uma abordagem que utiliza diferentes estratégias e técnicas para permitir que esses indivíduos se expressem, interajam e se comuniquem de forma efetiva. Segundo (LOJA et al., 2015), a comunicação aumentativa e alternativa (CAA) é uma abordagem utilizada para auxiliar pessoas com limitações na fala ou na linguagem a se comunicarem de forma efectiva. Essa abordagem é composta por diversos componentes que são essenciais para o seu funcionamento adequado.

Uma das principais utilizações da CAA é ajudar pessoas com limitações físicas ou mentais que têm dificuldades em utilizar a fala como principal meio de comunicação. Ela pode ser aplicada em diferentes faixas etárias e em uma variedade de condições, como paralisia cerebral, autismo, acidente vascular cerebral, distúrbios neuromusculares, entre outros ((ASHA), 2022).

De forma geral, a CAA compreende duas finalidades principais com dois públicos-alvo diferentes: (a) melhorar a capacidade comunicativa de pessoas com fala ininteligível ou não-fluente, e (b) proporcionar uma forma de comunicação alternativa para pessoas que não podem falar ou são incapazes de desenvolver uma linguagem capaz de realizar uma comunicação eficiente (SILVA, 2021).

A utilização dos recursos de CAA é indicada, normalmente, para pessoas com capacidade de fala ilimitada ou nula, sendo que essa perda pode ser causada por diversas condições médicas e de incapacidade (congênitas, adquiridas, progressivas ou temporárias) (SILVA, 2021).

Ainda, é importante destacar que a CAA não se limita apenas à comunicação oral. Ela abrange diferentes modalidades de comunicação, como a escrita, gestos, linguagem de sinais e outras formas de expressão. O objetivo principal da CAA é fornecer meios alternativos e complementares de comunicação para que as pessoas com limitações na fala possam se expressar, se conectar com o mundo ao seu redor e participar plenamente da sociedade (LOJA et al., 2015).

2.2.1 Estratégias de Comunicação

A CAA utiliza uma ampla gama de estratégias de comunicação, que podem incluir desde comunicadores de papel e lápis, até dispositivos eletrônicos especializados. Essas

estratégias podem ser divididas em duas categorias principais: comunicação auxiliada e suplementada ((ASHA), 2022).

A comunicação auxiliada envolve o uso de gestos, expressões faciais, linguagem corporal e sistemas de comunicação não verbal, como sistemas de símbolos ou representações gráficas, para complementar a comunicação verbal. Por exemplo, um indivíduo pode utilizar um livro de comunicação com figuras e símbolos gráficos para expressar suas necessidades ou preferências.

Já a comunicação suplementada envolve o uso de dispositivos eletrônicos, como computadores, tablets ou outros dispositivos de comunicação específicos. Esses dispositivos são programados para permitir que o usuário selecione e combine símbolos, palavras ou frases predefinidas para formar mensagens, que são vocalizadas ou exibidas na tela do dispositivo.

2.2.2 Componentes da CAA

Como já foi elucidado em topicos anteriores, a Comunicação Alternativa e Ampliada (CAA) é uma abordagem utilizada para auxiliar pessoas com dificuldades de comunicação. Ela engloba um conjunto de estratégias e técnicas, incluindo o uso de recursos e dispositivos, que visam facilitar a expressão e a compreensão da linguagem por meio de meios alternativos à fala (SILVA et al., 2021). Tendo em vista elucidar sobre alguns dos componentes que compoem um aparato que se utilize da CAA podemos citar:

- Interface de entrada: É o canal pelo qual a pessoa interage com o dispositivo ou recurso de CAA. Pode ser um teclado físico, um mouse, uma tela sensível ao toque, um joystick, entre outros.
- Módulo de reconhecimento e processamento: É responsável por interpretar os comandos ou entradas realizadas pelo usuário na interface de entrada. Pode utilizar técnicas de reconhecimento de fala, processamento de texto, inteligência artificial, entre outros, para transformar as ações do usuário em informações compreensíveis.
- Banco de dados de símbolos: É a base de informações que o sistema utiliza para representar palavras, frases, símbolos e outros elementos de linguagem. Pode incluir uma biblioteca de símbolos pictográficos, ícones, sons ou até mesmo representações textuais.
- Recursos de síntese de voz: São componentes responsáveis por transformar o texto ou os símbolos gerados pelo sistema em voz sonora. Através de algoritmos de síntese de fala, é possível proporcionar à pessoa uma forma de expressão vocal. Interface de saída: É o meio pelo qual as informações geradas pelo sistema são apresentadas ao

usuário. Pode ser uma tela de computador, um dispositivo auditivo, uma impressora em Braille, entre outros.

- Módulo de adaptação: É responsável por ajustar as configurações e características do sistema de acordo com as necessidades individuais da pessoa. Pode permitir a personalização de símbolos, a criação de atalhos, a modificação da velocidade da voz, entre outras customizações.

Em suma, a CAA é um campo de estudo e prática que busca oferecer alternativas de comunicação para pessoas com dificuldades de fala ou linguagem verbal. Os componentes descritos nos tópicos anteriores são exemplos de recursos utilizados na CAA, visando atender às necessidades e potencialidades de cada indivíduo.

2.2.3 Tecnologias de Apoio para a Comunicação

Um dos componentes centrais da CAA é o sistema de símbolos. Este sistema consiste em representações visuais, como símbolos gráficos, fotografias, desenhos ou formas gestuais, que são utilizadas para transmitir mensagens. Esses símbolos podem ser organizados em sistemas de comunicação alternativos, como placas com imagens, pranchas de símbolos ou sistemas de comunicação eletrônicos, que permitem que a pessoa possa selecionar os símbolos e formar suas frases (LOJA et al., 2015).

Outro componente importante são os auxílios de comunicação. Estes são dispositivos ou estratégias que ajudam as pessoas a se comunicarem de forma mais eficiente. Isso pode incluir desde pranchas de comunicação simples, até comunicação por meio de computadores, tablets ou outros dispositivos eletrônicos. Os auxílios de comunicação podem ser adaptados às necessidades e habilidades específicas de cada indivíduo (LOJA et al., 2015).

Além disso, a CAA também envolve estratégias de ensino e treinamento. Profissionais da área trabalham junto com as pessoas que utilizam a CAA para ensinar como usar os sistemas de símbolos e auxílios de comunicação, ajudando-as a desenvolver habilidades de comunicação e aprimorar sua capacidade de expressão. Essas estratégias podem incluir sessões de treinamento individualizadas, jogos interativos, atividades de modelagem de linguagem e suporte contínuo.

2.2.4 Público-Alvo

Entre as condições congênitas que podem estar relacionadas a necessidades complexas de comunicação estão a paralisia cerebral, a deficiência intelectual, a afasia e o autismo. Já com relação aos distúrbios adquiridos, a progressão ou a gravidade de algumas doenças podem levar a necessidade de usar mecanismos da CAA para se comunicar, como a esclerose lateral amiotrófica, doença do neurônio motor, a síndrome do encarceramento,

entre outras. Em condições progressivas, pode-se citar doenças como a distrofia muscular, e em condições temporárias, a utilização de CAA pode ocorrer em momentos pós-cirúrgicos ou ainda em situações que ocasionam a perda temporária da fala.

A escolha da estratégia de CAA mais adequada depende das necessidades e habilidades individuais de cada pessoa, bem como de fatores ambientais e contextuais. Além disso, é fundamental que haja uma avaliação cuidadosa das habilidades de comunicação e uma avaliação das capacidades cognitivas, motoras e linguísticas do indivíduo, a fim de desenvolver um plano personalizado de CAA ((ASHA), 2022).

2.3 Software Assistivo e Usabilidade

Existem diferentes abordagens para avaliar a usabilidade em softwares assistivos, sendo algumas:

- Métricas heurísticas: As métricas heurísticas são avaliações realizadas por especialistas em usabilidade com base em um conjunto de diretrizes ou princípios de usabilidade. Essas métricas são utilizadas para identificar problemas de usabilidade no software assistivo, como dificuldades na interação ou na compreensão das funcionalidades (SPOLSKY, 2008).

- Testes de usabilidade: Os testes de usabilidade envolvem a observação direta de usuários do software assistivo enquanto eles realizam tarefas específicas. Esses testes permitem identificar problemas de usabilidade e coletar métricas relacionadas a eficiência, eficácia, satisfação do usuário, entre outros aspectos (KRUG, 2009).

- Questionários de satisfação do usuário: Esses questionários buscam medir a satisfação geral dos usuários com o software assistivo, bem como aspectos específicos da usabilidade, como a facilidade de aprendizado, a eficiência de uso e a qualidade de ajuda e documentação. O questionário System Usability Scale (SUS) é uma referência amplamente utilizada nesse tipo de métrica e foi proposto por (BROOKE, 1996).

- Métricas de desempenho: As métricas de desempenho são utilizadas para medir aspectos relacionados à eficiência do uso do software assistivo, como o tempo necessário para realizar tarefas específicas, a quantidade de erros cometidos pelos usuários e a taxa de conclusão de tarefas (ALBERT; TULLIS, 2022).

2.4 Distúrbios Neuromotores

Os distúrbios neuromotores representam qualquer condição patológica onde existe a perda da função ou desestruturação das células nervosas, podendo ser denominados

distúrbios neuromotores ou ainda distúrbios neuromusculares (COELHO et al., 2016). Estes distúrbios, podem provocar restrições severas de comunicação, abrangendo a comunicação verbal e também a linguagem corporal.

Dentre os distúrbios neurodegenerativos mais graves está a Esclerose Lateral Amiotrófica (ELA) , a paralisia cerebral, tetraplegia, e a síndrome do encarceramento, que representa o caso mais extremo, pois o indivíduo perde praticamente todas as suas funções motoras, com exceção do movimento dos olhos (LOJA et al., 2015).

2.4.1 Esclerose Lateral Amiotrófica

De acordo com (BERTAZZI et al., 2017), a esclerose lateral amiotrófica (ELA) é uma doença degenerativa do sistema nervoso que afeta as células nervosas responsáveis pelo controle dos movimentos voluntários dos músculos. A ELA é caracterizada pela morte progressiva dos neurônios motores, resultando na perda da capacidade de controlar os músculos do corpo.

A incidência da ELA varia geograficamente, com taxas mais altas encontradas na América do Norte e Europa. A doença afeta geralmente pessoas na faixa dos 40 aos 70 anos, embora possa ocorrer em qualquer idade.

Os sintomas da ELA geralmente começam de forma sutil, com fraqueza muscular, cãibras, fadiga e dificuldades na fala, mastigação e deglutição. Com a progressão da doença, a fraqueza muscular se espalha para os membros superiores, inferiores e afeta também os músculos respiratórios. Eventualmente, a ELA leva à perda completa da capacidade de controlar os músculos, resultando em paralisia.

A progressão da doença é variável, mas geralmente é contínua e rápida. A ELA é uma doença fatal, com uma expectativa de vida média de 2 a 5 anos a partir do diagnóstico. No entanto, existem casos de sobrevida mais longa, e os sintomas e a velocidade de progressão podem variar entre os indivíduos.

2.4.2 Paralisia Cerebral

A paralisia cerebral é uma condição neurológica que afeta o controle motor do indivíduo, causando limitações na coordenação, equilíbrio e movimentação. Ela é resultado de uma lesão no cérebro que ocorre durante o desenvolvimento fetal ou nos primeiros anos de vida (LOJA et al., 2015).

Esta lesão no cérebro, muitas vezes, é decorrente de problemas durante a gestação, como infecções, desnutrição ou trauma. Além disso, complicações no parto, como falta de oxigênio, também podem ser um fator de risco.

Existem diferentes tipos de paralisia cerebral, que variam de acordo com a área

do cérebro afetada e a intensidade dos sintomas. Alguns indivíduos podem apresentar dificuldades apenas em um lado do corpo, enquanto outros podem ter acometimento em todo o corpo.

Os sintomas da paralisia cerebral podem incluir espasmos musculares, rigidez muscular, dificuldades de coordenação motora, alterações na fala, dificuldades para engolir, entre outros. Além disso, muitas pessoas com paralisia cerebral podem apresentar deficiências intelectuais ou distúrbios do desenvolvimento.

2.4.3 Tetraplegia

A tetraplegia é uma condição médica em que há uma perda de movimento e sensação nos membros superiores e inferiores, bem como no tronco. Também conhecida como paraplegia cervical, a tetraplegia é causada por danos à medula espinhal do pescoço para baixo (BALL; FAGER; FRIED-OKEN, 2012).

Existem várias causas possíveis para a tetraplegia, incluindo lesões traumáticas, como quedas ou acidentes de carro, doenças degenerativas, como a esclerose lateral amiotrófica (ELA) ou a esclerose múltipla, e complicações médicas, como infecções da coluna vertebral.

Uma das possíveis consequências da tetraplegia é a perda da fala. Isso ocorre porque a lesão na medula espinhal afeta os nervos responsáveis pelo controle dos músculos envolvidos na fala. Além disso, dependendo do nível da lesão, a tetraplegia pode afetar os músculos responsáveis pela respiração e pela produção de sons da fala (BALL; FAGER; FRIED-OKEN, 2012).

A perda da fala pode variar de pessoa para pessoa, algumas ainda conseguindo produzir alguns sons, enquanto outras podem ter dificuldade em falar ou perder completamente a capacidade de falar. Além disso, a tetraplegia também pode afetar a capacidade de articular palavras devido à fraqueza muscular nos músculos faciais e da boca.

Felizmente, existem opções de comunicação alternativa para pessoas com tetraplegia que perderam a capacidade de falar. Os dispositivos de tecnologia assistiva, como computadores controlados por voz e sistemas de comunicação por meio de luzes ou gráficos, podem ser utilizados para ajudar a pessoa a se comunicar com o mundo ao seu redor. Além disso, a terapia de fala e linguagem também pode ser útil para melhorar a articulação e a compreensão da fala.

No entanto, é importante destacar que cada caso de tetraplegia é único, e a possibilidade de perda da fala variará de acordo com a extensão e a localização da lesão. É fundamental que as pessoas com tetraplegia recebam cuidados especializados e uma equipe multidisciplinar para melhorar sua qualidade de vida e ajudá-las na comunicação.

Em suma, a tetraplegia é uma condição médica séria que pode resultar em perda da fala devido à lesão na medula espinhal. No entanto, com o avanço da tecnologia e os cuidados adequados, é possível encontrar maneiras alternativas de se comunicar e melhorar a qualidade de vida das pessoas afetadas por essa condição.

2.4.4 Síndrome do Encarceramento

Segundo (SMITH; DELARGY, 2005), a Síndrome do Encarceramento é um estado neurológico raro no qual o paciente fica completamente paralisado e incapaz de se comunicar, mas mantém a consciência e a capacidade de pensamento intactas. Essa condição resulta de uma lesão ou doença que afeta a conexão entre o cérebro e os músculos responsáveis pelos movimentos voluntários, mantendo o paciente aprisionado em seu próprio corpo.

A incidência da Síndrome do Encarceramento não é bem estabelecida, mas estima-se que seja extremamente rara, afetando apenas 1-2 pessoas por milhão de habitantes (SMITH; DELARGY, 2005). A condição pode ocorrer em qualquer idade, mas é mais comum em adultos jovens.

Os sintomas da Síndrome do Encarceramento incluem a completa paralisia dos músculos do corpo, exceto os movimentos oculares verticais e/ou piscar dos olhos. Os pacientes são totalmente dependentes de outras pessoas para realizar tarefas diárias, como comer, se vestir e se locomover. No entanto, sua capacidade cognitiva e de compreensão permanece preservada.

A progressão da Síndrome do Encarceramento é variável e depende da causa subjacente da condição. Em alguns casos, pode haver uma melhora gradual ao longo do tempo, permitindo que o paciente recupere algum controle motor. No entanto, em outros casos, a paralisia pode ser permanente e progressiva, levando a complicações como infecções respiratórias e dificuldade em respirar.

Segundo (BAUER; GERSTENBRAND; RUMPL, 1979), a Síndrome do Encarceramento, pode ser classificada em três subtipos diferentes:

- Síndrome do Encarceramento Clássica: nesse tipo, o paciente apresenta paralisia completa, exceto pelos movimentos oculares verticais e/ou o piscar dos olhos. A consciência e a capacidade de pensamento estão preservadas.
- Síndrome do Encarceramento Incompleta: nesse tipo, o paciente ainda possui algum controle motor além dos movimentos oculares. Isso pode incluir movimentos limitados dos membros, como os dedos das mãos.
- Síndrome do Encarceramento Total: nesse tipo, o paciente fica completamente paralisado e incapaz de realizar qualquer tipo de movimento, além dos movimentos

oculares. A comunicação se torna extremamente difícil e a dependência de outras pessoas é total.

2.5 Sistemas de Comunicação Aumentativa e Alternativa de Alta Tecnologia

O uso da Comunicação Aumentativa e Alternativa (CAA) em todas as doenças neuromusculares é limitado, mas os estudos em Esclerose Lateral Amiotrófica sugerem que o uso de um sistema CAA é normalmente bem aceito e promove melhorias na qualidade de vida (BALL; FAGER; FRIED-OKEN, 2012). O avanço da tecnologia está tornando os sistemas CAA mais acessíveis, de forma que, tem grande potencial para estender o acesso à comunicação até mesmo para os pacientes mais graves como doenças neuromusculares.

O estudo realizado em (OLIVEIRA, 2019) analisou a aplicabilidade dos sistemas CAA na Esclerose Lateral Amiotrófica na perspectiva dos doentes, cuidadores e profissionais de saúde, demonstrando que a percepção da utilidade destes sistemas é elevada por estes, e também identificou alguns fatores relevantes que contribuem para aumentar a utilidade desses sistemas.

Em (HWANG et al., 2014), os autores realizaram um estudo para analisar se um dispositivo de rastreamento ocular melhora a qualidade de vida dos pacientes, e também, analisar se este dispositivo afeta a carga experimentada pelos cuidadores. A pesquisa mostrou que o dispositivo auxiliar de rastreamento ocular melhorou significativamente a qualidade de vida dos pacientes quando comparados com pacientes que não fizeram o uso do dispositivo ($p < 0{,}01$), além disso, o dispositivo assistivo também reduziu a sobrecarga dos cuidadores ($p < 0{,}05$).

O desenvolvimento de um teclado virtual assistivo foi realizado em diversos estudos como em (SILVA; VEIGA, 2021; XIII, 1999; SILVA, 2021; GOMIDE et al., 2016), onde, além da criação do teclado assistivo, estes trabalhos tinham por objetivo otimizar seus resultados, ou melhorando a velocidade de digitação, ou realizando uma redução de erros e ou minimizando o esforço necessário realizado pelo usuário.

Em (KÄTHNER; KÜBLER; HALDER, 2015) foi realizada uma avaliação entre três métodos de acesso à comunicação aumentativa e alternativa, a eletrooculografia (EOG) , um rastreador ocular e uma interface auditiva cérebro-computador (Brain Computer Interface - BCI) por um indivíduo com síndrome do encarceramento. Foi demonstrado então, que essa pessoa foi capaz de utilizar todas as três interfaces propostas, porém, consideraria usar apenas o BCI, pois logo perderia o controle do músculo dos olhos devido a doença. Os autores concluíram então que abordagens centradas no usuário no desenvolvimento desses sistemas aumentará a probabilidade de que estes sejam usados

como tecnologia assistiva na vida diária.

O trabalho (ZHANG; KULKARNI; MORRIS, 2017) apresentou o desenvolvimento de um sistema de comunicação de gestos oculares, o GazeSpeak, que funciona em um smartphone e é projetado para ser de baixo custo, robusto, portátil e de fácil aprendizagem, e com uma maior largura de banda de comunicação que uma placa e-tran *(eye-gaze transfer)*. E sua avaliação final demonstrou que o GazeSpeak é robusto, apresentou uma boa satisfação do usuário e é capaz de fornecer uma melhoria de velocidade em relação a uma placa e-tran, além de melhorias adicionais como baixo custo e esforço.

Já em (COELHO et al., 2016) foi apresentado o desenvolvimento de um novo sistema de Comunicação Aumentativa e Alternativa (CAA) controlado pelo movimento dos olhos do usuário integrado a um aplicativo móvel para permitir uma comunicação a distância com o cuidador e familiares. O sistema foi testado por 17 voluntários, que avaliaram a usabilidade, de forma que, os resultados mostraram que o sistema foi classificado pelos usuários como excelente, com uma pontuação média de 92,79 na System Usability Scale (SUS) citecoelho2016novo.

3 Redes Neurais Artificiais

Ao longo desse trabalho são utilizados conceitos relacionados às áreas de Aprendizado de Máquina. Em um primeiro momento, serão abordados sobretudo os conceitos referentes à Redes Neurais, uma metodologia da técnica de Aprendizagem Supervisionada, posteriormente serão abordadas arquiteturas de redes neurais utilizadas na classificação de dados sequenciais.

3.1 Aprendizado de Máquina

De acordo com (LUDERMIR, 2021), o aprendizado de máquina (ou Machine Learning) é um campo da Inteligência Artificial que diz respeito ao desenvolvimento de algoritmos e técnicas que permitem que os computadores aprendam a partir de dados, sem serem explicitamente programados. O objetivo é capacitar as máquinas a tomar decisões precisas ou fazer previsões a partir desses dados.

O artigo apresenta três principais tipos de aprendizado de máquina:

- Aprendizado supervisionado: É o tipo mais comum, no qual existe um conjunto de dados anotado previamente por especialistas. O algoritmo aprende a partir desses exemplos rotulados, sendo capaz de classificar ou prever novos dados. Alguns exemplos de algoritmos supervisionados são a regressão linear, árvores de decisão e redes neurais.
- Aprendizado não supervisionado: Neste tipo, o algoritmo é exposto a um conjunto de dados não rotulados, sendo seu objetivo identificar padrões, agrupamentos ou relações ocultas presentes neles. Exemplos de algoritmos não supervisionados incluem o K-means, agrupamento hierárquico e análise de componentes principais.
- Aprendizado por reforço: Neste tipo de aprendizado, o algoritmo aprende a partir das ações e das recompensas ou punições recebidas em um ambiente virtual simulado. O objetivo é desenvolver uma política que maximize a recompensa ao tomar decisões. Algoritmos de aprendizado por reforço são frequentemente utilizados em jogos e robótica.

3.2 Aprendizado Supervisionado

A aprendizagem supervisionada é uma técnica de aprendizado de máquina que envolve o treinamento de um modelo com base em um conjunto de dados rotulados. O

modelo é alimentado com exemplos de entrada e saída, e ajusta seus parâmetros para minimizar o erro entre a saída prevista e a saída real. Uma vez treinado, o modelo pode ser usado para fazer previsões sobre novos dados (PAIXAO et al., 2022).

A aplicação da aprendizagem supervisionada é ampla e variada, e pode ser usada em uma variedade de contextos, como reconhecimento de fala, reconhecimento de imagem, detecção de fraudes, análise de sentimentos e muito mais.

O processo de Aprendizagem Supervisionada é bastante natural ao ser humano, pois grande parte de nosso próprio aprendizado é feito da mesma forma, tanto quando aprendemos diretamente de um professor quando usamos livros escritos por especialistas, e principalmente durante nossos primeiros anos de vida.

Muito antes de aprendermos as diferenças essenciais entre diferentes tipos de frutas, nossos familiares e pessoas ao nosso redor nos ensinam quais são maçãs, quais são bananas, quais são laranjas. Apenas com essa informação, somos capazes de "criar uma função f" capaz de diferenciá-las, mesmo que a maioria de nós sequer estude com profundidade as diferenças morfológicas entre elas.

Os problemas abordados pela Aprendizagem Supervisionada podem ser divididos em dois tipos: problemas de Classificação e problemas de Regressão. Problemas de Classificação se preocupam em separar os exemplos dados em grupos diferentes, com características semelhantes. Por outro lado, problemas de Regressão se preocupam em prever uma resposta numérica real.

Por exemplo, prever a quantidade de vendas de um produto baseado em sua popularidade e preço, ou prever a quantidade de pessoas que assistirão a um filme baseado em sua classificação e gênero, são problemas de regressão, que oferecem como valores de $f(x) = y$ qualquer valor numérico válido dentro de um certo alcance, dependendo do problema.

Por outro lado, prever se um paciente tem uma doença com base em seus sintomas, ou prever se um e-mail é spam ou não com base em seu conteúdo, são problemas de classificação, que possuem respostas $f(x) = y$ como valores discretos, como "Doença A" ou "Doença C", ou "sim" ou "não".

O propósito deste projeto é desenvolver um sistema que possa identificar qual gesto facial foi realizado diante da câmera, sendo um problema claramente discreto. Existem vários algoritmos disponíveis que seguem esses conceitos, cada um com suas vantagens e desvantagens, desde a simples Regressão linear até casos mais complexos como a Máquina de Vetores de Suporte (SVM) , Aprendizagem baseada em Árvores de Decisão ou Redes Neurais Artificiais. Portanto, a partir de agora, o foco será em problemas de Classificação e uso de Redes Neurais Artificiais.

3.3 Redes Neurais Artificiais (e deep learning)

As Redes Neurais Artificiais (RNAs) são sistemas computacionais inspirados no funcionamento do sistema nervoso humano. O cérebro humano é composto por cerca de 100 bilhões de neurônios que se comunicam entre si através de sinais eletroquímicos. Esses neurônios são conectados por meio de junções chamadas sinapses, e cada neurônio recebe milhares de conexões com outros neurônios, recebendo constantemente sinais para chegar ao corpo da célula. Se a soma resultante dos sinais ultrapassar um determinado limite, uma resposta é enviada através do axônio (ACADEMY, 2022).

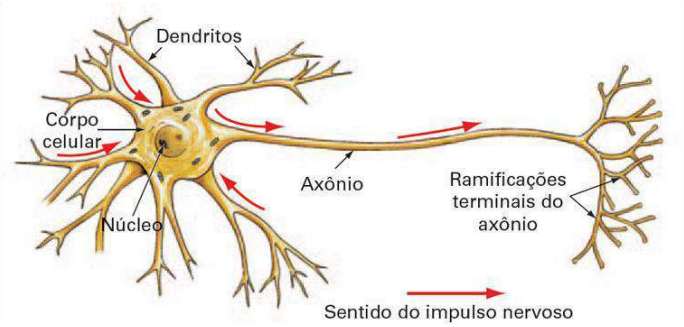


Figura 3.1 – Representação Simplificada do Neurônio Biolóogico

Fonte: (ACADEMY, 2022)

Pesquisadores tentaram imitar esse sistema em computadores, baseando-se na estrutura e funcionamento do neurônio biológico. O modelo mais popular foi proposto por Warren McCulloch e Walter Pitts em 1943 (MCCULLOCH; PITTS, 1943), e ele representa de forma simplificada os componentes e a operação de um neurônio biológico. Em resumo, um neurônio artificial é um elemento que realiza a soma ponderada de múltiplas entradas, aplica uma função e encaminha o resultado adiante.

Neste modelo de neurônio artificial, os sinais elétricos que vêm de outros neurônios são representados pelos chamados sinais de entrada $(x_1, x_2, x_3, ..., x_D)$, que são os dados que alimentam o modelo de rede neural artificial. Dentre os vários estímulos recebidos, alguns excitarão mais e outros menos o neurônio receptor. Essa medida de quão excitatório é o estímulo é representada no modelo de Warren McCulloch e Walter Pitts através dos pesos sinápticos w_D. Quanto maior o valor do peso, mais excitatório é o estímulo.

A soma ou corpo da célula é representada por uma composição de dois módulos. O primeiro é uma junção aditiva, somatório dos estímulos (sinais de entrada) multiplicado pelo seu fator excitatório (pesos sinápticos). Posteriormente, uma função de ativação definirá com base nas entradas e pesos sinápticos qual será a saída do neurônio. O axônio é aqui representado pela saída y obtida pela aplicação da função de ativação. Assim como no modelo biológico, o estímulo pode ser excitatório ou inibitório, representado pelo peso sináptico positivo ou negativo respectivamente.

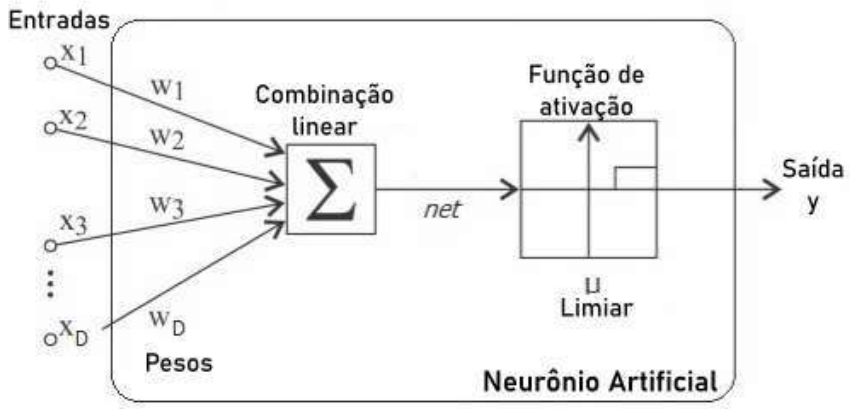


Figura 3.2 – Representação Simplificada do Neurônio Artificial

Fonte: (SILVA et al., 2022)

As Redes Neurais Artificiais são compostas por um conjunto de nós, conhecidos como neurônios artificiais, que são interconectados por camadas (RASCHKA; MIRJALILI, 2019). Cada neurônio em uma rede neural recebe um sinal como entrada, processa-o e o repassa para os neurônios subsequentes, como demonostrado na Figura 3.3.

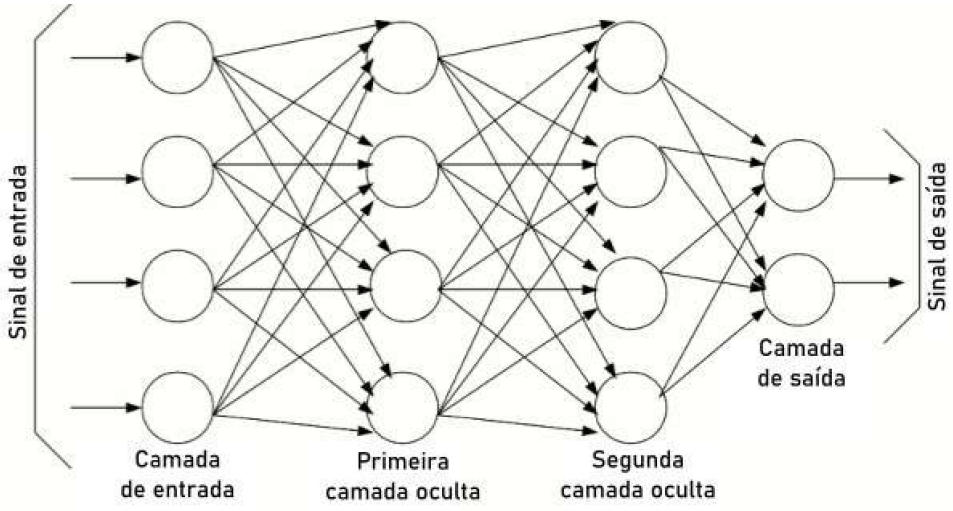


Figura 3.3 – Grafo Representativo de uma Rede Neural Artificial

Fonte: Autor, 2023

O conceito básico por trás das redes neurais artificiais é a capacidade de aprendizado. Assim como os neurônios do cérebro humano, os neurônios artificiais podem aprender e adaptar seus pesos sinápticos com base em dados de entrada. Essa habilidade de aprendizado é o que permite que as redes neurais artificiais sejam aplicadas em uma variedade de domínios, como reconhecimento de padrões, processamento de fala, visão computacional e muito mais.

Uma das vantagens das redes neurais artificiais é sua capacidade de lidar com problemas complexos e não-lineares. A estrutura em camadas permite que as redes neurais artificiais capturem relações e interações complexas entre as variáveis de entrada. Além disso, as redes neurais podem generalizar os padrões aprendidos para resolver problemas não vistos anteriormente.

Uma das arquiteturas mais comuns de redes neurais artificiais é a chamada rede neural profunda, ou *deep neural network*. Essa arquitetura é composta por diversas camadas ocultas, o que permite que a rede aprenda diversas representações hierárquicas dos dados de entrada. Esse tipo de rede neural profunda tem sido muito utilizado em aplicações de inteligência artificial, como reconhecimento de imagens, tradução automática, veículos autônomos e muito mais (RASCHKA; MIRJALILI, 2019).

No entanto, a eficiência e eficácia das redes neurais artificiais dependem de um bom treinamento e ajuste dos parâmetros. É necessário um conjunto de dados de treinamento adequado, além de uma escolha cuidadosa da arquitetura da rede e do algoritmo de aprendizado. Além disso, as redes neurais artificiais também podem exigir uma quantidade significativa de poder computacional para processar grandes quantidades de dados.

Em resumo, as redes neurais artificiais são poderosas ferramentas no campo da inteligência artificial, capazes de aprender e resolver problemas complexos. Com suas capacidades de aprendizado, as redes neurais artificiais têm o potencial de revolucionar uma ampla gama de aplicações, tornando-as uma área de pesquisa e desenvolvimento em constante evolução.

3.3.1 Backpropagation e Método do Gradiente Descendente

De maneira similar a outras técnicas de Aprendizagem Supervisionada, as Redes Neurais buscam definir uma função $f : X \to Y$ que possa mapear corretamente os exemplos do banco de dados, que são elementos do conjunto X, para os resultados esperados, que são elementos do conjunto Y. Essa função é explicitamente definida por meio de um conjunto específico de coeficientes $C = \{c_1, c_2, ..., c_p\}$, cuja ordem p depende diretamente da estrutura da rede.

3.3.1.1 Gradiente Descentente

O Método do Gradiente Descendente é um algoritmo de otimização utilizado para encontrar o mínimo global de uma função. Na aplicação em redes neurais, o objetivo é minimizar o erro da rede em relação aos dados de treinamento. O algoritmo calcula o gradiente da função de erro em relação aos pesos da rede neural e, então, ajusta os pesos em direção ao mínimo local, seguindo o gradiente descendente. Esse processo é repetido iterativamente até que um critério de parada seja alcançado (PINHEIRO, 2023).

O vetor gradiente de uma função $\phi : U \subset \mathbb{R}^m \to \mathbb{R}$ é um vetor $v \in \mathbb{R}^m$ cujas coordenadas correspondem às derivadas parciais da função: $\nabla\phi(x) = (\frac{\partial\phi}{\partial x_1}(x), ..., \frac{\partial\phi}{\partial x_m}(x))$. Uma das propriedades mais notáveis dele é a sua interpretação geométrica: quando avaliado em um ponto qualquer do domínio da função, a direção do gradiente indica a direção em que a imagem da função cresce com maior intensidade.

3.3.1.2 Backpropagation

O Backpropagation é um algoritmo de aprendizado supervisionado que é usado para ajustar os pesos da rede neural artificial. Ele é baseado na minimização da soma do erro quadrático usando o método do gradiente descendente (PINHEIRO, 2023).

O backpropagation é considerado a pedra angular das redes neurais modernas e do Deep Learning. Ele foi originalmente introduzido na década de 1970, mas sua importância não foi totalmente apreciada até um famoso artigo de 1986 de David Rumelhart, Geoffrey Hinton e Ronald Williams. O backpropagation pode ser considerado o algoritmo mais importante na história das redes neurais, sem o qual seria quase impossível treinar redes de aprendizagem profunda da forma que vemos hoje (ACADEMY, 2022).

3.3.2 Redes Neurais Convolucionais (CNN)

Neste contexto de avanços incessantes em inteligência artificial, as Redes Neurais Convolucionais (CNNs) emergem como uma peça central na transformação da análise de dados visuais. As CNNs representam uma evolução significativa das redes neurais convencionais, apresentando uma arquitetura especializada para lidar eficientemente com dados bidimensionais, como imagens (ACADEMY, 2022).

A análise das camadas convolucionais, camadas de pooling e camadas totalmente conectadas se torna essencial para a compreensão profunda das CNNs. Este estudo detalhado visa não apenas fornecer uma visão abrangente desses componentes fundamentais, mas também estabelecer as conexões entre teoria e aplicação prática, delineando os princípios subjacentes que moldam o funcionamento dessas redes.

3.3.2.1 Convolução e Convolução Discreta

A Convolução e a subsequente Convolução Discreta desempenham um papel central nas Redes Neurais Convolucionais (CNNs), representando a espinha dorsal de sua capacidade inigualável na análise de dados espaciais, especialmente em tarefas de visão computacional.

A convolução revela-se como a chave para a eficácia das CNNs na extração de características de dados bidimensionais, como imagens. Ela opera através de filtros ou kernels, que deslizam pela entrada, capturando padrões locais e construindo representações

hierárquicas de características. Este processo é essencial para a capacidade das CNNs de reconhecer padrões complexos e realizar tarefas sofisticadas (ACADEMY, 2022).

A Convolução, como conceito matemático, é definida como uma operação comutativa entre duas funções f e g, representada pelo operador $*$, que produz uma terceira função $f * g$. Essa terceira função representa a soma dos produtos entre as funções ao longo da região formada pela superposição delas. Esse resultado, além da operação em si, também é denominado Convolução.

A operação é definida formalmente como a integral sobre o produto de uma das funções com uma cópia deslocada e espelhada da outra. Essa operação pode ser percebida como a média ponderada da primeira função, com os pesos dados pela segunda função, depois de deslocada e espelhada:

$$\begin{aligned}(f * g)(t) &= \int_{-\infty}^{+\infty} f(\tau) \cdot g(t-\tau)\, d\tau \\ &= \int_{-\infty}^{+\infty} g(\tau) \cdot f(t-\tau)\, d\tau = (g * f)(t)\end{aligned} \tag{3.1}$$

Apesar de ter uso em diversas áreas de pesquisa, como Probabilidade, Estatística, Equações Diferenciais e Processamento de Sinais, para usar a Convolução em um sistema computacional, é necessário convertê-la para o universo discreto. Essa simplificação é feita por meio da convolução discreta, que é uma variação da convolução aplicada a sinais discretos (VOIGT et al., 2018). Tem-se então a seguinte simplificação:

$$\begin{aligned}(f * g)(n) &= \sum_{j=-k}^{k} f(j) \cdot g(n-j) \\ &= \sum_{j=-k}^{k} g(j) \cdot f(n-j) = (g * f)(n)\end{aligned} \tag{3.2}$$

Portando, a Convolução Discreta, por sua vez, é o processo específico pelo qual a convolução é aplicada na prática dentro das camadas convolucionais das CNNs. Este procedimento consiste em convoluir os dados de entrada com os pesos do filtro, somar os resultados e aplicar uma função de ativação. A convolução discreta é o alicerce sobre o qual as CNNs aprendem a identificar padrões, destacando características cruciais em diferentes níveis de abstração.

3.3.3 Filtros Convolucionais

Para tornar a notação mais simples, impõe-se uma condição razoável sobre a função g: ela deve ser simétrica. Isso significa que para qualquer valor j, tem-se que $g(j)$ é igual a $g(-j)$. Com essa condição, a partir da Equação 3.2, pode-se prosseguir com o seguinte

raciocínio:

$$\begin{aligned}(g * f)(n) &= \sum_{j=-k}^{k} g(j) \cdot f(n-j) \\ &= \sum_{j=-k}^{k} g(-j) \cdot f(n+j) \\ &= \sum_{j=-k}^{k} g(j) \cdot f(n+j)\end{aligned} \tag{3.3}$$

Para a utilização da operação de convolução em imagens, é preciso adaptá-la para um espaço bidimensional. Esse procedimento é bastante comum e faz parte do processo natural de desenvolvimento a partir da Equação 3.3.

$$(g * f)(m,n) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} g(i,j) \cdot f(m+i, n+j) \tag{3.4}$$

De maneira simplificada, a camada convolucional aplica um filtro (também conhecido como kernel) sobre a imagem de entrada, produzindo uma imagem filtrada com valores diferentes dos pixels. Em outras palavras, ela utiliza duas imagens como entrada (a entrada real e o kernel) e retorna uma imagem como resultado. O kernel desliza sobre a imagem de entrada, aplicando sua função matemática e retornando a imagem modificada. Sendo o funcionando desta operação ilustrado na Figura 3.4.

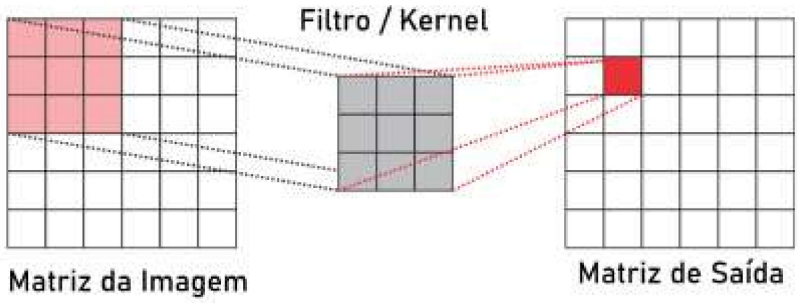


Figura 3.4 – Representação da Camada Convolucional

Fonte: (MARQUES, 2023)

3.3.3.1 Camadas Convolucionais

O funcionamento das camadas convolucionais nas CNNs é fundamental para a extração de características e a identificação dos padrões presentes nas imagens. Essas camadas são responsáveis por processar localmente as informações de uma imagem, tornando possível a detecção de bordas, texturas e outras características relevantes.

Uma camada convolucional é composta por um conjunto de filtros convolucionais, que são pequenas matrizes de pesos que deslizam sobre a imagem de entrada, capturando informações importantes em cada posição. Cada filtro é projetado para procurar por um determinado padrão na imagem, como linhas, curvas ou texturas específicas, e seu resultado é uma nova imagem, chamada de mapa de características.

Diversos efeitos podem ser obtidos ao se aplicar diferentes filtros (kernels) em uma imagem. Por exemplo, o filtro gaussiano é capaz de eliminar ruídos, enquanto outros filtros realçam os contornos ou removem qualquer elemento que não seja um contorno, deixando apenas o que se destaca na imagem.

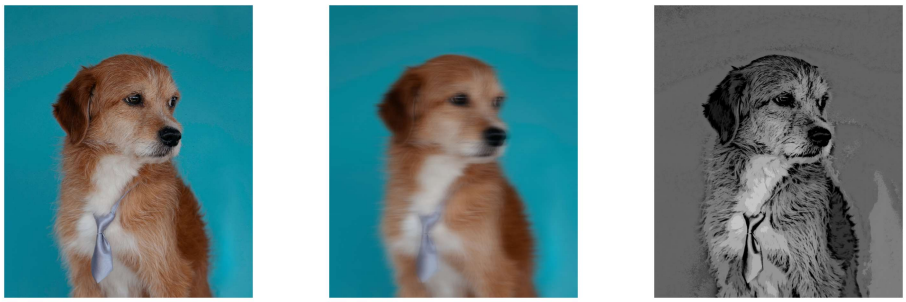


Figura 3.5 – Exemplos de diversos filtros aplicados sobre uma imagem

Fonte: Autor, 2023

Durante a convolução, o filtro é aplicado a várias regiões da imagem através de uma operação matemática conhecida como multiplicação do filtro pelos pixels da imagem e soma dos resultados. Esse valor resultante é colocado em uma nova imagem, que representa uma convolução específica. Essa operação de convolução é repetida para cada posição possível da imagem, gerando um conjunto de mapas de características.

Após a convolução, é comum aplicar uma função de ativação, como a função ReLU *(Rectified Linear Unit)* , que é não linear e ajuda a introduzir não linearidades nos mapas de características, aumentando a capacidade do modelo de aprender representações mais complexas.

Além das camadas convolucionais, as CNNs geralmente contêm camadas de pooling, que são responsáveis por reduzir a dimensionalidade dos mapas de características e fornecer invariantes de escala e translação. A operação de pooling realiza a subamostragem dos mapas de características, selecionando um valor representativo de cada região, como o valor minímo (*min pooling*), valor máximo (*max pooling*) ou a média (*average pooling*) (MARQUES, 2023). Normalmente, o *max pooling* é a técnica de pooling mais comumente empregada e que geralmente produz os resultados mais eficazes (VOIGT et al., 2018).

As camadas convolucionais e de pooling são combinadas em várias sequências para formar uma arquitetura de CNN. Geralmente, essas arquiteturas consistem em múltiplas

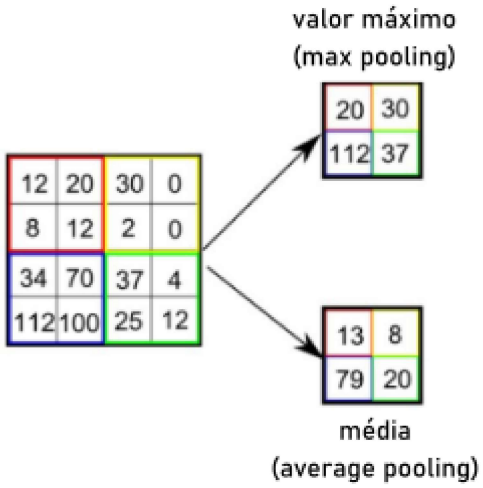


Figura 3.6 – Operação de pooling (valor máximo e média)

Fonte: Adaptada de (SALOMON, 2018)

camadas convolucionais seguidas por camadas de pooling e, ao final, uma ou mais camadas totalmente conectadas, que realizam a classificação final com base nas características extraídas.

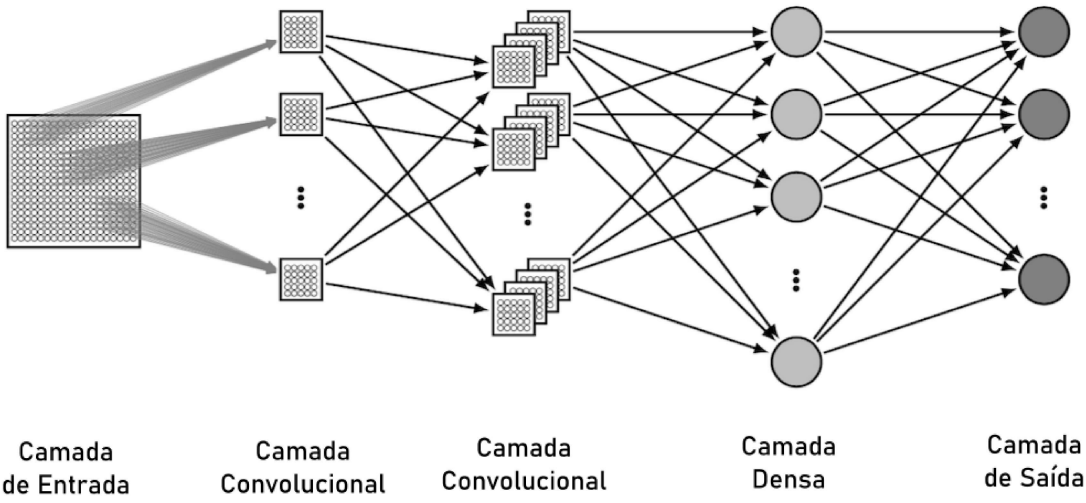


Figura 3.7 – Representação da arquitetura de uma CNN com duas camadas convolucionais e uma camada densa (camada completamente conectada)

Fonte: (SAKURAI, 2017)

3.3.4 Redes Neurais Recorrentes (RNN) e Long Short-Term Memory (LSTM)

As Redes Neurais Recorrentes (RNNs) são uma classe de algoritmos de aprendizado de máquina que se mostraram extremamente úteis na análise de dados sequenciais, seja em linguagem natural, séries temporais ou até mesmo em música (ACADEMY, 2022). Essas redes têm a capacidade de processar informações temporais, considerando o contexto passado para tomar decisões no presente.

Podemos entender uma RNN como um tipo especial de rede neural artificial, cuja arquitetura permite a retroalimentação dos dados. Diferente das redes neurais tradicionais, que possuem conexões apenas em uma direção, a RNN é capaz de armazenar informações sobre estados anteriores ao processar novas entradas. Esse tipo de memória permite que a rede aprenda a reconhecer padrões além da mera identificação de características isoladas.

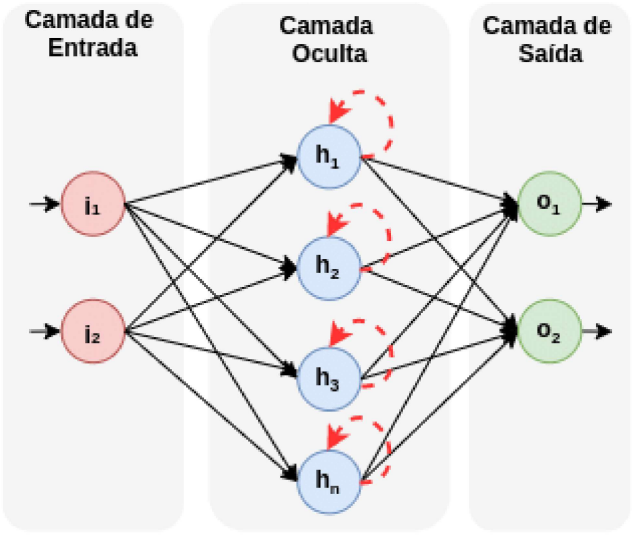


Figura 3.8 – Rede Neural Recorrente

Fonte: (BARBOSA et al., 2021)

Uma das grandes vantagens das RNNs é sua capacidade de processar sequências de comprimentos variados. Essa flexibilidade é especialmente útil em tarefas como tradução automática, onde as palavras de uma frase podem ter diferentes números de sinônimos ou quando as distâncias entre eventos sequenciais são variáveis, como em séries temporais. Essa adaptabilidade é resultado da estrutura interna das RNNs, que ajusta a sua memória para acomodar a complexidade dos dados.

No entanto, as RNNs tradicionais enfrentam desafios, como o desaparecimento do gradiente, que limitam sua capacidade de lidar eficazmente com dependências temporais de longo prazo. Essa é uma das razões pelas quais arquiteturas mais avançadas, como a chamada Long Short-Term Memory (LSTM) , ganharam destaque (ZHANG et al., 2021).

As LSTMs têm alcançado resultados impressionantes em tarefas de processamento

de linguagem natural, como tradução de textos, sumarização automática e até mesmo geração de texto (MADHAVAN; JONES, 2017).

3.3.4.1 Long Short-Term Memory (LSTM)

As redes LSTM *(Long Short-Term Memory)* são um tipo especializado de rede neural recorrente (RNN) que se destacam em tarefas que exigem a capacidade de lembrar e utilizar informações de longo prazo. Essas redes foram introduzidas por Sepp Hochreiter e Jürgen Schmidhuber em 1997 e desde então têm sido amplamente utilizadas em várias aplicações, como processamento de linguagem natural, reconhecimento de fala, tradução automática, entre outros (ACADEMY, 2022).

Uma das principais vantagens das redes LSTM é a sua capacidade de lidar com problemas de dependência de longo prazo, que são um desafio para as redes neurais tradicionais. Isso é possível graças à estrutura interna das unidades de memória LSTM, que utilizam uma combinação inteligente de portas de entrada, saída e esquecimento.

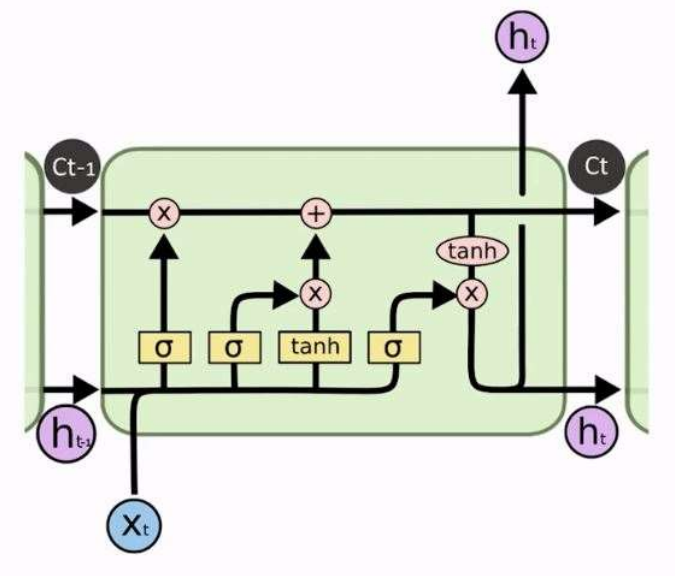


Figura 3.9 – Arquitetura da rede LSTM

Fonte: (VOIGT et al., 2018)

Cada unidade de memória LSTM mantém um estado interno que pode ser atualizado ou modificado ao longo do tempo. As portas de entrada permitem que a unidade de memória decida quais informações passarão para o estado interno, enquanto as portas de saída controlam quais informações serão exibidas na saída da unidade. A porta de esquecimento permite que a unidade de memória "esqueça"informações irrelevantes para a tarefa em questão.

Essa estrutura de portas permite que as redes LSTM evitem os problemas de gradiente desvanecente e gradiente explodindo que são comuns em redes neurais recorrentes mais simples. Além disso, as redes LSTM têm a capacidade de aprender relações temporais complexas, tornando-as especialmente úteis em tarefas que envolvem sequências de dados.

Em resumo, as redes LSTM são uma poderosa ferramenta para lidar com problemas que envolvem dependências de longo prazo, permitindo que as redes neurais processem efetivamente sequências de dados. Sua estrutura única de portas e unidades de memória faz com que sejam especialmente adequadas para tarefas de processamento de linguagem natural, reconhecimento de fala e tradução automática.

4 Metodologia

Podemos dividir esse trabalho em três etapas distintas, cada uma representada por um módulo de grande importância.

A primeira etapa é a captura de exemplos, que ocorre por meio de um módulo do aplicativo Web. Nessa etapa, o usuário define algumas configurações iniciais, como quais comandos serão ativados ou desativados (Continuar, Confirmar, Voltar e Emergência), a quantidade de exemplos a serem capturados para cada gesto e a quantidade de frames a serem capturados em cada amostra.

Em seguida, o usuário é direcionado ao módulo para fazer a captura de cada exemplo, vinculado à ação que representa. Durante a gravação da amostra, em cada frame, são detectadas as posições dos pontos faciais do usuário e calculada a distância entre os pontos-chave. Essas informações são armazenadas para compor o banco de dados de treinamento.

Após a etapa de aquisição do conjunto de treinamento, inicia-se a etapa de treinamento da Rede Neural. Outro módulo localiza os exemplos, efetua pré-processamentos no mesmo, define a arquitetura da rede neural e inicializa o processo de treinamento em si. Esse processo definirá o modelo classificador e armazenará-o em um arquivo local, tornando a rede neural já treinada pronta para uso.

Por fim, outro módulo presente no aplicativo web consulta o arquivo salvo anteriormente e carrega o modelo classificador já treinado. Em seguida, uma interface amigável exibirá ao usuário opções de interação e simultaneamente capturará as imagens do usuário com o objetivo de detectar os gestos realizados. Isso permitirá que o usuário manipule o sistema através de seus gestos.

4.1 O Sistema Proposto

4.1.1 Arquitetura

O sistema de Comunicação Aumentativa e Alternativa (CAA) proposto consiste em um sistema dividido em três módulos principais: Entrada e Processamento de Sinais, Análise da Leitura e Interface de Controle. A arquitetura do sistema proposto é apresentada na Figura 4.1.

O módulo de Entrada e Processamento de Sinais, baseado em um smartphone, utiliza a biblioteca MediaPipe Face Mesh para detectar os pontos de referência facial. Esse processo permite capturar com precisão os gestos e expressões faciais realizados pelo

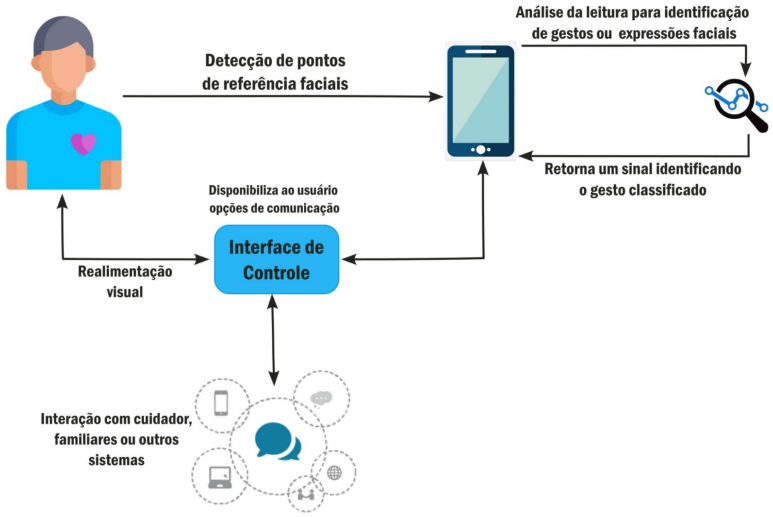


Figura 4.1 – Arquitetura do Sistema Proposto

Fonte: Autor, 2023

usuário.

Em seguida, no módulo de Análise da Leitura, é realizada uma análise detalhada desses gestos e expressões faciais identificados. O objetivo é identificar e compreender a intenção do usuário na comunicação. Para isso, é feita uma comparação dos gestos/expressões identificados com aqueles cadastrados previamente.

Uma vez concluída a análise, os resultados são enviados para a Interface de Controle, que está disponível no smartphone do usuário. Através dessa interface, o usuário tem acesso a diversas opções de interação, sejam elas com cuidadores, familiares ou outros sistemas. Essa interface recebe a classificação do sinal enviado pelo módulo de análise de leitura e, com base nessa informação, é capaz de determinar qual ação deve ser executada.

Dessa forma, o sistema de CAA proposto visa proporcionar uma comunicação mais eficiente e acessível para pessoas com dificuldades de comunicação verbal. Através da detecção e análise de gestos e expressões faciais, o usuário pode expressar suas necessidades, desejos e emoções de forma clara e objetiva. A interface de controle, por sua vez, permite uma interação mais facilitada e personalizável, atendendo às necessidades individuais de cada usuário.

4.1.2 Interface

4.1.2.1 Cadastro e configurações do Usuário

Ao iniciar o uso do sistema proposto, o usuário deve primeiro realizar um cadastro básico e fazer login para configurar os seus gestos personalizados e posteriormente utilizar o módulo principal, as telas podem ser observadas na Figura 4.2. Este procedimento é

essencial, pois os gestos de cada usuário são personalizados e esses dados são armazenados individualmente para cada usuário do aplicativo.

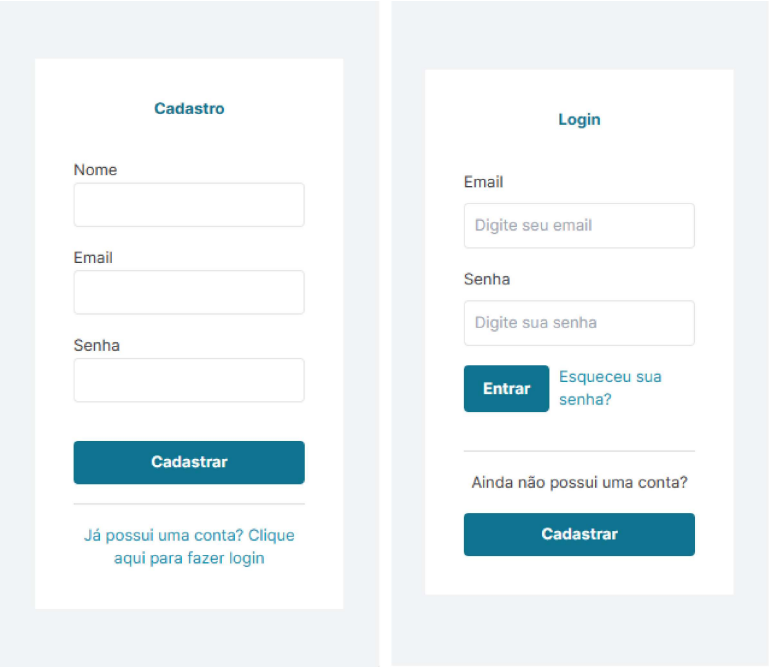


Figura 4.2 – Telas de Cadastro de Usuário e Login

Fonte: Autor, 2023

Após o login, o usuário deve configurar a coleta de amostras dos gestos faciais associados a cada comando, tela representada na Figura 4.3. As configurações disponíveis incluem:

- Quantidade de frames: representa o número de frames que serão armazenados para cada amostra, com um mínimo de 15 frames. Gestos mais longos podem exigir um maior número de frames.
- Quantidade de amostras: indica o número de amostras que serão gravadas para cada gesto/comando configurado, com um mínimo de 5 amostras.
- Comandos ativos: pelo menos dois comandos (Continuar e Confirmar) devem estar ativos para o uso do aplicativo. A versão inicial disponibiliza quatro possíveis comandos: Continuar, Confirmar, Voltar e Emergência.

A etapa seguinte envolve a gravação das amostras dos comandos ativos. O usuário pode escolher qual amostra correspondente a qual comando deseja gravar. Em caso de falhas, é possível excluir o gesto armazenado e realizar uma nova gravação. A visualização destas duas telas está representada na Figura 4.4, onde a primeira tela apresenta a lista de todas as amostras que devem ser gravadas e a segunda tela, a gravação de uma amostra.

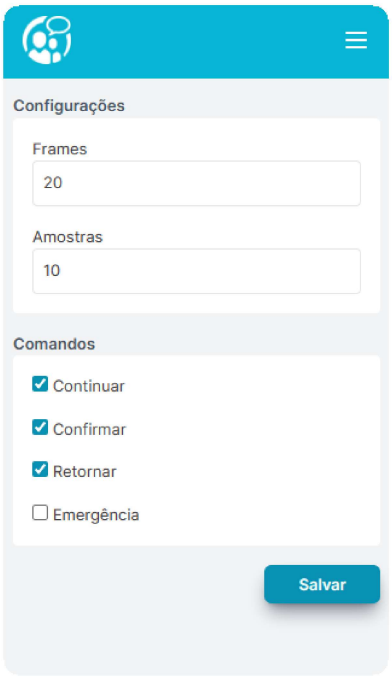


Figura 4.3 – Tela de Configurações

Fonte: Autor, 2023

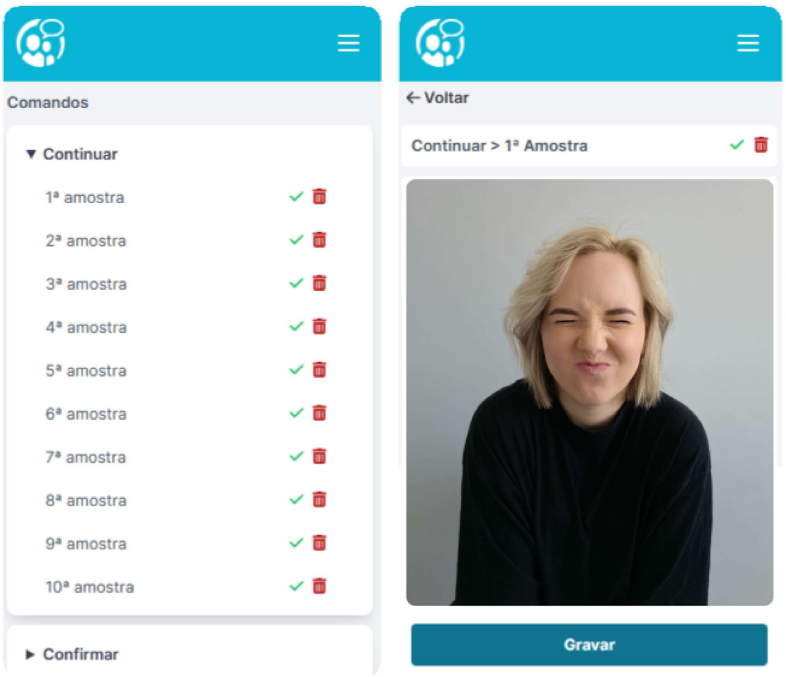


Figura 4.4 – Tela de Visualização da lista de amostras e tela de gravação da amostra

Fonte: Autor, 2023

4.1.2.2 Método de Interação

Existem diferentes métodos de seleção utilizadas na comunicação alternativa, projetados para auxiliar pessoas com deficiências físicas e motoras a interagir com dispositivos e meios digitais. Esses métodos se referem à maneira como o usuário indica os símbolos/opções nos dispositivos de comunicação alternativa.

Dentre os métodos existentes, o sistema proposto utilizará o movimento como meio de interação ao sistema, ao qual vamos nomear de *MotionAccess*. Esse método consiste no uso de gestos ou ações, realizados pela pessoa com deficiência através de uma câmera conectada a um dispositivo, como um computador, tablet ou smartphone, permitindo que o usuário navegue por diferentes opções ou funcionalidades apresentadas na tela, como representado na Figura 4.5.

Figura 4.5 – Representação de um sistema de comunicação alternativa

Fonte: Autor, 2023

Além disso, como pode ser observada na Figura 4.5 a escolha das ações demonstradas foi guiada por uma abordagem holística para abranger as necessidades e experiências abrangentes dos usuários. Cada ação apresentada representa elementos essenciais da vida cotidiana, desde a saúde e assistência médica até a interação social por meio de mensagens. A inclusão de aspectos como alimentação, bebidas, sono, higiene e condições ambientais, como luzes e temperatura, visa oferecer aos usuários uma gama completa de expressões para

comunicar suas necessidades, desejos e emoções. Ao incorporar elementos familiares, como a televisão, o sistema busca facilitar uma comunicação eficaz e significativa, promovendo assim uma interação mais inclusiva e personalizada.

O método de interação utilizado no sistema, o Motion Access, utiliza um padrão de exploração sequencial para navegar pelas opções exibidas no dispositivo, como pode ser observado na Figura 4.6. O usuário utiliza a câmera para iniciar a varredura e avançar ou não para a próxima opção. Quando a opção desejada é selecionada, o usuário ativa gestos pré-determinados para escolhê-la. Essa escolha pode ser feita por meio de uma piscada ou por outra ação específica.

Figura 4.6 – Representação do padrão de exploração sequencial, onde cada opção fica selecionada por um intervalo de tempo, aguardando a confirmação do usuário.

Fonte: Autor, 2023

O padrão de exploração sequencial é altamente personalizável e pode ser adaptado de acordo com as necessidades e capacidades individuais do usuário. Para isso, é necessário editar algumas configurações, como a velocidade de varredura (tempo em que uma opção permanece selecionada aguardando a possível confirmação do usuário), o tempo de atraso entre seleções e a disposição das opções na tela.

No entanto, esse processo pode ser lento e propenso a erros de seleção, o que pode prejudicar consideravelmente a eficiência do usuário. É importante projetar conjuntos de

seleção e configurar o sistema para minimizar erros e fornecer mecanismos para o usuário excluir ou cancelar seleções incorretas.

O sistema no qual este trabalho se baseia tenta contornar esses problemas através da simplificação de ações, como a utilização de comandos simples (como seguir e/ou retornar ao passo anterior), minimizando erros provenientes de múltiplas escolhas feitas através de uma interface multifacetada.

Outro proposta deste projeto é possibilitar a separação das opções em grupos, evitando assim que o usuário percorra todas as possíveis seleções até encontrar a opção desejada. Dessa forma, a própria aplicação age como um filtro, proporcionando ao usuário um caminho mais fácil de percorrer até a ação desejada.

A Figura 4.7 representa a utilização do sistema por parte do usuário com a separação das opções em grupos habilitada, na primeira imagem da figura observa-se a seleção das opções de grupo em andamento (o grupo de opções fica selecionado com a cor amarela), em seguida o usuário faz um gesto de confirmação (a confirmação representada pela cor vermelha), e então, a seleção passa a ser individual, agora apenas na opções do grupo selecionado, até que o usuário confirme a opção final desejada.

Um aspecto importante na utilização deste método é o uso de um suporte adequado para posicionar a câmera de forma acessível para o usuário. Além disso, o desenvolvimento de um sistema de comunicação utilizando este método requer a criação de um layout claro e intuitivo, que facilite a seleção das opções desejadas pelo usuário. Para isso, é essencial considerar o tamanho e a disposição dos elementos na tela, bem como a organização lógica das opções.

4.1.3 Aquisição e Pré-Processamento de Dados

Nesta fase, o usuário já definiu algumas configurações iniciais, como quais comandos serão ativados ou desativados (Continuar, Confirmar, Voltar e Emergência), a quantidade de exemplos a serem capturados para cada gesto e a quantidade de frames a serem capturados em cada amostra.

Para evitar problemas durante o treinamento da rede neural devido à quantidade insuficiente de amostras ou ao tempo limitado para a captura de cada amostra, foi necessário definir valores mínimos para cada opção nas configurações, como citado anteriormente.

4.1.3.1 Aquisição

O usuário realizar a gravação das amostras para cada ação/gesto pré-definido, para que então a rede neural após o treinamento seja capaz de reconhecer cada gesto quando realizado. A gravação de cada amostra segue os seguintes passos:

Figura 4.7 – Representação do método de separação das opções em grupos

Fonte: Autor, 2023

- Etapa de detecção e seleção do rosto do usuário: realizada utilizando a biblioteca FaceMesh (MediaPipe). Essa etapa tem como objetivo identificar o rosto do usuário na imagem.
- Detecção dos pontos faciais: Essa solução permite mapear 468 pontos faciais 3D em tempo real, possibilitando a identificação de diversas estruturas faciais, como olhos, nariz e boca.
- Seleção dos pontos-chave da estrutura facial: mapeamento do máximo de expressões possíveis. Esses pontos-chave incluem os olhos, sobrancelhas, nariz, boca e contorno do rosto.

De forma que, para cada amostra são armazenadas as informações referentes a quantidade total de frames, constituindo assim, o movimento do gesto/expressão, como pode ser observado, por exemplo, na Figura 4.8.

Figura 4.8 – Representação da captura de dados em diferentes frames ao longo do tempo durante a realização do gesto piscar

Fonte: Autor, 2023

4.1.3.2 Pré-Processamento

Após a aquisição dos dados relativos a face do usuário, o sistema utiliza pontos-chave e determina a distância entre dois pontos ao longo do tempo para concluir se houve ou não movimento, como pode ser observado na Figura 4.9.

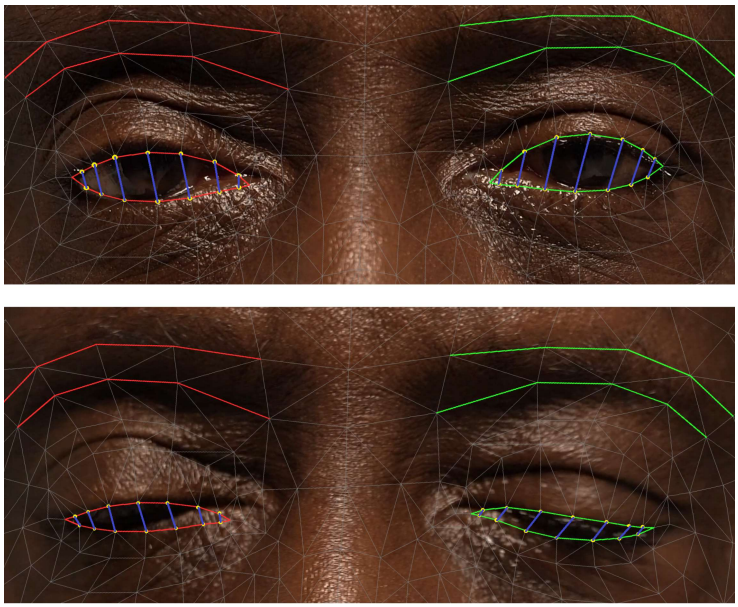


Figura 4.9 – Captura entre a distância entre pontos-chave essenciais em dois momentos diferentes (representando a mudança capturada na realização de cada gesto)

Fonte: Autor, 2023

A triangulação pode ser feita em quase qualquer região da face, mas para maior

confiabilidade na execução e identificação de movimentos, são utilizados pontos de triangulação onde a distância entre os pontos pré-determinados durante o movimento é maior. Por exemplo, os olhos, a boca e as sobrancelhas determinam uma grande movimentação e, portanto, uma maior diferença de distância entre os pontos. Nos olhos, temos a diferença de distância entre a pálpebra superior e inferior, e na boca, a diferença entre o lábio superior e inferior, como demonstram a Figura 4.10.

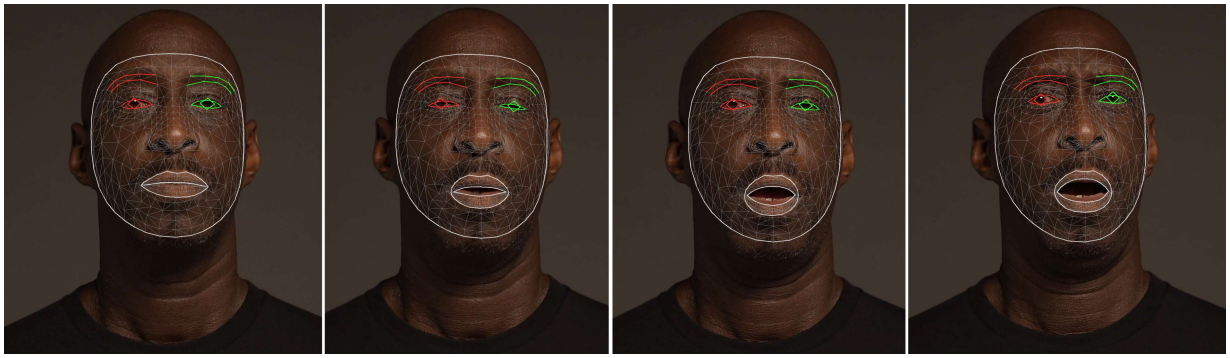


Figura 4.10 – Representação de diversos movimentos e os seus impactos na distância entre os pontos-chaves faciais

Fonte: Autor, 2023

Em seguida, é aplicado sobre o dado a normalização $Min - Max$, onde o valor da distância é linearmente transformado para um intervalo entre [0, 1]. Cada valor é subtraído pelo valor mínimo da característica e dividido pela diferença entre o valor máximo e mínimo. Isso garante que os dados estejam contidos em uma escala definida, preservando as relações proporcionais entre os valores.

4.2 Arquiteturas Adotadas

Para definir o modelo classificador de gestos a ser aplicado no sistema de comunicação alternativa em desenvolvimento, foram estudadas e testadas diversas estratégias. O objetivo era garantir a melhor eficiência da rede neural artificial.

Algumas configurações foram padronizadas e aplicadas de forma equivalente em todas as estratégicas testadas. Em primeiro lugar, na função de compilação, que é usada para preparar os modelos para a fase de treinamento, foram determinados os seguintes parâmetros: o otimizador ‘Adam’, a função de perda ‘categorical_crossentropy’ e a métrica de avaliação ‘categorical_accuracy’.

O otimizador Adam é um algoritmo popular que ajusta iterativamente os pesos do modelo com base nos gradientes calculados durante o treinamento, ajudando a convergir mais rapidamente para uma solução.

A função de perda é uma medida que indica quão bem o modelo está performando a tarefa desejada durante o treinamento. Sendo, escolhida a função de perda 'Categorical crossentropy' por ser comumente utilizada em problemas de classificação quando há várias classes. Ela compara a distribuição de probabilidade prevista pelo modelo com a distribuição real dos rótulos.

Para a métrica a 'categorical_accuracy' será usada para avaliar a precisão do modelo durante o treinamento. Essa métrica mede a proporção de predições corretas em relação ao total de predições feitas.

4.2.1 CNN

O modelo implementado é uma rede neural convolucional (CNN), cuja arquitetura é projetada para tarefas de classificação, onde o objetivo é categorizar dados de entrada (distância entre os pontos da face) em diferentes classes (gestos). A rede é organizada de maneira sequencial, com cada camada desempenhando um papel específico na extração e aprendizado de características. Uma visão geral da rede neural é apresentada na Figura 4.11.

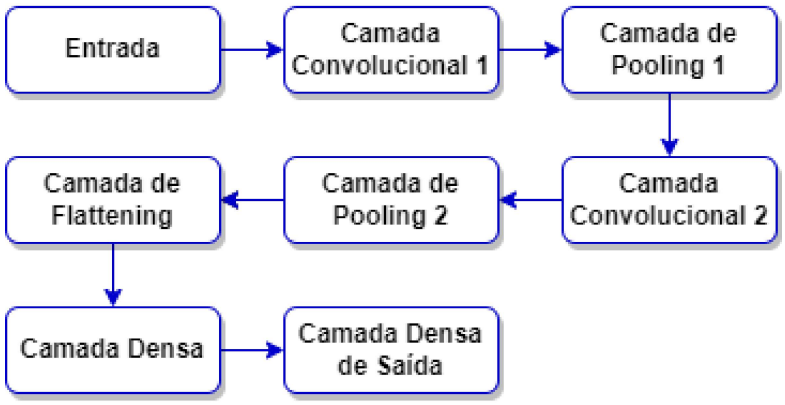


Figura 4.11 – Topologia da rede neural convolucional implementada

Fonte: Autor, 2023

A primeira camada é uma camada de convolução 1D com 32 filtros e um tamanho de núcleo de 3. Essa camada utiliza a função de ativação ReLU, que introduz não-linearidade na rede. A entrada para esta camada é definida pelo formato dos dados de entrada, especificamente, a quantidade de frames de cada amostra, e o tamanho do vetor de distâncias dos pontos faciais.

Logo após a camada de convolução, há uma camada de pooling 1D, que reduz a dimensionalidade dos dados pela aplicação de uma operação de max pooling com uma janela de tamanho 2. Essa operação ajuda a preservar as características mais relevantes e reduzir a complexidade computacional.

A arquitetura continua com uma segunda camada de convolução 1D, agora com 64 filtros e novamente usando a função de ativação ReLU. Após esta camada, há outra camada de pooling 1D, com uma janela de tamanho 2, realizando mais uma vez a redução da dimensionalidade.

A camada seguinte é uma camada de flattening, que transforma os dados em um vetor unidimensional, preparando o caminho para as camadas totalmente conectadas. A primeira dessas camadas densas possui 128 neurônios e utiliza a função de ativação ReLU.

A última camada é também densa, representando a camada de saída da rede. O número de neurônios nesta camada é determinado pelo número de classes no problema de classificação. A função de ativação utilizada é softmax, o que é comum em problemas de classificação, pois atribui probabilidades às diferentes classes, permitindo a escolha da classe com a probabilidade mais alta como a predição final da rede.

Em resumo, esse modelo segue uma abordagem convolucional para aprender padrões e características nos dados de entrada, culminando em uma camada de saída que fornece probabilidades para as diferentes classes do problema de classificação em questão.

4.2.2 LSTM

O modelo implementado é uma rede neural recorrente (RNN), sendo adotada a arquitetura Long Short-Term Memory (LSTM), que é particularmente adequada para classificar e prever série temporais. Uma visão geral da rede neural é apresentada na Figura 4.11.

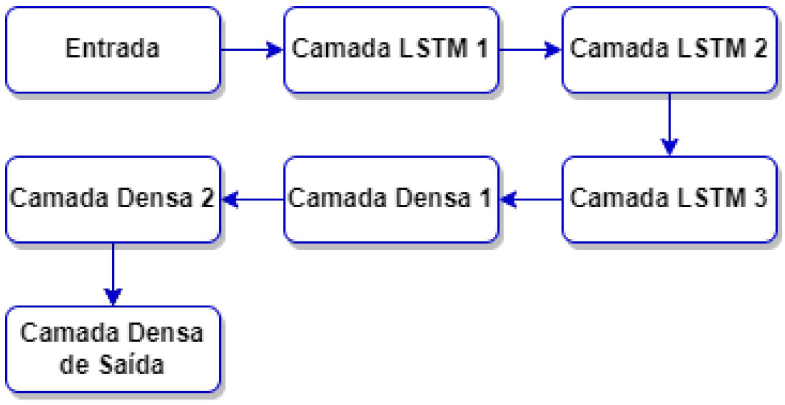


Figura 4.12 – Topologia da rede neural recorrente LSTM implementada.

Fonte: Autor, 2023

A primeira camada é uma camada LSTM (Long Short-Term Memory) com 256 unidades. Essa camada é configurada para retornar sequências, o que significa que ela

produzirá uma saída para cada passo de tempo na sequência de entrada. A função de ativação utilizada é a ReLU (Rectified Linear Unit), que introduz não-linearidade na rede.

A segunda camada LSTM tem 128 unidades e também retorna sequências. A função de ativação ReLU é novamente aplicada para promover a aprendizagem de padrões complexos nos dados sequenciais.

A terceira camada LSTM tem 64 unidades e, diferentemente das anteriores, está configurada para não retornar sequências, o que significa que produzirá uma única saída para toda a sequência. Isso geralmente é apropriado quando as camadas LSTM anteriores extraíram características relevantes e agora desejamos resumir essas informações.

Após as camadas LSTM, temos três camadas densas (Dense). A primeira tem 64 neurônios e utiliza a função de ativação ReLU. A segunda tem 32 neurônios, também com ativação ReLU. Essas camadas densas adicionam uma dimensão totalmente conectada à rede, permitindo a combinação de informações aprendidas pelas camadas LSTM.

A última camada densa, a camada de saída, tem um número de neurônios determinado pelo número de classes na tarefa de classificação. A função de ativação usada é softmax, que atribui probabilidades às diferentes classes, permitindo a classificação final.

Em resumo, este modelo utiliza camadas LSTM para processar dados sequenciais, seguidas por camadas densas para consolidar as informações e realizar a classificação final. A função de ativação ReLU é empregada ao longo das camadas para introduzir não-linearidade, enquanto a camada de saída utiliza softmax para gerar probabilidades associadas a cada classe.

4.3 Desenvolvimento

A metodologia descrita foi implementada em duas vertentes distintas. A primeira vertente é um aplicativo web desenvolvido para capturar os dados e fornecer a interface de controle. O aplicativo foi desenvolvido usando o framework $Vue.js$ em conjunto com a biblioteca $MediaPipeHolistic$. Essa combinação permite a detecção de poses e gestos do usuário, essenciais para a interação com o sistema. Além disso, o aplicativo utiliza os serviços $HostingeRealtimeDatabase$ do Firebase para hospedagem e armazenamento dos dados do usuário, incluindo os dados coletados para o treinamento da rede neural e o modelo classificador.

A segunda vertente é um backend implementado em Python, aproveitando as bibliotecas Keras, TensorFlow e Scikit-Learn. Essas ferramentas são fundamentais para o desenvolvimento e treinamento da rede neural. O backend também realiza consultas ao banco de dados para buscar os dados de treinamento, garantindo que o modelo esteja atualizado e eficiente.

Em resumo, essa abordagem híbrida combina tecnologias front-end e back-end para criar um sistema completo e funcional, capaz de processar dados, treinar modelos e fornecer uma experiência de usuário fluida.

4.3.1 Linguagens Utilizadas e Ferramentas

4.3.1.1 Vue.js

O Vue.js é um framework progressivo utilizado para a construção de interfaces de usuário (HANCHETT; LISTWON, 2018). Foi desenvolvido por Evan You, um engenheiro front-end que trabalhava anteriormente no Google. A ideia por trás do Vue.js era criar uma ferramenta que fosse simples e flexível o suficiente para ser adotada em projetos de qualquer tamanho [1].

O Vue.js é escrito em JavaScript, o que o torna uma opção acessível para a maioria dos desenvolvedores web. Ele utiliza uma abordagem baseada em componentes, o que significa que os componentes individuais podem ser reutilizados e combinados para criar interfaces complexas. Essa modularidade facilita a manutenção e o desenvolvimento de aplicativos.

Uma das principais vantagens do Vue.js é a sua curva de aprendizado suave. Ele possui uma sintaxe simples e intuitiva, o que facilita o seu uso por desenvolvedores iniciantes. Além disso, o Vue.js também fornece uma documentação abrangente que oferece guias, exemplos e tutoriais detalhados para orientar os desenvolvedores em seu aprendizado.

Outra vantagem do Vue.js é o seu desempenho. Ele é extremamente rápido e eficiente, o que permite a construção de aplicativos responsivos e de alta performance. Além disso, o Vue.js possui recursos avançados de renderização virtual, que otimizam a manipulação do DOM e garantem uma experiência mais suave para o usuário.

O Vue.js também é altamente flexível e adaptável. Ele oferece suporte tanto para a criação de projetos pequenos e simples quanto para aplicações complexas e de grande escala. Além disso, o Vue.js é compatível com a maioria das bibliotecas e ferramentas JavaScript, o que facilita a integração com outros frameworks e a expansão das funcionalidades do aplicativo.

4.3.1.2 Python

Python é uma linguagem de programação de alto nível, interpretada, que tem como objetivo principal produzir códigos de forma organizada, com uma sintaxe limpa e pouco verbosa. Além disso, é uma linguagem multiparadigma, suportando programação funcional, orientada a objetos e procedural (KUHLMAN, 2009). Uma característica importante do

[1] Informações disponíveis em: https://vuejs.org/

Python é a sua tipagem dinâmica, o que significa que não é necessário declarar os tipos de variáveis antes de utilizá-las. Além disso, a identação obrigatória ajuda a manter o código mais legível e estruturado.

Python é uma linguagem muito versátil e possui uma ampla biblioteca padrão que oferece suporte para muitas tarefas comuns, como manipulação de arquivos, comunicação de rede e processamento de string. Além disso, existem muitos pacotes e bibliotecas de terceiros disponíveis para Python, o que a torna uma escolha atraente para muitos tipos de projeto [2].

Python se tornou a linguagem preferida para muitos cientistas de dados e engenheiros de aprendizado de máquina por vários motivos. Em primeiro lugar, Python tem sido amplamente adotado na comunidade de aprendizado de máquina, com muitos tutoriais, documentações e recursos disponíveis para ajudar os desenvolvedores a aprender e usar Python para desenvolvimento de aprendizado de máquina.

Em segundo lugar, Python oferece uma série de bibliotecas populares específicas para aprendizado de máquina, como NumPy, pandas, scikit-learn e TensorFlow. Essas bibliotecas fornecem um conjunto poderoso de ferramentas para manipulação de dados, modelagem e treinamento de modelos de aprendizado de máquina.

Em terceiro lugar, Python possui uma sintaxe concisa e legível, o que facilita a escrita e a leitura do código. Isso é particularmente útil no desenvolvimento de algoritmos de aprendizado de máquina, onde a clareza do código é essencial para entender os processos envolvidos. Por fim, Python pode ser facilmente integrado com outras linguagens, como C++, o que permite aproveitar bibliotecas existentes implementadas nessas linguagens, aumentando ainda mais a capacidade de desenvolvimento.

Porém, existem algumas desvantagens ao usar a linguagem Python na implementação de algoritmos de aprendizado de máquina. Algumas delas incluem desempenho, uso de memória e suporte limitado para paralelismo (VANDERPLAS, 2016).

Python é uma linguagem interpretada, o que geralmente a torna mais lenta em comparação com linguagens de baixo nível, como C++ ou Java. Isso pode ser um problema quando se trabalha com conjuntos de dados grandes ou algoritmos complexos, onde o desempenho é crucial.

Além disso, a linguagem Python consome mais memória do que outras linguagens devido à alocação dinâmica de objetos e recursos adicionais da biblioteca padrão. Esse consumo excessivo de memória pode se tornar uma limitação quando se trabalha com servidores ou dispositivos com recursos limitados.

Python também tem suporte limitado para programação paralela e distribuída, o que pode ser um problema quando se lida com conjuntos de dados grandes e algoritmos

[2] Informações disponíveis em: https://www.python.org/about/

que podem se beneficiar do processamento paralelo. Embora existam bibliotecas como o "multiprocessing"disponíveis, outras linguagens como C++ oferecem melhores recursos nessa área.

No entanto, essas desvantagens podem ser contornadas de várias maneiras. É possível escrever partes críticas do código em linguagens de baixo nível, como C++ ou Cython, e integrá-las com o Python. Isso pode melhorar o desempenho do código. Python possui várias bibliotecas de terceiros, como NumPy e TensorFlow, que são otimizadas para desempenho e permitem a execução eficiente de algoritmos de aprendizado de máquina. É possível superar a limitação do Python para programação paralela usando técnicas de dimensionamento horizontal, como a distribuição de tarefas em vários servidores ou aproveitando serviços em nuvem (RASCHKA; MIRJALILI, 2019).

4.3.1.3 TensorFlow

O TensorFlow é um software de código aberto desenvolvido pelo Google e projetado para suportar desenvolvimento e treinamento de algoritmos de aprendizado de máquina (MCCLURE, 2017). Escrito na linguagem de programação Python, o TensorFlow é utilizado para criar, treinar e implantar modelos de aprendizado de máquina em uma ampla variedade de domínios [3].

O TensorFlow oferece uma estrutura flexível e eficiente para a construção de modelos de aprendizado de máquina. Ele inclui uma grande variedade de bibliotecas e ferramentas que auxiliam no processo de desenvolvimento e implementação de algoritmos de aprendizado de máquina. Além disso, oferece suporte para executar operações em GPUs (Graphics Processing Units) para acelerar o treinamento de modelos complexos.

Uma das principais características do TensorFlow é o seu uso de grafos computacionais. Ele permite que os desenvolvedores definam as operações e a arquitetura de um modelo por meio de um grafo, que representa as etapas de cálculo necessárias para realizar determinada tarefa. Essa abordagem possibilita a criação de redes neurais profundas e complexas, além de facilitar a otimização e o reuso de código.

O TensorFlow também possui uma ampla variedade de APIs e ferramentas que facilitam o processo de treinamento e implantação de modelos de aprendizado de máquina. Ele suporta o treinamento distribuído, permitindo que várias máquinas trabalhem juntas para treinar um modelo em grandes conjuntos de dados. Além disso, o TensorFlow oferece suporte para execução em dispositivos móveis e permite que os modelos sejam convertidos para formatos otimizados para esses dispositivos.

No campo do aprendizado de máquina, o TensorFlow é amplamente utilizado para tarefas como reconhecimento de imagem e de fala, tradução automática, processamento

[3] Informações disponíveis em: https://www.tensorflow.org/about/

de linguagem natural, previsão de séries temporais, entre outras. Sua flexibilidade e poder computacional permitem que seja aplicado em uma variedade de aplicações, desde análises de dados até desenvolvimento de sistemas de inteligência artificial avançados (MCCLURE, 2017).

4.3.1.4 Keras

Keras é uma biblioteca de código aberto para aprendizado de máquina e redes neurais, projetada para ser fácil de usar, modular e extensível. Ela foi desenvolvida com o objetivo de permitir uma prototipagem rápida e facilitar a implementação de modelos de aprendizado profundo [4].

Uma das principais vantagens do Keras é que ele pode ser usado em conjunto com o TensorFlow para criar modelos de aprendizado profundo de maneira mais intuitiva e eficiente. O Keras oferece uma interface de alto nível que facilita a construção de redes neurais, enquanto o TensorFlow oferece poder computacional e uma infraestrutura robusta para treinar e executar essas redes. Dessa forma, o Keras e o TensorFlow formam uma combinação poderosa para desenvolver soluções de aprendizado profundo.

Keras é altamente flexível e modular, o que significa que os usuários podem facilmente criar, modificar e conectar diferentes camadas em suas redes neurais. Com o Keras, é possível construir rapidamente uma variedade de arquiteturas de rede, como redes densamente conectadas (fully connected), convolucionais, recorrentes e até mesmo modelos GAN (Generative Adversarial Networks). Além disso, o Keras oferece suporte a técnicas avançadas de regularização, como dropout e normalização em lote (batch normalization), bem como otimizadores eficientes, como o Adam.

Outro aspecto importante do Keras é a sua capacidade de facilitar o processo de treinamento e avaliação de modelos. Ele fornece funções embutidas para pré-processamento de dados, como divisão de conjuntos de treinamento e validação, além de funcionalidades para treinamento em batch e monitoramento do desempenho do modelo durante o treinamento. Além disso, o Keras possui uma API muito intuitiva, o que torna mais fácil a tarefa de ajustar hiperparâmetros e analisar os resultados.

4.3.1.5 Media Pipe

A biblioteca *FaceMesh* do MediaPipe, desenvolvida pelo Google, utiliza redes neurais convolucionais (CNNs) para realizar o reconhecimento de pontos faciais em imagens e vídeos em tempo real. O *FaceMesh* é projetado para identificar e rastrear 468 pontos faciais, identificados na Figura 4.13, incluindo características como olhos, nariz, boca e contornos faciais.

[4] Informações disponíveis em: https://keras.io/

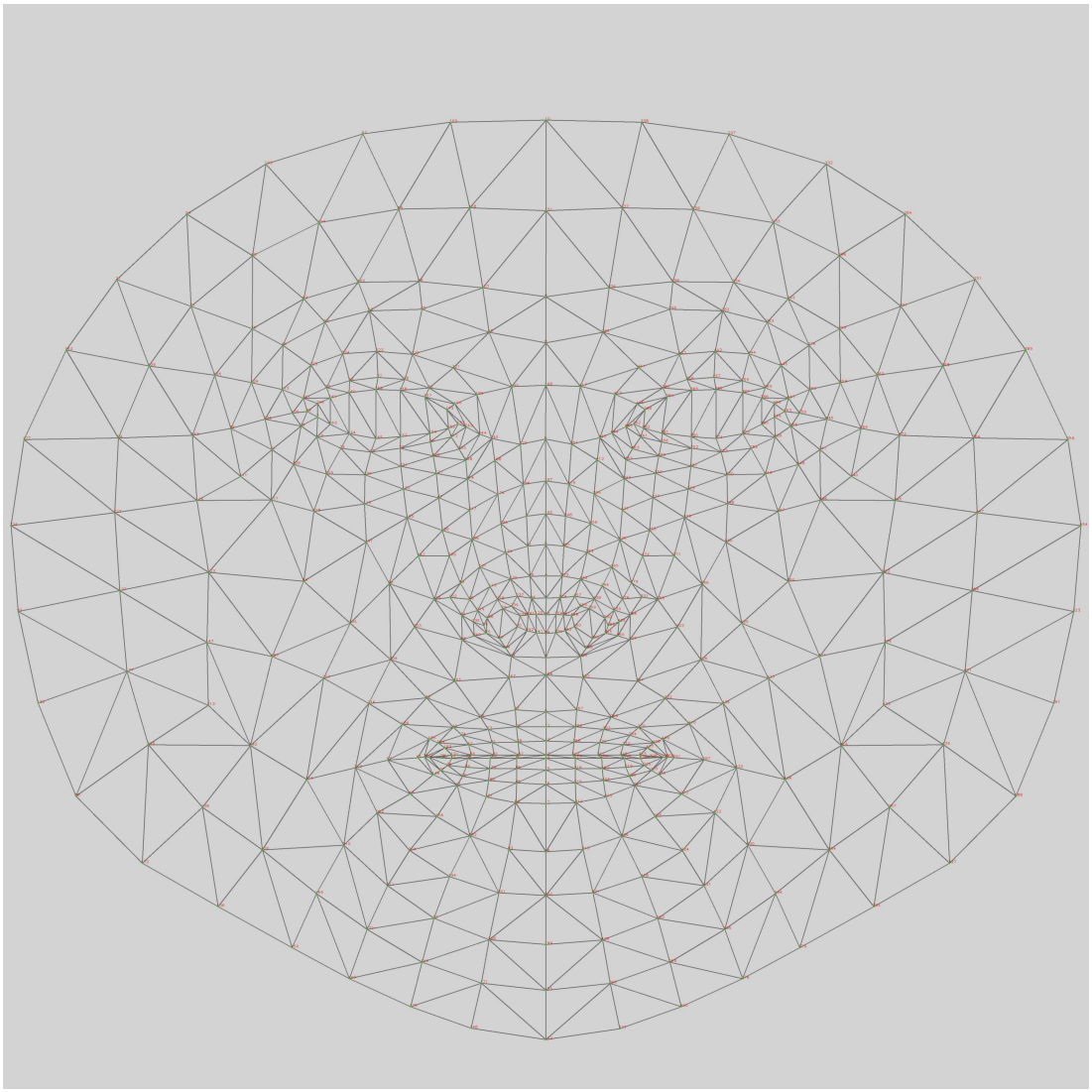


Figura 4.13 – Facemesh: mapa dos 468 pontos faciais

Fonte: Google

Como pode ser obervado na Figura 4.14 os pontos faciais são identificados corretamente mesmo com o uso de acessórios sobre a face, como por exemplo óculos, e com diferentes expressões faciais.

A abordagem do *FaceMesh* envolve uma arquitetura de rede neural convolucional treinada em um conjunto de dados extenso e diversificado de faces humanas. Essa arquitetura foi projetada para aprender representações hierárquicas das características faciais em diferentes níveis de abstração. Inicialmente, camadas convolucionais são utilizadas para a extração de características específicas. Posteriormente, camadas completamente conectadas interpretam o contexto facial de forma global. Essa abordagem permite que o modelo identifique e compreenda as características faciais com maior precisão e eficácia.

O treinamento com conjunto de dados anotado envolve o uso de um conjunto de

Figura 4.14 – Identificação dos 468 pontos faciais utilizando a biblioteca Facemesh

Fonte: Autor, 2023

dados abrangente e rotulado para treinar um modelo. Nesse contexto, o modelo em questão é voltado para a identificação de pontos faciais em imagens faciais. Esses pontos faciais são representados por 468 coordenadas específicas que indicam características como olhos, nariz, boca e contornos faciais. A anotação detalhada dessas coordenadas permite ao modelo aprender a mapear esses pontos com precisão nas imagens. Portanto, o treinamento é realizado com base em exemplos de imagens faciais anotadas, o que capacita o modelo a reconhecer e localizar esses pontos em novas imagens não vistas durante o treinamento.

Durante a execução, o *FaceMesh* utiliza a arquitetura treinada para realizar a detecção e inferência nos quadros de vídeo ou imagens de entrada. A rede convolucional (CNN) analisa localmente pequenas regiões da imagem, identificando padrões associados a diferentes pontos faciais.

Após a inferência, o *FaceMesh* realiza pós-processamento para otimizar a precisão dos resultados. O algoritmo utiliza técnicas de rastreamento para conectar os pontos faciais ao longo do tempo, proporcionando uma estimativa suave e contínua dos movimentos e expressões faciais.

A biblioteca *MediaPipe* oferece uma implementação eficiente e fácil de usar desse processo, permitindo que desenvolvedores incorporem facilmente o reconhecimento de pontos faciais em suas aplicações. O uso de CNNs na arquitetura do FaceMesh é crucial para a capacidade da biblioteca em lidar com a complexidade e variação nas expressões faciais em diferentes contextos e condições de iluminação.

Opções de Configuração

A biblioteca FaceMesh MediaPipe oferece uma ampla variedade de opções de configuração para atender às necessidades específicas de diferentes projetos. Neste texto, vamos abordar algumas dessas opções e explicar como elas podem ser úteis.

A primeira opção é o "*running_mode*", que determina o modo de execução da biblioteca. Existem dois modos disponíveis: "*lite*"e "*full*". O modo "*lite*"é mais rápido, mas pode fornecer menos detalhes e precisão nas informações do rosto. Já o modo "*full*"é mais lento, mas oferece uma maior quantidade de informações e detalhes. A escolha do modo depende das necessidades do projeto e das restrições de desempenho.

Em seguida, temos a opção "*num_faces*", que define o número máximo de faces que o algoritmo deve detectar e rastrear simultaneamente. Essa configuração é útil quando se deseja limitar o número de rostos processados pela biblioteca, economizando recursos computacionais.

Outra opção importante é o "*min_face_detection_confidence*", que define o nível de confiança mínimo necessário para considerar uma detecção de rosto como válida. Quanto maior o valor, mais confiável será a detecção, mas também pode resultar em uma taxa de detecção menor. Essa configuração é útil para controlar a sensibilidade do algoritmo em relação à detecção de rostos.

Da mesma forma, o "*min_face_presence_confidence*"define o nível de confiança mínimo necessário para considerar a presença de um rosto em um quadro de vídeo. Essa configuração é útil para evitar falsas detecções de rostos quando a confiança na detecção é baixa.

O "*min_tracking_confidence*"define o nível de confiança mínimo necessário para continuar rastreando um rosto detectado. Valores mais altos garantem uma maior confiabilidade no rastreamento, mas também podem resultar em perda de rastreamento de rostos com movimentos rápidos ou pouco visíveis.

A opção "*output_face_blendshapes*"permite habilitar ou desabilitar o retorno dos valores de blendshapes do rosto. Esses valores representam as expressões faciais e podem ser úteis para análise e animação de rostos.

Por sua vez, o "*output_facial_transformation_matrixes*"permite obter as matrizes de transformação facial. Essas matrizes podem ser usadas para ajustar a posição e a orientação de objetos virtuais relacionados ao rosto capturado.

Por fim, a opção "*result_callback*"permite definir uma função de retorno de chamada personalizada, que será executada sempre que um resultado do processamento de rosto estiver disponível. Essa opção é útil para realizar ações específicas com os dados obtidos, como a gravação de informações ou a atualização em tempo real de elementos

visuais.

4.3.1.6 Firebase Hosting

O Firebase Hosting é um serviço de hospedagem de aplicativos e sites estáticos, fácil de usar, rápido e seguro, fornecido pela Google. Com o Firebase Hosting, os desenvolvedores têm a possibilidade de implantar e hospedar facilmente seus aplicativos ou sites com apenas alguns comandos simples [5].

Uma das principais vantagens do Firebase Hosting é a facilidade de uso. Além disso, o Firebase Hosting oferece integração com outros serviços do Firebase, tornando a implantação e o escalonamento de aplicativos ainda mais simples.

Outra vantagem do Firebase Hosting é a velocidade. Os aplicativos e sites hospedados no Firebase são automaticamente distribuídos em uma CDN (Content Delivery Network) global, o que significa que seu conteúdo será entregue aos usuários através do servidor mais próximo, resultando em um carregamento mais rápido e uma melhor experiência do usuário.

Além disso, o Firebase Hosting também oferece segurança avançada. Todos os aplicativos e sites hospedados no Firebase são automaticamente servidos por HTTPS, garantindo uma comunicação segura entre o usuário e o servidor. Isso permite que você proteja seus dados e os dados dos usuários contra possíveis ataques.

4.3.1.7 Firebase Realtime Database

O Firebase Realtime Database é uma ferramenta de armazenamento e sincronização de dados em tempo real na nuvem, desenvolvida pela equipe do Google. Utilizando uma estrutura de árvore JSON, o Realtime Database permite que os desenvolvedores criem aplicativos escaláveis e em tempo real, que podem se adaptar às mudanças nos dados e ao crescimento do usuário [6].

Uma das principais vantagens do Firebase Realtime Database é a sua capacidade de sincronização em tempo real. Isso significa que qualquer alteração feita nos dados é imediatamente refletida em todos os dispositivos conectados ao banco de dados. Isso proporciona uma experiência de usuário fluída e consistente, mesmo em cenários de colaboração ou compartilhamento de dados em tempo real.

Além disso, o Firebase Realtime Database possui uma interface de programação simples e intuitiva, com suporte para várias plataformas, incluindo iOS, Android e web. Isso permite que os desenvolvedores criem aplicativos para uma ampla variedade de dispositivos

[5] Informações disponíveis em: https://firebase.google.com/
[6] Informações disponíveis em: https://firebase.google.com/

e plataformas, sem necessidade de se preocupar com a complexidade da infraestrutura de back-end.

Outra característica importante do Firebase Realtime Database é a sua escalabilidade. À medida que o número de usuários e a quantidade de dados aumentam, o serviço pode lidar facilmente com essa demanda, garantindo que os aplicativos continuem funcionando de maneira eficiente, sem lentidão ou interrupções.

Além de fornecer armazenamento de dados, o Firebase Realtime Database também oferece recursos avançados, como autenticação de usuários, análise de dados e notificações por push. Esses recursos permitem que os desenvolvedores criem aplicativos personalizados e ricos em recursos, que atendam às necessidades específicas dos usuários e ofereçam uma experiência excepcional .

5 Resultados e Discussões

5.1 Descrição das Métricas de Avaliação dos Resultados

Durante o processo de desenvolvimento de um modelo de Aprendizado de Máquina, é crucial avaliar sua eficácia de acordo com os objetivos específicos de cada tarefa. Para essa avaliação, recorremos a funções matemáticas conhecidas como Métricas de Avaliação, que desempenham um papel fundamental ao mensurar a capacidade de acerto e erro dos modelos.

A escolha adequada de uma métrica requer cuidado e consideração de diversos fatores, incluindo a distribuição proporcional dos dados em cada classe no conjunto de dados disponível e a natureza do objetivo de previsão (probabilidade, binário, ranking, etc.). Portanto, é essencial possuir um entendimento aprofundado da métrica selecionada (BERTONI et al., 2021).

Para uma compreensão mais abrangente das métricas de avaliação de resultados, torna-se necessário fornecer uma breve explanação sobre cada uma delas. Ao realizar previsões em uma população, os resultados obtidos podem ser categorizados de maneira distintiva em quatro partes fundamentais, como pode ser observado na Figura 5.1.

Figura 5.1 – Predições Negativas e Positivas (Falsas e Verdadeiras)

Fonte: (BERTONI et al., 2021)

- Verdadeiros Positivos (VP) : referem-se a amostras corretamente classificadas como Positivas.

- Verdadeiros Negativos (VN) : são amostras corretamente classificadas como Negativas.
- Falsos Positivos (FP) : são amostras erroneamente classificadas como Positivas
- Falsos Negativos (FN) : são amostras erroneamente classificadas como Negativas

A partir desses princípios, surgem quatro métricas essenciais: Acurácia (ou Exatidão), Sensibilidade, Especificidade e Precisão. Essas métricas formam a base para outras que visam avaliar o desempenho dos modelos. Detalhes adicionais sobre essas métricas são apresentados a seguir:

- **Acurácia**: também conhecida como Exatidão, representa a proporção das previsões corretas (Positivas + Negativas) em relação a todas as possibilidades, considerando erros e acertos.
- **Sensibilidade**: também chamada de **Recall**, representa a proporção de verdadeiros positivos, ou seja, a fração de positivos que foram corretamente previstos. Por outro lado, a **Especificidade** refere-se à fração de verdadeiros negativos, representando a capacidade do modelo em prever corretamente as instâncias negativas.
- **Precisão**: é a medida que indica a fração de positivos previstos que realmente são positivos. Em outras palavras, destaca a capacidade do modelo de evitar falsos positivos.
- **Medida-F** ($F - Measure$): uma média harmônica ponderada de Precisão e Sensibilidade, oferece uma visão equilibrada do desempenho do modelo, especialmente útil em situações em que ambas as métricas são cruciais.
- **Acurácia Balanceada**: em cenários binários, é calculada como a média aritmética entre Sensibilidade e Especificidade ou como a área sob a curva ROC, considerando previsões binárias em vez de pontuações. Quando o classificador apresenta desempenho uniformemente bom em ambas as classes, esse termo se simplifica para a precisão convencional (DEVELOPERS, 2020).
- **Matriz de Confusão**: é uma tabela que ajuda a entender como um modelo de classificação se sai em suas previsões. Ela é dividida em quatro partes: Verdadeiros Positivos (quando o modelo acerta ao prever uma classe positiva), Verdadeiros Negativos (acertos ao prever uma classe negativa), Falsos Positivos (erros ao prever uma classe positiva que era negativa) e Falsos Negativos (erros ao prever uma classe negativa que era positiva).
- **ROC, ou Curva Característica de Operação do Receptor)** : demonstra a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos em diferentes

limiares de decisão. Quanto mais a curva se aproxima do canto superior esquerdo, melhor é o desempenho do modelo, indicando uma maior capacidade de distinguir entre as classes. A área sob a curva ROC (AUC-ROC) é uma métrica comum usada para resumir o desempenho geral do modelo, sendo 1 o valor máximo possível.

5.2 Descrição dos resultados e seus significados

Os resultados serão delineados para duas abordagens distintas: o reconhecimento de gestos offline e o reconhecimento de gestos online em tempo real. Realizar a análise de reconhecimento de gestos tanto em um contexto offline quanto online oferece uma visão abrangente do desempenho do sistema em diferentes cenários. A distinção entre essas abordagens é significativa pelos seguintes motivos:

- Contexto de Uso:
  - Offline: Refere-se ao reconhecimento de gestos em dados que já foram previamente coletados. É útil para avaliar a capacidade do modelo de generalizar padrões em situações conhecidas.
  - Online (tempo real): Envolve o reconhecimento de gestos em tempo real, simulando condições práticas de interação. Isso é crucial para avaliar a eficácia do modelo em ambientes dinâmicos e sua capacidade de tomar decisões instantâneas.
- Avaliação de Desempenho Dinâmico:
  - Offline: Permite uma análise mais profunda da capacidade de aprendizado do modelo, pois os gestos são processados após a coleta de dados.
  - Online (tempo real): Avalia o desempenho em condições mais desafiadoras, onde o modelo deve lidar com variações em tempo real e tomar decisões instantâneas.

A comparação entre os resultados obtidos nessas duas abordagens fornece insights valiosos sobre a robustez e a adaptabilidade do modelo. Pode revelar se o modelo treinado offline mantém seu desempenho quando colocado em situações práticas em tempo real. Essa análise comparativa é crucial para validar a eficácia do sistema em cenários do mundo real e identificar possíveis lacunas entre as simulações offline e as condições online.

A coleta de dados para o reconhecimento offline foi conduzida de maneira análoga à obtenção de amostras para o treinamento da rede neural. Durante esse processo, as informações pertinentes ao gesto em questão foram registradas e posteriormente aplicadas ao modelo classificador. Em contrapartida, os testes de reconhecimento online foram conduzidos no aplicativo desenvolvido, simulando uma interação realista. Durante esses

testes, os usuários interagiram ativamente com o sistema, realizando os gestos previamente treinados e avaliando a resposta do aplicativo em tempo real.

Além disso, dentro de cada abordagem, serão estabelecidas comparações considerando variações em características específicas, como a quantidade de amostras empregadas no treinamento, a normalização ou não dos dados, a inclusão ou não de gestos semelhantes, e as diferentes arquiteturas utilizadas para treinar o modelo classificador. A inclusão destas variações possuem por objetivo:

- Quantidade de Amostras: O número de amostras no treinamento pode impactar diretamente na capacidade do modelo de generalizar padrões. Ter uma quantidade suficiente e representativa de dados é crucial para evitar overfitting (ajuste excessivo aos dados de treinamento) ou underfitting (modelo muito simplificado para capturar padrões complexos).

- Inclusão de Gestos Semelhantes: Incluir ou não gestos semelhantes no treinamento afeta a capacidade do modelo de discriminar entre gestos sutis. Isso é vital para cenários em que a distinção precisa ser feita entre movimentos similares, e a inclusão de gestos semelhantes pode melhorar a capacidade do modelo de aprender características distintivas.

- Diferentes Arquiteturas do Modelo: A escolha da arquitetura do modelo (como CNN, LSTM, etc.) impacta diretamente no modo como o modelo aprende e representa informações. Experimentar com diferentes arquiteturas permite encontrar a que melhor se adapta aos dados e à natureza da tarefa, otimizando assim o desempenho geral do modelo.

- Validação Cruzada: A utilização de técnicas como validação cruzada ajuda a mitigar o viés na avaliação do modelo, garantindo uma avaliação mais robusta e menos dependente da divisão específica entre dados de treinamento e teste. Isso é especialmente relevante quando o conjunto de dados é limitado.

Em resumo, essas variações na validação permitem uma avaliação abrangente do modelo, garantindo que ele seja robusto, preciso e capaz de generalizar para situações diversas. A consideração cuidadosa dessas variações visa contribuir para a construção de modelos mais confiáveis e eficazes.

Para uma compreensão mais aprofundada dos resultados apresentados, serão fornecidos a seguir detalhes sobre as abordagens utilizadas. Nos próximos tópicos, serão discutidos os resultados específicos de cada uma dessas abordagens, abrangendo tanto o reconhecimento de gestos offline quanto o reconhecimento de gestos online.

1ª Abordagem

Nesta fase, foram obtidas amostras de três gestos distintos, cada um desempenhando uma função única na interação com o aplicativo. Inicialmente, temos o gesto neutro, representando o comando "Continuar".

Seguindo, a coleta abrangeu uma piscada longa, associada ao comando "Confirmar". Este gesto adiciona uma dimensão expressiva à interação, proporcionando uma maneira natural de confirmar a escolha.Por fim, o terceiro gesto envolve a abertura da boca, caracterizando o comando "Retornar".

Essa variação nos gestos adiciona diversidade à interação, mas também representa um desafio considerável para o modelo classificador, dado o contraste significativo com os gestos anteriores. A presença de gestos distintos, como a abertura da boca para "Retornar", contribui para a riqueza e complexidade da interação gestual, destacando a capacidade do modelo em compreender e classificar movimentos variados.

2ª Abordagem

Nesta etapa, procedemos à coleta de amostras de três gestos distintos. Inicialmente, temos o gesto neutro, representando o comando "Continuar". Apesar de não interagir diretamente com o aplicativo, desempenha um papel crucial na classificação, especialmente quando o usuário aguarda a seleção desejada.

Em seguida, a coleta envolveu uma piscada longa, associada ao comando "Confirmar", e, por fim, duas piscadas rápidas, indicativas do comando "Retornar/Voltar". A presença de gestos semelhantes, como as duas formas de piscar, configura um desafio significativo para o modelo classificador nesta abordagem, que utiliza a normalização dos dados para padronizar as escalas e amplitudes, promovendo maior estabilidade no desempenho do modelo.

5.3 Reconhecimento de Gestos Offline

Esta abordagem concentra-se no reconhecimento de gestos offline, explorando a capacidade do sistema em interpretar e classificar movimentos sem a necessidade de uma transmissão em tempo real. O objetivo central desses experimentos é avaliar minuciosamente o desempenho do modelo, comparando sua eficácia no reconhecimento de gestos offline com os resultados obtidos no cenário em tempo real (online).

5.3.1 Resultados - 1ª Abordagem

Ao analisar os resultados dos experimentos de reconhecimento de gestos realizados para esta abordagem, observamos variações notáveis no desempenho dos modelos de redes

neurais convolucionais (CNN) e redes neurais recorrentes de longa memória (LSTM) com diferentes quantidades de amostras utilizadas no treinamento. A Tabela 5.1 apresenta a comparação entre as métricas obtidas nas variações realizadas nos experimentos, como no número de amostras e no arquitetura ao qual o modelo foi treinado.

Tabela 5.1 – Reconhecimento de Gestos Offline: Acurácia, Precisão, Recall e F1 Score - 1ª Abordagem

5 amostras	CNN	LSTM
Acurácia:	0,9333	0,7833
Precisão:	0,9349	0,7963
Recall:	0,9333	0,7833
F1 Score:	0,9337	0,7877

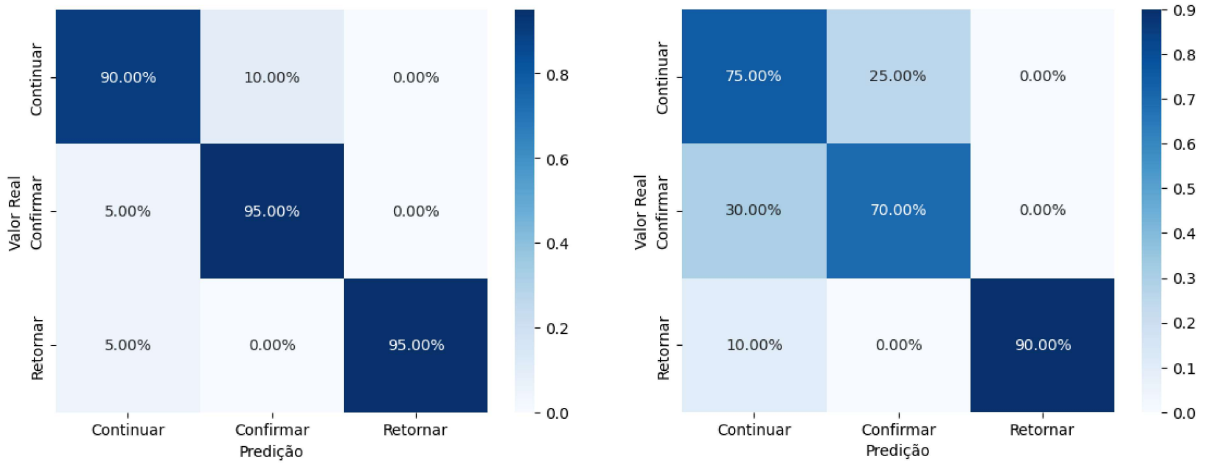
10 amostras	CNN	LSTM
Acurácia:	0,9000	0,7500
Precisão:	0,9023	0,7584
Recall:	0,9000	0,7500
F1 Score:	0,8997	0,7523

15 amostras	CNN	LSTM
Acurácia:	0,9667	0,8500
Precisão:	0,9697	0,8574
Recall:	0,9667	0,8500
F1 Score:	0,9670	0,8515

20 amostras	CNN	LSTM
Acurácia:	0,9833	0,8500
Precisão:	0,9841	0,8533
Recall:	0,9833	0,8500
F1 Score:	0,9833	0,8492

Para 5 amostras, a CNN apresentou um desempenho excepcional, atingindo uma precisão de 93,49%, recall de 93,33%, e um F1 Score impressionante de 93,73%. Por outro lado, a LSTM teve um desempenho inferior, com uma precisão de 79,63%, recall de 78,33%, e um F1 Score de 78,77%. Para mais detalhes a matriz de confusão comparativa entre os modelos CNN e LSTM é apresentada na Figura 5.2

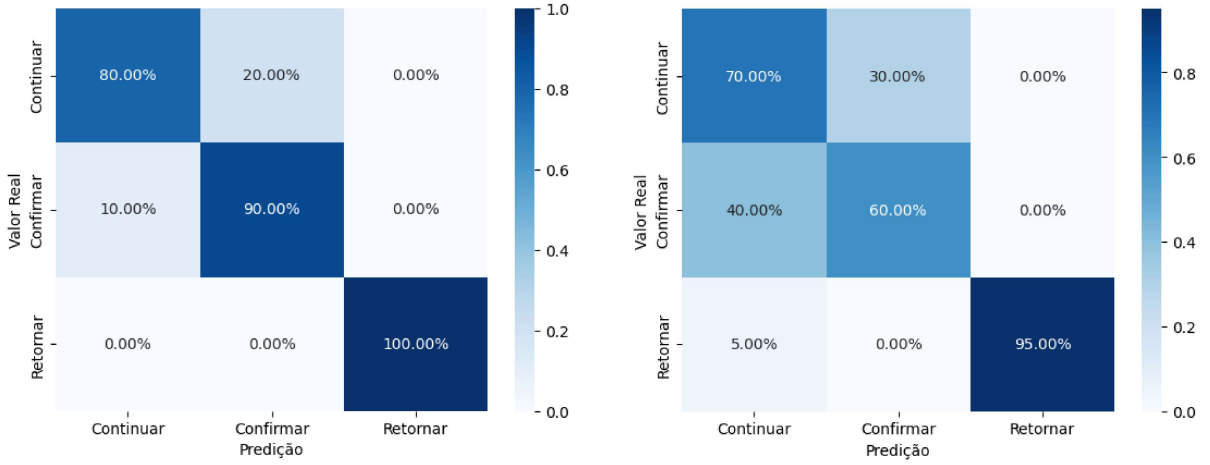
Figura 5.2 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Offline: 5 amostras (CNN x LSTM) - 1ª Abordagem



Fonte: Autor, 2023.

Com 10 amostras, a CNN manteve um desempenho robusto, com uma precisão de 90,24%, recall de 90%, e um F1 Score de 89,97%. Enquanto isso, a LSTM apresentou melhoria, alcançando uma precisão de 75%, recall de 75%, e um F1 Score de 75,24%. A matriz de confusão comparativa entre os modelos CNN e LSTM é fornecida na Figura 5.3.

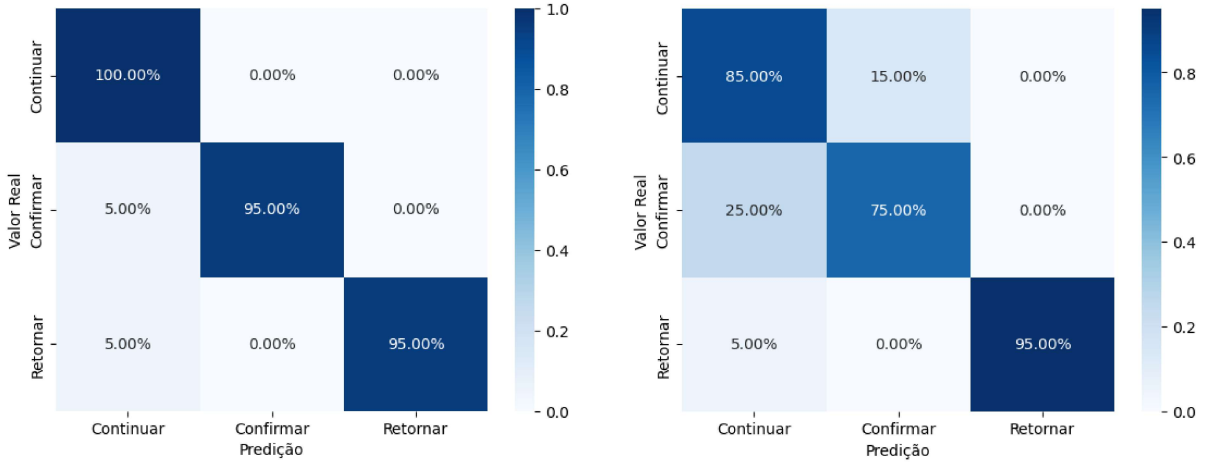
Figura 5.3 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Offline: 10 amostras (CNN x LSTM) - 1ª Abordagem



Fonte: Autor, 2023.

Aumentando para 15 amostras, a CNN continuou a se destacar, atingindo uma precisão de 96,97%, recall de 96,67%, e um F1 Score de 96,70%. A LSTM também melhorou, registrando uma precisão de 85%, recall de 85%, e um F1 Score de 85,15%. Para uma visão mais completa, a matriz de confusão comparativa entre os modelos CNN e LSTM é demonstrada na Figura 5.4.

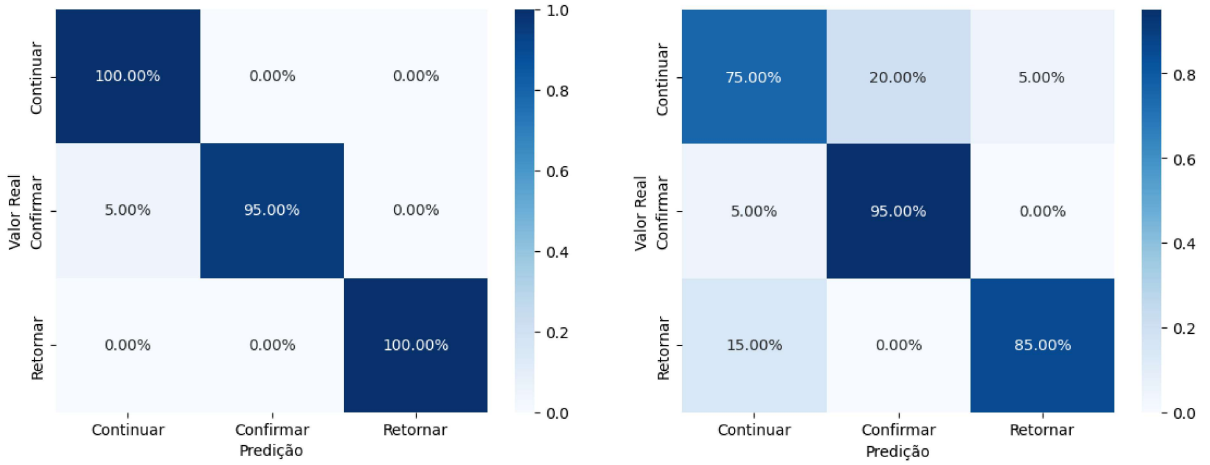
Figura 5.4 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Offline: 15 amostras (CNN x LSTM) - 1ª Abordagem



Fonte: Autor, 2023.

Finalmente, para 20 amostras, a CNN alcançou resultados excepcionais, com uma precisão de 98,41%, recall de 98,33%, e um F1 Score de 98,33%. Enquanto isso, a LSTM manteve um desempenho sólido, com uma precisão de 85,33%, recall de 85%, e um F1 Score de 84,92%. A matriz de confusão comparativa entre os modelos CNN e LSTM é fornecida na Figura 5.5.

Figura 5.5 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Offline: 20 amostras (CNN x LSTM) - 1ª Abordagem



Fonte: Autor, 2023.

Em resumo, os resultados sugerem que a CNN supera consistentemente a LSTM em todas as métricas avaliadas, e o aumento na quantidade de amostras tende a melhorar o desempenho de ambos os modelos. A escolha entre os dois dependerá dos requisitos específicos do problema e da disponibilidade de dados para treinamento.

5.3.2 Resultados - 2ª Abordagem

Ao analisar os resultados dos experimentos desta abordagem, observa-se variações significativas no desempenho entre os modelos de redes neurais convolucionais (CNN) e de redes neurais recorrentes de longa memória (LSTM), especialmente em relação à quantidade de amostras utilizadas. Na Tabela 5.2 visualiza-se as principais métricas obtidas a partir dos experimentos realizados. Em seguida, será apresentada a análise dos resultados obtidos e as suas respectivas matrizes de confusão, realizando um comparativo a partir da quantidade de amostras utilizadas no treinamento dos modelos.

Tabela 5.2 – Reconhecimento de Gestos Offline: Acurácia, Precisão, Recall e F1 Score - 2ª Abordagem

5 amostras	CNN	LSTM
Acurácia:	0,8167	0,4833
Precisão:	0,8423	0,4760
Recall:	0,8167	0,4833
F1 Score:	0,8119	0,4687

10 amostras	CNN	LSTM
Acurácia:	0,8667	0,5833
Precisão:	0,8702	0,6866
Recall:	0,8667	0,5833
F1 Score:	0,8650	0,5448

15 amostras	CNN	LSTM
Acurácia:	0,9167	0,8667
Precisão:	0,9157	0,8704
Recall:	0,9167	0,8667
F1 Score:	0,9158	0,8673

20 amostras	CNN	LSTM
Acurácia:	0,9333	0,9166
Precisão:	0,9387	0,9278
Recall:	0,9333	0,9166
F1 Score:	0,9331	0,9159

Para 5 amostras, a CNN demonstrou um desempenho superior, atingindo uma precisão de 84.23%, recall de 81.67%, e um F1 Score de 81.19%. Em contraste, a LSTM apresentou uma acurácia de apenas 48.33%, indicando uma dificuldade substancial na correta classificação dos gestos, com precisão, recall e F1 Score também inferiores. A matriz de confusão comparativa dos resultados obtidos entre os modelos CNN e LSTM pode ser observada na Figura 5.6.

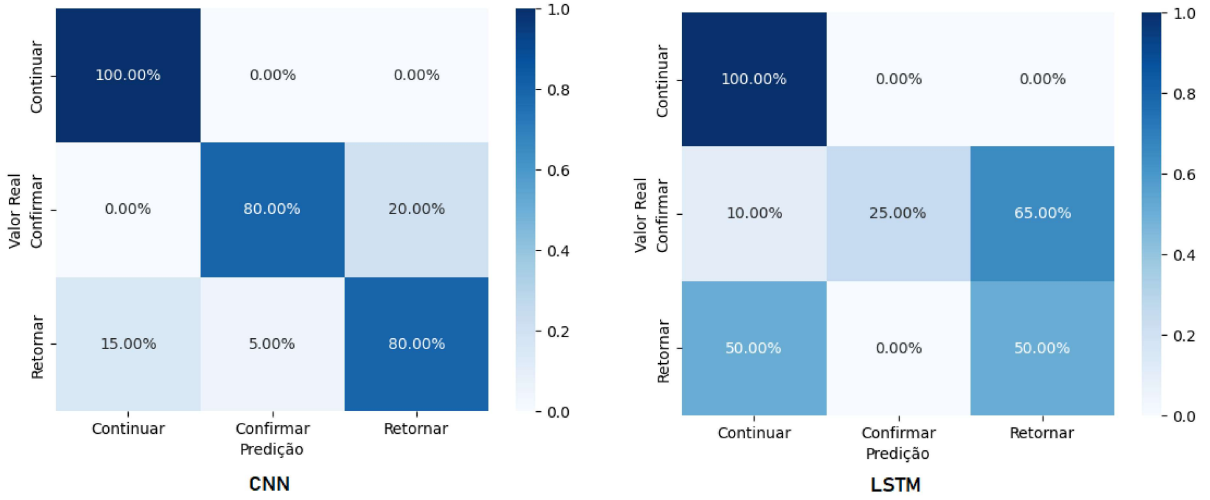
Figura 5.6 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Offline: 5 amostras (CNN x LSTM) - 2ª Abordagem

Fonte: Autor, 2023.

Com 10 amostras, a CNN manteve sua superioridade, alcançando uma acurácia de 86.67%, precision de 87.02%, recall de 86.67%, e F1 Score de 86.50%. A LSTM, embora tenha melhorado em comparação com 5 amostras, ainda apresentou um desempenho

inferior, com acurácia de 58.33% e métricas de precisão, recall e F1 Score abaixo dos valores obtidos pela CNN. Além disso, a matriz de confusão comparativa entre os modelos CNN e LSTM pode ser observada na Figura 5.7.

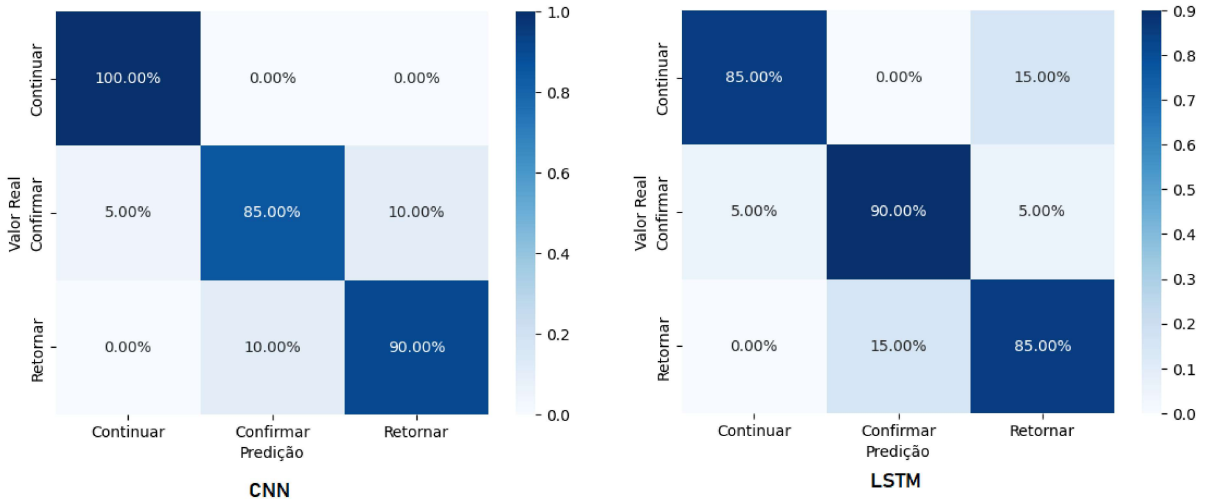
Figura 5.7 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Offline: 10 amostras (CNN x LSTM) - 2ª Abordagem



Fonte: Autor, 2023.

Ao aumentar para 15 amostras, a CNN continuou a se destacar, atingindo uma acurácia de 91.67% e mantendo métricas consistentes de precisão, recall e F1 Score, todos acima de 91.5%. A LSTM, embora tenha melhorado em relação a 10 amostras, ainda ficou atrás, com acurácia de 86.67% e métricas inferiores às da CNN. Pode ser observada também, na Figura 5.8, a matriz de confusão comparativa entre os modelos CNN e LSTM.

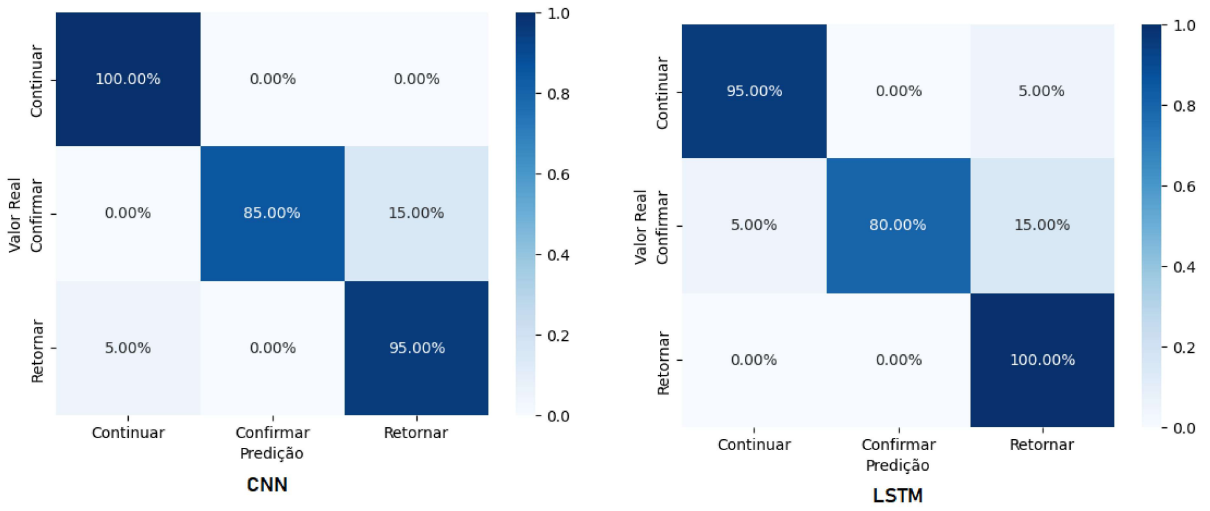
Figura 5.8 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Offline: 15 amostras (CNN x LSTM) - 2ª Abordagem



Fonte: Autor, 2023.

Finalmente, com 20 amostras, a CNN demonstrou um desempenho excepcional, alcançando uma acurácia de 93.33%, precisão de 93.87%, recall de 93.33% e F1 Score de 93.31%. A LSTM também melhorou, mas permaneceu atrás da CNN, com acurácia de 91.67% e métricas de precisão, recall e F1 Score inferiores. Pode-se observar também, através da Figura 5.9, a matriz de confusão comparativa entre os modelos CNN e LSTM.

Figura 5.9 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Offline: 20 amostras (CNN x LSTM) - 2ª Abordagem



Fonte: Autor, 2023.

Esses resultados destacam a eficácia da CNN no reconhecimento de gestos offline, especialmente à medida que a quantidade de amostras aumenta. A LSTM, embora melhore com mais dados, não consegue superar o desempenho consistente da CNN. Essa análise reforça a importância da escolha do modelo e da quantidade de dados no desenvolvimento de sistemas de reconhecimento de gestos robustos.

5.4 Reconhecimento de Gestos Online

Esta abordagem concentra-se no reconhecimento de gestos online, onde o usuário executa gestos de forma contínua, sem que a máquina saiba previamente quando cada gesto começa ou termina. Essa tarefa enfrenta desafios, como a execução de vários gestos distintos em sequência diante da câmera e a dificuldade em determinar automaticamente o término de um gesto e o início do próximo. Naturalmente, também, observa-se um intervalo entre gestos, que não representa claramente nem o primeiro nem o segundo gesto.

Com objetivo de contornar os desafios citados, foram adotadas soluções específicas. A primeira envolve o uso da técnica de varredura quadro a quadro e análise em intervalos para identificar qual gesto teve maior incidência. A segunda solução adotada é o estabelecimento

de um gesto "Neutro"representando a ação "Continuar", o qual, ao ser usado no aplicativo, não aciona nenhum comando, desta forma, o intervalo "sem gestos"também é classificado, contando como um gesto.

5.4.1 Resultados - 1ª Abordagem

Ao analisar os resultados dos experimentos de reconhecimento de gestos relativos a esta abordagem, percebemos variações distintas no desempenho dos modelos de redes neurais convolucionais (CNN) e redes neurais recorrentes de longa memória (LSTM) em diferentes quantidades de amostras utilizadas no treinamento. As principais métricas obtidas através dos experimentos podem ser observadas com mais detalhes em Tabela 5.3.

Tabela 5.3 – Reconhecimento de Gestos Online: Acurácia, Precisão, Recall e F1 Score - 1ª Abordagem

5 amostras		
	CNN	LSTM
Acurácia:	0,9500	0,7333
Precisão:	0,9507	0,7918
Recall:	0,9500	0,7333
F1 Score:	0,9499	0,7281

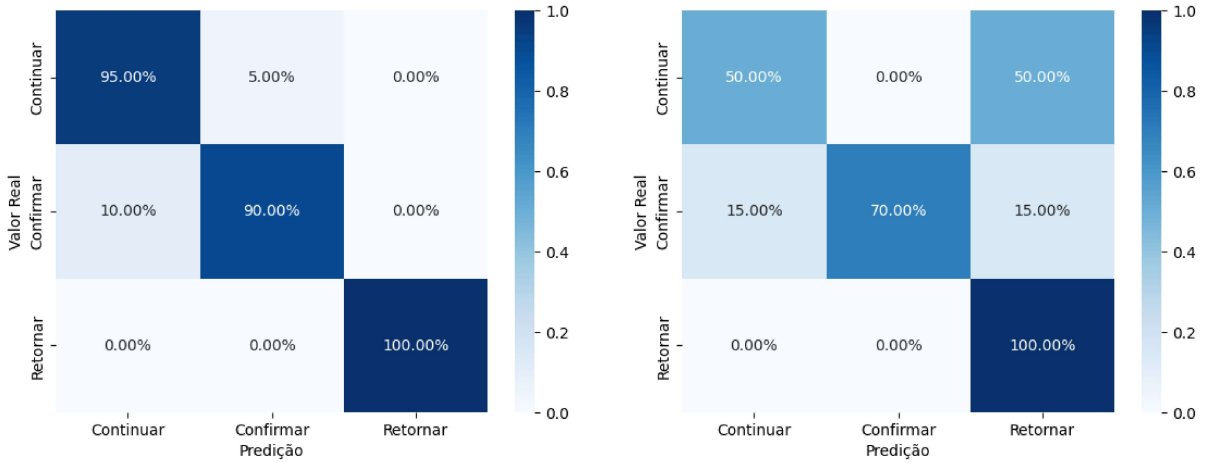
10 amostras		
	CNN	LSTM
Acurácia:	0,9000	0,7000
Precisão:	0,9137	0,7297
Recall:	0,9000	0,7000
F1 Score:	0,9003	0,6915

15 amostras		
	CNN	LSTM
Acurácia:	0,9333	0,8167
Precisão:	0,9444	0,8489
Recall:	0,9333	0,8167
F1 Score:	0,9346	0,8137

20 amostras		
	CNN	LSTM
Acurácia:	0,9667	0,7667
Precisão:	0,9696	0,7845
Recall:	0,9667	0,7667
F1 Score:	0,9666	0,7606

Com 5 amostras, a CNN se destacou com uma precisão de 95,07%, recall de 95%, e um F1 Score impressionante de 94,97%. Enquanto isso, a LSTM apresentou um desempenho inferior, alcançando uma precisão de 79,18%, recall de 73,33%, e um F1 Score de 72,81%. Caso haja interesse em dados mais específicos, a matriz de confusão comparativa entre os modelos CNN e LSTM pode ser encontrada na Figura 5.10.

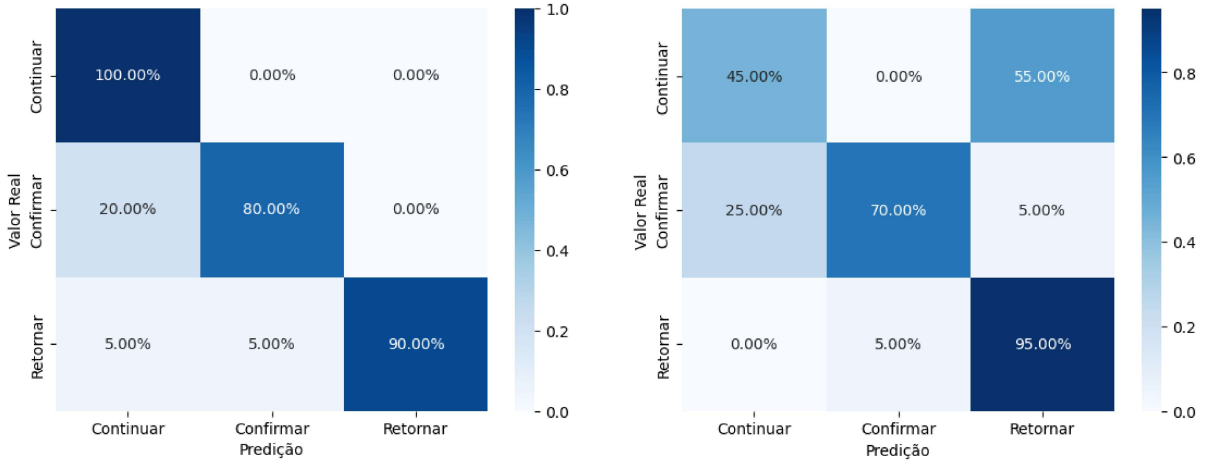
Figura 5.10 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Online: 5 amostras (CNN x LSTM) - 1ª Abordagem



Fonte: Autor, 2023.

Ao dobrar a quantidade de amostras para 10, a CNN manteve uma performance sólida, registrando uma precisão de 91,37%, recall de 90%, e um F1 Score de 90,04%. Já a LSTM apresentou melhora, atingindo uma precisão de 72,97%, recall de 70%, e um F1 Score de 69,15%. Para uma visão mais completa, a matriz de confusão comparativa entre os modelos CNN e LSTM é apresentada na Figura 5.11.

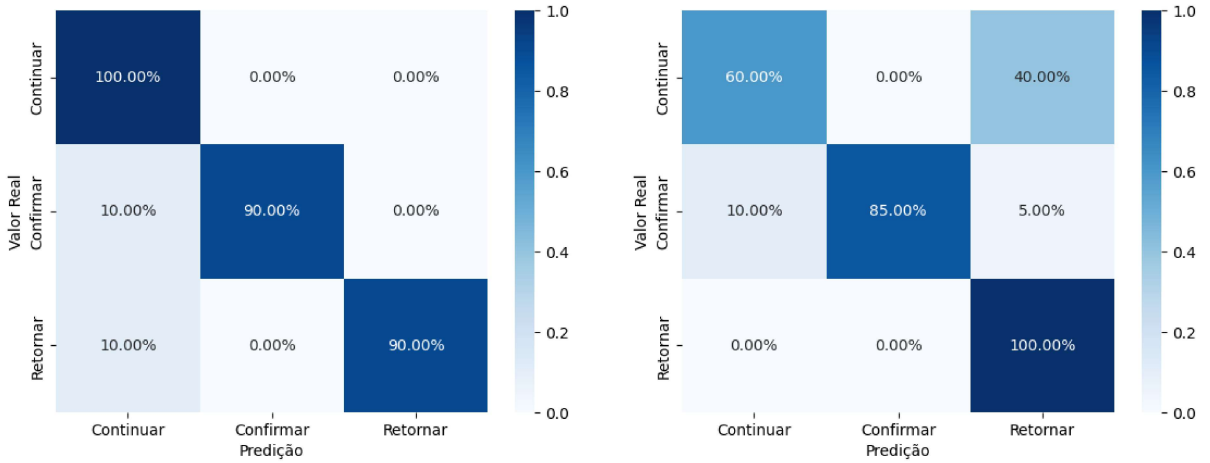
Figura 5.11 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Online: 10 amostras (CNN x LSTM) - 1ª Abordagem



Fonte: Autor, 2023.

Com 15 amostras, a CNN continuou a apresentar um desempenho robusto, com uma precisão de 94,44%, recall de 93,33%, e um F1 Score de 93,46%. A LSTM também melhorou, atingindo uma precisão de 84,89%, recall de 81,67%, e um F1 Score de 81,37%. A matriz de confusão comparativa entre os modelos CNN e LSTM é fornecida na Figura 5.12.

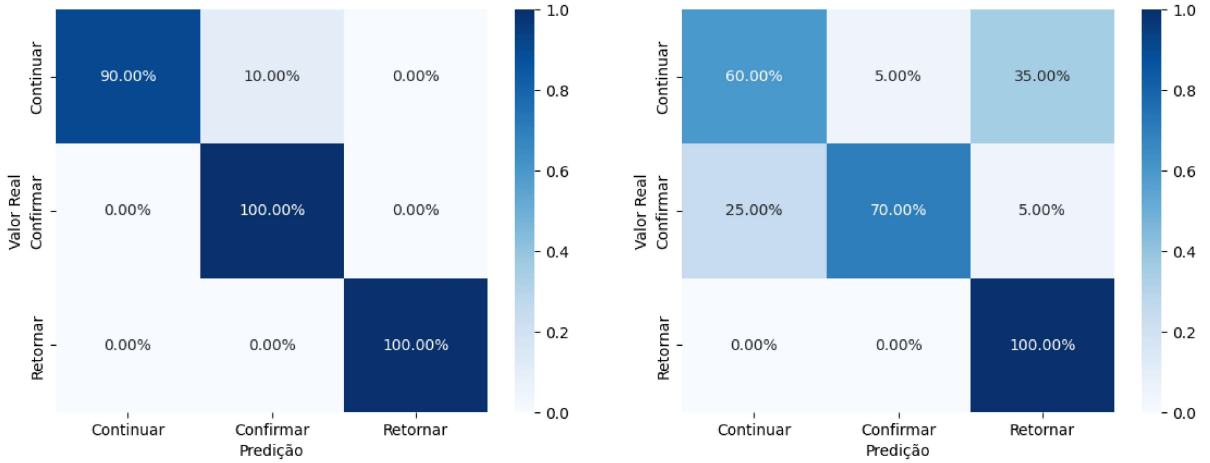
Figura 5.12 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Online: 15 amostras (CNN x LSTM) - 1ª Abordagem



Fonte: Autor, 2023.

Para 20 amostras, a CNN alcançou resultados excepcionais, com uma precisão de 96,97%, recall de 96,67%, e um F1 Score de 96,66%. A LSTM, por outro lado, manteve um desempenho sólido, registrando uma precisão de 78,45%, recall de 76,67%, e um F1 Score de 76,07%. Para mais detalhes a matriz de confusão comparativa entre os modelos CNN e LSTM é apresentada na Figura 5.13.

Figura 5.13 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Online: 20 amostras (CNN x LSTM) - 1ª Abordagem



Fonte: Autor, 2023.

Em resumo, os resultados indicam que a CNN supera consistentemente a LSTM em todas as métricas avaliadas durante o reconhecimento de gestos online. Aumentar a quantidade de amostras tende a beneficiar ambos os modelos, mas a escolha entre eles dependerá das especificidades do problema e dos requisitos de desempenho.

5.4.2 Resultados - 2ª Abordagem

Ao examinar os desdobramentos dos experimentos conduzidos nesta abordagem, é evidente a presença de variações significativas no desempenho entre os modelos de redes neurais convolucionais (CNN) e redes neurais recorrentes de longa memória (LSTM), especialmente ao considerar a variação na quantidade de amostras empregadas. A Tabela 5.4 oferece uma visão abrangente das principais métricas derivadas desses experimentos recentes. Posteriormente, procederemos com a análise aprofundada dos resultados obtidos, incluindo a apresentação das respectivas matrizes de confusão. Este enfoque permitirá uma comparação detalhada, destacando a influência da quantidade de amostras utilizadas no treinamento dos modelos sobre o desempenho alcançado.

Tabela 5.4 – Reconhecimento de Gestos Online: Acurácia, Precisão, Recall e F1 Score - 2ª Abordagem

5 amostras		
	CNN	LSTM
Acurácia:	0,7167	0,4000
Precisão:	0,7318	0,3960
Recall:	0,7167	0,4000
F1 Score:	0,7162	0,3560

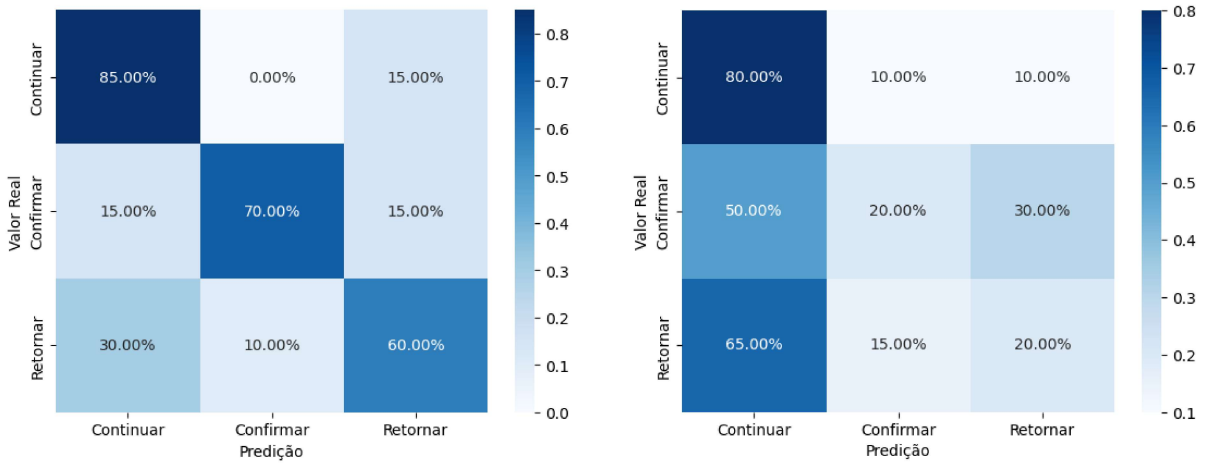
10 amostras		
	CNN	LSTM
Acurácia:	0,7667	0,5333
Precisão:	0,8056	0,7026
Recall:	0,7667	0,5333
F1 Score:	0,7651	0,4750

15 amostras		
	CNN	LSTM
Acurácia:	0,9167	0,7667
Precisão:	0,9245	0,8379
Recall:	0,9167	0,7667
F1 Score:	0,9180	0,7593

20 amostras		
	CNN	LSTM
Acurácia:	0,8000	0,8000
Precisão:	0,8131	0,8204
Recall:	0,8000	0,8000
F1 Score:	0,7952	0,7948

Para 5 amostras, CNN obteve uma acurácia de aproximadamente 71.7%, enquanto a LSTM apresentou um desempenho inferior, com uma acurácia de apenas 40%. A CNN também demonstrou melhor precisão, recall e F1 Score, indicando uma capacidade superior de classificação em comparação com a LSTM. Pode-se observar também, através da Figura 5.14, a matriz de confusão comparativa entre os modelos CNN e LSTM.

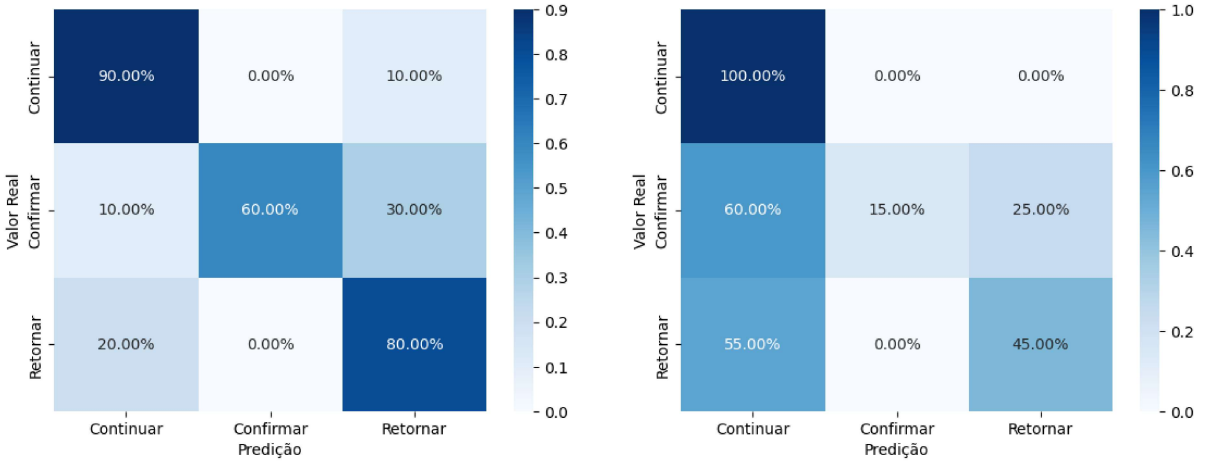
Figura 5.14 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Online: 5 amostras (CNN x LSTM) - 2ª Abordagem



Fonte: Autor, 2023.

Para 10 amostras os resultados melhoraram para ambas as arquiteturas em comparação com os experimentos de 5 amostras. A CNN alcançou uma acurácia de aproximadamente 76.7%, enquanto a LSTM melhorou para 53.3%. A CNN continuou a apresentar desempenho superior em termos de precisão, recall e F1 Score em comparação com a LSTM. Além disso, a matriz de confusão comparativa entre os modelos CNN e LSTM pode ser observada na Figura 5.15.

Figura 5.15 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Online: 10 amostras (CNN x LSTM) - 2ª Abordagem

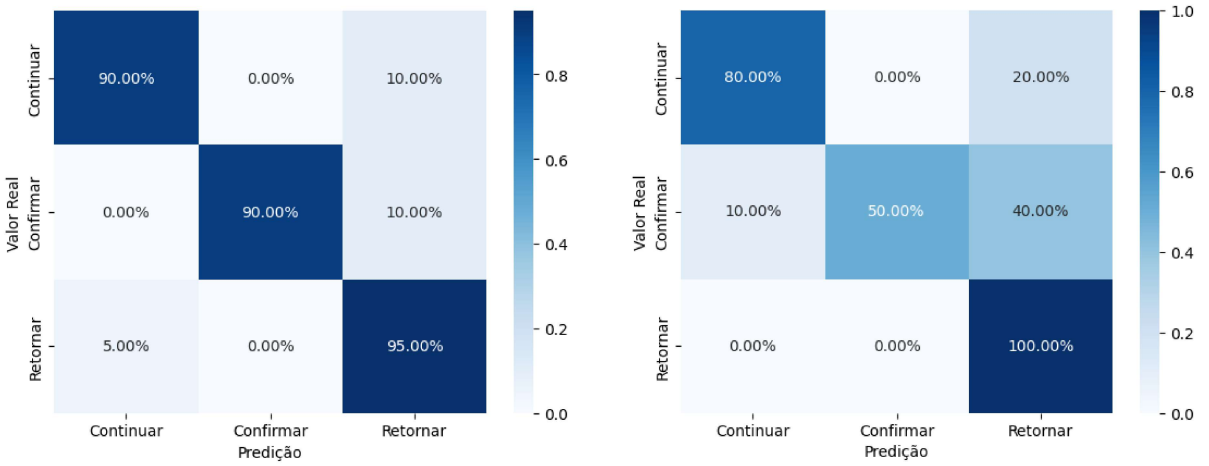


Fonte: Autor, 2023.

Com o aumento do número de amostras, agora 15, a CNN alcançou uma acurácia de 91.7%, mantendo uma alta precisão, recall e F1 Score. A LSTM também melhorou,

atingindo uma acurácia de 76.7%, embora ainda ficasse atrás da CNN em termos de métricas de desempenho, como pode ser observado também na matriz de confusão comparativa entre os modelos CNN e LSTM apresentada na Figura 5.16.

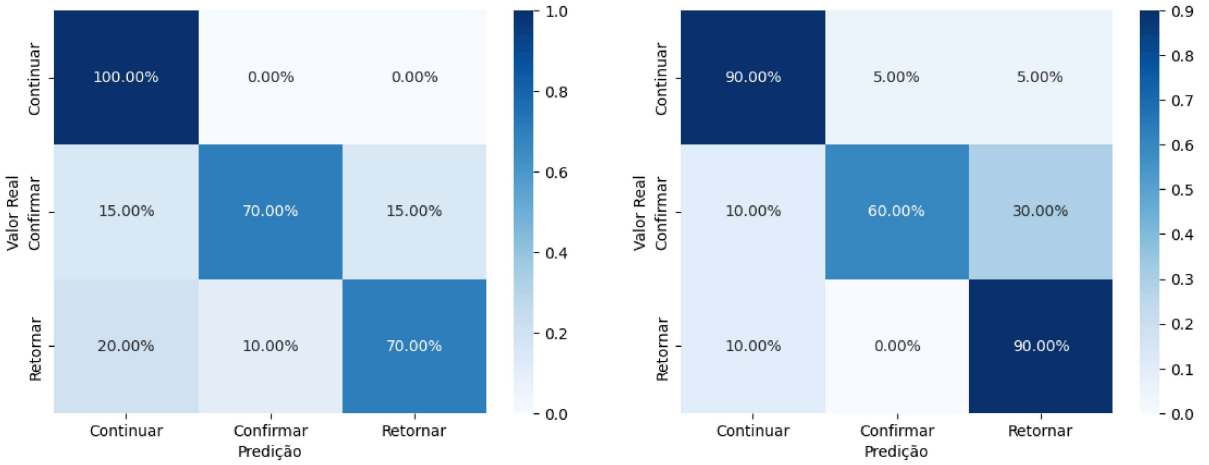
Figura 5.16 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Online: 15 amostras (CNN x LSTM) - 2ª Abordagem



Fonte: Autor, 2023.

Observou-se que ambas as arquiteturas mantiveram desempenho estável com 20 amostras. A CNN alcançou uma acurácia de 80%, enquanto a LSTM também atingiu 80%. A precisão, recall e F1 Score permaneceram consistentes para ambas as redes neurais. Para mais detalhes a matriz de confusão comparativa entre os modelos CNN e LSTM é apresentada na Figura 5.17.

Figura 5.17 – Apresentação da Acurácia através da Matriz de Confusão no Reconhecimento de Gestos Online: 20 amostras (CNN x LSTM) - 2ª Abordagem



Fonte: Autor, 2023.

Os resultados indicam que, em geral, a CNN superou a LSTM em termos de desempenho nas métricas avaliadas. O aumento no número de amostras resultou em melhorias significativas, principalmente para a LSTM, que mostrou maior sensibilidade à quantidade de dados. Este estudo destaca a importância do volume de dados no treinamento de modelos de reconhecimento de gestos online e sugere que a CNN é uma escolha mais robusta para esse tipo de tarefa.

5.5 Discussões

5.5.1 Resultados Offline x Online

Os resultados dos experimentos realizados para o reconhecimento de gestos, tanto em ambientes offline quanto em tempo real (online), proporcionaram insights valiosos sobre o desempenho dos modelos em diferentes contextos de aplicação. Ao comparar as métricas obtidas em ambas as configurações, observamos padrões distintos que refletem as nuances inerentes a cada cenário.

Nos experimentos offline, os modelos de reconhecimento de gestos exibiram desempenho notável, alcançando métricas superiores como era esperado. Esse ambiente oferece condições controladas, sem as restrições temporais e a complexidade adicionada do processamento em tempo real. Os resultados consistentemente elevados nesse contexto indicam que os modelos são capazes de aprender efetivamente e generalizar bem para dados previamente registrados.

Ao transicionar para o modo online, onde a captura e análise de gestos ocorrem em tempo real, observamos uma diminuição nas métricas em comparação com os experimentos offline. Essa disparidade é explicada pelas influências adicionais introduzidas por esse ambiente dinâmico. A latência na aquisição de dados, variações na iluminação e a presença de ruídos em tempo real são fatores que podem impactar a precisão dos modelos.

Apesar dos desafios inerentes ao reconhecimento de gestos em tempo real, notamos resultados particularmente promissores para o modelo classificador treinado a partir de uma Rede Neural Convolucional (CNN). Esse modelo apresentou métricas notáveis, estabilidade consistente e uma eficiência consideravelmente elevada mesmo sob as condições dinâmicas do modo online. Essa robustez sugere uma capacidade excepcional de generalização e adaptação do modelo a variações em tempo real.

Esses resultados indicam que, embora o desempenho em tempo real possa ser ligeiramente inferior devido às complexidades inerentes ao ambiente dinâmico, os modelos de reconhecimento de gestos ainda mantêm um desempenho satisfatório. A eficácia notável da CNN em tempo real sugere que essa arquitetura pode ser especialmente adequada para aplicações práticas que exigem respostas rápidas e precisas, como interfaces de usuário

baseadas em gestos e interações em tempo real.

Em resumo, os experimentos destacam a importância de avaliar o desempenho dos modelos em condições realistas e contextualmente relevantes. A compreensão das limitações e pontos fortes em ambientes offline e online é crucial para a aplicação bem-sucedida desses modelos em diversas situações do mundo real.

5.5.2 Quantidade de Amostras

Os experimentos realizados para o reconhecimento de gestos online e offline, utilizando as redes neurais convolucionais (CNN) e redes neurais recorrentes de longa memória (LSTM), oferecem uma visão detalhada sobre como a quantidade de amostras de treinamento influencia o desempenho da rede. A análise desses resultados destaca padrões distintos, proporcionando insights valiosos sobre a adaptabilidade e eficácia das mesmas em diferentes contextos de dados. Uma visão mais detalhada sobre os resultados obtidos nas diferentes abordagens é discutido a seguir.

5.5.2.1 1ª Abordagem

Nessa abordagem, o conjunto de treinamento é composto por amostras que representam gestos distintos e diversificados. Cada gesto apresenta características únicas, e a variedade nas amostras busca abranger uma gama ampla de possíveis movimentos. Embora essa abordagem promova uma ampla cobertura, ela pode resultar em um modelo mais robusto, capaz de generalizar eficientemente para uma variedade de gestos.

CNN

Os resultados apontam para a capacidade impressionante da CNN em generalizar a partir de quantidades variáveis de amostras de treinamento. A rede não apenas mantém um desempenho elevado com conjuntos menores, mas também demonstra uma notável adaptabilidade à medida que mais dados são fornecidos.

A análise desses resultados destaca a importância de encontrar um equilíbrio adequado entre a quantidade de amostras disponíveis e a acurácia desejada. Enquanto conjuntos menores podem oferecer eficácia aceitável, o aumento gradual na quantidade de dados pode resultar em melhorias incrementais, mas notáveis, no desempenho do modelo.

Essa compreensão é crucial ao projetar sistemas de reconhecimento de gestos, especialmente em cenários onde a coleta de dados pode ser desafiadora. O ajuste fino da quantidade de amostras de treinamento pode permitir a criação de modelos eficientes e precisos, adaptados às necessidades específicas de cada aplicação.

LSTM

Os resultados apontam para a sensibilidade da LSTM à quantidade de amostras de treinamento e sua habilidade de adaptação ao fornecimento de dados adicionais. A ligeira flutuação na acurácia, especialmente ao dobrar o número de amostras, destaca a importância de um equilíbrio adequado entre a quantidade de dados e a capacidade de generalização do modelo.

Embora a quantidade de amostras de treinamento seja crítica, é fundamental reconhecer que há um ponto em que a adição de mais dados pode não resultar em ganhos substanciais de desempenho. Essa análise orienta a criação de conjuntos de dados de treinamento eficazes, alinhando a quantidade de amostras disponíveis com as capacidades de aprendizado e generalização da LSTM.

A estratégia ideal pode envolver uma abordagem incremental ao aumento da quantidade de amostras, monitorando cuidadosamente as mudanças no desempenho para otimizar tanto a eficiência quanto a acurácia do modelo.

5.5.2.2 2ª Abordagem

Nessa abordagem, o foco é colocado em amostras que representam gestos intrinsecamente semelhantes. Isso pode ser útil para ensinar o modelo a reconhecer variações sutis dentro de uma categoria específica de gestos. No entanto, a desvantagem é que, se não houver uma representação adequada de gestos divergentes, a rede pode ter dificuldades em discriminar entre movimentos que compartilham características comuns. Isso pode levar a confusões durante o reconhecimento, especialmente em situações onde a distinção entre gestos é crucial.

CNN

A análise dos resultados destaca a capacidade adaptativa da CNN à quantidade de amostras de treinamento. A melhoria consistente na acurácia sugere uma habilidade robusta de aprendizado e generalização da CNN, especialmente quando apresentada a um conjunto de dados mais variado.

No entanto, a ligeira redução na acurácia com 20 amostras sugere a possibilidade de saturação no aprendizado, indicando que, em determinadas situações, a CNN pode atingir um ponto de estabilização no desempenho, mesmo com o aumento contínuo de dados.

A compreensão desses resultados é vital ao planejar estratégias de treinamento para redes CNN no reconhecimento de gestos semelhantes. Estratégias incrementais de aumento de dados podem ser eficazes até certo ponto, mas é crucial avaliar o ponto de equilíbrio onde mais dados deixam de fornecer ganhos significativos de desempenho.

Essa análise orienta a criação de conjuntos de dados de treinamento eficazes, equilibrando a quantidade de amostras disponíveis com a capacidade de aprendizado e generalização da CNN.

LSTM

A análise dos resultados evidencia a sensibilidade da LSTM à quantidade de amostras de treinamento. A melhoria progressiva na acurácia à medida que mais dados são fornecidos destaca a capacidade adaptativa da LSTM, mas também aponta para um ponto de saturação, onde ganhos significativos tornam-se menos pronunciados com a adição de mais amostras.

Estratégias incrementais de aumento de dados podem ser eficazes para melhorar o desempenho do modelo, mas é essencial reconhecer que há limites na eficácia dessa abordagem.

A análise detalhada da quantidade ótima de amostras pode orientar a criação de conjuntos de dados de treinamento eficazes, equilibrando o aumento de dados com a capacidade de aprendizado e generalização da LSTM.

5.5.3 CNN x LSTM

5.5.3.1 Treinamento

Uma das observações mais marcantes é a diferença significativa no tempo de treinamento entre a CNN e a LSTM. A rede CNN demonstrou um tempo de treinamento consideravelmente inferior em comparação com a LSTM. Essa eficiência temporal é atribuída à arquitetura específica da CNN, que se destaca na extração de características espaciais em dados, otimizando o processo de aprendizado.

A rápida convergência da CNN sugere que essa arquitetura é particularmente eficaz quando há a presença de padrões espaciais significativos nos dados. Em cenários onde a complexidade temporal é relativamente menor, a CNN pode oferecer uma solução mais eficiente.

Outro ponto digno de nota é a capacidade da CNN de atingir uma eficácia satisfatória com um número menor de cenários de treinamento em comparação com a LSTM. Isso sugere que a CNN pode ser mais rápida em aprender padrões distintivos e generalizar para novos exemplos, mesmo quando apresentados com um conjunto de dados menor.

A LSTM, com sua capacidade de lidar com dependências temporais mais complexas, pode exigir uma quantidade maior de dados para aprender efetivamente, especialmente em cenários onde a temporalidade desempenha um papel crucial.

De forma que, a escolha entre CNN e LSTM deve ser guiada pelas características

específicas da tarefa em questão e pelos recursos disponíveis. A eficiência e o desempenho de cada arquitetura podem ser otimizados ao alinhar as características do modelo com as demandas do problema, proporcionando assim soluções mais eficazes em termos de tempo e eficácia.

5.5.3.2 Resultados

Na primeira abordagem, que enfatizava a variedade de gestos, a CNN demonstrou consistentemente um desempenho superior. Mesmo com um número limitado de amostras, a CNN alcançou resultados excepcionais em comparação com a LSTM. À medida que o conjunto de dados se expandiu, a CNN continuou a destacar-se, superando a LSTM de maneira notável.

Na segunda abordagem, onde os gestos apresentavam semelhanças, a CNN também se destacou, mantendo um desempenho superior em relação à LSTM. Independentemente do número de amostras, a CNN demonstrou uma capacidade mais eficaz de lidar com gestos semelhantes, oferecendo resultados mais consistentes e precisos.

Os resultados globais sugerem que, em ambas as abordagens, a CNN mostrou uma vantagem distinta sobre a LSTM em termos de reconhecimento de gestos. A capacidade da CNN em capturar padrões complexos e nuances em gestos diversos ou semelhantes contribuiu para seu desempenho consistente e superior em comparação com a LSTM.

6 Conclusões

O desenvolvimento do Sistema de Comunicação Alternativa apresentado, proporcionou insights valiosos sobre a adaptabilidade e eficácia desses modelos em contextos específicos. A análise dos pontos faciais revelou-se uma abordagem eficaz e benéfica, pois, ao empregar dados consideravelmente simples, em contraste com imagens mais pesadas, otimizamos significativamente os requisitos de recursos computacionais, garantindo eficiência e rapidez no reconhecimento de gestos.

Esta escolha estratégica não apenas simplifica o processo de análise, mas também contribui para uma experiência ágil e responsiva. Esses benefícios destacam a importância de uma seleção criteriosa dos elementos para análise em sistemas de comunicação alternativa.

Além disso, ao analisar os resultados offline e online, observamos que, embora o desempenho em tempo real tenha sido ligeiramente inferior, os modelos de reconhecimento de gestos mantiveram um desempenho satisfatório. Destaca-se a notável eficiência da Rede Neural Convolucional (CNN) em ambientes online, sugerindo sua adequação para aplicações práticas que demandam respostas rápidas e precisas.

Na abordagem com gestos diversos, a CNN mostrou uma capacidade impressionante de generalização, adaptando-se eficientemente a diferentes quantidades de amostras. A análise destaca a importância de equilibrar a quantidade de amostras para otimizar a eficácia do modelo. Enquanto a CNN apresentou melhorias incrementais, a LSTM mostrou sensibilidade à quantidade de dados, ressaltando a necessidade de uma estratégia cuidadosa ao aumentar o conjunto de treinamento.

Na abordagem com gestos semelhantes, ambas as arquiteturas responderam positivamente ao aumento de amostras. A CNN destacou-se na generalização, mesmo com um conjunto de dados mais variado, enquanto a LSTM demonstrou adaptabilidade. A compreensão desses resultados é crucial para equilibrar a quantidade de amostras e otimizar o desempenho na detecção de gestos semelhantes.

A comparação entre CNN e LSTM revelou diferenças marcantes no tempo de treinamento e desempenho. A CNN demonstrou eficiência temporal superior, convergindo rapidamente e alcançando eficácia com conjuntos menores. A capacidade da CNN em destacar padrões espaciais foi evidente, tornando-a uma escolha eficaz em cenários com menor complexidade temporal. A LSTM, embora mais sensível à quantidade de amostras, destacou-se na captura de dependências temporais mais complexas.

Em ambas as abordagens, a CNN mostrou uma vantagem consistente sobre a LSTM em termos de reconhecimento de gestos. A capacidade da CNN em capturar

padrões complexos e nuances em gestos diversos ou semelhantes contribuiu para seu desempenho superior. Essa compreensão orienta a escolha da arquitetura de acordo com as demandas específicas da tarefa e destaca a importância de avaliar o desempenho em condições realistas.

Em resumo, os resultados fornecem uma base sólida para a implementação de um Sistema de Comunicação Alternativa, destacando a eficácia da CNN em ambientes online e offline e oferecendo insights valiosos para o ajuste fino da quantidade de amostras no treinamento de modelos de reconhecimento de gestos. Essas conclusões são cruciais para o avanço na criação de sistemas mais eficientes e adaptáveis para beneficiar indivíduos com distúrbios neuromotores severos.

6.1 Contribuições

O desenvolvimento do Sistema de Comunicação Alternativa proposto, apresentou diversas contribuições significativas para o campo de pesquisa e aplicação prática. A capacidade notável da Rede Neural Convolucional (CNN) em adaptar-se eficientemente a condições em tempo real, mesmo com a presença de desafios como latência na aquisição de dados e variações na iluminação, contribui para a aplicabilidade prática do sistema em ambientes dinâmicos.

Além disso, a observação do tempo de treinamento consideravelmente inferior da CNN em comparação com a LSTM destaca a eficiência dessa arquitetura na extração de características espaciais, sendo essa otimização no treinamento de relevância prática, especialmente em cenários onde a resposta rápida é crucial.

A capacidade de ambas as arquiteturas, CNN e LSTM, de reconhecer gestos diversos e semelhantes destaca a generalização eficaz dos modelos propostos. Essa generalização é fundamental para a aplicação do sistema em uma ampla variedade de contextos, adaptando-se a diferentes conjuntos de gestos.

A aplicação prática do sistema visa beneficiar diretamente indivíduos com distúrbios neuromotores severos, oferecendo-lhes uma alternativa eficaz de comunicação. A contribuição social e humanitária desse trabalho é inegável, fornecendo uma ferramenta acessível e adaptável.

A análise detalhada da quantidade ótima de amostras para o treinamento das redes neurais fornece uma orientação valiosa para o desenvolvimento de conjuntos de dados de treinamento eficazes. Isso permite uma adaptação precisa à capacidade de aprendizado e generalização de cada arquitetura, otimizando a eficácia do reconhecimento de gestos.

6.2 Sugestões de Trabalhos Futuros

Além das contribuições identificadas, algumas sugestões para trabalhos futuros podem aprimorar ainda mais o desenvolvimento e aplicação do Sistema de Comunicação Alternativa.

Uma possibilidade consiste em investigar estratégias para aprimorar a sensibilidade temporal da LSTM, permitindo uma melhor captura de dependências temporais complexas. Isso pode incluir o uso de arquiteturas híbridas ou técnicas específicas de pré-processamento, visando uma melhoria significativa na performance temporal.

Outra área promissora seria a exploração da adaptação do sistema para reconhecimento de outras modalidades de comunicação, como outros movimentos corporais. Essa expansão poderia proporcionar uma comunicação mais rica e inclusiva, abrangendo uma gama mais ampla de expressões e gestos.

Um aspecto crucial a ser explorado diz respeito ao aprimoraramento das instruções para configuração do equipamento associado ao sistema proposto, visando tornar o processo mais claro e intuitivo para os usuários. Investigações adicionais podem se concentrar na criação de diretrizes abrangentes, considerando a usabilidade para diferentes perfis de usuários, e na implementação de interfaces gráficas interativas. Essas melhorias não apenas otimizariam a eficiência operacional, mas também promoveriam a satisfação geral dos usuários, contribuindo para a aceitação mais ampla do sistema.

Adicionalmente, uma abordagem a ser explorada para aprimorar o sistema é a implementação de um mecanismo de ativação baseado no olhar do usuário. Este recurso permitiria que o sistema permanecesse em repouso, evitando a identificação constante de gestos, e fosse ativado somente quando o usuário direcionasse seu olhar para o dispositivo, proporcionando assim um controle mais preciso e eficiente sobre a interação. Essa funcionalidade não apenas conservaria recursos, mas também aprimoraria a experiência do usuário ao oferecer um sistema que responde de maneira imediata e sob demanda.

A realização de estudos clínicos constitui uma sugestão valiosa para avaliar a eficácia do sistema em ambientes do mundo real, envolvendo usuários com distúrbios neuromotores severos. Avaliar a usabilidade e a aceitação do sistema pelos usuários finais é crucial para sua implementação prática, fornecendo insights cruciais para ajustes e melhorias.

Outro ponto relevante é abordar questões éticas e de privacidade associadas ao uso do Sistema de Comunicação Alternativa. Garantir a proteção dos dados sensíveis dos usuários e promover práticas éticas na implementação do sistema são considerações essenciais para seu uso responsável.

A integração de feedback direto dos usuários no processo de treinamento do modelo representa uma abordagem inovadora para a personalização do sistema. Desenvolver

mecanismos que permitam a incorporação de preferências individuais dos usuários pode aprimorar significativamente a experiência e a eficácia do sistema.

Por fim, a investigação de técnicas avançadas de aumento de dados surge como uma estratégia para expandir a diversidade do conjunto de treinamento. Isso pode contribuir para melhorar ainda mais a capacidade de generalização das redes neurais, tornando-as mais robustas diante de diferentes contextos e variações.

Essas sugestões abrem caminhos promissores para a evolução contínua do Sistema de Comunicação Alternativa, buscando constantemente aprimorar sua eficácia, adaptabilidade e impacto positivo na vida dos usuários.

Referências

ACADEMY, D. S. *Deep Learning Book*. [S.l.: s.n.], 2022. <https://www.deeplearningbook.com.br/>. Citado 5 vezes nas páginas 35, 38, 39, 43 e 44.

ALBERT, B.; TULLIS, T. *Measuring the User Experience: Collecting, Analyzing, and Presenting UX Metrics*. [S.l.]: Morgan Kaufmann, 2022. Citado na página 27.

ANDZIK, N. R. et al. National survey describing and quantifying students with communication needs. *Developmental Neurorehabilitation*, Taylor & Francis, v. 21, n. 1, p. 40–47, 2018. Citado na página 16.

(ASHA), A. S.-L.-H. A. *Augmentative and Alternative Communication (Practice Portal)*. 2022. Disponível em: <www.asha.org/Practice-Portal/Professional-Issues/Augmentative-and-Alternative-Communication>. Acesso em: 11 set 2023. Citado 4 vezes nas páginas 15, 24, 25 e 27.

BALL, L. J.; FAGER, S.; FRIED-OKEN, M. Augmentative and alternative communication for people with progressive neuromuscular disease. *Physical Medicine and Rehabilitation Clinics*, Elsevier, v. 23, n. 3, p. 689–699, 2012. Citado 2 vezes nas páginas 29 e 31.

BARBOSA, G. et al. Segurança em redes 5g: Oportunidades e desafios em detecção de anomalias e predição de tráfego baseadas em aprendizado de máquina. In: _____. [S.l.: s.n.], 2021. p. 145–189. ISBN 9786587003658. Citado na página 43.

BAUER, G.; GERSTENBRAND, F.; RUMPL, E. Varieties of the locked-in syndrome. *Journal of neurology*, Springer, v. 221, p. 77–91, 1979. Citado na página 30.

BERSCH, R. de C. R. *Design de um serviço de tecnologia assistiva em escolas públicas*. Dissertação (Dissertação de Mestrado) — Universidade Federal do Rio Grande do Sul, 2009. Citado na página 15.

BERSCH, R. de C. R. *Introdução à Tecnologia Assistiva*. 2018. Disponível em: <www.assistiva.com.br>. Citado 2 vezes nas páginas 22 e 23.

BERTAZZI, R. N. et al. Esclerose lateral amiotrófica. *Revista de Patologia do Tocantins*, v. 4, n. 3, p. 54–65, 2017. Citado na página 28.

BERTONI, A. A. et al. Avaliação de características e previsão de sucesso de canções populares brasileiras por meio de aprendizado de máquina. Universidade Federal de Goiás, 2021. Citado na página 68.

BILESAN, A. et al. Markerless human motion tracking using microsoft kinect sdk and inverse kinematics. In: IEEE. *2019 12th Asian Control Conference (ASCC)*. [S.l.], 2019. p. 504–509. Citado na página 17.

BROOKE, J. Sus: a "quick and dirty'usability. *Usability evaluation in industry*, Taylor & Francis, v. 189, n. 3, p. 189–194, 1996. Citado na página 27.

BROWN, M. N.; GRAMES, L. M.; SKOLNICK, G. B. Augmentative and alternative communication (aac) use among patients followed by a multidisciplinary cleft and craniofacial team. *The Cleft Palate-Craniofacial Journal*, SAGE Publications Sage CA: Los Angeles, CA, v. 58, n. 3, p. 324–331, 2021. Citado na página 16.

COELHO, Y. et al. Um novo sistema de comunicação aumentativa e alternativa baseado em rastreamento do olhar. 2016. Citado 2 vezes nas páginas 28 e 32.

CREER, S. et al. Prevalence of people who could benefit from augmentative and alternative communication (aac) in the uk: determining the need. *International Journal of Language & Communication Disorders*, Wiley Online Library, v. 51, n. 6, p. 639–653, 2016. Citado na página 16.

DEVELOPERS scikit-learn. *Metrics and scoring: quantifying the quality of predictions*. 2020. Disponível em: <https://scikit-learn.org/stable/modules/model_evaluation.html>. Citado na página 69.

ELLIOTT, E. et al. An epidemiological profile of dysarthria incidence and assistive technology use in the living population of people with mnd in scotland. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, Taylor & Francis, v. 21, n. 1-2, p. 116–122, 2020. Citado na página 16.

FREEMAN-SANDERSON, A.; MORRIS, K.; ELKINS, M. Characteristics of patient communication and prevalence of communication difficulty in the intensive care unit: an observational study. *Australian Critical Care*, Elsevier, v. 32, n. 5, p. 373–377, 2019. Citado na página 16.

FUNKE, A. et al. Provision of assistive technology devices among people with als in germany: a platform-case management approach. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, Taylor & Francis, v. 19, n. 5-6, p. 342–350, 2018. Citado na página 16.

GARCIA, J. C. D.; FILHO, T. A. G. Pesquisa nacional de tecnologia assistiva. *São Paulo: ITS Brasil/MCTI-Secis*, v. 68, 2012. Citado na página 21.

GOMIDE, R. d. S. et al. A new concept of assistive virtual keyboards based on a systematic review of text entry optimization techniques. *Research on Biomedical Engineering*, SciELO Brasil, v. 32, p. 176–198, 2016. Citado na página 31.

HANCHETT, E.; LISTWON, B. *Vue. js in Action*. [S.l.]: Simon and Schuster, 2018. Citado na página 59.

HAUSAMANN, P. et al. Evaluation of the intel realsense t265 for tracking natural human head motion. *Scientific reports*, Nature Publishing Group UK London, v. 11, n. 1, p. 12486, 2021. Citado na página 17.

HWANG, C.-S. et al. An eye-tracking assistive device improves the quality of life for als patients and reduces the caregivers' burden. *Journal of motor behavior*, Taylor & Francis, v. 46, n. 4, p. 233–238, 2014. Citado na página 31.

IACONO, T.; TREMBATH, D.; ERICKSON, S. The role of augmentative and alternative communication for children with autism: current status and future trends. *Neuropsychiatric disease and treatment*, Taylor & Francis, p. 2349–2361, 2016. Citado na página 16.

JUDGE, S. et al. Provision of powered communication aids in the united kingdom. *Augmentative and Alternative Communication*, Taylor & Francis, v. 33, n. 3, p. 181–187, 2017. Citado na página 16.

JÚNIOR, I. A. d. M. P.; KNOP, I. de O. Construcao de um software media center com reconhecimento de gestos e comandos de voz utilizando o microsoft kinect e os principios de natural user interface. *Caderno de Estudos em Sistemas de Informação*, v. 1, n. 1, 2015. Citado na página 17.

KÄTHNER, I.; KÜBLER, A.; HALDER, S. Comparison of eye tracking, electrooculography and an auditory brain-computer interface for binary communication: a case study with a participant in the locked-in state. *Journal of neuroengineering and rehabilitation*, BioMed Central, v. 12, n. 1, p. 1–11, 2015. Citado na página 31.

KRISTOFFERSSON, E.; SANDBERG, A. D.; HOLCK, P. Communication ability and communication methods in children with cerebral palsy. *Developmental Medicine & Child Neurology*, Wiley Online Library, v. 62, n. 8, p. 933–938, 2020. Citado na página 16.

KRUG, S. *Rocket surgery made easy: The do-it-yourself guide to finding and fixing usability problems*. [S.l.]: New Riders, 2009. Citado na página 27.

KUHLMAN, D. *A python book: Beginning python, advanced python, and python exercises*. [S.l.]: Dave Kuhlman Lutz, 2009. Citado na página 59.

LIN, S. C.; GOLD, R. S. Assistive technology needs, functional difficulties, and services utilization and coordination of children with developmental disabilities in the united states. *Assistive Technology*, Taylor & Francis, v. 30, n. 2, p. 100–106, 2018. Citado na página 16.

LOJA, L. F. B. et al. Tecnologia assistiva: um teclado virtual evolutivo para aplicação em sistemas de comunicação alternativa e aumentativa. Universidade Federal de Uberlândia, 2015. Citado 3 vezes nas páginas 24, 26 e 28.

LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, SciELO Brasil, v. 35, p. 85–94, 2021. Citado na página 33.

MADHAVAN, S.; JONES, M. T. *Deep learning architectures*. 2017. Disponível em: <https://developer.ibm.com/articles/cc-machine-learning-deep-learning-architectures/>. Citado na página 44.

MARQUES, G. *Redes neurais convolucionais na classificação de imagens médicas*. 2023. Disponível em: <https://profissaobiotec.com.br/redes-neurais-convolucionais-classificacao-imagens-medicas/>. Citado 2 vezes nas páginas 40 e 41.

MCCLURE, N. *TensorFlow machine learning cookbook*. [S.l.]: PACKT publishing Ltd, 2017. Citado 2 vezes nas páginas 61 e 62.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, p. 115–133, 1943. Citado na página 35.

MELO, A. R. L. L. d. *Um estudo sobre o mapeamento de gestos do Leap motion para a língua brasileira de sinais-(Libras)*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2015. Citado na página 17.

NARCIZO, H. F.; CÂMARA, C. E. Desenvolvimento de protótipo de software para reconhecimento de posições e gestos utilizando microsoft kinect. *Revista Engenho*, v. 5, n. 7, p. 61–86, 2013. Citado na página 17.

NISBET, P. Alternative access technologies. In: *Handbook of Electronic Assistive Technology*. Elsevier, 2019. p. 105–148. Disponível em: <https://doi.org/10.1016/b978-0-12-812487-1.00005-3>. Citado na página 21.

OLIVEIRA, A. M. F. Sistemas aumentativos e alternativos de comunicação na esclerose lateral amiotrófica: Aplicabilidade e utilidade nos doentes, cuidadores e profissionais de saúde. 2019. Citado na página 31.

PAIXAO, G. M. d. M. et al. Machine learning na medicina: Revisão e aplicabilidade. *Arquivos Brasileiros de Cardiologia*, Sociedade Brasileira de Cardiologia - SBC, v. 118, n. 1, p. 95–102, Jan 2022. ISSN 0066-782X. Disponível em: <https://doi.org/10.36660/abc.20200596>. Citado na página 34.

PINHEIRO, R. P. *Mecanismo de cibervigilância baseado em aprendizado de máquina para detecção de malwares*. [S.l.]: Editora Dialética, 2023. Citado 2 vezes nas páginas 37 e 38.

RASCHKA, S.; MIRJALILI, V. *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. [S.l.]: Packt Publishing Ltd, 2019. Citado 3 vezes nas páginas 36, 37 e 61.

SAKURAI, R. *Implementando a estrutura de uma Rede Neural Convolucional utilizando o MapReduce do Spark*. 2017. Disponível em: <https://www.sakurai.dev.br/cnn-mapreduce/>. Citado na página 42.

SALOMON, J. *Lung Cancer Detection using Deep Convolutional Networks Final Year Project Report*. Tese (Doutorado) — Tese de doutorado, 2018. Citado na página 42.

SILVA, R. *Desenvolvimento de um teclado virtual para comunicação por meio de gestos visuais*. Tese (Doutorado) — Universidade Federal de Uberlândia, 2021. Disponível em: <https://doi.org/10.14393/ufu.te.2021.19>. Citado 2 vezes nas páginas 24 e 31.

SILVA, R. A. d. et al. Desenvolvimento de um teclado virtual para comunicação por meio de gestos visuais. Universidade Federal de Uberlândia, 2021. Citado na página 25.

SILVA, R. A. da; VEIGA, A. C. P. Algorithm for decoding visual gestures for an assistive virtual keyboard. *IEEE Latin America Transactions*, v. 18, n. 11, p. 1909–1916, Mar. 2021. Disponível em: <https://latamt.ieeer9.org/index.php/transactions/article/view/3858>. Citado 2 vezes nas páginas 15 e 31.

SILVA, R. V. T. e et al. Inteligência artificial e o teste de turing: uma análise do prêmio loebner de 2017 e 2018. *Revista Científica Multidisciplinar Núcleo do Conhecimento*, Revista Cientifica Multidisciplinar Nucleo Do Conhecimento, p. 121–141, mar. 2022. Disponível em: <https://doi.org/10.32749/nucleodoconhecimento.com.br/tecnologia/premio-loebner>. Citado na página 36.

SMITH, E.; DELARGY, M. Locked-in syndrome. *Bmj*, British Medical Journal Publishing Group, v. 330, n. 7488, p. 406–409, 2005. Citado na página 30.

SOUSA, L. et al. Interface natural de utilizador baseado em reconhecimento de gestos usando o sensor leap motion. *Dos Algarves: Tourism, Hospitality & Management Journal*, v. 1, n. 26, p. 106–129, 2017. Citado na página 17.

SPOLSKY, A. J. *User interface design for programmers*. [S.l.]: Apress, 2008. Citado na página 27.

VANDERPLAS, J. *Python data science handbook: Essential tools for working with data*. [S.l.]: "O'Reilly Media, Inc.", 2016. Citado na página 60.

VASCONCELOS, T. G. d. Leap motion como tecnologia assistiva para pessoas com deficiência motora nos membros superiores. Universidade Federal da Paraíba, 2017. Citado na página 17.

VOIGT, J. F. et al. Aprendizagem profunda para reconhecimento de gestos da mão usando imagens e esqueletos com aplicações em libras. Universidade Federal de Alagoas, 2018. Citado 4 vezes nas páginas 17, 39, 41 e 44.

XIII, C. E.-D. Educação em tecnologias de apoio para utilizadores finais. *Linhas de Orientação para Formadores*, 1999. Citado 3 vezes nas páginas 21, 23 e 31.

ZHANG, A. et al. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021. Citado na página 43.

ZHANG, X.; KULKARNI, H.; MORRIS, M. R. Smartphone-based gaze gesture communication for people with motor disabilities. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. [S.l.: s.n.], 2017. p. 2878–2889. Citado na página 32.

ZUBOW, L.; HURTIG, R. A demographic study of aac/at needs in hospitalized patients. *Perspectives on Augmentative and Alternative Communication*, ASHA, v. 22, n. 2, p. 79–90, 2013. Citado na página 16.