FACULDADE DE TECNOLOGIA DE SÃO JOSÉ DOS CAMPOS
FATEC PROFESSOR JESSEN VIDAL

Felipe Menino Carlos

Desenvolvimento de recursos de tecnologia assistiva utilizando técnicas de Deep Learning: um estudo de casos

São José dos Campos
2018

Felipe Menino Carlos

Desenvolvimento de recursos de tecnologia assistiva utilizando técnicas de Deep Learning: um estudo de casos

Trabalho de Graduação apresentado à Faculdade de Tecnologia São José dos Campos, como parte dos requisitos necessários para a obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

FACULDADE DE TECNOLOGIA DE SÃO JOSÉ DOS CAMPOS

FATEC PROFESSOR JESSEN VIDAL

Orientador: Me. Giuliano Araujo Bertoti

São José dos Campos

2018

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

RNA - Redes Neurais Artificiais

AM - Aprendizado de Máquina

DL - Deep Learning

CNN - *Convolutional Neural Network*

IBGE - Instituto Brasileiro de Geografia e Estatística

Sumário

1 Introdução

Este capítulo demonstra a motivação para o desenvolvimento deste trabalho, os objetivos deste e a metodologia adotada.

1.1 Motivação

De acordo com o censo do IBGE, realizado em 2010, no Brasil, há cerca de 45 milhões de pessoas com algum tipo de deficiência. E todas estas pessoas necessitam de uma vida independente e de inclusão (SARTORETTO; BERSCH, 2017).

Uma das maneiras de permitir que deficientes sejam inclusos na sociedade e tenham uma vida autônoma é com a utilização de recursos de tecnologias assistivas, estas aliadas a serviços de tecnologia assitiva. Isto porque, estes recursos assitivos deixam de lado a deficiência, e focam nas habilidades presentes no individuo.

No Brasil, há uma grande dificuldade de acesso aos recursos de tecnologia assitiva, causadas por diversos fatores, a citar, o alto custo e a necessidade de importação (ANDRIOLI, 2017). O alto custo, na maioria dos casos pode ser justificado pela necessidade de desenvolvimento e construção de equipamentos específicos, o que acaba gerando um alto valor de compra, com equipamentos chegando em valores próximos a 15 mil reais.

Por outro lado, tem-se técnicas de *Deep Learning*, que são atualmente o estado-da-arte da solução de problemas com aprendizado de máquina (PONTI, 2017), isto por conta da grande capacidade de generalização diante de diferentes conjuntos de dados. Um de seus grandes benefícios é a possibilidade de alta personalização frente a diferentes tipos de usuários e aplicações.

Desta forma, este trabalho foi motivado pela possibilidade da realização de um estudo de casos, onde técnicas de *Deep Learning* são aplicadas para possibilitar a criação de recursos de tecnologias assistivas de baixo custo.

1.2 Objetivo Geral

Implementar recursos de tecnologias assistivas de baixo custo, para usuários com deficiências auditiva e motora, utilizando técnicas de *Deep Learning*

1.3 Objetivo Específico

Para a consecução deste objetivo foram estabelecidos os seguintes objetivos específicos:

- Desenvolvimento de uma ferramenta que permite a movimentação do *cursor* do *mouse* através de movimentos da cabeça, com foco em usuários com deficiência motora;
- Desenvolvimento de uma ferramenta que permite a realização de alguns processos no computador através de comandos de voz, com foco em usuários com deficiência motora;
- Desenvolvimento de uma ferramenta que permite a escrita de textos utilizando LIBRAS, para usuários com deficiência auditiva;
- Integração das ferramentas desenvolvidas.

1.4 Metodologia

A realização dos objetivos específicos do trabalho é feita através da aplicação de modelos de DL no desenvolvimento das ferramentas. A linguagem de programação empregrada para o desenvolvimento das ferramentas é o Javascript, junto ao *framework* de desenvolvimento de DL *Tensorflow.js*

Por contar com diferentes estudos de caso, todas as ferramentas são desenvolvidas de maneira modular, a permitir que, no momento da integração entre as ferramentas desenvolvidas, injeções de dependências sejam realizadas para tal.

Cada uma das ferramentas desenvolvidas, faz a utilização de um modelo de DL, no caso do controle do *mouse* o modelo Posenet (KENDALL, 2015) é utilizado, para facilitar a identificação de pontos faciais do usuário, e permitir que cada gesto seja mapeado em movimentos do *mouse*, o reconhecimento de voz, por sua vez, é feito com o *Web Speech API*, uma *API* livre que facilita a sintetização de som em texto. Por fim, para a tradução de LIBRAS em texto, será utilizado uma rede neural convolucional (LECUN et al, 1998), que apresenta bons resultados na classificação de imagens, junto a um conjunto de imagens de LIBRAS criado pelo autor.

1.5 Organização do trabalho

Este Trabalho está organizado nos seguintes capítulos:

- Capítulo 2: Revisão bibliográfica
- Capítulo 3: Desenvolvimento
- Capítulo 4: Resultados
- Capítulo 5: Considerações finais

2 Fundamentação Teórica

Neste capítulo serão fundamentados os conhecimentos básicos para o entendimento do trabalho.

2.1 Deficiência

De acordo com o censo do IBGE, realizado em 2010, cerca de 6.2% da população brasileira possui algum tipo de deficiência. E a necessidade de inclusão destas pessoas na sociedade é extremamente importante. Do grupo citado anteriormente, cerca de 1.3% tem algum tipo de deficiência auditiva, e 1.1% tem deficiências auditivas

Para aqueles com deficiência auditiva, a comunicação e inclusão pode ser feita através da Linguagem Brasileira de Sinais, segunda língua oficial do Brasil desde 2005. Mas, pode-se encontrar problemas com a comunicação através de LIBRAS principalmente pelo fato de, boa parte dos ouvintes não falar esta língua o que acarreta também na baixa utilização desta em diversos meios de comunicação. Um ponto importante apontado no documentário feito pela TVE RS, é que, pessoas com deficiência auditiva, normalmente são alfabetizadas somente com LIBRAS, por terem muita difículdade e falta de estrutura para o aprendizado da Língua Portuguesa. Ainda de acordo com o documentário, para as pessoas com deficiências motoras há os recursos de tecnologias assitivas, que aumentar a facilidade do acesso destas pessoas aos meios sociais, principalmente os digitais.

2.2 Tecnologias assistivas

Uma das formas de realizar a inclusão social de pessoas com deficiência é através da inclusão digital, utilizando técnologias assitivas, estes que visam ampliar as habilidades presentes no indivíduo, não o forçando a ter características específicas para a inclusão (NTAAI, 2016). As tecnologias assitivas, de acordo com o Núcleo de Tecnologia Assistiva, Acessibilidade e Inovação da Universidade de Brasilia, podem ser divididas em dois grupos, os recursos, que representam equipamentos que expandem as habilidades dos indivíduos com deficiência, e os serviços, que normalmente são aqueles relacionados a facilitação e capacitação para o uso correto dos recursos assistivos.

Para aqueles que possuem deficiência motora, a inclusão digital vem através de *mouses* adaptados, formas diferentes de utilizar teclados, ou até mesmo a utilização do computador por comandos de voz. E para os surdos ferramentas que ajudam no processo de interação já levando em consideração problemas com a alfabetização.

2.3 Inteligência artificial

Sistemas inteligentes de forma geral são aqueles que apresentam a capacidade de planejar e resolver problemas através de dedução e indução utilizando conhecimentos de situações anteriores (ZUBEN, 2013), e a inteligência artificial, é um campo da ciência e engenharia de computação (ZUBEN, 2013), que possibilitam a sistemas computacionais, perceber, raciocionar e agir (WINSTON, 1992).

Uma das técnicas computacionais mais utilizadas para o desenvolvimento e aplicação de inteligência artificial, são aquelas relacionadas ao aprendizado de máquina. Esta que é uma área que tem como objetivo principal, desenvolver técnicas que permitam aos sistemas adquirir conhecimento de forma automática e com estes conhecimentos tomar decisões (AUGUSTO, 2007).

Para a realização do aprendizado de máquina, existem diversas técnicas, que vão de simples regressões estatísticas, até modelos complexos, como às redes neurais artificiais (NG, 2016).

2.4 Redes neurais artificiais

Como descrito na seção anterior, uma das áreas de aplicação do aprendizado de máquina mais avançadas atualmente são as RNA, que imitam principalmente aspectos do funcionamento do corpo humano, neste caso, o cérebro e suas redes neuronais (CINTRA, 2015).

2.4.1 Neurônio biológico

Todo o processamento de informações no cérebro humano, é feito através de elementos biológicos de processamento, que operam em paralelo para a produção de ações apropriadas para cada estimulo recebido pelo corpo. A célula base do sistema nervoso cerebral é o neurônio (Figura 1), e sua principal função é conduzir impulsos (Representando os estimulos) levando em consideração as condições do corpo e assim produzindo ações. Os neurônios também são os resposáveis pelos atos do pensamento e armazenamento de informações (NUNES et al, 2016).

O neurônio podem ser divididos em três partes elementares, os dendritos, que captam de forma continua os impulsos vindos de outros neurônios, o corpo celular, que processa todas as informações captadas e os axônio que enviam as informações processadas no corpo celular para outros neurônios.

Estima-se que a rede neural cerebral, possui cerca de 100 bilhões de neurônios, cada um destes mantendo conexão com uma média de 6.000 outros neurônios, gerando

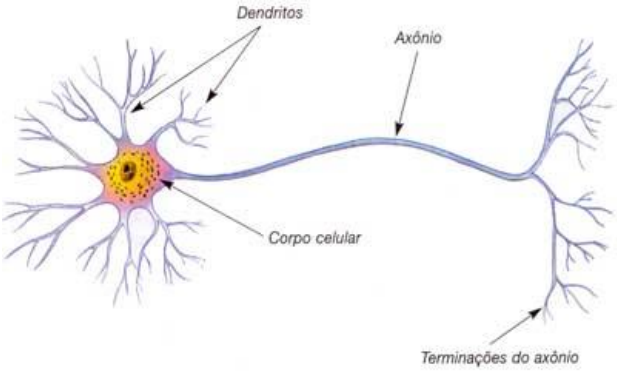

Figura 1 – Ilustração do neurônio biológico

cerca de 600 trilhões de conexões (SHEPHERD, 1990). As conexões entre os neurônios são chamados de sinapses, estas que como apresentado por Donald Hebb em 1949, em seu livro *The Organization of Behavior* tem os caminhos fortalecidos toda vez que é utilizado, assim, pode-se entender que, neurônios tem propensões para certas atividades, quando os neurônios utilizados por esta tem suas sinapses bem fortalecidas (XAVIER, 2017).

A Figura 2 demonstra um exemplo de uma pequena parte das redes neuronais responsáveis pelo cortex auditivo.

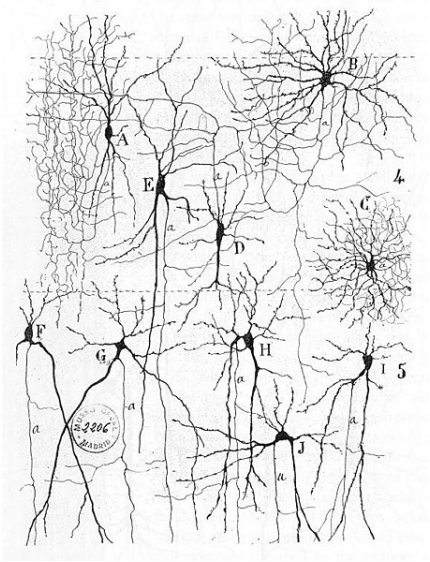

Figura 2 – Rede neural do cotex auditivo

A representação inicial deste conjunto de neurônios em sistemas de computação foram implementadas através de circuitos elétricos (MCCULLOCH; PITTS, 1943), estes que foram utilizados como base para a criação dos modelos de neurônios artificiais (HODGKIN; HUXLEY, 1952).

2.4.2 Neurônio artificial

Os neurônios artificias, que são modelos de representações dos neurônios biológicos compoem a RNA. O principal modelo de neurônio artificial utilizado, mesmo em arquiteturas mais atuais, é o proposto por McCulloch e Pitts em 1943 (Figura 3). Neste há componentes que fazem referência direta ao neurônio biológico visto anteriormente.

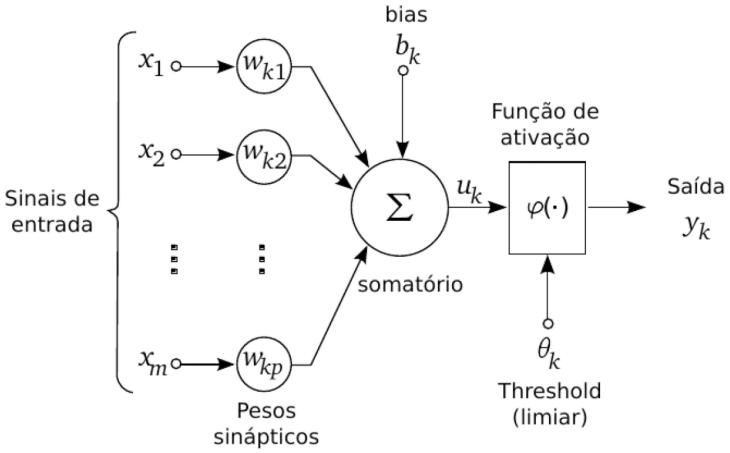

Figura 3 – Modelo de Neurônio artificial

Os sinais de entrada, apresentados na Figura 3, advindos do meio externo, normalmente uma aplicação, são análogos aos impulsos elétricos captados pelos dendritos no neurônio biológico. Os pesos sinápticos representam a importância do sinal recebido para o neurônio, o que representa as ponderações exercidas pelas junções sinápticas do modelo biológico, ou seja, a força do caminho entre as sinapses, citados anteriormente. O campo de somatório junto a função de ativação, representam o corpo celular do neurônio biológico, é nesta parte que os resultados criados pelo neurônio são calculados (NUNES et al, 2016).

Através destes modelos as redes neurais são compostas com dezenas ou até mesmo centenas de neurônios, dependendo exclusivamente da complexidade do problema a ser resolvido. As RNA, ainda são divididas em camadas, onde cada uma delas tem uma função específica, uma RNA simples, normalmente tem-se três camadas apenas (Figura 4), uma camada de entrada, uma camada oculta, que contém um conjunto de neurônios e a camada de saída dos resultados.

Na camada de entrada, os dados são recebidos e enviados para a camada oculta, na camada oculta, os dados são processados pelos neurônios, e seus resultados são unidos na camada de saída (PONTI, 2017).

Um tipo muito comum de rede neural com esta estrutura básica são os chamados **Perceptron**, que é uma rede neural simples, criada por Frank Rosenblat em 1957. Porém redes neurais com apenas uma camada de processamento (Camada oculta) não são

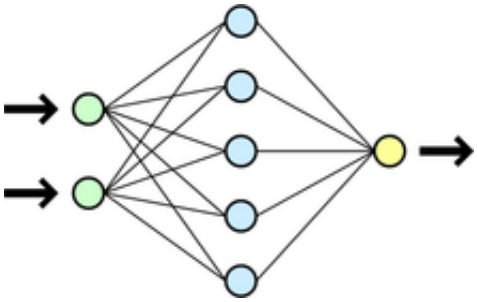

Figura 4 – Rede neural artificial simples

aplicáveis em diversos casos, principalmente pela limitação no nível de complexidade de problemas que podem ser resolvidos com este tipo de rede (PONTI, 2017). Problemas como classificação de imagens, reconhecimento de objetos e detecção de fraudes podem ser um grande desafio para estes tipos de arquitetura, desta forma, mais camadas devem ser inseridas para tratamentos mais sofisticados dos dados, gerando resultados mais acertivos e principalmente, resolvendo problemas mais complexos.

2.4.3 Processo de aprendizado

Um dos pontos mais relevantes das RNA é seu poder de generalização, assim, após aprender a realizar alguma atividade, levando em consideração um determinado conjunto de dados, estas redes conseguem realizar atividades com diferentes conjuntos de dados. Porém isto exige um processo de treinamento bem definido, seguindo um algoritimo de treinamento, este algoritimo é o processo de treinamento. (NUNES et al, 2016).

Estes processos de treinamento podem adotar diferentes estratégias para ensinar as RNA, e cada estratégia gera um algoritimo de aprendizado diferente, sendo os principais, os algoritimos de aprendizado supervisionado e não-supervisionado.

2.4.3.1 Aprendizado supervisionado

No aprendizado supervisionado, o usuário indica o comportamento que a RNA deve apresentar dado um conjunto de dados qualquer, desta forma, a RNA pode ir ajustando os pesos sinápticos de seus neurônios com o objetivo de produzir o mesmo resultado apresentado pelo usuário (OSÓRIO, 1999). Levando em consideração um problema de classificação, onde um conjunto de dados é apresentado para a RNA, e ela deve informar ao usuário o que cada um dos dados daquele conjunto representa.

2.4.3.2 Aprendizado não-supervisionado

O aprendizado não supervisionado é completamente o oposto do supervisionado, neste é apresentado para a RNA apenas o conjunto de dados, e a RNA se encarrega de

aprender sobre aquele conjunto de dados. Este tipo de aprendizado pode ser utilizado para deixar a RNA identificar os padrões presentes nos dados, e tirar informações destes padrões (NG, 2013).

2.5 Deep Learning

O *Deep Learning* apresenta uma abordagem diferente para os problemas resolvidos com técnicas de RNA, porém, no caso de DL, muitas camadas são empregadas (GOODFELLOW, 2016) nas arquiteturas das redes neurais (Figura 5).

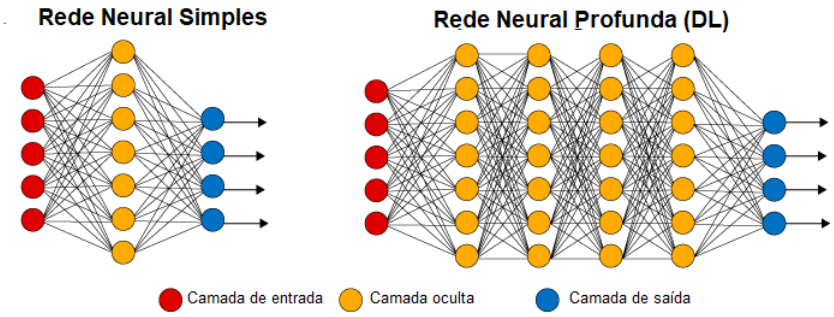

Figura 5 – Diferenças entre redes neurais simples e Deep Learning

A utilização de múltiplas camadas, cada um com dezenas de neurônios, permitiu as técnicas de DL chegarem ao estado-da-arte em muitos problemas que envolvem o AM (SHANKAR, 2017). Além disto, de acordo com Andrew NG, o processo de aprendizado destas redes melhora muito com o aumento dos dados (Figura 6), diferente do que ocorria com arquiteturas e algoritimos de aprendizado de máquinas antigos.

Ainda de acordo com Andrew, isto ocorre pois ao utilizar múltiplas camadas, diversos recursos são captados dos dados, fazendo com que o processo de aprendizado se torne eficaz, e tende a melhorar ainda mais com o aumento da quantidade de dados utilizados no processo de treinamento.

A utilização de múltiplas camadas, permitiram que diferentes técnicas pudessem ser utilizadas dentro de uma rede neural, e isto fez com que diversas arquiteturas, para os mais váriados fins fossem criados.

2.5.1 Redes neurais convolucionais

Redes neurais convolucionais, do inglês, *Convolutional Neural Network* são um tipo de rede neural profunda, especializadas em análise de elementos visuais, tais como imagens e vídeos (SAVARESE, 2018). Sua especialidade em dados visuais permitiu um grande avanço nas áreas de visão computacional, especialmente por estar serem mais fáceis de

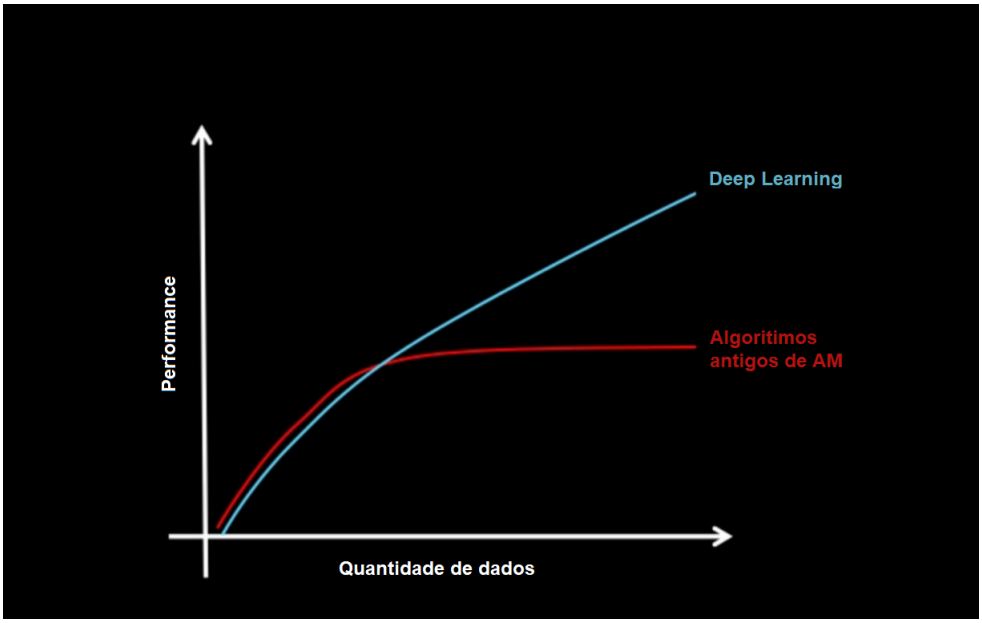

Figura 6 – Deep Learning X Quantidade de dados

treinar, quando comparado a redes neurais comuns em trabalhos com imagens (ARAÚJO, 2017).

Um dos primeiros modelos de CNN propostos foi a LeNet (LECUN et al, 1998), proposta por Yann LeCunn em 1998, e mesmo com a evolução dos conceitos deste tipo de rede, os conceitos apresentados por LeCun continuam sendo aplicados. Nesta arquitetura, uma sequência de camadas convolucionais, de *pooling* e totalmente conectadas são utilizadas (ARAÚJO, 2017).

As camadas convolucionais, que são a grande diferença das CNN para outros tipos de RNA, trabalham como filtros, recuperando apenas pontos importantes da imagem para a classificação, isto através de uma matriz de pesos que é utilizada nas convoluções (ARAÚJO, 2017). Após o filtro realizado por esta camada, as imagens resultantes do filtro são passadas para a camada de *pooling*, estas camadas que básicamente reduzem a dimensionalidade das resultantes. Por fim, as camadas totalmente conectadas são as responsáveis em realizar a multiplicação ponto a ponto dos sinais recebidos (imagens) e aplicar uma função de ativação, que produzirá a probabilidade de cada uma das classes esperadas na classificação (ARAÚJO, 2017).

A Figura 7 demonstra a arquitetura de LeCun sendo utilizada para a classificação de imagens de tumores, podendo ter como resultado às classes **normal** ou **anormal**.

Veja que, o diferencial citado acima, na utilização das convoluções está justamente na quantidade de elementos que são utilizados para a classificação, em RNA comuns, ao realizar a classificação de imagens, deve-se ter de neurônios na RNA a mesma quantidade de píxels presentes na imagem a ser classificada, o que nas CNN não ocorre, exatamente

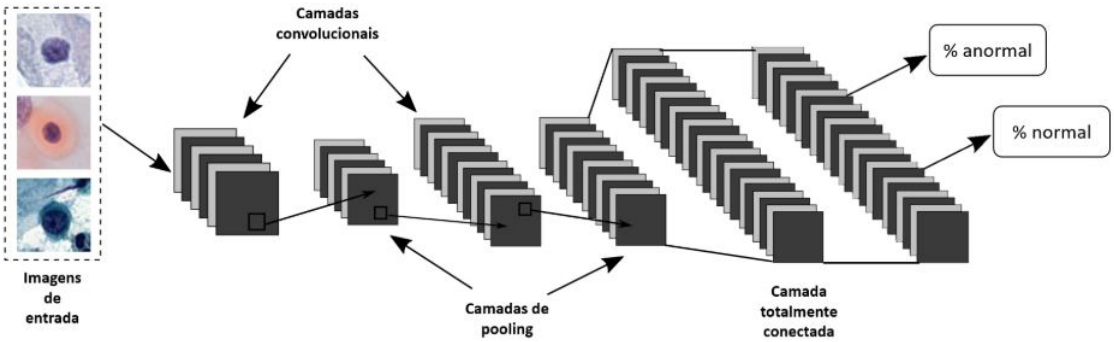

Figura 7 – Estrutura básica de CNN proposta por LeCun, 1998

por conta dos filtros que são criados (PONTI, 2017).

2.5.2 Posenet

Posenet é um tipo de CNN, para a identificação em tempo real de pontos do corpo dos usuários, o modelo desenvolvido para esta rede é extremamente robusto e permite a identificação das poses mesmo quando há problemas com luz e iluminação do ambiente que está sendo levado em consideração na classificação (KENDALL, 2015).

2.6 Tecnologias

Esta seção demonstra as tecnologias utilizadas durante a implementação do presente trabalho.

2.6.1 Tensorflow.js

Tensorflow.js é uma biblioteca para a linguagem de programação Javascript, que permite a implementação de modelos de AM com grande facilidade de expressão. A biblioteca é flexível e pode ser usada para expressar uma ampla variedade de algoritmos, incluindo algoritmos de treinamento e inferência para modelos de DL, e tem sido usado para conduzir pesquisas e implantar sistemas de aprendizado de máquina nas mais diversas áreas, envolvendo trabalhos como reconhecimento de fala, visão computacional e robótica (ABADI et al, 2015).

2.6.2 Google Colaboratory

Colaboratory é uma ferramenta criada pelo Google, que permite a fácil execução de algoritimos de aprendizado de máquina. O ambiente é criado sobre o pacote Jupyter,

um ambiente interativo e simples para a execução de código, com a diferença de que, no Colaboratory, toda a execução pode ser feita utilizando máquinas disponibilizadas pelo Google.

2.6.3 Processamento de voz

O processamento de voz é uma área que engloba diversas técnicas diferentes, como o reconhecimento de fala natural e síntese de voz (GUILHOTO, 2001). Para a aplicação de reconhecimentos de comandos por voz, utiliza-se o reconhecimento de palavras, este que é caracterízado pelo processamento de um pequeno trecho da fala, com o objetivo de identificar qual ação o sistema deve tomar (GUILHOTO, 2001).

2.6.3.1 Web speech API

Normalmente a utilização de comandos voz em sistemas pode ser muito complexa, pois esta envolve diversos conceitos, das mais variadas áreas de estudo. Porém, a utilização de comandos de voz torna todo sistema que o utiliza, mais acessível a pessoas com deficiência, desta forma o *Web Speech API* foi desenvolvido. Esta é uma ferramenta que facilita o processo de análise e sintetização de voz em texto (ADORF, 2013).

Sua utilização permite uma rápida aplicação do processamento de voz em sistemas criados utilizando a linguagem de programação Javascript.

Referências