



38º JORNADA DE ATUALIZAÇÃO EM INFORMÁTICA (JAI) AUTENTICAÇÃO USANDO SINAIS BIOMÉTRICOS: FUNDAMENTOS, APLICAÇÕES E DESAFIOS

BELÉM, PA
Brasil



RNP



Apresentação Disponível



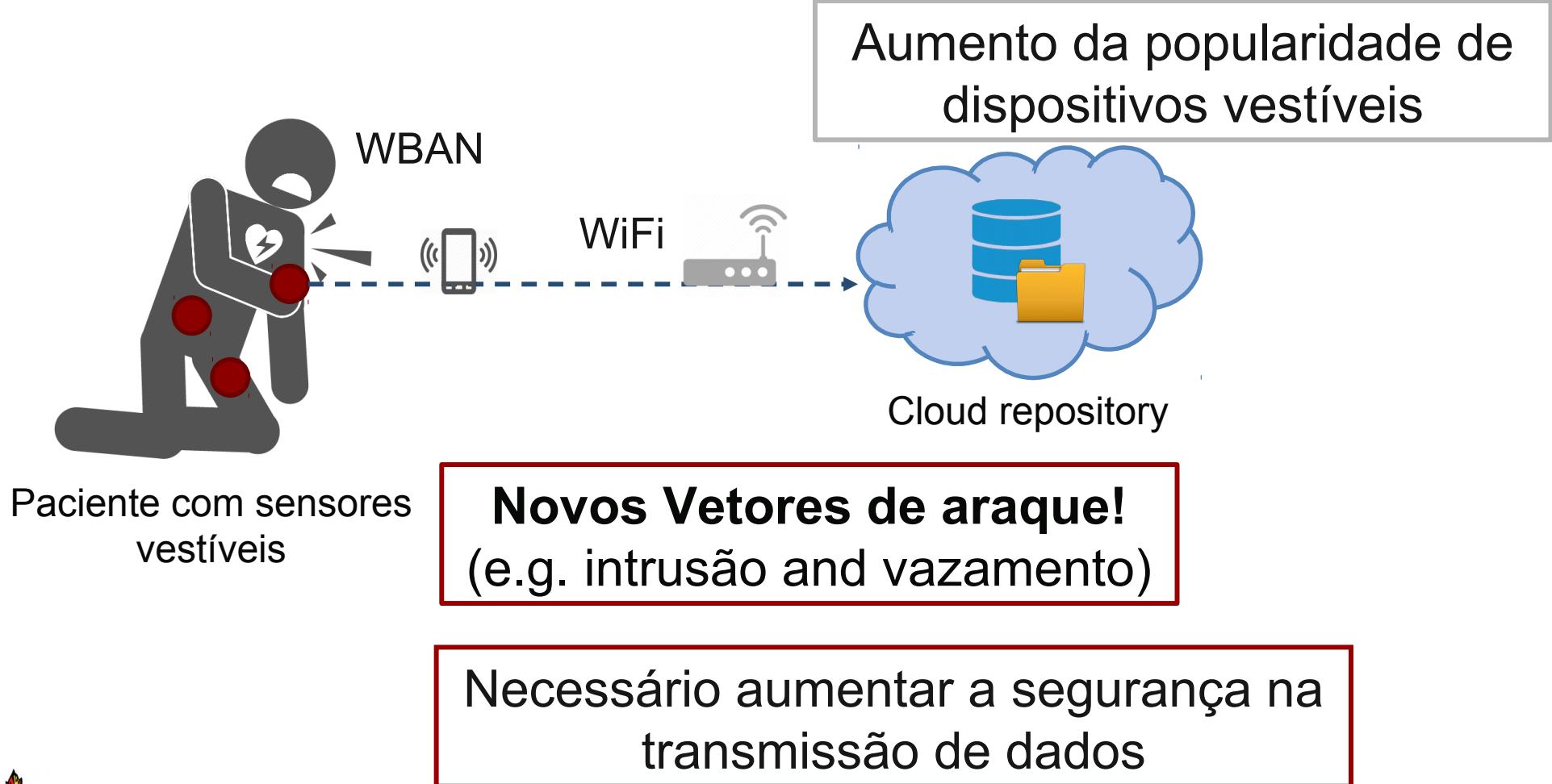
Apresentação:

<https://github.com/TheHealthsenseProject/JAI/blob/master/2019-JAI-ProjetoHealthsense.pdf>



Problema Investigado pelo Projeto

Privacidade dos dados fisiológicos dos pacientes



Problema Investigado



Dispositivos Vestíveis

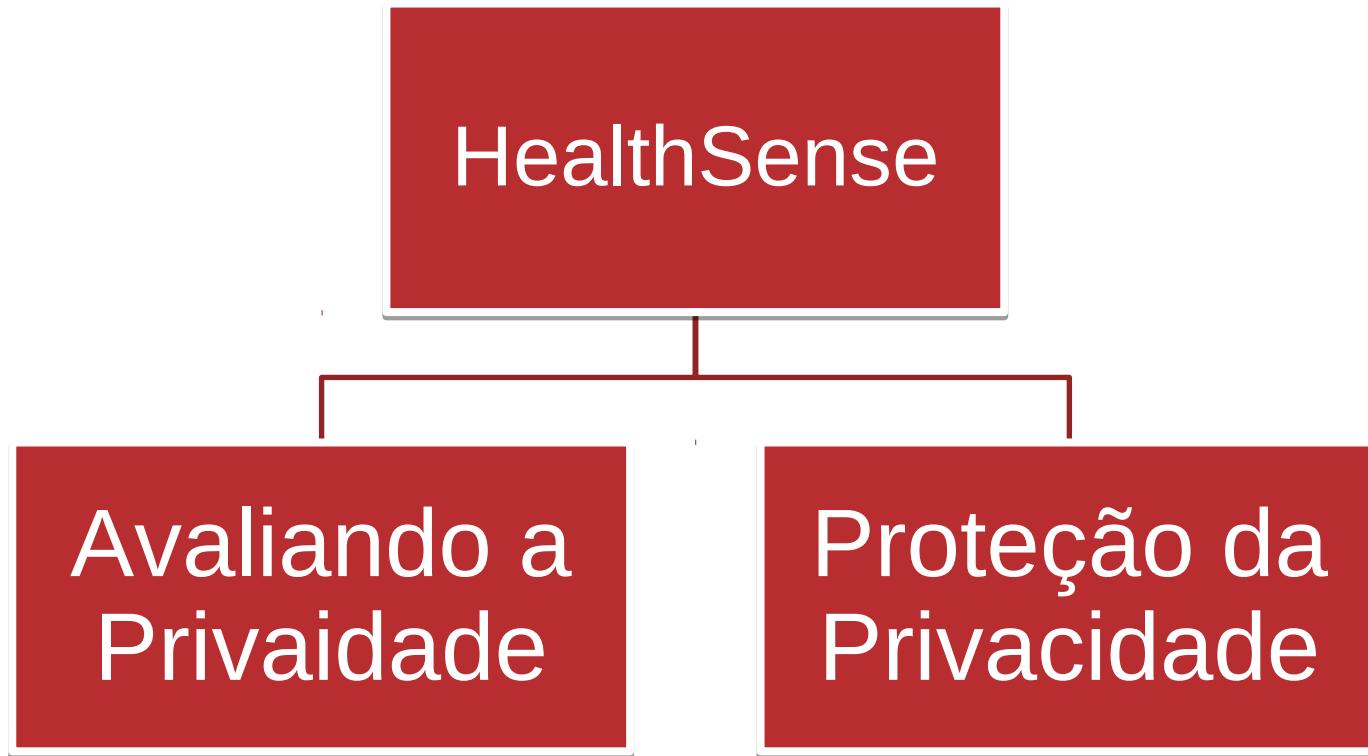
Dados sensíveis

Limitações de Recursos
Computacionais

Vulnerabilidades de segurança em
hardware e software

**Esses três fatores fazem com que os dispositivos
vestíveis tornem-se alvo de ataques**

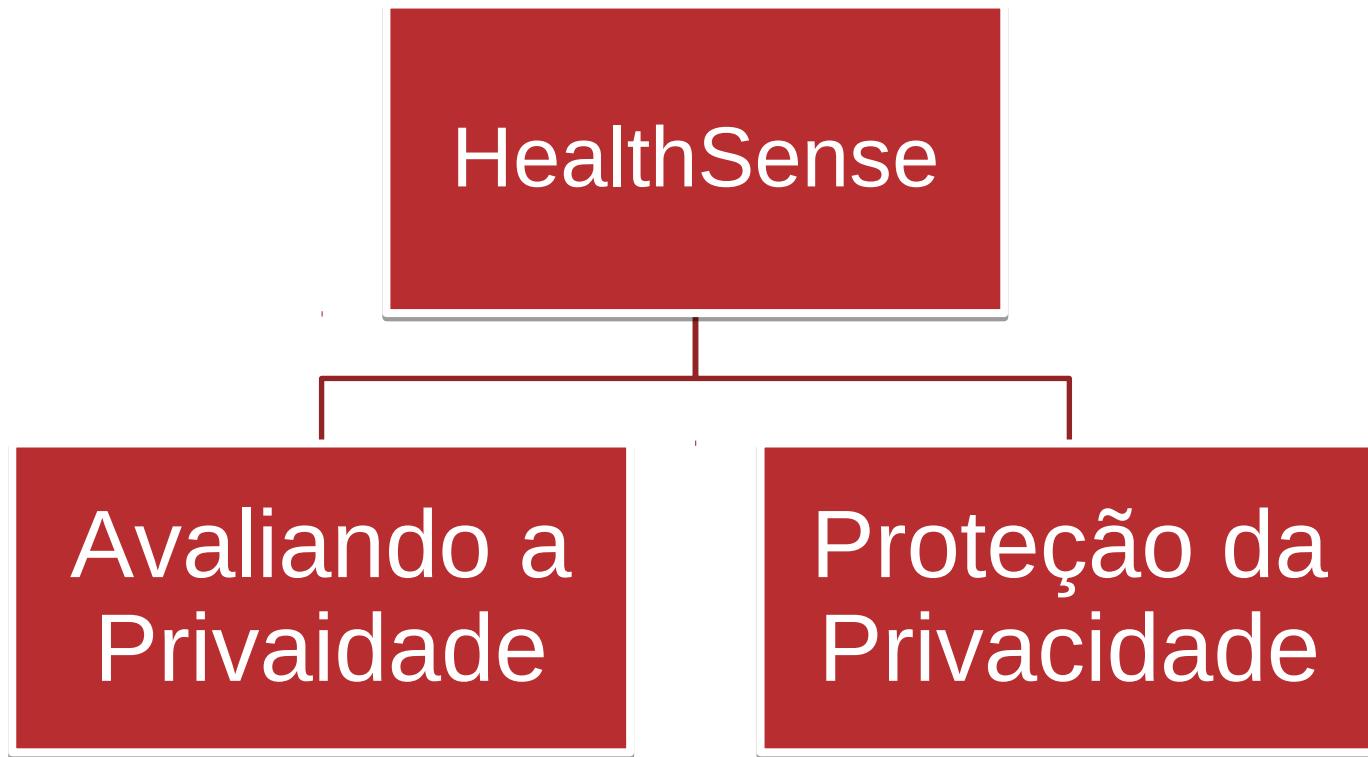
Escopo do Projeto



Cooperação entre
Northeastern University,
UFPA e UFPR.

1 - Analisar e explorar
características de
dispositivos vestíveis,
aplicações e pilha de
protocolos;

Escopo do Projeto



Cooperação entre
Northeastern University,
UFPA e UFPR.

2- Propor técnicas de
resiliência para proteger o
canal de comunicação
entre o corpo e o
dispositivo.



www.healthsenseproject.net
healthsense.project@ufpr.br

Agenda

Conteúdo do Mini-curso

I. Parte da Manhã:

- I.1: Introdução à Biometria;
- I.2: Evolução Histórica;
- I.3: Biometria através da mão;
- I.4: Biometria através dos olhos;
- I.5: Biometria através do rosto;
- I.6: Biometria através de sinais vitais;

Agenda

Conteúdo do Mini-curso

II. Parte da Tarde - Técnicas de Aprendizado de Máquina:

- II.1 Artificial Neural Networks (ANN);
- II.2 k-Nearest Neighbor (k-NN);
- II.3 Support Vector Machine (SVM);
- II.4 Naïve Bayes (NB);
- II.5 K-means clustering
- II.6 Decision Tree (DT)
- II.7 Random Forest (RF)

III- Demonstração Prática

Introdução a IoT e a expansão da biometria



Os dispositivos inteligentes estão presentes em quase todos os lugares.

Por exemplo, temos sensores em edifícios e em veículos e alguns itens embarcados em sistemas eletrônicos com conexões entre si ou com a Internet, permitindo, assim, a coleta e o compartilhamento de dados.

A IoT permite que mais objetos sejam controlados e os dados sejam sensoreados remotamente através de uma infraestrutura de rede preexistente.

Isto permite que haja maior interação de sistemas computacionais com o mundo real, aumentando, assim, a eficiência em serviços nas cidades inteligentes.

Introdução a IoT e a expansão da biometria



Introdução a IoT e a expansão da biometria



Introdução a IoT e a expansão da biometria

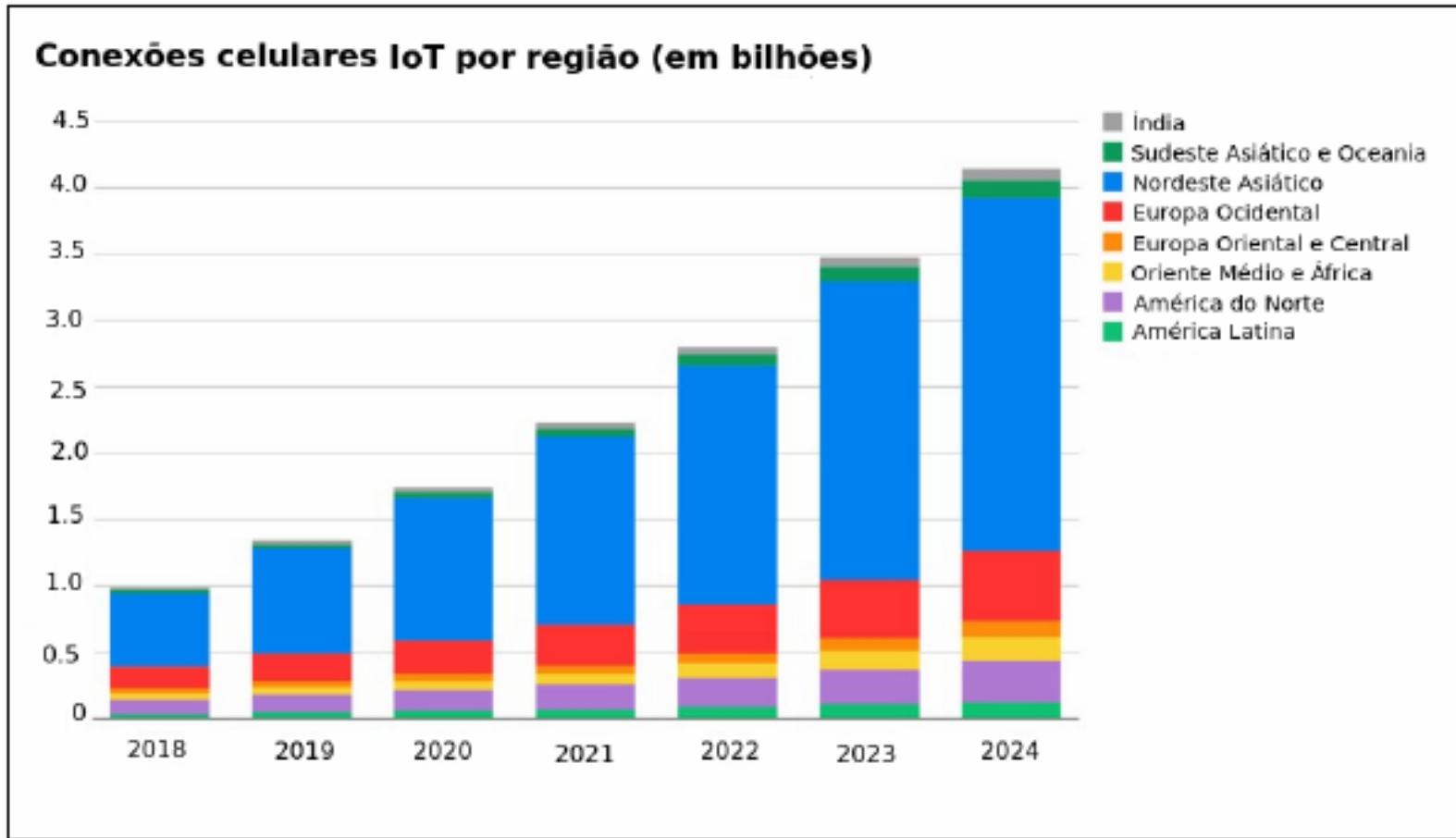


Introdução a IoT e a expansão da biometria



- I. Redução drástica do custo de produção dos dispositivos eletrônicos
- II. Forte expansão, principalmente, nos países em desenvolvimento na África e Ásia.
- III. Gastos devem chegar a 1 trilhão de dólares em 2022.
- IV. Segundo a Ericsson, teremos 3.5 bilhões de dispositivos celulares em 2023.
- V. Grande potencial de novos serviços de realidade aumentada e inteligência artificial.

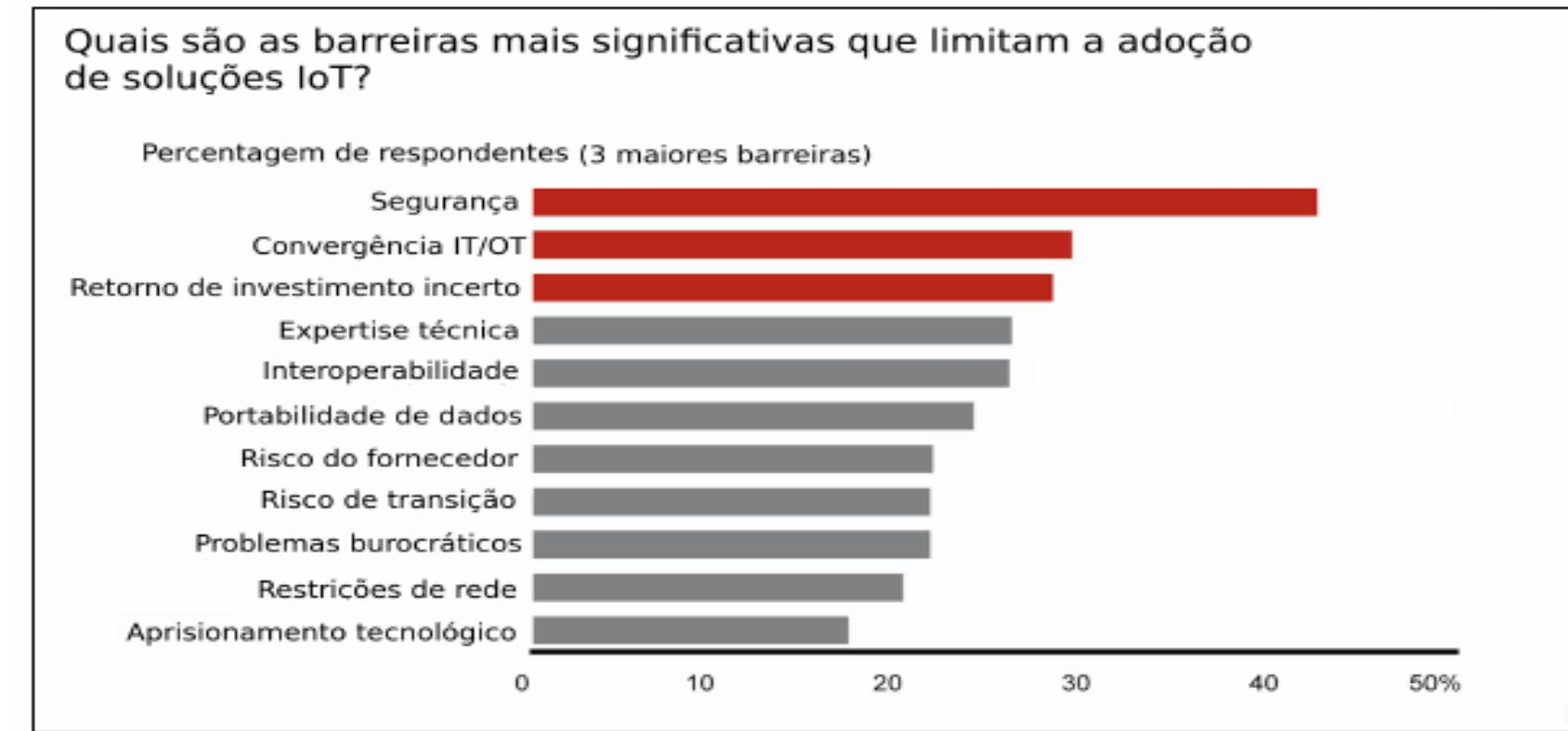
Introdução a IoT e a expansão da biometria



Fonte: Ericsson, Projeção da quantidade de dispositivos IoT conectados.

Introdução a IoT e a expansão da biometria

I. Apesar do cenário bastante promissor, há ainda uma grande preocupação com a segurança.



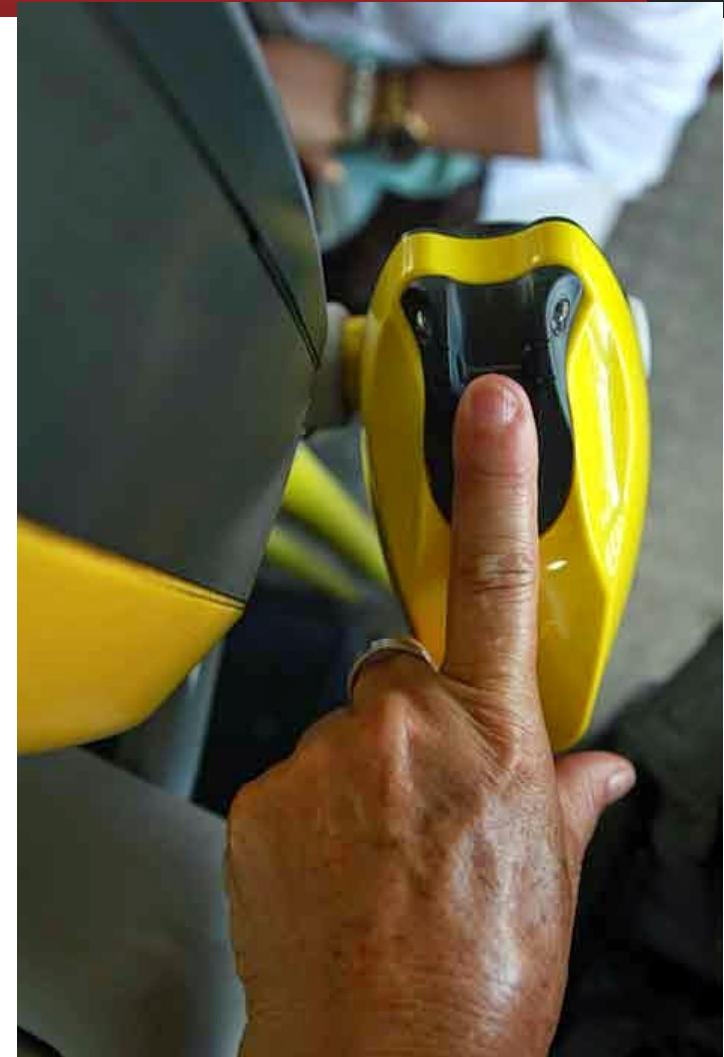
Fonte: Bosche et Al, Barreiras para adoção de IoT.

Introdução a IoT e a expansão da biometria

Hoje em dia esses serviços estão geralmente ligados ao **transporte, à segurança e ao lazer da população**, proporcionando o uso mais eficiente de recursos públicos e melhorando a qualidade de vida da população;

Os sistemas biométricos surgem como uma excelente solução para cenário que exijam o acesso físico direto, **podendo oferecer uma solução escalável com a IoT**, protegendo-a de acesso sem autorização, de troca de identidades ou evitando a checagem manual de credenciais, por exemplo.

Introdução a IoT e a expansão da biometria



Introdução a IoT e a expansão da biometria

Os sistemas biométricos podem reconhecer indivíduos com base em seu comportamento ou características biológicas.

O potencial para a IoT é imenso e está presente em diversos domínios como healthcare, sistemas de transporte, monitoramento ambiental, cidades inteligentes, controle industrial e outros, ainda mais quando a associamos com o aumento de vazão e redução de latência esperados com as tecnologias de 5G.

Introdução a IoT e a expansão da biometria



A biometria pode ser aplicada em investigação forense, controle de imigração, transações financeiras e segurança computacional.

Nesse contexto, os dispositivos móveis possuem um papel importante na vida cotidiana, não somente pela comunicação mas também pelo entretenimento e pelas relações sociais.

É necessário garantir a proteção dos dados bancários, e-mails, fotografias, vídeos e diversos outros dados confidenciais.

Introdução a IoT e a expansão da biometria



Hacker em ação!

Introdução a IoT e a expansão da biometria

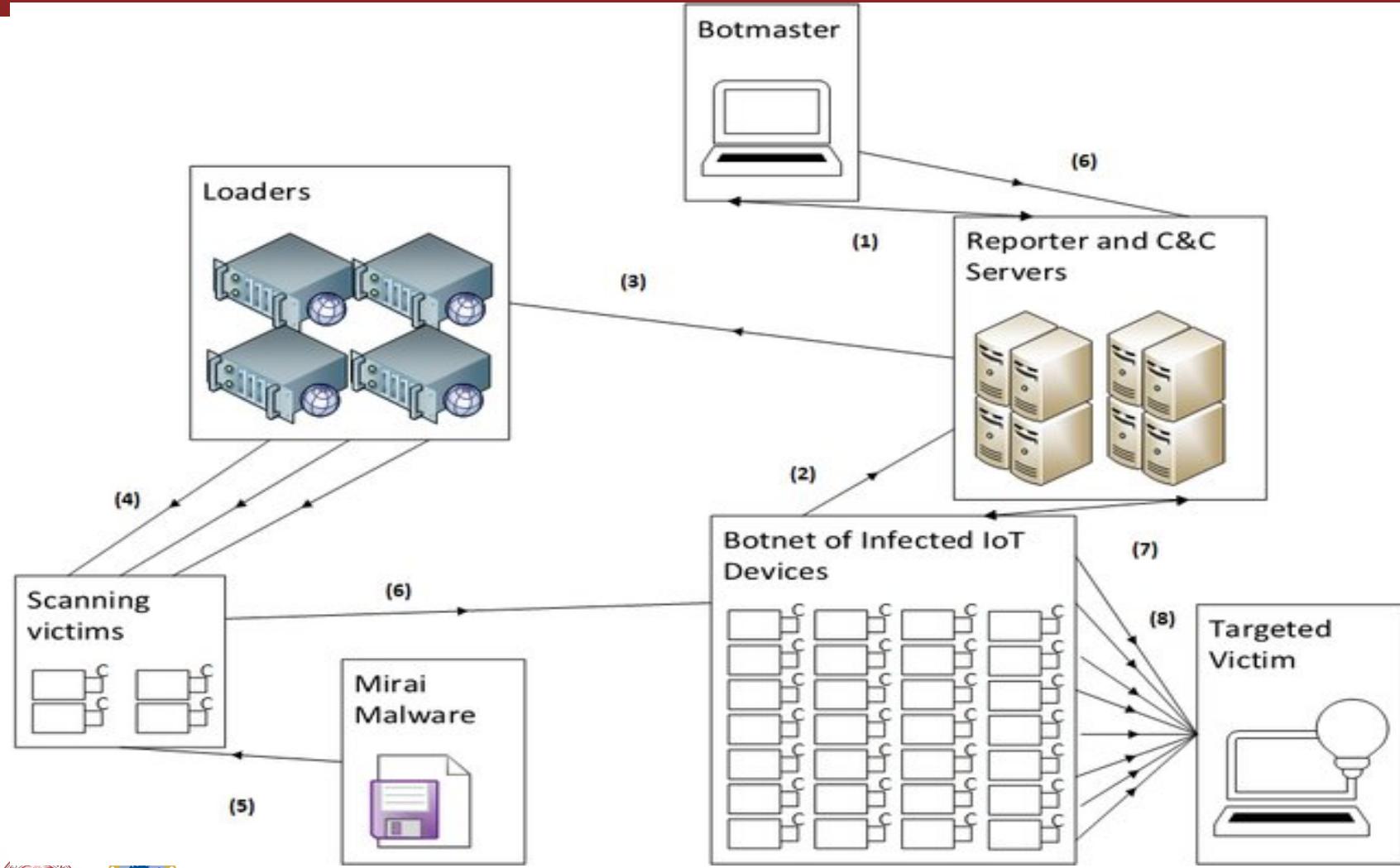


Com o aumento de dispositivos conectados à rede, o escopo por potenciais ataques hackers ou outros crimes cibernéticos também aumentou.

Há o risco de utilização de dispositivos controlados remotamente para ataques de botnet, **onde milhares de dispositivos podem ser utilizados em conjunto para um ataque em rede.**

Introdução a IoT e a expansão da biometria

Ataques Botnet



Introdução a IoT e a expansão da biometria



A cada ano, temos um aumento de ocorrência no roubo de dados e de identidades, quando o usuário utiliza de modo descuidado os dispositivos, como telefones celulares, smartwatches e outros.

O acúmulo de muitas informações em dispositivos pessoais deixa as pessoas vulneráveis a grandes danos caso esses dados sejam violados.

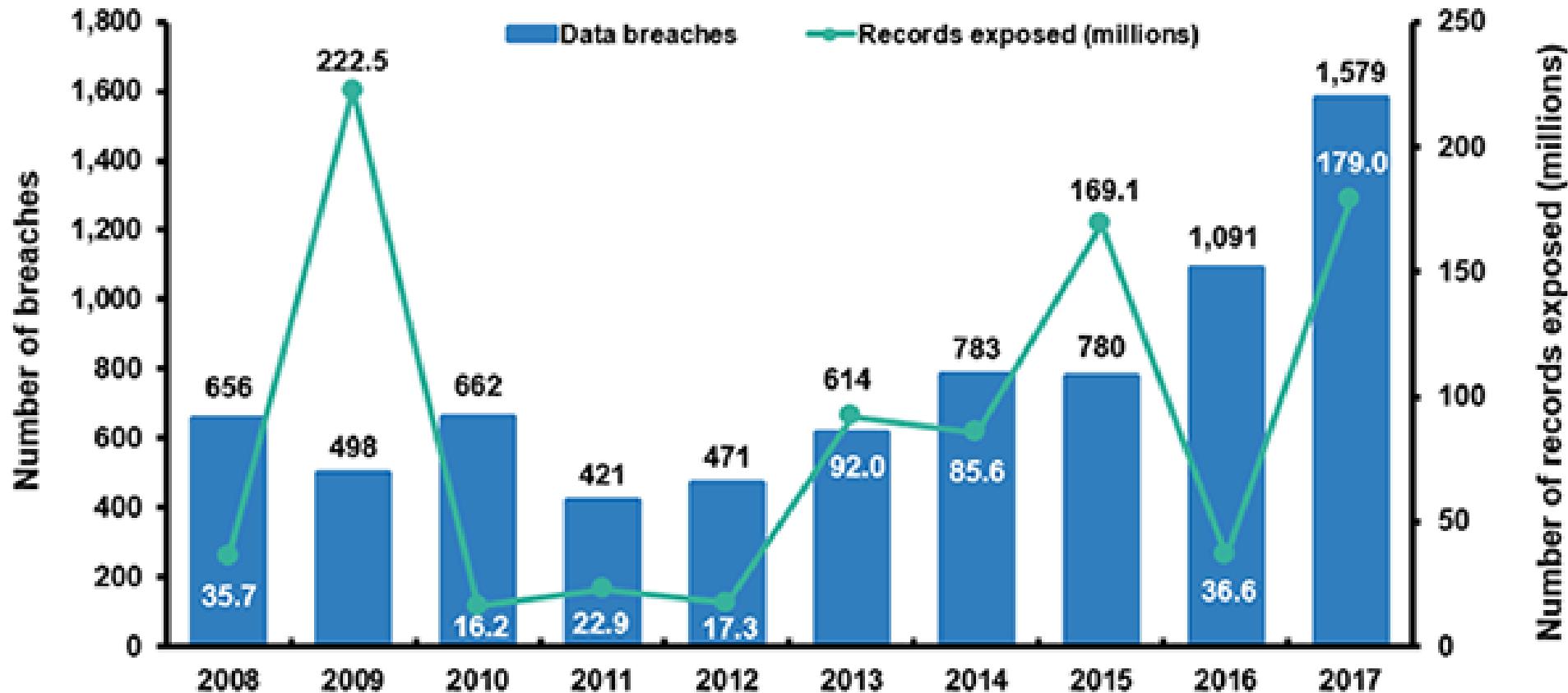
Assim, garantir a segurança desses dados somente com o uso de senhas não tem sido mais suficiente frente aos ataques possíveis.

Introdução a IoT e a expansão da biometria

Aumento nas ocorrências de ataques clássicos de *man-in-the-middle*, no qual o invasor faz de forma transparente o intermédio na comunicação entre dois dispositivos, podem enviar informações falsas ou coletar informações sensíveis.

- Reutilização de senhas em diversos sites online. 82% dos americanos adultos, 61% usam a mesma senha em metade dos sites e 22% usam somente uma senha.
- Utilização de Wi-Fi publica. 48% afirmam utilizar.
- Salvam as senhas e informações de pagamento ao fazer compras online.
- Andar com o cartão de crédito na carteira ou bolsa. (33%)

Introdução a IoT e a expansão da biometria



Source: Identity Theft
Resource Center.

Introdução a IoT e a expansão da biometria

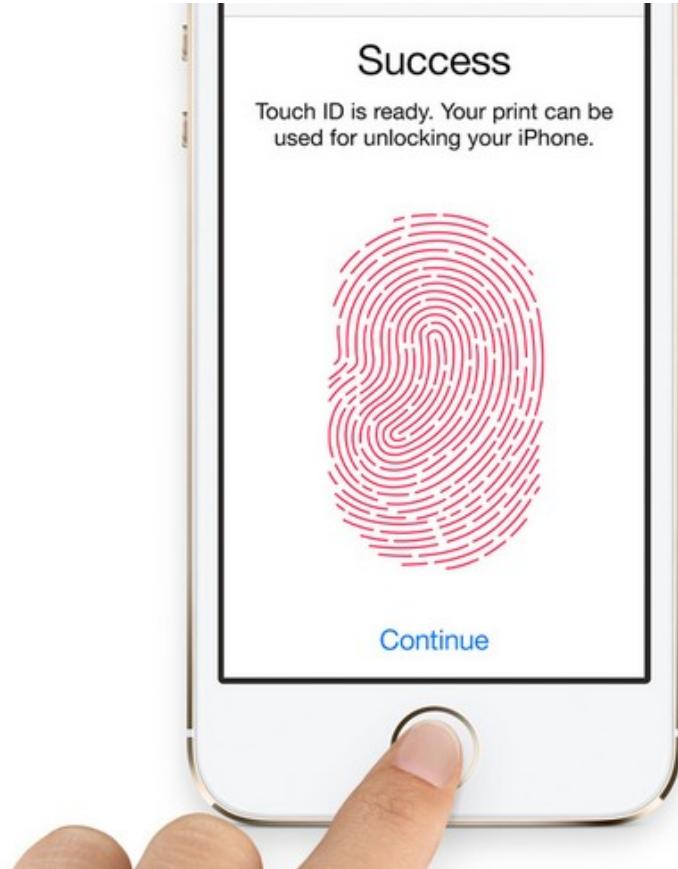


Essa necessidade crescente por proteção dos dados sensíveis fez com que a busca por soluções mais seguras aumentasse, com destaque maior para a biometria.

Os sistemas de segurança biométrica podem substituir métodos tradicionais que utilizam senhas ou PINs gestuais na tela do dispositivo.

Os métodos embarcados em smartphones são reconhecimento por digitais, por face, assinatura, de voz e por íris.

Introdução a IoT e a expansão da biometria



Iphone 5S com Touch ID



Galaxy Note com sensor de digitais

Introdução a IoT e a expansão da biometria



Iphone X com a tecnologia FaceID;

Introdução a IoT e a expansão da biometria

Atualmente, os dispositivos vestíveis podem medir sinais biométricos vitais e não vitais, tais como temperatura corporal, frequência cardíaca, pressão arterial, eletromiograma (EMG), eletrocardiograma (ECG), fotopletismograma (PPG), frequência respiratória, dentre outros sinais.

Introdução a IoT e a expansão da biometria



Apple Watch Series 4



Amazfit Verge da Xiaomi

Introdução a IoT e a expansão da biometria



Nesse contexto, a biometria refere-se a tecnologias usadas para medir características físicas ou comportamentais humanas, tais como as fornecidas pela íris, face, impressões digitais, retina, geometria da mão, voz ou assinaturas para detectar e reconhecer indivíduos.

Por exemplo, os dispositivos vestíveis no antebraço podem obter sinais de ECG/PPG, processá-lo para extrair as características que identifiquem o usuário e utilizar tal informação tanto para identificação quanto para autenticação.

Introdução a IoT e a expansão da biometria

Um sistema biométrico possui basicamente duas etapas:

1- Etapa ligada a **captura dos dados**:

O sistema biométrico coleta o traço biométrico e extrai um conjunto de features (características) relevantes e armazena o modelo desses dados extraídos em um banco de dados.

2- Etapa ligada ao **reconhecimento dos padrões**:

Na etapa de reconhecimento, o sistema captura novamente o traço biométrico de um indivíduo, extrai as características desse sinal e compara esse conjunto de características com os padrões armazenados no banco de dados de modo que possa afirmar de quem é a identidade.

Introdução a IoT e a expansão da biometria



Quando você teve contato pela primeira vez com a biometria?



Sinais Vitais e Segurança

Evolução Histórica



Na Babilônia em 1900 A.C., as digitais eram utilizadas em contratos para dar validade aos mesmos.

Porém, foram os chineses que aproveitaram melhor o potencial da biometria, utilizando as digitais dos dedos para uma variedade de funcionalidades, incluindo desde registro da população e em cenas de crimes até para validar documentos importantes utilizando as digitais de um lado e um selo oficial do outro lado em documentos para casamento, divórcio, registros do exército e outros.

Isso ocorria pois o domínio da escrita era restrita a uma parcela bem pequena da população.

Até hoje, a digital é utilizada em documentos em casos que a pessoa não sabe ou não pode assinar.

Sinais Vitais e Segurança

Evolução Histórica



Gregos



Chineses

Sinais Vitais e Segurança

Evolução Histórica



Apesar de ter surgido há tanto tempo, somente em 1901, foi fundado o primeiro escritório de investigação de digitais, a polícia da Inglaterra, conhecida como Scotland Yard que foi o primeiro órgão de investigação oficial a ter um departamento dedicado para digitais.

Depois disso, a utilização de digitais começou a se espalhar pelas polícias do mundo todo.

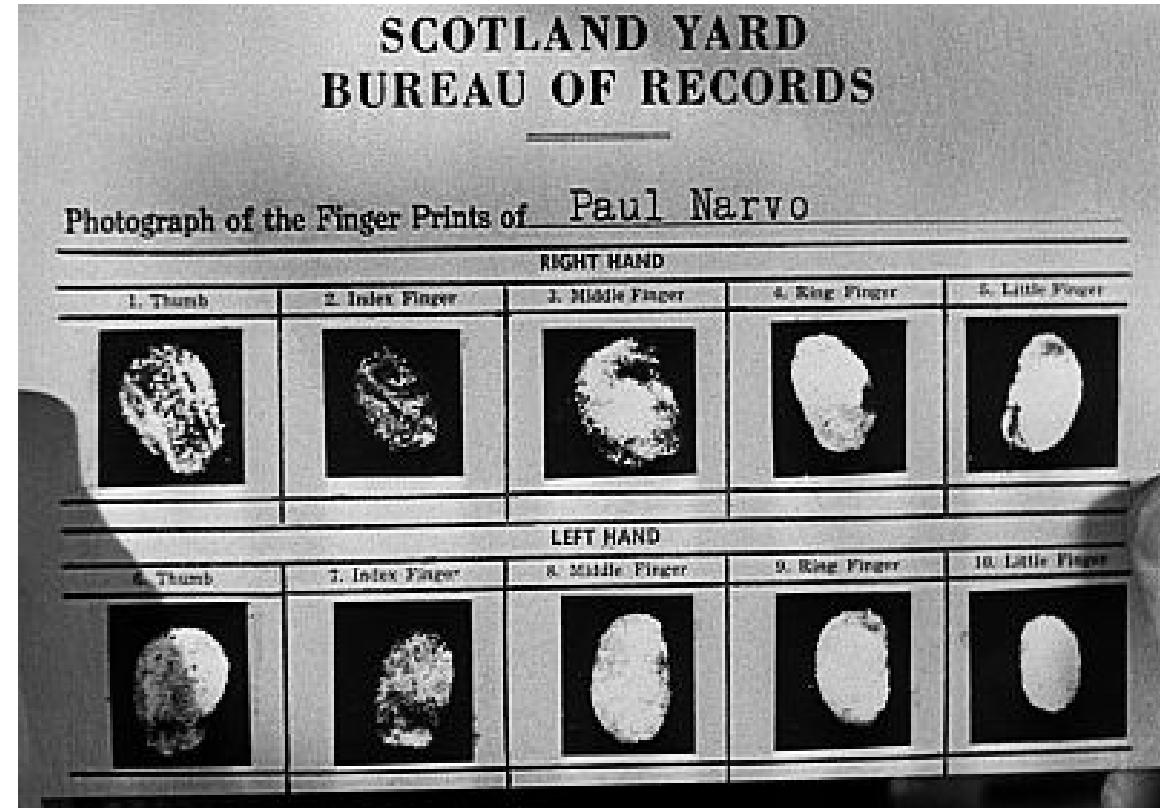
Nessa época, o reconhecimento através de digitais não era automático, pois era um trabalho complicado, manual e demorado. Além dessas dificuldades, a quantidade de registros ainda era pequena.

Sinais Vitais e Segurança

Evolução Histórica



O processo de identificação de digitais exigia uma análise cuidadosa e manual das imagens



Registros de digitais de um indivíduo no inicio do século XX

Sinais Vitais e Segurança

Evolução Histórica



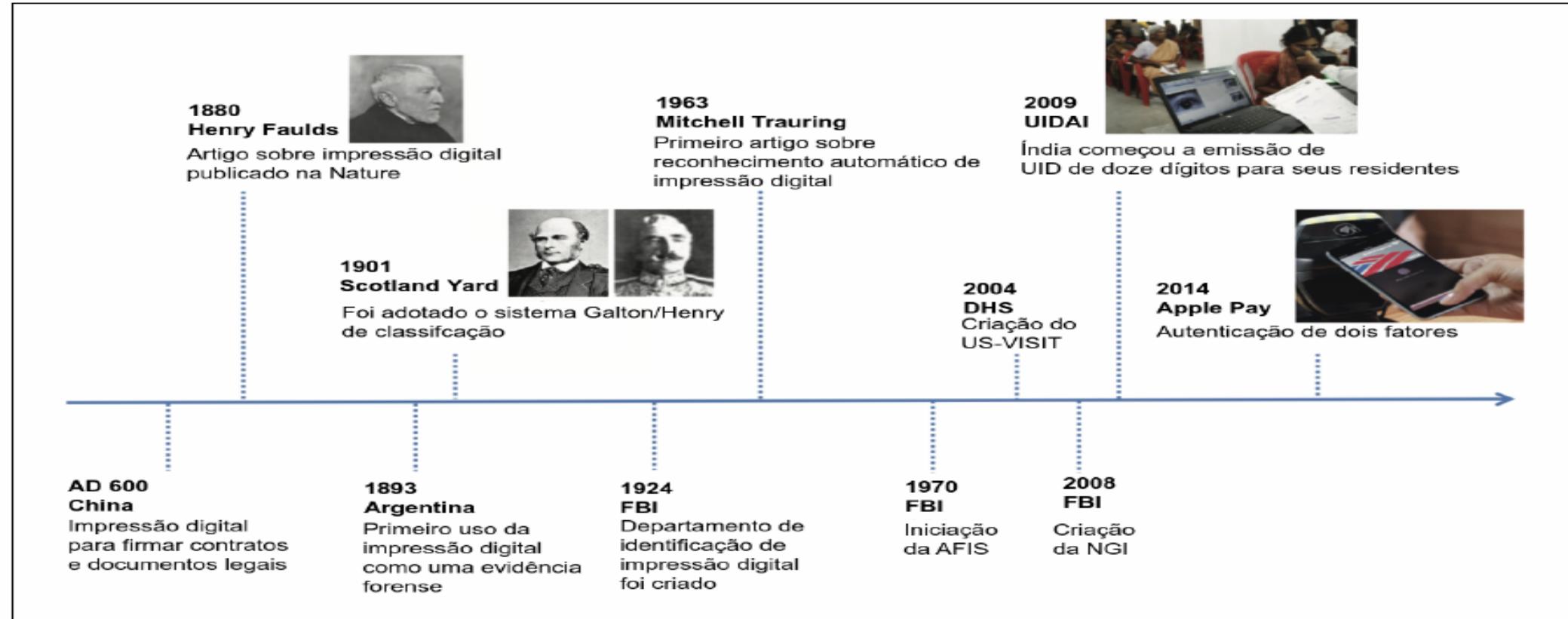
O primeiro sistema automático de digitais do dedo veio em 1963 com Mitchell Trauring.

O desenvolvimento de sistemas automatizados baseados em outros traços como voz, rosto e assinatura também começaram por volta dos anos 60.

Somente depois surgiram os sistemas baseados em geometria da mão e íris, assim, a biometria vem evoluindo nos últimos 60 anos.

Sinais Vitais e Segurança

Evolução Histórica



Linha do tempo da evolução de aplicações com biometria

Sinais Vitais e Segurança

Evolução Histórica



“A quick overview of biometric history shows that much of what we consider to be “new” in biometrics was really considered decades ago.”

Entendemos que a maioria dos aparelhos móveis já possuem sistemas de biometria com no mínimo a utilização de digitais.

E mais recentemente a utilização de reconhecimento facial. Contudo, o primeiro dispositivo com a utilização dessa tecnologia foi lançado em 2011. o Galaxy Nexus. Após diversas melhorias pelos fabricantes, tivemos o Galaxy S8 e finalmente o iPhone X, este último oferecendo uma qualidade melhor para essa nova funcionalidade.

Sinais Vitais e Segurança

Evolução Histórica



Além do Iphone X, outros dispositivos já utilizam o reconhecimento facial, Galaxy S8, Galaxy Note 8, LG V30, OnePlus 5T, HTC U11, Huawei P10, Moto G5, Xiaomi Mi 6 e Xiaomi Mi MIX 2.

O reconhecimento de voz ainda não apareceu como uma opção disponível em smartphones, apesar das grandes marcas como Apple, Google, Amazon e Microsoft estarem investindo bastante em soluções com assistentes inteligentes de processamento de voz.

Atualmente, os dispositivos vestíveis podem medir sinais biométricos vitais e não vitais, tais como temperatura corporal, frequência cardíaca, pressão arterial, ECG, EMG, frequência respiratória, dentre outras informações.

Sinais Vitais e Segurança

Evolução Histórica



A biometria fornece não apenas uma alternativa para IDs ou números PIN (esquemas baseados em conhecimento) para autenticar alguém em um sistema, mas também um método de autenticação contínua.

Além disso, ela possui vantagens relacionadas à clonagem, perda de dispositivos e adivinhação de senhas.

Embora a biometria possa reduzir limitações de segurança associadas a senhas, os sistemas biométricos também são vulneráveis a ataques de falsificação, ataques de vinculação equivocada de usuários (ou seja, quando um impostor tenta se passar por outro usuário), além de possivelmente aumentar os custos com hardware e software comparado com o uso de senha ou token.

Sinais Vitais e Segurança

Evolução Histórica



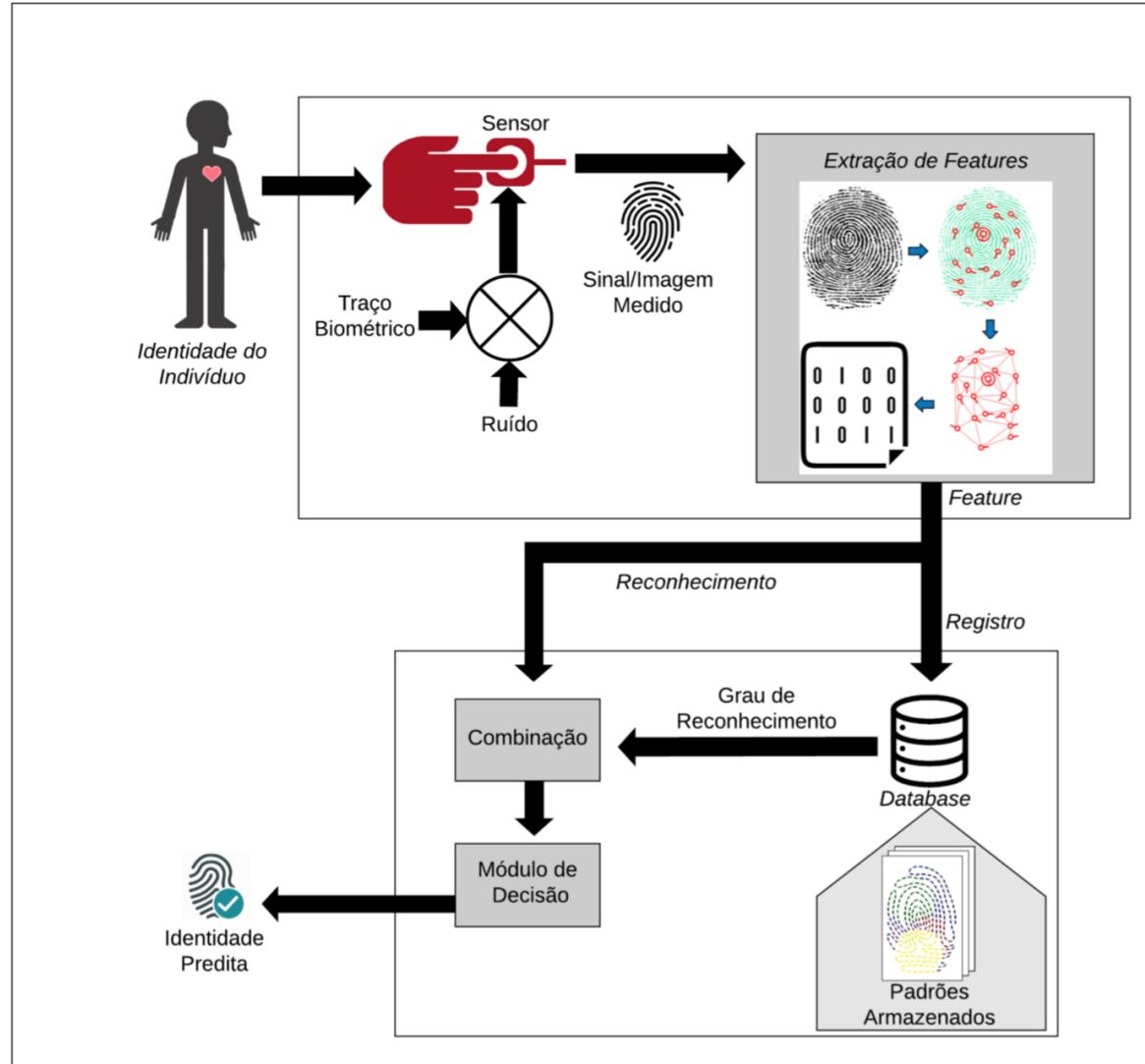
Mesmo a identificação facial, a qual passou a ser utilizada recentemente em dispositivos móveis, possui riscos associados a falta de confidencialidade, com o uso de imagens publicadas na Internet em sistemas que usam as imagens da face, além das limitações na distinção de gêmeos legítimos.

Por esses motivos, julga-se importante a avaliação dos diversos tipos de biometrias existentes, onde a combinação entre elas e o uso adicional de senhas tradicionais seja a solução adotada pela maioria dos sistemas no futuro.

De um modo geral, a maioria dos sistemas biométricos possuem o mesmo esquema.

Sinais Vitais e Segurança

Evolução Histórica



Sistema
Biométrico
Típico.

Extração de *features* dos Sinais

Os sinais biométricos, tais como EMG, ECG, PPG e outros, desempenham um papel crucial na autenticação, fornecendo características individuais como uma forma de controle de acesso e identificação de seres humanos por suas características, tais como sinais vitais.

Portanto, é necessário considerar técnicas exclusivas de extração de características para sinais biométricos adquiridos de dispositivos vestíveis.

Por exemplo, os dispositivos vestíveis obtêm sinais de ECG/PPG, processam os para extrair as características que identifiquem o usuário e utilizam tal informação para identificação e autenticação.

Extração de *features* dos Sinais

Teoricamente, qualquer sinal anatômico, comportamental ou fisiológico de um indivíduo pode ser usado como um marcador biométrico.

Contudo, a escolha de qual marcador utilizar depende dos requisitos da aplicação e do grau de algumas propriedades a serem satisfeitas:

- (i) exclusividade ou distinção,
- (ii) durabilidade,
- (iii) universalidade,
- (iv) coleta,
- (v) desempenho,
- (vi) aceitação do usuário,
- (vii) invulnerabilidade e
- (viii) integração.

Extração de *features* dos Sinais



Características mais utilizadas em sistemas de identificação biométrica

Extração de *features* dos Sinais

Reconhecimento de Mão



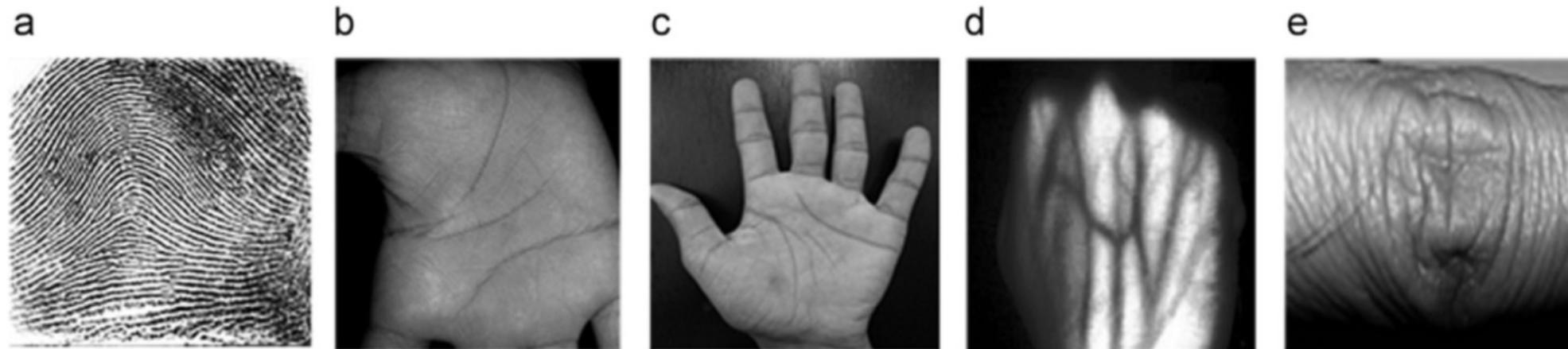
A impressão digital é uma informação humana usada como biometria.

Porém, há vários atributos, além das impressões digitais, que foram identificados e testados, tais como *palmprint* (impressão da palma), geometria da mão, impressão das juntas dos dedos, região abaixo das unhas e o padrão das veias da mão.

Os atributos presentes nas mãos são uma extensão da tecnologia de impressão digital.

Extração de *features* dos Sinais

Reconhecimento de Mãos



Modalidades das regiões da mão: (a) impressão digital (b) impressão da palma, (c) geometria da mão, (d) padrão das veias da mão, (e) articulação dos dedos

Extração de *features* dos Sinais

Reconhecimento de Mãos



O sistema de reconhecimento de impressão digital caracteriza as **texturas de cumes e sulcos (linhas)** presentes nas pontas dos dedos.

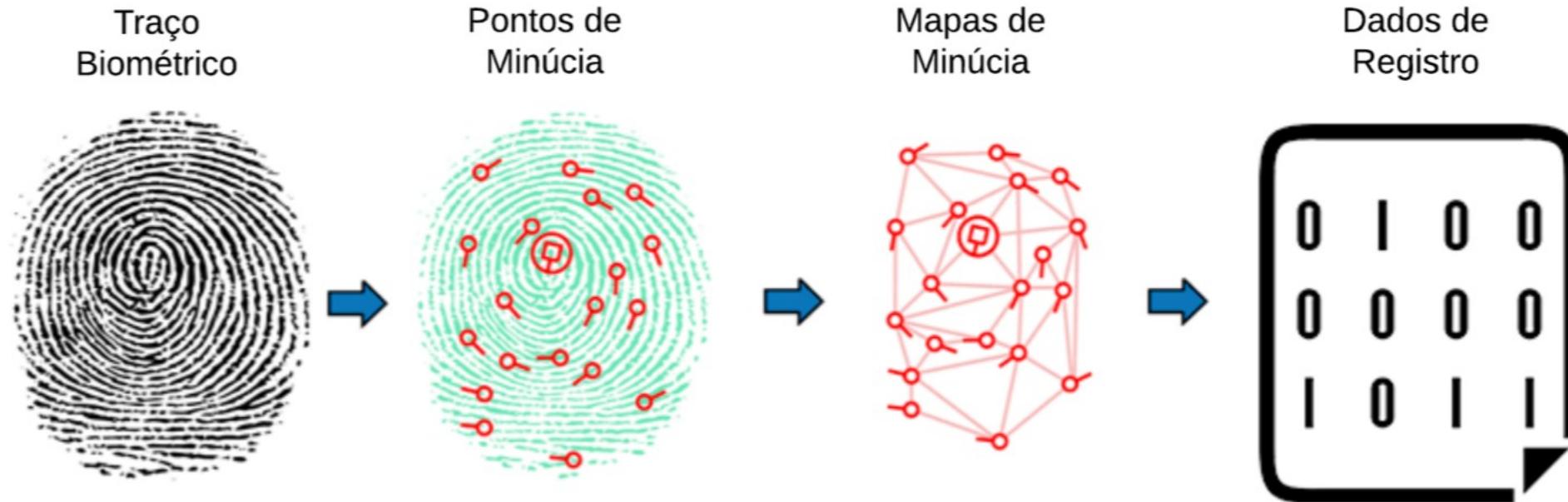
As linhas são quase paralelas, com exceção de alterações do padrão, chamadas de **minúcias**.

Existem categorias para estes pontos característicos (minúcias), como **arco, presilha interna, presilha externa e verticilo**. Individualmente, os pontos de minúcia executam a tarefa de reconhecimento.

Como todos os outros sistemas biométricos, um módulo de aquisição captura as pontas dos dedos, de preferência, a partir de imagens de **alta resolução** e o sistema extrai os sulcos, alguns pontos singulares e pontos de minúcia.

Extração de *features* dos Sinais

Reconhecimento de Mãos



A figura acima mostra o processo de extração de features para a impressão digital, onde são detalhados a coleta, o mapeamento e o registro das informações, como dados a serem inseridos no database para reconhecimento. **A qualidade da imagem é fator decisivo nesse tipo de sistema.**

Extração de *features* dos Sinais

Reconhecimento de Mãos



Estes métodos de aquisição usam dados de sensores ópticos, térmicos, de silício ou imagens de ultrassom.

Devido à redução do custo dos sensores de imagem, assim como o pequeno tamanho para guardar o template, torna os atributos da região da mão como uma boa opção para muitos tipos de aplicações, em comparação com assinaturas biométrica mais complexas.

Extração de *features* dos Sinais

Reconhecimento de Mão



No entanto, a biometria manual tem desafios.

- I. As imagens distorcidas.
- II. O contato físico com o dispositivo de imagem.
- III. A necessidade de cooperação do usuário.
- IV. doenças da mão (artrite), contaminantes naturais para a imagem (como células mortas, cicatrizes, cortes, pele úmida e/ou seca),
- V. Sensores de imagens com superfícies sujas ou oleosas que comprometem a eficácia do sistema.

Extração de *features* dos Sinais

Reconhecimento de Mão



Podem ser utilizadas técnicas de Deep Learning para realizar o reconhecimento de mãos, principalmente envolvendo impressões digitais e impressão de veias da palma.

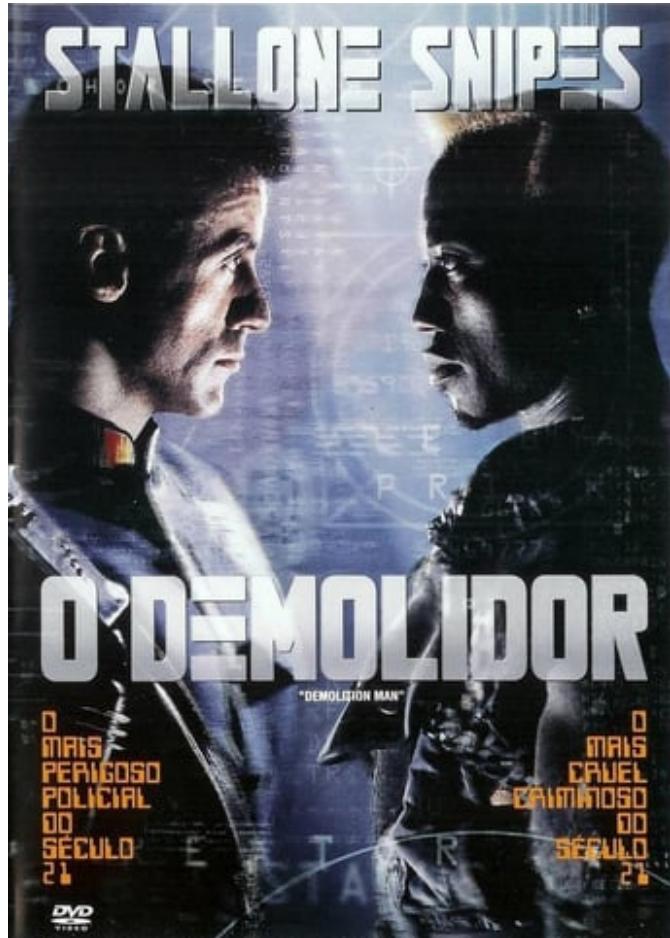
Alguns autores propõe a utilização de reconhecimento de mãos em conjunto com o reconhecimento facial. **Veremos que a utilização de vários sinais biométricos combinados pode ser a alternativa para o aumento da eficiência desses sistemas.**

Alguns sistemas precisam usar técnicas de machine learning como o CNN, PCA e k-Nearest Neighbor (k-NN) para melhorar o resultado com a utilização de imagens de baixa qualidade.

Devido ao baixo custo e a praticidade, tem sido um dos sistemas biométricos de maior expansão na última década.

Extração de *features* dos Sinais

Reconhecimento de Mãos



Filme O Demolidor - 1993

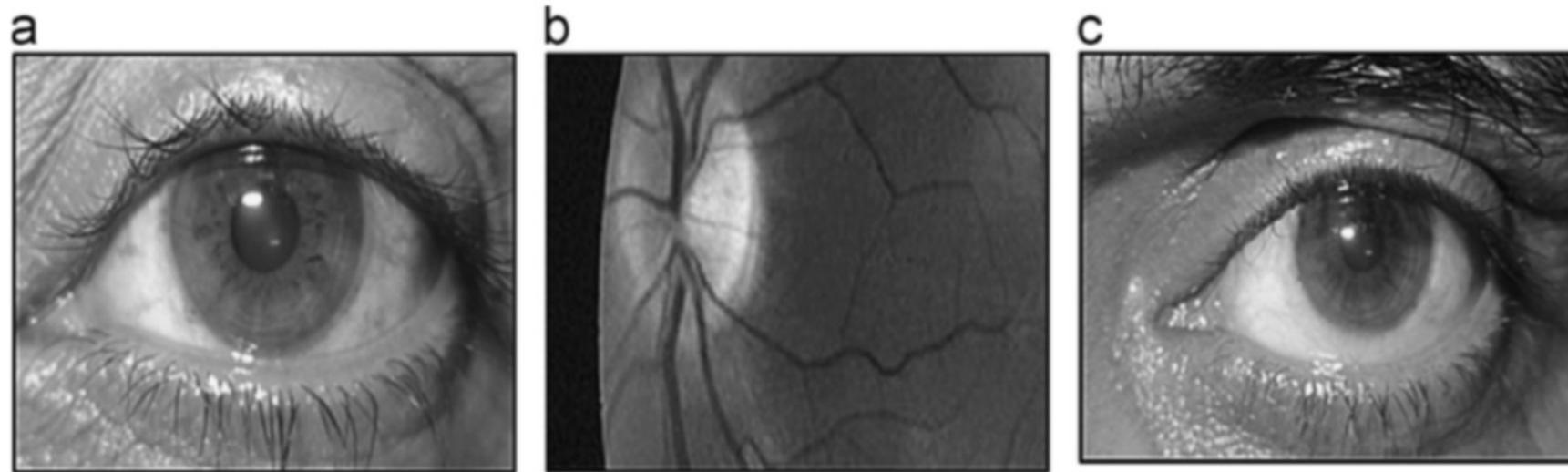
Extração de *features* dos Sinais

Reconhecimento Ocular



Para o reconhecimento ocular, três modalidades foram as mais utilizadas ao longo da história e estão ilustradas na Modalidades de região ocular como **retina**, **íris** e **padrão das veia da esclera**, ganharam considerável atenção de pesquisadores devido a região ocular possuir sinais mais precisos, altamente confiáveis, bem protegidos, estável e quase impossíveis de forjar assinaturas biométrica;

Extração de *features* dos Sinais Reconhecimento Ocular



Modalidades da
região ocular:
(a) íris, (b) retina,
(c) esclera

Extração de *features* dos Sinais

Reconhecimento Ocular



Um sistema de identificação da **retina** leva conta a estrutura única e invariante de veias sanguíneas presentes na retina humana para estabelecer a identidade. O processo de escaneamento captura algumas features relacionadas às marcas (como posição e bifurcações dos vasos sanguíneos) ou medidas da área de referência (fóvea ou o disco ótico).

Já um sistema de identificação baseado em **escleras** considera o padrão vascular dos vasos sanguíneos presentes na região da esclera do olho humano.

O sistema baseado em **íris** cria um modelo usando os padrões únicos presentes na íris como as criptas, linhas radiais, área da pupila, área ciliar e anel limite.

Extração de *features* dos Sinais

Reconhecimento Ocular



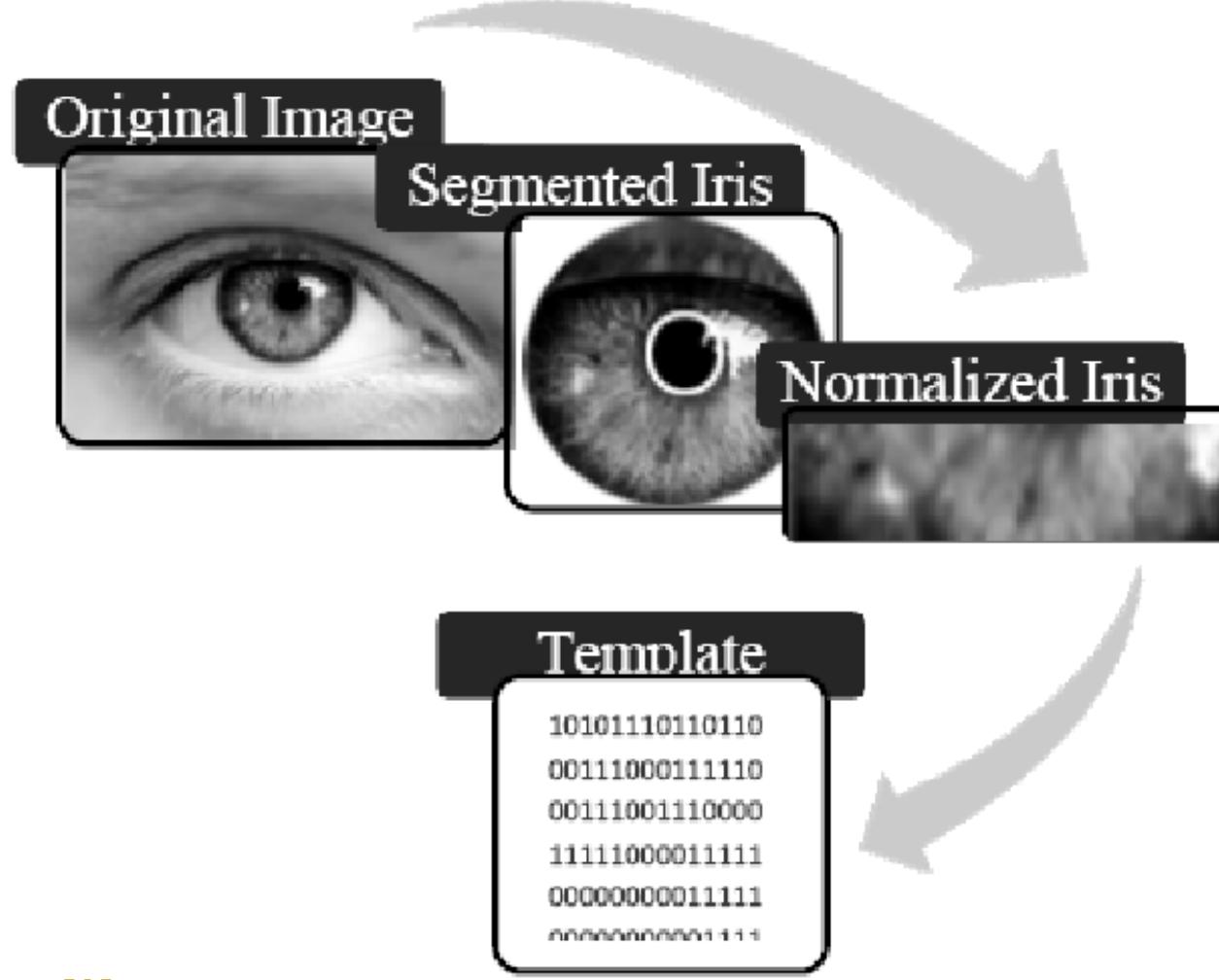
Foram propostos modelos onde tanto as cores quanto as formas eram utilizadas para distinguir as pessoas.

As primeiras propostas iniciaram em 1936 com Frank Burch, o primeiro sistema completo foi implementado somente no início dos anos 1990 com uma câmera para capturar a imagem da íris, algoritmos para processar as imagens e retirar a região da íris e código de representação da íris capaz de convertê-la em um código binário compacto.

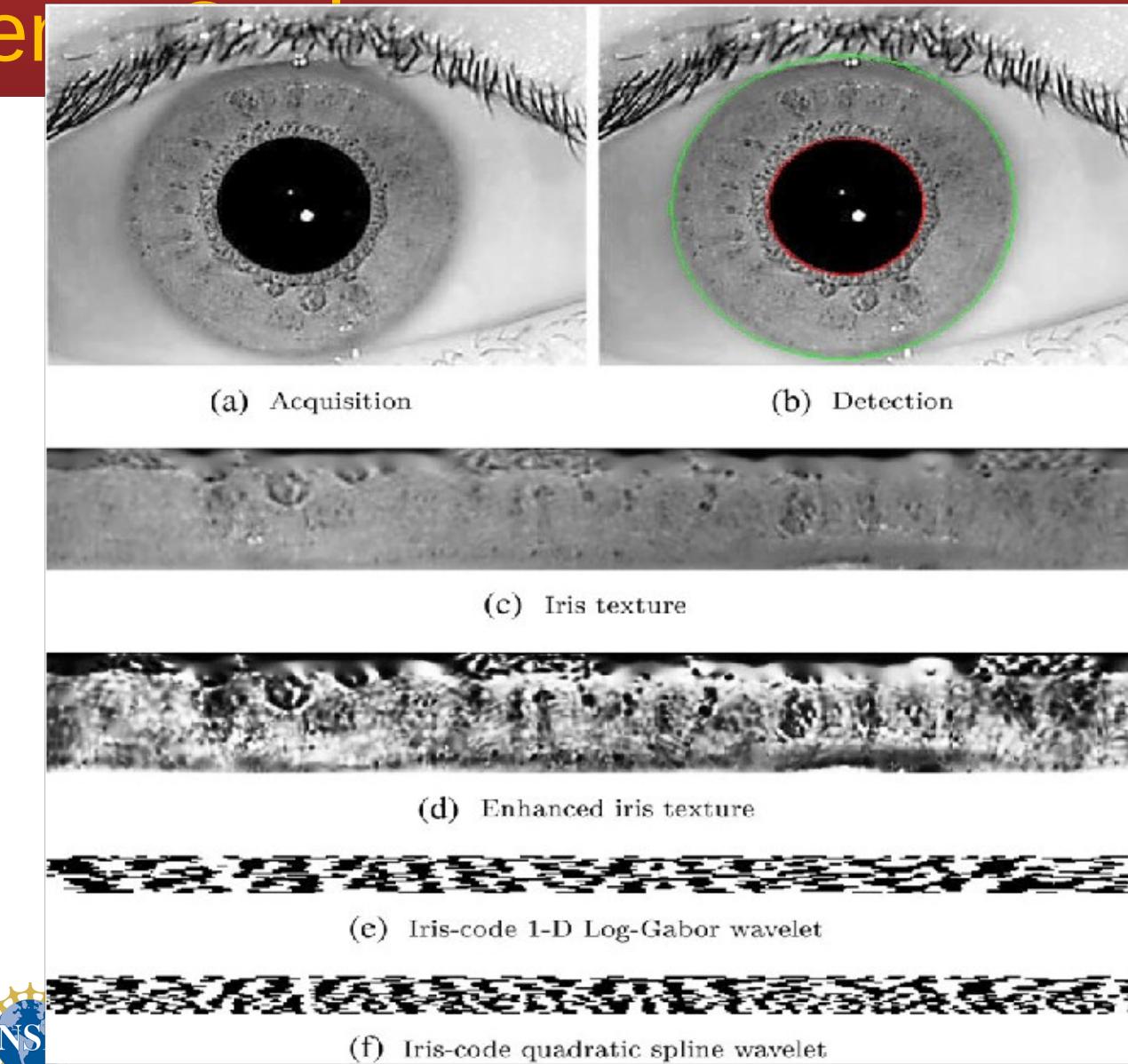
Ou seja, cada íris de um indivíduo era convertido para uma espécie de hash e utilizado a distribuição de distâncias hamming para classificá-las.

Extração de *features* dos Sinais

Reconhecimento Ocular



Extração de *features* dos Sinais Reconhecimento



Extração de *features* dos Sinais

Reconhecimento Ocular



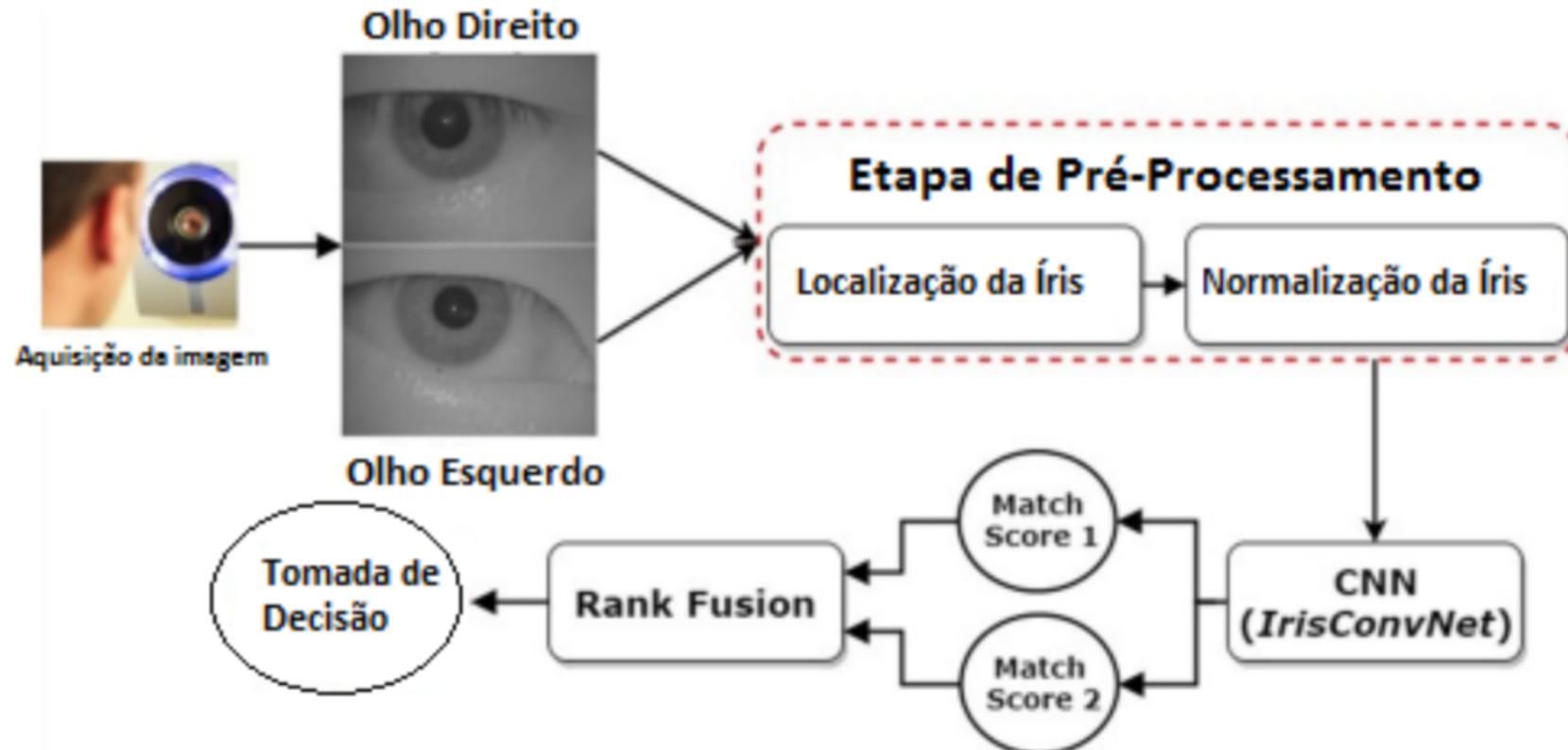
Recentemente, as técnicas de Deep Learning, especialmente utilizando CNNs, têm mostrado grande potencial para a classificação de imagens.

Desde 2016, alguns trabalhos já foram publicados com a utilização de Deep Learning para o reconhecimento de íris, sendo capaz de reconhecer íris heterogêneas, obtidas com diferentes tipos de sensores, e estabelece a similaridade entre um par de imagens de íris usando CNNs.

Outras abordagens propostas utilizaram um sistema multi-biométrico, usando as íris esquerda e direita de uma pessoa.

Extração de *features* dos Sinais

Reconhecimento Ocular



Esquema de Identificação utilizando CNN e Íris dos olhos Direito e Esquerdo.

Extração de *features* dos Sinais

Reconhecimento Ocular



De um modo geral, os trabalhos de reconhecimento de íris utilizam uma imagem de entrada que passa por um processo de normalização, localização da íris e segmentação antes de ser utilizada em uma técnica de inteligência computacional.

Todas essas abordagens têm limitações práticas.

Por exemplo, o padrão vascular da retina só pode ser visto expondo o olho humano a uma luz infravermelha, enquanto a textura da íris pode ser adquirida através da iluminação do olho humano com uma luz perto da infravermelha ou luz de comprimento de onda invisível.

Extração de *features* dos Sinais

Reconhecimento Ocular



A íris se torna uma abordagem promissora em assinaturas biométrica oculares devido a sua aquisição de imagens menos invasiva.

As modalidades da região ocular exigem um alto grau de cooperação por parte dos sensores médicos/químicos, o alto custo de sensores de imagem, e reflexões de fontes de luz ambiente.

Esses requisitos eventualmente limitam a sua utilização em ambientes industriais ou a sua utilização em dispositivos com **limitações de recursos**.

Extração de *features* dos Sinais

Reconhecimento Facial



O reconhecimento facial humano é uma técnica bem conhecida, já que o rosto humano é o mais natural traço biométrico usado para reconhecer indivíduos por séculos.

Um sistema de reconhecimento facial leva em conta algumas características/features, como distância entre os olhos, boca, lado do nariz, imagem da face inteira, pontos de canto, contornos, pelos faciais, redondeza de face, etc

Extração de *features* dos Sinais

Reconhecimento Facial



Extração de *features* dos Sinais

Reconhecimento Facial



Extração de *features* dos Sinais

Reconhecimento Facial

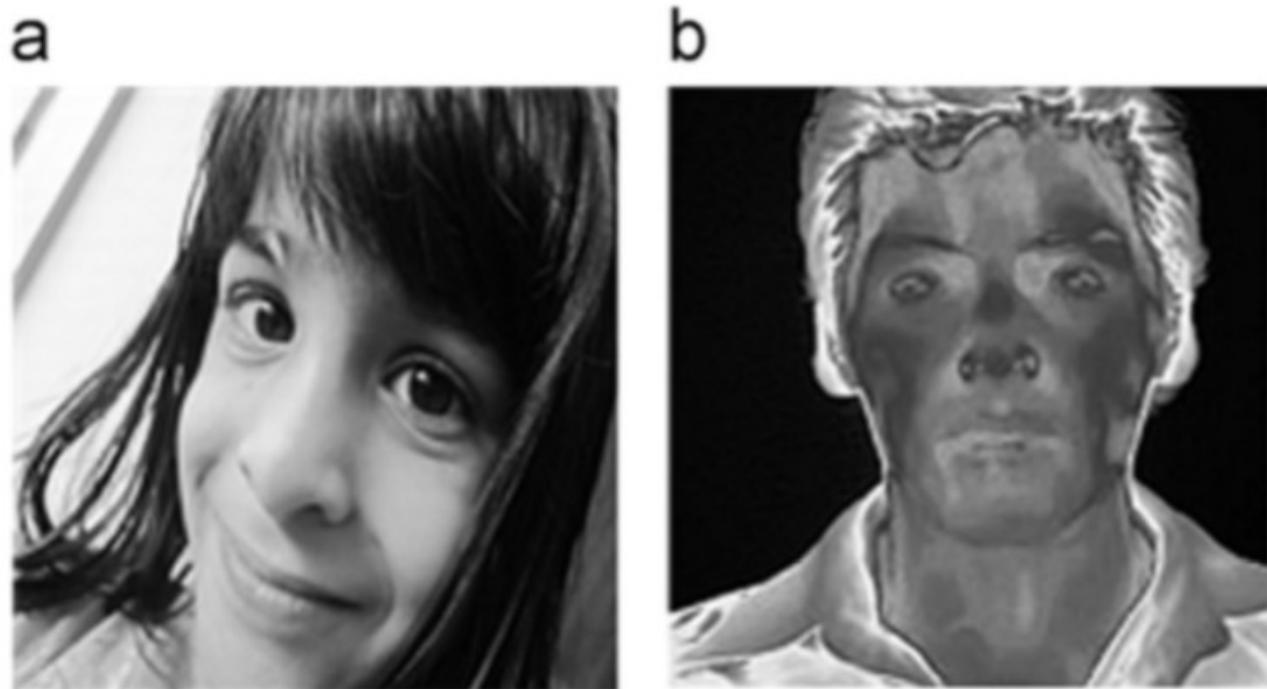


Mesmo com muitos recursos disponíveis, esses sistemas não garantem uma identificação confiável na presença de alguns artefatos, como o uso de cirurgias plásticas, sendo necessários alguns novos algoritmos para mapear essas possíveis alterações.

Por exemplo, o sistema deve estar preparado para reduzir o impacto das mudanças pessoais ao longo do tempo sobre a precisão de tais sistemas.

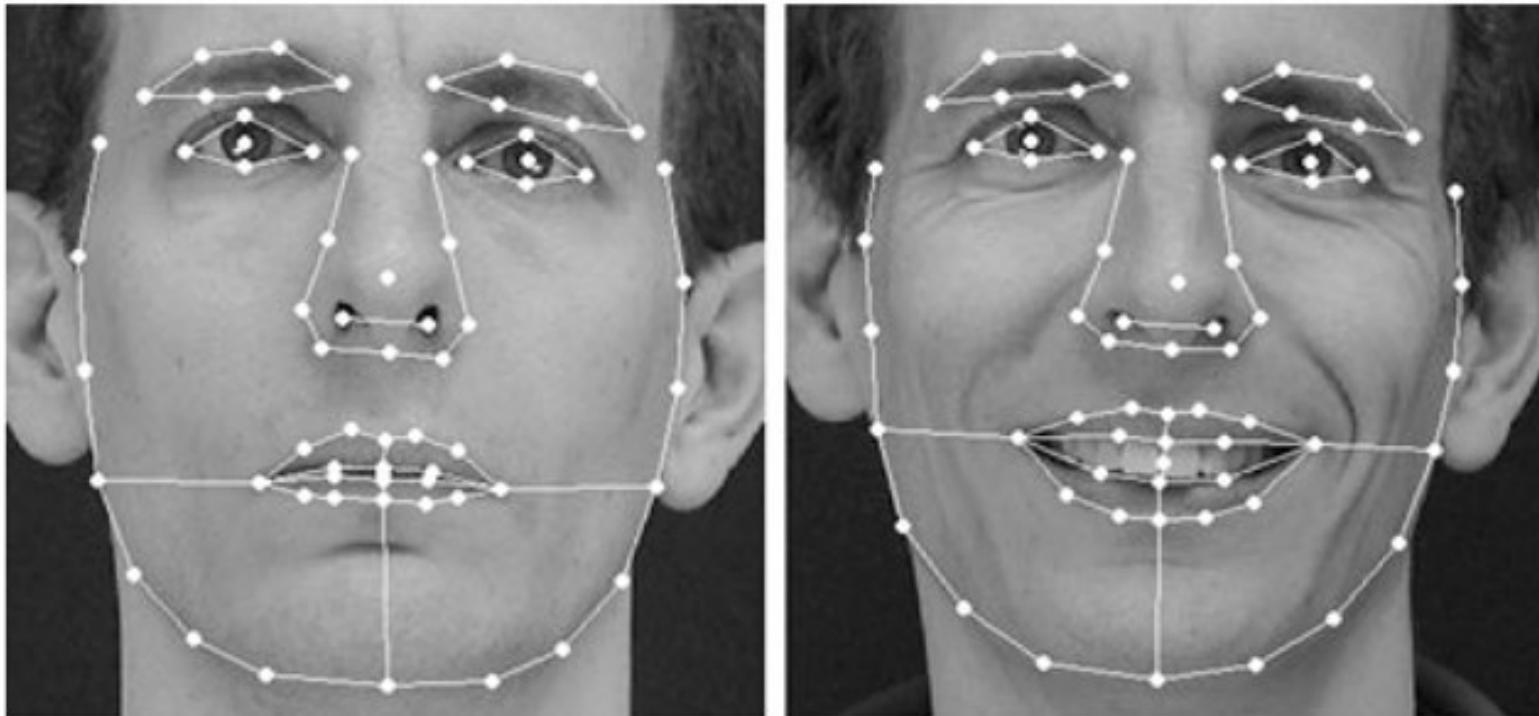
A comunidade de pesquisa propôs a ideia de reconhecimento humano baseado em termografia facial com objetivo de fornecer um sistema robusto de reconhecimento de face.

Extração de *features* dos Sinais Reconhecimento Facial



Modalidades na região facial: (a) face, (b) termografia facial

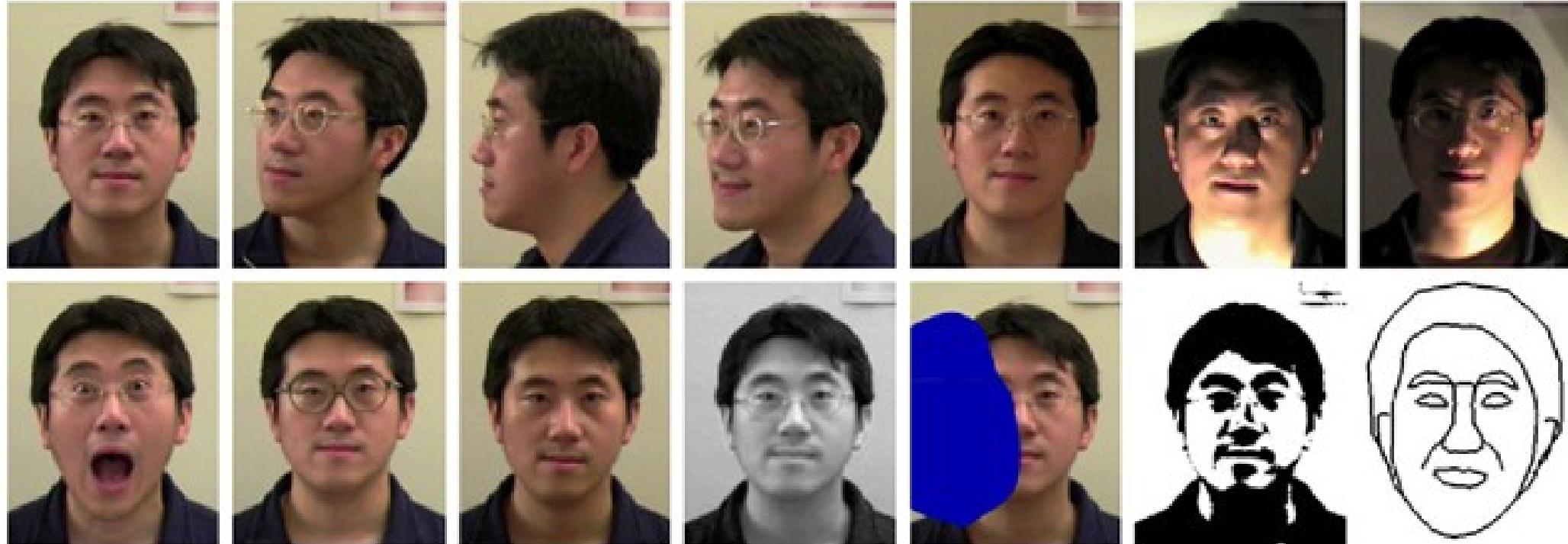
Extração de *features* dos Sinais Reconhecimento Facial



Identificação de pontos de referência em diferentes expressões faciais

Extração de *features* dos Sinais

Reconhecimento Facial



Para uma mesma pessoa temos variações de poses, iluminação, acessórios, brilho, etc

Extração de *features* dos Sinais Reconhecimento Facial



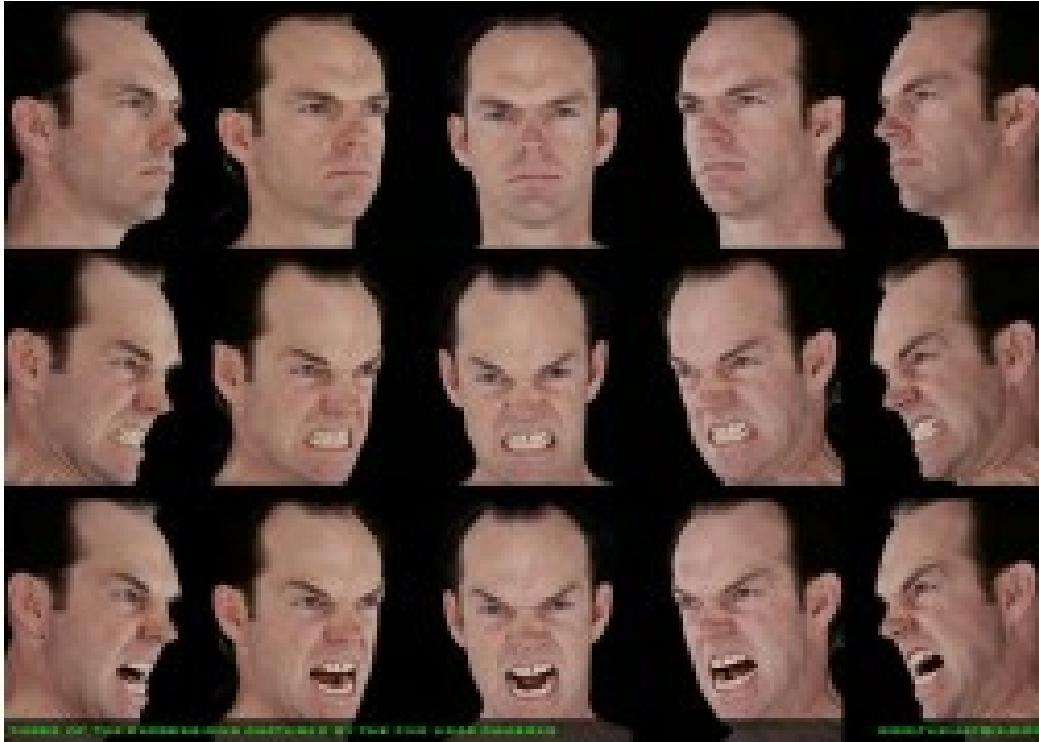
(a)



(b)

Para pessoas diferentes podem ocorrer similaridades entre irmãos gêmeos; e pais e filhos

Extração de *features* dos Sinais Reconhecimento Facial



Quanto maior a quantidade de imagens para treinamento, melhor o modelo

Extração de *features* dos Sinais

Reconhecimento Facial



Tendo em vista o crescimento no uso de técnicas de inteligência artificial em reconhecimento facial, várias empresas estão lançando serviços que se beneficiam dessa tecnologia.

A China é um país líder nesse setor em várias frentes: seja permitir sacar dinheiro em caixas eletrônicos de bancos sem uso de cartão, fazer compras em lojas de conveniência, viajar sem passagem ou identidade em um aeroporto e outros.

Extração de *features* dos Sinais

Reconhecimento Facial



Entretanto, alguns atos podem ser considerados polêmicos;

Como cobrar multas para as pessoas que não atravessam na faixa de pedestre, ou guardas usarem smart glasses para avaliar quem o sistema considera como infrator.

E quando ocorrerem casos de falso positivos?

O sistema pode identificar pessoas que realmente cometem crimes graves como assassinato ou roubo, pessoas que simplesmente atravessaram a rua fora da faixa, ou até mesmo pessoas que foram identificadas erroneamente.

Extração de *features* dos Sinais Reconhecimento Facial



CARNAVAL

CARNAVAL 2019 NA BAHIA

Procurado por homicídio vai para o carnaval de Salvador vestido de mulher e é preso após ser flagrado por câmera

Jovem de 19 anos teve imagem gravada por câmera de reconhecimento facial instalada nas ruas da festa. Marcos Vinicius estava fantasiado para sair no Bloco As Muquiranas.

Extração de *features* dos Sinais Reconhecimento Facial



Extração de *features* dos Sinais Reconhecimento Facial



G1

POP & ARTE

San Francisco proíbe a polícia de usar reconhecimento facial

Oito dos nove conselheiros municipais são contrários à tecnologia.

G1

CAMPINAS E REGIÃO 



Por France Presse

16/05/2019 09h05 · Atualizado há uma semana

Campinas testa câmera de reconhecimento facial como a que flagrou procurado no carnaval

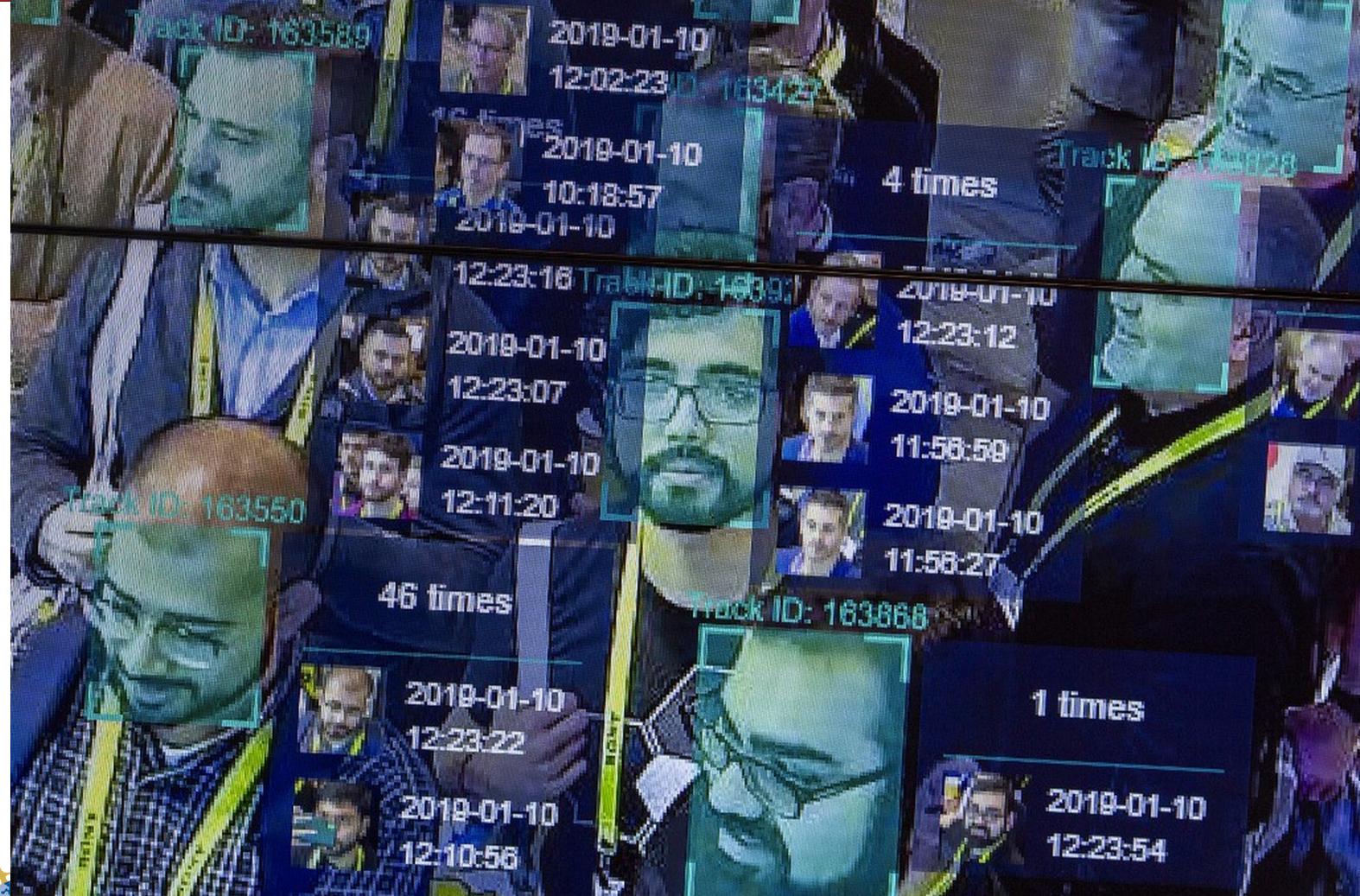
Em Salvador (BA), procurado por homicídio foi identificado mesmo fantasiado de mulher. Implantado no Terminal Central e em instalação na rodoviária e no centro, sistema será testado até junho.

Por Helen Sacconi, EPTV 2

09/03/2019 19h43 · Atualizado há 4 meses



Extração de *features* dos Sinais Reconhecimento Facial



Extração de *features* dos Sinais Reconhecimento Facial



Extração de *features* dos Sinais

Reconhecimento Facial



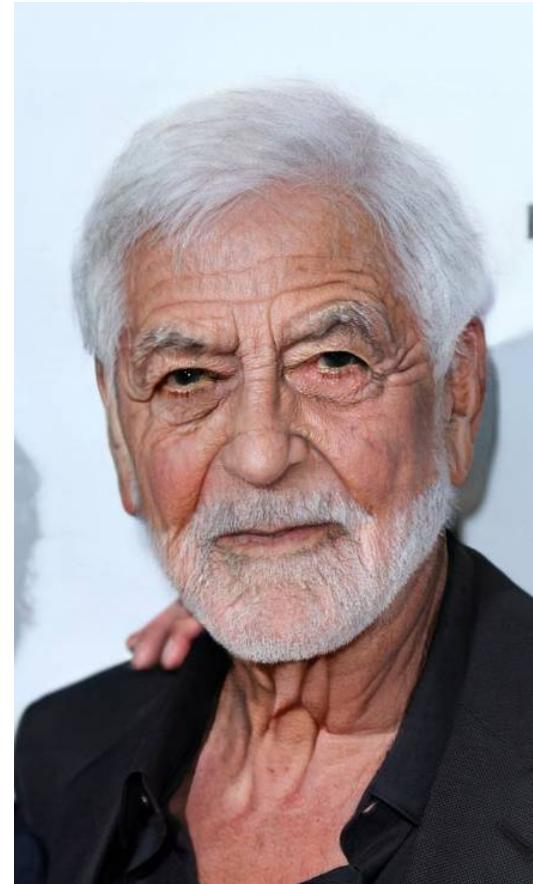
Ainda teremos muitos debates sobre o tema e aplicação consciente do uso de reconhecimento facial, sendo que alguns lugares são mais favoráveis (como em Londres) e outros já entraram em processo de banimento (como em San Francisco).

O uso de reconhecimento facial é uma das técnicas biométrica mais famosas atualmente por estas aplicações.

Apesar de polêmico, cabe ao governo, empresas legais e sociedade verem formas de aplicar seu uso de forma benéfica para a população, sem tentar infringir questões de privacidade.

Extração de *features* dos Sinais

Reconhecimento Facial



Extração de *features* dos Sinais

Reconhecimento através de sinais vitais



A maioria dos sistemas pode ser dividida em cinco etapas principais: aquisição, pré-processamento do sinal, extração de features, correspondência e classificação.

Geralmente, os sinais vitais são capturados por um sensor e pré-processados para remoção de ruídos.

Depois da segmentação e normalização, aplica-se a extração de features. As features resultantes são processadas para formar um template, o qual é comparado com o template de usuários autorizados.

Extração de *features* dos Sinais

Reconhecimento através de sinais vitais

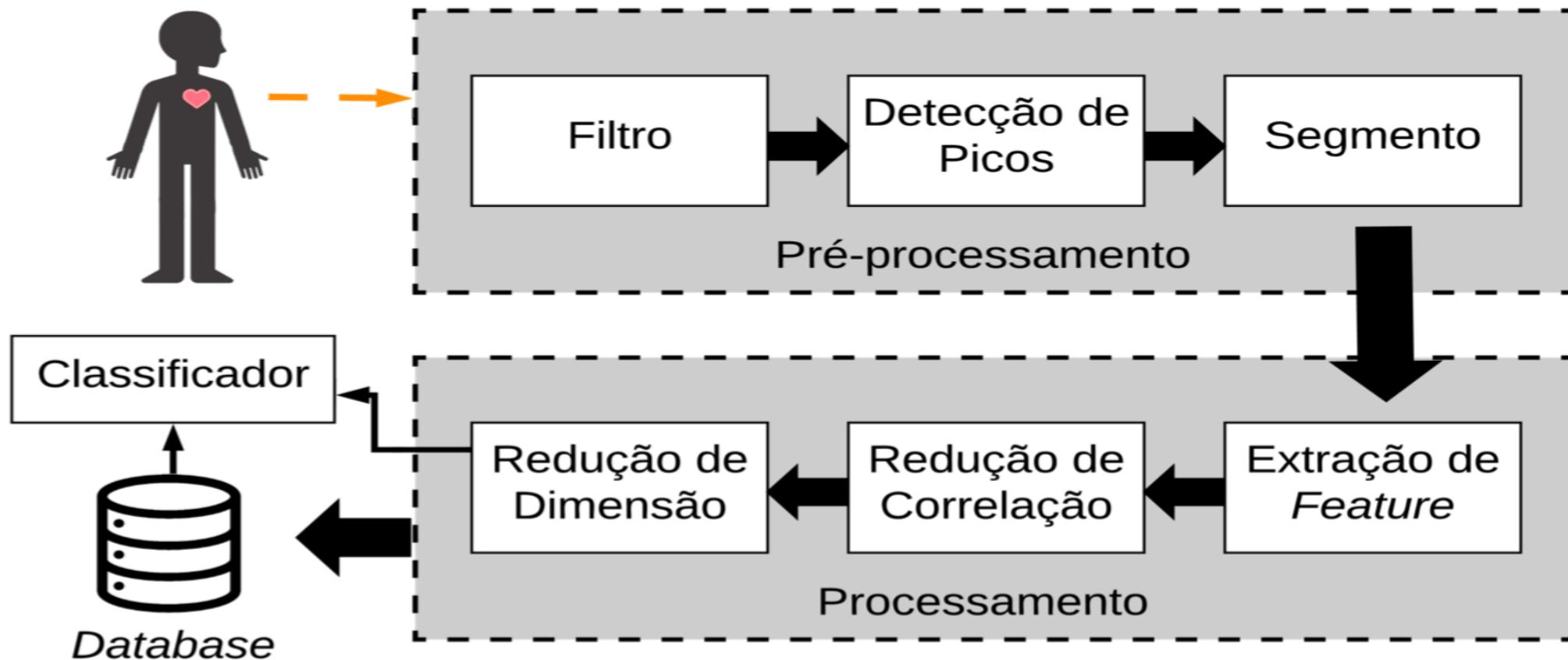


Finalmente, a classificação é aplicada para distinguir os dados de sinais vitais genuínos dos dados de sinais vitais impostores.

Por conta disso, a extração das features é a etapa mais importante, pois é quando as características do usuário são extraídas do sinal vital para que o processo de autenticação seja realizado.

Extração de *features* dos Sinais

Reconhecimento através de sinais vitais



Extração de *features* dos Sinais

Reconhecimento através de PPG



A fotopletismografia (PPG) é um método eletro-óptico, não invasivo, que mede o volume sanguíneo que flui através da parte do corpo humano em análise (exemplo: pulso, dedo, lóbulos auriculares, etc).

Os sinais PPG refletem as ações pulsativas das artérias através da interação entre a hemoglobina oxigenada e os fótons.

Acredita-se que cada pessoa tenha uma hemodinâmica e um sistema cardiovascular únicos.

Por conta disso, os sinais PPG podem ser utilizados para autenticação biométrica.

Extração de *features* dos Sinais

Reconhecimento através de PPG



Os sinais PPG são registrados através de uma combinação de LED, que emite luminosidade em uma parte do corpo, e Foto-Diodo (PD), que mede a luz absorvida pelos tecidos epiteliais.

Esta combinação proporciona maior flexibilidade para o projeto de sistemas de autenticação. As medições indicam as mudanças no volume sanguíneo.

Como o registro do PPG requer apenas LED e PD, ele é muito econômico, comparado aos outros traços biométricos.

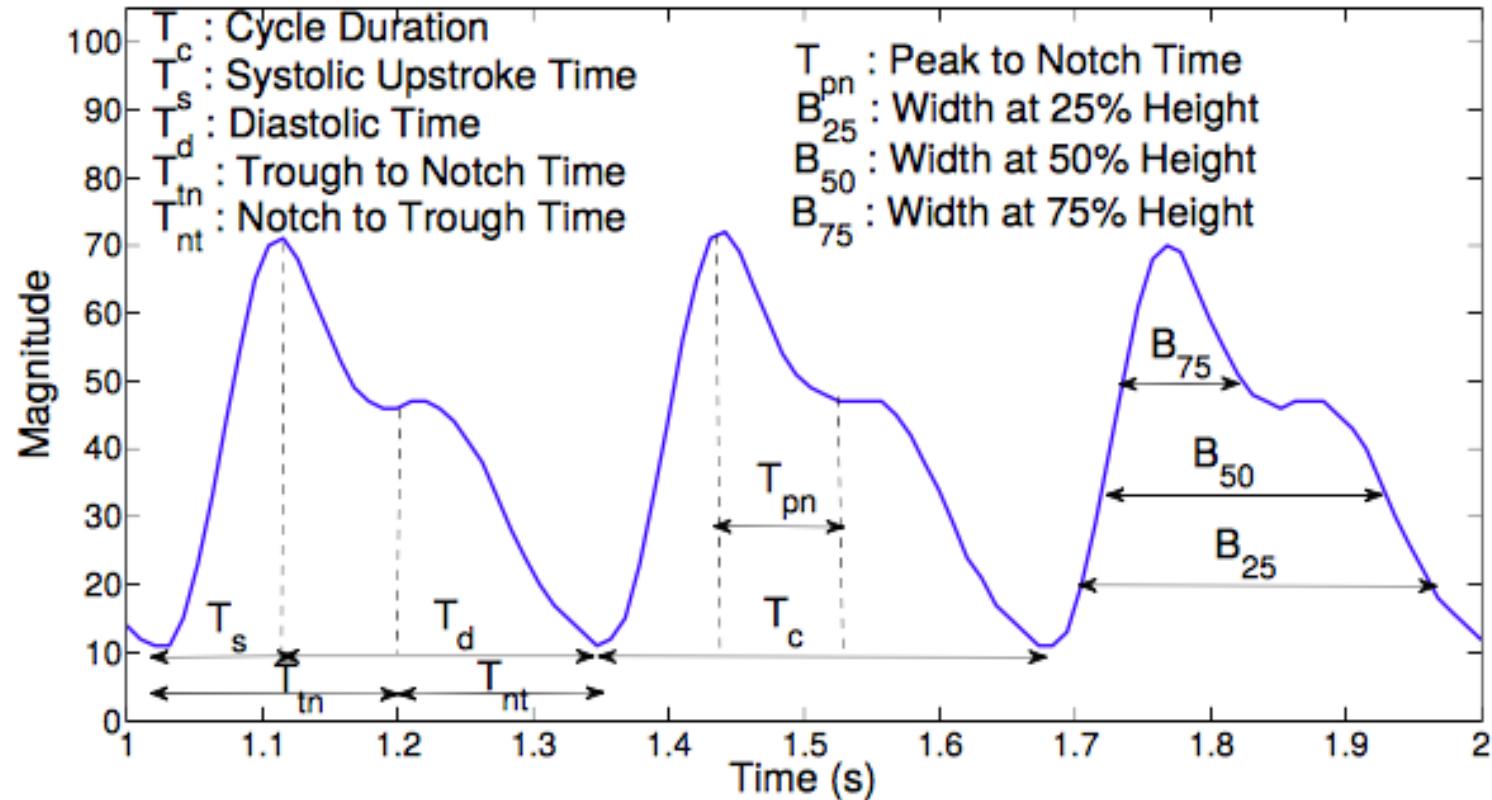
Extração de *features* dos Sinais

Reconhecimento através de PPG



Extração de *features* dos Sinais

Reconhecimento através de PPG



Extração de *features* dos Sinais

Reconhecimento através de PPG



No contexto da biometria médica, o registro do PPG não requer nenhum tipo de gel, estímulo externo ou vários eletrodos e pode ser convenientemente registrado de praticamente qualquer parte do corpo.

As features de um sinal PPG podem ser utilizadas para identificar diferentes indivíduos, ao mesmo tempo, similares o suficiente para reconhecer uma mesma pessoa.

Extração de *features* dos Sinais

Reconhecimento através de PPG



Os sinais PPG têm diversas vantagens para autenticação de usuários quando comparados com outras abordagens biométrica.

Eles possuem baixo custo de desenvolvimento e são acessíveis a várias partes do corpo humano (dedo, lóbulo da orelha, pulso ou braço).

Por conta disso, muitos trabalhos concentram-se em pesquisar sobre o uso dos sinais PPG como identificadores biométricos.

Extração de *features* dos Sinais

Reconhecimento através de PPG



Em primeiro lugar, as gravações PPG desses dispositivos podem frequentemente ter ruídos devido aos movimentos contínuos dos usuários.

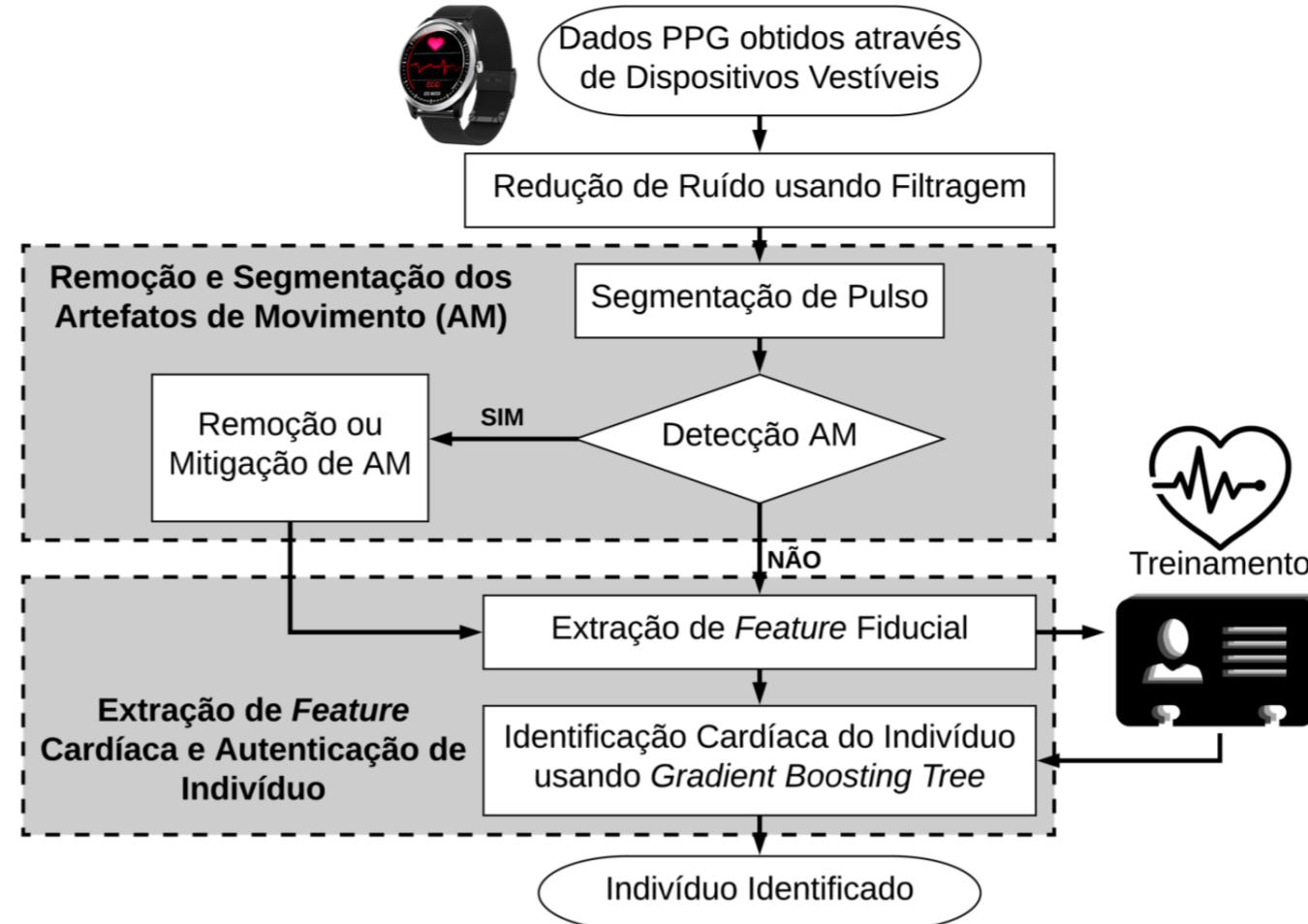
Em segundo lugar, a ocorrência de eventos nas artérias podem registrar ruídos. Assim, existe a necessidade de desenvolver algoritmos que respondam aos fatores acima de detecção de ruídos.

Pesquisas na área de processamento de sinais vitais envolvem o estudo de sinais como ECG, eletroencefalograma (EEG) e PPG para prever ampla gama de eventos fisiológicos no corpo humano.

Algumas das aplicações incluem reconhecimento de emoções, detecção de crises e detecção do estágio do sono.

Extração de *features* dos Sinais

Reconhecimento através de PPG



Extração de *features* dos Sinais



Reconhecimento através de ECG

Extração de *features* dos Sinais

Reconhecimento através de ECG



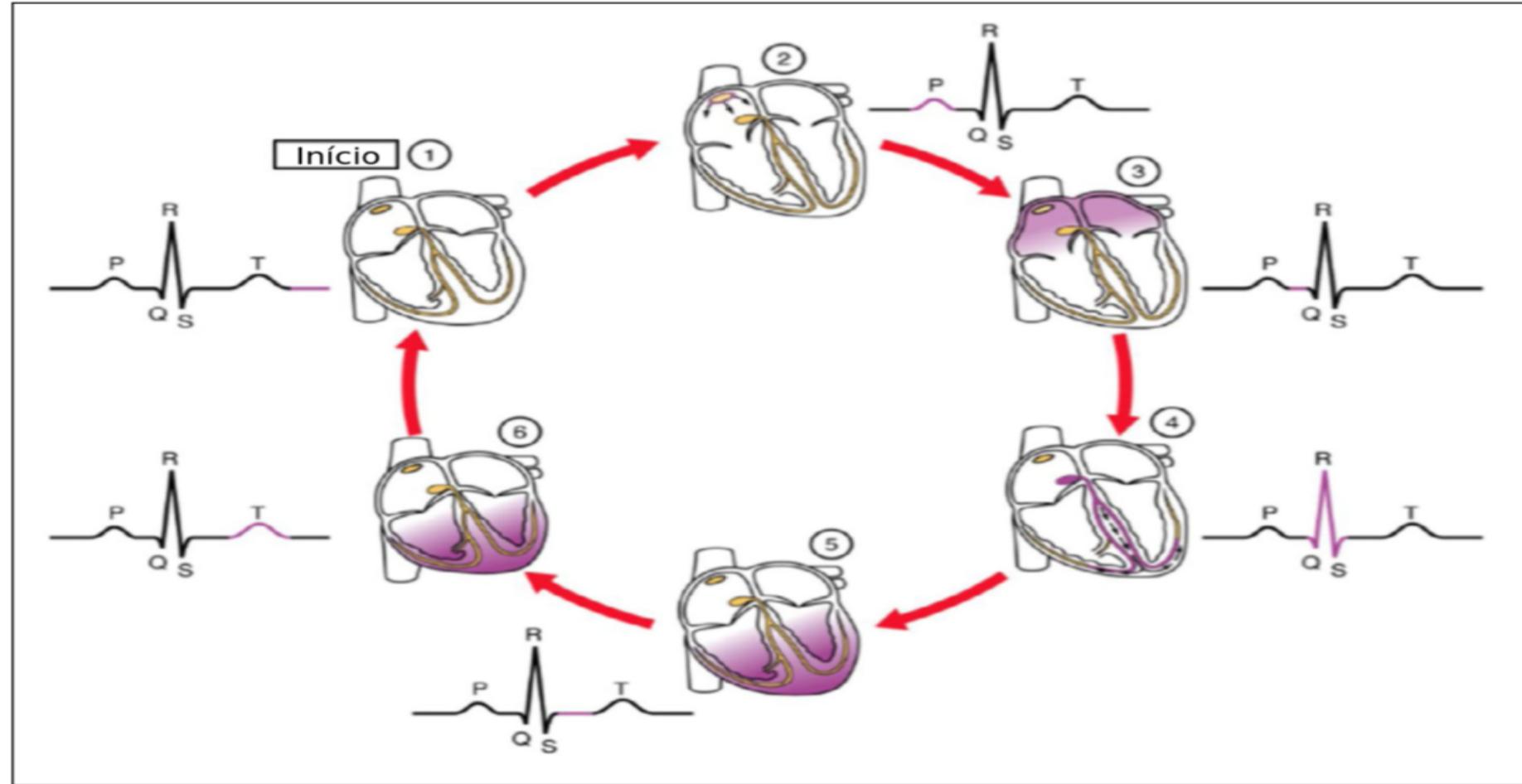
As batidas do coração geram ondas de polarização e despolarização nas fibras musculares.

O eletrocardiograma (ECG) é feito de uma forma não evasiva, o qual representa simplesmente o registro da atividade elétrica cardíaca baseada nas diferenças de potencial resultantes.

Sua amostra está associada ao ciclo cardíaco.

Extração de *features* dos Sinais

Reconhecimento através de ECG



Etapas de funcionamento do coração e forma de onda captada pelo ECG.

Extração de *features* dos Sinais

Reconhecimento através de ECG



O ECG é bastante útil para várias aplicações biomédicas, tais como a **medição da taxa de frequência cardíaca**, exame de ritmo das batidas do coração em busca de arritmias, **diagnóstico de anormalidades do coração**, reconhecimento de emoção e mais recentemente identificação biométrica.

Para a extração de features do sinal de ECG é necessário que o sinal elétrico coletado seja tratado previamente, passando por um processo de filtragem para retirada de ruídos, normalização e amostragem.

Extração de *features* dos Sinais

Reconhecimento através de ECG

O ECG apresenta três componentes predominantes: onda P, complexo QRS e onda T.

Primeiro, a despolarização do átrio gera um pulso registrado como onda P.

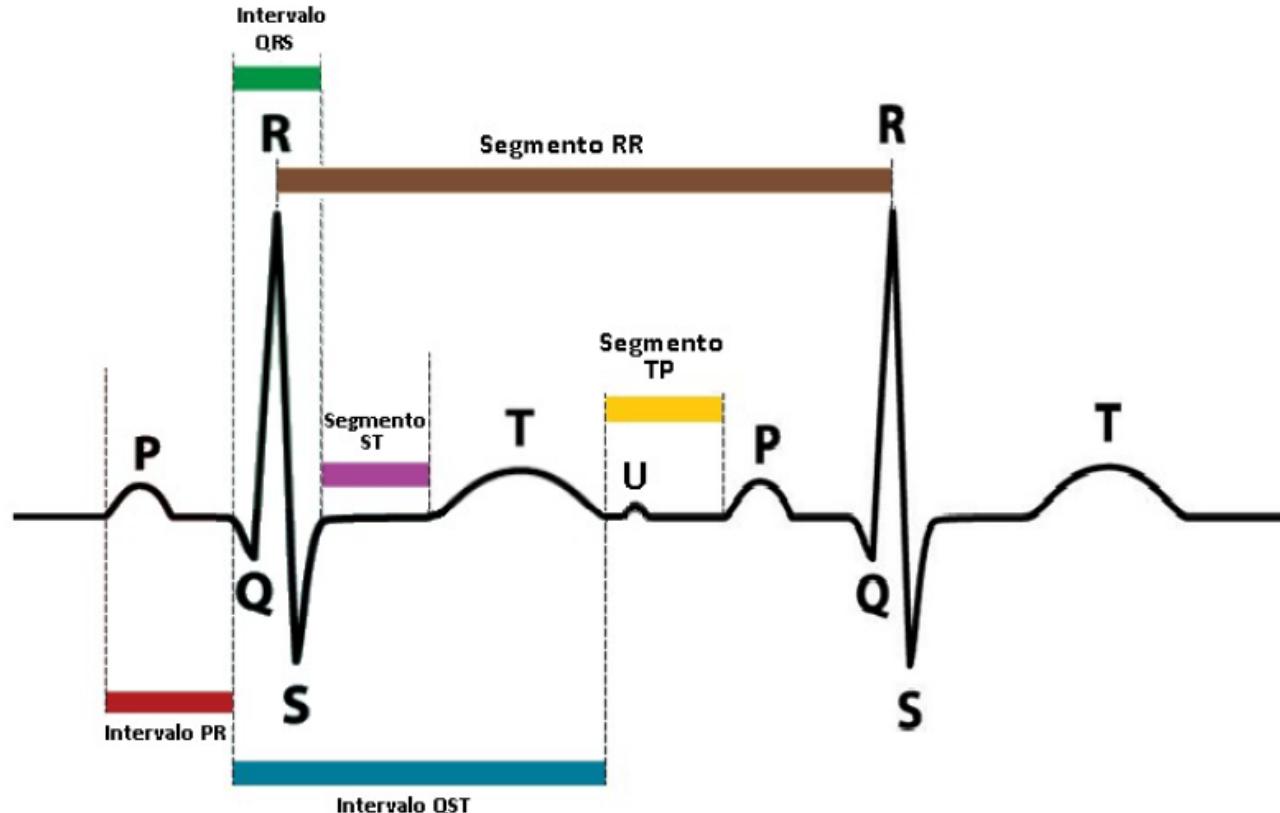
A série de pulsos seguinte a onda P é o complexo QRS e está associada com a atividade ventricular.

Finalmente, a onda T está associada à repolarização ventricular.

Este complexo P-QRS-T é o mais utilizado para identificação de pessoas. Ele basicamente corresponde às localizações, durações, amplitudes e formas de onda do sinal coletado do coração (ou seja, ECG).

Extração de *features* dos Sinais

Reconhecimento através de ECG



A literatura apresenta três tipos de features: fiduciais, não fiduciais e híbridas.

As **features fiduciais** extraem características no **domínio do tempo** das formas de onda ECG, as **features não fiduciais** aplicam uma **função de transformação** aos pontos característicos, e as **features híbridas** são a **combinação** das features fiduciais com as features não fiduciais.

Extração de *features* dos Sinais

Reconhecimento através de ECG



Há pesquisas que usam ECG como identificador biométrico, um dos primeiros foi proposto em 2001.

Especificamente, avaliamos um survey que pesquisou cinquenta estudos dedicados à identificação humana, onde 66% dos artigos pesquisados empregaram características não-fiduciais, 26% aplicaram características fiduciais, e 8% dos trabalhos de pesquisa usaram a abordagem híbrida.

Extração de *features* dos Sinais

Reconhecimento através de ECG



Quanto ao método de classificação, 44% dos trabalhos de pesquisa selecionaram os algoritmos k-Nearest Neighbour (k-NN) ou Nearest Center (NC), 16% implementaram Redes Neurais (ANN) e 16% utilizaram Linear Discriminant Analysis (LDA). Finalmente, 12% das pesquisas atingiram uma acurácia superior a 99% e 20% dos artigos pesquisados atingiram 100% de acurácia.

De um modo geral, as características fiduciais, não-fiduciais ou híbridas não influenciam diretamente na acurácia.

Extração de *features* dos Sinais

Reconhecimento através de ECG



Quais as aplicações e vantagens de utilizar o ECG?

Em trabalhos recentes (2018), tivemos algumas aplicações para cenário de uma torre de controle de tráfego aéreo.

Nesse conceito, os pesquisadores assumiram que os controladores exigiam monitoramento permanente para evitar incidentes de segurança, como um intruso tentando usar o sistema de controle, um controlador assumindo a posição de um colega e a alta variação da frequência cardíaca do controlador devido a uma situação estressante.

- Autenticação Contínua

Extração de *features* dos Sinais Reconhecimento através de ECG



Os autores consideraram um dispositivo médico implantável para capturar o sinal de ECG e uma etapa de filtragem para limpar a corrente contínua (CC).

Nesse sentido, o módulo de extração de recursos analisa a janela de ECG e a envia para uma Transformada Walsh-Hadamard (HT).

O modelo de aprendizado de máquina implementou um algoritmo k-NN, usando o conjunto de dados MIT-BIH Normal Sinus Rhythm.

Para a estratégia de monitoramento contínuo, o sistema atingiu 96% de acurácia.

Extração de *features* dos Sinais

Reconhecimento através de ECG



Outros trabalhos já demonstraram a possibilidade de capturar o ECG através de eletrodos de dois dedos associados a um aplicativo de smartphone.

Nos dois casos, a aplicação do sistema considera três etapas: etapa de filtragem de pré-processamento, etapa de extração de features e etapa de classificação. A filtragem elimina o ruído associado à linha de energia, interferência, movimento muscular e ruído de alta frequência.

O módulo de detecção funciona como uma pré-etapa para a extração de features. Os autores consideraram o complexo de computação de extração de features não fiduciais e, assim, decidiram implementar o modelo usando formas de onda divididas por fiduciais (WF).

Extração de *features* dos Sinais

Reconhecimento através de ECG



The EKG device that
goes where you go.

Peace of mind
in 30 seconds.

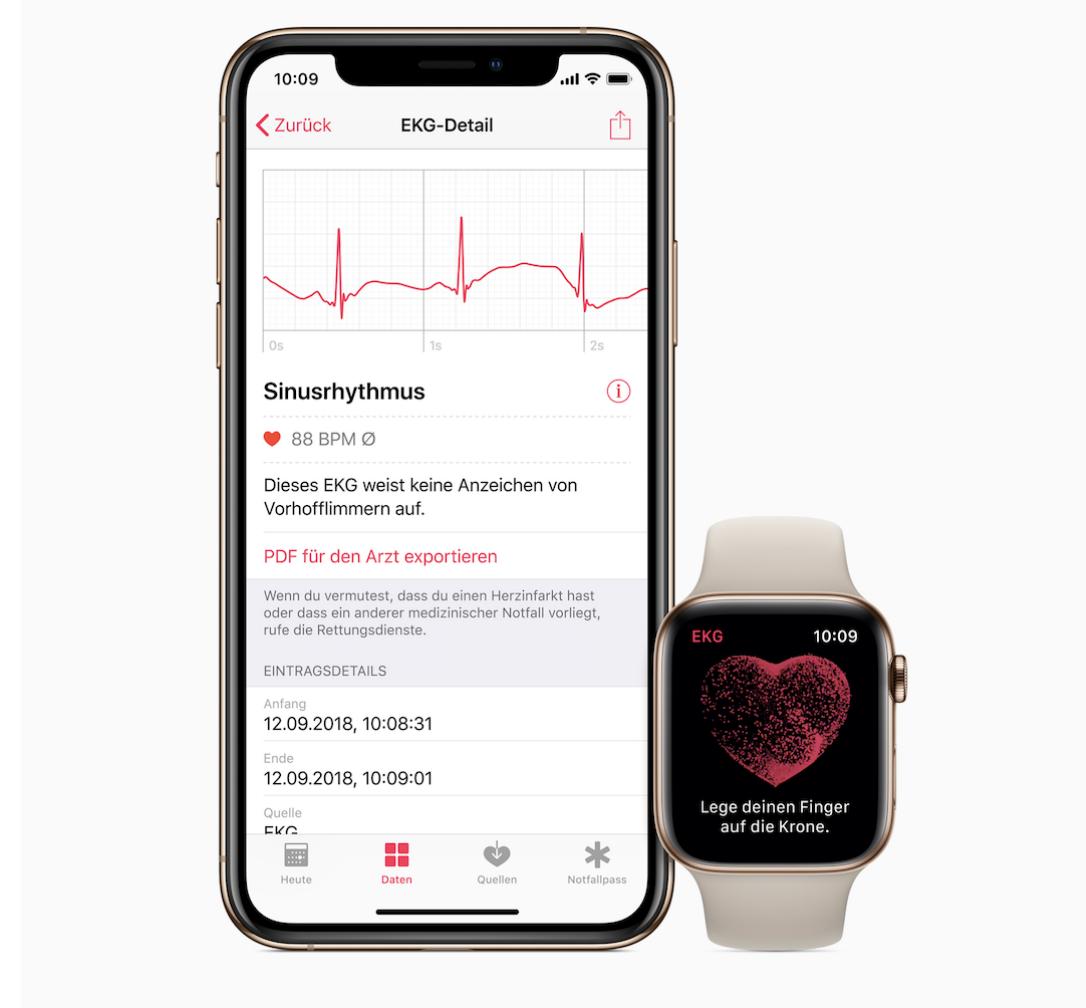


Compatible
with most Apple
or Android
devices.



Extração de *features* dos Sinais

Reconhecimento através de ECG



Extração de *features* dos Sinais

Reconhecimento através de ECG



A etapa de classificação refere-se a dois métodos biométricos: autenticação e identificação.

Durante a autenticação, o sistema compara a biometria do usuário com um modelo armazenado que calcula a distância euclidiana como uma métrica de comparação.

Para identificação, o sistema realiza uma classificação usando Support Vector Machine (SVM) e Redes Neurais (ANN).

Para ambos os casos, um mecanismo de votação exigiu que mais da metade dos eleitores validassem o assunto do teste. Projetos já atingem 97,55% de acurácia e executou a autenticação em quatro segundos.

Extração de *features* dos Sinais

Reconhecimento através de ECG



A seleção de features e a extração combinada com o aprendizado de máquina apresentam um papel relevante no reconhecimento de indivíduos.

As features do ECG contêm uma quantidade significativa de informações, mas algumas características são altamente correlacionadas.

É preciso conduzir a presente pesquisa por meio de simulações para concluir a melhor combinação entre o algoritmo de aprendizado de máquina versus o tipo de features no contexto de dispositivos vestíveis.

Além disso, a identificação correta de um número significativo de indivíduos é uma condição obrigatória para a avaliação do modelo.

Extração de *features* dos Sinais

Reconhecimento através de ECG



Apesar de pesquisadores já terem provado o conceito de usar um smartphone para reconhecimento de identidade baseado em um sinal ECG.

Os testes consideraram três conjuntos de dados ECG abertos e, portanto, os resultados práticos podem ser diferentes.

Considerando que os dispositivos portáteis, como o AppleWatch, já fornecem ECG, outras empresas seguirão a tendência em breve.

Nesse sentido, o uso do ECG como método de autenticação se tornará frequente.

Extração de *features* dos Sinais Sistemas Multimodais



O monitoramento de sinais ECG pode ser usado como uma ferramenta essencial para monitorar condições de saúde de pacientes, assim como para identificação de indivíduos.

As principais limitações dos sistemas biométrico:

- i) condições ambientais variáveis (i.e., ruído, mudanças na iluminação posicionamento da impressão digital ou da face em relação ao sensor);
- ii) grandes variações intra-classe causadas pela aquisição o em diferentes condições ou efeitos de envelhecimento;
- iii) não-universalidade de alguma credencial biométrica, devido a doença ou deficiência,
- iv) ataques de fraudes que são realizados falsificando um traço biométrico e, em seguida, apresentando essas informações falsificadas ao sistema biométrico

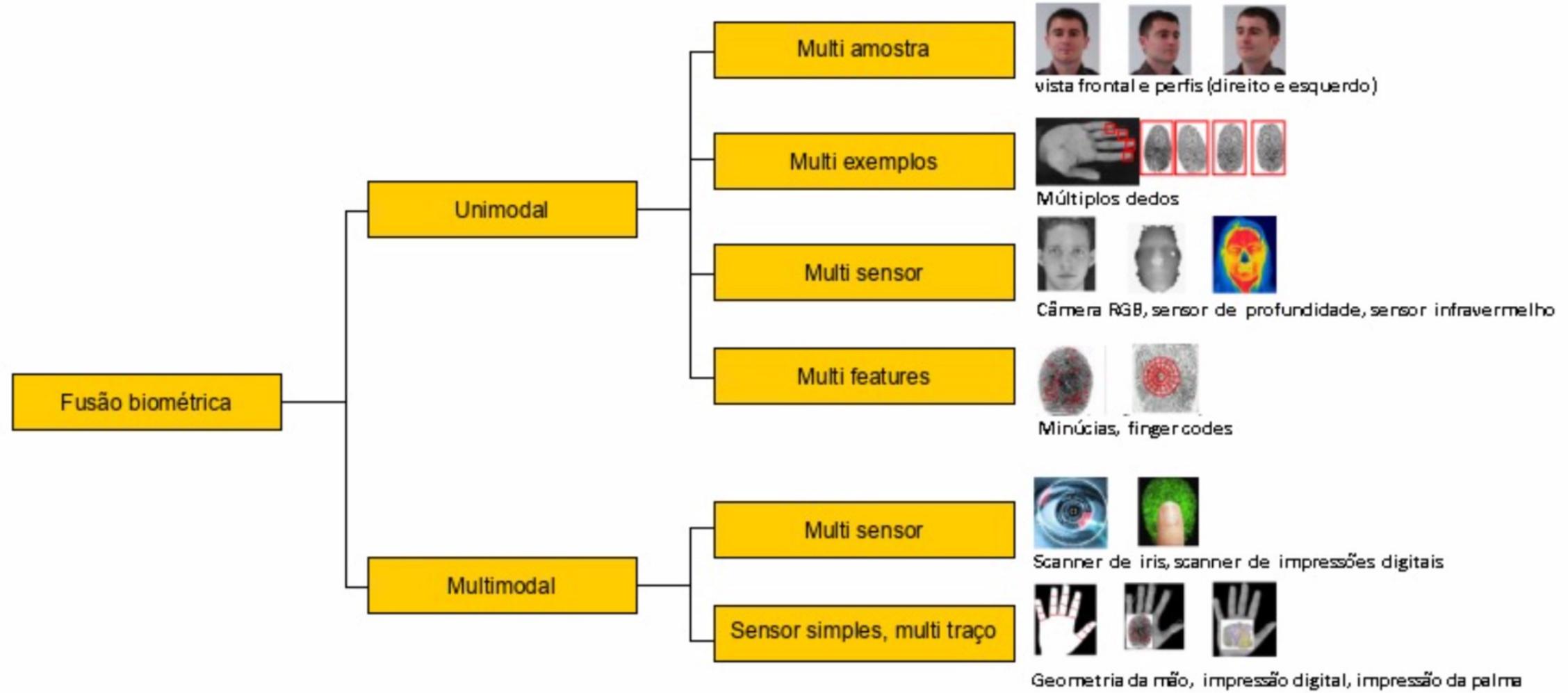
Extração de *features* dos Sinais Sistemas Multimodais



Em alguns casos, o desempenho de uma simples modalidade biométrica é insuficiente e, por isso, combinar biomarcadores tem se tornado um interesse acadêmico.

Essa combinação é chamada “fusão biométrica”, a qual pode ser classificada em dois grupos: sistemas biométricos unimodais e sistemas biométricos multimodais.

Extração de *features* dos Sinais Sistemas Multimodais



Extração de *features* dos Sinais Sistemas Multimodais



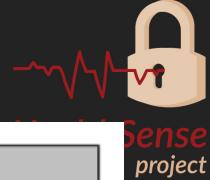
A multimodalidade tem diversas vantagens sobre a unimodalidade, devido à sua capacidade de reduzir a taxa de falha de registro (do inglês Failure-to-Enroll Rate (FTER)) e a taxa de falha de captura (do inglês Failure-to-Capture Rate (FTCR)) e garantir uma cobertura populacional suficiente.

Por exemplo, este processo estima que 2% da população pode não estar apta para fornecer uma impressão digital, devido a condições médicas/ genéticas/ accidentais/ temporárias.

Os sistemas multimodais são também resistentes a ataques fraudulentos porque é mais difícil para o atacante fraudar múltiplas fontes biométrica simultaneamente.

O futuro dos sistemas biométricos está relacionado aos sistemas multimodais.

Extração de *features* dos Sinais



Sistema biométrico	Prós	Contras
Mão	O pequeno <i>template</i> torna os atributos da região da mão uma boa opção para muitos tipos de aplicações, comparados com assinaturas biométricas mais complexas.	Pode comprometer facilmente a eficácia do sistema.
Ocular	É mais preciso, altamente confiável, bem protegido, estável e quase impossível de forjar as assinaturas biométricas.	Alto custo de sensores de imagem; Requer cuidados com fontes de luz ambiente.
Facial	Aquisição de imagem não intrusiva e sem contato, resultando em maior aceitação pública.	Esses sistemas não podem garantir identificação confiável na presença de ruídos.

Extração de *features* dos Sinais

Sistema	Feature	Descrição
ECG	<i>Features</i> ECG não podem ser copiadas ou manipuladas.	É necessário o uso de vários eletrodos durante a aquisição dos sinais ECG, causando desconforto no indivíduo. Muda de acordo com a frequência cardíaca (repouso ou exercício)
PPG	Sua coleta requer apenas LED e Foto-Diodo, ou seja, não requer nenhum tipo de gel, estímulo externo ou vários eletrodos; pode ser coletado de qualquer parte do corpo; é econômico, comparado aos outros traços biométricos.	Geralmente é prejudicado por ruídos durante a aquisição (artefatos de movimento, movimentos de sensor, respiração, contração ventricular prematura e luz ambiente); muda junto com a frequência cardíaca (repouso ou exercício); as emoções influenciam o funcionamento do sistema nervoso autônomo e do coração.

Extração de *features* dos Sinais Sistemas Multimodais



O sistema biométrico que faz uso da mão e o que faz uso facial, apesar de serem não intrusivos, não passam muita segurança quanto a probabilidade de serem forjados.

Já o sistema que considera os olhos como identificador biométrico, apesar de ser altamente confiável e protegido, acaba sendo custoso devido aos sensores de imagem exigidos.

Entre os sinais ECG e PPG, apesar de o PPG ser mais fácil e mais confortável de ser coletado, ele é mais prejudicado por ruídos durante a aquisição que o ECG. Sinais ECG são também mais fáceis de serem interpretados, de acordo com a natureza da sua forma de onda.

Parte II



- Técnicas de Aprendizado de Máquina

Aprendizado de Máquina



Conhecer as diversas técnicas de Aprendizado de Máquina possibilita muitas oportunidades em diversas aplicações.

Veremos algumas como:

- I. Artificial Neural Networks (ANN);
- II. k-Nearest Neighbor (k-NN);
- III. Support Vector Machine (SVM);
- IV. Naïve Bayes (NB);
- V. K-means clustering
- VI. Decision Tree (DT)
- VII. Random Forest (RF)

Aprendizado de Máquina

Redes Neurais



As redes neurais artificiais (ANN) são algoritmos de aprendizado de máquina capazes de classificar dados lineares e não lineares.

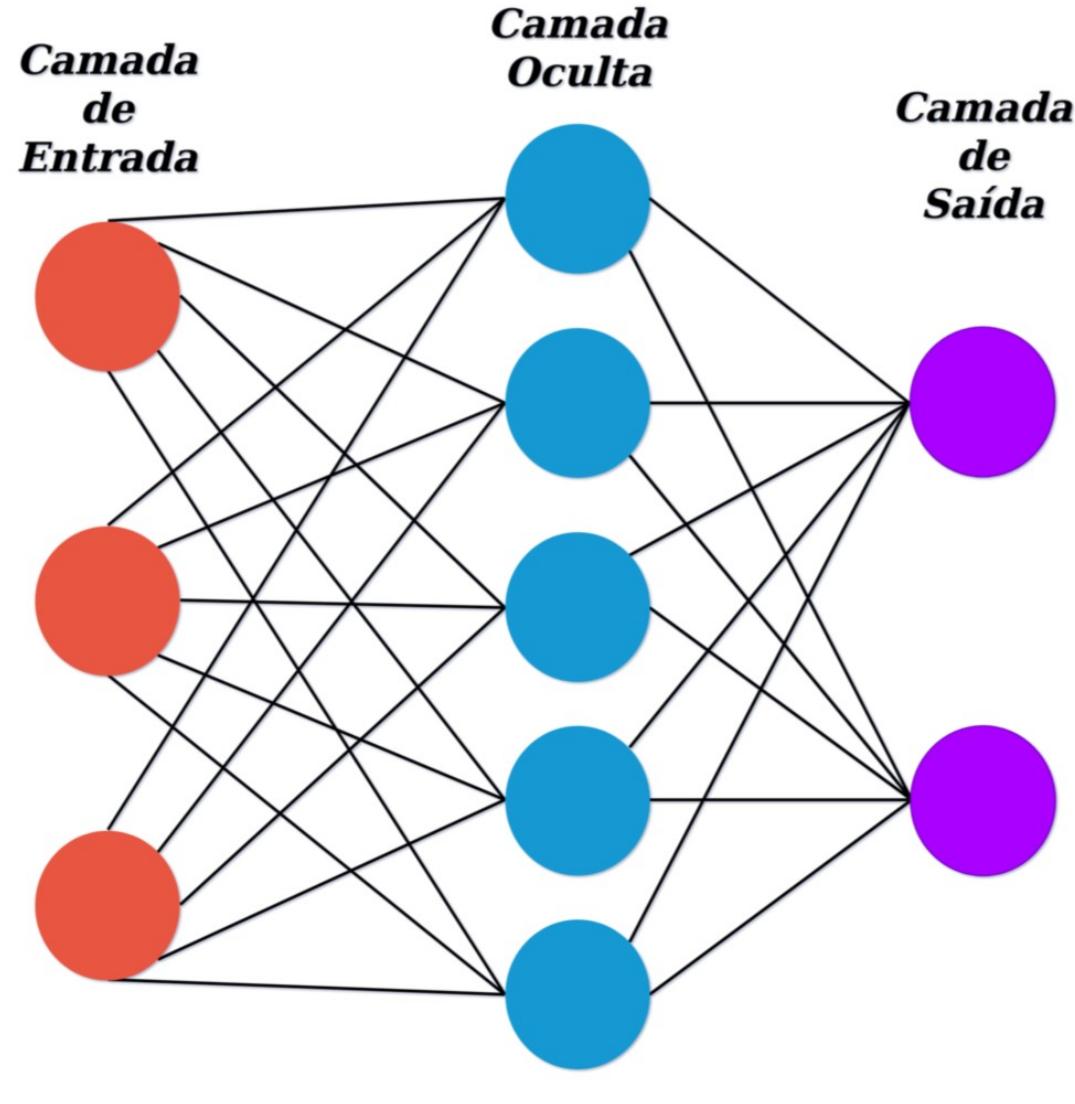
A arquitetura típica é inspirada nas redes neurais biológicas e apresenta um determinado número conectado de neurônios dispostos em diferentes camadas.

A estrutura de dados representa um neurônio artificial, geralmente configurado com ponteiros conectados a outros neurônios e um valor de peso (número real) para ponderar cada conexão

Aprendizado de Máquina

Redes Neurais

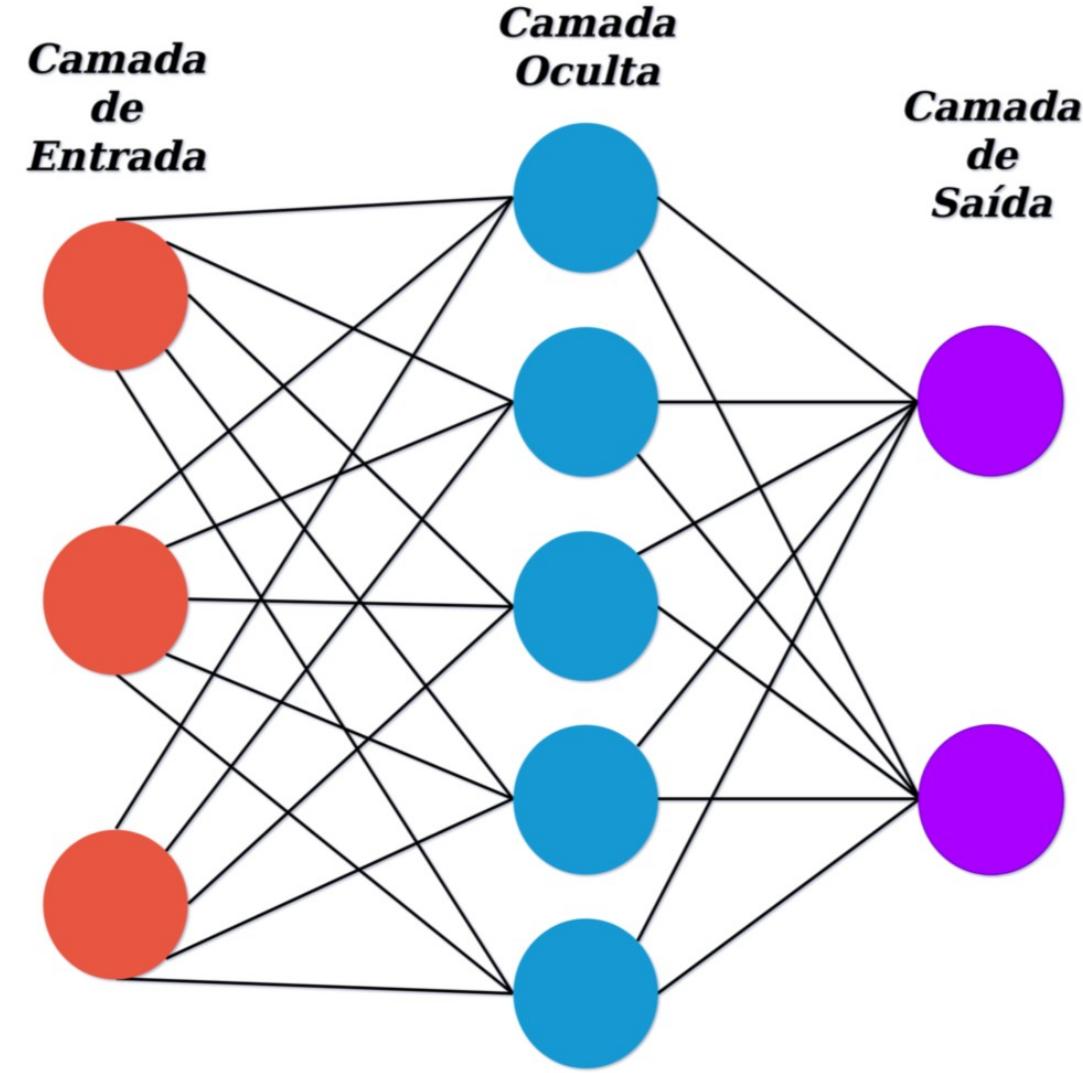
Multilayer Perceptron (MLP) refere-se a uma rede neural configurada com um número variável de neurônios dispostos em uma camada de entrada, uma ou mais camadas ocultas, e uma camada de saída



Aprendizado de Máquina

Redes Neurais

Especificamente, o MLP propaga um estímulo injetado nos neurônios na camada de entrada, através dos neurônios conectados nas camadas ocultas, e reflete na camada de saída.

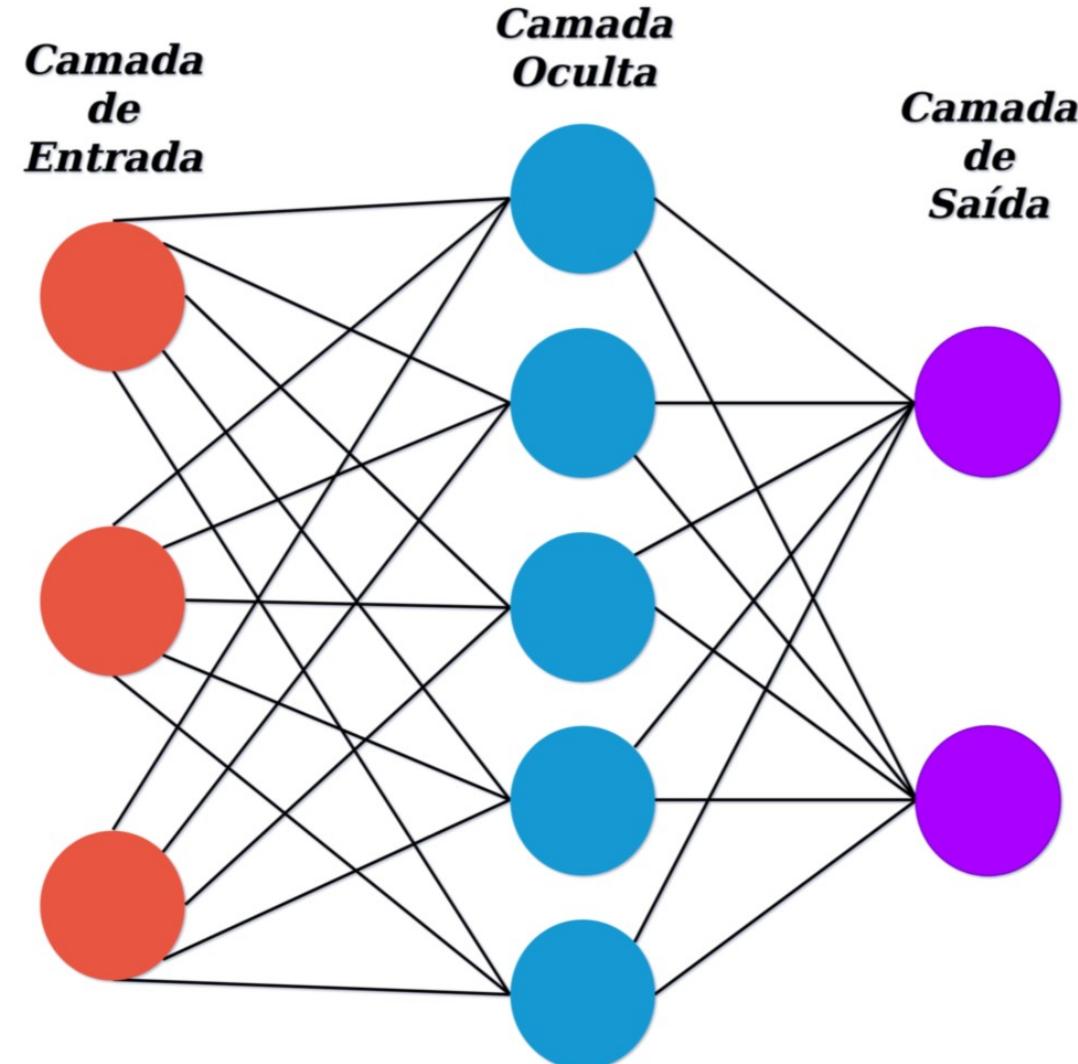


Aprendizado de Máquina

Redes Neurais

Processamento Direto: Inicialmente, cada neurônio recebe um valor de peso fixo. O algoritmo transfere o sinal injetado da camada de entrada para a camada de saída usando os pesos para ponderar o resultado.

Processamento Reverso (Back Propagation): O algoritmo calcula o erro entre as saídas obtidas e as saídas desejadas, e retro-propaga o erro na rede ajustando os pesos do neurônio no caminho de volta.

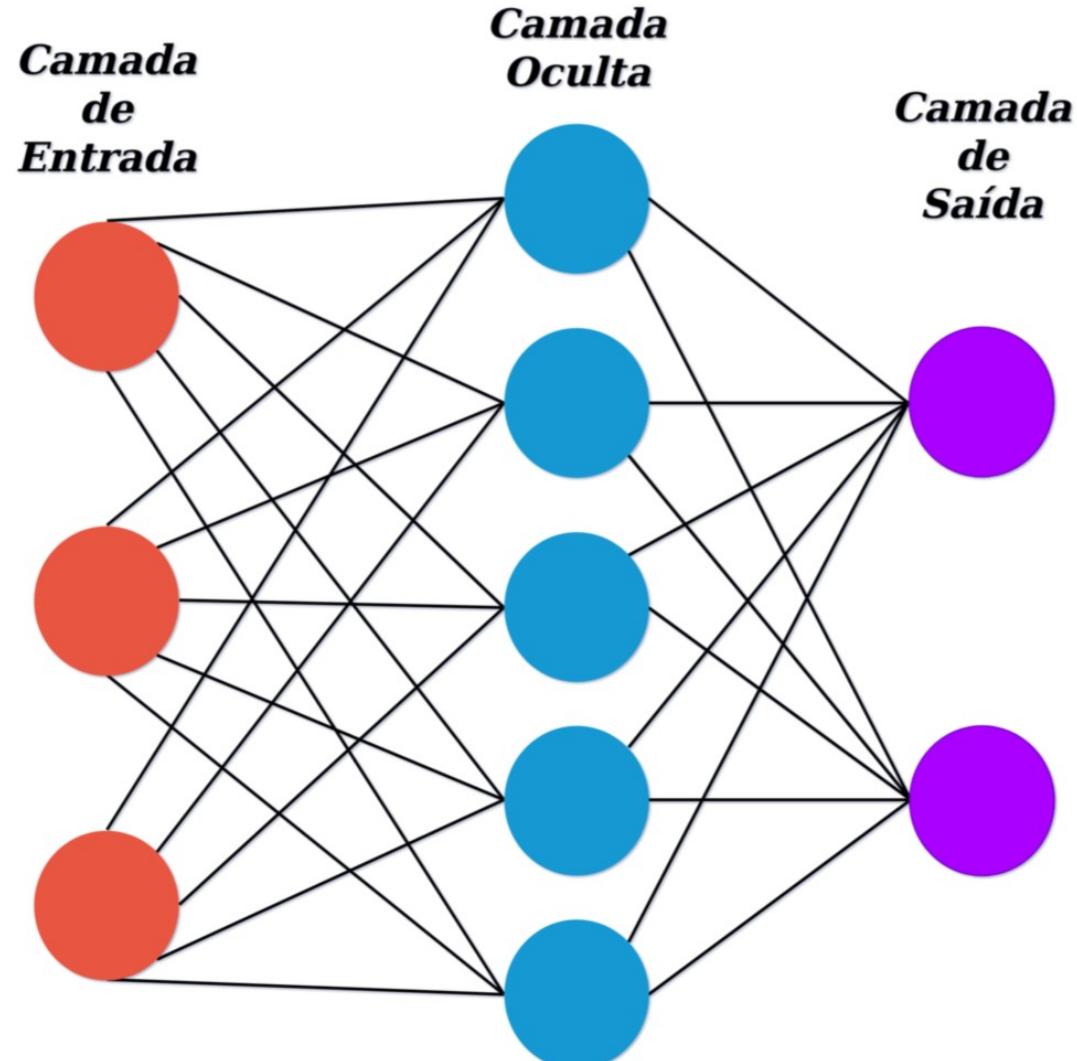


Aprendizado de Máquina

Redes Neurais

O processo de treinamento repete as fases de processamento direto e reverso por um determinado número de iterações.

Cada iteração se concentra em minimizar o erro entre a saída real e a saída desejada. O processo de treinamento termina quando o erro cai abaixo de um determinado limite ou quando o algoritmo atinge o número máximo de iterações estabelecidas.



Aprendizado de Máquina

Redes Neurais



Convolutional Neural Network (CNN) é um tipo de rede neural artificial (ANN) que apresenta algumas características distintas.

CNN apresenta uma ou mais camadas de unidades de convolução, que reduz as unidades em mapeamentos muitos-para-um.

Devido a sua menor complexidade, CNN acaba precisando de menos amostras, o que torna o tempo de treinamento menor e deixa o modelo mais rápido.

Aprendizado de Máquina

Redes Neurais



Deep Learning é uma técnica de rede neural que consiste em transformar um dado de entrada em várias camadas de representações abstratas.

Quanto mais camadas tiver na rede, maior será o número de features que vai aprender.

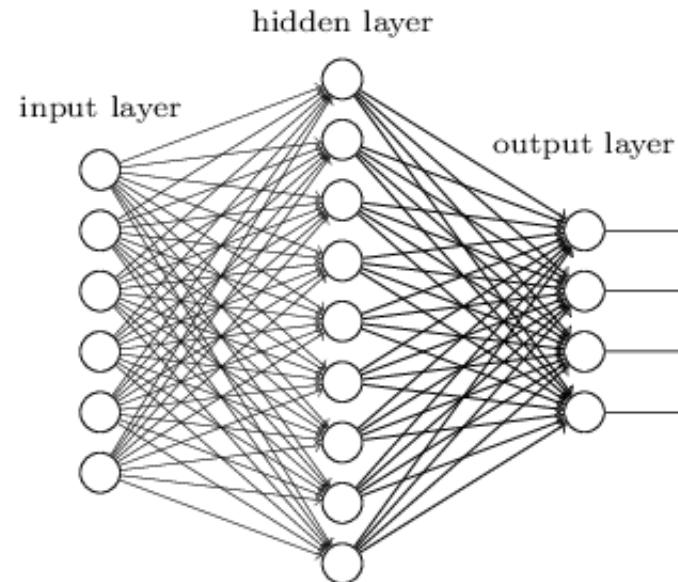
A camada de saída reunirá todas estas e features e fará a previsão.

Um dos impedimentos para se usar deep learning é que dependendo do número de neurônios, aumenta também o número de features e consequentemente a necessidade de aumentar a base de dados de treinamento.

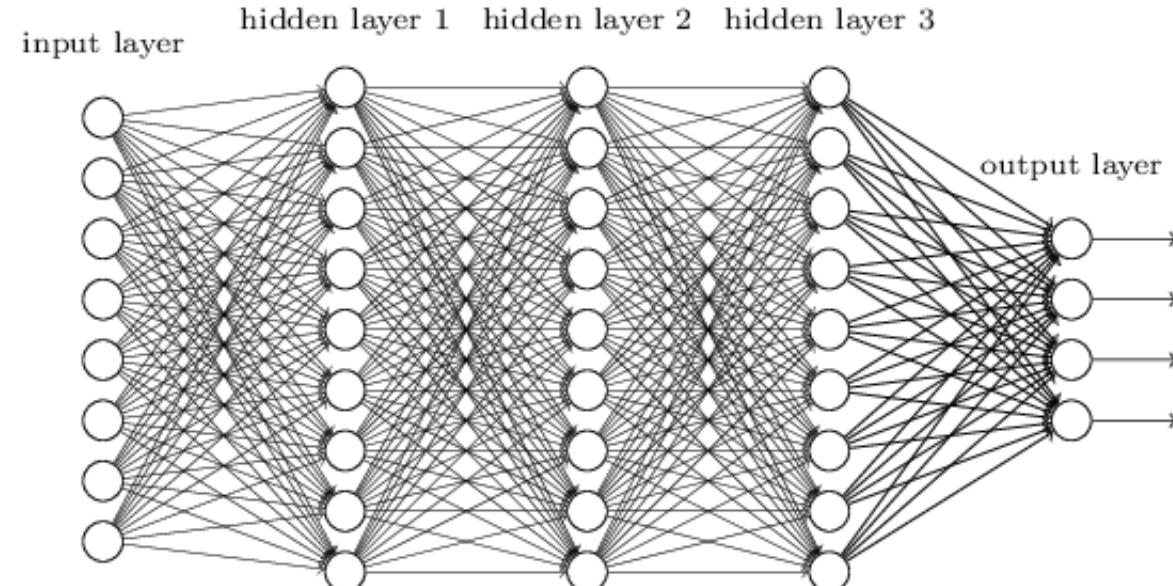
Aprendizado de Máquina

Redes Neurais

"Non-deep" feedforward neural network



Deep neural network



Aprendizado de Máquina

k-Nearest Neighbor (k-NN)



O k-NN classifica os vetores de características de acordo com os rótulos das amostras de treinamento mais recentes no espaço de recursos.

Para um vetor de característica desconhecida, as distâncias deste vetor para todos os vetores no conjunto de treinamento são calculadas usando uma medida de distância, e.g., distância euclidiana, entre um ponto de dados particular e seus k-vizinhos.

Em seguida, um vetor de recurso desconhecido é atribuído à classe na qual as amostras k mais próximas pertencem. Assim, um tipo de abordagem por maioria de votos é aplicado.

Aprendizado de Máquina

k-Nearest Neighbor (k-NN)



O valor de k é um inteiro positivo e é conhecido por ser um fator que influencia fortemente a precisão da classificação.

O k-NN tem a vantagem de não fazer uma suposição inicial sobre o conjunto de dados, pois apenas agrupa os pontos de dados baseados em uma vizinhança.

No entanto, o algoritmo armazena todos os dados de treinamento e só libera-os quando todos os dados são classificados. Normalmente, mas nem sempre, o algoritmo consegue uma melhor precisão com k valores mais altos.

Aprendizado de Máquina

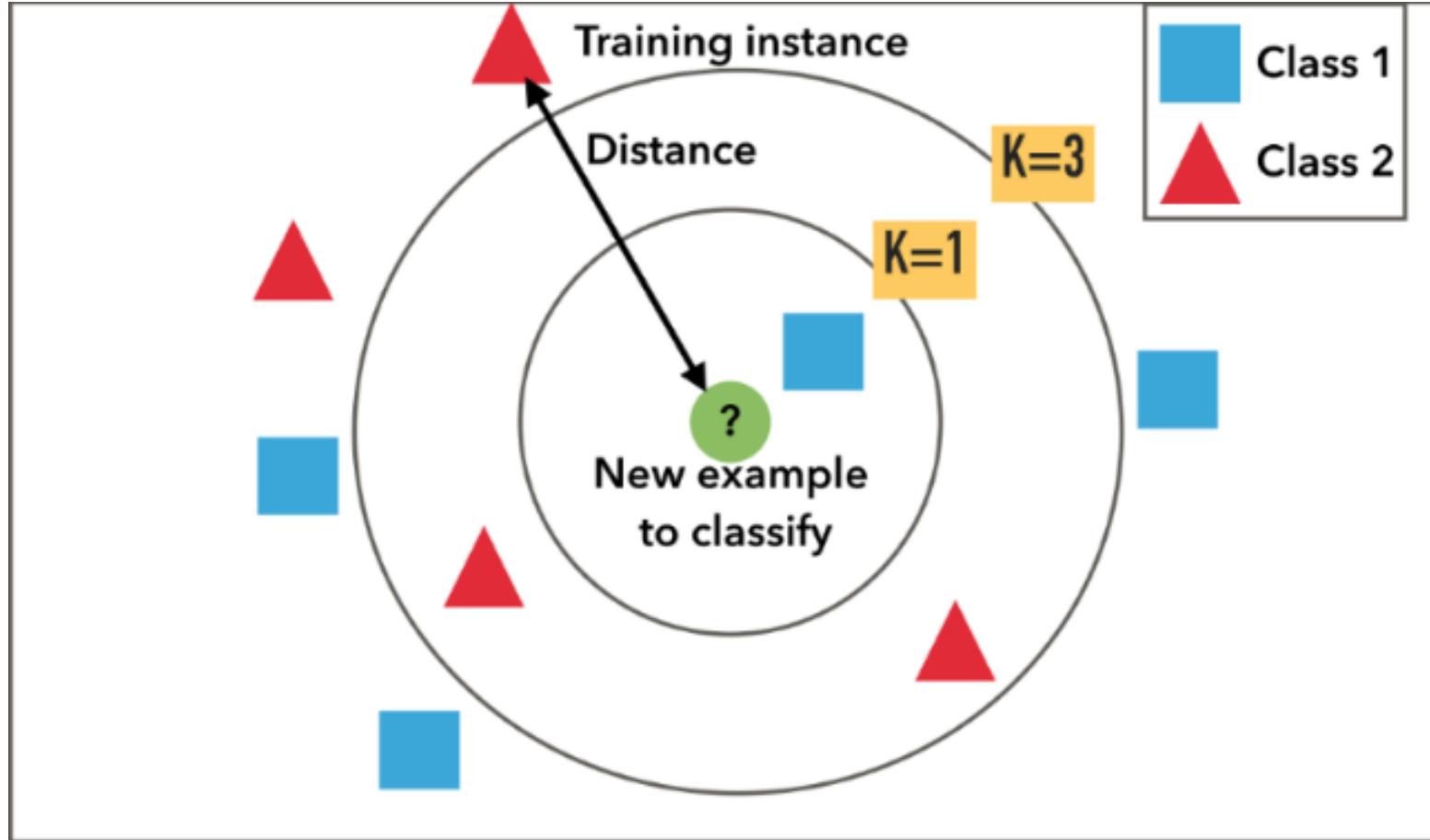
k-Nearest Neighbor (k-NN)



O algoritmo k-NN herda formas de tratar características de aplicações, como previsão meteorológica, detecção de quedas de idosos, detecção de crime, além de um uso amplo na maioria dos problemas de reconhecimento de padrões, como também é empregado em alguns estudos recentes de classificação de ECG ou de detecção de convulsões epilépticas

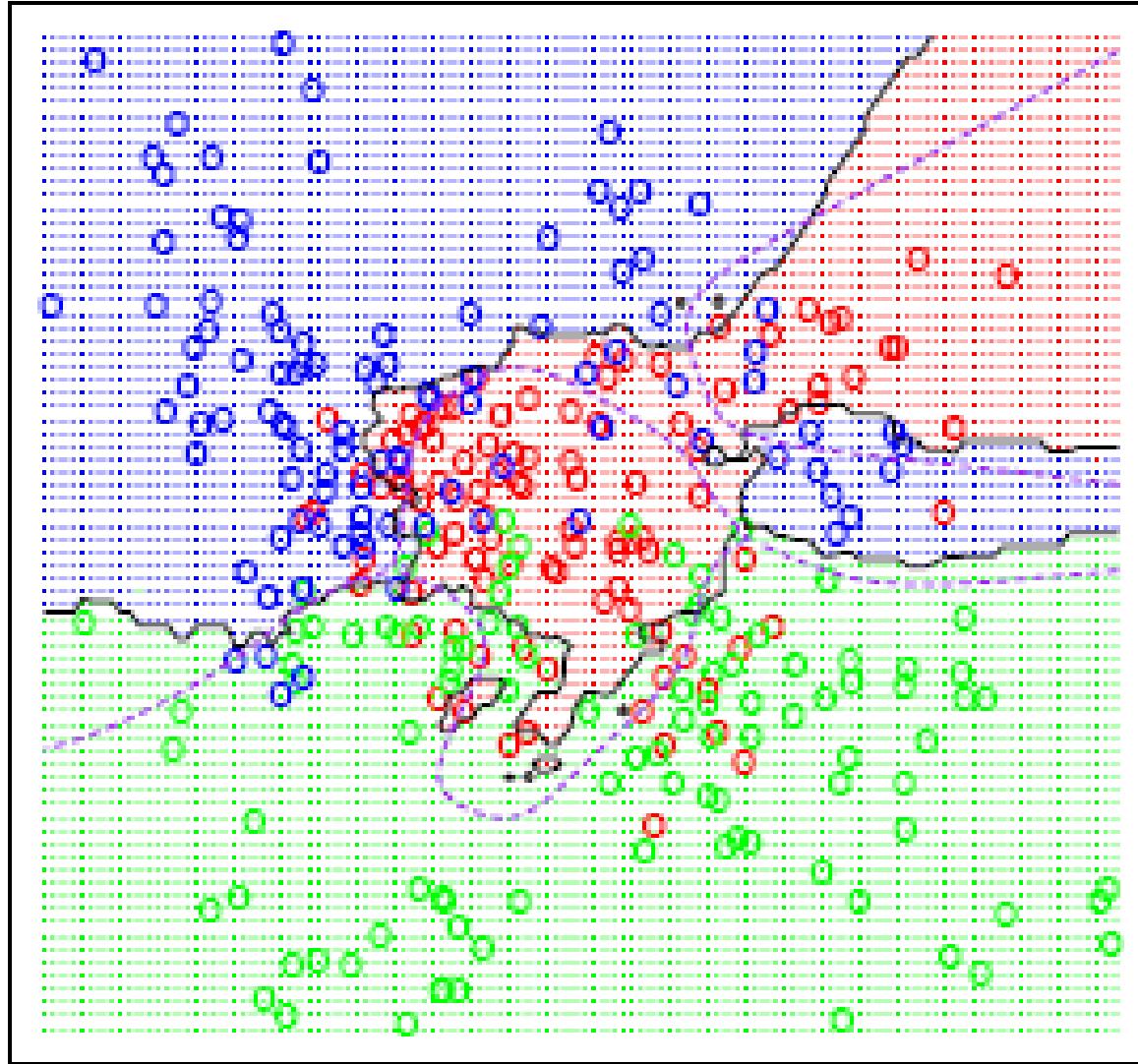
Aprendizado de Máquina

k-Nearest Neighbor (k-NN)



Aprendizado de Máquina

k-Nearest Neighbor (k-NN)



Aprendizado de Máquina

Support Vector Machine



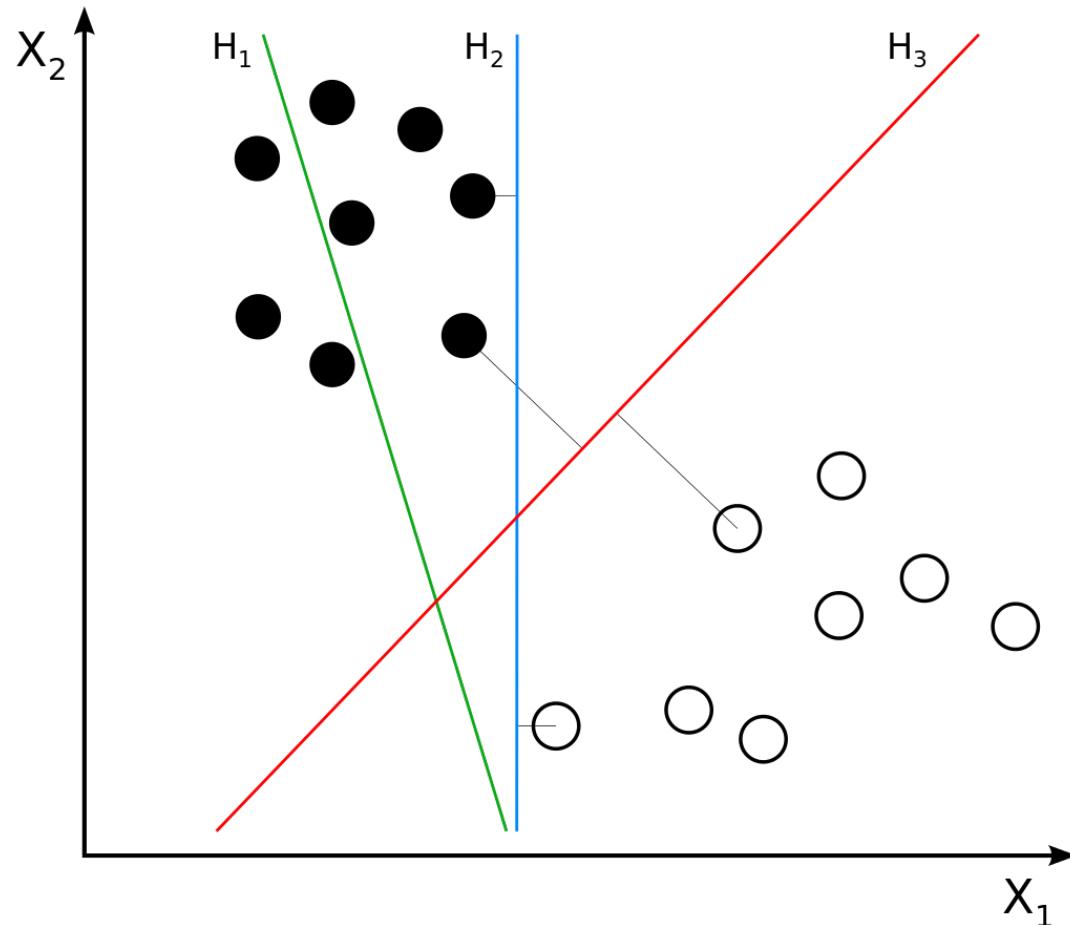
O SVM é uma ferramenta amplamente usada para resolver problemas de classificação binária devido ao seu excelente desempenho de generalização.

A ideia principal do SVM é encontrar uma margem máxima entre os dados de treinamento e o limite de decisão. Os vetores de suporte, que são as amostras de treinamento mais próximas do limite de decisão que são usados para a maximização da margem. O SVM pode ser considerado como um classificador linear ou não linear de acordo com o tipo de sua função kernel.

Enquanto uma função de kernel linear torna o SVM um classificador linear, outras funções do kernel, como base radial de gaussian, polinômio e sigmoide, fazem dele um classificador não linear. O SVM é utilizado na maioria dos estudos de classificação do ECG.

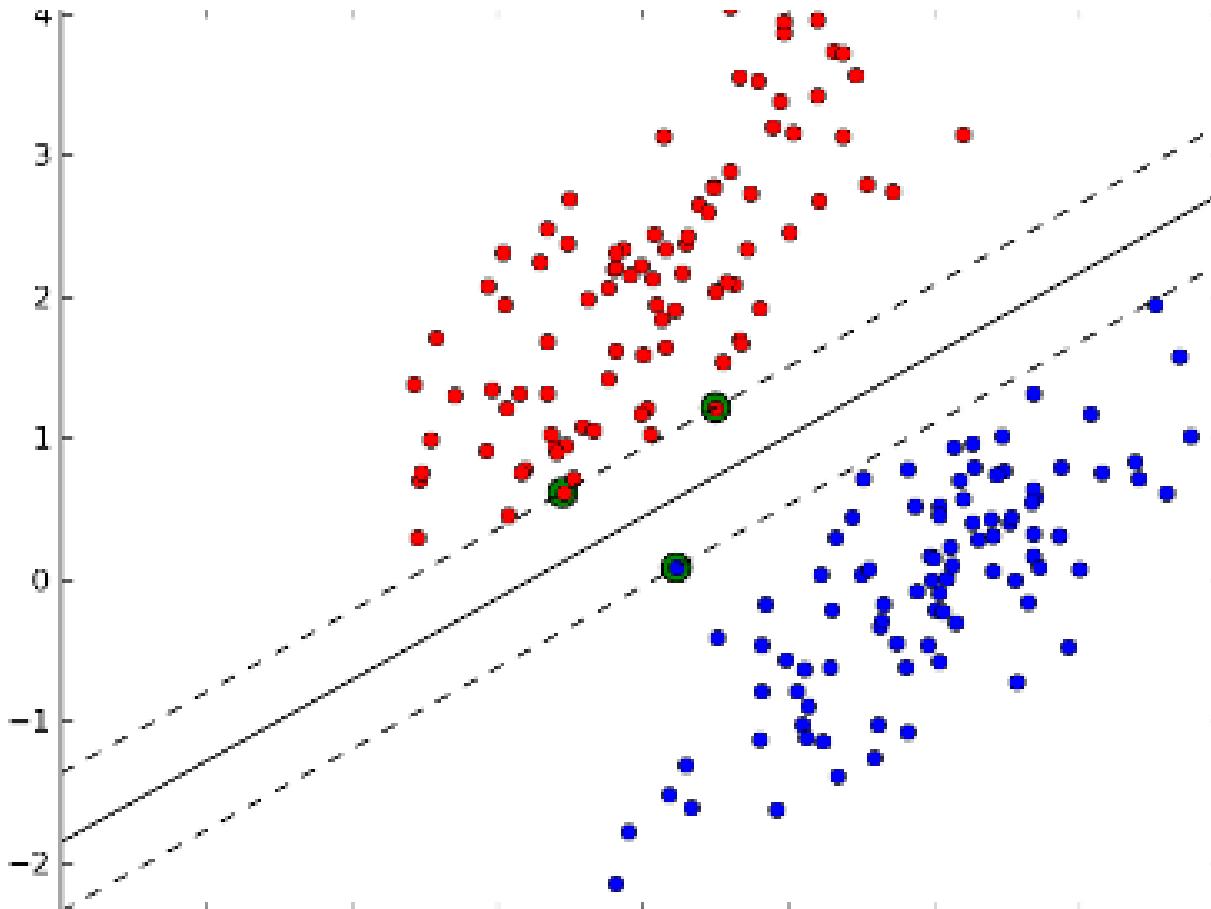
Aprendizado de Máquina

Support Vector Machine



Aprendizado de Máquina

Support Vector Machine



Aprendizado de Máquina

Naïve Baye(NB)



NB é um método de classificação probabilístico e estatístico baseado nas regras do Teorema de Bayes.

A definição do teorema deriva-se da definição probabilísticas condicional, que estabelece a probabilidade de um evento A de ocorrer baseado em uma ocorrência prévia de um evento B, como mostrado na Eq. 1.

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)} \quad (1)$$

Aprendizado de Máquina

Naïve Baye(NB)



$P(A|B)$, é a probabilidade condicional de B ocorrer sabendo que A já ocorreu.

$P(A)$, é a probabilidade do evento A de acontecer.

$P(B|A)$, é a probabilidade condicional de A ocorrer sabendo que B já ocorreu.

$P(B)$, é a probabilidade de evento B de ocorrer.

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)} \quad (1)$$

Aprendizado de Máquina

Naïve Baye(NB)



Esta teoria é uma abordagem estatística fundamental na qual a ideia por trás é que, se a classe for conhecida, os valores dos outros recursos podem ser previstos. No caso em que a classe não é conhecida, a regra de Bayes pode ser usada para prever o rótulo da classe de acordo com os valores de recurso fornecidos.

O algoritmo inicializa as probabilidades para as variáveis de resultado e as ajusta em cada interação baseada no que aconteceu com as outras variáveis do conjunto de dados.

Em classificadores bayesianos, os modelos probabilísticos dos recursos são criados para prever o rótulo de classe de uma nova amostra. Eles são um dos métodos mais utilizados para problemas de reconhecimento de padrões.

Aprendizado de Máquina

k-Means

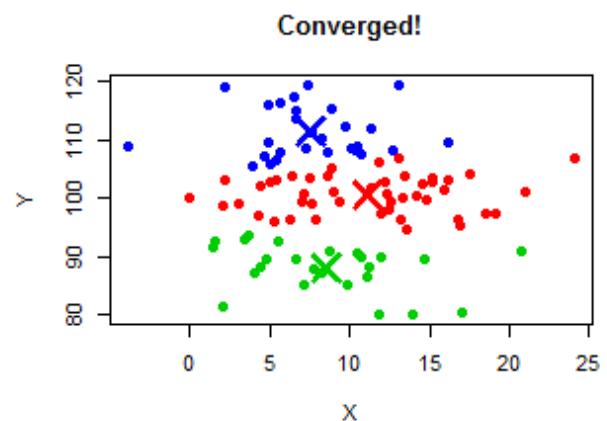
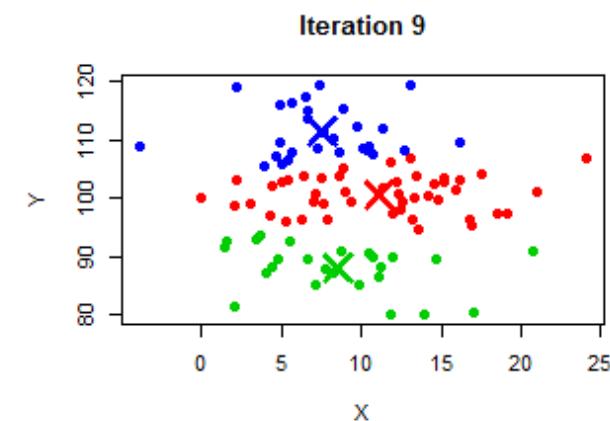
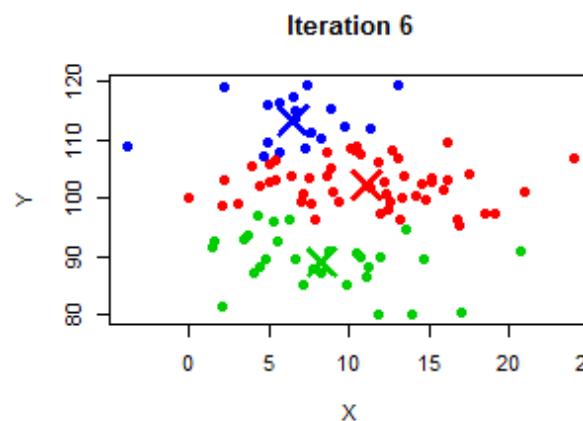
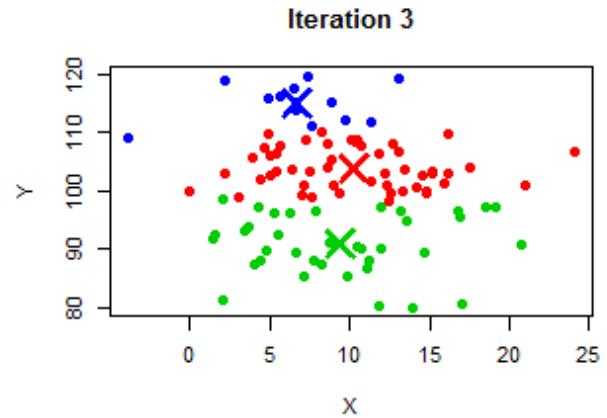
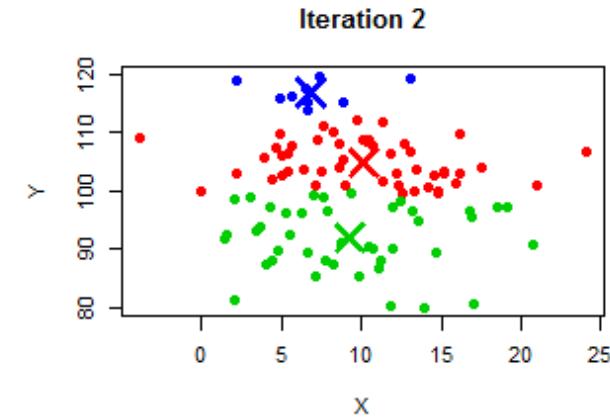
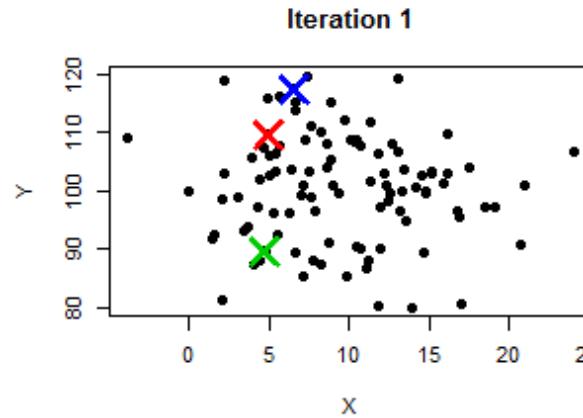


K-Means é um clusterizador que divide n-instâncias do conjunto de dados em k-clusters. A associação de cada dado é feita de acordo com a distância mais próxima do centro de cada cluster. Portanto, cada cluster agrupa os conjuntos de dados mais próximos.

No final, o K-Means divide todo o conjunto de dados usados em um diagrama de Voronoi. É importante ressaltar que K-Means é **uma técnica não supervisionada**, diferente de todos os outros modelos de aprendizado de máquina apresentados aqui. Por sua natureza não supervisionada, ele não precisa de pontos rotulados.

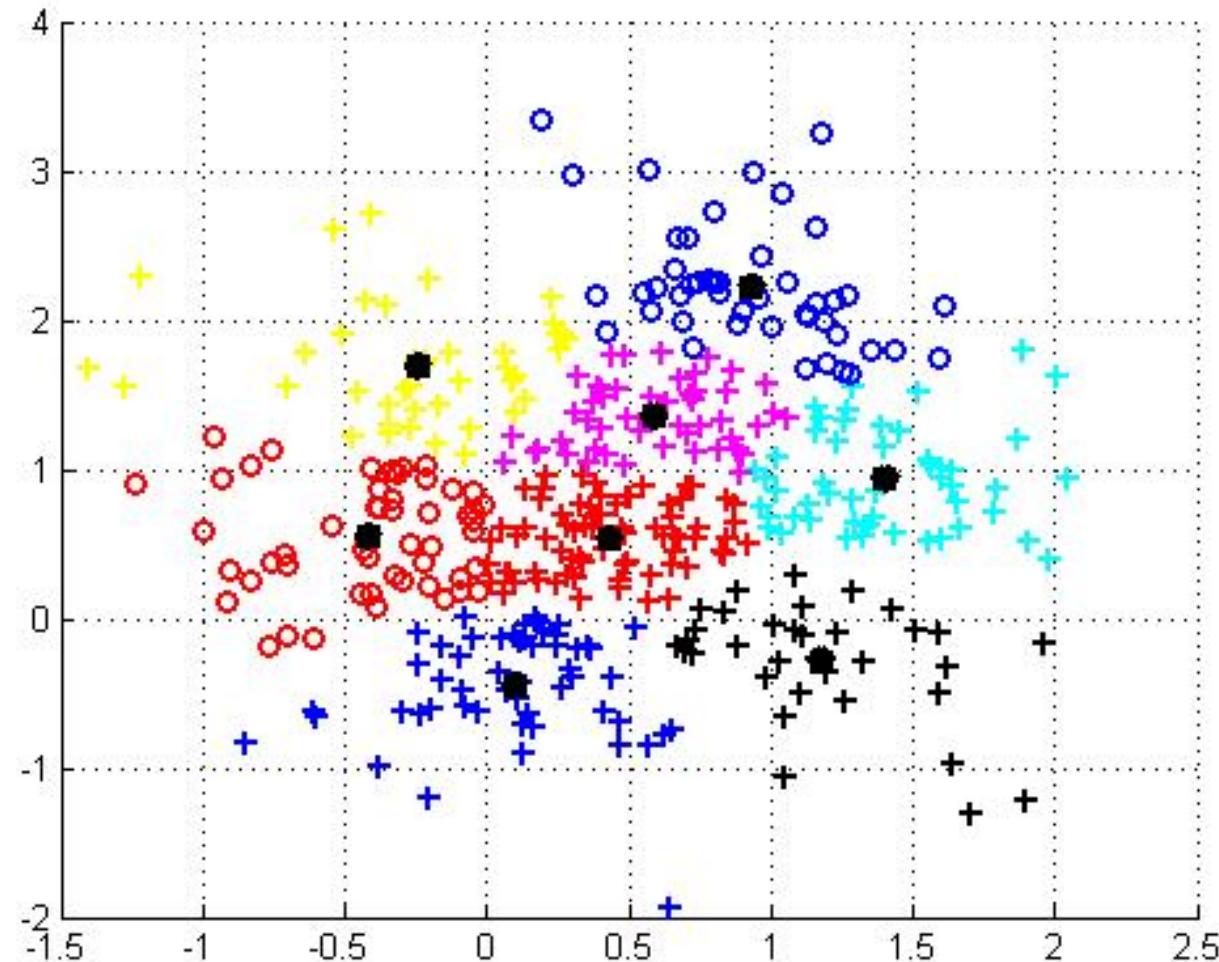
Aprendizado de Máquina

k-Means



Aprendizado de Máquina

k-Means



Aprendizado de Máquina

k-Means



Portanto, K-Means é ótimo para ser usado em cenários onde não sabemos muito a respeito do conjunto de dados processado, sendo assim muito útil para utilizá-lo inicialmente em grandes datasets como demografia da população, tendências nas redes sociais, detecção de anomalia, entre outros.

Aprendizado de Máquina

Decision Tree(DT)



A DT tornou-se um popular método de classificação de aprendizado de máquina devido à sua versatilidade para aplicações em muitos problemas, desde a identificação de objetos até diagnósticos médicos como análise de digitais, íris ou até ECG.

O conceito é usar a estrutura em árvore para dividir as features em classes diferentes com base em critérios probabilísticos e em limites numéricos. Features ou atributos definem a classe.

Aprendizado de Máquina

Decision Tree(DT)



As estruturas de DT são chamadas de árvores de classificação ou regressão. Enquanto as folhas das árvores de classificação representam rótulos de classe, as folhas das árvores de regressão representam valores contínuos.

Existem muitos algoritmos para desenvolver um DT, como ID3(Iterative Dichotomiser 3), C4.5(alternativa para ID3), Cart (Árvore de Classificação e Regressão) e Chaid (Chi-Squared Automatic Interaction Detector).

DT tem a vantagem de ser fácil de implementar e interpretar quando comparado a outros métodos de classificação. Entretanto, dependendo das features escolhidas e da forma como os dados são divididos na árvore, o modelo pode perder a capacidade de generalização devido ao overfitting.

Aprendizado de Máquina

Decision Tree(DT)

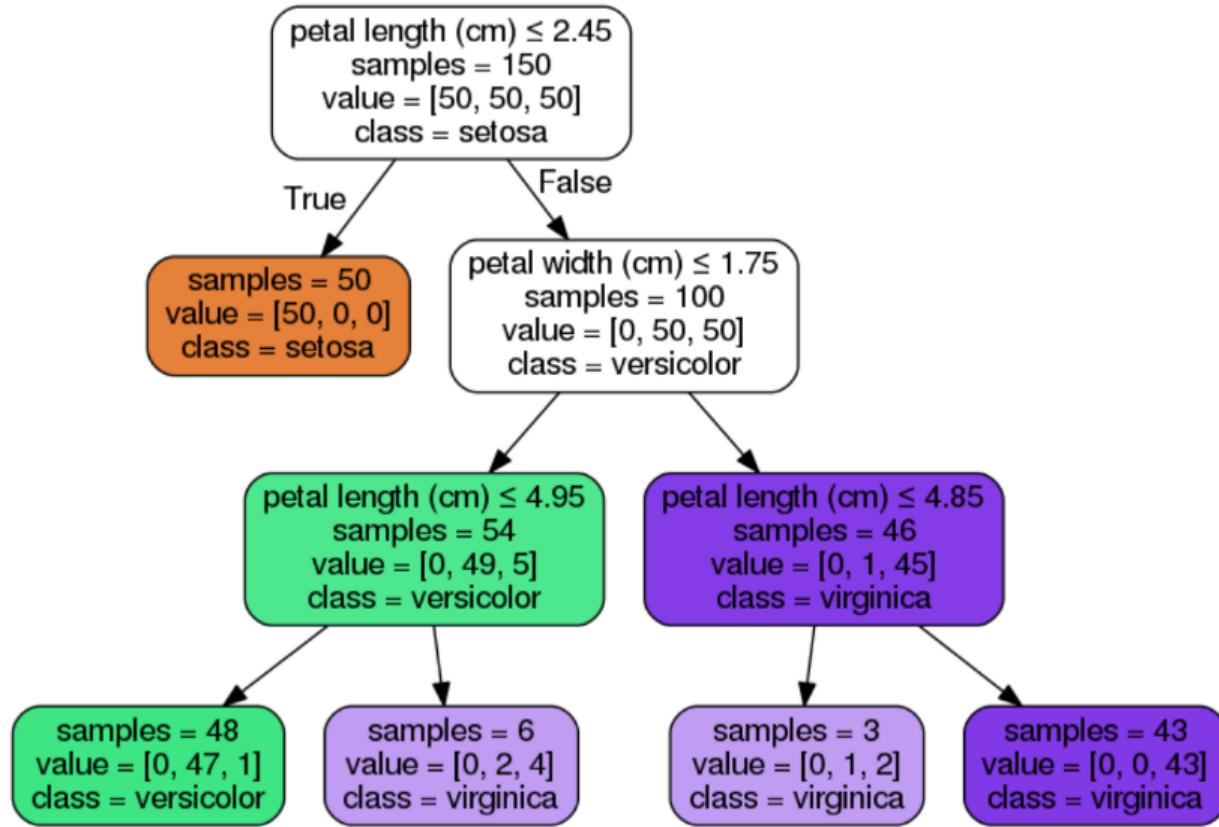
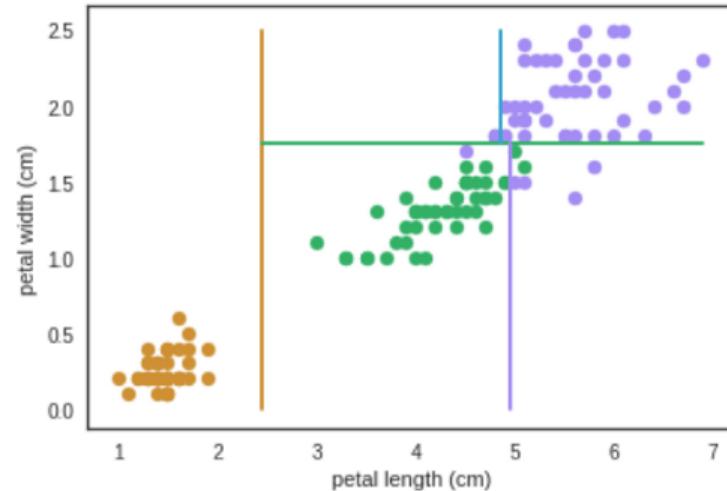


Há duas maneiras comuns de lidar com o overfitting: limitar o número de divisões ou permitir a divisão apenas se houver um número mínimo de pontos de dados na ramificação da árvore. Um aspecto essencial da configuração do DT é como definir a importância de cada feature para maximizar os resultados da classificação. O algoritmo pega a característica que melhor representa a classe e a posiciona na raiz.

Existem alguns índices propostos na literatura para identificar as características mais relevantes: Coeficiente de Gini e índice de Entropia. o Coeficiente de Gini é uma medida da importância da variável para o conjunto de dados e o índice de entropia é uma medida de incerteza associada com a variável.

Aprendizado de Máquina

Decision Tree(DT)



Aprendizado de Máquina

Decision Tree(DT)



Além das abordagens de árvores de decisão comuns, existem algumas estruturas de árvores de decisão mais específicas que são usadas frequentemente para classificação de ECG.

Uma abordagem mais complexa é a utilização da floresta aleatória, onde várias árvores de decisão são treinadas com subconjuntos de dados e será explicada a seguir.

Aprendizado de Máquina

Random Forest(RF)



Um conjunto de árvores de decisão (DT) pode receber a nomeação de RF, o qual treina cada DT com dados selecionados aleatoriamente.

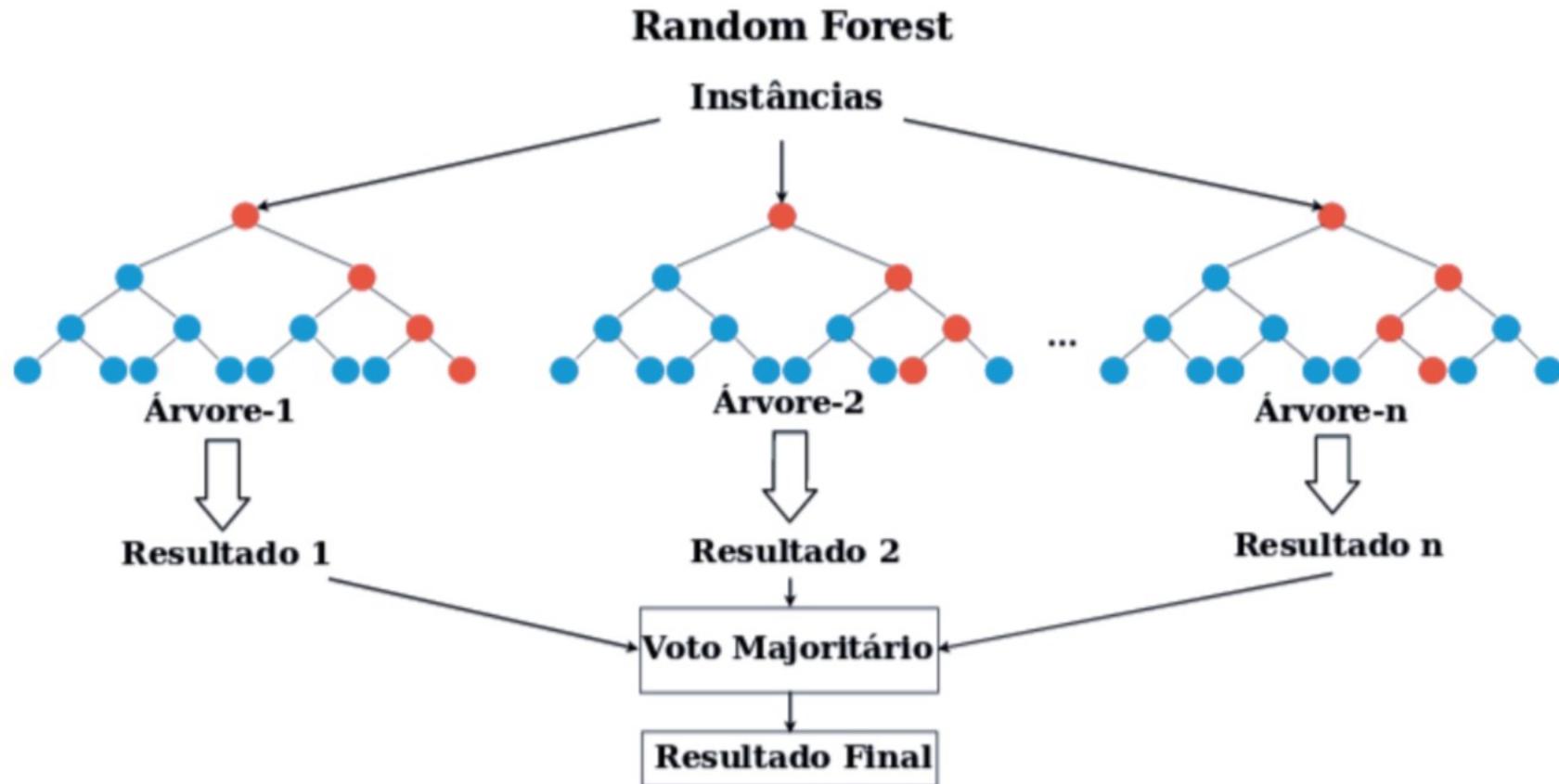
Essa metodologia garante que cada árvore seja ligeiramente diferente uma da outra. Assim, cada árvore pode retornar um resultado distinto para um conjunto de dados. O algoritmo de RF classifica os dados com base em um sistema de votação envolvendo os resultados de árvores individuais

O sistema de votação pode calcular o voto direto ou ponderado. Especificamente, o voto direto conta quantas árvores classificaram uma determinada feature sob uma classe específica.

A votação ponderada retorna a proporção de elementos pertencentes a uma determinada classe.

Aprendizado de Máquina

Random Forest(RF)



Aprendizado de Máquina

Random Forest(RF)



O RF executa melhor que DT em dois aspectos críticos: detecção de anomalia e overfitting. Devido ao processo de treinamento, os outliers estarão presentes em algumas das árvores, mas não em todas elas.

Assim, o sistema de votação garante que os resultados anômalos sejam isolados.

O sistema de votação também minimiza o efeito do overfitting em relação ao DT individual.

Aprendizado de Máquina

Random Forest(RF)



No entanto, tanto RF quanto DT tem problemas para extrapolar dados. Especificamente, os valores de atributo no conjunto de validação devem estar dentro dos limites de valor do conjunto de treinamento.

Os atributos não treinados ou fora do limite levam a resultados imprevisíveis quando incluídos no conjunto de validação.

Aprendizado de Máquina

Random Forest(RF)



O algoritmo de RF aceita que o número de árvores cresça como um parâmetro configurável.

Não há um melhor valor e o limite deve ser a capacidade de armazenamento para salvar a DT. No entanto, um número maior de árvores de decisão não reflete necessariamente nos resultados da classificação.

Uma abordagem é começar com poucas árvores e aumentar gradualmente seu número até que os benefícios não valham os aumentos.

Estudo de Caso

Parte Prática



- Parte Prática



www.healthsenseproject.net
healthsense.project@ufpr.br