

CENTRO UNIVERSITÁRIO FEI

GABRIEL P. DE SOUZA SPINARDI

KELVIN MOZART DE PAULA

LUCAS EDUARDO DA SILVA BISPO

VICTOR DE MORAES MILANI

**METODOLOGIA PARA PREVENÇÃO E PREDIÇÃO DE QUEDAS DE PACIENTES
INTERNADOS EM AMBIENTES HOSPITALARES**

São Bernardo do Campo

2021

GABRIEL P. DE SOUZA SPINARDI
KELVIN MOZART DE PAULA
LUCAS EDUARDO DA SILVA BISPO
VICTOR DE MORAES MILANI

**METODOLOGIA PARA PREVENÇÃO E PREDIÇÃO DE QUEDAS DE PACIENTES
INTERNADOS EM AMBIENTES HOSPITALARES**

Trabalho de Conclusão de Curso apresentado ao Centro Universitário FEI, como parte dos requisitos necessários para obtenção do título de Bacharel em Ciência da Computação. Orientado pelo Prof. Dr. Paulo Sérgio Silva Rodrigues.

São Bernardo do Campo

2021

Às nossas famílias, cujo suporte e ajuda guiaram nossa trajetória. Ao orientador, à banca avaliadora e à universidade.

RESUMO

A monitoração de pacientes em estado de internação em ambientes hospitalares é fundamental para que o paciente, seja da rede pública ou privada, tenha o atendimento básico necessário para sua rápida recuperação. Muitas vezes, esta monitoração é realizada presencialmente por diversos membros da equipe médica de um hospital, onerando esses profissionais que poderiam estar realizando outras tarefas necessariamente presenciais, como realização de exames em massa ou até mesmo participando de campanhas de vacinação, utilizando como exemplo o cenário atual de pandemia que vivenciamos. Em cenários como este, o uso de tecnologias baratas e de fácil programação, como câmeras, modelo bayesiano para probabilidades e esqueletização para o corpo humano, podem ser atrativos para modernizar a estrutura de monitoração atualmente presente em hospitais e clínicas de reabilitação física, cruzando informações médicas já conhecidas com análise de cenas utilizando visão computacional, de modo a não apenas possibilitar um monitoramento de ações básicas do paciente, mas também realizar uma verificação de seus movimentos para definir um perfil comum e prevenir quedas nestes ambientes. Neste trabalho foi proposto um novo método capaz de entender um padrão de possíveis quedas, utilizando modelo bayesiano na probabilidade de uma queda, analisando as informações disponíveis pelo hospital dos seus pacientes e analisando movimentos através das posturas obtendo 17 pontos do esqueleto humano. Como resultado, foi possível estimar a probabilidade de risco de queda de pacientes que já possuíam algum histórico de queda, tendo casos onde o nível de assertividade de predição pode chegar até 90,7% conforme o movimento que está sendo realizado. Foram também gerados dados gráficos importantes capazes de identificar o movimento que está sendo realizado, utilizando ferramentas de detecção de pontos-chave semânticos do corpo humano capazes de obter uma precisão de até 83,0% no nível de detecção do corpo humano.

Palavras-chave: detecção de queda, monitoramento hospitalar, predição de queda, prevenção de queda, aprendizado de máquina, visão computacional.

LISTA DE ILUSTRAÇÕES

Ilustração 1 – Na Figura acima, tem-se uma imagem RGB (Red – Green – Blue) sendo separada por seus três planos coloridos – Vermelho, Verde e Azul. Existem vários desses espaços de cores nos quais existem imagens – Escala de cinza, RGB, HSV, CMYK, etc.(Fonte: Nielsen e Michael (2019))	21
Ilustração 2 – Entrada de 28x28 dimensões com receptive field de área 5x5.(Fonte: Nielsen e Michael (2019))	21
Ilustração 3 – Exemplo de captura de imagens do dataset. (Fonte: Auvinet et al. (2010))	23
Ilustração 4 – Diagrama de relacionamento entre as técnicas de Aprendizado de Máquina, Inteligência Artificial e Visão Computacional. (Fonte: Brownlee (2019))	25
Ilustração 5 – Cena analisada utilizando segmentação semântica. (Fonte: Le (2018)) .	28
Ilustração 6 – Exemplo de arquitetura de camadas FCN. (Fonte: Le (2018))	29
Ilustração 7 – Cena analisada utilizando segmentação de instância. (Fonte: Le (2018))	30
Ilustração 8 – Cena analisada utilizando segmentação de instância. (Fonte: Le (2018))	31
Ilustração 9 – Uso do RoIAlign, com interpolação bilinear. (Fonte: He et al. (2018)) .	31
Ilustração 10 –Exemplos de retângulos delimitadores em segmentação de instância, utilizando o Mask R-CNN. (Fonte: Le (2018))	32
Ilustração 11 –Configuração de Hardware do Kinect. (Fonte: Han et al. (2013)) . . .	33
Ilustração 12 –Esqueletização Kinect. (Fonte: Han et al. (2013))	33
Ilustração 13 –Imagem processada utilizando <i>OpenPose</i> . (Fonte: Chen et al. (2020))	34
Ilustração 14 –Exemplo de estágio de treinamento e classificação. (Traduzido da : Htike, Lai e Khalida (2011))	35
Ilustração 15 –Estágio de aplicação do sistema. (Traduzido da Fonte: Htike, Lai e Khalida (2011))	36
Ilustração 16 –Fluxograma Geral da Metodologia.	38
Ilustração 17 –Teste com <i>dataset</i> utilizando <i>openpose</i> (Fonte: autores)	41
Ilustração 18 –Posturas humanas - Sentando (Fonte: autores)	42
Ilustração 19 –Posturas humanas - Levantando (Fonte: autores)	43
Ilustração 20 –Posturas humanas - Levantado (Fonte: autores)	43

Ilustração 21 –Gráfico de movimentação em geral. Eixo <i>x</i> : <i>frames</i> , eixo <i>y</i> : <i>keypoints</i> . (Fonte: autores)	51
Ilustração 22 –Gráfico de movimentação - Sentando. Eixo <i>x</i> : <i>frames</i> , eixo <i>y</i> : <i>keypoints</i> . (Fonte: autores)	52
Ilustração 23 –Gráfico de movimentação - Levantando. Eixo <i>x</i> : <i>frames</i> , eixo <i>y</i> : <i>keypoints</i> . (Fonte: autores)	52
Ilustração 24 –Gráfico de movimentação - Em pé. Eixo <i>x</i> : <i>frames</i> , eixo <i>y</i> : <i>keypoints</i> . (Fonte: autores)	53

LISTA DE TABELAS

Tabela 1 – Tabela dos <i>KeyPoints</i> do esqueleto humano. (Fonte: Chen et al. (2020))	44
Tabela 2 – Informações referentes à queda. (Fonte: Souza et al. (2019))	46
Tabela 3 – Probabilidades com maiores e menores valores de risco de quedas. (Fonte: autores)	47
Tabela 4 – Médias com o nível alto de risco de queda. (Fonte: autores)	47
Tabela 5 – Médias com o nível baixo de risco de queda. (Fonte: autores)	47
Tabela 6 – Valores com variação de turno e dia. (Fonte: autores)	48
Tabela 7 – Valores com variação de ambiente. (Fonte: autores)	48
Tabela 8 – Valores com variação de protocolo e medicação. (Fonte: autores)	48
Tabela 9 – Probabilidades de queda analisando informações dos pacientes. (Fonte: autores)	48
Tabela 10 – Probabilidades de queda analisando informações médicas. (Fonte: autores)	49
Tabela 11 – Probabilidades de queda analisando informações dos ambientes hospitalares. (Fonte: autores)	49
Tabela 12 – Tabela de <i>keypoints</i> do esqueleto humano. (Fonte: autores)	50
Tabela 13 – Detecção de movimento. (Fonte: autores)	51
Tabela 14 – Menores valores dos dados com os movimentos entre as posições e probabilidade de queda. (Fonte: autores)	54
Tabela 15 – Média dos pacientes (sem severidade). (Fonte: autores)	54
Tabela 16 – Média das informações médicas. (Fonte: autores)	55
Tabela 17 – Média das informações hospitalares. (Fonte: autores)	55

SUMÁRIO

1	INTRODUÇÃO	9
1.1	Objetivo	9
1.2	Estrutura do Trabalho	9
2	TRABALHOS RELACIONADOS	11
2.1	Detecção de Queda	11
2.2	Reconhecimento de Postura	13
2.3	Reabilitação	14
2.4	Sensores Vestíveis	16
3	CONCEITOS FUNDAMENTAIS	19
3.1	Modelo Bayesiano	19
3.2	Rede neural Convolucional	20
3.3	Base de Dados	22
3.3.1	<i>Dataset de detecção de queda</i>	22
3.3.2	Informações médicas do Hospital Moinho de Vento	23
3.4	Hardware Disponível	24
3.4.1	Computadores	24
3.4.2	Câmera	24
3.5	Visão Computacional e Processamento de Imagens	25
3.5.1	Técnicas e aplicação de Visão Computacional	26
3.5.2	Classificação da imagem	27
3.5.2.1	<i>Segmentação semântica</i>	27
3.5.2.2	<i>Segmentação de instância</i>	29
3.6	Esqueletização de Imagens	32
3.6.1	<i>OpenPose</i>	33
3.7	Reconhecimento de Postura	34
3.8	Reconhecimento de Ações	36
4	METODOLOGIA	38
4.1	Base de dados	39
4.2	Informações dos pacientes	39
4.3	Informações Médicas	39
4.4	Informações Hospitalares	39

4.5	Câmera	40
4.6	Extração de Características Visuais	40
4.7	Composição das informações	40
4.8	Interface Médica	40
5	PROPOSTA EXPERIMENTAL	41
5.1	Aquisição de informações dos pacientes, médicas e hospitalares	41
5.2	Experimento de extração de características visuais	41
5.3	Experimento de reconhecimento de movimento	42
5.3.1	<i>KeyPoints</i>	43
5.4	Composição de informações	45
6	RESULTADOS E DISCUSSÃO	46
6.1	Informações do paciente, médicas e hospitalares	46
6.2	Teste de reconhecimento de movimento	49
6.2.1	Movimentação entre as posições	50
6.3	Composição das informações	53
7	Conclusão	56
	REFERÊNCIAS	57

1 INTRODUÇÃO

A queda de pacientes no ambiente hospitalar está entre os principais e mais difíceis eventos a serem evitados atualmente, conforme (IBSP (2019)). As consequências vão desde danos físicos graves, como fraturas, a psicológicos, que derrubam a confiança do paciente e atrasam a recuperação na sua totalidade. Desta forma, a prevenção de queda tem sido enfatizada em sistemas de saúde internacionais como um indicador importante de segurança do paciente, conforme divulgado no Instituto Brasileiro de Segurança do Paciente (IBSP (2019)).

Um dos maiores desafios para monitoração de queda é a vigilância que depende fortemente do ambiente que o paciente está localizado. Esse monitoramento pode ser efetuado através de câmeras sensoriais. No entanto, ainda é um desafio a generalização desse problema visto que depende da iluminação local, interpretação da cena, ruídos visuais, oclusão, entre outros fatores.

Neste trabalho de conclusão de curso foram realizados testes utilizando dados médicos hospitalares coletados do Hospital Moinho de Vento para calcular um modelo bayesiano e gerar um padrão de probabilidades de queda e, com a extração de características visuais utilizando *keypoints*, foi possível realizar testes e simulações de movimentos e posturas, como sentado e levantado, o que proporcionou maior acurácia para se determinar a probabilidade de uma queda através do movimento do paciente.

Os resultados obtidos com a metodologia proposta mostram que o método proposto é promissor, visto que contribui com o estado-da-arte em predição de risco de quedas em ambientes hospitalares.

1.1 OBJETIVO

Este projeto propôs uma metodologia que combina informações hospitalares de diversos níveis, como aquelas fornecidas tanto pela base de dados hospitalar quanto por câmeras para a previsibilidade de risco de queda em ambientes médicos durante a internação do paciente.

1.2 ESTRUTURA DO TRABALHO

O restante deste trabalho foi dividido da seguinte maneira:

No Capítulo 2, foram descritos trabalhos relacionados disponíveis na literatura, para apresentar o cenário atual de pesquisa da área.

No Capítulo 3, foram apresentados todos os conceitos utilizados e relacionados ao tema abordado, para que o leitor pudesse entender com clareza as técnicas tratadas no trabalho e compreender os termos descritos posteriormente.

O Capítulo 4 detalhou a metodologia utilizada para o desenvolvimento deste trabalho, demonstrando as técnicas utilizadas e os passos realizados para atingir o objetivo final.

No capítulo 5 foi definido o procedimento experimental de implantação da metodologia proposta ao longo deste trabalho.

No capítulo 6 foram apresentados os resultados dos testes e discussões a respeito dos números alcançados.

No capítulo 7 houve a apresentação da conclusão deste trabalho, reafirmando os principais resultados obtidos.

2 TRABALHOS RELACIONADOS

A estratégia apresentada no levantamento bibliográfico desse projeto contempla duas linhas principais. A primeira explora os trabalhos que envolvem ferramentas, métricas e base de dados que possam dar suporte à metodologia proposta; e o segundo avalia trabalhos baseados em fisioterapia relacionadas ao posicionamento de pacientes. Esses trabalhos estão divididos a seguir em seções específicas.

2.1 DETECÇÃO DE QUEDA

No trabalho de Tsinganos e Skodras (2017) foi apresentada uma proposta de desenvolver um detector de quedas para idosos usando um celular. Tal sistema pode diferenciar as quedas das atividades do dia-a-dia. Essa solução é baseada no trabalho de Abbate et al. (2012), incorporando uma implementação personalizado do sistema, alcançando uma precisão na detecção da queda de 97.53% na sensibilidade e 94.89% de especificidade, sendo compatível com trabalhos similares. O estudo apresenta uma solução para identificar sinais e extrair parâmetros classificando-os como uma possível queda ou uma atividade rotineira do dia. Também é incorporado um método no sistema para o gerenciamento de energia do celular. Finalmente, o sistema final é montado usando todas as informações nos dados. A partir dessas informações, o sistema FSM com os parâmetros escolhidos conseguiu detectar 932 possibilidades de quedas, 637 dessas sendo quedas reais. Essas possíveis quedas usaram as seguintes métricas: sensibilidade, especificidade, precisão. Foram capturadas usando o classificador KNN e o dataset MobiAct.

No trabalho de Abobakr, Hossny e Nahavandi (2018) foi apresentada uma proposta de detecção de quedas usando um novo método de reconhecimento de postura sem esqueleto e reconhecimento de atividades. A solução proposta faz o reconhecimento da queda e postura através de três pontos: aumentar as condições de luz e atmosfera na imagem e suas oclusões. Reduzir a complexidade para garantir um melhor tempo de resposta com os aparelhos; garantir medidas de privacidade para integrar o sistema em diversos lugares, ate mesmo banheiros, sendo um dos maiores lugares aonde ocorre uma queda. Finalmente, os resultados apresentados possuem os seguintes métodos propostos: primeiro, a segmentação dos frames são passados por diversos módulos; segundo, um RDF é usado para identificar a postura atual no quadro. O RDF é treinado e avaliado usando conjuntos de dados sintéticos para superar a limitação de obten-

ção de conjuntos de dados de detecção de queda variável; terceiro, um modelo de classificação SVM é usado para analisar a mudança da confiança da postura deitada temporalmente e indicar a ocorrência de eventos de queda. Os resultados obtidos tiveram uma acurácia de 98% com uma resolução de 80 x 60 e 100% com uma resolução de 640 x 480.

No trabalho de Rhuma, Yu e Chambers (2013) é apresentado um sistema de detecção de queda para idosos em um ambiente doméstico. Esse sistema usa dados reais para alcançar uma boa detecção de queda. Os métodos técnicos para conseguir essa detecção são: métodos baseados em acelerômetros, som ou sensor de vibração e visão computacional. Esse trabalho propõem um novo método de detecção baseado em classificação de dados de postura, assim como uma versão melhorada do *Directed Acyclic Graph Support Vector Machine* (DAGSVM) foi adotada para conseguir avançar com o desempenho de classificação. Os resultados obtidos 96.88% das quedas foram detectadas, e apenas 4.7% das atividades simuladas que não tiveram quedas forma confundidas como quedas. Esses resultados finais mostram que o sistema proposto funciona em um ambiente domestico real.

No trabalho de Gjoreski, Lutrek e Gams (2011) é apresentada uma proposta de um detector de quedas usando um acelerômetro que pode reconhecer posturas que podem resultar em quedas. É proposto o uso de diversos sensores no ambiente assim como também sensores que o paciente possa vestir para determinar qual é a melhor solução. Foi possível reconhecer as posturas e classifica-las em dois grupos: deitado e de quadro; em pé, sentado e sentado no chão. O experimento provou que acelerômetros são melhores que os sensores localizados no ambiente. E em quedas um acelerômetro na cintura provou ter um desempenho similar a de um, acelerômetro vestido no peito. As detecções de quedas foram melhores reconhecidas com sensores no ambiente apenas por já saberem antemão a localização de uma cama ou cadeira.

No trabalho de Kianoush et al. (2016) foi apresentado um sistema livre de aparelhos (sem necessidade de celular, eletrônicos, etc junto à pessoa) para evitar quedas e lesões em ambiente industrial. Para isso, foi criada uma rede wireless com câmeras e sensores em ambientes de trabalho. Assim, o objetivo é, a partir da movimentação de uma pessoa que gera interferência na banda wireless, detectar risco de queda e lesões. Como método experimental, utilizou-se sensores em máquinas e câmeras num ambiente industrial, junto com o Modelo Oculito de Markov para realizar os cálculos necessários. Os resultados são positivos e confirmam a vantagem na prevenção de quedas e acidentes, conforme proposto neste sistema.

No trabalho de Daher et al. (2016) foi proposta uma forma de detecção de queda através de sensores de força e acelerômetros triaxiais em apartamentos de idosos. A proposta se baseia em um resultado a partir da fusão entre os sensores de força e as decisões do acelerômetro

triaxial do ladrilho inteligente. O método abordado é eficiente, podendo facilmente ser utilizado em apartamentos de idosos.

No trabalho de Zerrouki e Houacine (2017) foi proposta uma abordagem inovadora para a detecção confiável de incidentes de queda com base na variação da forma da silhueta humana. O método identifica posturas utilizando SVM e adapta um Modelo Oculto de Markov para classificar sequências de vídeos em eventos de queda e não queda. Os resultados obtidos através dos cálculos, médias e análises de imagens demonstram a eficácia da proposta quando comparada a algumas abordagens de última geração.

2.2 RECONHECIMENTO DE POSTURA

No trabalho de Mehta et al. (2017) é apresentado o primeiro método para capturar uma postura em 3D em tempo real, usando uma única câmera RGB. Esse método combina uma CNN baseado em regressão de posturas com esqueleto cinemático. O método proposto obteve os seguintes resultados para previsões de 82.5 e 80.5, utilizando as CNN ResNet 100 e ResNet 50 no dataset Human3.6m. As ações testadas foram conversar, sentar, andar, levantar, comer, dentre outras atividades.

No trabalho de Chen e Ramanan (2017) é apresentada uma estimativa de postura 3D do corpo, com base em postura em 2D. Foram utilizadas imagens baseadas em CNN prevendo pontos chaves em uma imagem 2D. A simplicidade do método e eficiência, combinado com o desempenho e benchmark estado da arte, sugere que essas linhas de base simples devem ser usadas para benchmarking futuros na estimativa de postura 3D. Uma vantagem notável das representações 2D intermediárias é o treinamento modular - conjuntos de dados 2D (que são normalmente maiores e mais diversos devido à facilidade de anotação) e que podem ser usados para treinar o módulo de processamento de imagem inicial, enquanto os dados de captura de movimento 3D podem ser usados para treinar o módulo de raciocínio 3D subsequente.

No trabalho de Tran e Trivedi (2012) foi apresentado um sistema que reconhece gestos humanos utilizando diversas câmeras. Assim, foi proposta uma solução de um sistema para interação humana que reconheça esses gestos, focando apenas na parte superior do tronco até a cabeça, considerando que os gestos da cintura para cima possuem mais relevância em interações humanas. Os autores mostraram que o reconhecimento tem bons resultados com vários sujeitos em diferentes ambientes. Com relação ao reconhecimento de gestos com base nas dinâmicas dos ângulos das articulações, o trabalho foi avaliado com 5 sujeitos fazendo seis gestos com

a parte de cima do corpo, alcançando uma acurácia acima de 90% indicando uma promessa e viabilidade do sistema.

2.3 REABILITAÇÃO

No trabalho de Chiarovano et al. (2015) foi apresentada uma proposta para o estudo de uma análise biométrica de artigos acerca de reabilitação de pacientes usando tecnologia de realidade virtual. Os autores apresentaram uma metodologia em três passos. Inicialmente, fizeram uma análise nas bases de dados de artigos da PubMed, Scopus e Web of Science. Em seguida, para analisar, foram considerados tópicos como: resultados, locais, instituições, autores, principais periódicos e citações. Finalmente, os resultados foram apresentados em formas de textos, tabelas e gráficos. Como resultados, os autores mostraram que a realidade virtual é uma das tecnologias mais impactantes e que vem crescendo nos últimos anos.

No trabalho de Anton et al. (2018) os resultados obtidos até o momento com o KiReS (Sistema de Tele-Reabilitação) mostram sua adequação para tele-reabilitação e uma experiência de usuário de boa qualidade. Os pacientes que o utilizaram descobriram que a interação com ele foi amigável, consideraram-o um complemento à sua terapia que pode melhorar o atendimento médico; além de mostrarem uma predisposição para o uso do sistema novamente. O KiReS pode ser estendido e, como trabalho futuro, os desenvolvedores planejam aprimorar as informações que o KiReS recupera adicionando dispositivos de rastreamento de bio-sinais. Assim, seria possível estender os recursos de raciocínio e análise de dados do sistema com essas novas entradas.

No trabalho de Yang et al. (2018) foi apresentado um sistema IoT de reabilitação de AVC baseado em uma braçadeira inteligente, algoritmos de machine learning e um braço destros impresso em 3D. Um estudo comparativo foi feito com 3 algoritmos, com o objetivo de selecionar o algoritmo com a melhor classificação e o menor tempo de treino. Como resultado demonstrou-se que o braço robô consegue imitar os gestos do usuário em tempo real, mostrando que o sistema proposto pode ser usado como uma ferramenta de treinamento para facilitar o processo de reabilitação para pacientes depois de um AVC.

No trabalho de Piraintorn e Sa-ing (2020) foi proposto um novo sistema de orientação que auxilia na reabilitação do paciente através do fisioterapeuta e do médico. O sistema de interação de inteligência é proposto para detecção e monitoramento da reabilitação do paciente com AVC acamado. Este sistema detecta um paciente com AVC usando uma câmera 3D, que é a Intel Realsense D415, para colocar no final do leito do paciente para extraí-lo da cama medindo

a distância entre ele e a cama. O resultado representa o monitoramento e avaliação eficazes da reabilitação do AVC para que o programa possa executar com precisão uma terapia de gestos. Portanto, o sistema inteligente de interação pode ajudar o fisioterapeuta a monitorar e avaliar a reabilitação do AVC na cama.

No trabalho de Vourganas, Stankovic e Stankovic (2020) é apresentada uma abordagem individual do paciente com sistema de apoio para uma reabilitação em casa. Para tal, foi apresentado um método para gerar conjuntos de dados sintéticos, bem como um algoritmo de machine learning híbrido, combinando aprendizagem de conjunto e empilhamento híbrido usando árvores de decisão com gradiente extremo. Os resultados mostram uma melhora de 5% e 15% para os testes FTSTS e TUG, respectivamente, em relação às abordagens anteriores que usam meios intrusivos de monitoramento, como câmeras.

No trabalho de Lee et al. (2020) foi apresentado um modelo híbrido que integra machine learning (ML), Redes Bayesianas (RB), e inteligência artificial (AI) para avaliação quantitativa de exercícios de reabilitação de AVC. Esse modelo usa uma técnica de conjunto de média ponderada para avaliar a qualidade do movimento. Os resultados indicam um potencial para o modelo híbrido com a técnica de IA explicável para apoiar a interpretação de um modelo e ajustar com regras específicas do usuário para personalização.

No trabalho de Fidalgo-Herrera et al. (2020) foi apresentada uma forma de avaliar o desempenho de uma ANN usando variáveis cinemáticas para prever a mudança da amplitude de movimento após um período da reabilitação de um paciente com whiplash associated disorder (WAD), que é um conjunto de sintomas que afetam o pescoço que podem ser desencadeados por acidente em que a pessoa sofra uma desaceleração brusca no pescoço. Um total de 1082 pacientes foram testados sendo 55.1% mulheres e 44.9% homens. Esses pacientes foram classificados do início da avaliação conforme a *Neck Functional Holistic Analysis Score* (NFHAS) uma forma de pontuar a análise holística funcional do pescoço, NFHAS: 3.69% NFHAS tipo 1, 16.82% NFHAS tipo 2, 24.67% NFHAS tipo 3, 31.23% NFHAS tipo 4 e 23.56% NFHAS tipo 5. A gravidade da lesão aumenta segundo o tipo NFHAS classificado.

No trabalho de Liu et al. (2019) é apresentado um tipo de reconhecimento de gestos por meio de um sensor capacitivo de baixo custo em uma atividade de reabilitação. Para o reconhecimento de gestos propostos, uma estrutura de rede neural de memória (C-LSTM) é aplicada e hiperparâmetros são analisados para determinar quais recursos serão necessários para classificar as tarefas. Como resultado, os sensores capacitivos, o processamento de sinal e o classificador de *Machine Learning* combinam-se para realizar uma classificação precisa dos movimentos das mãos.

No trabalho de Esfahlani e Wilson (2017) é apresentada a criação de um jogo chamado RehabGame. O objetivo do jogo é combater a deficiência física de vítimas de acidente vascular cerebral/lesão cerebral, a fim de melhorar o movimento do membro superior. Através do processo de reabilitação em tempo real jogo, o jogador decide os caminhos que podem ser percorridos por seu membro superior, para alcançar objetos virtuais. Assim, é utilizado o Algoritmo de Monte-Carlo Tree Search para desenvolvimento do jogo. Os resultados de duas amostras mostram a eficácia do Monte-Carlo Tree Search no RehabGame por ser capaz de construir um movimento coerente da mão. O paciente progride a partir de caminhos altamente alcançáveis para os menos alcançáveis, configurando e personalizando o processo de reabilitação.

2.4 SENSORES VESTÍVEIS

No trabalho de Al-khafajiy et al. (2019) foi proposto o desenvolvimento de um sistema de monitoramento de saúde inteligente, capaz de observar os idosos à distância e rastrear seus dados fisiológicos para detectar distúrbios específicos por meio do processamento e análise precisa dos dados sensoriais adquiridos, com o objetivo de melhorar o suporte à decisão clínica, enquanto facilita as Práticas de Intervenção Precoce. O sistema consiste em diferentes sensores médicos vestíveis, smartphones, sensores e atuadores sem fio, hardware de computador, redes de computadores, aplicativos de software, que são interconectados para trocar dados e fornecer serviços em um ambiente assistido. A taxa geral de descarte de todos os pacotes enviados é de cerca de 2,2%, sendo essa uma taxa promissora e menos do que o limite aceitável de 15%, uma vez que tudo acima dessa taxa deteriora visivelmente a disponibilidade da rede.

No trabalho de Wang et al. (2019) foi realizada a captura e análise de movimentos na natação usando sensores inerciais vestíveis. O método proposto utiliza um algoritmo de fusão de dados multissensor para estimativa de atitude dos nadadores, e a postura de natação é reconstruída em combinação com um modelo biomecânico humano. Quinze pessoas foram recrutadas para participar do experimento. A maior taxa de amostragem do NDI é 60 Hz, e a taxa de amostragem do sistema é 200 Hz. Assim a interpolação de dados unidimensional foi usada para ajustar os dados do NDI.

No trabalho de AbhishekVaish, SaifAhmed e AnanShetty (2016) a terapia D + R é uma plataforma que permite que os pacientes possam ser monitorados remotamente. Consiste em um aplicativo iPhone vinculado a um site. Os exercícios são prescritos pelo fisioterapeuta através de uma tela chamada quando o aplicativo é iniciado para ser realizado o download. Uma vantagem

distinta desta plataforma é o uso do hardware próprio do paciente, ou seja, o iPhone está sempre com os pacientes.

Na pesquisa de Choe et al. (2017) é explorado o rastreamento semiautomático para apoiar o automonitoramento em três domínios distintos: sono, humor e estresse, e dieta. São utilizadas abordagens manuais e automatizadas para a viabilidade da captura de dados visando diminuir as cargas de captura, coletar dados que normalmente são difíceis de rastrear automaticamente e promover a conscientização para ajudar as pessoas a atingirem seus objetivos de automonitoramento. Os resultados sugerem que uma abordagem semiautomática de baixa carga pode atingir uma visão mais abrangente do estresse, incluindo respostas subjetivas e fisiológicas necessárias, em contraste com uma abordagem totalmente automatizada ou totalmente manual. Por sua vez, a análise nutricional semiautomática pode reduzir as cargas de rastreamento e promover a conscientização necessária para facilitar os efeitos reativos.

No trabalho de Ofli et al. (2012) é proposta uma nova abordagem que relaciona uma ação ao movimento real do corpo ou de suas partes através de uma representação das ações humanas denominada Sequência das Articulações Mais Informativas (SMIJ). A cada instante de tempo, são selecionadas automaticamente algumas articulações do esqueleto humano que são consideradas as mais informativas para realizar uma determinada ação atual. A seleção das articulações é baseada em medidas altamente interpretáveis, como a média ou variância dos ângulos das articulações, velocidade angular máxima das articulações, etc. Em seguida, uma ação é representada como uma sequência dessas articulações mais informativas. É possível observar que usar $N > 1$ para as articulações mais informativas é melhor que usar apenas uma única articulação mais informativa. O melhor desempenho de classificação é obtido para diferentes valores de N para diferentes conjuntos de dados; especificamente, 94,91% quando $N = 6$ para o conjunto de dados 1, 84,40% quando $N = 2$ para o conjunto de dados 2 e 33,33% quando $N = 5, 6$ para o conjunto de dados 3. O resultado do melhor desempenho ao usar um valor intermediário de N não é inesperado. Em geral, à medida que N aumenta, a representação proposta captura cada vez mais informações sobre a ação que está sendo executada. Ao mesmo tempo, o número de parâmetros de classificação aumenta com N , enquanto a quantidade de dados de treinamento permanece a mesma. Portanto, existe o risco de sobreajuste quando N é grande.

No trabalho de Crossley et al. (2011) a proposta é determinar se o monitoramento remoto sem fio com alertas clínicos automáticos reduz o tempo de um evento clínico até uma decisão clínica em resposta a arritmias, progressão de doença cardiovascular (CV) e problemas com o dispositivo em comparação com pacientes que recebem atendimento padrão em consultório. Uma avaliação multicêntrica, prospectiva e randomizada foi realizada envolvendo 1.997

pacientes de 136 centros clínicos que foram observados durante 15 meses após serem submetidos à inserção de um cardioversor desfibrilador implantável (incluindo dispositivos de terapia de ressincronização cardíaca). Os dados de utilização de cuidados de saúde incluíram todas as hospitalizações relacionadas ao CV, visitas ao departamento de emergência e visitas ao consultório clínico. O tempo médio desde o evento clínico até a decisão clínica por paciente foi reduzido de 22 dias em consultório para 4,6 dias em ambiente remoto. Os dados de utilização de cuidados de saúde revelaram uma diminuição no tempo médio de permanência por visita de hospitalização CV de 4,0 dias em consultório para 3,3 dias em ambiente remoto.

3 CONCEITOS FUNDAMENTAIS

Este capítulo aborda conceitos, ferramentas, bases de dados, hardwares e outros trabalhos já conhecidos no meio acadêmico e profissional, que servirão como base fundamental para a realização deste projeto.

3.1 MODELO BAYESIANO

Segundo Schoot et al. (2021), a estatística bayesiana é uma abordagem para análise de dados e estimativa de parâmetros com base no teorema de Bayes. O único requisito para a estatística bayesiana é que todos os parâmetros observados e não observados em um modelo estatístico recebam uma distribuição de probabilidade conjunta, denominada distribuição *a priori* de dados.

O objetivo é tirar uma conclusão do todo, adicionando informações subjetivas ao modelo preditivo e extraíndo parte das informações dos dados coletados. Nesse caso, os pesquisadores admitem que há algum conhecimento de que os dados coletados não conseguem prever e podem ter impacto na previsão.

O fluxo de trabalho Bayesiano típico consiste em três etapas principais: capturar o conhecimento disponível sobre um determinado parâmetro em um modelo estatístico por meio da distribuição anterior, que é normalmente determinada antes da coleta de dados; determinar a função de probabilidade usando as informações sobre os parâmetros disponíveis nos dados observados; e combinar a distribuição anterior e a função de verossimilhança usando o teorema de Bayes na forma da distribuição posterior. A distribuição posterior reflete o conhecimento atualizado, equilibrando o conhecimento anterior com os dados observados, e é usada para conduzir inferências. As inferências Bayesianas são ótimas quando calculada a média dessa distribuição de probabilidade conjunta e a inferência para essas quantidades é baseada em sua distribuição condicional, considerando os dados observados.

O teorema de Bayes (Equação 1) prova, matematicamente, como a informação não contida na amostra (chamada de informação prévia) deve ser incluída no modelo de previsão. Segundo o teorema, tem-se:

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

onde:

- a) $P(A|B)$ é a probabilidade a posteriori;
- b) $P(B|A)$ é a Verossimilhança;
- c) $P(A)$ é a probabilidade a priori;
- d) $P(B)$ é a marginalização.

3.2 REDE NEURAL CONVOLUCIONAL

O termo Deep Learning ou Deep Neural Network se refere a uma rede neural artificial (ANN) com múltiplas camadas. A Rede Neural Convolutiva (CNN) é uma das redes neurais profundas mais populares. Este nome é derivado da operação matemática linear entre matrizes chamada convolução. Uma CNN possui várias camadas, incluindo a camada convolutiva, camada de não linearidade, camada de *pooling* e camada totalmente conectada. A camada convolutiva e a camada totalmente conectada têm parâmetros, mas a camada de *pooling* e a camada de não linearidade não os possuem.

A CNN mostra excelente desempenho em problemas de aprendizado de máquina, principalmente em aplicações que processam dados de imagens, como os maiores conjuntos de dados para classificação de imagens (Image Net), visão computacional e processamento de linguagem natural (PNL).

A ideia principal de uma rede convolutiva é, a cada camada, transformar uma imagem mais complexa e detalhada em imagens mais simples. Geralmente, essas redes são compostas de duas partes principais. A primeira é a chamada camada convolutiva, sendo responsável por reduzir o tamanho e informações das imagens de entrada. A segunda é geralmente uma rede totalmente conectada, utilizada para classificação das imagens produzidas pela camada convolutiva. Quando falamos em reconhecimento/classificação de imagens, as entradas são geralmente uma matriz tridimensional com altura e largura (de acordo com o tamanho da imagem) e profundidade, cuja profundidade depende do número de canais de cores, como visto na Figura 1. De modo geral, a imagem usa três canais de RGB, e o valor de cada *pixel* é o mesmo.

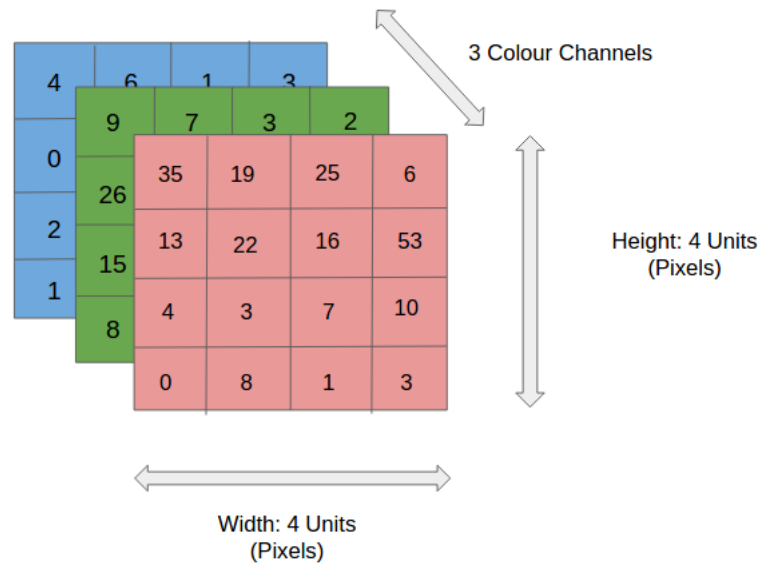


Figura 1 – Na Figura acima, tem-se uma imagem RGB (Red – Green – Blue) sendo separada por seus três planos coloridos – Vermelho, Verde e Azul. Existem vários desses espaços de cores nos quais existem imagens – Escala de cinza, RGB, HSV, CMYK, etc.(Fonte: Nielsen e Michael (2019))

As convoluções funcionam como filtros. Elas podem ver pequenos quadrados e "deslizar"na imagem para capturar os recursos mais marcantes. Usando uma imagem de $32 \times 32 \times 3$ e um filtro que cobre a área 5×5 da imagem com um movimento de 2 saltos (denominado *stride*), o filtro passará por toda a imagem, através de cada canal, e finalmente formará um *feature map* ou *activation map* de $28 \times 28 \times 1$, como visto na Figura 2.

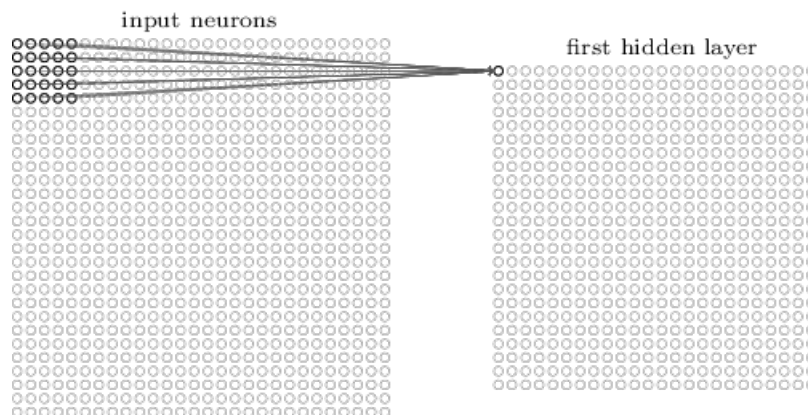


Figura 2 – Entrada de 28x28 dimensões com receptive field de área 5x5.(Fonte: Nielsen e Michael (2019))

A profundidade da saída de uma convolução é igual ao número de filtros aplicados. Quanto mais profunda a camada convolucional, mais detalhados são os traços identificados pelo mapa de ativação. O filtro (também conhecido como kernel) consiste em pesos inicializados aleatoriamente e os pesos, por sua vez, são atualizados com cada nova entrada durante o

processo de retro propagação. A pequena área da entrada, à qual o filtro é aplicado, é chamada de campo de recebimento.

As funções de ativação servem para trazer a não-linearidades ao sistema, para que a rede consiga aprender qualquer tipo de funcionalidade. Há muitas funções, como sigmoid, tanh e softmax, mas a mais indicada para redes convolucionais é a Relu por ser mais eficiente computacionalmente sem grandes diferenças de acurácia quando comparada a outras funções. Essa função zera todos os valores negativos da saída da camada anterior.

A camada de *pooling* é usada para simplificar as informações da camada superior. Como na convolução, o *pooling* seleciona uma unidade de área (por exemplo, 2×2) para passar toda a saída da camada anterior. Esta unidade é responsável por resumir as informações desta área com um único valor. Se a saída da camada anterior for 24×24 , a saída combinada será 12×12 . Além disso, é necessário escolher o método de execução do resumo. O método mais comum é o *maxpooling*, onde apenas o número máximo de unidades é passado para a saída. Essa agregação de dados pode reduzir os pesos a serem aprendidos e também pode evitar *overfitting*.

No final da rede, é colocada uma camada *Fully connected*, cuja entrada é a saída da camada convulacional anterior, e sua saída são N neurônios, onde N é a quantidade de classes do seu modelo para finalizar a classificação.

3.3 BASE DE DADOS

Esta seção abordará duas bases de dados disponíveis para uso acadêmico, que serão utilizadas para coletar dados e imagens compatíveis com o tema na realização de testes e simulações durante o desenvolvimento deste trabalho.

3.3.1 *Dataset* de detecção de queda

De acordo com o Auvinet et al. (2010), o laboratório de tratamento de imagens da Universidade de Montreal disponibiliza gratuitamente um *dataset* com diversos vídeos de queda para serem utilizados em trabalhos acadêmicos.

Um dos temas de investigação deste laboratório é a visão assistida por computação aplicada à área médica, para cuidados domiciliários a idosos e de imagens 3D. A outra temática do laboratório é a modelagem estatística de imagens para segmentação, rastreamento e reconhecimento de forma por meio de modelos bayesianos e dados estatísticos.



Figura 3 – Exemplo de captura de imagens do dataset. (Fonte: Auvinet et al. (2010))

Especificamente para uso desde projeto, de acordo com a Figura 3, o conjunto de dados visuais contém 24 cenários registrados com 8 câmeras de vídeo posicionadas de forma inteligente. Os primeiros 22 cenários contêm uma queda e eventos de confusão, os últimos 2 contêm apenas eventos de confusão.

3.3.2 Informações médicas do Hospital Moinho de Vento

De acordo com o trabalho de Souza et al. (2019), esta base de dados pode ser obtida através da Gestão de Riscos do Hospital Moinhos de Vento (HMV), uma instituição com 497 leitos localizada na cidade de Porto Alegre/RS. Este hospital está entre os seis hospitais no Brasil classificados como Centro de Excelência pelo Ministério da Saúde. Planejamento organizacional e uma forte cultura de segurança são aspectos importantes do plano estratégico desta instituição. Neste hospital, todas as ações são pautadas na busca por excelência, e a instituição tem investido na contínua melhoria para ampliar a qualidade e segurança do atendimento fornecido.

Os dados desta base foram obtidos em janeiro de 2018, contendo informações desde primeiro de janeiro de 2012 até 31 de dezembro de 2017. As informações coletadas são importantes para análises e identificação de grupos de riscos para possibilitar uma implementação de estratégias de prevenção de quedas de pacientes. As informações têm apenas uma limitação dos dados referente a que medicação os pacientes tomaram que, por sua vez, possa ter levado a uma queda. Porém, isso não limita o estudo desses dados já que contém outras variáveis importantes relacionadas com os acidentes de quedas.

Este conjunto de dados abrange 1.071 quedas de pacientes internados, caracterizadas pelas seguintes variáveis: ano, data, nascimento do paciente, dia da semana, turno, departamento,

local do incidente, gravidade, presença de acompanhante, idade, sexo, nível de risco, medicamentos associados ao risco de queda, protocolo de prevenção de quedas implantado, tipo de lesão, motivo, prescrição de contenção, prescrição de fisioterapia. Esses dados podem ser analisados usando estatísticas descritivas e apresentados como medidas de tendência central (média e mediana) e variabilidade (desvio padrão e intervalos), bem como distribuições absolutas e relativas.

3.4 HARDWARE DISPONÍVEL

Nesta seção são detalhados os hardwares disponíveis para uso no desenvolvimento deste projeto.

3.4.1 Computadores

Para esse trabalho está disponível os próprios computadores dos autores desse trabalho, e no caso da extração dos vídeos e sua esqueletização e detecção de pontos chaves do esqueleto humano tem a disposição um computador com a seguinte configuração:

- a) Processador AMD Ryzen 7 5800x, 4.8GHz;
- b) 32GB de memória RAM, ADATA XPG SPECTRIX D60G, DDR4, 4133MHz;
- c) Placa de Vídeo GEFORCE RTX 3060 TI 8GB;
- d) Sistema Operacional Windows 10 PRO;

3.4.2 Câmera

A captura de vídeo e reconhecimento de postura e ação pode ser feito através da própria câmera do computador ou vídeo gravado pelo celular. O uso de um algoritmo para detectar os principais pontos do corpo humano consegue atuar facilmente com uma imagem capturada de qualquer aparelho celular moderno.

Para esse trabalho, especificamente, foram utilizados dispositivos como Samsung Galaxy S10e, iPhone 11 e uma câmera de computador para gravar as poses.

3.5 VISÃO COMPUTACIONAL E PROCESSAMENTO DE IMAGENS

A visão computacional (VC), a princípio descrita pelo trabalho de Brownlee (2019), é definida como um campo de estudo que busca desenvolver técnicas para ajudar os computadores a reconhecerem e compreenderem o conteúdo de imagens digitais, como fotografias e vídeos. É um campo multidisciplinar que pode ser descrito como subcampo da inteligência artificial e do aprendizado de máquina (Figura 4), podendo envolver o uso de métodos especializados e fazer uso de algoritmos de aprendizado geral.

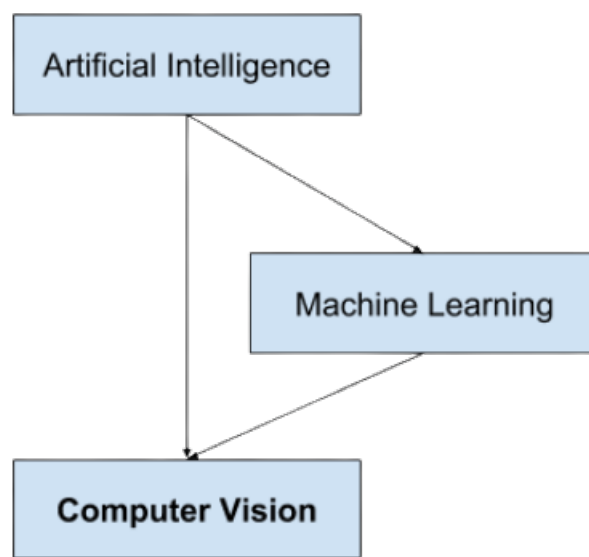


Figura 4 – Diagrama de relacionamento entre as técnicas de Aprendizado de Máquina, Inteligência Artificial e Visão Computacional. (Fonte: Brownlee (2019))

O objetivo da visão computacional é entender o conteúdo das imagens digitais para posterior realização de processamento de imagem, assunto a ser analisado ainda nesta seção. Normalmente, na VC trabalha-se com o desenvolvimento de métodos que tentam reproduzir a capacidade da visão humana no espectro virtual e, desta forma, a compreensão do conteúdo das imagens digitais pode envolver a extração de uma descrição da imagem, uma descrição de texto, um modelo tridimensional e assim por diante.

Um problema específico de visão computacional pode ser facilmente resolvido com um método estatístico feito à mão, enquanto outro pode exigir um conjunto grande e complexo de algoritmos de aprendizado de máquina.

Considerando o melhor entendimento do assunto, é importante destacar que a técnica anteriormente citada, chamada de Processamento de Imagem, complementa o trabalho da visão computacional, mesmo sabendo que ambas possuem comportamentos e funções distintas. Para

entender melhor as diferenças dessas técnicas e como elas se complementam no tratamento de imagens digitais, tem-se um breve resumo do processamento de imagens.

O processamento de imagem é o processo de criação de uma nova imagem a partir de uma imagem existente, geralmente simplificando ou aprimorando o conteúdo de alguma forma. É um tipo de processamento de sinal digital que não se preocupa em entender o conteúdo de uma imagem, por assim dizer.

Dada uma situação utilizando as duas técnicas em atuação conjunta, um determinado sistema de visão computacional pode receber uma figura e realizar um pré-processamento, a fim de facilitar o reconhecimento pela VC da seguinte forma, por exemplo:

- a) Normalizando as propriedades fotométricas da imagem, como brilho ou cor.
- b) Recortando os limites da imagem, como centralização de um objeto em uma fotografia.
- c) Removendo ruído digital de uma imagem, como artefatos digitais de baixos níveis de luz.

Nesta situação, verifica-se que o processamento de imagens pôde preparar imagens para a aplicação de métodos de visão computacional, sendo estes últimos os responsáveis por analisar e interpretar o conteúdo das imagens extraindo informações relevantes. Ambos os processos possuem algoritmos com variáveis de entrada e saída. No processamento de imagens, a entrada é uma imagem e a saída é outra imagem especialmente modificada. Em visão computacional, a entrada também é uma imagem, mas a saída possui informações sobre o que está acontecendo na cena e descrição de seus elementos.

As subseções a seguir abordarão técnicas e aplicações de visão computacional em mais detalhes.

3.5.1 Técnicas e aplicação de Visão Computacional

De acordo com o trabalho de Le (2018), pode-se destacar 3 etapas no processo de reconhecimento de imagem no campo computacional de visão, que serão usadas como base conceitual teórica do trabalho. Estas etapas compõem as seções a seguir.

3.5.2 Classificação da imagem

A classificação de imagens em VC ocorre da seguinte forma: dado um conjunto de imagens rotuladas com uma categoria específica, é solicitada a alteração dessa categoria para um novo conjunto de imagens, utilizando uma medição e previsão. Existem desafios para a máquina em reconhecer um possível objeto, incluindo variação de ponto de vista, variação de escala, deformação de imagem, oclusão de imagem, condições de iluminação e confusão de fundo.

A literatura de visão computacional apresenta abordagens baseadas em dados para resolver esses desafios. Essa literatura fornece exemplos de cada categoria de imagem, desenvolvendo algoritmos de aprendizagem para analisar exemplos e entender a aparência visual de cada categoria. Em outras palavras, eles primeiro acumulam um conjunto de dados de treinamento de imagens rotuladas e, em seguida, alimentam o computador para processar os dados, como a seguir:

- a) A entrada é um conjunto de dados de treinamento que consiste em n imagens, cada uma rotulada com uma das k classes diferentes.
- b) Em seguida, é usado este conjunto de dados para treinar um classificador, a fim de aprender a aparência de cada uma das classes.
- c) Por fim, é avaliada a qualidade do classificador solicitando que ele preveja rótulos para um novo conjunto de imagens que não fora utilizado para o treinamento. Sendo assim, é feita a comparação dos rótulos verdadeiros dessas imagens com os previstos pelo classificador.

3.5.2.1 Segmentação semântica

Um dos processos da visão computacional é a segmentação semântica, que divide imagens em grupos de *pixels* a serem rotulados e classificados. Esta segmentação busca entender o sentido de cada *pixel* na imagem analisada.



Figura 5 – Cena analisada utilizando segmentação semântica. (Fonte: Le (2018))

Por exemplo, na Figura 5, além de se definirem os limites de cada objeto, também se definem as classes pertencentes à pessoa, ruas, carros e árvores, separadamente. Portanto, ao contrário da classificação, são necessárias previsões mais densas em termos de *pixels* dos modelos analisados.

Esta técnica pode ser realizada com a *Fully Convolutional Networks* (FCN), vista no trabalho de Long, Shelhamer e Darrell (2015), que popularizou arquiteturas convolucionais de ponta a ponta para previsões densas, sem ter a necessidade de camadas completamente conectadas. Isso permitiu, a princípio, que mapas de segmentação fossem gerados para imagens de qualquer tamanho, utilizando densa predição por *pixel*, como visto na Figura 8.

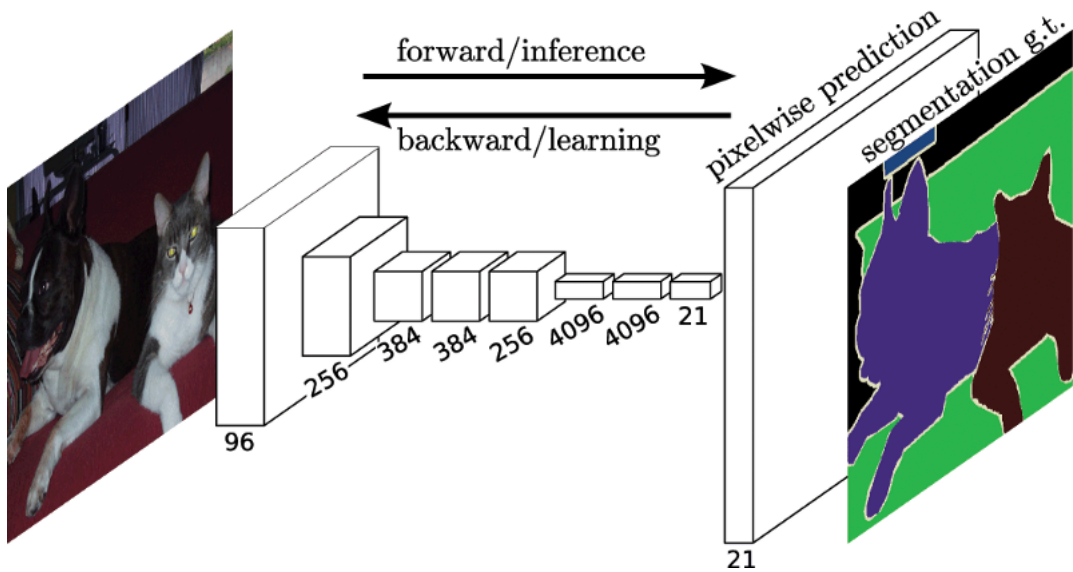


Figura 6 – Exemplo de arquitetura de camadas FCN. (Fonte: Le (2018))

O FCN usa *downsampling* (redução artificial da taxa de amostragem) e *upsampling* (aumento artificial da taxa de amostragem) dentro da rede. A camada de *downsampling* é conhecida como convolução listrada, enquanto a camada de *upsampling* é conhecida como convolução transposta.

3.5.2.2 Segmentação de instância

A segmentação de instância separa pessoas das cenas em diferentes instâncias de classes, como por exemplo, definir rótulos de 5 pessoas com 5 cores diferentes. Na classificação, geralmente há uma imagem com um único objeto como foco e a tarefa é dizer o que é essa imagem. Na segmentação semântica, é possível separar classes inteiras de objetos utilizando cadeias de *pixels*. Já na segmentação de instâncias, por sua vez, as tarefas são mais complexas, visto que a separação é individual para cada *pixel* (Le (2018)).



Figura 7 – Cena analisada utilizando segmentação de instância. (Fonte: Le (2018))

Na Figura 7, existem regiões de difícil segmentação, com múltiplos objetos sobrepostos e diferentes fundos, onde o objetivo de instanciar os objetos da imagem não apenas se resume na classificação das diferenças de itens, mas também na identificação de seus limites e suas relações uns com os outros.

Até esta seção, foi possível verificar como utilizar recursos da CNN de maneiras interessantes para localizar efetivamente diferentes objetos em uma imagem com caixas delimitadoras, como na Figura 8. Pode-se estender essas técnicas para localizar *pixels* exatos de cada objeto em vez de apenas caixas delimitadoras, como na arquitetura Mask R-CNN, presente no artigo de He et al. (2018).

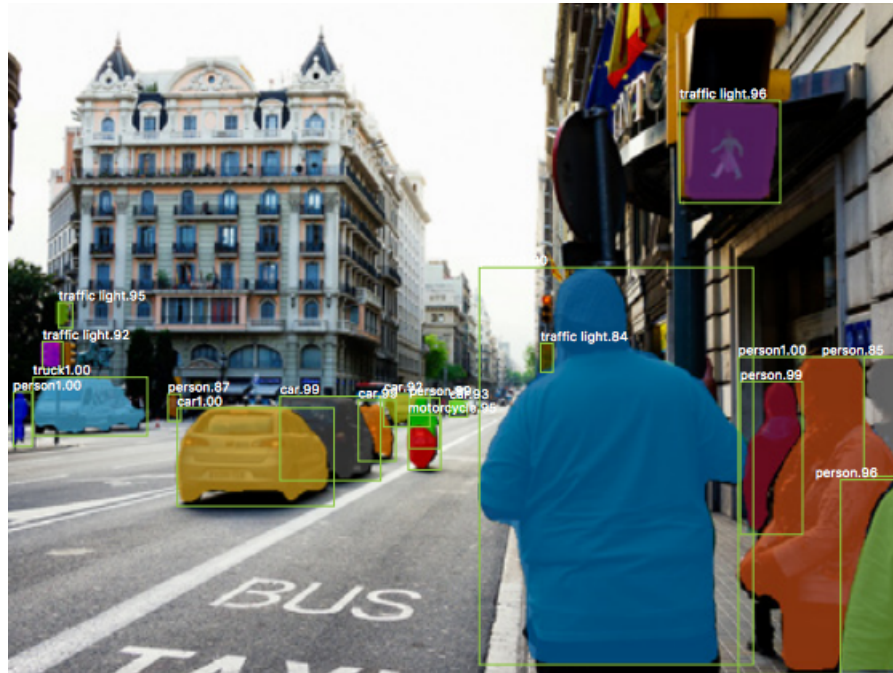


Figura 8 – Cena analisada utilizando segmentação de instância. (Fonte: Le (2018))

Uma técnica para segmentar instâncias é o Mask R-CNN, de He et al. (2018). Ele produz uma máscara binária que diz se um determinado *pixel* é ou não parte de um objeto. Sua ramificação é uma rede totalmente convolucional, onde se produz uma matriz com o número 1 em todos os locais onde o *pixel* pertence ao objeto e 0 em outros lugares.

Além disso, esse último método, ao ajustar um RoIPool (Figura 9), é alinhado com mais precisão usando um método conhecido como RoIAlign (*Region of Interests Align*). Particularmente, o RoIAlign usa interpolação bilinear para evitar erros de arredondamento, corrigindo possíveis imprecisões na detecção e segmentação.

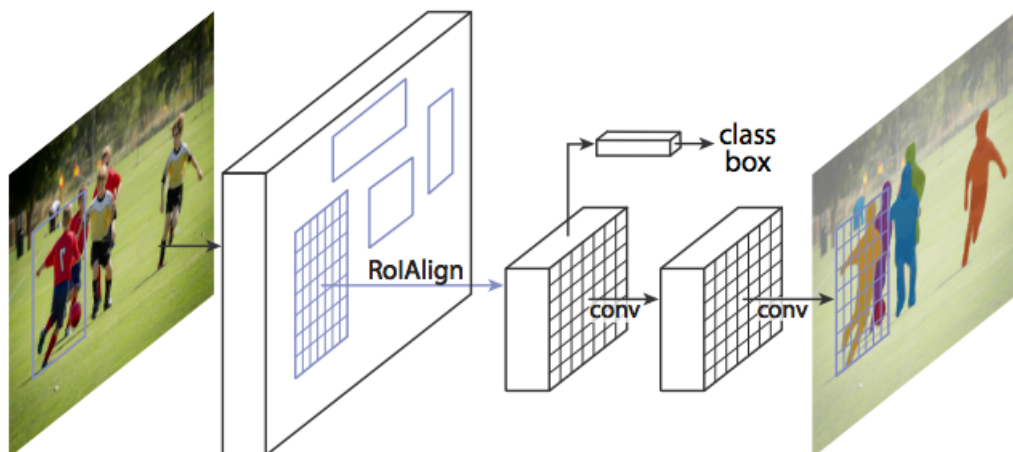


Figura 9 – Uso do RoIAlign, com interpolação bilinear. (Fonte: He et al. (2018))

Sendo assim, uma vez que essas “máscaras” são geradas, o Mask R-CNN é combinado com as classificações e retângulos delimitadores para gerar segmentações precisas, conforme Figura 10.



Figura 10 – Exemplos de retângulos delimitadores em segmentação de instância, utilizando o Mask R-CNN. (Fonte: Le (2018))

3.6 ESQUELETIZAÇÃO DE IMAGENS

A criação do Microsoft Kinect proporcionou detecção de profundidade a um baixo custo e conseqüentemente abriu novas oportunidades para resolver diversos problemas na visão computacional. O sensor pode combinar equipamentos avançados para a atividade de detecção. São eles: um projetor de infravermelho, uma câmera infravermelha e outra RGB, além de quatro microfones. Estes hardwares são responsáveis por fornecerem a captura de movimentos 3D, o reconhecimento facial e o reconhecimento de voz.

A tecnologia utilizada na detecção de profundidade é resultado da combinação do projetor infravermelho e uma câmera infravermelho, ambos representados na Figura 11. O padrão de pontos projetados e a geometria entre o projetor e a câmera são conhecidos, possibilitando o relacionamento entre um ponto padrão do projetor e um ponto observado na imagem.

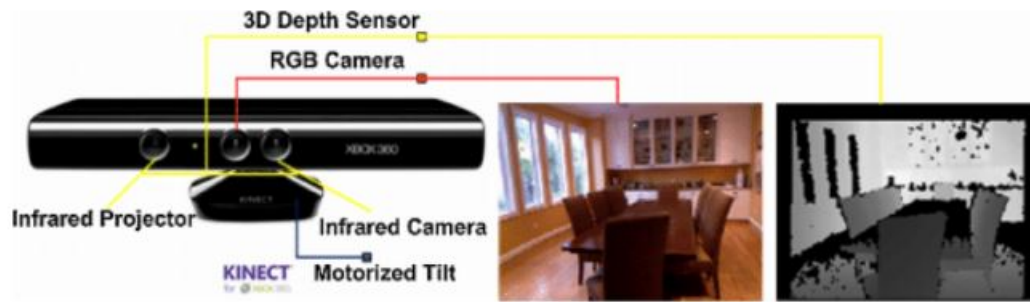


Figura 11 – Configuração de Hardware do Kinect. (Fonte: Han et al. (2013))

Como resultado do processo, obtém-se um mapa de profundidade (Figura 12), onde os valores são codificados em escalas de cinza. Quanto mais escuro é o *pixel*, mais próximo o ponto está da câmera. Os pontos pretos indicam que não há valores de profundidade.

O Kinect também trouxe evolução quanto à tarefa de detecção de articulações ou esqueleto. Para isto ocorrer, é necessário realizar dois estágios. O primeiro é construir um mapa de profundidade, obtido através da análise de um padrão de pontos infravermelho. No segundo estágio, são inferidas partes do corpo com uso de uma floresta de decisão aleatória, treinada a partir de imagens de profundidade de partes corporais. A etapa anterior combinada a um algoritmo de deslocamento médio, responsável por calcular os modos de distribuições de probabilidade, transformam a imagem da parte do corpo em esqueleto (Zhang (2012)).

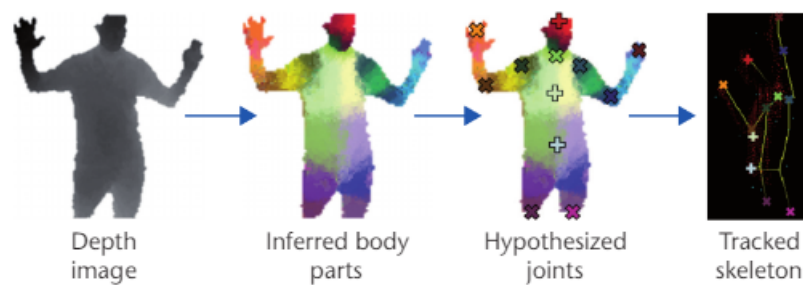


Figura 12 – Esqueletização Kinect. (Fonte: Han et al. (2013))

3.6.1 *OpenPose*

O projeto de reconhecimento de gestos humanos *OpenPose*, visto no trabalho de Chen et al. (2020), compõe uma biblioteca de código aberto desenvolvida pela Universidade Carnegie Mellon (CMU), com base em uma rede neural convolucional e sistema de aprendizado supervisionado. Em 2017, cientistas e pesquisadores dessa universidade lançaram um código fonte do sistema para reconhecer um esqueleto humano com o *OpenPose*, com a finalidade de

rastrear o movimento em tempo real de pessoas por vídeo. O sistema da *OpenPose* foi capaz de detectar informações do esqueleto de várias pessoas, e para isso adotou a estimativa de postura do corpo humano de cima para baixo, passando por um algoritmo de detecção da posição de pontos-chave do corpo humano e, em seguida, usou o vetor de características de afinidade para determinar os nós chaves. Esse algoritmo pode realizar distinção de movimentos humanos, expressões faciais, movimento dos dedos e outras estimativas de postura.

Conforme mostrado na Figura 13, a tela capturada pela câmera de vigilância usa *OpenPose* para obter informações de nós chaves humanos.

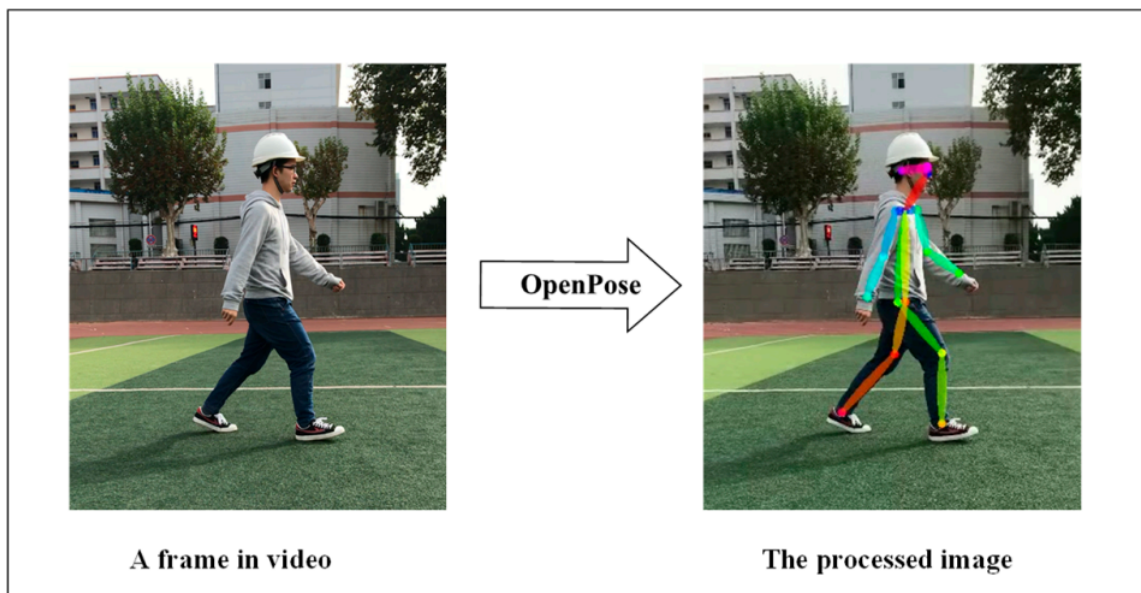


Figura 13 – Imagem processada utilizando *OpenPose*. (Fonte: Chen et al. (2020))

3.7 RECONHECIMENTO DE POSTURA

Reconhecimento de postura é um tópico atrativo e desafiador em computação visual e vem sendo aplicado em diversas áreas, sendo que a área médica é uma das que possui uma maior demanda. Uma das principais aplicações nessa área é na saúde de idosos como, por exemplo, detecção de quedas. Como citado no trabalho de Huang et al. (2019), nos Estados Unidos, a população maior de 65 anos alcançou mais de 50 milhões em 2017, representando 15.41% da população do país. A China também aparece com um número crescente de idosos com mais de 158 milhões, representando 10.64% da população em 2017, se tornando o país com maior número de idosos no mundo.

Essa categoria de reconhecimento consiste em basicamente dois estágios: o primeiro é um treinamento do sistema usando uma base de dados de posturas para poder aprender a classificar as posturas humanas. Htike, Lai e Khalida (2011)

Na Figura 14 pode-se ver melhor como é esse estágio de treinamento e classificação.

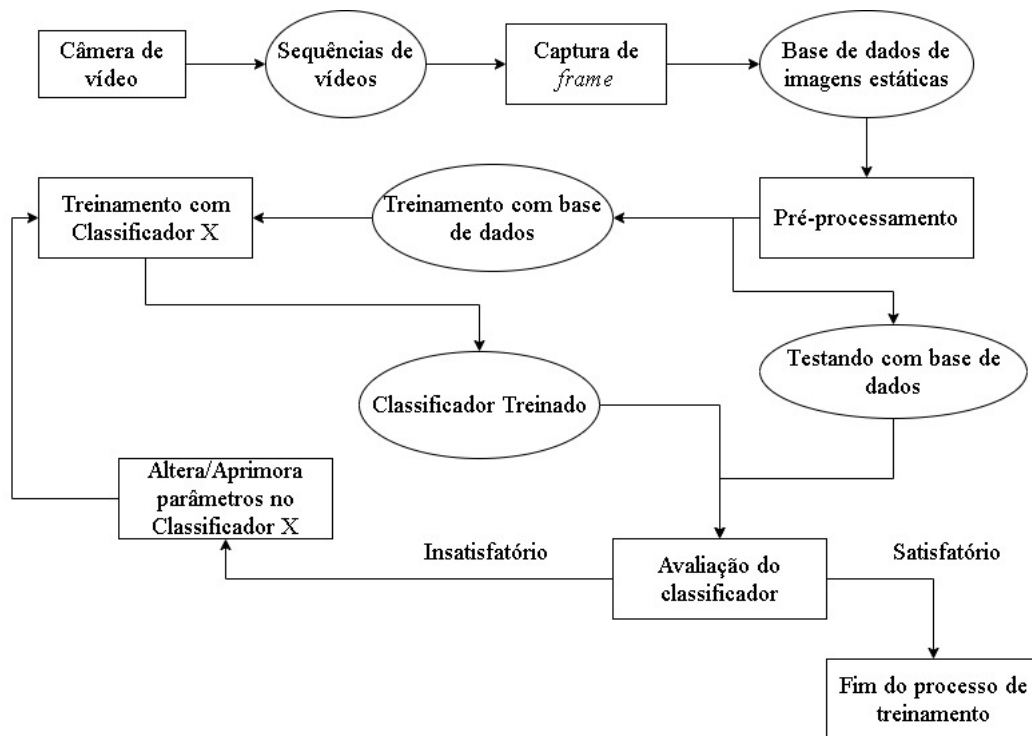


Figura 14 – Exemplo de estágio de treinamento e classificação. (Traduzido da : Htike, Lai e Khalida (2011))

Inicia com uma câmera de vídeo fazendo uma captura, e essas imagens capturadas passam por uma sequência: *Video Camera*, *Video Sequences*, *Frame Grabber*, criando assim uma base de dados de imagens estáticas coloridas. O pré-processamento é um passo com diversos algoritmos que aplicam o treinamento desses *datasets* e testes. Em seguida, é executado um *loop* por diversos algoritmos de classificação, treinados, testados e pontuados, até que o resultado seja satisfatório para o fim do processo. Essa fase é passada de maneira a aprimorar ainda mais a classificação a cada teste. Cada algoritmo de classificação é treinado e testado individualmente.

Após o teste e treinamento, o modelo treinado está pronto para ser aplicado.

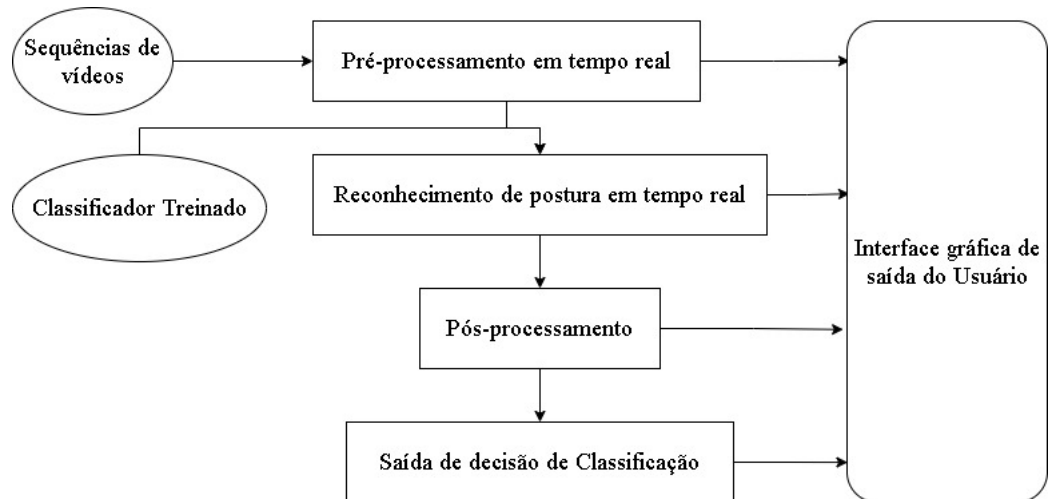


Figura 15 – Estágio de aplicação do sistema. (Traduzido da Fonte: Htike, Lai e Khalida (2011))

Nessa fase, mostrada na Figura 15, cada *frame* da sequência do vídeo, que está a 30 *frames* por segundo (fps), é diminuída em 6 fps, visto que cada quadro não necessita ser processado, já que há uma redundância entre eles. Para o reconhecimento, já é suficiente o processo de apenas alguns *frames* em um segundo.

No final desse processo é possível ver o resultado de cada *frame* capturado que foi processado, feito sua extração e classificado em uma interface do sistema (a *interface* vai depender de sistema para sistema).

3.8 RECONHECIMENTO DE AÇÕES

O reconhecimento de ações ou gestos é uma *interface* entre usuário e computador que possibilita interpretar os gestos humanos como comandos, sendo capaz de entendê-los para executar comandos apenas e através de suas ações.

Como descrito em Schechter (2014), o reconhecimento de gestos é definido com mais detalhes como funciona o reconhecimento de gestos:

- a) Uma câmera gera imagens, sendo passadas para um dispositivo de detecção conectado ao computador. Esse dispositivo de detecção usa um sensor infravermelho ou projetor que calcula a profundidade.
- b) É implementado um software específico de reconhecimento de gestos que possui um banco de dados com diversos gestos, sendo igualados a comandos para o computador.

c) O software correlaciona cada gesto registrado, interpretando e usando o banco de dados com gestos para identificá-los nas suas correspondências.

d) Finalmente, o gesto é interpretado e o computador executa o comando daquela específica ação.

De forma resumida, o reconhecimento de gestos pode ser similar a um reconhecimento de postura, mas possui uma finalidade diferente de usar essas imagens processadas que são segmentadas e, quando reconhecidas, possui um comando específico para ser executado.

4 METODOLOGIA

A metodologia proposta neste trabalho consiste em oito etapas, como apresentada na Figura 16. A primeira etapa (bloco 1 da Figura 16) contempla a base de dados, que alimenta as informações para as etapas subsequentes. A segunda, terceira e quarta etapas (bloco 2, 3 e 4 da Figura 16), se referem à inserção de dados referentes às informações dos pacientes, médicas e hospitalares. Na quinta etapa (bloco 5 da Figura 16), foi realizada a captura de imagens através de uma câmera. Durante a sexta etapa (bloco 6 da Figura 16), realizou-se a extração de características visuais contendo a esqueletização da imagem e tirando seus pontos chaves, passados para a etapa de composição de informações (bloco 7 da Figura 16), de modo a cruzar as informações importantes. Na etapa 8, por fim, foi feita a entrega dos resultados finais do conjunto das informações dos blocos 2, 3 e 4 com as informações de extração de características visuais (bloco 8 da Figura 16).

A seguir, cada uma dessas etapas foi melhor descrita.

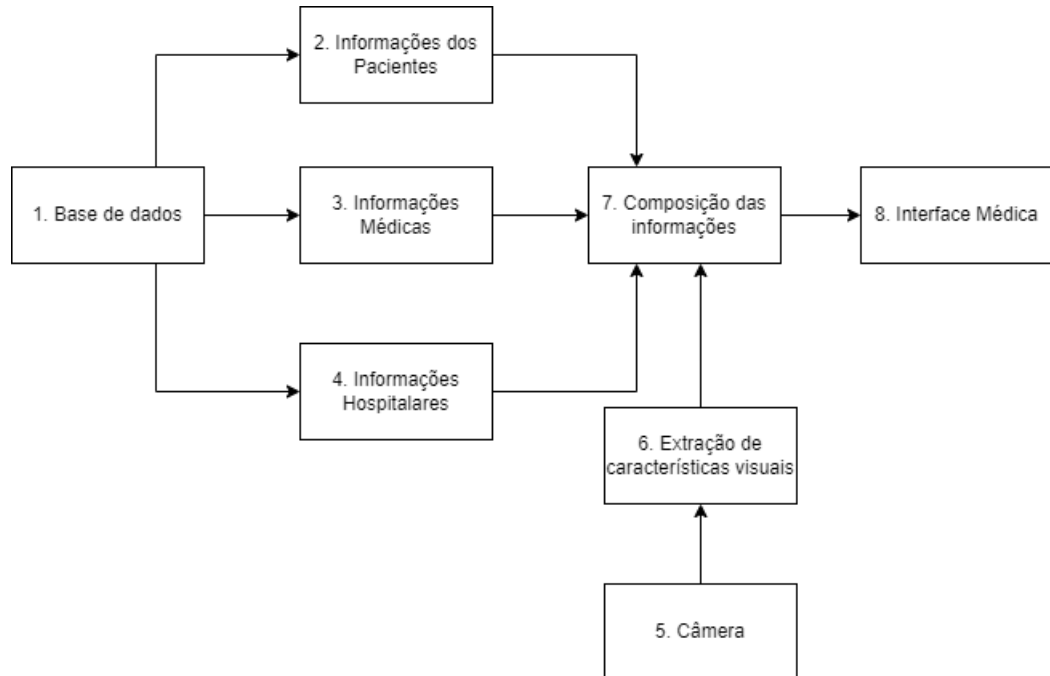


Figura 16 – Fluxograma Geral da Metodologia.

4.1 BASE DE DADOS

A primeira etapa deste capítulo (bloco 1 da Figura 16) abordou dados textuais e visuais pré cadastrados, utilizados para alimentar toda a estrutura de informações do projeto. Para esta alimentação inicial, foi utilizada a base de dados com quedas presentes na Seção 3.3.1, assim como também imagens dos próprios autores desse trabalho para a detecção de movimento entre as poses e os dados de livre acesso do Hospital Moinhos de Vento, para análise de informações dos pacientes, presentes na Seção 3.3.2.

4.2 INFORMAÇÕES DOS PACIENTES

A segunda etapa (bloco 2 da Figura 16) da metodologia proposta neste trabalho foi a obtenção das informações dos pacientes para uma análise e estudo de tais dados. Essas informações são referentes às quedas de pacientes contendo informações importantes dos mesmos, como: idade, sexo do paciente, severidade do acidente e ferimento. Esse estudo serviu para identificar probabilidades de possíveis quedas futuras com base nesses dados.

4.3 INFORMAÇÕES MÉDICAS

A terceira etapa (bloco 3 da Figura 16) da metodologia proposta neste trabalho foi a obtenção das informações médicas, com a finalidade destes dados serem utilizados para a composição das informações, presente no bloco 7 da Figura 16. Neste módulo, houve a aquisição de informações importantes a respeito de nível de risco do paciente, como: se houve implementação de algum protocolo de queda, medicação envolvida com determinada queda, razão do acidente, turnos e dia do acidente.

4.4 INFORMAÇÕES HOSPITALARES

A quarta etapa (bloco 4 da Figura 16) da metodologia proposta neste trabalho foi a obtenção das informações hospitalares, que se basearam em dados referentes à infraestrutura hospitalar, como setores e ambientes onde ocorreram os acidentes. Estas informações foram organizadas em tabelas e enviadas à composição de informações, presente no bloco 7 da Figura 16.

4.5 CÂMERA

A quinta etapa da metodologia (bloco 5 da Figura 16) foi implementar uma rede de captura de imagens através de uma câmera, onde se aplicou um posicionamento cenográfico estratégico para a obtenção completa de movimentos humanos em um determinado ambiente e posterior realização da extração de características visuais.

4.6 EXTRAÇÃO DE CARACTERÍSTICAS VISUAIS

A sexta etapa da metodologia (bloco 6 da Figura 16) recebeu imagens da câmera (bloco 5 da Figura 16), sendo realizado o tratamento de imagem de acordo com métodos propostos na Seção 3.5, onde, utilizando dados coletados da câmera e com a base de dados pré alimentada, foi feito o reconhecimento de movimento das posições: sentado, em pé e levantado. Dessas movimentações, foi possível extrair seus *keypoints*, *frame* por *frame*, e houve a verificação do seu movimento através da diferença entre os seus *keypoints* x e y .

4.7 COMPOSIÇÃO DAS INFORMAÇÕES

Na sétima etapa (bloco 7 da Figura 16) da metodologia proposta neste trabalho, foi feita a composição das informações dos blocos 2, 3 e 4, assim como das extrações de características visuais. Resultando em uma probabilidade de queda entre as movimentações detectadas, essa probabilidade foi utilizada com os menores valores possíveis dos dados. Havendo então uma porcentagem mínima de risco de queda, esta pode variar de acordo com a quantidade de pacientes e variedade das informações nos dados.

4.8 INTERFACE MÉDICA

Na oitava etapa (bloco 8 da Figura 16) da metodologia proposta neste trabalho, houve a disposição dos dados recolhidos nas etapas anteriores para expor seus resultados e poderem, futuramente, ser utilizados para uma interface médica visual.

5 PROPOSTA EXPERIMENTAL

Conforme metodologia proposta neste trabalho de oito etapas, descritas na Figura 16, os experimentos aqui identificados visaram quatro tarefas importantes para a execução estratégica do trabalho. A primeira etapa teve como base a aquisição de informações dos pacientes, médicas e hospitalares para tratamento de dados com modelo bayesiano. Na segunda etapa houve a extração de características visuais. A terceira etapa compôs o experimento com reconhecimento da movimentação das posturas. Na quarta etapa, por fim, houve a composição das informações geradas com o modelo bayesiano e com as movimentações detectadas.

5.1 AQUISIÇÃO DE INFORMAÇÕES DOS PACIENTES, MÉDICAS E HOSPITALARES

Para iniciar a aplicação da metodologia proposta, foi necessário realizar a coleta de informações dos pacientes, informações médicas e hospitalares. Foram utilizados dados do Hospital Moinhos de Vento (Seção 3.3.2), todos de livre acesso. Posteriormente, houve uma análise destes dados para melhor familiarização a respeito dos pacientes que sofreram uma queda, bem como um tratamento para extrair informações importantes como relações entre idade, risco de queda e ambientes.

Esses resultados foram dispostos em tabelas com as probabilidades das quedas e gráficos com informações relacionadas usando um modelo bayesiano, vistos na Seção 3.1.

5.2 EXPERIMENTO DE EXTRAÇÃO DE CARACTERÍSTICAS VISUAIS

Neste experimento, foram extraídas as características visuais, vistas na Seção 4.6, que foram as mais adequadas para a proposta desse trabalho. Foi utilizado um *dataset* de imagens de queda (Seção 3.3.1) para testes iniciais do *OpenPose*, conforme Figura 17.



Figura 17 – Teste com *dataset* utilizando *openpose* (Fonte: autores)

Após ter sucesso, foi realizada a extração das imagens dos próprios autores desse trabalho, através de vídeos, para passar por um processo de reconhecimento de movimento, como fora visto na Seção 3.7. O resultado deste experimento foi cruzado com as informações da Seção 5.1 para os cálculos probabilísticos de uma possível queda.

5.3 EXPERIMENTO DE RECONHECIMENTO DE MOVIMENTO

Neste experimento, houve o teste da capacidade de usar informações do esqueleto fornecidas por uma câmera para o reconhecimento da esqueletização humana na Seção 3.6, e extrai informações para detectar um movimento, em um contexto de uma estrutura de monitoramento para prevenção de uma queda.

Foram conduzidos experimentos com recursos extraídos do esqueleto humano. Foram feitos experimentos com diversas posturas (como, por exemplo: sentando, levantando e levantado, conforme mostrado nas Figuras 18, 19 e 20) para verificar o movimento entre essas posturas. Utilizando filtros de processamento de imagens e segmentação semântica/segmentação de instância, foi possível determinar os limites da postura de forma genérica, para assim, identificar e diferenciar um esqueleto humano de demais objetos e concentrar o resto do trabalho na forma de movimentação humana.

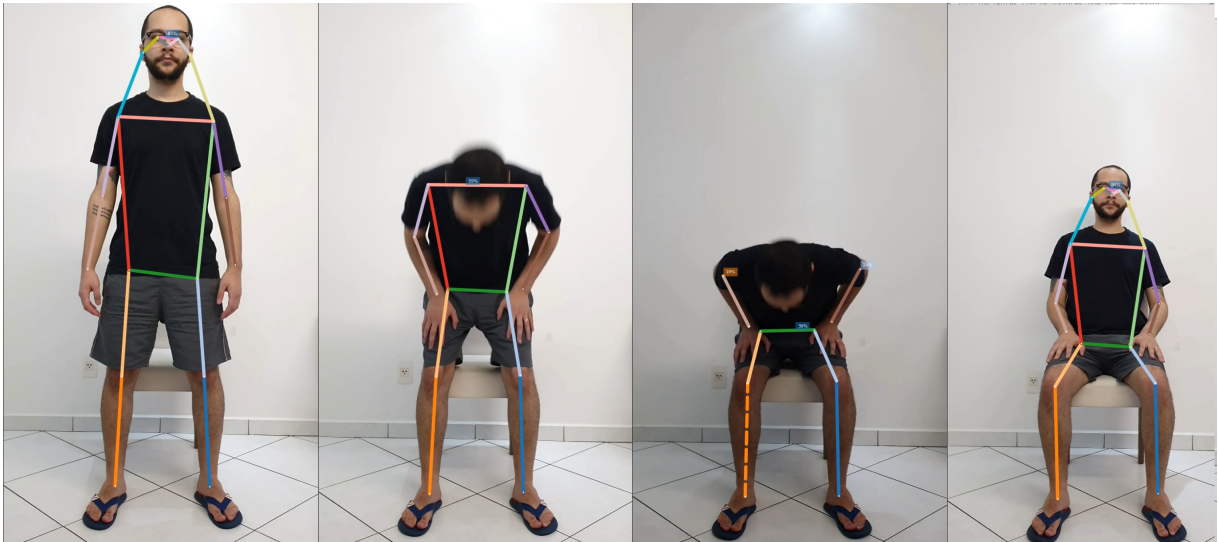


Figura 18 – Posturas humanas - Sentando (Fonte: autores)

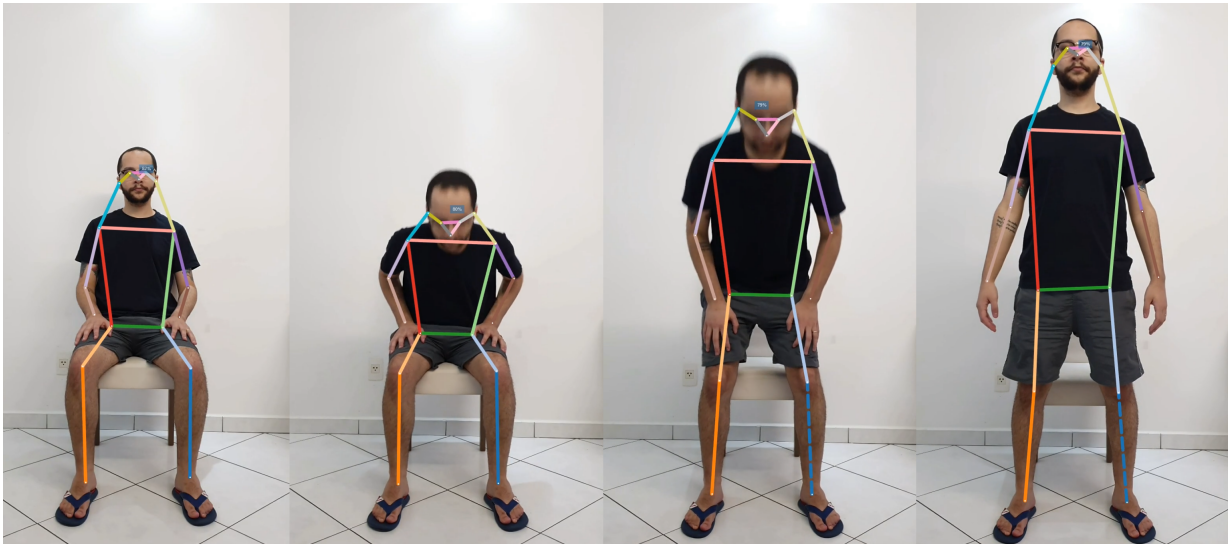


Figura 19 – Posturas humanas - Levantando (Fonte: autores)



Figura 20 – Posturas humanas - Levantado (Fonte: autores)

5.3.1 *KeyPoints*

Os *keypoints* foram obtidos utilizando o *OpenPose*, da Seção 3.6.1. Exemplos de *keypoints* são cabeça, ombros ou joelhos, conforme mostrado na Tabela 1. E cada *keypoint* foi representado pela sua posição no eixo X e Y do plano cartesiano.

Index	<i>keypoint</i>
0	Nariz
1	Olho esquerdo
2	Olho direito
3	Orelha esquerda
4	Orelha direita
5	Ombro esquerdo
6	Ombro direito
7	Cotovelo esquerdo
8	Cotovelo direito
9	Pulso esquerdo
10	Pulso direito
11	Quadril esquerdo
12	Quadril direito
13	Joelho esquerdo
14	Joelho direito
15	Tornozelo esquerdo
16	Tornozelo direito

Tabela 1 – Tabela dos *KeyPoints* do esqueleto humano. (Fonte: Chen et al. (2020))

Primeiramente, ao se utilizar os pontos X e Y referentes aos dados da Tabela 1, pôde-se iniciar os cálculos através da Equação 2:

$$Dif = \sum_{n=1}^K difX, difY \quad (2)$$

onde,

a) $difX$ é a diferença dos pontos do eixo X ;

b) $difY$ é a diferença dos pontos do eixo Y ;

Em seguida, o resultado Dif é passado para a Equação 3:

$$T = \frac{K}{R} \quad (3)$$

onde,

- a) K representa os *keypoints*;
- b) R representa a resolução da tela;

Então, a probabilidade da ação P é calculada utilizando a Equação 4:

$$P = \frac{Dif}{T} \quad (4)$$

onde,

- a) P é a probabilidade da ação realizada;
- b) Dif é a diferença dos pontos X e Y ;
- c) T é a diferença dos *keypoints* pela resolução da tela;

5.4 COMPOSIÇÃO DE INFORMAÇÕES

Nesta seção, por fim, foram analisados os resultados da movimentação e das probabilidades dos dados, e foram reutilizados no modelo bayesiano para compor dois resultados juntos, tendo uma probabilidade mínima de queda com as médias de todos os dados da Seção 5.1 e os movimentos detectados com as poses, citados na Seção 5.3.

6 RESULTADOS E DISCUSSÃO

Neste capítulo foram apresentados os resultados obtidos através dos experimentos propostos no Capítulo 5.

6.1 INFORMAÇÕES DO PACIENTE, MÉDICAS E HOSPITALARES

Nesta seção foram discutidos os resultados sobre os dados adquiridos da Seção 3.3.2. Os dados coletados, após análise, identificaram grupos de riscos de modo a possibilitar uma implementação de estratégias de prevenção de quedas para pacientes em situações similares. Estas informações têm apenas algumas limitações dos dados referentes a maior variedade de informações e detalhes como, por exemplo, que medicação o paciente tomou que tem envolvimento com o risco de queda. A limitação da quantidade de dados também pôde influenciar nas porcentagens, porém foi possível notar a diferença entre os casos.

O primeiro passo foi fazer uma limpeza dos dados e entender o que cada informação significava, conforme Tabela 2. Cada informação nessa base se refere aos pacientes, informações medicas e hospitalares.

Hospital Moinhos de Vento
Dia da semana
Turno
Setor
Idade
Presença de companhia
Local do acidente
Nível de risco
Razão do acidente
Implementado protocolo de risco de queda
Sexo
Medicação
Severidade do acidente
Tipo de ferimento

Tabela 2 – Informações referentes à queda. (Fonte: Souza et al. (2019))

Tendo estes dados, foi feito um cálculo utilizando o modelo bayesiano com as informações coletadas, utilizando a Equação 5.

$$P(A, B, \dots, n) = 1 - (1 - P(A)) * (1 - P(B)) * \dots * (1 - P(n)) \quad (5)$$

onde:

- a) P é a probabilidade de risco de queda (de 0 a 1);
- b) A, B, \dots, n , representa a média dos valores na Tabela 2

Com base na Equação 5, o primeiro experimento com os dados obteve sucesso ao pegar os menores valores da tabela após se normalizar, onde obteve-se um resultado de 86,85%. Já com os maiores valores, notou-se um resultado de 100,00%, utilizando uma escala arredondada de duas casas decimais. Abaixo, na Tabela 3, pôde-se ver com maior detalhamento os números obtidos:

Valores	$P(A, B, \dots, n)$	Porcentagem de risco de queda
Menores	0,868578985	86,85%
Maiores	0,999998668	100,00%

Tabela 3 – Probabilidades com maiores e menores valores de risco de quedas. (Fonte: autores)

Os seguintes experimentos foram necessários para buscar uma variação entre os níveis de risco de queda, agrupando apenas as informações diretas ao paciente.

Idade	Sexo	Porcentagem de risco de queda
0,214950	0,464486	57,95%

Tabela 4 – Médias com o nível alto de risco de queda. (Fonte: autores)

Idade	Sexo	Porcentagem de risco de queda
0,068220	0,5335514	56,72%

Tabela 5 – Médias com o nível baixo de risco de queda. (Fonte: autores)

Referente aos resultados das Tabelas 4 e 5, vale considerar que, no caso das médias com nível de risco de queda baixo dos dados, contém um total de apenas 50 pacientes contra 634 pacientes com o nível alto de risco de queda, influenciando assim o resultado com uma diferença de apenas 1,23%.

Nas Tabelas 6, 7, 8, foram feitas médias dos 1070 pacientes com os variados valores entre eles, obtendo uma porcentagem final de mais de 90% no geral.

Dia	Turno	Idade	Presença de companhia	Nível de risco	Sexo
0,149530	0,349530	0,181310	0,276640	0,592520	0,535514
Porcentagem de risco de queda					
93,79%					

Tabela 6 – Valores com variação de turno e dia. (Fonte: autores)

Idade	Presença de companhia	Ambiente	Nível de risco	Sexo
0,181310	0,276640	0,542991	0,592520	0,535514
Porcentagem de risco de queda				
94,87%				

Tabela 7 – Valores com variação de ambiente. (Fonte: autores)

Idade	Nível de risco	Protocolo	Sexo	Medicação
0,181310	0,592520	0,700930	0,535514	0,610000
Porcentagem de risco de queda				
98,19%				

Tabela 8 – Valores com variação de protocolo e medicação. (Fonte: autores)

Por último, ainda nesta seção de relação entre os dados, foi calculada a probabilidade de queda entre os valores dos dados de paciente, médico e hospitalar. Os resultados podem ser vistos nas Tabelas 9, 10 e 11.

Tipo	Probabilidade de queda
Paciente	99,75%
Paciente (sem severidade)	89,12%
Paciente (risco baixo, sem severidade)	87,62%
Paciente (risco alto, sem severidade)	87,97%

Tabela 9 – Probabilidades de queda analisando informações dos pacientes. (Fonte: autores)

Tipo	Probabilidade de queda
Médico	97,93%
Médico (risco baixo)	80,59%
Medico (risco alto)	97,93%

Tabela 10 – Probabilidades de queda analisando informações médicas. (Fonte: autores)

Tipo	Probabilidade de queda
Hospitalar	87,78%
Hospitalar (risco baixo)	16,86%
Hospitalar (risco alto)	87,78%

Tabela 11 – Probabilidades de queda analisando informações dos ambientes hospitalares. (Fonte: autores)

6.2 TESTE DE RECONHECIMENTO DE MOVIMENTO

O experimento de reconhecer um movimento teve como objetivo analisar pontos chaves das imagens capturadas e passadas para o algoritmo *OpenPose*, visto na Seção 3.6.1, que extraiu 17 pontos do corpo humano e formou uma esqueletização da imagem. Nesse algoritmo foi extraído um arquivo *.json* com esses pontos, *frame a frame*.

Para análise dos dados, foram utilizadas as Equações 2, 3 e 4, e os resultados foram os seguintes:

Para *Dif*, que foi chamado de somatória das diferenças dos pontos X e Y , pôde-se obter os valores para posterior passagem para T , que foi o cálculo da divisão dos 34 *keypoints* pela resolução da tela de 800 *pixels*. Então, por fim, foi realizado o cálculo de P , a probabilidade da ação realizada.

Visto isso, de acordo com os testes utilizando os *keypoints* vistos na Tabela 1 e as coordenadas X e Y , foi obtido o resultado abaixo, na Tabela 12, onde utilizando um parâmetro de 0 a 1, foi definido pelos autores um grau de satisfação acima de 0,9 de acurácia (Seção 5.3).

Index	Coordenada-X	Coordenada-Y	Acurácia
0	423,23	47,93	0,954543
1	433,04	35,14	0,922112
2	405,62	37,89	0,953331
3	449,32	38,56	0,891242
4	442,23	48,22	0,904421
5	387,11	52,93	0,833412
6	481,22	128,95	0,895478
7	367,74	134,23	0,862399
8	505,02	220,42	0,898577
9	342,32	215,32	0,934531
10	516,34	308,45	0,913673
11	325,32	296,45	0,855732
12	463,22	319,90	0,973621
13	383,25	319,12	0,784673
14	480,86	437,27	0,845792
15	481,34	415,87	0,883657
16	415,32	354,14	0,953664

Tabela 12 – Tabela de *keypoints* do esqueleto humano. (Fonte: autores)

De acordo com a Tabela 12, dos 17 *keypoints* analisados, 8 estão acima de 0,9 na escala de acurácia, onde caracterizam uma alta fidelidade na sua detecção. O restante, embora não esteja totalmente satisfatório, mostra um mínimo de 0,78, um valor alto se considerar a escala de 0 a 1.

6.2.1 Movimentação entre as posições

Nesta seção foram detectados os movimentos entre as posições citadas na Tabela 13, havendo uma porcentagem mais alta na movimentação do paciente de pé por ter uma diferença maior entre os seus *keypoints*, calculados *frame a frame*. Nas outras posições, o resultado foi baixo pela razão dos testes terem sido feitos para cada posição com o mínimo de movimento possível, para conseguir diferenciar o sentado do levantado. Esta diferença ficou ainda mais visível quando o realmente esta de pé se movimentando.

Posições	Porcentagem de detecção de movimento
Levantando (com poucos movimentos)	04,73%
Levantando (normalmente)	07,33%
Movimentando (de pé)	29,25%
Sentado	01,80%
Sentando	08,49%

Tabela 13 – Detecção de movimento. (Fonte: autores)

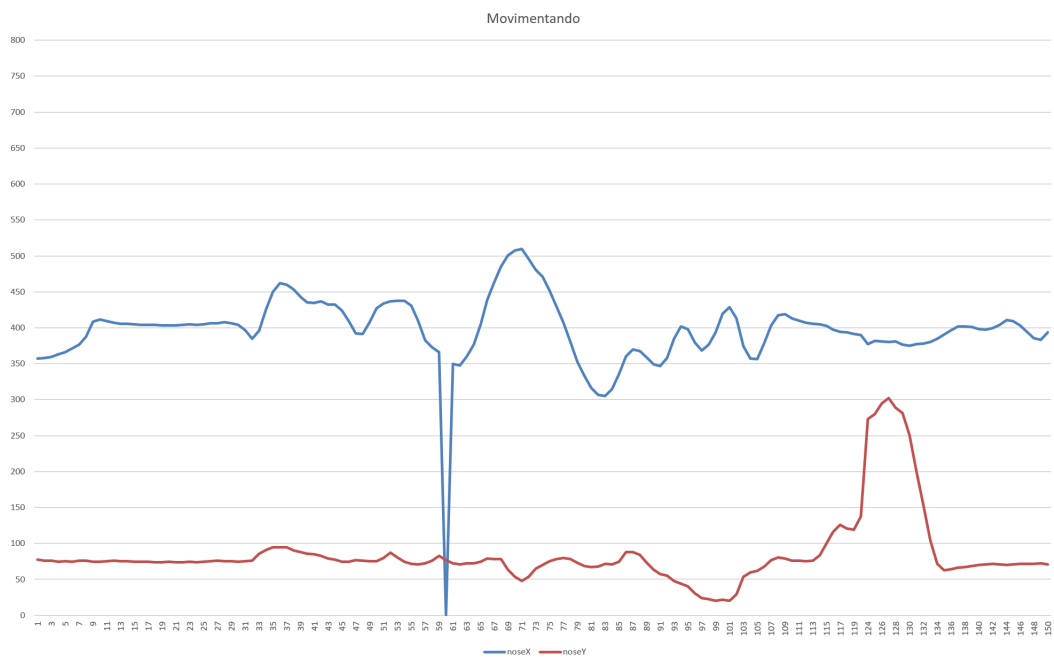


Figura 21 – Gráfico de movimentação em geral. Eixo x : frames, eixo y : keypoints. (Fonte: autores)

Importante notar que na Figura 21, representando a movimentação do paciente, no *frame* 60 o valor Y do gráfico cai para zero. Isso se dá pela razão do *keypoint* não ter sido detectado, fazendo assim o *OpenPose* zerar nesse determinado ponto. Em outros testes de movimentos, isso ocorre com bastante frequência, visto que alguns pontos podem não estar visíveis na câmera ou a sua acurácia estar baixa.

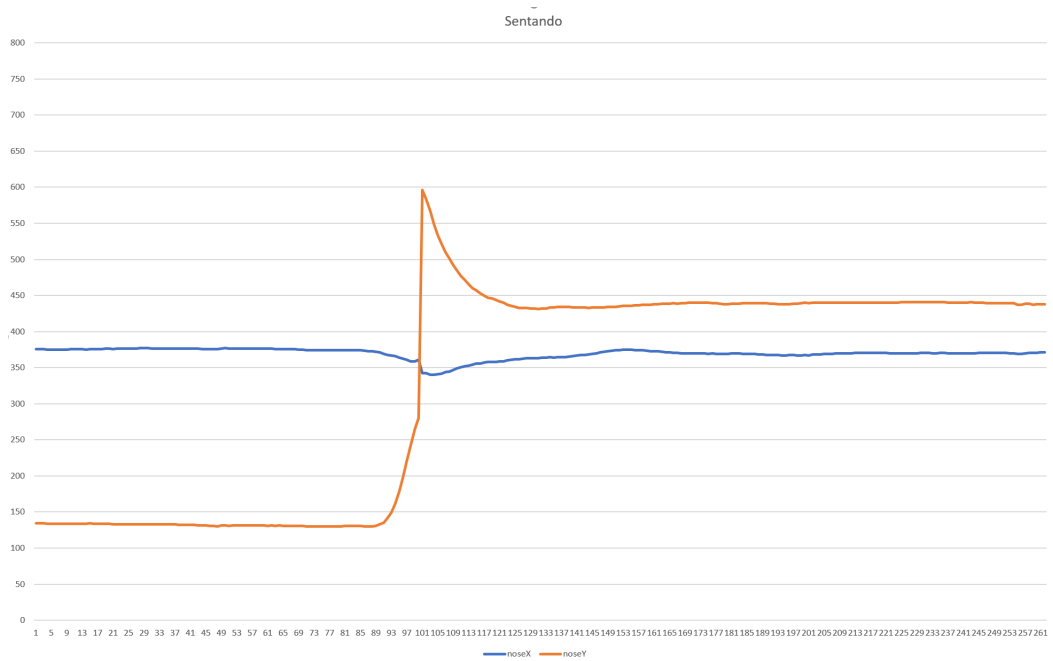


Figura 22 – Gráfico de movimentação - Sentando. Eixo x : *frames*, eixo y : *keypoints*.
(Fonte: autores)

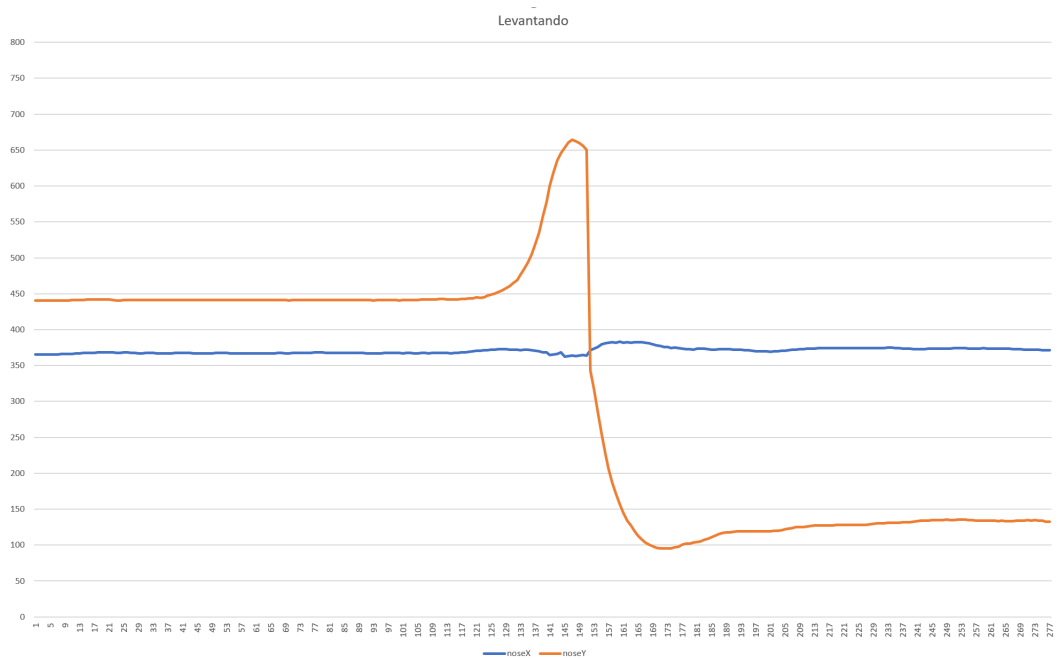


Figura 23 – Gráfico de movimentação - Levantando. Eixo x : *frames*, eixo y : *keypoints*.
(Fonte: autores)

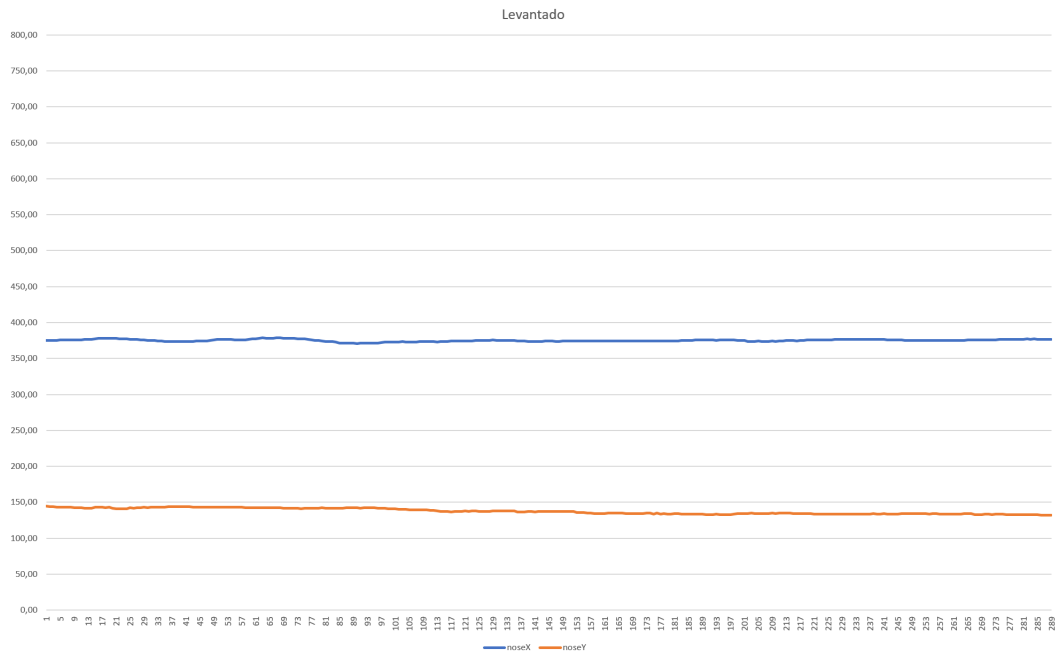


Figura 24 – Gráfico de movimentação - Em pé. Eixo x : *frames*, eixo y : *keypoints*. (Fonte: autores)

Nas Figuras 21, 22, 23 e 24, foram representadas as variações dos deslocamentos dos *keypoints* da cabeça nos eixos X e Y do plano cartesiano, considerando movimentação geral, sentando, levantando e em pé, respectivamente. Estas representações foram identificadas em um dos testes de reconhecimento dos *keypoints* do corpo humano. Após o processo de obtenção dos *keypoints*, conforme descrito na Seção 5.3.1, eles foram plotados de forma a representar a variação de cada ponto de acordo com a passagem dos *frames* e dessa forma, com uma análise gráfica é possível identificar o padrão de deslocamento do *keypoint* e fazer a identificação do movimento que está sendo realizado. No caso das figuras supracitadas, foi possível identificar a variação em ambos os eixos, indicando assim que o corpo estava em movimento.

6.3 COMPOSIÇÃO DAS INFORMAÇÕES

Nesta seção foi mostrado o resultado da composição das informações dos dados, com o reconhecimento de movimento usando o modelo bayesiano, da Equação 5.

Na Tabela 14, foram utilizados os menores valores contidos nos dados com os movimentos das posições. O resultado foi o menor caso possível com todas as variáveis dispostas nas informações do Hospital Moinho de Vento.

Menores valores dos dados	Posições	Probabilidade de queda
0,868579	Levantando (com poucos movimentos)	87,48%
0,868579	Levantando (normalmente)	87,82%
0,868579	Movimentando (de pé)	90,70%
0,868579	Sentado	87,09%
0,868579	Sentando	87,97%

Tabela 14 – Menores valores dos dados com os movimentos entre as posições e probabilidade de queda. (Fonte: autores)

O resultado indicou uma diferença pequena entre os movimentos. Porém, qualquer movimento entre estas posições do paciente tem uma probabilidade alta de queda. Mesmo no caso do paciente estar sentado, a probabilidade foi de 87,09%, indicando que o mesmo tem uma alta chance de cair, havendo aumento desta probabilidade conforme sua movimentação.

Média dos Pacientes	Posições	Probabilidade de queda
0,89125	Levantando (com poucos movimentos)	89,64%
0,89125	Levantando (normalmente)	89,92%
0,89125	Movimentando (de pé)	92,30%
0,89125	Sentado	89,32%
0,89125	Sentando	90,04%

Tabela 15 – Média dos pacientes (sem severidade). (Fonte: autores)

Na Tabela 15 foi utilizada a média dos valores relacionados aos pacientes, como: idade, sexo e ferimento depois da queda. Neste caso, houve a ausência do fator severidade, que é um dado referente ao resultado da queda. Esse dado não possui muita variedade, gerando alta interferência no resultado final da probabilidade. O resultado da probabilidade de queda neste caso é satisfatório, levando em conta um resultado mínimo de 89,32% e máximo de 92,30%.

Média das informações médicas	Posições	Probabilidade de queda
0,97931	Levantando (com poucos movimentos)	98,02%
0,97931	Levantando (normalmente)	98,08%
0,97931	Movimentando (de pé)	98,53%
0,97931	Sentado	97,96%
0,97931	Sentando	98,10%

Tabela 16 – Média das informações médicas. (Fonte: autores)

Na Tabela 16 foi utilizada a média dos valores das informações médicas, como: risco de queda, protocolo de risco de queda, medicação, razão do acidente, turno e dia. O resultado foi satisfatório, variando de 97,96% até 98,53%.

Hospitalar	Posições	Probabilidade de queda
0,87785	Levantando (com poucos movimentos)	88,36%
0,87785	Levantando (normalmente)	88,68%
0,87785	Movimentando (de pé)	91,35%
0,87785	Sentado	88,00%
0,87785	Sentando	88,82%

Tabela 17 – Média das informações hospitalares. (Fonte: autores)

Na Tabela 17 foi utilizada a média dos valores das informações hospitalares, como: setor onde ocorreu o acidente e ambiente. O resultado foi satisfatório, obtendo um mínimo de 88,00% e um máximo de 91,35%.

7 CONCLUSÃO

Considerando o cenário atual dos hospitais e estudos direcionados a problemas hospitalares, determinar o risco de uma queda pôde envolver diversos fatores, como posição específica do esqueleto humano, movimentos realizados e características que, se correlacionadas, põe o paciente a uma situação mais propícia de queda, como severidade de patologia diagnosticada e idade, por exemplo.

Diante deste contexto, neste trabalho foi proposta uma metodologia de 8 etapas capaz de analisar dados de livre acesso e extrair informações visuais importantes para definir uma queda e uma não queda. Através de informações médicas dos pacientes, obtidas através do trabalho de Souza et al. (2019), pôde-se entender melhor quais os pontos chaves para a decisão e a probabilidade de uma queda com o cálculo dos *keypoints* que foram pré-determinados para os testes de movimentos do corpo humano. Os intervalos adquiridos através do uso do modelo bayesiano nos dados do Hospital Moinhos de Vento foram de 86.85% a 100.00%, e indicaram uma alta acurácia na medição das probabilidades de risco de queda, enquanto os cálculos utilizando os *keypoints* mostraram uma diferença significativa entre os movimentos, principalmente com o paciente se mexendo. Na composição dos dados com os movimentos, foi possível ver uma probabilidade alta de queda mesmo em um paciente com baixo risco.

No entanto, para executar o experimento da análise de dados, seriam necessárias bases de dados maiores, pois quanto mais dados e maior variedade de informações entre eles, maior a influência na porcentagem final. Pôde-se notar que a diferença entre os resultados dos dados foi pequena e, durante o experimento, foi feita uma busca pelos maiores e menores valores possíveis com os dados de forma aleatória, em conjuntos como pacientes, informações médicas, estrutura hospitalar e níveis de risco de queda.

A principal contribuição desse trabalho para a comunidade científica foi um passo a mais no caminho já traçado por Souza et al. (2019) na obtenção de um padrão específico para uma possível queda, de modo a desenvolver estratégias e ferramentas para evitar que uma queda ocorra.

Portanto, a escolha por esse estudo foi satisfatória visto que houve um avanço no grau de confiabilidade em futuras implementações visuais de predição e prevenção de queda, possibilitando a criação de uma interface barata em sistemas independentes de monitoramento hospitalares, reduzindo quedas e desonerando profissionais de saúde.

REFERÊNCIAS

- ABBATE, S. et al. A smartphone-based fall detection system. **Pervasive Mob. Comput.**, v. 8, p. 883–899, 2012.
- ABHISHEKVAISH; SAIFAHMED; ANANSHETTY. Remote physiotherapy monitoring using the novel d+r therapy iphone application. **Journal of clinical orthopaedics**, v. 4 pages, 2016.
- ABOBAKR, A.; HOSSNY, M.; NAHAVANDI, S. A skeleton-free fall detection system from depth images using random decision forest. **IEEE Systems Journal**, v. 12, p. 2994–3005, 2018.
- AL-KHAFAJIY, M. et al. Remote health monitoring of elderly through wearable sensors. **Multimedia Tools and Applications**, p. 24681–247065, 2019.
- ANTON, D. et al. A telerehabilitation system for the selection, evaluation and remote management of therapies. **Department of Electrical Engineering**, v. 21 pages, 2018.
- AUVINET, E. et al. **Multiple cameras fall dataset**. DIRO - Université de Montréal, 2010. Disponível em: <<http://www.iro.umontreal.ca/~labimage/Dataset/>>.
- BROWNLEE, J. **A Gentle Introduction to Computer Vision**. 2019. Disponível em: <<https://machinelearningmastery.com/what-is-computer-vision/>>.
- CHEN, C.-H.; RAMANAN, D. 3d human pose estimation = 2d pose estimation + matching. **2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, p. 5759–5767, 2017.
- CHEN, W. et al. Fall detection based on key points of human-skeleton using openpose. **Symmetry**, v. 17 pages, 2020.
- CHIAROVANO, E. et al. Maintaining balance when looking at a virtual reality three-dimensional display of a field of moving dots or at a virtual reality scene. **Frontiers in Neurology**, v. 6, 2015.
- CHOE, E. K. et al. Semi-automated tracking: A balanced approach for self-monitoring applications. **IEEE Pervasive Computing**, v. 16, p. 74 – 84, 2017.
- CROSSLEY, G. H. et al. The connect (clinical evaluation of remote notification to reduce time to clinical decision) trial: the value of wireless remote monitoring with automatic clinician alerts. **Journal of the American College of Cardiology**, 2011.
- DAHER, M. et al. Elder tracking and fall detection system using smart tiles. **IEEE Sensors Journal**, p. 11, 2016.

- ESFAHLANI, S. S.; WILSON, G. Development of rehabilitation system (rehabgame) through monte-carlo tree search algorithm using kinect and myo sensor interface. **Computing Conference 2017 18-20 July 2017 | London, UK**, p. 1021–1028, 2017.
- FIDALGO-HERRERA, A. et al. Artificial intelligence prediction of the effect of rehabilitation in whiplash associated disorder. **PLoS ONE**, v. 15, 2020.
- GJORESKI, H.; LUTREK, M.; GAMS, M. Accelerometer placement for posture recognition and fall detection. **2011 Seventh International Conference on Intelligent Environments**, p. 47–54, 2011.
- HAN, J. et al. Enhanced computer vision with microsoft kinect sensor: A review. **IEEE Transactions on Cybernetics**, v. 43, n. 5, p. 1318–1334, 2013.
- HE, K. et al. **Mask R-CNN**. 2018. Disponível em: <arXiv:1703.06870v3>.
- HTIKE, K. K.; LAI, W. K.; KHALIDA, O. O. Human posture recognition: Methodology and implementation. **Journal of Electrical Engineering and Technology**, 2011.
- HUANG, X. et al. **A Posture Recognition Method Based on Indoor Positioning Technology**. 2019. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6471703/>>.
- IBSP. **Prevenção de queda: um case de sucesso em um hospital paulistano**. 2019. Disponível em: <<https://www.segurancadopaciente.com.br/qualidade-assist/prevencao-de-queda-um-case-de-sucesso-em-um-hospital-paulistano/>>.
- KIANOUSH, S. et al. Device-free rf human body fall detection and localization in industrial workplaces. **IEEE Internet of Things Journal**, p. 11, 2016.
- LE, J. **The 5 Computer Vision Techniques That Will Change How You See The World**. 2018. Disponível em: <<https://heartbeat.fritz.ai/the-5-computer-vision-techniques-that-will-change-how-you-see-the-world-1ee19334354b>>.
- LEE, M. H. et al. An exploratory study on techniques for quantitative assessment of stroke rehabilitation exercises. **Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization**, 2020.
- LIU, H. et al. **Gesture Classification with Low-Cost Capacitive Sensor Array for Upper Extremity Rehabilitation**. 2019. 4 p.
- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. **arXiv:1411.4038**, 2015.
- MEHTA, D. et al. Vnect: Real-time 3d human pose estimation with a single rgb camera. **ArXiv**, abs/1705.01583, 2017.

NIELSEN; MICHAEL. **Neural Networks and Deep Learning**. 2019. Disponível em: <<http://www.neuralnetworksanddeeplearning.com>>.

OFLI, F. et al. Sequence of the most informative joints (smij): A new representation for human skeletal action recognition. **2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops**, 2012.

PIRAINTORN, P.; SA-ING, V. Stroke rehabilitation based on intelligence interaction system. **2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)**, p. 648–651, 2020.

RHUMA, A.; YU, M.; CHAMBERS, J. **Posture Recognition Based Fall Detection System**. 2013. Disponível em: <<https://ieeexplore.ieee.org/document/6279483>>.

SCHECHTER, S. **What is gesture recognition? Gesture recognition defined**. 2014. Disponível em: <<https://www.marxentlabs.com/what-is-gesture-recognition-defined/>>.

SCHOOT, R. van de et al. Bayesian statistics and modelling. **NATURE**, 2021.

SOUZA, A. B. de et al. In hospital falls of a large hospital. **BMC Research Notes**, v. 3 pages, 2019.

TRAN, C.; TRIVEDI, M. 3-d posture and gesture recognition for interactivity in smart spaces. **IEEE Transactions on Industrial Informatics**, v. 8, p. 178–187, 2012.

TSINGANOS, P.; SKODRAS, A. A smartphone-based fall detection system for the elderly. **Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis**, p. 53–58, 2017.

VOURGANAS, I.; STANKOVIC, V.; STANKOVIC, L. Individualised responsible artificial intelligence for home-based rehabilitation. **MDPI, Basel, Switzerland.**, p. 25, 2020.

WANG, Z. et al. Using wearable sensors to capture posture of the human lumbar spine in competitive swimming. **IEEE Transactions on Human-Machine Systems**, v. 49, p. 194 – 205, 2019.

YANG, G. et al. An iot-enabled stroke rehabilitation system based on smart wearable armband and machine learning. **IEEE Journal of Translational Engineering in Health and Medicine**, v. 6, p. 1–10, 2018.

ZERROUKI, N.; HOUACINE, A. Combined curvelets and hidden markov models for human fall detection. **Springer Science+Business Media New York 2017**, p. 6405–6424, 2017.

ZHANG, Z. Microsoft kinect sensor and its effect. **IEEE Multimedia - IEEEEMM**, v. 19, p. 4–10, 02 2012.