

## Aplicação de Algoritmos de Aprendizagem Profunda com Foco na Detecção de Quedas Através da Classificação de Imagens Térmicas

Rebeca M. Silva\*. Elias T. da Silva, Jr.\*\*

\**Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Campus Fortaleza.  
Brasil (Tel: (85) 9 9708-5563; e-mail: rebeca.matos.silva05@aluno.ifce.edu.br).*

\*\**Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Campus Fortaleza.  
Brasil (Tel: (85) 9 9158-2332; e-mail: elias@ifce.edu.br).*

---

**Abstract:** Falls tend to cause physical and psychological damage, especially for people over 65 years of age. Such damage can be aggravated if they live alone and do not get help in the right time. This is one of the motivations for investigating smart devices in homes, acting as a way of monitoring some action or incident. Machine learning techniques have been applied to increase the effectiveness of these devices by looking for patterns in audio signals and in images. When using DNN (Deep Neural Networks) the success rates increase, but these solutions usually use cloud processing, which can bring a feeling of insecurity to the user. This work investigates an environment monitoring system, capable of detecting fall incidents, without invading the user's privacy, as it is based on low resolution thermal images. Some DNN models were tested to choose a model that best fits the problem, looking for architectures with few processing layers so that they can be embedded for local processing. Images from two different sensors were used, resized to the standard size of 64x64 *pixels*, resulting in an average accuracy of 91.66% for the bidirectional LSTM (Long short-term memory) network, and of 97.88 % using the VGG16 network.

**Resumo:** A queda tende a gerar danos físicos e psicológicos, principalmente para as pessoas com idade acima dos 65 anos. Tais danos podem ser agravados caso as mesmas morem sozinhas e não consigam ajuda no tempo adequado. Esta é uma das motivações para se investigar dispositivos inteligentes nas residências, atuando como forma de monitorar alguma ação ou incidente. Técnicas de aprendizagem de máquina têm sido aplicadas para aumentar a eficácia desses dispositivos, procurando por padrões em sinais de áudio e em imagens. Quando se recorre às DNN (*Deep Neural Networks*) as taxas de sucesso aumentam, mas essas soluções costumam utilizar processamento em nuvem, o que pode trazer uma sensação de insegurança ao usuário. Este trabalho investiga um sistema de monitoramento do ambiente, capaz de detectar incidentes de queda, sem invadir a privacidade do usuário, por ser baseado em imagens térmicas de baixa resolução. Alguns modelos de DNN foram testados para escolha de um modelo que melhor se adequa ao problema, buscando arquiteturas com poucas camadas de processamento para que possam ser embarcadas para processamento local. Foram utilizadas imagens de dois sensores distintos, redimensionadas para o tamanho padrão de 64x64 *pixels*, resultando em uma acurácia média de 91,66% para a rede LSTM (*Long short-term memory* ou memória de longo prazo) bidirecional, e de 97,88% utilizando a rede VGG16.

**Keywords:** Thermal imaging; Image Classification; Deep Learning; Bi-LSTM; VGG16.

**Palavras-chaves:** Imagens térmicas; Classificação de Imagens; Aprendizado Profundo; Bi-LSTM; VGG16.

---

### 1. INTRODUÇÃO

De acordo com a *United Nations* (2017), em 2030 a população mundial com idade acima de 60 anos atingirá de cerca de 1,4 bilhão de pessoas, e de acordo com o *World Health Statistics* (2008), 28% a 35% das pessoas com idade acima de 65 anos caem ao menos uma vez ao ano, sendo que este percentual pode chegar a 42% se considerarmos o grupo de pessoas com idade acima de 70 anos.

Baseado na Fig. 1, extraída do relatório anual de 2015 das Nações Unidas, é possível observar que a população idosa (com 60 anos ou mais) terá um crescimento maior do que a população com idade entre 15 e 59 anos. Se a tendência continuar, não haverá pessoas suficientes para cuidar dos idosos em um futuro distante, portanto, tecnologias de vida assistida serão necessárias no futuro para cuidar de pessoas idosas e ajudá-las a viver de forma independente (Uddin, Khaksar and Torresen, 2018).

Além da real necessidade de se ter um acompanhante, outro fator importante a se observar no que tange os cuidados a população idosa é a independência, que muitas vezes lhe é privada por excesso de cuidados de familiares. O medo de cair e não conseguir ter uma devida recuperação posterior também amplifica as consequências negativas do pós-queda pode diminuir a confiança do paciente (Friedman *et al.*, 2002). Tais situações acarretam na limitação do paciente, redução na interação social e depressão.

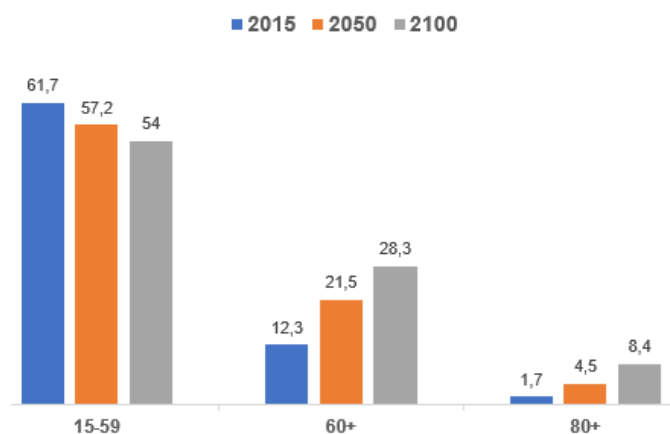


Fig. 1: Crescimento da população idosa

É possível encontrar diversas soluções que visam realizar o monitoramento de atividades e emitir alertas, caso necessário, dando uma sensação de segurança ao usuário. Conforme cita Mubashir (2013), tais soluções são geralmente separadas em 3 grupos: sistemas baseados em dispositivos vestíveis, sistemas que capturam informações obtidas a partir de sensores instalados no ambiente e câmeras de monitoramento. Os sistemas baseados em dispositivos vestíveis têm sua eficácia limitada ao uso direto do dispositivo por parte do paciente, ou seja, caso o mesmo venha a esquecer de acoplar a solução ao seu corpo tal dispositivo passa a não ter eficiência. Sensores instalados no ambiente possuem seu campo de atuação limitado ao local de instalação, assim como as câmeras de monitoramento, esta por sua vez, podendo se tornar desagradável por invadir a privacidade do usuário e inviabilizar seu uso em compartimentos mais restritos da residência, como quartos e banheiros.

Uma das formas de realizar a análise das imagens obtidas através dos sensores térmicos é através de algoritmos de aprendizagem de máquina. Através de um processo chamado treinamento o algoritmo de aprendizagem de máquina cria um modelo baseado nos dados fornecidos. Novos dados são passados por esse modelo para fazer previsões, em um processo chamado de inferência. As soluções mais comuns costumam usar o dispositivo embarcado para capturar os dados e realizam a inferência em um servidor em nuvem, tornando a conexão à internet uma característica relevante a ser observada nas soluções microcontroladas apresentadas. A figura 2 apresenta um exemplo das diferentes plataformas que podem executar o processamento em cada etapa de um processo de aprendizagem de máquina.

Dentre as vantagens de se realizar um processamento de dados locais, pode-se destacar a confiabilidade, a privacidade

e baixo consumo de energia. Apesar disso, as limitações físicas de uma plataforma embarcada impõem limites no desenvolvimento de soluções. Tais plataformas costumam ser inflexíveis no tamanho, memória, capacidade de processamento, potência, temperatura, durabilidade e, principalmente, custo (Nadeski, 2019).

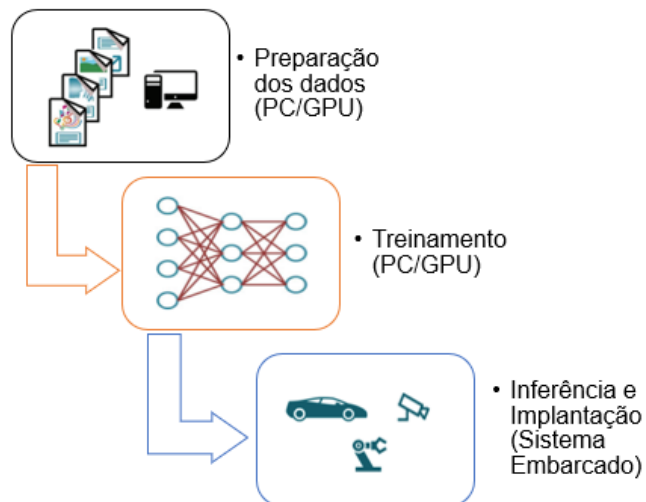


Fig. 2: Fluxo de desenvolvimento de uma solução com aprendizado profundo tradicional

Neste trabalho, será feito um estudo de técnicas que podem ser utilizadas no desenvolvimento de uma solução baseada em sensores térmicos de baixa resolução para detecção de quedas, realizando a identificação do acidente através de algoritmos de aprendizado profundo, com poucas camadas de abstração, visando a possibilidade de realizar inferências localmente, dispensando o envio das informações capturadas para um servidor externo via internet.

## 2. TRABALHOS RELACIONADOS

O estudo realizado por Uddin, Khaksar e Torresen (2018) apresenta um comparativo de diversas soluções baseada no uso de sensores para monitoramento de idosos. Em relação aos sensores vestíveis o estudo destaca sensores que medem parâmetros como aceleração, velocidade, força magnética, frequência cardíaca, temperatura corporal, saturação de oxigênio, frequência respiratória, eletrocardiograma e etc, e aponta como principal desvantagens de tais sensores a dificuldade de instalação, que muitas vezes requer a ajuda de um profissional. Por outro lado, os sensores externos são instalados no ambiente, não afetando a rotina do usuário, e coletam vários tipos de dados para modelar os eventos ou atividades, como quedas, localização, maneira de andar, conseguindo antecipar as necessidades do usuário e melhorar a qualidade de vida.

Outro fator importante está relacionado a privacidade do usuário. Soluções que realizam o monitoramento das atividades diárias e aspectos de saúde dos usuários precisam manter os dados capturados sob sigilo. Os serviços de saúde exigem que os dados sejam armazenados em segurança,

podendo ser disponibilizados apenas para profissionais de saúde, parentes ou o próprio usuário. Câmeras de monitoramento são utilizadas em algumas situações e se destacam por apresentarem uma visão mais clara da situação, mas acabam não sendo utilizadas por conta da violação de privacidade, e se tornam inviável para uso em ambientes privados, como quarto e banheiro (Noury *et al.*, 2007).

Como forma de preservar a segurança e privacidade dos usuários, o uso de sensores térmicos de baixa resolução é cada vez mais aceito, detectando postura e movimentação dos usuários no ambiente. A Tabela 1 apresenta um resumo dos trabalhos similares encontrados na literatura e seus respectivos resultados.

### 2.1. Soluções utilizando sensores de matriz quadrada

Uma das primeiras soluções encontradas foi proposta por Sixsmith e Johnson (2004), usando o IRISYS (*InfraRed Integrated Systems*), que gera uma imagem de 16x16 pixels. A Fig. 2 apresenta uma comparação de uma imagem extraída de uma câmera de monitoramento e a imagem térmica gerada pelo sensor IRISYS. O estudo propõe a solução embarcada denominada SIMBAD (*Smart Inactivity Monitor using Array-Based Detectors*), que identifica a queda a partir da velocidade vertical calculada baseada nos dados coletados pelo sensor. Em testes realizados em laboratório, a detecção da queda só ocorreu em 30% dos cenários de queda simulados. O mau resultado é atrelado pelo autor a uma má generalização dos dados de treinamento, já que nos testes o sensor foi posicionado de forma a capturar um ângulo diferente do ângulo das imagens obtidas para treinamento do modelo.

Outro sensor encontrado na literatura é a série de sensores de matriz infravermelho Grid-EYE AMG88, fabricado pela empresa Panasonic, que produz uma matriz de 64 pixels dispostos em 8x8. Em seu estudo, Liu *et al.* (2020) utiliza o sensor AMG8853, destacando como vantagens desse tipo de solução o baixo custo, a preservação da privacidade, a baixa complexidade, a detecção em tempo real e a alta acurácia obtida. Seu sistema considera a captura dos dados em tempo real e transferência para um host. Utilizando o algoritmo *Random Forest* os autores apresentaram uma acurácia de 94% em um sistema de classificação binária, queda e não queda.

### 2.2 Soluções utilizando sensores de matriz retangular

Um dos sensores muito utilizado em trabalhos similares, é a série de sensores MLX906, da Melexis. Uma das soluções encontradas é proposta por Wei-Han Chen e Hsi-Pin Ma (2015), utilizando o MLX90620, que gera uma imagem com resolução de 16x4 pixel. Nesse estudo, os autores propõem o uso de 2 sensores no ambiente, posicionados a 3.3m de distância horizontal entre eles, como forma de obter uma visão mais clara do ambiente, gerando uma imagem de 32x8 pixel. Utilizando o algoritmo classificador K-NN (*Nearest Neighbors*) os resultados apontaram que o sistema consegue

distinguir um evento de queda com sensibilidade de 95,25%, especificidade de 90,75% e acurácia total de 93%.

Outra solução encontrada é proposta por Tateno *et al.* (2020), utilizando o sensor MLX90640, que gera uma imagem com resolução de 32x24 pixel. Para efetuar a classificação, nesse trabalho, o autor utiliza dois métodos como forma de realizar um comparativo dos resultados: 3D-CNN (*Convolutional Neural Network*) e LSTM (*Long Short-Term Memory*), apresentando 94,2% e 82,5% de acurácia, respectivamente.

**Tabela 1. Trabalhos similares e resultados encontrados**

Autor	Resolução (pixel)	Método de Classificação	Acurácia
Sixsmith e Johnson (2004)	16x16	MLP	30,0%
Mashiyama, Hong e Ohtsuki (2015)	8x8	SVM	100,0%
Wei-Han Chen e Hsi-Pin Ma (2015)	4x16	Classificador k-NN	93,0%
Hayashida, Moshnyaga e Hashimoto (2017)	8x8	Limiar de Distância	94,0%
Riquelme <i>et al.</i> (2019)	1x8/ 24x32	Bi-LSTM	93,0%
Liu <i>et al.</i> (2020)	8x8	Random Forest	94,0%
Tateno <i>et al.</i> (2020)	24x32	LSTM / 3D-CNN	82,5% / 94,2%

## 3. METODOLOGIA

Nesta seção será descrito o processo adotado para obtenção de um modelo que realize a classificação das imagens térmicas. A Fig. 3 apresenta um diagrama da metodologia utilizada.

Inicialmente foi realizado buscas na literatura para se basear nos modelos que estão sendo utilizados como solução em problemas similares, assim como, obter um *dataset* adequado para treinamento dos modelos de aprendizagem de máquina escolhidos. Dois modelos serão treinados de forma a possibilitar uma comparação dos resultados e definição do melhor modelo. As imagens escolhidas para treinamento do modelo devem passar por um pré-processamento, de forma a utilizar técnicas que facilitem o processo de classificação.

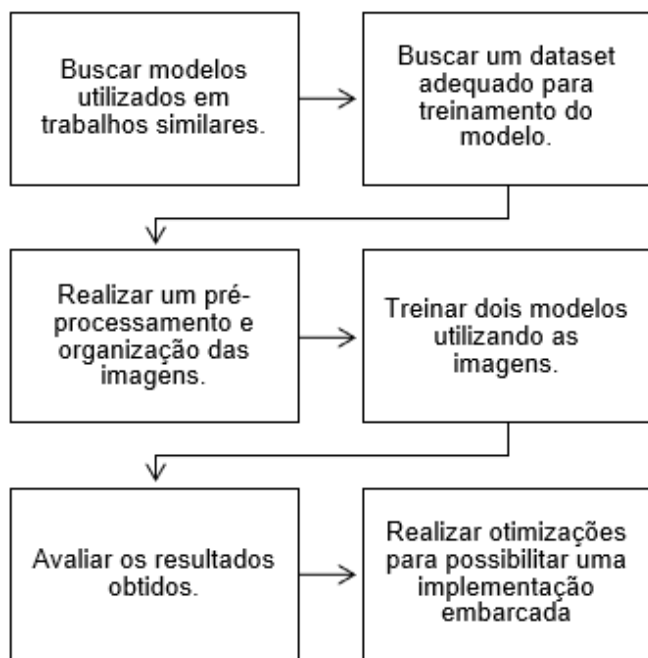


Fig. 3: Diagrama da metodologia utilizada

### 3.1 Dataset Utilizado

Riquelme et al. (2019) apresenta um trabalho voltado especificadamente para a montagem de 2 *datasets* que pudessem vir a ser utilizados em trabalhos futuros com abordagem similar. Foram utilizados dois modelos de sensores térmicos, conforme descrito na Tabela 2. O sensor MLX90640, da Melexis, com resolução de imagem de 32x24 *pixels*, e o sensor D6T-8L-06, da Omron Eletronics, com resolução de 1x8 *pixels*. Ambos sensores foram posicionados na parede, sendo o sensor MLX90640 a uma altura de 1,2m, enquanto os sensores D6T-8L-06 foram posicionados dois a 10cm do chão e outros dois a 1m, formando uma imagem de 2x16 *pixel*.

**Tabela 2. Sensores utilizados na criação do dataset.**

Sensor	Resolução ( <i>pixel</i> )	Frame Rate	Qtd. de sensores utilizados
MLX90640	32x24	16fps	1
D6T-8L-06	1x8	5fps	4

Dois grupos de voluntários participaram da captura das imagens, o grupo 01 formados por artistas performáticos que receberam instruções de como cair e o grupo 02 que não recebeu nenhuma instrução, sendo estes de gênero, idade, peso e altura distintas. O diferencial desses dados é que eles foram construídos baseado em dois grupos de voluntários: um contendo pessoas normais e outro contendo artistas performáticos orientados por fisioterapeutas para simularem idosos. O *dataset* é formado por 180 arquivos em formato .csv, um para cada tipo de queda. A Fig. 4 e a Fig. 5 apresentam um exemplo de imagem de cada sensor.

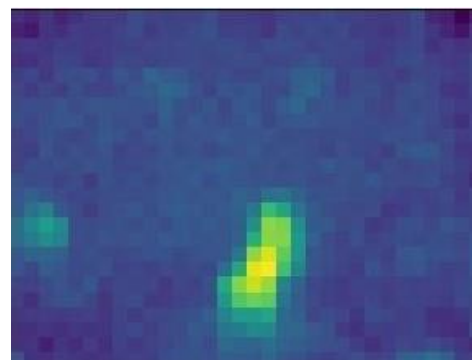


Fig. 4: Imagem 24x32: um sensor MLX90640



Fig. 5: Imagem 2x16: quatro sensores D6T-8L-06

As imagens do *dataset* passaram por um processo de redimensionamento para um tamanho adequado a arquitetura que será treinada, conforme explicado abaixo. Os modelos devem ser treinados utilizando os dois *datasets*, de forma distinta.

### 3.2 Pré-processamento dos dados

Por conta da limitação do tamanho de imagem solicitada pelos modelos de classificação utilizados, as imagens passaram por processo de ampliação, tanto usando o algoritmo de *Resize*, da própria biblioteca OpenCV, quanto usando o algoritmo HQX (Filtro de ampliação de alta qualidade), desenvolvido por Maxim Stepin (Kopf and Lischinski, 2011).

No caso das imagens obtidas pelo sensor MLX90640, por apresentarem mais detalhes de contorno do objeto detectado, observou-se que ao ser utilizada a função *Resize*, ocorria a perda de detalhes da imagem. Portanto, para evitar a perda de tais detalhes, visando inclusive uma melhoria dos contornos do objeto, foi utilizado o algoritmo HQX para ampliação dessas imagens. Na Fig. 6 se pode observar a diferença da imagem obtida originalmente com o sensor e suas ampliações com *Resize* e HQX.

No caso das imagens extraídas do sensor D6T-8L-06, por serem imagens com menor resolução e com um padrão mais simples, o redimensionamento foi realizado primeiramente por duplicação das suas linhas, de forma a deixar a imagem mais simétrica e, em seguida, aplicado a função *Resize* da biblioteca OpenCV, deixando a imagem com o tamanho mínimo recomendado para treinamento na arquitetura da VGG16, 32x32 *pixels*. A Fig. 7 apresenta a evolução da ampliação das imagens.

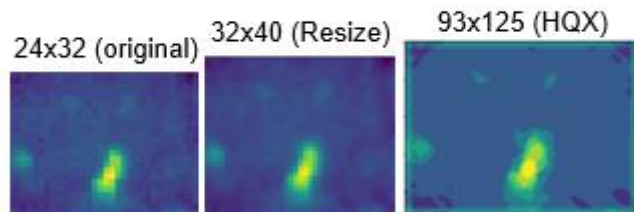


Fig.6: Ampliação de imagens do sensor MLX90640

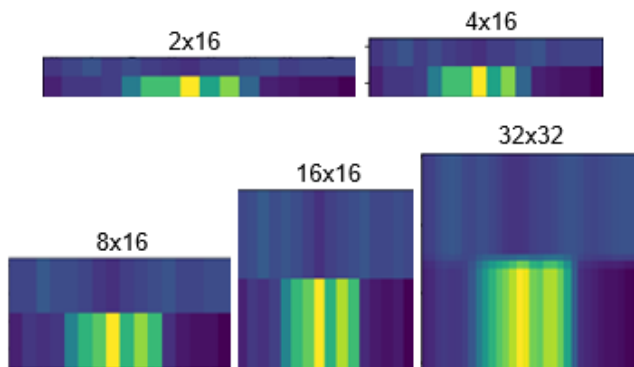


Fig. 7: Ampliação de imagens do sensor D6T-8L-06

### 3.3 Treinamento das arquiteturas

Após realizado o pré-processamento das imagens a serem utilizadas, as imagens foram manualmente separadas em duas classes: ocorreu queda e não ocorreu queda. A base de dados final utilizada para treinamento foi composta de 2500 imagens, sendo divididas igualmente entre as classes determinadas, e 20% das imagens foram utilizadas para validação.

As arquiteturas treinadas foram a Bi-LSTM e a VGG16. A motivação para realizar o treino com a arquitetura Bi-LSTM partiu do seu uso no trabalho de Riquelme et. Al (2019), onde se obteve acurácia de 93,0%. Já no caso da VGG16, foi escolhida a realização de testes com a mesma, pois é uma rede amplamente conhecida, com pesos que podem ser reutilizados, diminuindo o tempo de treinamento da rede.

#### 3.3.1 Bi-LSTM

A arquitetura do modelo conta inicialmente com uma camada de convolução, onde filtros recebem o vetor de imagens na entrada com a finalidade de extrair características hierárquicas que facilitem a classificação. Esta camada tem como resultado um vetor de 1 dimensão com tais características. Em seguida, a camada de *Maxpooling* recebe este vetor e reduz o número de características encontradas, deixando as mais importantes, que serão repassadas para a camada Bi-LSTM. Para finalizar, têm-se a camada densa, que consiste em uma rede neural do tipo *perceptron* usada para obter uma saída entre 0 e 1, seguida da camada de ativação.

A camada de convolução unidimensional conta com um kernel linear de valor 3, e extrai características dos *pixels* capturados, que são passados em forma de vetor (o tamanho

do vetor de entrada é ajustado de acordo com o tamanho da imagem redimensionada). Usando esta estrutura, a próxima camada extrai informações mais precisas que serão usadas para estabelecer os elementos essenciais que podem ser processados em formato temporal, reduzindo o processamento computacional que é realizado na camada recorrente. Esta camada é implementada com a função de ativação ReLU, pois a mesma possui características que lhe permitem reduzir os efeitos do *vanishing gradient*, alcançando um resultado mais rápido.

A camada de *Maxpooling* também conta com 1 dimensão e possui tamanho de pool 2. Como forma de obter uma melhor capacidade de generalização, esta camada reduz o número de características encontradas pela camada de convolução, deixando apenas as de interesse para a rede.

A arquitetura Bi-LSTM possui uma estrutura similar a LSTM, porém realiza o reconhecimento das sequências temporais nas duas direções. Já a camada de ativação usa uma função *sigmoid*, considerando que será realizado apenas uma classificação binária, identificando se houve ou não a queda.

Para compilação do modelo, como se trata de um problema de classificação binária, foi utilizado *binary\_crossentropy* como parâmetro de perda, otimizador RMSprop e acurácia como métrica.

#### 3.3.2 VGG16

Para realizar o treinamento com a arquitetura de rede da VGG16, as imagens foram redimensionadas para um tamanho padrão de 64x64 *pixels* e escala de cor cinza. Com o intuito de utilizar os pesos já conhecidos da *Imagenet* para a VGG16, a arquitetura da rede foi importada diretamente da biblioteca do Keras. Esse procedimento, conhecido é como *transfer learning* e ajuda a diminuir o tempo de treinamento da rede. Para isso é preciso tornar as camadas do meio em camadas não-treináveis e remover as últimas camadas. As camadas de saída que melhor se adequem a solução do nosso problema são então inseridas no modelo ajustado. Para compilação do modelo, como temos um problema de classificação binária, foi utilizado *binary\_crossentropy* como parâmetro de perda, otimizador Adam e acurácia como métrica.

## 4. RESULTADOS EXPERIMENTAIS

Através do processo de treinamento detalhado na seção anterior, foram obtidos os resultados apresentados na Tabela 3. As imagens obtidas através do sensor D6T-8L-06, possuem uma resolução menor e formam imagens mais triviais. É possível notar, a olho nu, que as imagens que apresentam o evento de queda apresentam, em sua maioria, o mesmo padrão: os *pixels* que preenchem a metade inferior da imagem possuem temperaturas mais elevadas do que os *pixels* que preenchem a metade superior. Dessa forma, torna-se fácil também para o modelo de aprendizado profundo

detectar o fenômeno, chegando a uma acurácia média de 97,88%, quando usado a arquitetura da VGG16.

**Tabela 3. Acurácia média obtida a partir do treinamento**

Arquitetura	Resolução da imagem original ( <i>pixel</i> )	Acurácia Média
Bi-LSTM	2x16	91,66%
	24x32	85,35%
VGG16	2x16	97,88%
	24x32	87,49%

No caso das imagens do sensor MLX90640, se consegue obter mais detalhes do ambiente, capturando detalhes do movimento e tornando as imagens mais complexas para análise. Utilizando a VGG16, obtêm-se a acurácia média de 87,49%. O valor mais alto de desvio padrão, observado na Tabela 4, aponta para uma dispersão maior dos resultados obtidos a cada época de treinamento.

**Tabela 4. Desvio Padrão obtido a partir do treinamento**

Arquitetura	Resolução da imagem original ( <i>pixel</i> )	Desvio Padrão
Bi-LSTM	2x16	3,95
	24x32	5,98
VGG16	2x16	3,11
	24x32	6,47

Os resultados foram observados durante o treinamento dos modelos, utilizando um *dataset* composto por 2500 imagens, sendo 1250 de cada classe (queda e não queda), com 20% das imagens separadas para validação. Os dados de validação foram utilizados como forma de limitar o número de épocas. Foram realizadas 5 rodadas de treinamento, sendo que a arquitetura VGG16 alcançava a convergência com um número bem menor de épocas, em torno de 100 enquanto com a Bi-LSTM ficou em torno de 500. Essa performance da VGG16 se dá em parte, devido a utilização dos pesos já treinados da Imagenet.

É importante notar que, mesmo a VGG16 apresentando uma acurácia média superior a acurácia média da rede Bi-LSTM, de acordo com o intervalo de acurácias, apresentados na Tabela 5, em algumas rodadas o modelo da Bi-LSTM alcançou resultados melhores que o da VGG16, mostrando assim ser uma arquitetura viável para esta aplicação. Vale lembrar que a rede VGG16 possui 16 camadas com pesos e possui aproximadamente 138 milhões de parâmetros, tornando mais difícil embarcar este modelo em um microcontrolador se comparado a Bi-LSTM, que conta apenas com 4 camadas.

**Tabela 5. Acurácia mínima e máxima obtida a partir do treinamento**

Arquitetura	Resolução da imagem original ( <i>pixel</i> )	Intervalo de Acurácia
Bi-LSTM	2x16	88,45% a 97,93%
	24x32	71,75% a 98,37%
VGG16	2x16	91,65% a 100%
	24x32	74,55% a 100%

Em relação a execução de inferências, utilizando o ambiente virtual Google Colab com auxílio de GPU, foram passadas 20 imagens para avaliar o comportamento do modelo e os resultados podem ser vistos na Tabela 6. É possível observar que, apesar dos tempos de execução serem bem próximos, o modelo VGG16 apresenta uma maior taxa de acertos na classificação.

**Tabela 6. Resultado da Inferência dos modelos**

Modelo	Resolução da imagem original ( <i>pixel</i> )	Tempo de CPU ( <i>user e sys</i> )	Tempo de relógio	Acurácia
Bi-LSTM	2x16	3 $\mu$ s	7,39 $\mu$ s	78,95%
	24x32	3 $\mu$ s	7,20 $\mu$ s	82,00%
VGG16	2x16	3 $\mu$ s	7,15 $\mu$ s	100,00%
	24x32	4 $\mu$ s	7,15 $\mu$ s	95,00%

## 5. CONCLUSÕES E TRABALHOS FUTUROS

Os resultados apresentados na seção anterior mostram que com o tratamento adequado nas imagens e um modelo de classificação baseado em redes neurais, tanto recorrentes quanto convolucionais, é possível obter um modelo capaz de realizar a detecção de quedas em um ambiente monitorado, de forma não invasiva.

É possível constatar que uma rede VGG16 apresenta uma boa performance durante a classificação, sendo o modelo melhor para implementação, porém a estrutura do modelo pode trazer dificuldades ao tentar implementar a solução em uma plataforma embarcada.

O modelo Bi-LSTM, apesar de apresentar, de maneira geral, resultados inferiores ao resultado do modelo VGG16, consegue ter resultados superiores em rodadas particulares e apresenta uma estrutura bem mais simples e propícia a implementação em um sistema embarcado.

Pretendemos realizar a otimização do modelo Bi-LSTM utilizando as ferramentas disponíveis através da biblioteca do TensorFlow Lite para microcontroladores, de forma a embarcar o modelo.

REFERÊNCIAS

- Friedman, S.M. et al. (2002) 'Falls and fear of falling: Which comes first? A longitudinal prediction model suggests strategies for primary and secondary prevention', *Journal of the American Geriatrics Society*, 50(8), pp. 1329–1335.
- Kopf, J. and Lischinski, D. (2011) 'Depixelizing pixel art', *ACM Transactions on Graphics*, 30(4), pp. 1–8.
- Liu, Z. et al. (2020) 'Fall Detection and Personnel Tracking System Using Infrared Array Sensors', *IEEE Sensors Journal*, 20(16), pp. 9558–9566.
- Mubashir, M., Shao, L. and Seed, L. (2013) 'A survey on fall detection: Principles and approaches', *Neurocomputing*, 100, pp. 144–152.
- Nadeski, M. (2019) 'Embedded Processing: Bringing machine learning to embedded systems', *Texas Instruments*.
- Noury, N. et al. (2007) 'Fall detection - Principles and Methods', in 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. *IEEE*, pp. 1663–1666.
- Riquelme, F. et al. (2019) 'Ehomeseniors dataset: An infrared thermal sensor dataset for automatic fall detection research', *Sensors (Switzerland)*, 19(20), pp. 1–16.
- Sixsmith, A. and Johnson, N. (2004) 'A smart sensor to detect the falls of the elderly', *IEEE Pervasive Computing*, 3(2), pp. 42–47.
- Taramasco, C. et al. (2018) 'A novel monitoring system for fall detection in older people', *IEEE Access*, 6, pp. 43563–43574.
- Tateno, S. et al. (2020) 'Privacy-preserved fall detection method with three-dimensional convolutional neural network using low-resolution infrared array sensor', *Sensors (Switzerland)*, 20(20), pp. 1–22.
- Uddin, M.Z., Khaksar, W. and Torresen, J. (2018) 'Ambient sensors for elderly care and independent living: A survey', *Sensors (Switzerland)*, 18(7), pp. 1–31.
- United Nations, Department of Economic and Social Affairs, Population Division (2015). *World Population Ageing 2015 (ST/ESA/SER.A/390)*
- United Nations, Department of Economic and Social Affairs, Population Division (2017). *World Population Ageing 2017 - Highlights (ST/ESA/SER.A/397)*.
- Wei-Han Chen and Hsi-Pin Ma (2015) 'A fall detection system based on infrared array sensors with tracking capability for the elderly at home', 2015 17th *International Conference on E-Health Networking, Application and Services, HealthCom 2015*, pp. 428–434.
- World Health Organization. (2008). *World health statistics 2008*. World Health Organization.