

Association for Information Systems

AIS Electronic Library (AISeL)

ISLA 2023 Proceedings

Latin America (ISLA)

Fall 8-7-2023

Aplicação de técnicas de Aprendizado de Máquina para o Reconhecimento de Atividades Humanas com dados de pontos de referência do corpo humano

Rubens de Oliveira Santos

Eric Ferreira de Paula

Fernando Silva Parreiras

Fábio Corrêa

Vinícius Figueiredo de Faria

See next page for additional authors

Follow this and additional works at: <https://aisel.aisnet.org/isla2023>

This material is brought to you by the Latin America (ISLA) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ISLA 2023 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Authors

Rubens de Oliveira Santos, Eric Ferreira de Paula, Fernando Silva Parreiras, Fábio Corrêa, Vinícius Figueiredo de Faria, and Renata de Souza França



Aplicação de técnicas de Aprendizado de Máquina para o Reconhecimento de Atividades Humanas com dados de pontos de referência do corpo humano

Artigo Completo

Rubens de Oliveira Santos
Universidade FUMEC
ruivo68@yahoo.com.br

Eric Ferreira de Paula
Universidade do Estado de
Minas Gerais (UEMG)
eric.p.f@gmail.com

**Fernando da Silva
Parreiras**
Universidade FUMEC
fernando.parreiras@fumec.br

Fábio Corrêa
Universidade FUMEC
fabiocontact@gmail.com

**Vinícius Figueiredo de
Faria**
Universidade FUMEC
profrenatafranca@gmail.com

Renata de Souza França
Universidade do Estado de Minas
Gerais (UEMG)
profrenatafranca@gmail.com

Abstract

The adoption of machine learning methods for recognizing human activities has shown remarkable results in the extraction of discriminative characteristics from sequences acquired through images. Thus, this research aims to carry out the recognition of activities carried out by students undergoing training at UniverCemig in the process of certifying their learning. This is a quantitative-exploratory research, which applied the Support Vector Machine, K-Nearest Neighbors, Gated Recurrent Unit and Long Short Term Memory techniques. As a result, K-Nearest Neighbors and Support Vector Machine achieved 91% and 96% performance, respectively, while Gated Recurrent Unit and Long Short Term Memory achieved 96% performance for both. It is believed that this research can bring an opportunity to implement the recognition of human activities in the training of students at UniverCemig, helping instructors during training and, consequently, bringing agility to the student's learning certification process.

Keywords

Human Activity Recognition, Machine Learning, Deep Machine Learning

Resumo

A adoção de métodos de aprendizado de máquina para reconhecimento de atividades humanas tem apresentado notáveis resultados na extração de características discriminativas de sequências adquiridas por intermédio de imagens. Assim, esta pesquisa visa realizar o reconhecimento de atividades executadas por alunos em treinamento na UniverCemig no processo de certificação de aprendizado desses. Esta é uma pesquisa quantitativa-exploratória, que fez uso da aplicação das técnicas *Support Vector Machine*, *K-Nearest Neighbors*, *Gated Recurrent Unit* e *Long Short Term Memory*. Por resultado, o *K-Nearest Neighbors* e o *Support Vector Machine* atingiram 91% e 96% de performance, respectivamente, enquanto

o *Gated Recurrent Unit* e *Long Short Term Memory* tiveram performance de 96% para ambos. Acredita-se que esta pesquisa pode trazer oportunidade de implantação no reconhecimento de atividades humanas no treinamento de alunos na UniverCemig, auxiliando os instrutores durante o treinamento e, conseqüentemente, levando agilidade ao processo de certificação de aprendizado do aluno.

Palavras-chave

Reconhecimento de Atividade Humana, Aprendizado de Máquina, Aprendizagem de Máquina Profundo.

Introdução

O Reconhecimento da atividade humana (RAH) é um campo de pesquisa ativo em visão computacional. Seu foco principal é detectar, automaticamente, informações sobre atividades executadas por pessoas através de dispositivos digitais (POPPE, 2010). Com o avanço da tecnologia, diversos dispositivos como: câmera RGB, câmera de profundidade, câmera RGB-D e Microsoft Kinect foram criados, possibilitando a extração de diferentes informações, dentre essas destaca-se os pontos de referências do corpo humano (TASNIM *et al.*, 2020).

A extração de pontos de referências do corpo humano é uma técnica que tem sido aplicada dentro da visão computacional para realizar o RAH, a exemplo de estimar poses de um ser humano (STRAKA *et al.*, 2011), identificar o comportamento humano (SAPINSKI *et al.*, 2019) e detecção de expressão facial (FILNTISIS *et al.*, 2019).

Técnicas de Aprendizagem de Máquina (AM) e Aprendizagem de Máquina Profunda (AMP) podem ser aplicadas ao processo de treinamento para o reconhecimento das atividades executadas por seres humanos. Com o uso de AM e AMP, atividades executadas podem ser automaticamente reconhecidas, auxiliando no processo de treinamento e certificação do alunos, onde os dados das atividades executadas são extraídos após o treinamento. A adoção de instrumentos automatizados que executem tarefas como RAH contribui para um processo de certificação célere.

Os dados utilizados para treinamento de modelo de AM e AMP podem extrair pontos de referência do corpo humano, utilizando processamento de imagem a partir de atividades executadas por alunos em treinamento. A partir dessas reflexões, este estudo tem como objetivo implementar técnicas de aprendizado de máquina para realizar o reconhecimento de atividades humanas com dados obtidos de imagens a partir de atividades executadas por um ser humano. Para atingir tal objetivo, esta pesquisa será dividida em três passos: a) criar um *dataset* (conjunto de dados); b) treinar os modelos de máquinas de aprendizado para classificação de RAH; e c) comparar os resultados dos algoritmos citados anteriormente.

A coleta de dados desta investigação foi realizada a partir de uma filmagem no campo de treinamento de eletricitistas oferecido pela UniverCemig. Assim, após esta introdução, segue a estrutura do texto: (seção 2) revisão teórica sobre as temáticas que suportam o estudo; (seção 3) procedimentos metodológicos; (seção 4) análise dos resultados; e (seção 5) considerações finais.

Referencial teórico

Reconhecimento de Atividade Humana

Visão Computacional é a área da ciência que estuda e procura desenvolver teorias e métodos voltados à extração automática de informações úteis contidas em imagens que são capturadas por dispositivos como câmera de vídeo e *scanner* (CROWLEY; CHRISTENSEN, 1995). A visão computacional é uma sub área da inteligência artificial, com o objetivo de buscar um corpo de conhecimentos que modele artificialmente a visão humana, realizando ações por meio de suas suas funções através de desenvolvimento de *softwares* e *hardwares* avançados.

O Reconhecimento de Atividades Humanas (RAH) é a tarefa de identificar ou prever atividades humanas, baseando-se em dados de sensores (DEHGHANI *et al.*, 2019) ou captura de imagens (VAHORA; CHAUHAN, 2019), podendo utilizar técnicas de aprendizado de máquina para realizar a inferência. Os dados adquiridos a partir de movimentos executados por um usuário são capturados e convertidos em padrões característicos por meio de uma combinação de componentes de *hardware* e *software*. Após a captura dos dados de observações, é feita a inferência da atividade executada.

Conforme Lara e Labrador (2013), o RAH é um problema de séries temporais. Os dados utilizados referem-se ao movimento executado pelo usuário em um certo momento. Sendo assim, o RAH pode ser abordado das seguintes maneiras: 1) abordagem de séries temporais sequenciais; e 2) abordagem de séries temporais com janelas deslizantes.

A abordagem de séries temporais sequenciais pode ser dada a partir de um conjunto de dados $S = S_0 \dots S_k - 1$, onde k é definido por uma série temporal e cada característica está contida dentro de um intervalo de tempo $l = [t\alpha, t\omega]$. O objetivo é olhar a partição do tempo $\langle t_0 \dots t_r - 1 \rangle$ de l , de acordo com os dados no conjunto S e encontrar as classes que representam a atividade executada. Os dados são consecutivos, não podem ser vazios e não se sobrepõem. Essa definição implica que uma pessoa não executa duas atividades ao mesmo tempo (LARA; LABRADOR, 2013).

A abordagem de séries temporais com janelas deslizantes pode ser dada pelo conjunto de dados $W = W_0 \dots W_m - 1$ para m sendo o tamanho da janela deslizante, total ou parcialmente rotulada, para cada W_i contendo um conjunto de séries temporais, dado por $S = S_0 \dots S_k - 1$ e um conjunto $A = a_0 \dots, a_n - 1$ de atividades rotuladas, onde o objetivo é uma função de mapeamento $f: S_i \rightarrow A$ que pode ser avaliado para todos os valores possíveis de S_i , de modo que $f(S_i)$ seja o mais semelhante possível à atividade real realizada durante W_i (LARA; LABRADOR, 2013).

Um movimento humano é coordenado por diferentes partes do corpo e com a utilização de tecnologia é possível realizar a representação através de coordenadas adquiridas de cada ponto do corpo humano que, em um conjunto de dados, representa uma ação executada, conforme Figura 1. Tal representação pode ser definida como um conjunto de coordenadas que são utilizadas para descrever a pose de uma pessoa em um determinado espaço temporal (MANZI *et al.*, 2017).

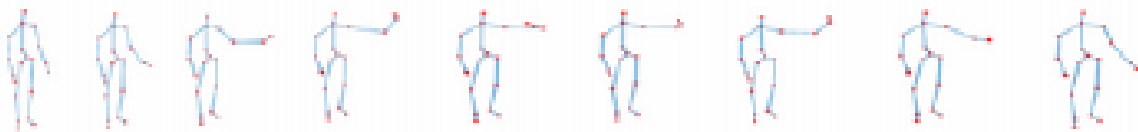


Figura 1. Pontos de Referência do Corpo Humano

Pontos chaves que compõem um corpo humano podem ser extraídos por câmera de profundidade com rastreador específico ou utilizando outro modelo que faz o mapeamento da imagem. O modelo do esqueleto humano possui algumas variedades de acordo com o *software* e o dispositivo adotado.

Um modelo esquelético pode ser modelado com vários pontos de junção (SHOTTON *et al.*, 2011), onde cada ponto descreve coordenadas em relação a posição de cada membro (MANZI *et al.*, 2017). Para Tasnim *et al.* (2020), modelos de dados podem ser criados a partir dos pontos de junção extraídos do modelo

esquelético para formar um *dataset* que possa ser utilizado por técnicas de aprendizado de máquina para realizar o RAH.

Aprendizado de Máquina

Aprendizado de Máquina (AM) é um sub campo da Inteligência Artificial que visa explorar modelos de dados e construir algoritmos que possibilitem compreender os dados de maneira autônoma. Utilizando algoritmos de AM, através de um conjunto de dados de entrada e tendo como base determinados padrões, é possível inferir as saídas. Para Sugiyama e Kawanabe (2012), o principal propósito da AM é a construção de sistemas que podem se adaptar em ambientes com constantes mudanças e aprender com sua experiência. Outra definição trazida por Mitchell *et al.* (2013) reforça o que foi dito anteriormente, caracterizando AM como um sistema de computador criado para otimizar a performance utilizando dados e experiência passada. A aprendizagem de máquina pode ser dividida em várias categorias, tais como:

- **Aprendizagem Supervisionada:** nesta categoria, AM recebe todos os dados de entrada rotulados como um conjunto de treinamento e faz previsões sobre dados não vistos anteriormente. Uma desvantagem dessa categoria é que na maioria das vezes os dados são rotulados manualmente para orientar o processo de aprendizado (DAVILA *et al.*, 2017).
- **Aprendizagem Não Supervisionada:** ao contrário do método supervisionado, esta categoria não conhece os dados de saída (MITCHELL *et al.*, 2013).
- **Aprendizado semi-supervisionado:** esta categoria usa tanto dados rotulados quanto não rotulados. A vantagem desta categoria é o fato de que apenas um número limitado de amostras rotuladas é necessário (DAVILA *et al.*, 2017).

Alguns métodos de aprendizagem de máquina são *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN) e *Redes Neurais Artificiais* (RNA). Conforme Rosati *et al.* (2018), o SVM é um classificador binário, capaz de diferenciar dados entre duas classes. Sua estrutura é fundamentada na teoria estatística, tendo como objetivo propor técnicas de aprendizagem de máquina para maximizar a capacidade de generalização do modelo e minimizar o risco estrutural. Para realizar a classificação, os dados são projetados em um espaço multidimensional, possibilitando realizar a classificação entre duas classes.

No espaço multidimensional, a separação entre as classes é obtida com um hiperplano que maximiza sua distância dos chamados vetores de suporte. De acordo com Janidarmian *et al.* (1995), SVM consiste em conjuntos de modelos supervisionados de aprendizagem e algoritmos associados que analisam dados usados para problemas de classificação e regressão.

Segundo Janidarmian *et al.* (1995), é possível construir, a partir de um conjunto de dados, um modelo capaz de aprender com novos exemplos, atribuindo novas classes aos novos exemplos, com o objetivo de separar categorias distintas por uma margem, sendo esta mais larga possível, no intuito de minimizar o erro de generalização do classificador. Este modelo trabalha com duas perspectivas, a classificação de classes linearmente separáveis e não separáveis.

Um conjunto de dados pode ser considerado linearmente separável quando seus dados podem ser separados em um plano multidimensional. Sendo assim, quando o modelo de Janidarmian *et al.* (1995) encontrar um hiperplano que separe perfeitamente os dados de cada classe e cuja margem de separação seja a maior possível, este hiperplano é denominado de hiperplano ótimo. Margem é a região delimitada por um par de hiperplanos paralelos que separam as duas classes de dados com a maior distância possível.

Quando os dados de treinamento são linearmente separáveis, é possível encontrar um hiperplano de margem rígida que atenda ao requisito de separabilidade máxima desde que não haja erros e/ou outliers nos dados. Entretanto, caso os dados não sejam linearmente separáveis e, ou, se deseja permitir certa tolerância de erros de classificação (visando evitar *overfitting*), deve-se, então, encontrar um hiperplano de margem suave, que não se ajustará totalmente e apresentará capacidade de generalização potencialmente superior. Neste caso, é necessário encontrar um hiperplano que minimize a probabilidade de erro de classificação.

De acordo com Sáez *et al.* (2016), o KNN é um algoritmo de aprendizagem de máquina para classificação e regressão. O método baseia-se na premissa de que exemplos com características similares devem mantêm-

se próximos. Quando um novo exemplo é adicionado, KNN realiza uma análise de vizinhos mais próximos, através do cálculo de distância para classificar uma instância desconhecida (AGUILETA *et al.*, 2019). Durante a classificação, KNN analisa a classe de menor custo. Caso tenha classe com o mesmo menor custo, então é realizado um desempate aleatório entre os grupos de menor custo (JANIDARMIAN *et al.*, 2017). A fórmula de início de KNN é: $eftKin = \sqrt{n}$, onde n é o número de amostras dos dados de treino.

O trabalho publicado por McCulloch e Pitts (1943) deu início a diversos estudos sobre as RNAs, também conhecido como Aprendizagem de Máquina Profundo (AMP), uma vez que se buscou entender qual era o funcionamento dos neurônios. Após este entendimento, foi criado um modelo computacional para redes neurais baseadas em matemática e algoritmos, denominado lógica limiar simulando o funcionamento dos neurônios. Haykin (2001) classifica RNA como uma máquina poderosa e complexa que é capaz de processar uma grande quantidade de informações em tempo mínimo e tem como característica possuir a capacidade de simular o comportamento do cérebro ao realizar uma tarefa. Para Janidarmian *et al.* (2017), a RNA é um sistema de neurônios interconectados capazes de criar modelos matemáticos para realizar o aprendizado de máquina.

Para Haykin (2001), RNA é formada por unidades básicas de processamento, chamadas de perceptrons, sendo compreendida como uma unidade de tomada de decisão com várias entradas e uma única saída. McCulloch e Pitts (1943) classificam um neurônio como uma unidade de processamento da informação que é fundamental para a operação do cérebro. Com os avanços da tecnologia e a evolução sobre os conceitos de RNA, diversas arquiteturas para aprendizagem de máquina foram sendo criadas, tais como *Long Short Term Memory (LSTM)* e *Gated Recurrent Unit (GRU)*.

Com o objetivo de resolver o problema de explosão do gradiente ou *vanish gradient*, Hochreiter e Schmidhuber (1997) propuseram a arquitetura LSTM. Este modelo é capaz de aprender e armazenar dependências de longo prazo por armazenar as informações em uma memória interna (WANG *et al.*, 2017). A LSTM utiliza-se do conceito de portões: um mecanismo baseado na multiplicação da entrada, o qual define o comportamento de cada célula de memória individualmente (ORDONEZ; ROGGEN, 2016).

Proposta por Chung *et al.* (2014), o GRU é capaz de armazenar os dados e os pesos em memória, tendo sua arquitetura similar ao LSTM; porém com uma implementação e poder de computação mais simples. Sua arquitetura é formada por dois portões, um de atualização (z) e um de esquecimento (r), ambos repassam informações por meio de muitas janelas de dados, buscando atingir uma melhor previsão ou classificação (LEI *et al.*, 2018). O GRU também foi criado com o objetivo de resolver o problema de explosão do gradiente ou *vanish gradient*.

Procedimentos metodológicos

A pesquisa caracteriza-se como aplicada. De acordo com Gil (2008) a pesquisa aplicada objetiva investigar, gerar conhecimentos para aplicação prática, comprovar ou rejeitar hipóteses. A natureza desta pesquisa é quantitativa. A abordagem quantitativa tem por objetivo garantir uma precisão sobre os dados analisados, assegurando um aumento da margem de confiança sobre esses para uma tomada de decisão com a utilização de técnicas estatísticas (BAPTISTA; CUNHA, 2007). Nessa pesquisa, a estatística será utilizada durante a etapa de análise e resultados.

A pesquisa é de caráter exploratório. Conforme Gil (2008), a pesquisa exploratória caracteriza-se por desenvolver, esclarecer e modificar conceitos e idéias, tendo em vista a formulação de problemas ou hipóteses pesquisáveis para estudos posteriores. O caráter de pesquisa exploratório se justifica devido à natureza desta pesquisa, que tem como objetivo a implementação de técnicas de aprendizado de máquina capazes de reconhecer as atividades executadas por um ser humano.

O *locus* desta pesquisa é a UniverCemig. A UniverCemig define um conjunto de atividades essenciais à formação do electricista. No entanto, para esta pesquisa foram coletados dados da atividade de montagem de estrutura denominada como B1. Nessa atividade, o aluno deve seguir todos os procedimentos e normas que foram desenvolvidos na parte teórica do curso, sendo estes dados coletados para aplicação das técnicas de aprendizado de máquina SVM e KNN, GRU e LSTM.

Assim, busca-se verificar, através de métricas os resultados das técnicas de aprendizagem de máquina implementadas, os quais trabalhos que envolvem comparações de qualquer natureza requerem a definição sobre qual ótica os dados devem ser analisados (CORDOVA; SILVEIRA, 2008). É imprescindível que o reconhecimento de atividades humanas no contexto do treinamento do electricista não rotule atividades erroneamente. Isso pode trazer prejuízos no processo de certificação do aprendiz do aluno. Desta maneira, é factível utilizar métricas de acurácia, precisão e revocação para avaliação dos modelos de aprendizagem de máquina que correspondem às ausências de resultados falso-positivos e falso-negativos, respectivamente.

Resultados

Nas seções abaixo serão detalhadas as configurações das técnicas utilizadas e os resultados que foram divididos em duas seções: Experimentos com técnicas de AM e Experimentos com técnicas de RNA, também denominada AMP.

Experimentos com técnicas de Aprendizado de Máquina

Esta seção apresenta os resultados e configurações provenientes das implementações de duas técnicas de AM, o KNN e SVM. As técnicas foram treinadas com dados coletados sendo separados em teste e treino, 25% e 75% respectivamente. Os melhores hiper parâmetros encontrados foram alcançados através de testes manuais. Foi escolhida uma janela de 44 amostras por segundo, com o tempo variando entre 3 e 60 segundos de dados para extração de características que podem capturar ciclos em atividades. Para reduzir a perda de informações nas bordas das janelas, utilizou-se uma sobreposição de 50%. Conforme Su *et al.* (2014), é possível encontrar dados relevantes através de cálculos que agrupe-os retornando informações que possam ser utilizadas pelas técnicas de AM. Assim, para cada eixo foi calculada a raiz quadrada, o desvio padrão, máximo e o mínimo (SU *et al.*, 2014; NGUYEN *et al.*, 2015).

Para que se chegasse ao melhor valor dos parâmetros, diversos testes manuais foram feitos, onde o valor de k variou entre 3 e 20. Os parâmetros da KNN para a execução das técnicas KNN são apresentados na tabela 1.

Parâmetro	Valor
Nneighbors	8
Weights	uniform
Metric	minkowski
P	1

Tabela 1. Hiper parâmetros KNN

Os testes para chegar ao melhor resultado com SVM foram realizados de forma manual utilizando 4 funções para testar o kernel, sendo estas: radial, linear, polinomial e sigmoide. Para a função C os testes foram variados entre 1 e 10. Os parâmetros da SVM que apresentaram melhor resultados são apresentados na tabela 2.

Parâmetro	Valor
Kernel	polinomial
C	5
Gamma	1

Tabela 2. Hiper parâmetros de Configuração da Técnicas SVM

Na Tabela 3, são exibidos os resultados do experimento para nosso modelo, projetado em termos de precisão, revocação e acurácia para o conjunto de dados UniverCemig. Foram alcançadas precisões altas de

aproximadamente 97% com o SVM e 95% com KNN. Os testes foram executados com janelas de tempo entre 3 e 160 segundos.

Janela em Segundos	Algoritmo	Acurácia	Revogação	Precisão
3(S)	KNN	0.9175	0.9176	0.9178
	SVM	0.9657	0.9657	0.9658
5(S)	KNN	0.9020	0.9026	0.9015
	SVM	0.9328	0.9328	0.9329
10(S)	KNN	0.8750	0.8750	0.8895
	SVM	0.9459	0.9454	0.9472
20(S)	KNN	0.8571	0.8671	0.8675
	SVM	0.8918	0.8917	0.8975
30(S)	KNN	0.8378	0.8378	0.8412
	SVM	0.9183	0.9183	0.9339
60(S)	KNN	0.8887	0.8888	0.9206
	SVM	0.9583	0.9583	0.9642

Tabela 3. Resultados das Técnicas de SVM e KNN

A técnica de SVM apresentou melhores resultados quando comparada com a técnica de KNN, mesmo quando a janela de tempo aumenta. Conforme Tasnim *et al.* (2020), a perda de desempenho ocorre porque uma janela de tempo maior pode capturar mais informações, o que pode afetar significativamente no desempenho do classificador. A técnica SVM não foi afetada pela janela de tempo, porém, KNN apresentou uma leve perda de desempenho quando a janela fica entre 10(S) e 60(s).

Experimentos com técnicas de Aprendizado de Máquina Profundo

Nesta seção, serão detalhadas as configurações e resultados de duas técnicas de aprendizagem de máquina profunda para realizar o RAH. Os arquivos gerados pelo conjunto de dados foram separados em treino e teste, sendo 3 arquivos para treino e 1 para teste, com uma margem de 31% e 69% para teste e treino. Para iniciar o experimento das técnicas propostas, primeiro foi definido o estilo arquitetural de cada uma, após a definição se iniciou a busca pelo melhor tamanho do *batch* e números de épocas que iriam apresentar a melhor acurácia. Foi iniciado com o número de época 1 e foi aumentando gradualmente até chegar em um melhor resultado. Nas seções abaixo será detalhado o estilo arquitetural proposto por cada modelo, seus resultados, como foram realizados o treinamento e as configurações.

Após a definição do estilo arquitetural, conforme apresentado na Tabela 4, será iniciada a execução do experimento. Durante os treinos, foram alternados os valores janela de tempo entre 1 e 6 segundos e o número de épocas foi definido em 500 com tamanho de *batch* em 300. Foram definidos também o parâmetro *paciente* com o valor 30, isso significa que caso o modelo pare de aprender, a execução do treino irá parar imediatamente após 30 tentativas.

Camada	Função de Ativação	Número de Neurônios	Percentual de Dropout
GRU	Relu	32	
GRU	Relu	32	0.5
GRU	Relu	16	0.5
Dense	Sigmoid	1	

Tabela 4. Configuração Hiper parâmetros do modelo GRU

Assim como em GRU, o primeiro passo foi definir o estilo arquitetural de LSTM, que pode ser visto na Tabela 5. Feito isto, foram iniciados os treinos variando a janela de tempo, entre 1 e 6 segundos, o número de neurônios e camadas com o objetivo de encontrar um melhor tempo de treino e aprendizagem do modelo. Os treinos foram executados com 500 épocas e o tamanho do *batch* em 300. Durante os treinos, percebe-se que quando é adicionado um maior número de camadas e neurônios o modelo aprende mais rápido; no entanto, não se percebe uma melhora nos resultados. Isso pode ser explicado pela quantidade de dados do conjunto de dados Univercemig. Quando comparado com o resultado do treino de GRU, percebe-se que este modelo aprendeu mais rápido, no entanto utilizou mais poder computacional, aumentou o tempo de treino e não apresentou diferença nos resultados para o conjunto de dados.

Camada	Função de Ativação	Número de Neurônios	Percentual de Dropout
LSTM	Relu	16	
LSTM	Relu	16	0.2
LSTM	Relu	8	0.5
Dense	Relu	100	
Dense	Sigmoid	1	

Tabela 5. Configuração Hiper parâmetros do modelo LSTM

Conforme pode ser observado na Tabela 6, foram alcançadas precisões de aproximadamente 96% com LSTM e GRU, quando se chega a um valor de 6 segundos de janela de tempo. Quando se compara os resultados apresentado pelos modelos, percebe-se que não existe uma diferença expressiva. No entanto, LSTM utilizou mais recursos computacionais em termos de camada e neurônios, implicando em um aprendizado mais rápido, mas não refletido em melhores resultados.

Janela em Segundos	Algoritmo	Acurácia	Revogação	Precisão
6(S)	LSTM	0.9681	1	1
	GRU	0.9681	1	1
5(S)	LSTM	0.8051	1	1
	GRU	0.8053	1	1
4(S)	LSTM	0.6971	1	1
	GRU	0.6972	1	1
3(S)	LSTM	0.7725	1	1
	GRU	0.7725	1	1
2(S)	LSTM	0.5493	0.50	0.4140
	GRU	0.5493	0.3532	0.3532
1(S)	LSTM	0.3514	0.3334	0.1171
	GRU	0.3513	0.3332	0.1171

Tabela 6. Resultados das Técnicas de GRU e LSTM

A arquitetura de GRU utilizou menos recursos computacionais, menor número de camadas e neurônios; porém, não teve seu resultado alterado. Os dois modelos aplicados suportaram uma maior quantidade de dados, o que era esperado; no entanto, a janela de dados para se alcançasse bons resultados foi de 360 segundos. Isso também pode ser explicado devido a quantidade de dados no conjunto de dados UniverCemig. Para ambos os modelos, varia-se as funções de ativação nas camadas de entrada e saída. A função de ativação Relu apresentou uma melhor performance nas camadas de entrada e na camada de saída; a função Sigmoid também foi a que apresentou a melhor performance.

Considerações finais

O objetivo desta pesquisa foi implementar técnicas de AM e AMP para reconhecer as atividades desempenhadas por alunos da UniverCemig durante o processo de capacitação. A análise dos resultados apontaram que a aplicação de técnicas de aprendizagem profundo apresentaram boa performance quando o tamanho da janela supera 6 segundos.

Constata-se que as técnicas SVM e KNN demonstraram ótimos resultados, mesmo quando a janela de tempo é maior, o que pode ser indicado para a aplicação com o conjunto de dados UniverCemig. A técnica SVM apresentou excelentes resultados em todas as variações de janelas, indicando ser uma ótima opção para o conjunto de dados.

Acredita-se que esta pesquisa pode trazer oportunidade de implantação no reconhecimento de atividades humanas no treinamento de alunos na UniverCemig, auxiliando os instrutores durante o treinamento e, consequentemente, levando agilidade ao processo de certificação de aprendizado do aluno.

Esta pesquisa se limitou a tratar o reconhecimento de atividades humanas de treinamento desempenhadas por alunos da UniverCemig, tendo como base a atividade de montagem de estrutura B1. Todavia, para a execução de trabalhos futuros, faz-se necessário coletar uma maior quantidade de dados e aplicá-los as técnicas, com o objetivo de atingir uma boa performance e diminuir a janela de tempo.

Referências

- Aguileta, A., Brena, R., Mayora, O., Molino Minero Re, E., and Trejo, L. 2019. "Multisensor fusion for activity recognition - a survey," *Sensors* (19:17), pp. 1-14.
- Baptista, S. and Cunha, M. 2007. "Estudo de usuários: visão global dos métodos de coleta de dados," *Perspectivas em Ciência da Informação* (12), pp. 168-184.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. 2014. "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proceedings of the NIPS 2014 Deep Learning and Representation Learning Workshop*, pp. 1-14.
- Gil, A. C. 2008. "Métodos e técnicas de pesquisa social," 6 ed., São Paulo: Editora Atlas.
- Crowley, J. I., Christensen, H. I. 1995. "Vision as process: basic research on computer vision systems," J. L. Crowley, H. I. Christensen (eds.); with editorial assistance by: Alain Chehikian e. a., pp. 341-371. Berlin: Springer, 1995.
- Davila, J., Cretu, A.-M., and Zaremba, M. 2017. "Wearable sensor data classification for human activity recognition based on an iterative learning framework," *Sensors* (17:6), pp. 1287.
- Dehghani, A., Sarbishei, O., Glatard, T., and Shihab, E. 2019. "A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors," *Sensors* (19:22), pp. 5026.
- Filntisis, P., Efthymiou, N., Koutras, P., Potamianos, G., and Maragos, P. 2019. "Fusing body posture with facial expressions for joint recognition of affect in child-robot interaction," *IEEE Robotics and Automation Letters* (4:4), pp. 4011-4018.
- Haykin, S. 2001. "Redes Neurais: Princípios e prática," São Paulo: Bookman Editora.
- Hochreiter, S. and Schmidhuber, J. 1997. "Long short-term memory," *Neural computation* (9), pp. 1735-80.
- Janidarmian, M., Roshan Fekr, A., Radecka, K., and Zilic, Z. 1995. "Support-vector networks," *Machine learning* (20), pp. 273-297.
- Lara, O. D. and Labrador, M. A. 2013. "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys Tutorials* (15:3), pp. 1192-1209.
- Lei, T., Zhang, Y., Wang, S., Dai, H., and Artzi, Y. 2018. "Simple recurrent units for highly parallelizable recurrence," *Preprint*, pp. 1-15.
- Manzi, A., Dario, P., and Cavallo, F. 2017. "A human activity recognition system based on dynamic clustering of skeleton data," *Sensors* (17:5), pp. 1-14.
- McCulloch, W. S. and Pitts, W. 1943. "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*. Cambridge, Massachusetts, USA, 1 edition.

- Mitchell, R., Michalski, J., and Carbonell, T. 2013. "An artificial intelligence approach," Cambridge, Massachusetts, USA, 1 Edition.
- Nguyen, P., Akiyama, T., Ohashi, H., Nakahara, G., Yamasaki, K., and Hikaru, S. 2015. "User-friendly activity recognition using svm classifier and informative features," IEEE, pp. 1-8.
- Ordóñez, F. and Roggen, D. 2016. "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors* (16:1), pp. 115.
- Poppe, R. 2010. "A survey on vision-based human action recognition," *Image and Vision Computing* 28:6), pp. 976-990.
- Rosati, S., Balestra, G., and Knaflitz, M. 2018. "Comparison of different sets of features for human activity recognition by wearable sensors," *Sensors* (18:12), pp. 4189.
- Sapiński, T., Kamińska, D., Pelikant, A., and Anbarjafari, G. 2019. "Emotion recognition from skeletal movements," *Entropy* (21:7), pp. 646.
- Sáez, Y., Gómez, A. B., and Viñuela, P. I. 2016. "A comparison study of classifier algorithms for cross-person physical activity recognition," *Sensors* (17:1), pp. 66.
- Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. 2011. "Real-time human pose recognition in parts from single depth images", In *Proceedings of the CVPR 2011*, pp. 1297–1304.
- Straka, M., Hauswiesner, S., Rüdter, M., and Bischof, H. 2011. "Skeletal graph based human pose estimation in real-time," In *Proceedings of the BMVC 2011*, p. 1-12.
- Su, X., Tong, H., and Ji, P. (2014). "Activity recognition with smartphone sensors," *Tsinghua Science and Technology* (19:2), pp. 35-249.
- Sugiyama, M. and Kawanabe, M. 2012. "Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation," volume 4 of 10. The name of the publisher, Cambridge, MA, USA, 1 edition.
- Tasnim, N., Islam, M. M., and Baek, J.-H. 2020. "Deep learning-based action recognition using 3d skeleton joints information," *Inventions*, (5:3), pp. 49.
- Vahora, S. and Chauhan, N. 2019. "Deep neural network model for group activity recognition using contextual relationship," *Engineering Science and Technology, an International Journal* (22:1), pp. 47-54.
- Wang, Y., Velswamy, K., and Huang, B. (2017). "A long-short term memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems," *Processes* (5:3), pp. 1-15.