



Reconhecimento de Gestos de Mão em Sequência a partir de Sensores Inerciais

Sequence Hand Gesture Recognition from Inertial Sensors

Reconocimiento de Secuencia de Gestos con las Manos a partir de Sensores Inerciales

Carlos Henrique Gomes Correia¹, Karin Satie Komati², Francisco de Assis Boldt²

RESUMO

Descritores: Auxiliares de Comunicação para Pessoas com Deficiência; Processamento de Sinais Assistido por Computador; Aprendizado de Máquina

Objetivo: O propósito desse trabalho é o de reconhecer os gestos da mão de uma base de dados de sensores inerciais. **Métodos:** A proposta é dividida em duas partes, na primeira, o processo de segmentação utiliza o algoritmo não supervisionado KMeans para identificar os segmentos correspondentes aos gestos. A segunda parte usa a técnica Floresta Aleatória (FA), com e sem extração de características, para classificação dos gestos dos segmentos encontrados na fase anterior. **Resultados:** O melhor resultado foi obtido com extração de características, apresentando 83% de acurácia. **Conclusão:** Os resultados mostram que é possível utilizar técnicas de aprendizado de máquina em conjunto para identificar, segmentar e classificar gestos de mãos em uma base de dados adquirida por sensores inerciais.

ABSTRACT

Keywords: Communication Aids for Disabled; Signal Processing, Computer-Assisted; Machine Learning

Objective: The purpose of this work is to recognize hand gestures from a database of inertial sensors. **Methods:** The proposal has two parts, where the first is the segmentation process that uses the unsupervised algorithm KMeans to identify the segments corresponding to the gestures. The second part uses the Random Forest algorithm, with and without feature extraction, to classifying the gestures segmented in the previous phase. **Results:** The best result was obtained with feature extraction, achieving 83% of accuracy. **Conclusion:** The results show that it is possible to combine machine learning techniques to identify, segment and classify hand gestures in a database acquired by inertial sensors.

RESUMEN

Descriptores: Equipos de Comunicación para Personas con Discapacidad; Procesamiento de Señales Asistido por Computador; Aprendizaje Automático

Objetivo: El propósito de este trabajo es reconocer los gestos de las manos de una base de datos de sensores inerciales. **Métodos:** La propuesta consta de dos partes, donde la primera es el proceso de segmentación que utiliza el algoritmo no supervisado KMeans para identificar los segmentos correspondientes a los gestos. La segunda parte utiliza el algoritmo Random Forest con y sin extracción de características para clasificar los gestos segmentados en la fase anterior. **Resultados:** El mejor resultado se obtuvo con la extracción de características, logrando 83% del exactitud. **Conclusión:** Los resultados muestran que es posible combinar técnicas de aprendizaje automático para identificar, segmentar y clasificar los gestos de las manos en una base de datos adquirida por sensores inerciales.

¹ Aluno do Programa de Pós-Graduação em Computação Aplicada, Instituto Federal do Espírito Santo – IFES, Serra (ES), Brasil.

² Professor associado ao Programa de Pós-Graduação em Computação Aplicada, Instituto Federal do Espírito Santo – IFES, Serra (ES), Brasil.

INTRODUÇÃO

De acordo com dados do último censo realizado em 2010⁽¹⁾, considerando a população residente no país, 23,9% (45.606.048 de brasileiros) possuíam pelo menos uma das deficiências investigadas: visual, auditiva, motora e mental ou intelectual. Sendo a deficiência motora, ocorrendo em 7% da população, ou seja, mais de 13 milhões de pessoas.

Para melhorar a qualidade de vidas das pessoas com deficiência motora, uma solução é o uso de Tecnologia Assistiva (TA). A expressão Tecnologia Assistiva⁽²⁾ é definida como “qualquer item, equipamento ou sistema de produto, adquirido comercialmente, modificado ou personalizado, usado para aumentar, manter ou melhorar as capacidades funcionais de pessoas com deficiência”. Uma das formas de TA é o reconhecimento de gestos humanos por um sistema computacional, que facilita a interação homem-máquina, melhorando a usabilidade de pessoas que podem ter restrições motoras ou facilitando a movimentação em locais restritos⁽³⁾.

Os pesquisadores Stancic, Music e Grujic⁽⁴⁾ desenvolveram um sistema eficaz baseado em sensor inercial, com um microprocessador e módulo sem fio para comunicação com um robô a uma distância de até 250 m. Os dados da base passam por quatro etapas: (i) extração de características, (ii) redução de características, (iii) geração de sequência de primitivas (que o trabalho denomina de “Motion Primitives”) através do KMeans⁽⁵⁾ e que são (iv) classificadas em 9 diferentes gestos. A redução de características é feita por PCA (do inglês *Principal Component Analysis*) para criar um novo conjunto com dimensionalidade reduzida, no caso, diminuiu de 12 para 5 atributos. O algoritmo KMeans tem como resultado 13 agrupamentos que são codificados em “strings”. O sistema demonstrou uma precisão de 90,05% implementado em um microcontrolador baseado em arduino para controle de uma plataforma de robô móvel rastreada, com a possibilidade de detectar gestos e classificação de classe nula (por exemplo, gestos indefinidos foram descartados da análise).

O artigo de Capucho, Boldt e Godinez Tello⁽⁶⁾, o único outro trabalho encontrado que utilizou a mesma base, fez a

classificação de 5 dos 9 gestos para o controle de uma cadeira de rodas. O trabalho teve como objetivo fazer um estudo comparativo entre extratores de características, e os resultados demonstraram medida-F1 de 94,92% para combinação dos extratores *Mean Absolute Value* (MAV) e *Waveform Length* (WL) para o classificador Análise Discriminante Linear.

Neste trabalho, a abordagem é composta por até três etapas: (i) uso do KMeans direto sobre os dados crus para segmentar o início e o final de cada gesto, (ii) etapa de “Extração de Características” que transforma o sinal em um vetor de características e (iii) a classificação em 10 gestos, os 9 gestos existentes mais o gesto em repouso entre os movimentos.

A Figura 1 ilustra um diagrama geral da proposta deste trabalho. A partir da base de dados de sensores inerciais, é feita a segmentação (ou separação) de cada gesto. A extração de características está representada em um retângulo tracejado, pois pode ou não ser realizada. Ao final, o resultado do classificador, isto é, o reconhecimento do gesto pode ser usado para controlar diversos dispositivos, como uma TV, a iluminação, um braço robótico ou mesmo uma cadeira de rodas.

Assim, a proposta desse trabalho é identificar gestos de mãos em uma base de dados gerada por sensores inerciais. A base de dados utilizada contém dados de indivíduos realizando nove diferentes gestos com a mão direita, com um sensor no pulso e outro no indicador⁽⁴⁾. Na etapa de classificação, a técnica RF⁽⁷⁾ (Floresta Aleatória, em inglês *Random Forest*) é testada de três formas: uma sem extração de características e as outras duas com diferentes técnicas de extração de características.

O artigo está dividido da seguinte forma: a próxima seção Materiais e Métodos discorre sobre a base de dados utilizada nesse trabalho, bem como as técnicas de segmentação e classificação. Na seção Experimentos, Resultados e Discussão, são detalhados os experimentos realizados e seus resultados e na seção Conclusão são relatadas as conclusões sobre esses resultados.

MATERIAIS

Inicia-se a seção com a descrição da base de dados. A

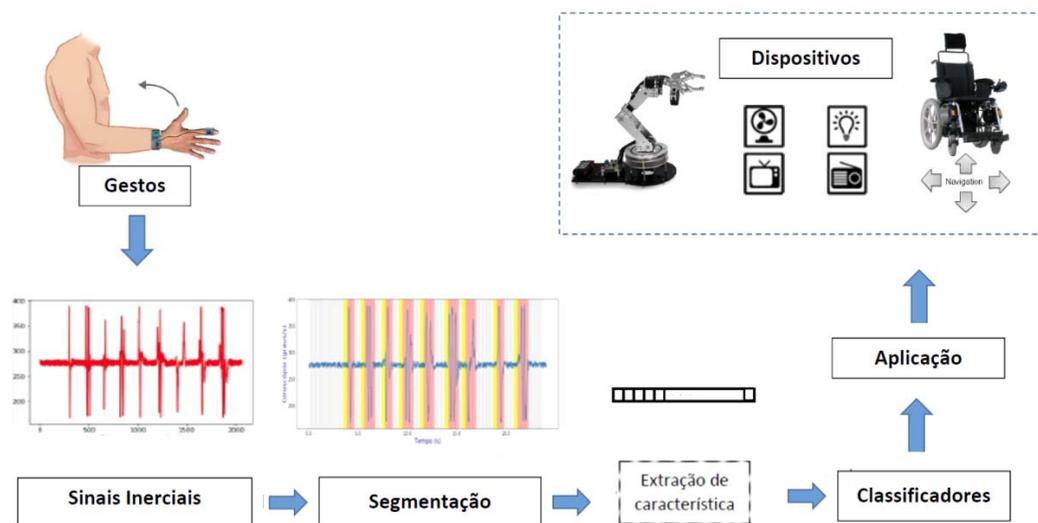


Figura 1 – Arquitetura geral proposta. Baseado no trabalho de Piol², com permissão concedida para adaptação pelos autores.

Classe	Sequências de movimentos	Movimento
1		N F N
2		N F N F N
3		N E N
4		N E N E N
5		(N+P) F (N+P)
6		(N+P) F (N+P) F (N+P)
7		(N + P) E (N + P) E (N + P)
8		(N+S) F (N+S)
9		(N+S) F (N+S) F (N+S)

Figura 2 – Sequência dos 9 (nove) gestos realizados por um indivíduo em uma sessão⁽⁸⁾

base de dados é composta por dados registrados por sensores inerciais durante sequências de gestos de mãos realizados por vários indivíduos⁽⁴⁾. Vinte indivíduos contribuíram para a aquisição dos dados, realizando os gestos com a mão direita, onde os sensores estavam fixados. Cada indivíduo realizou dez sessões de gestos simples. Em cada sessão, cada indivíduo realizou nove gestos em sequência, mantendo a mão em posição de repouso durante alguns segundos depois de cada gesto. No total, a base de dados possui 200 leituras: dez leituras realizadas por vinte indivíduos. Considerando os nove gestos de cada leitura, chega-se ao número de 1.800 gestos armazenados na base de dados.

A Figura 2 apresenta a sequência de gestos realizados pelos indivíduos. São nove gestos formados por pequenos movimentos da mão: flexão (F), extensão (E), pronação (P) e supinação (S). O gesto de repouso é identificado pela mão na posição neutra (N). O Gesto 1 (Classe 1), por exemplo, é composta pela sequência de gestos - N F N - que significa posição neutra, movimento de flexão e posição neutra. O mesmo movimento pode ser visualmente diferente, tal como se observa no movimento de flexão (F), no segundo movimento do Gesto 1, do Gesto 5 e do Gesto 8.

Durante as sessões de gestos realizadas por um indivíduo, foram utilizados dois sensores acoplados em sua mão direita: um no pulso e outro no dedo indicador. Cada um desses sensores fornece dados gerados por um acelerômetro de três eixos e um giroscópio de dois eixos. Então, cada registro de dados dos sensores produz um conjunto de dez valores, um para cada eixo de cada sensor. Ao incluir o instante de tempo em que o registro foi realizado (milissegundo passados desde o início da sessão), cada leitura é composta vários registros de onze valores: (i) eixo X do acelerômetro do pulso, (ii) eixo Y do acelerômetro do pulso, (iii) eixo Z do acelerômetro do pulso, (iv) eixo X do giroscópio do pulso, (v) eixo Z do giroscópio do pulso, (vi) eixo X do acelerômetro do dedo, (vii) eixo Y do acelerômetro do dedo, (viii) eixo Z do acelerômetro do dedo, (ix) eixo X do giroscópio do dedo, (x) eixo Z do giroscópio do dedo, (xi)

instante de tempo. Todas as leituras dos sensores foram tratadas para minimizar efeitos ruído e vibração. Nesse estudo foram utilizados os dados de leitura de todos os indivíduos, totalizando 200 leituras e 1800 gestos.

MÉTODOS

A proposta do trabalho pode ser dividida em três partes: a segmentação usando algoritmo de aprendizado de máquina não supervisionado, a extração de características e a classificação usando algoritmo de aprendizado de máquina supervisionado.

Segmentação

O objetivo da segmentação é separar as partes das leituras registradas pelos sensores inerciais que correspondem aos nove gestos realizados pelos indivíduos durante as sessões de captura.

A base utilizada não passou por nenhum tipo de pré-processamento. O algoritmo não supervisionado escolhido foi o algoritmo de clusterização KMeans, devido sua simplicidade de uso e velocidade⁽⁵⁾. Foi utilizada a implementação disponibilizada pelo módulo *Scikit-learn*⁽⁹⁾ para a linguagem de programação *Python*. O KMeans funciona agrupando uma base de dados em grupos de dados semelhantes chamados *clusters* e o único parâmetro de inicialização do KMeans definido nesse estudo foi a quantidade de *clusters*. O algoritmo *KMeans* foi usado apenas com os dados de leituras de um indivíduo, e inicializado com o número de *clusters* igual a 3 (três).

O uso de três *clusters* foi motivado pela posição dos registros das leituras em cada instante do tempo, observado nas plotagens das leituras. Observou-se que os registros ocorriam em três regiões, ao longo do período de duração da leitura, uma região central, uma região acima da região central e outra abaixo da região central. Na região central são encontrados, predominantemente, os registros dos sensores

que ocorreram enquanto o indivíduo estava com a mão em posição de repouso. Nas regiões acima e abaixo são encontrados os registros dos sensores que correspondem aos gestos realizados pelo indivíduo.

Após a execução do algoritmo KMeans e a associação dos registros com seus respectivos *clusters*, observou-se que existem, em sua maioria, blocos seguidos de registros de repouso, blocos seguidos de registros de movimento e alguns registros de movimentos intercalados com registros de repouso. Nesta última situação, decidiu-se que se a diferença de tempo entre dois registros de movimentos for menor que um determinado valor, então todos os registros de repouso entre eles serão considerados registros de movimento. A diferença de tempo entre dois registros de movimentos deve ser menor ou igual a 900 milissegundos, e esse valor de tempo foi escolhido empiricamente por ter produzido a maior quantidade de processos de segmentação bem sucedidos nos experimentos preliminares.

Em alguns casos, o processo de segmentação não é bem sucedido. O indicativo utilizado para determinar se a segmentação foi realizada com sucesso é o número de segmentos encontrados. Como se sabe que existem registros de nove gestos em cada leitura de um indivíduo, então um processo de segmentação é considerado sem falhas se ele encontra nove segmentos de movimentos. Uma falha no processo de segmentação pode ocorrer porque dois gestos distintos foram realizados com pouco tempo de repouso entre eles, fazendo com que eles sejam considerados como um gesto. Essa situação é mostrada no Gráfico 1A, onde no lugar do segundo segmento deveriam existir três segmentos.

Outro motivo é quando um gesto é composto por mais de um movimento e eles são executados com muito tempo de repouso entre eles, resultando em mais de um segmento onde deveria existir apenas um (Gráfico 1B). Essas situações poderiam ser contornadas ajustando a distância mínima aceitável entre dois registros de movimentos, mas um ajuste para diminuir as chances de acontecer a primeira situação aumentam as chances da segunda situação, e vice-versa. Ainda é possível que exista falha na segmentação quando a leitura possui muito ruído e o KMeans não consegue distinguir, a partir dos dados utilizados, os registros de repouso dos registros de movimento, gerando um número de segmentos diferente da quantidade de gestos existente.

O processo de segmentação foi desenvolvido para gerar uma base de dados segmentada, divididos entre classes correspondentes

aos gestos de 1 a 9 e o movimento de repouso, considerado como o Gesto 0. Este resultado foi utilizado para treinar algoritmos de aprendizado de máquina supervisionados e avaliar sua performance na classificação de novos dados.

Extração de Características

Só foram utilizados os dados gerados pelos processos de segmentação sem falhas. Um dos requisitos dos algoritmos de classificação é que todos os dados de treinamento tenham o mesmo tamanho. Por esse motivo, todos os dados gerados pelo processo de segmentação possuem o mesmo tamanho, definido pelo tamanho do maior movimento dos indivíduos de treinamento, mais um acréscimo de 10%. Os movimentos de repouso maiores foram divididos, produzindo vários movimentos de repouso com o tamanho requerido.

O Gráfico 1A apresenta o resultado da segmentação da Leitura 1 do Indivíduo 1. As partes em branco representam os períodos onde o indivíduo manteve a mão em repouso e as partes roxas representam os períodos onde o indivíduo realizou os gestos. O Gráfico 1B apresenta o resultado da segmentação da Leitura 1 do Indivíduo 1 com o tamanho do segmento determinado no processo, observa-se que as faixas roxas estão em tamanho fixo diferente do resultado apresentado no Gráfico 2A.

A extração de características é um método utilizado para tentar transformar os dados disponíveis mantendo suas informações mais relevantes⁽¹⁰⁾. Foram utilizadas extração de características no domínio do tempo, no domínio da frequência e no domínio de tempo-frequência. No domínio do tempo, foram calculadas as características de valor médio absoluto (MAV), raiz quadrada média (RMS), curtose (*kurtosis*), assimetria (*skewness*), valor entre picos (*peak to peak value*), fator de pico (*crest factor*), fator de impacto (*impact factor*), fator de margem (*margin factor*), fator de forma (*shape factor*) e fator de curtose (*kurtosis factor*). No domínio da frequência foram calculadas as características de média e a variância dos valores absolutos da transformada de Fourier. No domínio de tempo-frequência foi utilizado um método que permite utilizar uma função wavelet para decompor sinais por níveis, resultando em $2.n$ sinais, onde n é o número de níveis. A energia calculada nas folhas é usada como característica⁽¹¹⁾.

Classificação

A base de dados gerada pelo processo de segmentação foi utilizada para treinar os algoritmos de classificação. De

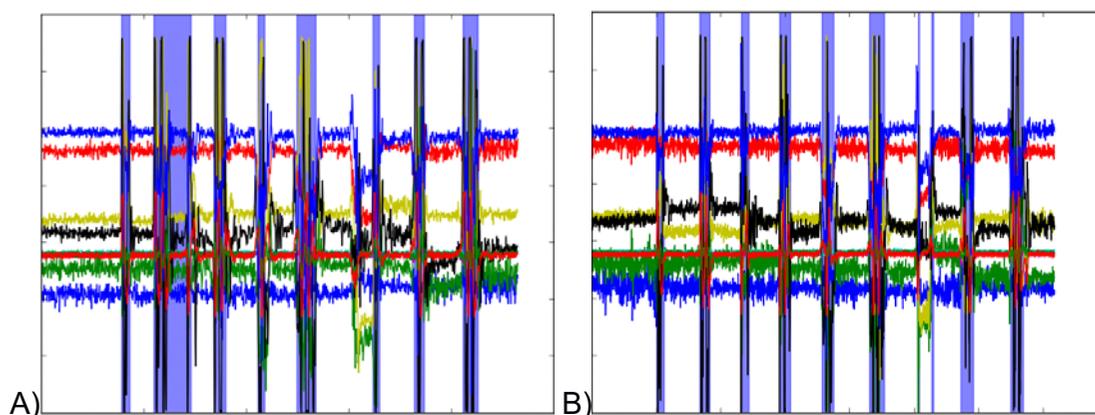


Gráfico 1 – Falhas na segmentação da Leitura 1 do Indivíduo 2 (esquerda), identificada no segundo segmento, e na segmentação da Leitura 2 do Indivíduo 3 (direita), identificada nos segmentos sete e oito.

forma arbitrária, foram selecionados os dados dos indivíduos 1 a 12 como dados de treinamento e os dados dos indivíduos 13 a 20 como dados de teste, de forma a favorecer a replicabilidade do experimento.

O algoritmo de classificação⁽⁷⁾ utilizado nesta etapa foi o RF, que consiste em uma coleção de classificadores estruturados como árvores de decisão. O resultado final é uma votação realizada entre todas as árvores de decisão. Foi utilizada a implementação do algoritmo disponibilizada pelo *Scikit-learn*. Foram treinadas três instâncias do RF: a instância RF utilizou dados sem extração de características, a versão RF_S foi treinada extraindo somente as características no domínio do tempo, e a versão RF_H foi treinada utilizando todas as formas de extração de características⁽¹¹⁾. Os parâmetros de inicialização da RF foram determinados de forma automatizada entre vários valores possíveis em um treinamento de validação cruzada com três divisões. O Quadro 1 mostra o conjunto de valores testados para cada tipo de classificador.

Com os melhores parâmetros definidos, a RF passou pelas fases de treino e teste 50 vezes, gerando uma matriz de confusão acumulada, que contém a soma dos valores das matrizes de confusão geradas a cada iteração. A partir da matriz de confusão foram geradas duas métricas de desempenho: acurácia e f1 macro⁽¹²⁾. A acurácia é a porcentagem das classificações corretas em relação às classificações realizadas e pode ser entendido com que frequência o classificador acerta uma classificação. Na matriz de confusão essa métrica é calculada pela divisão da soma dos valores da diagonal pela soma de todos os valores da matriz. F1 macro é a média aritmética de medida-F1 (ou *f1 score*) para cada classe, e a medida-F1 é a média harmônica dos valores de revocação e precisão, tendendo para o menor entre os dois valores. Precisão é a porcentagem de instâncias classificadas para uma classe que são realmente daquela classe e revocação é a porcentagem de instâncias de uma classe classificada na classe correta.

EXPERIMENTOS, RESULTADOS E DISCUSSÃO

O experimento foi realizado com 200 leituras que

possuíam registros de nove gestos, totalizando 1.800 gestos de movimento, a quantidade de gestos de repouso depende do tamanho dos segmentos determinado pelo processo de segmentação. O processo de segmentação não foi bem sucedido para 59 aquisições, ou seja, foi identificado um número de gestos de movimentos diferente de nove. O processo de segmentação foi bem sucedido para 141 leituras, identificando um total de 1.269 gestos de movimentos. Além dos gestos de movimentos, o processo de segmentação produziu 108 gestos de repouso. No total, o processo de segmentação gerou 1.377 gestos de movimento e repouso.

Os dados gerados pelo processo de segmentação foram utilizados para treinar um conjunto de algoritmos de classificação. As leituras dos indivíduos 1 a 12 resultaram em 954 gestos e foram utilizadas como dados de treinamento, as leituras dos indivíduos 7 a 10 resultaram em 423 gestos e foram utilizadas como dados de testes. Cada instância do RF foi avaliada 50 vezes.

Para avaliar o desempenho do algoritmo são apresentadas as matrizes de confusão de cada versão. Os valores da matriz de confusão apresentada em cada tabela são a soma dos valores da matriz de confusão calculada em cada uma das 50 iterações do experimento. Nas Tabelas 1, 2 e 3 são apresentados os resultados das versões RF, RF_H e RF_S, respectivamente e a Tabela 4 apresenta os resultados de acurácia obtidos a partir das matrizes de confusão.

Pelos valores das Tabelas 2 e 3 é possível avaliar que há muita troca entre o Gestor 1 (N F N) e o Gestor 3 (N E N), que possuem diferença no segundo movimento da mão, assim como o par Gestor 2 (N F N F N) e Gestor 4 (N E N E N). Outros movimentos em que houve erros, apresentam o mesmo início, tais como os pares: Gestor 1 (N F N) / Gestor 2 (N F N F N), Gestor 3 (N E N) / Gestor 4 (N E N E N), Gestor 5 (N+P F N+P) / Gestor 6 (N+P F N+P F N+P) e Gestor 8 (N+S F N+S) / Gestor 9 (N+S F N+S F N+S).

Calculando a medida-F1 a partir das matrizes de confusão, observa-se que as versões do algoritmo apresentaram alta variação no desempenho por gestos, chegando a uma variação de até 24,72%, o que mostra uma dificuldade em identificar corretamente todos os gestos. A média das medidas-F1 dos gestos mostra que as versões

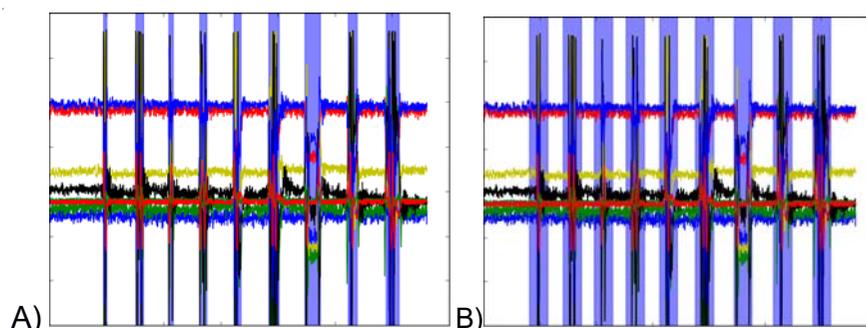


Gráfico 2 – Resultado da segmentação da Leitura 1 do Indivíduo 1, e A) com períodos sem tamanho fixo e em B) com o tamanho do segmento determinado no processo.

Quadro 1 – Conjunto de valores testados pelo processo de automatizado de determinou os parâmetros de inicialização dos classificadores.

Classificador	Parâmetros	Valores testados	Descrição
Random Forest	n_neighbors	10, 20	Número de árvores da floresta
	max_features	4, 8, todas	Número máximo de características

RF_H e RF_S tiveram desempenhos próximos, alcançando 83,92% e 83,62%, respectivamente, e acima da RF, que atingiu 79,42%.

A Tabela 4 mostra a versão RF_H com melhor desempenho, alcançando um percentual de 83,77%, com performance muito próxima, a versão RF_S obteve um percentual de 83,47%, enquanto a versão RF chegou a um percentual de 79,53%. Conclui-se que a extração de características contribui para uma melhor classificação dos segmentos, mas não houve diferença nos resultados pela forma diferenciada de extração de características.

CONCLUSÃO

O propósito desse trabalho foi de reconhecer os gestos da mão de uma base de dados de sensores inerciais. A proposta foi dividida em duas partes, o processo de segmentação que utilizou o algoritmo não supervisionado

KMeans para identificar os segmentos correspondentes aos gestos e apresentou um bom resultado quantitativo: foram identificados 1.269 dos 1.800 gestos existentes. A segunda parte foi o processo de classificação dos segmentos dos gestos encontrados. Essa análise mostrou que o *Random Forest* com extração de características obteve o melhor desempenho, com percentual de 83% de acurácia. Os resultados obtidos mostram que é possível utilizar técnicas de aprendizado de máquina em conjunto para identificar, segmentar e classificar gestos de mãos em uma base de dados gerada por sensores inerciais, e são promissores pois podem ser estendidos para a aplicação de próteses robóticas, braços robóticos em cirurgia, e reconhecimento de linguagem baseado em sinais.

Apesar dos resultados, a taxa de extração de gestos dos sensores tem uma perda de aproximadamente 30% dos movimentos. Realizar uma análise das estatísticas sobre o tempo médio de repouso entre gestos dos voluntários pode ajudar a entender se o problema é a falta de uniformidade,

Tabela 1 – Matriz de confusão da versão RF do algoritmo.

Gesto	0	1	2	3	4	5	6	7	8	9
0	1899	7	2	33	7	1	0	0	1	0
1	122	2055	12	39	56	11	0	4	1	0
2	135	108	1712	244	49	21	12	4	4	11
3	243	28	27	1938	29	15	2	16	2	0
4	86	295	72	219	1537	59	5	14	12	1
5	50	199	6	18	126	1503	383	12	3	0
6	95	5	110	9	39	363	1668	7	3	1
7	18	12	7	12	11	35	53	2107	44	1
8	51	46	12	14	35	14	5	13	1769	341
9	50	5	28	6	14	35	14	20	302	1826

Tabela 2 – Matriz de confusão da versão RF_H do algoritmo.

Gesto	0	1	2	3	4	5	6	7	8	9
0	1950	0	0	0	0	0	0	0	0	0
1	105	1767	42	344	40	0	0	0	2	0
2	51	213	2002	0	20	0	0	0	1	13
3	1	173	0	2126	0	0	0	0	0	0
4	0	53	148	282	1811	1	0	0	5	0
5	0	115	0	25	16	2024	80	0	40	0
6	0	0	82	0	265	294	1634	0	1	24
7	0	0	0	5	0	55	40	2080	120	0
8	1	161	0	13	2	0	0	51	2070	2
9	0	0	66	0	1	0	0	32	690	1511

Tabela 3 – Matriz de confusão da versão RF_S do algoritmo.

Gesto	0	1	2	3	4	5	6	7	8	9
0	1950	0	0	0	0	0	0	0	0	0
1	111	1836	2	324	8	2	0	0	17	0
2	55	254	1758	34	160	0	1	0	16	22
3	6	125	0	2135	1	1	0	27	5	0
4	2	84	62	425	1665	9	1	5	46	1
5	0	114	0	35	16	1996	74	22	43	0
6	0	0	34	3	104	313	1772	13	9	52
7	0	0	0	12	3	3	49	2184	49	0
8	5	75	17	29	49	0	2	118	1892	113
9	0	1	118	1	7	0	2	4	449	1718

Tabela 4 – Resultados de desempenho das versões do algoritmo.

	Acurácia
RF	79,53%
RF_H	83,77%
RF_S	83,47%

se os sensores não estão bem calibrados ou se os voluntários não executaram os gestos apropriadamente.

Como trabalhos futuros, é possível desenvolver pesquisas para identificar o valor do tempo que separa os movimentos, utilização de extração de características no processo de

segmentação, utilização de somente dois *clusters*, um para repouso e um para movimento, comparação com outros algoritmos não supervisionados e teste de classificação com uma LSTM (*Long-Short Term Memory*) ou RNN (*Recurrent Neural Network*), devido à natureza sequencial dos dados.

REFERÊNCIAS

1. Secretaria Nacional de Promoção dos Direitos da Pessoa com Deficiência. Cartilha do Censo 2010 - Pessoas com Deficiência [cartilha na Internet]. Portal Inclusão. 2012 [acesso em 27 de março de 2020]. Disponível em: <https://inclusao.enap.gov.br/wp-content/uploads/2018/05/cartilha-censo-2010-pessoas-com-deficiencia-reduzido-original-eleitoral.pdf>.
2. Isabelle S, Bessey SF, Dragas KL, Blease P, Shepherd JT, Lane SJ. Assistive technology for children with disabilities. *Occupational Therapy in Health Care*. 2003 Jan 1;16(4):29-51.
3. Chakraborty BK, Sarma D, Bhuyan MK, MacDorman KF. Review of constraints on vision-based gesture recognition for human-computer interaction. *IET Computer Vision*. 2017 Nov 16;12(1):3-15.
4. Stančič I, Musič J, Grujić T. Gesture recognition system for real-time mobile robot control based on inertial sensors and motion strings. *Engineering Applications of Artificial Intelligence*. 2017 Nov 1;66:33-48.
5. Xu, J, Lange, K. Power k-Means Clustering. *Proceedings of 36th International Conference on Machine Learning 2019* (pp. 6921–6931). PMLR.
6. Capucho RP, de Assis Boldt F, Tello RJ. Reconhecimento de Sequência de Movimentos de uma Mão a partir de Sensores Inerciais para o Controle de uma Cadeira de Rodas Robotizada. *Anais do II Congresso de Tecnologia da Informação do IFSUL Passo Fundo (IFSUL) – Passo Fundo, RS*. 2019:48-57.
7. Ceravolo I, Brasil A, Komati K. Classificação de dislexia a partir de movimentos oculares durante a leitura usando aprendizado de máquina e wavelets. *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional 2019* Out 15 (pp. 880-891). SBC.
8. Capucho, R. P. Um Sistema de Reconhecimento de Gestos e Sequências de Movimentos a Partir de Sensores Inerciais para o Controle de uma Cadeira de Rodas Robotizada Usando Redes Neurais Convolutivas. Serra: Instituto Federal do Espírito Santo, 2019. Tese de Mestrado
9. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*. 2011 Nov 1; 12:2825-30.
10. Choi S, Jiang Z. Comparison of envelope extraction algorithms for cardiac sound signal segmentation. *Expert Systems with Applications*. 2008 Feb 1;34(2):1056-69.
11. Boldt FD, Rauber TW, Varejao FM. Feature extraction and selection for automatic fault diagnosis of rotating machinery. *Anais do X Encontro Nacional de Inteligência Artificial e Computacional (ENIAC) - Fortaleza, Ceará*. 2013:213-20.
12. Zhang X, Graepel T, Herbrich R. Bayesian online learning for multi-label and multi-variate performance measures. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics 2010* Mar 31 (pp. 956-963).