

# Comparative Study of Methods of Feature Extraction from Keypoints for the Classification of General Movements in Newborns

André Valente de Cristo<sup>1</sup>, Eulanda Miranda dos Santos<sup>1</sup>,  
Rafael Giusti<sup>1</sup>, Ayrles Silva Gonçalves Barbosa Mendonça<sup>2</sup>

<sup>1</sup>Instituto de Computação (IComp), Universidade Federal do Amazonas (UFAM),  
Manaus, Amazonas (AM), Brasil.

<sup>2</sup>Faculdade de Educação Física e Fisioterapia (FEFF), UFAM, Manaus, AM, Brasil.

{andre.cristo, emsantos, rgiusti}@icomp.ufam.edu.br, ayrles@ufam.edu.br

**Abstract.** *General movements assessment (GMA) is a widely used method for predicting motor dysfunction, particularly cerebral palsy. In recent years, several studies have made progress in automating GMA by using manually extracted features from key points identified in videos combined with either shallow or deep machine learning models. However, there is still no consensus in the literature regarding the best feature extraction method, and there is a lack of studies comparing different methods, especially for newborns. In this context, this work aims to compare three methods of feature extraction from keypoints found in the literature for the classification of general movements in newborns.*

**Resumo.** *A avaliação de movimentos gerais (GMA) é um método muito usado para prever disfunção motora, principalmente paralisia cerebral. Nos últimos anos, vários trabalhos têm feito progresso na automatização da GMA utilizando características extraídas manualmente de pontos-chaves identificados em vídeos, combinadas com modelos de aprendizado de máquina rasos ou profundos. Entretanto, não existe um consenso na literatura quanto ao melhor método de extração de características e há uma falta de trabalhos comparando métodos diferentes, especialmente para recém-nascidos. Nesse contexto, este trabalho visa comparar três métodos encontrados na literatura de extração de características a partir de pontos-chaves para a classificação de movimentos gerais de recém-nascidos.*

## 1. Introdução

Movimentos gerais (GM—*general movements*), são movimentos espontâneos que ocorrem desde a vida fetal até o fim do primeiro meio ano de vida. Tratam-se de movimentos frequentes que duram tempo suficiente para serem observados. Eles envolvem o corpo inteiro em uma sequência de movimentos do tronco, pescoço, braços e pernas, com aparência complexa e variável decorrente de rotações ao longo do eixo dos membros e leves mudanças na direção dos movimentos [Einspieler and Prechtel 2005]. GMs podem ser caracterizados em dois tipos: do nascimento até 2 meses de idade os GMs apresentam características semelhantes a movimentos fetais e são chamados *writhing movements* (WMs); entre 3 a 5 meses de idade, os GMs consistem em movimentos contínuos, de

pequena amplitude e velocidade moderada, que ocorrem em todas as direções, e são chamados de *fidgety movements* (FMs). Entre 2 e 3 meses, lactentes apresentam tanto WMs quanto FMs, com a frequência de WMs diminuindo e a de FMs aumentando.

A avaliação de movimentos gerais (GMA) é muito usada para prever disfunção motora, pois é não invasiva, tem bom custo-benefício e é de relativamente fácil aprendizado. Entre os padrões que a GMA identifica, a presença contínua de câimbra sincronizada (CS) na fase de WMs e a ausência de FMs (FM-) são os de maior valor preditivo, indicando grande chance de disfunção motora significativa [Einspieler et al. 2016].

Para se analisar os GMs, é primeiro necessário detectar de alguma forma os movimentos dos lactentes. Isso pode ser feito com sensores conectados fisicamente aos bebês, como giroscópios e acelerômetros, ou com sensores não intrusivos, como câmeras RGB e dispositivos de profundidade. Devido à facilidade de uso, baixo custo, portabilidade e o seu uso não ser invasivo, câmeras RGB são os dispositivos mais comuns na literatura.

Uma forma de extrair informações de movimento de vídeos é são os de extração de pontos-chaves, como *OpenPose* e *MediaPipe BlazePose*. Algumas vantagens desses métodos sobre outras formas de extração de informações de movimento, tais como fluxo óptico, é que eles filtram informações desnecessárias, como mudanças de iluminação, plano de fundo e tamanho de partes do corpo, deixando apenas a postura dos lactentes, além de anonimizá-los [Leo et al. 2022, Huang et al. 2024, McCay et al. 2022].

Para fazer a avaliação com base nas informações de movimento obtidas via pontos-chaves, trabalhos recentes têm obtido resultados promissores com modelos de aprendizado de máquina (AM) profundos utilizando como entrada as posições dos pontos-chaves ou características extraídas manualmente a partir deles [McCay et al. 2020, Passmore et al. 2024]. No entanto, esse tipo de modelo exige uma grande quantidade de dados para ser treinado e, devido a questões de proteção de dados sensíveis, não há muitas bases de dados disponíveis publicamente, especialmente para WMs [Leo et al. 2022, Huang et al. 2024, McCay et al. 2022]. Devido a isso, a maior parte dos trabalhos usa bases de dados próprias, que não são publicadas e muitas vezes têm tamanho limitado, como é o caso da base usada neste artigo. Dado esse problema, vários artigos recentes utilizam métodos mais tradicionais, com modelos rasos e características extraídas manualmente dos pontos-chaves [Huang et al. 2024, Doroniewicz et al. 2020, McCay et al. 2022].

Atualmente, não há na literatura consenso quanto ao melhor tipo de característica, sendo que cada autor tipicamente propõe seu próprio conjunto [Huang et al. 2024, Doroniewicz et al. 2020, Letzkus et al. 2024, McCay et al. 2022, Ji et al. 2024]. Além disso, normalmente os trabalhos não promovem comparações com métodos de outros autores em suas próprias bases, o que, combinado com a falta de bases públicas, torna comparações entre métodos difíceis. Considerando essas lacunas, neste artigo é feita uma comparação entre os métodos de extração de características a partir de pontos-chaves propostos em três artigos recentes [Huang et al. 2024, Letzkus et al. 2024, McCay et al. 2022], dos quais o método apresentado em [Letzkus et al. 2024] não foi originalmente testado com modelos de AM e o apresentado em [McCay et al. 2022] não foi originalmente testado com WMs. Os dados representados por esses três conjuntos de características são utilizados para treinar diferentes modelos de AM clássicos, de forma a classificar WMs entre típicos e atípicos, englobando movimentos dos tipos PR e CS.

O restante deste artigo está organizado da seguinte forma. Na Seção 2 são descritos alguns trabalhos relacionados. A metodologia empregada e detalhes da base de dados usada são apresentados na Seção 3. Os resultados dos experimentos são destacados na Seção 4. Por fim, a Seção 5 discute as conclusões e trabalhos futuros.

## 2. Trabalhos Relacionados

Em [Huang et al. 2024] é descrita uma metodologia para classificação de WMs entre típico e atípico. Os autores descrevem as instalações físicas e o processo de filmagem dos vídeos dos lactentes. A base gerada é composta por 103 vídeos de 10 a 20 minutos de 96 recém-nascidos de 40 semanas de idade concepcional, dos quais foram selecionados aleatoriamente 37 vídeos de movimentos típicos e 37 de atípicos, com atípicos englobando PR e CS. Em seguida, o *MediaPipe BlazePose* foi submetido a uma fase de ajuste fino para dados de vídeos de recém-nascidos. O modelo resultante foi utilizado para extrair 23 pontos-chaves 3D. Na fase de pré-processamento, os pontos obtidos em cada quadro com taxa de confiança abaixo de 90% da média de confiança foram removidos e substituídos via interpolação linear. Em seguida, foi aplicado um filtro Gaussiano unidimensional. O próximo passo envolveu o cálculo de 12 ângulos de juntas com base nos pontos pré-processados e o ponto médio entre os quadris foi tomado como centro. Por fim, foram extraídas 27 características da série temporal de cada ponto e ângulo e esses dados foram inseridos em 8 modelos de AM rasos. O melhor resultado foi obtido com *logistic regression*, com 93,33% de acurácia.

Já em [Doroniewicz et al. 2020] o foco foi classificar WMs entre normal e PR. A base de dados usada inclui 31 vídeos de lactentes de 2 ou 3 dias de idade, dos quais 17 pertenciam à classe normal e 14 à classe PR, com os segmentos específicos dos movimentos definidos nos vídeos. O *OpenPose* foi usado para extrair 25 pontos-chaves 2D. Para pré-processar os dados, foi utilizado um filtro Savitzky–Golay, o ponto do pescoço foi tomado como centro e o esqueleto foi rotacionado para que seu eixo coincidisse com o eixo da imagem. Foram então calculadas 4 características para cada membro do corpo, as quais foram utilizadas por 3 modelos clássicos de AM. O melhor resultado foi o do *random forest*: 80,93% de acurácia.

Os autores em [Letzkus et al. 2024] avaliaram a possibilidade de criar um modelo para classificar WMs entre normais e CS em lactentes de muito baixo peso ao nascer (MBPN). A base usada é composta por 76 vídeos de lactentes MBPN (idade gestacional inferior a 32 semanas ou peso ao nascer inferior a 1500 g) de  $5370,93 \pm 1353,28$  quadros. Primeiramente foi feito o treinamento de um modelo estimador de pose baseado no *Detec-tron2*, pré-treinado na base de imagens de adultos MS COCO, e submetido a ajuste fino na base de lactentes. Posteriormente, foi calculada a autocorrelação de 8 ângulos de 5 vídeos de GMs normais e 5 CS GMs, mostrando que CS GMs têm um valor de autocorrelação maior em todos os níveis de atraso em comparação com GMs normais, especialmente nas extremidades inferiores. Os autores não utilizaram métodos de AM para realizar a classificação dos GMs. Porém, neste trabalho nós utilizamos as características propostas em [Letzkus et al. 2024] com métodos clássicos de AM para realizar essa tarefa.

Em [McCay et al. 2022] foi feita a classificação de FMs entre presentes (FM+) e ausentes (FM-) em bebês de 3 a 5 meses. Foram usadas a base disponível publicamente MINI-RGBD [Hesse et al. 2018] e uma base própria, chamada de RVI-38 e dis-

ponibilizada publicamente, composta por 38 vídeos de 40 segundos a 5 minutos, dos quais 32 apresentam FM+ e 6 FM-. O *OpenPose* foi usado para extrair 25 pontos-chaves 2D. Para o pré-processamento foi feita uma correção de dados e normalização. Na fase de correção, pontos com confiança abaixo de 95% da média de confiança do ponto no vídeo foram eliminados e interpolados com o método Akima modificado, proposto em [McCay et al. 2020]. Por fim, foi aplicado um filtro de média móvel de 5 quadros em cada ponto. Para a normalização, os pontos foram centralizados no ponto médio dos quadris e rotacionados para alinhar a coluna com o eixo y da imagem. Os autores então calcularam 7 características baseadas em histogramas, testando-as individualmente e em conjunto com 6 modelos clássicos de AM. A fusão de todas as características e um comitê de todos os modelos obteve os melhores resultados: 100% de acurácia na base MINI-RGBD e 97,37% na base RVI-38.

Por fim, em [Ji et al. 2024] foi feita a predição do desenvolvimento cognitivo de lactentes de 8 meses de idade, os dividindo em baixo risco ou alto risco, com base em seus FMs aos 3 a 4 meses. A base usada é composta por 94 vídeos de pelo menos 5 minutos, dos quais 48 apresentam lactentes posteriormente classificados como baixo risco e 46 como alto risco. Uma versão adaptada do *OpenPose* foi usada para extrair 13 pontos-chaves 2D. Na fase de pré-processamento, inicialmente, pontos não reconhecidos foram interpolados. Então, uma janela deslizante de 25 quadros e passo 1 foi usada em cada série temporal, retirando a média dos valores em cada passo para gerar novas séries com menos ruído. Foi então usado um filtro *Butterworth* passa-baixa de quarta ordem para suavizar as séries. Por fim, os pontos foram centralizados no pescoço e normalizados pela altura do lactente. Após o pré-processamento, foram calculadas as séries temporais de velocidade e aceleração dos pulsos e tornozelos, assim como velocidade e aceleração angular dos cotovelos e joelhos. Foram então calculados 3 grupos de características a partir dessas séries, totalizando 48 características básicas. A partir dessas características, eles obtiveram características que representam a simetria de movimento bilateral (BMS), divididas em 3 categorias, com 24 características de cada categoria, totalizando 72 características BMS. As características básicas e BMS foram usadas como entrada para 8 modelos clássicos de AM. O melhor resultado, área sob a curva ROC de 0,851, foi obtido utilizando apenas as características BMS e Naive Bayes.

Dentre os trabalhos mencionados nesta seção, foram escolhidos os conjuntos de características apresentados em [Huang et al. 2024], [Letzkus et al. 2024] e [McCay et al. 2022] para se fazer os testes. O conjunto de características, etapas de pré-processamento e modelos de AM utilizados em [Huang et al. 2024] serviram como base para nosso trabalho, pois o artigo foi o que apresentou os melhores resultados entre os trabalhos recentes encontrados na literatura que fazem classificação de WMs com características extraídas manualmente e modelos rasos. Já o conjunto proposto em [Letzkus et al. 2024] foi escolhido pois, apesar de os autores não terem testado a característica proposta em modelos de AM, ela é simples e eles demonstraram que havia uma diferença considerável entre os seus valores para movimentos típicos e CS. Embora em [Doroniewicz et al. 2020] os autores também tenham trabalhado com WMs, o método usado é mais complexo do que os descritos em [Huang et al. 2024] e [Letzkus et al. 2024], e os resultados não foram tão bons quanto os apresentados em [Huang et al. 2024]. Por fim, o conjunto apresentado em [McCay et al. 2022] também foi escolhido pois, embora tenha sido testado com FMs, obteve resultados promissores, e

o uso de características baseadas em histogramas tem obtido sucesso na identificação de movimentos de forma geral, como em [Xia et al. 2012], sugerindo seu uso para WMs.

### 3. Metodologia

#### 3.1. Base de dados

Neste trabalho, foram realizados experimentos utilizando uma base de dados própria, criada a partir da gravação de vídeos de 26 recém-nascidos<sup>1</sup>. Esses vídeos foram gravados com um telefone celular e rotulados por um especialista em GMA, e todos mostram lactentes realizando WMs. Os vídeos originais, com durações entre 1 minuto e meio e 3 minutos, foram divididos em segmentos de 30 segundos, resultando em 900 quadros por segmento a uma taxa de 30 quadros por segundo (fps). Entre os lactentes da classe atípico, apenas os segmentos nos quais os lactentes realizaram movimentos atípicos foram marcados como exemplos da classe, com os outros sendo descartados. No total, foram obtidas 54 instâncias, com 25 da classe típico e 29 da atípico. Assim como em [Huang et al. 2024], desses vídeos foram extraídos 23 pontos-chaves 3D utilizando o *MediaPipe BlazePose* e 12 ângulos foram calculados a partir desses pontos, como indicado na Figura 1, também retirada do artigo original.

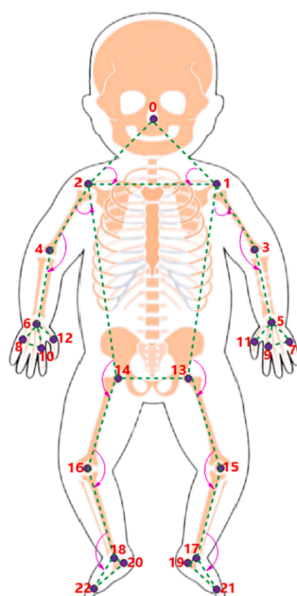


Figura 1. Pontos-chaves e ângulos utilizados [Huang et al. 2024].

#### 3.2. Pré-processamento

Embora cada um dos três métodos de extração de características investigados neste trabalho tenha sido desenvolvido utilizando uma estratégia de pré-processamento diferente, como o objetivo deste artigo é comparar o conjunto de características proposto em cada trabalho, todos os testes foram feitos utilizando a mesma estratégia. Idealmente, múltiplas estratégias deveriam ser testadas separadamente para a identificação da melhor, porém,

<sup>1</sup>Estudo aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do Amazonas, sob o CAAE nº 78265824.5.0000.5020 e parecer nº 6.765.526, em conformidade com a Resolução CNS nº 466/2012.

isto está fora do escopo deste artigo. Como os outros dois métodos não podem ser replicados diretamente, devido aos autores de [Letzkus et al. 2024] não terem feito testes com modelos de AM e os de [McCay et al. 2022] terem trabalhado com FMs, o pré-processamento apresentado em [Huang et al. 2024] foi escolhido para ser usado em nossos experimentos, permitindo que possamos testar da forma mais aproximada possível o impacto dos conjuntos de características em nossa base de dados.

Como especificado em [Huang et al. 2024], os pontos de confiança abaixo de 90% da média do ponto no vídeo foram inicialmente removidos, e os valores ausentes foram interpolados linearmente. Foi então empregado um filtro Gaussiano unidimensional para suavizar os dados. Por fim, estes foram centralizados no ponto médio entre os pontos-chaves dos quadris.

### 3.3. Extração de características

Neste trabalho são comparados três métodos diferentes de extração de características extraídos de diferentes domínios.

**Características básicas em domínio do tempo e frequência:** Este conjunto é composto por 17 características em domínio do tempo e 10 em domínio da frequência, como especificado em [Huang et al. 2024]. Como os valores de alguns parâmetros não estão claramente definidos no artigo original, como a ordem usada nos momentos e a etapa  $f(k)$  usada nas características de *waveform*, neste trabalho foi usada a ordem 5, que demonstrou melhor desempenho em comparação com outros valores testados, e  $f(k)$  foi considerada como a fase do sinal, visto que é uma grandeza principal em função da frequência.

**Autocorrelação:** A autocorrelação representa o grau de semelhança entre uma série temporal e uma versão atrasada dela mesma em intervalos de tempo sucessivos. Em [Letzkus et al. 2024], foi demonstrado que CS GMs apresentam valores maiores de autocorrelação, embora os autores não tenham chegado a testar os dados em modelos de AM. Adicionalmente, eles não demonstraram a distribuição de valores para PR GMs, que compõem a maior parte de GMs atípicos da base de dados usada neste trabalho. Como não existe na literatura um padrão para o atraso  $k$ , neste artigo os testes foram feitos com números ímpares entre 1 e 900.

**Características baseadas em histogramas:** O conjunto proposto em [McCay et al. 2022] é composto por 7 características baseadas em histogramas, das quais 4 são baseadas em pose: HOJO3D, HOAD3D, HORJO3D e FFT-JO e 3 são baseadas em velocidade: HOJD3D, HORJAD3D e FFT-JD. As características tiveram que ser modificadas devido ao nosso uso de pontos-chaves 3D, resultando em um aumento de dimensionalidade em várias delas. Em particular, as características de orientação, antes representadas por um ângulo, precisaram de dois (elevação e azimuth), com seus histogramas calculados separadamente por ponto e concatenados. Assim como no artigo original, testes foram feitos com cada característica individualmente e com a fusão das características baseadas em velocidade, pose e ambos os tipos.

### 3.4. Abordagem *Leave-One-Subject-Out* (LOSO)

Dada a limitação de dados disponíveis para a realização dos experimentos, nós optamos por utilizar a estratégia *Leave-One-Subject-Out* (LOSO), que é uma técnica de validação

cruzada, para a avaliação do desempenho dos modelos propostos. Nesse contexto, a abordagem LOSO implica que, em cada iteração, os vídeos de um lactente específico são designados como o conjunto de teste, enquanto os vídeos dos demais constituem o conjunto de treinamento. Essa estratégia permite uma melhor avaliação do desempenho dos modelos, mesmo com o baixo volume de dados. Além disso, ao utilizar LOSO, é possível obter uma estimativa mais confiável da capacidade do modelo de lidar com diferentes indivíduos, o que é essencial em cenários de reconhecimento de atividades humanas com dados limitados [Gholamiangonabadi et al. 2020].

### 3.5. Métodos de Classificação

Assim como em [Huang et al. 2024], em nossos experimentos foram utilizados os seguintes modelos clássicos de AM: *logistic regression* (LR), *naïve Bayes* (NB), *support vector machine* (SVM), *decision tree* (DT), *gradient boosting* (GB), *k-nearest neighbour* (KNN) e *random forest* (RF). Os hiperparâmetros dos métodos de AM foram definidos no contexto da validação cruzada LOSO, descrito na subseção anterior. As seguintes métricas foram calculadas após os testes com a estratégia LOSO: acurácia (AC), precisão (PR), revocação (RE), F1-score (F1) e ROC-AUC (AUC). Com exceção dos testes iniciais com autocorrelação para definir o valor de k, todos os diferentes métodos de AM foram utilizados em todos os testes.

## 4. Experimentos e Resultados

### 4.1. Características básicas em domínio do tempo e frequência

A Tabela 1 apresenta os resultados com apenas as características em domínio do tempo, apenas as características em domínio da frequência e com ambos os grupos, mostrando apenas os resultados do melhor modelo. Os melhores resultados estão destacados em negrito. Conforme pode ser observado na tabela, a maior acurácia obtida foi alcançada pelo método GB (79,63%) utilizando apenas as características no domínio da frequência.

**Tabela 1. Resultado dos testes com as características em domínio do tempo, frequência e ambos os grupos.**

Carac.	Modelo	AC	PR	RE	F1	AUC
Frequência	GB	<b>0.7963</b>	<b>0.7812</b>	0.8621	<b>0.8197</b>	<b>0.7910</b>
Tempo	NB	0.7407	0.7027	<b>0.8966</b>	0.7879	0.7283
Ambos	GB	0.7222	0.7059	0.8276	0.7619	0.7138

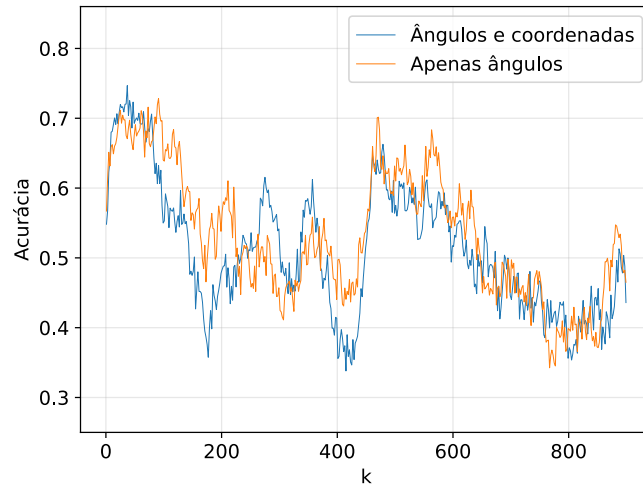
Como mencionado anteriormente, [Huang et al. 2024] é o trabalho mais relacionado ao nosso, pois foca em WMs, utiliza uma base de tamanho comparável, modelos rasos e pontos-chaves 3D. No entanto, não foi possível replicar por completo o protocolo do artigo original. As principais diferenças entre os dois trabalhos são: processo de captura dos vídeos; modelo estimador de pose; e hiperparâmetros das características de *waveform*. Diferentemente da base usada pelos autores, a nossa base é composta por vídeos gravados via celular não padronizados. A diferença poderia ser compensada através do pré-processamento, porém, uma tentativa inicial de replicar o pré-processamento descrito em [Passmore et al. 2024], que também utiliza vídeos de celular, piorou consideravelmente nossos resultados, talvez porque a perda de informação por fixar a coluna do lactente tenha um impacto maior em WMs do que com FMs. Mais testes devem ser feitos

para que conclusões possam ser tomadas quanto a esse aspecto. Em relação ao modelo estimador de pose, os autores fizeram um ajuste fino no *MediaPipe BlazePose* que melhorou consideravelmente o desempenho do modelo. Em [Jahn et al. 2024], vários modelos são avaliados, e o *MediaPipe BlazePose* sem ajuste fino que nós usamos não apresentou um bom desempenho, sugerindo que o uso de um modelo diferente poderia melhorar nossos resultados.

## 4.2. Autocorrelação

Para encontrar um valor ideal para  $k$ , foram feitos testes com todos os modelos, com exceção de GB e RF por demandarem tempos de treinamento mais longos. Dois grupos de testes foram realizados: calculando a autocorrelação apenas dos ângulos, como em [Letzkus et al. 2024] e calculando a autocorrelação dos ângulos e coordenadas juntos, como em [Huang et al. 2024]. O valor de  $k$  variou entre 1 e 900, com passo igual a 2.

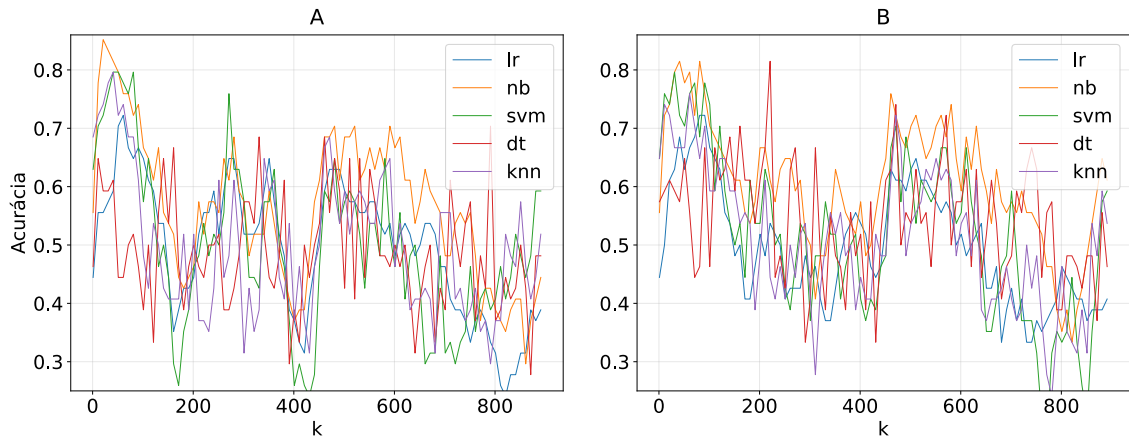
A média das acurácias dos modelos obtida ao variar o valor de  $k$  para cada grupo de teste é sumarizada na Figura 2. Já a Figura 3 mostra a variação das acurácias obtidas por cada modelo separadamente. O gráfico A mostra os valores para o grupo coordenadas e ângulos, enquanto o gráfico B mostra os resultados somente para o grupo ângulos. É importante mencionar que os gráficos mostram valores com intervalos de 10 em 10 para melhorar a legibilidade. Os melhores resultados foram obtidos usando o grupo coordenadas e ângulos e  $k = 19$ . Os resultados dos testes dos modelos para  $k = 19$  estão na Tabela 2. Como pode ser observado, o melhor valor geral foi obtido por NB, precisamente 85,19% de acurácia.



**Figura 2. Média das acurácias de cada modelo para cada valor de  $k$ .**

Como os autores desse método não fizeram testes com modelos de AM, nós não conseguimos comparar os nossos resultados aos obtidos pelos autores, apesar de ambos os trabalhos focarem em WMs e usarem pontos-chaves. No entanto, resultados promissores foram obtidos com autocorrelação, indicando não apenas a utilidade da característica como proposta pelos autores, mas também seu valor para a identificação de padrões diferentes de CS GMs, neste caso os PR GMs que compõem a maior parte dos exemplos atípicos de nossa base.





**Figura 3. Acurácia de cada modelo para cada valor de  $k$  com coordenadas e ângulos (A) e com apenas ângulos (B).**

**Tabela 2. Resultado dos testes com autocorrelação para  $k = 19$ .**

Modelo	AC	PR	RE	F1	AUC
NB	<b>0.8519</b>	0.8621	<b>0.8621</b>	<b>0.8621</b>	<b>0.8510</b>
KNN	0.7593	<b>0.8636</b>	0.6552	0.7451	0.7676
RF	0.7407	0.7419	0.7931	0.7667	0.7366
SVM	0.7222	0.7333	0.7586	0.7458	0.7193
GB	0.6667	0.7037	0.6552	0.6786	0.6676
DT	0.6667	0.7619	0.5517	0.6400	0.6759
LR	0.5185	0.5385	0.7241	0.6176	0.5021

#### 4.3. Características baseadas em histogramas

Os resultados dos experimentos executados para as características baseadas em histogramas estão sumarizados na Tabela 3. Tanto as características de pose quanto as de velocidade foram avaliadas individualmente e agrupadas. A última linha da tabela corresponde ao resultado da combinação dos dois grupos. É importante destacar que a Tabela 3 mostra apenas o resultado do melhor modelo de AM para cada cenário testado.

Esses resultados demonstraram que características baseadas em histograma são também relevantes na classificação de WMs. Dentre as combinações testadas, o agrupamento das características de velocidade gerou os melhores resultados, sendo 88,89% de acurácia com SVM e LR, e 90,00% de F1-Score com SVM. Novamente não é possível estabelecer uma comparação direta dos nossos resultados com os resultados do artigo original devido aos seguintes fatores: o foco do artigo original eram FMs, enquanto o nosso são os WMs; e pontos-chaves 3D não são utilizados no artigo original.

#### 4.4. Comparação

Os resultados obtidos para cada um dos três métodos de extração de características investigados neste trabalho são comparados na Tabela 4. É importante destacar que são exibidos apenas os resultados do melhor modelo e característica ou subconjunto de características. Além disso, a Tabela 5 mostra as matrizes de confusão correspondentes a cada resultado.

**Tabela 3. Resultado dos testes com características baseadas em histogramas.**

Carac.	Modelo	AC	PR	RE	F1	AUC
HOJO3D	DT	0.7778	0.7429	0.8966	0.8125	0.7683
HOJD3D	LR	0.6667	0.7037	0.6552	0.6786	0.6676
FFT-JO	NB	0.7037	0.7097	0.7586	0.7333	0.6993
FFT-JD	NB	0.7963	0.7812	0.8621	0.8197	0.7910
HOAD3D	KNN	0.7593	0.7500	0.8276	0.7869	0.7538
HORJO3D	NB	0.5926	0.5814	0.8621	0.6944	0.5710
HORJAD3D	KNN	0.6852	0.7727	0.5862	0.6667	0.6931
Pose	NB	0.7222	0.7059	0.8276	0.7619	0.7138
Velocidade	SVM	<b>0.8889</b>	<b>0.8710</b>	<b>0.9310</b>	<b>0.9000</b>	<b>0.8855</b>
Pose & Vel.	NB	0.7407	0.7143	0.8621	0.7812	0.7310

**Tabela 4. Comparação entre os melhores resultados obtidos com cada método.**

Carac.	Modelo	AC	PR	RE	F1	AUC
Histogramas (Velocidade)	SVM	<b>0.8889</b>	<b>0.8710</b>	<b>0.9310</b>	<b>0.9000</b>	<b>0.8855</b>
Autocorrelação ( $k = 19$ )	NB	0.8519	0.8621	0.8621	0.8621	0.8510
Básicas (Frequência)	GB	0.7963	0.7812	0.8621	0.8197	0.7910

Como pode ser observado nas tabelas, o melhor resultado entre todos os testes foi 88,89% de acurácia e 90,00% de F1-Score. Esses valores foram obtidos com todas as características baseadas em velocidade propostas em [McCay et al. 2022] e SVM. No entanto, o teste de Friedman aplicado aos resultados (p-valor igual a 0,657, maior que o nível de significância convencional  $\alpha = 0,05$ ), mostrou que as diferenças de desempenho observadas não são robustas o suficiente para afirmar a superioridade de um método sobre os outros.

Vale apontar que o melhor resultado obtido com o método proposto em [Huang et al. 2024] foi resultante do uso apenas das características em domínio da frequência. Já o melhor resultado obtido com o método de [McCay et al. 2022] sem fusão de características foi com FFT-JD. Isso indica que o uso de características em domínio da frequência é um caminho promissor para a classificação de WMs.

Por fim, é importante destacar que todos os métodos obtiveram elevadas taxas

**Tabela 5. Matrizes de confusão para os melhores resultados obtidos com cada método.**

Carac.	Modelo	VN	FP	FN	VP
Histogramas (Velocidade)	SVM	21	4	2	27
Autocorrelação ( $k = 19$ )	NB	21	4	4	25
Básicas (Frequência)	GB	18	7	4	25

quanto à classificação correta de instâncias da classe atípico (25 ou 27 de 29 instâncias). Essa métrica é fundamental no problema de classificação investigado, pois é preferível aumentar a taxa de verdadeiros positivos mesmo que isso implique em aumento na taxa de falso positivo. Isso mostra que as características geradas nos três métodos conseguem representar padrões significativamente discriminativos entre os GMs atípicos.

## 5. Conclusão

Neste trabalho nós fizemos o teste e a comparação de três métodos de extração de características a partir de pontos-chaves com modelos de AM clássicos para a classificação de WMs entre típico e atípico com uma base de dados própria. Os melhores resultados foram obtidos com as características baseadas em histogramas, mais especificamente com todas as características baseadas em velocidade. Porém, os três métodos geraram características que levaram os métodos de AM a obterem taxas de classificação estatisticamente equivalentes, especialmente na classe de movimentos atípicos. Considerando que a base utilizada nos experimentos é bem pequena, os resultados são promissores, sendo a melhor taxa de F1-Score de 90,00%.

Em trabalhos futuros, seria interessante testar diferentes modelos estimadores de pose, como os testados em [Jahn et al. 2024], e métodos de pré-processamento, como o de [Passmore et al. 2024], assim como outros métodos de extração de características, como os de [Doroniewicz et al. 2020] e [Ji et al. 2024].

## Agradecimentos

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (AUXPE-CAPES-PROEX) - Código de Financiamento 001 e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Adicionalmente, este trabalho foi parcialmente financiado pela Fundação de Amparo à Pesquisa do Estado do Amazonas - FAPEAM - por meio do projeto PDPG-CAPES.

## Referências

- Doroniewicz, I., Ledwoń, D. J., Affanasowicz, A., Kieszczyńska, K., Latos, D., Matyja, M., Mitas, A. W., and Myśliwiec, A. (2020). Writhing movement detection in newborns on the second and third day of life using pose-based feature machine learning classification. *Sensors*, 20(21).
- Einspieler, C., Peharz, R., and Marschik, P. B. (2016). Fidgety movements – tiny in appearance, but huge in impact. *Jornal de Pediatria*, 92(3, Supplement 1):S64–S70.
- Einspieler, C. and Prechtl, H. F. R. (2005). Prechtl’s assessment of general movements: A diagnostic tool for the functional assessment of the young nervous system. *Mental Retardation and Developmental Disabilities Research Reviews*, 11(1):61–67.
- Gholamiangonabadi, D., Kiselov, N., and Grolinger, K. (2020). Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection. *IEEE Access*, 8:133985–133994.
- Hesse, N., Bodensteiner, C., Arens, M., Hofmann, U. G., Weinberger, R., and Schroeder, A. S. (2018). Computer vision for medical infant motion analysis: State of the art and rgb-d data set. In *ECCV Workshops*.

- Huang, X., Huang, C., Yin, W., Huang, H., Xie, Z., Huang, Y., Chen, M., Fan, X., Shang, X., Peng, Z., Wan, Y., Han, T., and Yi, M. (2024). Automatic quantitative intelligent assessment of neonatal general movements with video tracking. *Displays*, 82:102658.
- Jahn, L., Flügge, S., Zhang, D., Poustka, L., Bölte, S., Wörgötter, F., Marschik, P. B., and Kulvicius, T. (2024). Comparison of marker-less 2d image-based methods for infant pose estimation.
- Ji, S., Ma, D., Pan, L., Wang, W., Peng, X., Amos, J. T., Ingabire, H. N., Li, M., Wang, Y., Yao, D., and Ren, P. (2024). Automated prediction of infant cognitive development risk by video: A pilot study. *IEEE Journal of Biomedical and Health Informatics*, 28(2):690–701.
- Leo, M., Bernava, G. M., Carcagnì, P., and Distantè, C. (2022). Video-based automatic baby motion analysis for early neurological disorder diagnosis: State of the art and future directions. *Sensors*, 22(3).
- Letzkus, L., Pulido, J., Adeyemo, A., Baek, S., and Zanelli, S. (2024). Machine learning approaches to evaluate infants’ general movements in the writhing stage—a pilot study. *Scientific Reports*, 14.
- McCay, K. D., Ho, E. S. L., Shum, H. P. H., Fehringer, G., Marcroft, C., and Embleton, N. D. (2020). Abnormal infant movements classification with deep learning on pose-based features. *IEEE Access*, 8:51582–51592.
- McCay, K. D., Hu, P., Shum, H. P. H., Woo, W. L., Marcroft, C., Embleton, N. D., Munteanu, A., and Ho, E. S. L. (2022). A pose-based feature fusion and classification framework for the early prediction of cerebral palsy in infants. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:8–19.
- Passmore, E., Kwong, A. L., Greenstein, S., Olsen, J. E., Eeles, A. L., Cheong, J. L. Y., Spittle, A. J., and Ball, G. (2024). Automated identification of abnormal infant movements from smart phone videos. *PLOS Digital Health*, 3(2):1–21.
- Xia, L., Chen, C.-C., and Aggarwal, J. K. (2012). View invariant human action recognition using histograms of 3d joints. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 20–27.