

RehabMotion: Síntese de Dados de Movimento para Fisioterapia usando Modelos de Linguagem

Lucas D. Brandão
lucas.d.brandao@edu.ufes.br
Universidade Federal do Espírito
Santo (Ufes)
Vitória, ES, Brasil

Giovanni Comarela
gc@inf.ufes.br
Universidade Federal do Espírito
Santo (Ufes)
Vitória, ES, Brasil

Marcelo P. Guimarães
marcelo.guimaraes@unifesp.br
Universidade Federal de São Paulo
(Unifesp)
São Paulo, SP, Brasil

Leonardo Rocha
lrocha@ufsj.edu.br
Universidade Federal de São João
del-Rei (UFSJ)
São João del-Rei, MG, Brasil

Eduardo Zambon
zambon@inf.ufes.br
Universidade Federal do Espírito
Santo (Ufes)
Vitória, ES, Brasil

Diego R. C. Dias
diego.dias@ufes.br
Universidade Federal do Espírito
Santo (Ufes)
Vitória, ES, Brasil

ABSTRACT

The scarcity of labeled human motion data in physical rehabilitation is a major bottleneck for developing intelligent systems for remote patient monitoring, as data collection requires specialized infrastructure and expert annotation. To address this limitation, this article presents RehabMotion, an end-to-end pipeline that specializes large-scale motion language models for the clinical domain without training them from scratch. The proposed approach integrates physical rehabilitation datasets, such as KIMORE and UI-PRMD, with the MotionGPT3 architecture through multi-task fine-tuning based on programmatically generated textual descriptions. To evaluate the quality of the synthetic data, a multi-level assessment combining biomechanical, distributional, and discriminative metrics was implemented. The fine-tuned model achieved high discriminative fidelity, generating synthetic exercises that highly agreed with a reference classifier trained on real data, outperforming the pre-trained baseline which often proved ineffective or harmful due to its generic motion generation. In downstream utility tasks under strict data scarcity scenarios, data augmentation with the fine-tuned synthetic movements substantially increased the classification accuracy, with performance gains saturating as real data became abundant. The complete pipeline, models, and synthetic datasets were made publicly available.

KEYWORDS

Synthetic data generation, physical rehabilitation, motion language models, biomechanical analysis, telehealth

1 INTRODUÇÃO

Até o momento, avaliar se um paciente executa corretamente um exercício de reabilitação é uma tarefa que depende da presença de um fisioterapeuta. Levantar essa avaliação para fora da clínica, permitindo que o acompanhamento aconteça na casa do paciente, é uma das promessas da telessaúde [3], o que tem motivado o desenvolvimento de sistemas capazes de julgar a qualidade de um

movimento de forma automática [2, 6, 18–20, 23, 25]. O obstáculo está nos dados. Treinar esses sistemas de classificação exige exemplos rotulados de movimento humano, e obter tais dados não é trivial: a captura depende de infraestrutura especializada (sistemas como Kinect, Vicon, câmeras RGB ou IMUs) e a rotulagem precisa do julgamento de profissionais de saúde, capazes de distinguir entre uma execução de movimento correta e uma compensação patológica. Como principal consequência de tais dificuldades na obtenção de dados tem-se que *datasets* públicos, como o KIMORE [3] e o UI-PRMD [24], são pequenos em número de pacientes e pouco diversos em padrões patológicos, exatamente a matéria-prima de que os modelos de aprendizado mais carecem.

Diante dessa limitação, a geração de dados sintéticos surge como caminho natural [7, 9, 12]. Em imagens médicas, redes adversárias generativas e modelos de difusão já ampliam conjuntos de treinamento com bons resultados [8]. No domínio do movimento humano, modelos de linguagem especializados, como o MotionGPT [11] e o MotionGPT3 [29], mostram que é possível gerar sequências de movimento 3D plausíveis a partir de descrições textuais. A principal questão que motiva este trabalho é verificar se os modelos citados, quando submetidos a um ajuste fino (*fine-tuning*) sobre dados de reabilitação, conseguem sintetizar movimentos terapêuticos com as características cinéticas específicas do domínio, e, mais importante, se esses movimentos sintéticos são de fato úteis para treinar classificadores quando há poucos dados reais disponíveis.

Este artigo apresenta o RehabMotion, um *pipeline* completo que integra o *dataset* KIMORE ao MotionGPT3 para a geração automatizada de dados sintéticos de reabilitação. O *pipeline* executa: (i) a conversão dos dados de esqueletos do formato KIMORE para o formato HumanML3D; (ii) a geração programática de anotações textuais; (iii) o ajuste fino multitarefa (*multi-task fine-tuning*) no MotionGPT3; (iv) a geração de movimentos sintéticos via descrição textual; (v) a avaliação multinível combinando métricas biomecânicas, distribucionais e discriminativas; e (vi) a validação de utilidade *downstream* com classificadores de exercícios em regimes controlados de escassez de dados reais. Duas versões dos dados sintéticos obtidos neste trabalho foram disponibilizadas publicamente¹: a

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2026). Lavras, Brazil. Porto Alegre: Brazilian Computer Society, 2026.

© 2026 SBC – Brazilian Computing Society.

ISSN 2966-2753

¹<https://zenodo.org/records/20671628>

primeira, indicada como baseline, corresponde ao resultado direto do modelo MotionGPT3 pré-treinado, sem refinamento; e a segunda, denominada fine-tuning, corresponde ao resultado após o ajuste fino em 200 épocas sobre o *dataset* KIMORE. Essas versões permitem a reprodução dos experimentos e a comparação direta do efeito do ajuste fino na qualidade dos resultados. Para avaliar a generalização do método, replicou-se o *pipeline* completo e a validação *downstream* sobre um segundo *dataset*, o UI-PRMD [24].

As principais contribuições deste trabalho são: (i) a elaboração de um *pipeline* reproduzível e modular para adaptar modelos de linguagem de movimento ao domínio de reabilitação física; (ii) uma avaliação de resultados multinível combinando métricas biomecânicas, distribucionais e discriminativas; (iii) uma validação de utilidade *downstream* com classificadores sob escassez de dados controlada, mostrando que o ajuste fino gera dados genuinamente úteis; e (iv) a disponibilização de conjuntos sintéticos públicos para KIMORE e UI-PRMD, em variantes baseline e fine-tuned, juntamente com os *checkpoints* de treinamento dos modelos, todos sob licença CC BY-NC 4.0.

O restante do artigo está organizado da seguinte forma. A Seção 2 discute os trabalhos relacionados. Já a Seção 3 descreve o *pipeline* proposto em detalhes. Os resultados são apresentados na Seção 4 e a Seção 5 discute implicações e limitações do método. Por fim, a Seção 6 conclui o trabalho e indica direções futuras de pesquisa.

2 TRABALHOS CORRELATOS

O *dataset* KIMORE [3] contém dados esqueléticos de 78 participantes, divididos em um grupo de controle e um grupo de patologias, cujos movimentos foram capturados utilizando o Kinect v2 e a seguir anotados com escores clínicos. Por outro lado, o *dataset* UI-PRMD [24] oferece exercícios de reabilitação realizados por dez participantes, todos saudáveis, com captura via uma combinação Kinect+Vicon. Ao contrário do KIMORE, que foca em pacientes reais com patologias, a arquitetura e a proposta do UI-PRMD baseiam-se em capturar movimentos ideais e desvios técnicos simulados.

Também utilizando captura pelo Kinect, o *dataset* IntelliRehabDS [5, 15] engloba 29 indivíduos, com o diferencial de que alguns pacientes realizam os exercícios sentados, por restrições de mobilidade. Por fim, o *dataset* FineRehab [13], composto por 50 participantes, reúne dados multimodais (Kinect+IMU) com rótulos de posição, velocidade e *feedback* do terapeuta. Sua proposta é tentar superar restrições dos *datasets* anteriores através de dados e anotações clínicas mais detalhadas.

De forma geral, todos os *datasets* citados acima compartilham das mesmas limitações: um número reduzido de participantes (menos de 100 por estudo), uma cobertura limitada de padrões de execução e uma coleta presencial cara. Uma vez que tais limitações são difíceis de serem superadas, estratégias para aumento sintético dos dados surgem como uma alternativa natural.

A área de pesquisa em aumento sintético dos dados (*data augmentation*) evoluiu de modelos autorregressivos (e.g., redes neurais recorrentes) para modelos generativos mais robustos, como o CVAE (*Conditional Variational Autoencoder*) [17] e modelos de difusão [4, 22]. O atual estado da arte envolve a integração com modelos de linguagem (LLMs), como T2M-GPT [27], MotionGPT [11] e MotionGPT3 [29].

O T2M-GPT [27] discretiza os movimentos utilizando a técnica de VQ-VAE (*Vector Quantized Variational Autoencoder*), que transforma uma sequência de poses 3D em uma sequência de números inteiros (“*palavras de movimento*” quantizadas em um vetor). A seguir, tais sequências podem ser empregadas em comandos de texto para gerar novos movimentos físicos realistas. Por sua vez, o MotionGPT [11] unifica tanto a geração quanto a compreensão do movimento sob uma única interface de linguagem natural, permitindo comandos como “*Produza uma caminhada*” (geração) ou “*O que essa pessoa está fazendo?*” (compreensão) em uma mesma interface. Por fim, o MotionGPT3 [29], usado neste trabalho, amplia esse paradigma com um modelo maior e suporte multitarefa nativo, permitindo ao modelo classificar se um exercício de reabilitação está certo ou errado, gerar novas variações desse exercício e traduzir movimentos em texto descritivo. Tais modelos de linguagem empregam o formato HumanML3D [10], que tornou-se o padrão de fato no domínio. Esse formato descreve o esqueleto humano usando uma lista de 263 propriedades físicas (dimensões) e representa a rotação de articulações por seis números (6D) [28].

A literatura de geração de movimento especificamente para reabilitação é recente. Abedi *et al.* [1] aplicam técnicas de *augmentation* aos vídeos RGB do KIMORE, re-extraindo esqueletos dos vídeos aumentados. A abordagem diversifica o espaço visual mas não gera padrões novos. Por sua vez, Mennella *et al.* [14] geram imagens RGB de exercícios via modelos de difusão, sem fornecer esqueletos 3D. A técnica mais próxima a do presente trabalho é a *Error-Guided Pose Augmentation* (EGPA) [21], que sintetiza esqueletos simulando erros clinicamente relevantes via regras pré-definidas combinadas com modelos de A-GCN (*Adaptive Graph Convolutional Network*). O EGPA reduz o erro absoluto dos dados sintéticos em 27.6% para os *datasets* KIMORE e UI-PRMD, mas requer especificação manual dos padrões de erro por exercício, limitando a sua escalabilidade.

O RehabMotion difere dessas abordagens em dois aspectos: (i) a técnica emprega um modelo de linguagem de movimento pré-treinado em larga escala para gerar movimentos novos via descrição textual, abordagem não explorada anteriormente em reabilitação; e (ii) o controle semântico via texto habilita especificação granular de padrões de execução sem intervenção manual por exercício, tornando o *pipeline* escalável para novos protocolos.

3 METODOLOGIA

Esta seção descreve o *pipeline* RehabMotion em detalhes. A Figura 1 apresenta uma visão geral de todas as etapas de execução. Antes do seu detalhamento, vale destacar algumas etapas: a conversão de formato (Seção 3.2) e a anotação textual (Seção 3.3) são pré-requisitos de engenharia, ao passo que duas decisões concentram o impacto sobre os resultados, o ajuste fino multitarefa (Seção 3.4) e o alinhamento de representação entre real e sintético no classificador *downstream* (Seção 3.7).

3.1 O Dataset KIMORE

Os dados presentes no KIMORE [3] foram coletados com sensor Kinect v2 em ambiente clínico controlado, rastreamento 25 articulações do esqueleto humano a uma taxa de 30 quadros por segundo (*fps – frames per second*). O *dataset* contém dados de 78 participantes, divididos em um grupo controle, formado por participantes

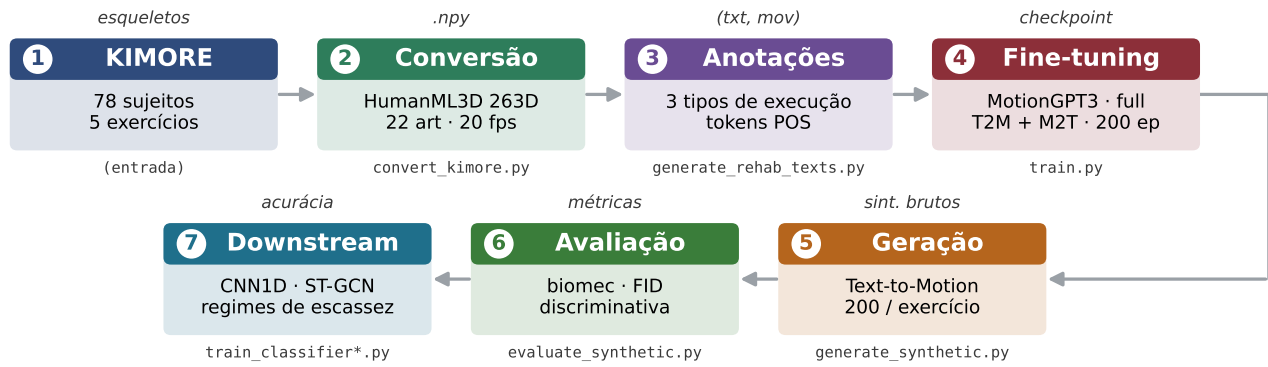


Figura 1: Visão geral do *pipeline* RehabMotion. As etapas são executadas sequencialmente, desde o *dataset* bruto até a avaliação biomecânica dos movimentos sintéticos.

Tabela 1: Exercícios do protocolo KIMORE utilizados neste trabalho. As classes Es2, Es3 e Es4 envolvem movimentos bilaterais registrados em duas sub-gravações (esquerda/direita).

ID	Descrição	Foco anatômico
Es1	Elevação dos braços no plano sagital (frente, sobre a cabeça)	Ombro
Es2	Inclinação lateral do tronco com braços em extensão	Coluna torácica/lombar
Es3	Rotação do tronco, posição sentada	Coluna torácica
Es4	Rotação pélvica no plano transversal, em pé	Coluna lombar
Es5	Agachamento (flexão de joelhos no plano sagital)	Membros inferiores

sem patologias; e em cinco grupos com patologias documentadas do tronco (GP1–GP5). Este trabalho utiliza todo o protocolo de cinco exercícios presentes no *dataset*, detalhados na Tabela 1. Cada exercício bilateral é gravado separadamente para os lados esquerdo e direito; após conversão e remoção de sequências com quadros corrompidos (nove exclusões com valores inválidos), o conjunto completo (sem *split* de dados) totalizou 2.797 sequências.

3.2 Conversão para o Formato HumanML3D

O MotionGPT3 foi originalmente treinado sobre o *dataset* HumanML3D [10], que adota o modelo corporal SMPL (*Skinny Multi-Person Linear model*), capturando 22 articulações esqueléticas a uma taxa de 20 *fps*. Devido às diferenças no número de articulações e na taxa de captura, torna-se necessária uma conversão dos dados do KIMORE para o formato SMPL. Esta conversão é descrita brevemente a seguir e está implementada na etapa 2 do *pipeline* da Figura 1 (programa `convert_kimore.py`).

3.2.1 Mapeamento KIMORE para SMPL e reamostragem. O mapeamento entre os dois esqueletos foi construído por correspondência anatômica direta para 21 das 22 articulações do SMPL. A única exceção é a articulação *spine3* do SMPL, que foi aproximada por *SpineMid* do KIMORE.

Quanto às taxas de amostragem, as sequências originais do KIMORE em 30 *fps* foram reamostradas para os 20 *fps* do SMPL por interpolação linear das coordenadas 3D de cada articulação. O erro introduzido pela interpolação é desprezível frente à precisão do Kinect v2 ($\pm 2\text{--}4$ mm).

3.2.2 Extração de Características 263D. Cada quadro de captura $\mathbf{p} \in \mathbb{R}^{22 \times 3}$ (posição das 22 articulações em um espaço 3D) é transformado em um vetor $\mathbf{h} \in \mathbb{R}^{263}$, com cinco grupos funcionais: (1) dados da raiz (quadril/centro de gravidade – $\mathbb{R}^4$): velocidade angular, velocidades lineares em x e z , altura em y ; (2) posições RIC ($\mathbb{R}^{63}$): posições relativas das 21 articulações no sistema local da raiz; (3) rotações em 6D [28] ($\mathbb{R}^{126}$): informações de rotação de cada articulação; (4) velocidades articulares ($\mathbb{R}^{66}$): derivadas temporais discretas das 22 articulações; e (5) contatos com o solo ($\mathbb{R}^4$): indicadores binários de pé e tornozelo. A concatenação resulta em $4 + 63 + 126 + 66 + 4 = 263\text{D}$. As sequências são normalizadas por $\mathbf{h}' = (\mathbf{h} - \boldsymbol{\mu}) / \sigma$, onde $\boldsymbol{\mu}, \sigma$ são a média e desvio padrão calculados sobre todo o conjunto de quadros do participante.

3.3 Geração de Anotações Textuais

A etapa 3 do *pipeline* RehabMotion (arquivo `generate_rehab_texts.py`) gera descrições textuais em inglês cobrindo três tipos de execução para cada exercício: (i) execução normal, e.g., “*a person raises both arms forward and up to shoulder height*”; (ii) amplitude reduzida, e.g., “*a person performs lateral trunk tilt with limited range of motion*”; e (iii) com movimentos compensatórios, e.g., “*a person performs a squat with slight compensatory trunk movement*”. Foram preparados 10 *prompts* distintos por exercício (totalizando 50 *prompts* base), cobrindo essa diversidade semântica. As descrições seguem o formato do HumanML3D, com pares de *tokens* POS (*Part-of-Speech* – substantivos, verbos, etc.) e marcadores temporais para determinar a evolução dos movimentos.

3.4 Ajuste Fino Multitarefa do MotionGPT3

3.4.1 Arquitetura. O MotionGPT3 [29] adota uma arquitetura MoT (*Motion-Tokens*) que trata o movimento como segunda modalidade linguística, integrando “*tokens* de movimento” com o processamento de linguagem natural, no caso, o inglês. A arquitetura combina:

(i) um mecanismo de *Motion Latent Diffusion* (MLD-VAE) que codifica/decodifica sequências 263D em um espaço latente; (ii) um transformador autorregressivo GPT-2 que opera sobre *tokens* unificados de texto e movimento; e (iii) uma cabeça de difusão que reconstrói o espaço latente de movimento nas saídas do GPT-2. Essa decomposição do MotionGPT3 permite tradução bidirecional entre texto e movimento.

3.4.2 Estratégia de Treinamento. O ajuste fino partiu dos pesos pré-treinados do MotionGPT3 sobre o HumanML3D e cobriu duas tarefas conjuntas para evitar o esquecimento catastrófico do conhecimento geral de movimento: **T2M** (texto para movimento) e **M2T** (movimento para texto). Na primeira tarefa, dado um *prompt* textual $\mathbf{t}$, o refinamento busca a maximização de $p(\mathbf{m} | \mathbf{t})$, isto é, a probabilidade p do movimento $\mathbf{m}$ acontecer dado o texto $\mathbf{t}$. A segunda tarefa é a dual da primeira e busca maximizar a probabilidade do texto $\mathbf{t}$ estar correto dado o movimento $\mathbf{m}$, isto é, maximizar $p(\mathbf{t} | \mathbf{m})$. Ambas as tarefas compartilham o mesmo conjunto de parâmetros e foram amostradas a partir de um modelo unificado de instruções, conforme especificado nos arquivos `rehab_finetune.yaml` e `template_t2m_instructions.json` da implementação².

3.4.3 Detalhes de Implementação. A implementação do ajuste fino foi desenvolvida em Python 3.11 e PyTorch 2.0, e executada em uma GPU NVIDIA RTX 5090 com 32 GB de VRAM. Utilizou-se o otimizador AdamW, o escalonador CosineAnnealingLR, um *batch size* de 16 sequências e 200 épocas, das quais o melhor *checkpoint* foi selecionado pelo FID (*Fréchet Inception Distance*) de validação (veja Seção 3.6). O FID é uma métrica amplamente usada na área de Inteligência Artificial Generativa. No caso do movimento humano, o FID compara os movimentos gerados com movimentos feitos por humanos reais: um FID baixo significa que os movimentos são fluidos e naturais; já um FID alto significa que os movimentos estão robóticos, bizarros ou repetitivos.

O treinamento completo do ajuste fino sobre todo conjunto de treinamento levou cerca de 2 horas e a inferência **T2M** leva cerca de 3 segundos por amostra (Seção 3.5). Na configuração principal, foram gerados 1.000 movimentos sintéticos balanceados (200 por exercício) a partir de 10 *prompts* por exercício, número ampliado para 30 no estudo de ablação (Seção 4.5). A Tabela 2 resume os parâmetros utilizados na implementação do ajuste fino (arquivo `train.py` – etapa 4 do *pipeline* da Figura 1).

3.5 Geração de Dados Sintéticos

A inferência é realizada em modo **T2M** pelo programa `generate_synthetic.py`, que estabelece a etapa 5 do *pipeline* do RehabMotion. Dado um *prompt* textual, o transformador gera espaços latentes que são decodificados pelo MLD-VAE em vetores 263D e desnormalizados via $\mathbf{h} = \mathbf{h}' \cdot \boldsymbol{\sigma} + \boldsymbol{\mu}$, o inverso da normalização realizada na Seção 3.2. Foram geradas 200 amostras por exercício, totalizando 1.000 movimentos balanceados. O *checkpoint* usado para a geração foi aquele que minimizou o FID sobre o conjunto de validação durante o ajuste fino, selecionado automaticamente pelo mecanismo de *model checkpointing* da biblioteca PyTorch Lightning.

Tabela 2: Parâmetros de implementação da etapa de ajuste fino no RehabMotion.

Parâmetro	Valor
Modelo base	MotionGPT3 (pesos HumanML3D)
Framework	PyTorch 2.0 / Python 3.11
GPU	NVIDIA RTX 5090 32 GB
Otimizador	AdamW ($\beta_1 = 0,9$, $\beta_2 = 0,99$)
Taxa de aprendizado	5×10^{-6}
Escalonador	CosineAnnealingLR ($T_{\max} = 500$, $\eta_{\min} = 10^{-7}$)
Batch size	16 sequências
Épocas	200 (melhor <i>checkpoint</i> via FID de validação)
Tempo de treinamento	≈ 2 horas
Taxa de reamostragem	30 $\rightarrow$ 20 fps (interpolação linear)
Dimensão de entrada	263 (vetores HumanML3D)
Articulações	22 (SMPL)
Movimentos gerados	1.000 (200 por exercício, balanceado)
Prompts por exercício	10 (principal); 30 (ablação, Seção 4.5)

3.6 Avaliação Biomecânica e Distribucional

A avaliação da etapa 6 do *pipeline* combina três famílias de métricas: biomecânicas, distribucionais e discriminativas.

3.6.1 Métricas biomecânicas. Foram empregadas três métricas físicas e biológicas para avaliação biomecânica. A primeira é o *jerk* médio, calculado como a derivada de terceira ordem das posições articulares com $\Delta t = 0.05$ s. Esta métrica serve como um indicador de ruído/suavidade da trajetória: um *jerk* médio alto indica espasmos de movimento, enquanto um valor baixo caracteriza um movimento fluido e suave.

A segunda métrica empregada é a validade anatômica de ROM (*Range of Motion*), que calcula a fração de quadros em que os ângulos dos cotovelos e joelhos permanecem dentro do intervalo $[0^\circ, 180^\circ]$, conforme os limites naturais de dobra dessas articulações [16].

A terceira métrica biomecânica, chamada de *foot skating*, avalia que o esqueleto gerado não está flutuando/deslizando no chão. Ela mede a fração de quadros em que um pé em contato com o solo (distância $y < 5$ cm da posição mais baixa) exibe uma velocidade horizontal (para os lados, para a frente ou para trás) superior a 7.5 cm por quadro.

3.6.2 Métricas distribucionais. Métricas distribucionais medem se o conjunto sintético se parece, em agregado, com um conjunto real.

A primeira métrica empregada analisa a velocidade articular agregada via diferenças finitas, isto é, pela variação de posição 3D entre quadros sucessivos. A seguir, todas as velocidades computadas dão origem a dois histogramas, um real e outro sintético, que são finalmente comparados pela divergência Kullback-Leibler D_{KL} . O valor desejável de $D_{KL} = 0$ indica histogramas idênticos; por outro lado, quanto maior o valor de D_{KL} , mais os dados sintéticos diferem de um movimento real.

A segunda métrica distribucional utilizada é a Distância de Fréchet (FID) por classe de movimento. Cada sequência é reduzida ao vetor médio temporal (ao longo de todos os quadros) das suas características 263D, de forma que uma sequência é representada por um único ponto no espaço 263D. A seguir, os pontos de cada classe c dão origem a uma distribuição Gaussiana $\mathcal{N}^c(\boldsymbol{\mu}^c, \boldsymbol{\Sigma}^c)$, onde

²Disponível em <https://zenodo.org/records/20671626>

a média μ^c representa o “centro” do movimento, e a matriz de covariância Σ^c captura a diversidade de cada gesto. Finalmente, as distribuições por classe dos dados reais $\mathcal{N}_r^c(\mu_r^c, \Sigma_r^c)$ e sintéticos $\mathcal{N}_s^c(\mu_s^c, \Sigma_s^c)$ podem ser comparadas pelo FID da classe, calculado por

$$FID = \|\mu_r - \mu_s\|^2 + \text{Tr}(\Sigma_r + \Sigma_s - 2(\Sigma_r \Sigma_s)^{1/2}), \quad (1)$$

onde Tr é o traço (soma da diagonal principal) da matriz quadrada resultante da diferença geométrica das duas matrizes de covariância. A Equação 1 acima é definida para cada classe de movimento c , mas o sobrescrito foi deixado implícito para facilitar a legibilidade. Diferentemente do FID padrão da literatura de geração de movimento, esta formulação opera diretamente sobre o espaço HumanML3D 263D, eliminando dependência de extratores neurais auxiliares e permitindo o cálculo por classe mesmo em conjuntos pequenos.

3.6.3 Métricas discriminativas. As métricas anteriores não verificam se um classificador treinado em dados reais consegue atribuir cada amostra sintética à sua classe pretendida. Entretanto, tal característica é essencial para prever a utilidade *downstream* dos dados sintéticos. Assim, alguma métrica discriminativa torna-se necessária. Este trabalho propõe um teste simples: treinar uma CNN 1D leve ($\approx 113k$ parâmetros) sobre o conjunto real e utilizá-la para avaliar o conjunto sintético, comparando a classe predita com a induzida pelo *prompt*. Uma taxa de concordância alta indica que o sintético é discriminativamente fiel, e não apenas distribucionalmente próximo.

3.7 Validação de Utilidade Downstream

Para verificar se o ajuste fino produz dados sintéticos úteis, foi avaliado o ganho em uma tarefa *downstream* de classificação de exercícios sob regimes controlados de escassez de dados reais. A validação constrói subconjuntos estritamente disjuntos por paciente, estratificados por grupo clínico, com uma separação de 55 sujeitos utilizados para treinamento, 12 para validação e 11 para teste, totalizando os 78 sujeitos originais do *dataset* KIMORE.

Foram treinadas duas arquiteturas discriminativas: (i) uma CNN 1D (diferente da definida na Seção 3.6.3) de três camadas convolucionais sobre poses centradas na pélvis ($\approx 113k$ parâmetros); e (ii) uma ST-GCN (*Spatial-Temporal Graph Convolutional Network*) [26] de três blocos espaço-temporais que opera sobre o modelo SMPL ($\approx 385k$ parâmetros). Para alinhar a representação entre real e sintético, cada movimento passa por um ajuste que envolve: (i) a normalização de escala por amostra (divisão pela raiz do segundo momento das posições centradas na pélvis) e (ii) uma reamostragem uniforme para 50 quadros.

O *grid* de validação varre cinco regimes de escassez de sujeitos reais no treino, $n_{\text{real}} \in \{2, 5, 10, 20, 55\}$, e três tamanhos de *augmentation* sintético, $n_{\text{aug}} \in \{250, 500, 1.000\}$, totalizando 15 possíveis combinações. Cada combinação é avaliada sob três sementes aleatórias e a ferramenta reporta a média e o desvio padrão da acurácia no conjunto de teste real (sujeitos nunca vistos). Os resultados da validação estão apresentados na Seção 4.4.

4 RESULTADOS

A avaliação compara três configurações: **Real**, o conjunto de teste do KIMORE convertido para HumanML3D (395 sequências para o

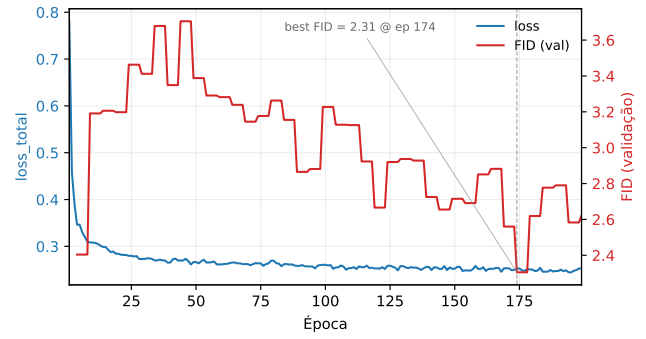


Figura 2: Curvas de perda total e FID de validação ao longo das 200 épocas de ajuste fino. Linha tracejada vertical: época do melhor *checkpoint* selecionado para geração.

conjunto de teste); **BL** (*baseline*), um conjunto de 1.000 movimentos gerados pelo MotionGPT3 pré-treinado, sem o ajuste fino da Seção 3.4; e **FT** (*fine-tuning*), um conjunto de 1.000 movimentos gerados pelo MotionGPT3 após o ajuste fino multitarefa descrito na Seção 3.4. Cada variante sintética é balanceada (200 amostras por exercício).

4.1 Convergência do Treinamento

Durante o ajuste fino, a perda total caiu de 0.78 (época 1) para 0.25 (época 200), e o FID computado de validação do MotionGPT3 caiu de 3.46 (época inicial) para 2.31 (época 174 – melhor *checkpoint*), uma redução de aproximadamente 33%. A Figura 2 apresenta as curvas de perda e FID ao longo das 200 épocas.

4.2 Avaliação Biomecânica e Distribucional

A Tabela 3 consolida os valores obtidos com as métricas de avaliação biomecânica e distribucional definidas na Seção 3.6. O *baseline* já apresenta validade anatômica perfeita (100% de quadros dentro do ROM de cotovelo e joelho) e *foot skating* desprezível (0.01%), confirmando que o modelo pré-treinado não produz poses anatomicamente impossíveis nem inconsistências grosseiras de contato com o solo. Após o ajuste fino, essas duas garantias se preservam: validade de ROM de 100% e *foot skating* de 0.08%, um valor ainda desprezível em termos absolutos. O ganho do ajuste fino se materializa nas demais métricas, todas relacionadas à fidelidade da distribuição sintética em relação à real.

4.2.1 Velocidade articular. A velocidade articular média do KIMORE real é 0.245 m/s. As duas variantes sintéticas se posicionam próximas deste valor: **BL** em 0.287 m/s (+17%) e **FT** em 0.258 m/s (+5%). A Figura 3 apresenta as distribuições agregadas, evidenciando que o ajuste fino aproxima a distribuição sintética do regime real. Como mostrado adiante nas Seções 4.3 e 4.4, esta semelhança distribucional agregada não é suficiente: o **BL** e o **FT** podem ter velocidades agregadas similares e ainda assim diferir substancialmente em termos discriminativos e nos resultados *downstream*.

4.2.2 FID por classe de movimento. O FID 263D agregado é -11% menor para o **FT** (7.87 vs 8.83 para **BL**). O maior ganho (acima de 50%) concentra-se nas classes Es1 e Es5 do HumanML3D, que

Tabela 3: Métricas biomecânicas e distribucionais por configuração. Os valores das colunas BL e FT foram calculados sobre as 1.000 sequências sintéticas de cada variante. A coluna Δ FT/BL indica a variação relativa do ajuste fino em relação ao *baseline*.

Métrica	BL	FT	Δ FT/BL
Validade ROM cotovelo/joelho (%)	100	100	–
Foot skating médio (% de quadros)	0.01	0.08	+700%
Velocidade articular média (m/s) [Real: 0.245]	0.287	0.258	-10%
FID agregado	8.83	7.87	-11%
FID Es1 (elevação sagital de braços)	15.63	7.11	-55%
FID Es2 (inclinação lateral de tronco)	11.96	11.96	–
FID Es3 (rotação de tronco)	10.42	10.10	-3%
FID Es4 (rotação pélvica)	10.34	11.50	+11%
FID Es5 (agachamento)	13.40	5.34	-60%

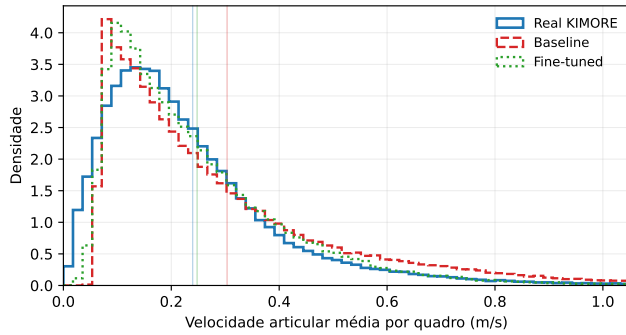


Figura 3: Distribuições agregadas de velocidade articular (m/s) para os dados reais (azul sólido), baseline (vermelho tracejado) e fine-tuned (verde pontilhado). Após o ajuste fino, a distribuição sintética torna-se substancialmente mais próxima da real, refletindo a redução de 72% na divergência D_{KL} .

são também aquelas em que o ajuste fino mais se aproxima do KIMORE em valor absoluto, evidenciando a especialização de domínio adquirida. Já as classes de rotação do tronco (Es3) e rotação pélvica (Es4) tiveram, respectivamente, um ganho marginal e uma pequena perda, consistente com o fato de serem compostas por pequenas variações de orientação articular dificilmente discrimináveis por estatísticas distribucionais globais. Essa limitação motiva a avaliação discriminativa da seção a seguir.

4.3 Avaliação Discriminativa

Para verificar se o sintético é discriminativamente fiel e não apenas distribucionalmente próximo do real, aplicou-se o teste descrito na Seção 3.6.3. Um classificador CNN 1D foi treinado nos 55 sujeitos de treino do KIMORE, obtendo um valor de acurácia de 92.8%. A seguir, esse classificador foi executado sobre cada uma das 1.000 amostras sintéticas e a classe predita é comparada à classe induzida pelo *prompt*.

A Tabela 4 consolida os resultados para ambas as variantes sintéticas. Percebe-se que a fidelidade discriminativa separa **FT** e **BL** de forma muito mais clara do que o FID agregado. O ajuste fino atinge

Tabela 4: Avaliação discriminativa das duas variantes sintéticas BL e FT (1.000 amostras cada). Classificador filtro: CNN 1D ($\approx$ 113k parâmetros) treinado nos 55 sujeitos de treino reais. Para cada variante, tem-se: (i) o percentual de concordância (coluna % Conc.), que indica a proporção de amostras cuja classe predita coincide com o *prompt*; e (ii) a confiança (coluna Conf. – valores próximos a 1.0 são melhores).

Classe	BL		FT	
	% Conc.	Conf.	% Conc.	Conf.
Es1 (elevação sagital de braços)	80.5	0.81	88.5	0.87
Es2 (inclinação lateral de tronco)	9.5	0.11	89.0	0.89
Es3 (rotação de tronco)	4.5	0.06	85.5	0.84
Es4 (rotação pélvica)	8.0	0.10	83.0	0.78
Es5 (agachamento)	60.0	0.60	100	1.00
Geral	32.5	0.34	89.2	0.88

89.2% de concordância geral (com confiança média de 0.88), contra apenas 32.5% (0.34 de confiança) para o *baseline* pré-treinado. Tal diferença é ainda mais pronunciada nas classes específicas de reabilitação: *inclinação lateral de tronco* (89.0% **FT** vs. 9.5% **BL**), *rotação de tronco* (85.5% vs. 4.5%) e *rotação pélvica* (83.0% vs. 8.0%), exercícios estes que o HumanML3D original não conhece. Por outro lado, o agachamento é familiar a ambas as variantes e por isso ambas o reproduzem bem (**FT** 100%, **BL** 60%). Esses números indicam que o ajuste fino produz amostras genuinamente reconhecíveis como o exercício pretendido: o classificador treinado em dados reais “enxerga” o sintético **FT** da mesma maneira que veria outros exemplos reais daquela classe.

4.4 Utilidade Downstream

A Tabela 5 apresenta a acurácia de teste (média $\pm$ desvio de 3 sementes aleatórias) das duas arquiteturas (CNN 1D e ST-GCN, ver Seção 3.7) sob três condições: somente dados reais (coluna **RE**), dados reais acrescidos dos dados sintéticos *baseline* (coluna **RE +BL**), e dados reais acrescidos dos dados com ajuste fino (coluna **RE +FT**). Em todas as condições variou-se o número de sujeitos reais entre 2 e 55.

Pelos números na Tabela 5 percebe-se que os dados sintéticos **FT** elevam substancialmente a acurácia em escassez de dados reais, sobretudo na arquitetura ST-GCN. Com apenas dois sujeitos reais, o **FT** leva a uma melhora de quase 15% no resultado, e este ganho persiste também para $n_{\text{real}} = 5$, atenuando-se quando o número de sujeitos reais aumenta. A arquitetura CNN 1D, mais simples e já próxima do teto (86.9% de acurácia com $n_{\text{real}} = 2$), ainda tem um ganho de 6.8% e satura em $n_{\text{real}} \geq 5$.

Os dados sintéticos do *baseline* (**BL**) não são uma fonte de *augmentation* confiável. Seu efeito depende da arquitetura: para a CNN 1D, o **BL** prejudica nos cenários com dados reais escassos e é inócuo quando n_{real} cresce. Essa segunda situação se repete para a ST-GCN. Por outro lado, a vantagem do **FT** sobre o **BL** é grande em todos os regimes. A Tabela 4 explica o contraste: apenas 32.5% das amostras do **BL** são reconhecíveis como o exercício pretendido (contra 89.2% do **FT**), e esse “movimento humano genérico” introduz ruído de rótulo que a CNN 1D absorve mais que a ST-GCN.

Tabela 5: Acurácia de teste (%) do classificador *downstream* para regimes crescentes de dados reais, com *augmentation* por 1.000 amostras sintéticas. Três sementes aleatórias por configuração. As colunas FT-RE e FT-BL indicam a variação absoluta do percentual de acurácia entre o ajuste fino (FT) e os valores reais (RE) e *baseline* (BL), respectivamente. Um valor positivo indica melhora.

n_{real}	Arquit.	RE	RE+BL	RE+FT	FT-RE	FT-BL
2	CNN 1D	86.9 ± 6.0	68.8 ± 3.5	93.7 ± 0.9	+6.8	+24.9
	ST-GCN	55.7 ± 1.6	56.1 ± 5.9	70.4 ± 2.5	+14.7	+14.3
5	CNN 1D	92.8 ± 1.9	83.2 ± 3.3	91.9 ± 0.9	-0.9	+8.7
	ST-GCN	66.2 ± 8.6	68.6 ± 5.6	81.9 ± 1.6	+15.7	+13.2
10	CNN 1D	95.8 ± 2.5	92.6 ± 1.3	94.9 ± 0.6	-0.8	+2.4
	ST-GCN	86.2 ± 3.0	84.6 ± 3.7	87.8 ± 3.5	+1.5	+3.1
20	CNN 1D	97.1 ± 1.4	95.2 ± 1.0	96.5 ± 1.2	-0.7	+1.3
	ST-GCN	89.3 ± 4.3	89.1 ± 4.3	91.5 ± 4.1	+2.2	+2.4
55	CNN 1D	97.5 ± 0.7	97.6 ± 0.6	98.4 ± 0.1	+0.9	+0.8
	ST-GCN	95.7 ± 1.5	97.7 ± 0.5	98.5 ± 0.2	+2.8	+0.8

Estes resultados confirmam que é o ajuste fino, e não o volume de dados sintéticos em si, que gera a utilidade do *dataset*.

Os dados da Tabela 5 confirmam que o efeito de *augmentation* satura quando dados reais ficam abundantes. A partir de $n_{\text{real}} \geq 10$ o sintético oferece ganhos diminutos ou estabiliza-se em torno do RE. Esse padrão é desejado e esperado: o processo de *augmentation* não substitui dados reais; ele compensa a sua escassez. A Figura 4 ilustra o comportamento das duas arquiteturas, onde pode-se perceber visualmente: o estouro do FT e a nocividade do BL para os cenários com poucos sujeitos reais, e a convergência da acurácia quando o volume de dados reais aumenta.

4.5 Ablação: Escala e Filtro dos Dados Sintéticos

Os estudos da seção anterior utilizam 1.000 amostras sintéticas balanceadas (200 por exercício). Esta seção descreve a análise de ablação realizada, onde se investigou se essas amostras já saturam o ganho *downstream* ou se uma variação no conjunto (seja maior, ou um subconjunto filtrado por qualidade ou diversidade), traria uma melhora adicional. Para tal, foram geradas 5.000 amostras FT adicionais (denotadas **Raw**), separadas de forma balanceada em 1.000 por exercício, sob os mesmos *prompts* por classe usados na configuração principal. A seguir, foram construídos dois subconjuntos filtrados: **Q1k**, formado pelos 200 melhores valores por exercício, segundo a confiança do classificador discriminativo da Seção 4.3; e **D1k**, formado pelos 200 melhores valores por exercício escolhidos por seleção *greedy max-min* no espaço 263D para maximizar diversidade inter-amostra.

A Tabela 6 reporta os resultados do estudo de ablação para o regime mais sensível, com $n_{\text{real}} = 2$. Esta ablação usa o *dataloader* original, sem o alinhamento de representação da Seção 3.7. Por isso, os valores de ΔFT da configuração principal (primeira linha da Tabela 6) correspondem ao valor pré-alinhamento, e o objetivo é comparar o tamanho e a filtragem dos conjuntos sob condições idênticas, e não os valores absolutos da Tabela 5.

Tabela 6: Ablação de conjuntos sintéticos para $n_{\text{real}} = 2$, e três sementes aleatórias.

Conjunto sintético	$\Delta\text{FT CNN 1D}$	$\Delta\text{FT ST-GCN}$
FT 1.000 (principal)	+3.3	+11.5
Raw (5.000 não-filtrado)	+1.5	+12.8
Q1k (<i>top</i> 1.000 por confiança)	+5.5	+5.3
D1k (<i>top</i> 1.000 por diversidade)	+3.9	+5.7

Tabela 7: Utilidade *downstream* no UI-PRMD ($n_{\text{aug}} = 1.000$, três sementes aleatórias). O ganho do ajuste fino replica o achado do KIMORE no regime de escassez extrema ($n_{\text{real}} = 1$).

n_{real}	Arquit.	RE	RE+BL	RE+FT	FT-RE	FT-BL
1	CNN 1D	71.0 ± 5.7	65.7 ± 11.2	85.2 ± 5.6	+14.2	+19.5
	ST-GCN	51.3 ± 8.8	58.2 ± 11.8	66.8 ± 4.0	+15.5	+8.7
2	CNN 1D	84.7 ± 2.2	86.8 ± 2.9	84.0 ± 3.9	-0.7	-2.8
	ST-GCN	70.8 ± 9.0	76.2 ± 2.7	75.3 ± 6.6	+4.5	-0.8
3	CNN 1D	90.2 ± 5.0	89.7 ± 2.9	86.2 ± 4.7	-4.0	-3.5
	ST-GCN	81.5 ± 5.0	83.0 ± 0.0	79.3 ± 6.2	-2.2	-3.7
5	CNN 1D	96.7 ± 3.0	94.0 ± 2.4	99.3 ± 0.2	+2.7	+5.3
	ST-GCN	90.7 ± 3.1	97.5 ± 0.4	94.3 ± 2.0	+3.7	-3.2
6	CNN 1D	97.3 ± 0.2	96.2 ± 1.3	99.5 ± 0.0	+2.2	+3.3
	ST-GCN	94.5 ± 6.4	94.2 ± 3.3	98.3 ± 0.6	+3.8	+4.2

Três observações resultam da Tabela 6. Primeiro, percebe-se que escalar de 1.000 para 5.000 amostras proporciona um ganho mínimo na ST-GCN e reduz o ganho na CNN 1D (+3.3 $\rightarrow$ +1.5), sugerindo que o benefício marginal de mais amostras converge rápido para a escala deste estudo. Segundo, a filtragem por confiança discriminativa não é universalmente benéfica: ajuda a CNN 1D (que sofre mais com ruído de rótulo) mas degrada a ST-GCN (que aparentemente extrai mais sinal de amostras imperfeitas via redundância). Terceiro, a diversidade explícita no espaço de características 263D não compensa a perda absoluta de amostras quando o conjunto é reduzido para 1.000. Simultaneamente, essas observações indicam que a configuração principal de 1.000 amostras não-filtradas está próxima do ótimo prático para esse regime de tarefa, e que estratégias de filtragem mais elaboradas se tornariam relevantes apenas com conjuntos sintéticos substancialmente maiores ou em tarefas mais sensíveis a ruído de rótulo.

4.6 Validação para *Dataset* UI-PRMD

Para verificar se o ganho *downstream* é específico para o KIMORE ou generaliza para outros *datasets*, foi replicado todo o *pipeline* sobre o *dataset* UI-PRMD [24], composto por dez exercícios, realizado por dez sujeitos saudáveis. Como o UI-PRMD possui poucos sujeitos, a grade de escassez é diferente: $n_{\text{real}} \in \{1, 2, 3, 5, 6\}$. A Tabela 7 reporta a acurácia de teste.

O resultado central obtido para o KIMORE se repete no UI-PRMD. No regime de escassez extrema de dados reais ($n_{\text{real}} = 1$), a inclusão dos dados sintéticos do ajuste fino eleva ambas as arquiteturas substancialmente (coluna FT-RE): +14.2 na CNN 1D e +15.5 na ST-GCN. Como esperado, o ganho é maior na escassez extrema e atenua conforme os dados reais crescem, mantendo-se positivo

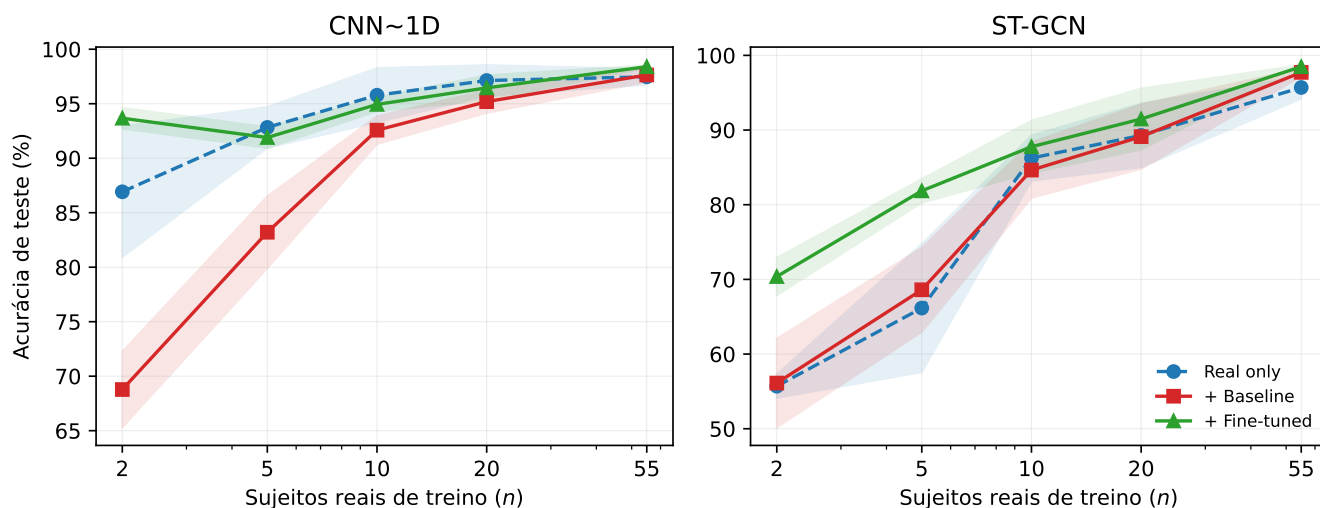


Figura 4: Acurácia de teste (eixo y) em função do número de sujeitos reais de treino (eixo x , escala logarítmica) para as três condições, considerando $n_{\text{aug}} = 1\,000$ e apresentando a CNN 1D (esquerda) e o ST-GCN (direita). Banda sombreada: desvio padrão sobre 3 sementes aleatórias.

para $n_{\text{real}} \in \{5, 6\}$. Os regimes intermediários ($n_{\text{real}} \in \{2, 3\}$) são ruidosos, com desvios de até ± 9 pontos percentuais, mas uma diferença em relação ao KIMORE é que o *baseline* pré-treinado não é uniformemente nocivo no UI-PRMD (coluna **FT-BL**); isso é consistente com o fato dos sujeitos serem saudáveis e os movimentos mais próximos do caso genérico do HumanML3D, ao invés dos exercícios patológicos do KIMORE. Os dois conjuntos sintéticos do UI-PRMD e o *checkpoint* com ajuste fino correspondente também foram disponibilizados publicamente³.

5 DISCUSSÃO

Os resultados da Seção 4 sustentam quatro conclusões interligadas. Primeiro, o ajuste fino do MotionGPT3 sobre o KIMORE produz dados sintéticos úteis para *data augmentation* em escassez extrema de dados reais, sobretudo sob arquiteturas espaço-temporais mais expressivas. Segundo, o *baseline* pré-treinado (sem ajuste fino) não é fonte confiável para *augmentation*, pois gera movimentos genéricos que prejudicam a discriminação fina em ambos os modelos (CNN 1D e ST-GCN) utilizados para avaliação de utilidade *downstream*.

A terceira conclusão é que a avaliação discriminativa é mais informativa que as métricas distribucionais agregadas para decidir qual gerador adotar: a diferença de FID (-11%, de 8.83 para 7.87) seria insuficiente para justificar o ajuste fino, enquanto a diferença de concordância *prompt*-classificador (+57 pontos percentuais) correlaciona-se diretamente com o ganho *downstream*. E finalmente, quarto, a melhoria satura rapidamente com o tamanho do conjunto de dados sintéticos: ampliar de 1.000 para 5.000 amostras leva a um ganho marginal de acurácia na arquitetura ST-GCN, e estratégias de filtragem por confiança ou diversidade não superam a configuração não-filtrada (Seção 4.5), indicando saturação prática em aproximadamente 1.000 amostras, neste regime de exercícios.

Como observação adicional, percebe-se que a aplicação de uma avaliação discriminativa, no caso, treinar uma CNN 1D leve em dados reais e medir a concordância entre *prompt* e classificador, custa aproximadamente 30 segundos de GPU comum e é recomendada como métrica complementar rotineira. O controle semântico via texto permite especificar modificadores nos *prompts* (e.g., “*with limited range of motion*”) e gerar padrões raros sem coleta adicional, algo que é particularmente relevante quando a distribuição é desbalanceada.

6 CONCLUSÃO

Este artigo apresentou o RehabMotion, um *pipeline* completo para gerar e validar dados sintéticos de movimento de reabilitação via ajuste fino multitarefa do MotionGPT3, integrando a conversão do formato de dados do KIMORE para HumanML3D, a anotação textual dos movimentos, a geração *text-to-motion* (T2M) e *motion-to-text* (M2T), a avaliação dos resultados multinível com métricas biomecânicas, distribucionais e discriminativas, e a validação de utilidade *downstream* em regime de escassez extrema de dados reais.

Na prática, verificou-se que modelos de linguagem de movimento de larga escala podem ser especializados para domínios clínicos com poucos dados rotulados e a um custo acessível (aproximadamente duas horas de GPU para execução completa), sem a necessidade de se treinar geradores do zero. A avaliação discriminativa, de custo desprezível, emerge como componente recomendado em *pipelines* de geração sintética de movimento, complementando métricas tradicionais como o FID.

O código completo implementado, bem como todos os *datasets* discutidos neste artigo, estão disponíveis publicamente. Dessa forma, acredita-se que este trabalho abre um caminho promissor para minimizar o gargalo de dados em sistemas inteligentes de monitoramento de reabilitação, com impacto potencial na acessibilidade e qualidade da tele saúde.

³<https://zenodo.org/records/20671628>

7 CUIDADOS ÉTICOS

Não houve pesquisa envolvendo seres humanos conduzida pelos autores; sem recrutamento de participantes, intervenção, entrevista, questionário ou coleta primária de dados, atendendo à Resolução CNS nº 510/2016. Foram utilizados apenas os *datasets* públicos KIMORE [3] e UI-PRMD [24], cuja coleta original foi conduzida sob os protocolos éticos das respectivas instituições de origem. Ambos contêm somente coordenadas articulares anonimizadas, sem imagens de face, áudio ou identificadores diretos, com sujeitos referenciados por índices numéricos atribuídos na origem. Em conformidade com a Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018), não houve tentativa de reidentificação nem cruzamento com fontes externas, e os artefatos disponibilizados publicamente contêm apenas sequências sintéticas, que não correspondem a nenhum indivíduo real.

AGRADECIMENTOS

Os autores agradecem o financiamento parcial da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001; do Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq) – Códigos de Financiamento 442344/2023-6, 444602/2024-0 e 407466/2025-8; da Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES) – projetos nº 2026-7597P e nº 2025-1H3F; e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Os autores declaram o uso de ferramentas de Inteligência Artificial Generativa como apoio à redação, restrito à revisão gramatical, ao aprimoramento de estilo e à tradução de trechos do texto. A concepção do trabalho, os experimentos e a análise dos resultados não envolveram o emprego de tais ferramentas, e todo o conteúdo foi revisado e validado integralmente pelos autores.

REFERÊNCIAS

- [1] Ali Abedi, Mobin Malmirian, and Shehroz S Khan. 2023. Cross-modal video to body-joints augmentation for rehabilitation exercise quality assessment. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Torino, Italy, 320–327.
- [2] Luiz Felipe de Camargo, Gabriel Rissatti Miguel, Juliana da Costa Feitosa, Luis Guilherme Silva Rodrigues, Diego Roberto Colombo Dias, and José Remo Ferreira Brega. 2025. Combining computer vision and artificial intelligence to support the rehabilitation of post-stroke patients. In *International Conference on Computational Science and Its Applications*. Springer, 415–432.
- [3] Marianna Capecci, Maria Gabriella Ceravolo, Francesco Ferracuti, Sabrina Iarlori, Andrea Monteriu, Luca Romeo, and Federica Verdini. 2019. The KIMORE dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27, 7 (2019), 1436–1448.
- [4] Xin Chen, Biao Jiang, Wen Liu, Zilong Huang, Bin Fu, Tao Chen, and Gang Yu. 2023. Executing your commands via motion diffusion in latent space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Vancouver, Canada, 18000–18010.
- [5] Swakshar Deb, Md Fokhrul Islam, Shafin Rahman, and Sejuti Rahman. 2022. Graph convolutional networks for assessment of physical rehabilitation exercises. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30 (2022), 410–419.
- [6] Gisela Miranda Difini, Marcio Garcia Martins, and Jorge Luis Victória Barbosa. 2021. Human Pose Estimation for Training Assistance: a Systematic Literature Review. In *Proceedings of the Brazilian Symposium on Multimedia and the Web* (Belo Horizonte, Minas Gerais, Brazil) (WebMedia '21). Association for Computing Machinery, New York, NY, USA, 189–196. doi:10.1145/3470482.3479633
- [7] Aline PA Ferreira, Lucas O Alves, Fabio AF Molinares, Marcelo P Guimarães, Vinicius FS Mota, Leonardo CD Rocha, and Diego RC Dias. 2026. A Multidimensional Quality Assessment of Synthetic Movement Data for Post-stroke Rehabilitation. In *International Conference on Computational Science and Its Applications*. Springer, 233–250.
- [8] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. 2018. Synthetic data augmentation using GAN for improved liver lesion classification. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, Washington, USA, 289–293.
- [9] Jader Gonçalves, Wyllgner Amorim, Nicolas Sales, and Lucas Cunha. 2025. Framework Procedural para Geração de Dados Sintéticos Baseados em Simulação Estatística: Uma Abordagem para Análise de Comportamento Humano. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web* (Rio de Janeiro/RJ). SBC, Porto Alegre, RS, Brasil, 275–283. doi:10.5753/webmedia.2025.16041
- [10] Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. 2022. Generating diverse and natural 3D human motions from text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New Orleans, USA, 5152–5161.
- [11] Biao Jiang, Xin Chen, Wen Liu, Jingyi Yu, Gang Yu, and Tao Chen. 2023. MotionGPT: Human motion as a foreign language. *Advances in Neural Information Processing Systems* 36 (2023), 20067–20079.
- [12] Victor YK Kwon, Luis GS Rodrigues, Bartolomeu Zamprogno, Marcelo P Guimarães, Leonardo CD Rocha, Eduardo Zambon, Vinicius FS Mota, and Diego RC Dias. 2026. Semantic Noise Filtering in Post-Stroke Kinematic Data: An Automated Deep Learning Approach. In *International Conference on Computational Science and Its Applications*. Springer, 601–617.
- [13] Jianwei Li, Jun Xue, Rui Cao, Xiaoxia Du, Siyu Mo, Kehao Ran, and Zeyan Zhang. 2024. FineRehab: A multi-modality and multi-task dataset for rehabilitation analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Seattle, USA, 3184–3193.
- [14] Ciro Mennella, Umberto Maniscalco, Giuseppe De Pietro, and Massimo Esposito. 2023. Generating a novel synthetic dataset for rehabilitation exercises using pose-guided conditioned diffusion models: A quantitative and qualitative evaluation. *Computers in Biology and Medicine* 167 (2023), 107665.
- [15] Alina Miron, Nouredin Sadawi, Waidah Ismail, Hafez Hussain, and Crina Grosan. 2021. IntelliRehabDS (IRDS) – A dataset of physical rehabilitation movements. *Data* 6, 5 (2021), 46.
- [16] Cynthia C Norkin and D Joyce White. 2016. *Measurement of Joint Motion: A Guide to Goniometry* (5th ed.). F.A. Davis Company, Philadelphia, USA.
- [17] Mathis Petrovich, Michael J Black, and Gül Varol. 2021. Action-conditioned 3D human motion synthesis with transformer vae. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Nashville, USA, 10985–10995.
- [18] Elvis Ribeiro, Alexandre Brandão, Marcelo Guimarães, Leonardo Rocha, José Remo Brega, Rodolfo Villaça, and Diego Dias. 2025. A body tracking framework for neuromotor rehabilitation with multimedia application support. In *International Conference on Computational Science and Its Applications*. Springer, 345–360.
- [19] Elvis Ribeiro, Alexandre Brandão, Marcelo Guimarães, Leonardo Rocha, José Brega, and Diego Dias. 2024. Um framework de rastreamento corporal para reabilitação neuromotora com suporte a aplicativos multimídia. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web* (Juiz de Fora/MG). SBC, Porto Alegre, RS, Brasil, 292–300. doi:10.5753/webmedia.2024.241537
- [20] Luis. G. S. Rodrigues, Diego R. C. Dias, Marcelo P. Guimarães, Alexandre F. Brandão, Leonardo C. D. Rocha, Rogério L. Iope, and José R. F. Brega. 2021. Upper Limb Motion Tracking and Classification: A Smartphone Approach. In *Proceedings of the Brazilian Symposium on Multimedia and the Web* (Belo Horizonte, Minas Gerais, Brazil) (WebMedia '21). Association for Computing Machinery, New York, NY, USA, 61–64. doi:10.1145/3470482.3479618
- [21] Omar Sherif and Ali Hamdi. 2025. Error-Guided Pose Augmentation: Enhancing Rehabilitation Exercise Assessment through Targeted Data Generation. In *Intelligent Methods, Systems, and Applications (IMSA)*. IEEE, Giza, Egypt, 278–283.
- [22] Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H. Bermano. 2022. Human Motion Diffusion Model. *ArXiv abs/2209.14916* (2022), 12 pages.
- [23] Tiago Trotta, Luis Rodrigues, José Remo Brega, Marcelo Guimarães, Leonardo Rocha, Alexandre Brandão, and Diego Dias. 2024. ReBase: a framework for acquiring and managing body tracking data. In *Proceedings of the 26th Symposium on Virtual and Augmented Reality*. 152–160.
- [24] Aleksandar Vakanski, Hyung-pil Jun, David Paul, and Russell Baker. 2018. A data set of human body movements for physical rehabilitation exercises. *Data* 3, 1 (2018), 2.
- [25] Felipe Reis Valente, Marcelo de Paiva Guimarães, Elder José Reioli Cirilo, and Diego Roberto Colombo Dias. 2022. A multi-agent body tracking application framework applied to physical and neurofunctional rehabilitation. In *International Conference on Computational Science and Its Applications*. Springer, 459–472.
- [26] Sijie Yan, Yuanjun Xiong, and Dahua Lin. 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, New Orleans, Louisiana, USA, Article 912, 9 pages.
- [27] Jianrong Zhang, Yangsong Zhang, Xiaodong Cun, Yong Zhang, Hongwei Zhao, Hongtao Lu, Xi Shen, and Ying Shan. 2023. Generating human motion from textual descriptions with discrete representations. In *Proceedings of the IEEE/CVF*

- Conference on Computer Vision and Pattern Recognition*. IEEE, Vancouver, Canada, 14730–14740.
- [28] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. 2019. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Long Beach, USA, 5745–5753.
- [29] Bingfan Zhu, Biao Jiang, Sunyi Wang, Shixiang Tang, Tao Chen, Linjie Luo, Youyi Zheng, and Xin Chen. 2025. MotionGPT3: Human motion as a second modality. *ArXiv abs/2506.24086* (2025), 26 pages.