

Predição de Congelamento de Marcha por Meio da Detecção de Eventos de Pré-Congelamento

João Pedro A. Amoedo¹, Paulo H N Gonçalves¹,
Rafael Albuquerque¹, Eulanda M. dos Santos¹, Rafael Giusti¹, Renato C F Junior²

¹Instituto de Computação – Universidade Federal do Amazonas (UFAM)

²Faculdade de Educação Física e Fisioterapia – UFAM

Av. Gen. Rodrigo Octávio 6200, Coroado I – 69080-900 – Manaus – AM – Brasil

{pedro.abensur,paulo,rafael.albuquerque,emsantos}@icomp.ufam.edu.br

rgiusti@icomp.ufam.edu.br,renatocfjunior@ufam.edu.br

Abstract. *Freezing of gait (FOG) is a debilitating symptom commonly observed in patients with Parkinson's disease. Using data usually captured by inertial sensors, the automatic prediction of FOG may help to improve patients' quality of life—e.g., by preventing falls. Among the strategies applied to predict FOG, machine learning methods are used to solve a classification problem involving three classes: FOG, Non-FOG and Pre-FOG. The Pre-FOG class refers to time series segments immediately preceding the occurrence of a FOG event. It is expected that Pre-FOG instances present unique features that allow identifying signs of an impending FOG episode. However, the Pre-FOG class is neither well defined nor well characterized in the literature due to its transitional nature. In this paper we analyze the behavior of the Pre-FOG class using two machine learning models, SVM and the combination of CNN with LSTM, on data collected from accelerometers and gyroscopes. The results show that the Pre-FOG class can be detected, although with low recognition rates, which indicates the lack of well-defined patterns to discriminate this class from the two remaining classes.*

Resumo. *Congelamento de marcha (FOG) é um sinal clínico debilitante comumente observado em pacientes com doença de Parkinson. Utilizando dados normalmente capturados por sensores inerciais, a predição automática do FOG pode ajudar a melhorar a qualidade de vida dos pacientes—e.g., evitando quedas. Dentre as estratégias aplicadas, destaca-se o uso de métodos de aprendizagem de máquina para resolver um problema de classificação em três classes: FOG, Não-FOG e Pré-FOG. A classe Pré-FOG refere-se a segmentos de série temporal imediatamente anterior à ocorrência de um evento de FOG. A expectativa é que instâncias de Pré-FOG apresentem características únicas que permitam identificar aspectos de um episódio iminente de FOG. No entanto, essa classe não é bem definida e nem bem caracterizada na literatura devido à sua natureza transicional. Neste artigo nós analisamos o comportamento da classe Pré-FOG usando dois modelos de aprendizagem de máquina, SVM e a combinação de CNN com LSTM, em dados coletados por acelerômetros e giroscópios. Os resultados mostram que Pré-FOG pode ser detectada, mas com taxas de reconhecimento baixas, indicando a falta de padrões discriminantes da classe Pré-FOG em relação às instâncias de FOG ou de Não-FOG.*

1. Introdução

O congelamento da marcha (FOG—*Freezing of Gait*) é um sinal clínico comum e debilitante que afeta indivíduos com a doença de Parkinson (DP). É caracterizado por episódios repentinos e breves de incapacidade de produzir passos efetivos para a frente, os quais ocorrem normalmente durante o início da marcha ou quando o indivíduo tenta girar durante a caminhada. Esses bloqueios de marcha são um dos principais fatores de risco para quedas de pacientes com DP, pois durante a caminhada o tronco continua em movimento enquanto os pés ficam presos [Mazilu et al. 2013].

A automatização da detecção e da predição do FOG pode melhorar significativamente a qualidade de vida das pessoas acometidas por DP. Nos últimos anos, avanços na área de aprendizado de máquina (AM) têm possibilitado a obtenção de resultados promissores nessas duas tarefas, as quais diferenciam-se da seguinte maneira. Detecção de FOG refere-se ao processo de identificação em tempo real do FOG. Tendo-se como base os dados utilizados para treinamento do modelo de AM, este classificará novos dados como sendo casos de FOG ou de não FOG. Já a abordagem de predição de FOG busca identificar um evento de FOG que ainda ocorrerá. Portanto, considerando o tempo de reação do paciente em uma situação de evento de FOG identificado em tempo real, a abordagem de predição é mais interessante por possibilitar o aumento da janela de tempo de reação do paciente a fim de evitar que este sofra ferimentos devido à quedas. Nesse caso, a expectativa é a predição preventiva do FOG.

Os dados utilizados nessas tarefas são majoritariamente coletados por sensores de diferentes tipos e normalmente vestíveis, tais como: palmilhas de solas plantar, acelerômetros e giroscópios, condutância de pele, etc. A literatura na área de detecção de FOG é composta por diversos estudos que utilizam predominantemente modelos de AM clássicos/rasos para aprendizado supervisionado e atributos facilmente interpretáveis e extraídos manualmente [Mazilu et al. 2013, Naghavi et al. 2019]. Entretanto, há desvantagens claras no uso desses métodos, como o elevado tempo de treinamento dos modelos, dada a grande quantidade de dados coletados pelos sensores, e a dificuldade em gerar manualmente os atributos.

Em termos de predição, as principais abordagens propostas envolvem o uso de modelos de AM profundo, os quais analisam os dados de séries temporais para prever o FOG e realizam extração automática de atributos [Sigcha et al. 2020, Shalin et al. 2021]. Nesse contexto, uma forma recente de abordar o problema busca prever FOG por meio da detecção de uma terceira classe, Pré-FOG. Essa classe é descrita como uma janela de tempo observada antes do FOG acontecer. Os estudos de predição de FOG variam na abordagem empregada e no desempenho apresentado, sendo a maioria um tanto preliminar e com foco menos voltado para o desempenho e mais na compreensão das complexidades da previsão de FOG [Pardoel et al. 2019]. Por exemplo, não há um padrão quanto ao tamanho da janela temporal de uma instância Pré-FOG. Em [Shalin et al. 2021], é utilizada uma janela de tempo de 2 segundos, enquanto em [Mazilu et al. 2013] esse tamanho é de 3 segundos. Em tais estudos, é possível verificar que o uso de janelas muito curtas dificulta a utilização de modelos de classificação três classes e que, devido à natureza de transição da classe Pré-FOG, torna-se difícil distingui-la das classes Não-FOG e FOG. Já com a utilização de janelas maiores, a classificação correta da classe Pré-FOG aumenta, entretanto reduz consideravelmente as classificações de FOG e caminhada típica.

Assim como os modelos de AM rasos, o uso de modelos profundos para predição também apresenta problemas, como o fato de demandarem elevado poder computacional, dificultando sua incorporação em sistemas vestíveis de hardware limitado, sendo necessárias técnicas de redução de modelos e de dados. Além disso, demandam maior quantidade de dados de treinamento para obterem resultados mais precisos. Tal problema é particularmente importante em aplicações de predição por meio da classificação de Pré-FOG, pois normalmente há pouca quantidade de dados rotulados dessa classe.

Considerando o contexto descrito, o foco deste trabalho é a predição de FOG, mais precisamente, o objetivo é estudar o comportamento da classe Pré-FOG. Para tanto, duas técnicas de AM são empregadas: Máquinas de Vetores de Suporte (SVM—*Support Vector Machines*) que é um modelo clássico de AM; e um modelo de AM profundo composto por uma Rede de Convolução (CNN—*Convolutional Neural Network*) e uma Rede Neural Recorrente do tipo Memória de Curta Duração de Longo Prazo (LSTM—*Long Short Term Memory*). Um processo de extração manual de características foi empregado para gerar atributos a serem utilizados pelo SVM. Os atributos extraídos são amplamente utilizados na literatura, tais como desvio padrão, entropia da amostra etc. Já no caso do segundo modelo, os atributos são extraídos de forma automatizada pela CNN e em seguida enviados para as camadas LSTM para processamento das informações temporais. Neste trabalho foi feita a otimização dos hiperparâmetros de ambos os modelos. Além disso, foi empregada uma estratégia de aumento de dados para melhor lidar com a classe Pré-FOG. Os dados utilizados nos experimentos foram coletados por sensores vestíveis, precisamente acelerômetros e giroscópios. Os dados dos sensores foram combinados para aumentar a gama de informações fornecidas aos modelos de AM.

O artigo está dividido da seguinte forma. A Seção 2 discute alguns trabalhos relacionados. Já a Seção 3 especifica a metodologia seguida na pesquisa, enquanto a Seção 4 detalha e discute os resultados obtidos nos experimentos. Por fim, a Seção 5 conclui o artigo e detalha possíveis trabalhos futuros a partir dos resultados obtidos.

2. Trabalhos Relacionados

Recentemente, várias abordagens de AM foram propostas para detectar e prever a ocorrência do FOG em pacientes com DP. Os primeiros estudos nesse campo de aplicação utilizam algoritmos de AM clássicos, como Árvores de Decisão, Florestas Aleatórias [Mazilu et al. 2013] e SVM [Polat 2019]. O objetivo desses trabalhos é a detecção em tempo real utilizando sensores vestíveis, como acelerômetros e giroscópios, sendo realizadas extrações manuais de atributos dos dados de cada sensor. Esses algoritmos obtiveram resultados promissores, com especial destaque para SVM. Porém, em cada estudo diferentes processos de extração de atributos foram conduzidos, dado que algoritmos clássicos são normalmente treinados com atributos extraídos manualmente. Algumas das técnicas mais comumente usadas para extração de atributos incluem análise no domínio do tempo, no domínio da frequência e na análise tempo-frequência [Mazilu et al. 2013, Naghavi et al. 2019].

No domínio do tempo, várias medidas estatísticas, como média, desvio padrão e alcance são usadas para caracterizar os dados de cada sensor [Mazilu et al. 2015]. Além disso, alguns estudos usam medidas de entropia [Zhang et al. 2021]. Já no domínio da frequência, técnicas como a transformada rápida de Fourier (FFT—*Fast Fourier Trans-*

form) e análise de densidade espectral de potência são empregadas para converter os sinais no domínio do tempo para o domínio da frequência [Rezvanian and Lockhart 2016, Tahafchi et al. 2017]. Por fim, em termos de análise no domínio da frequência do tempo, tais técnicas são usadas para extrair componentes de frequência e energia de frequência de tempo, como a amplitude máxima de aceleração [Zhao et al. 2012].

Apesar de apresentar limitação para capturar a dinâmica temporal dos dados obtidos pelos sensores, dificultando a predição de FOG, algoritmos como SVM têm a vantagem de requererem menos processamento em dispositivos de memória limitada. Sendo um classificador que consegue lidar bem com os diversos atributos extraídos por cada sensor, a utilização de SVM para detectar FOG apresenta resultados significativos.

Em termos de predição de FOG, um dos primeiros métodos publicados na literatura foi apresentado em [Mazilu et al. 2013] e, posteriormente aprofundado e ampliado em [Mazilu et al. 2015]. Nesse estudo foi utilizada uma classe adicional: Pré-FOG, considerada como um período de tempo imediatamente antes da ocorrência de um evento de FOG. Atualmente esse período de tempo varia de 1 a 5 segundos na literatura. Posteriormente, muitos trabalhos passaram a utilizar a detecção da classe Pré-FOG para prever FOG, como [Pardoel et al. 2021], [Pardoel et al. 2022], [Shalin et al. 2021].

Os principais estudos de predição de FOG utilizam algoritmos de AM profundo para avaliar dados de séries temporais, como: CNN [Filtjens et al. 2021], Redes Neurais Recorrentes (RNN—*Recurrent Neural Network*) [Hu et al. 2019] e, mais recentemente, uma variante específica de RNN, a LSTM. Um modelo que tem produzido resultados promissores em trabalhos recentes é a combinação de CNN e LSTM (CNN-LSTM) [Sigcha et al. 2020, Li et al. 2020, Zhang and Gu 2019]. Essa combinação provê um modelo capaz de perceber nuances e características de séries temporais, pois captura dependências temporais em atributos extraídos por operações convolucionais. Sua principal vantagem é a extração automática de atributos feita pelas camadas de convolução, posteriormente tratados pelas camadas LSTM, que possuem alta eficácia para classificar, processar e prever séries temporais com intervalos de tempo de duração desconhecida.

Entretanto, apesar de sua eficiência, a CNN-LSTM é um modelo que requer maior poder computacional em relação ao SVM. Além disso, pode apresentar maior dificuldade em aprender dados muito desbalanceados, com classes minoritárias muito reduzidas, como a classe Pré-FOG. Neste trabalho é feito um estudo comparativo entre CNN-LSTM e SVM para detecção de Pré-FOG utilizando técnicas de redução e de aumento de dados no intuito de aumentar a taxa de detecção das instâncias da classe Pré-FOG.

3. Metodologia

As etapas realizadas neste trabalho para executar o estudo comparativo estão descritas nesta seção. Inicialmente nós detalhamos a base de dados empregada. Em seguida, as etapas de pré-processamento são descritas. Por fim, os modelos de AM são apresentados.

3.1. Base de Dados

A base de dados utilizada foi selecionada a partir da coleta feita por [Ribeiro De Souza et al. 2022] utilizando sensores de aceleração, giroscópios e câmeras. A coleta foi feita enquanto os pacientes realizavam uma tarefa de giro em uma posição determinada. Neste trabalho, apenas os dados de aceleração e dos giroscópios

são utilizados: 3 acelerômetros e 3 giroscópios do tipo médio-lateral, anteroposterior e vertical. Os dados apresentam o momento da coleta, os valores coletados pelos sensores, e a classe—0 para Não-FOG e 1 para FOG. Os dados coletados foram rotulados por profissionais de saúde especializados em identificar FOG em pacientes de PD. A base possui os dados filtrados por meio do filtro Butterworth passa-baixa de quarta-ordem fase zero, que atenua as frequências mais altas do sinal de entrada de forma gradual. Esses dados fornecem informações sobre os padrões de movimento dos pacientes durante a realização da tarefa específica. Na base há um total de 35 pacientes, cada paciente realizou de 1 a 3 sessões, variando conforme o grau de FOG detectado. Cada sessão possui uma duração de 2 minutos e foi catalogada e organizada. Dentre os 35 pacientes, 9 não apresentaram eventos de FOG.

Os dados de todos os pacientes foram utilizados nos experimentos, mesmo dos 9 pacientes que não apresentaram nenhum evento de FOG. Entretanto, tal fato foi levado em consideração durante o particionamento da base de dados, para melhorar a generalização do modelo. Os dados desse 9 pacientes foram distribuídos igualmente entre os conjuntos de treino, teste e validação. Dada a frequência de 128hz de cada sensor, um segundo equivale a 128 linhas de dados coletados. Cada linha possui um rótulo que indica se o paciente estava em FOG naquele determinado instante. Neste trabalho, os dados são submetidos a uma etapa de pré-processamento, a qual é descrita na próxima subseção.

3.2. Pré-Processamento

Duas tarefas foram realizadas nesta etapa: os dados foram divididos em segmentos e as instâncias da classe Pré-FOG foram rotuladas.

3.2.1. Segmentação dos Dados

Os dados brutos coletados pelos seis sensores disponíveis na base foram divididos em janelas (ou segmentos) para permitir uma análise do comportamento temporal dos dados. Em trabalhos recentes, o intervalo de janelamento varia de 1 a 4 segundos, assim como a sobreposição, que varia de 0,2 a 1,5 segundos, quando utilizada. Por exemplo, em [Reches et al. 2020] foram usadas janelas de três segundos e sobreposição de 1,5. Neste trabalho, os dados foram segmentados em janelas de 1 segundo, com sobreposição de 0,20 segundos entre cada janela, valores utilizados em [Pardoel et al. 2022] e [Shalin et al. 2021]. A escolha do tamanho dos segmentos foi decidida em função da natureza dos dados, pois, dessa forma, a quantidade de segmentos (instâncias) gerados foi maior e possibilitou maior detalhamento, o que é importante devido à quantidade de dados disponibilizados por sessão e por paciente. O processo de segmentação produziu o número de 42090 instâncias, com valores coletados por cada sensor, com duração de 1 segundo cada e com sobreposição, organizadas em ordem temporal.

3.2.2. Rotulagem da Classe Pré-FOG

Conforme descrito anteriormente, a base de dados empregada nos nossos experimentos possui instâncias rotuladas em apenas duas classes: FOG e Não FOG. Torna-se, portanto, necessário obter a classe Pré-FOG para podermos nos concentrar na detecção desta classe.

Como não há consenso na literatura quanto à extensão de um evento de Pré-FOG, nós assumimos neste trabalho o intervalo de tempo de 2 segundos imediatamente antes de cada evento de FOG como extensão do Pré-FOG. Essa escolha deve-se a trabalhos relacionados recentes que relatam bons resultados ao adotar esse valor, como [Pardoel et al. 2021] e [Shalin et al. 2021]. Portanto, todo segmento originalmente rotulado como Não-FOG que ocorra 2 segundos antes de cada evento de FOG foi rotulado como Pré-FOG. A Tabela 1 mostra como as instâncias estão distribuídas entre as duas classes originais e também a distribuição das instâncias em 3 classes, resultante do processo de rotulagem da classe Pré-FOG. Uma vez segmentadas as instâncias, estas são enviadas como entrada para os modelos de AM. No caso de SVM, as instâncias são submetidas à etapa de extração manual de atributos, descrita na próxima seção.

Tabela 1. Quantidade de instâncias das classes da base de dados.

Classe	Rotulagem Original	Rotulagem Modificada
Não-FOG	33897	32302
Pré-FOG	0	1595
FOG	8193	8193

3.3. Extração de Atributos para Uso com SVM

Levando em consideração o levantamento dos principais atributos empregados na literatura para representar os dados em problemas de detecção de FOG, os atributos utilizados neste trabalho são: Desvio Padrão, Entropia de Amostra e Potência Total, cujas fórmulas estão sumarizadas na Tabela 2.

Tabela 2. Lista de atributos, e suas respectivas fórmulas, utilizados para representar os dados para experimentos com SVM.

Atributo	Fórmula
Desvio Padrão	$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
Entropia de Amostra	$-\ln \frac{C_{m+1}(r)}{C_m(r)}$
Potência Total	$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

O desvio padrão é uma medida da dispersão ou variabilidade de um conjunto de dados. O resultado indica o quanto os dados se desviam do valor médio. Na equação, n é o número de pontos de dados no segmento, x_i é o valor do i -ésimo ponto de dados e \bar{x} é o valor médio dos pontos de dados.

A entropia de amostra é uma medida da regularidade ou previsibilidade de um conjunto de dados de séries temporais. Essa métrica quantifica a probabilidade de que os padrões nos dados de um determinado comprimento sejam seguidos por padrões semelhantes do mesmo comprimento. Na fórmula, m representa o número de pontos de dados sequenciais que são usados para definir um padrão, r representa o critério de similaridade ou valor de tolerância, que é usado para determinar se dois padrões em uma série temporal são considerados semelhantes, e $C_m(r)$ representa o número de padrões de comprimento m que são semelhantes entre si dentro de uma tolerância de r .

A potência total é uma medida da quantidade total de energia em uma série temporal. É calculada como a soma dos quadrados dos pontos de dados, normalizada pelo

número de pontos de dados. Na equação, n representa o número de pontos de dados no segmento, x_i o valor do i -ésimo ponto de dados e \bar{x} é o valor médio dos pontos de dados.

3.4. Modelo CNN-LSTM

O modelo de rede utilizado neste trabalho é o DeepConvLSTM, proposto em [Ordóñez and Roggen 2016] e que se encontra disponibilizado publicamente¹. Cada bloco da CNN contém uma camada de convolução 1D, seguida de uma camada de normalização em lote (*Batch Normalization*) e função de ativação ReLU. Após o último bloco convolucional há uma camada de Dropout de 20% para diminuir as chances de *overfitting* e melhorar a generalização. Nos experimentos realizados neste trabalho, são usados quatro blocos convolucionais. Na Tabela 3, são resumidos os hiperparâmetros da CNN empregada neste trabalho.

Tabela 3. Hiperparâmetros da CNN do modelo CNN-LSTM empregado.

Hiperparâmetro	Valor	Descrição
Dimensão do Kernel	3x3	Tamanho do filtro de convolução
Preenchimento	same	Preenchimento zero para corresponder à entrada
Função de ativação	ReLU	Função de ativação para cada camada
Taxa de Dropout	0.2	Taxa de neurônios desabilitados durante o treinamento

No caso da componente LSTM, os hiperparâmetros especificados na Tabela 4 foram empregados neste trabalho. Por fim, é possível visualizar a arquitetura completa do modelo DeepConvLSTM na Figura 1.

Tabela 4. Hiperparâmetros da componente LSTM do modelo empregado.

Hiperparâmetro	Valor
Número de Camadas LSTM	3
Total de Unidades	128, 128, 128
Taxa de Dropout	0.2
Função de Ativação	TanH
Função de Otimização	Adam
Taxa de Aprendizado	0.01
Tamanho do Lote	64
Número de Épocas	20

4. Experimentos e Resultados

Os experimentos realizados para analisar o comportamento da classe Pré-FOG foram divididos em duas séries: 1) Classificação envolvendo três classes (FOG, Pré-FOG e Não-FOG); e 2) Classificação binária (FOG e Não-FOG). Nessa segunda série de experimentos, dois cenários são considerados: *A*, no qual as instâncias de Pré-FOG são consideradas como pertencentes à classe Não-FOG, correspondente à rotulagem original da base; e *B*, em que instâncias de Pré-FOG são consideradas FOG.

¹<https://github.com/STRCWearlab/DeepConvLSTM>

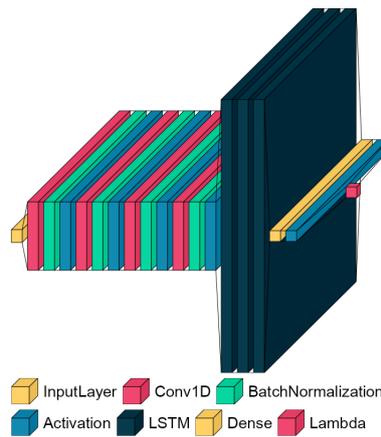


Figura 1. Arquitetura do modelo CNN-LSTM empregado nos experimentos.

4.1. Protocolo Experimental

A base de dados utilizada nos nossos experimentos é composta por dados de 35 pacientes. Os dados foram particionados por *hold-out* em subconjuntos de treino, teste e validação, contendo respectivamente 70%, 15% e 15% das instâncias. A base foi preparada para experimentação inter-paciente, de modo que os dados de cada paciente estão todos no treino, na validação ou no teste—ou seja, o modelo é treinado com dados de alguns pacientes e validado e testado em outros. Dentre os 35 pacientes, 9 não apresentam casos de FOG, representando apenas dados de caminhada normal. Esses 9 pacientes foram distribuídos em igual número nos três subconjuntos. A Tabela 5 resume as principais informações sobre o particionamento dos dados.

Tabela 5. Partições da Base Utilizadas nos Experimentos.

Partição	% de Instâncias	Número de Pacientes	Número de Instâncias
Treino	70	25	16536
Teste	15	5	4270
Validação	15	5	4270

Dada a natureza totalmente desbalanceada da base de dados, como pode ser visto na Tabela 1, foi adotada uma técnica híbrida de redução e aumento de dados para balancear os dados das três classes no conjunto de treino. Primeiramente, a quantidade de instâncias da classe majoritária (Não-FOG) foi reduzida à quantidade da classe FOG por meio do descarte de instâncias escolhidas aleatoriamente. Em seguida, a quantidade de instâncias da classe minoritária (Pré-FOG) foi aumentada por meio da técnica de geração de dados sintéticos SMOTE [Chawla et al. 2002]. Ao final, as três classes passaram a conter exatamente a mesma quantidade de instâncias. É importante mencionar que o balanceamento foi feito apenas no conjunto de treino.

O conjunto de validação foi utilizado para ajustar hiperparâmetros dos modelos. No caso do SVM, três tipos de kernel foram testados (Linear, RBF e Polinomial), sendo os hiperparâmetros de cada kernel definidos com o algoritmo *Grid Search* executado no conjunto de validação. Os melhores resultados foram obtidos pela versão com kernel RBF e os seguintes hiperparâmetros: $C = 0.1$ e $\Gamma = 0.1$. Já em relação ao modelo

CNN-LSTM, os hiperparâmetros apresentados na Tabela 3 também foram definidos com dados de validação. Além disso, foi utilizada a estratégia de parada antecipada para evitar *overfitting* e melhorar a generalização do modelo.

Os hiperparâmetros dos modelos foram definidos na primeira série de experimentos, mas foram também utilizados na segunda série por simplificação. Entretanto, houve mudança no número de instâncias de cada classe, pois na segunda série de experimentos há dois cenários testados: *A*, que considera a classe Pré-FOG como Não-FOG, mantendo a rotulagem original da base; e *B*, que considera a classe Pré-FOG como FOG. Nesse caso, a redução da quantidade de dados de Não-FOG foi mantida, mas o uso de SMOTE para aumento de dados não, pois a classe Pré-FOG deixou de ser considerada. Os resultados obtidos nas duas séries de experimentos são apresentados na próxima subseção.

4.2. Resultados

Os resultados obtidos na primeira série de experimentos estão resumidos na Tabela 6. O melhor resultado está em negrito para cada métrica. Em nossas análises, nós consideramos o F1-score, pois o conjunto de testes é desbalanceado. Observa-se uma superioridade do SVM. Porém, se considerarmos os valores de Especificidade e Sensitividade calculados para a classe Pré-FOG, mostrados na Tabela 7, a superioridade da CNN-LSTM é destacada, ainda que com taxas baixas, indicando que poucas instâncias de Pré-FOG foram detectadas. Essas duas métricas de avaliação binária calculam a proporção de classificações corretas em relação às falsamente positivas, e a proporção de classificações corretas em relação às falsamente negativas, respectivamente.

Tabela 6. Resultados dos modelos SVM e CNN-LSTM no conjunto de teste da primeira série de experimentos.

Modelo	Acurácia	Precisão	Recall	F1-Score
SVM RBF	75.0%	81.0%	75.4%	77.0%
CNN-LSTM	76.9%	82.1%	64.8%	70.2%

Tabela 7. Métricas de avaliação binária da classe Pré-FOG.

Modelo	Especificidade	Sensitividade
SVM RBF	30.4%	6.5%
CNN-LSTM	32.7%	8.7%

De fato, como pode ser visto nas matrizes de confusão exibidas na Figura 2, as instâncias de Pré-FOG foram confundidas com as duas outras classes, tanto pelo SVM quanto pela CNN-LSTM, em proporções muito semelhantes. Portanto, o padrão de Pré-FOG parece se assemelhar tanto a FOG quanto a Não-FOG. Por essa razão, na segunda série de experimentos nós comparamos os resultados quando instâncias de Pré-FOG são consideradas como FOG ou como Não-FOG.

Dois cenários são testados na segunda série: *A*, classificação binária considerando as instâncias de Pré-FOG como Não-FOG; e *B*, instâncias de Pré-FOG consideradas como FOG. Os resultados obtidos estão destacados na Tabela 8 e mostram uma superioridade da CNN-LSTM em relação ao SVM. Entretanto, se compararmos o cenário *A* com o *B*, a

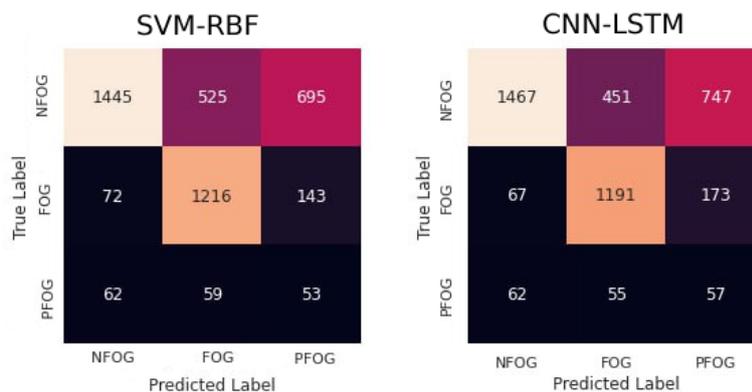


Figura 2. Matrizes de confusão geradas pelos dois modelos.

Tabela 8. Comparação dos dois cenários da segunda série de experimentos.

Modelo	Acurácia	Precisão	Revocação	F1-Score
SVM-RBF A	78.9%	91.3%	79.0%	84.7%
SVM-RBF B	74.9%	74.7%	94.0%	83.4%
CNN-LSTM A	85.1%	85.7%	89.0%	87.3%
CNN-LSTM B	82.2%	84.4%	92.1%	88.1%

diferença de F1-score é pouco significativa, fato que confirma que a classe Pré-FOG pode ser confundida igualmente entre as duas demais classes.

A partir das duas séries de experimentos é possível constatar a difícil caracterização da classe Pré-FOG. Quando esta é considerada uma terceira classe, os classificadores confundem-a tanto com Não-FOG quanto com FOG em quantidade considerável, mesmo fazendo aumento de dados para balancear as classes. No caso da segunda série, é evidente que os resultados são bem semelhantes nos dois cenários testados. Portanto, o Pré-FOG não foi bem caracterizado nem pelo processo de extração de atributos automático (CNN-LSTM) e nem pelo processo manual, utilizado com SVM. Isso pode indicar que o tamanho da janela adotado não é o ideal. Por exemplo, o tamanho pode ser muito grande ou muito pequeno para diferenciar Pré-FOG das demais classes, por ser um momento de transição cuja duração específica pode variar devido à inconstante marcha humana.

5. Conclusão

Este trabalho tem como foco a predição de FOG em pacientes com Doença de Parkinson por meio de uma análise experimental da detecção da classe Pré-FOG. Os experimentos mostram que as taxas de detecção de Pré-FOG são baixas devido a muitos fatores. O principal fator que dificulta a caracterização dessa classe é o período de tempo necessário para observá-la. Uma janela curta de Pré-FOG pode não ser suficiente, enquanto janelas mais longas podem misturar muito Pré-FOG com eventos de FOG e com Não-FOG. Assim como há eventos de FOG que variam em tempo e intensidade de acordo com cada paciente, o período da janela de Pré-FOG provavelmente deve ser avaliado levando em consideração esses fatores também. Além disso, o Pré-FOG também pode variar em função do avanço da Doença de Parkinson no paciente. Portanto, essa classe precisa ser melhor definida antes de poder ser caracterizada.

Para trabalhos futuros, pretende-se analisar com mais detalhes o Pré-FOG para que possa ser melhor definido, aumentando as taxas de detecção. Dentre as atividades pretendidas destacam-se: variar o tamanho da janela e interpretá-la de forma mais individual em relação ao paciente, pois é provavelmente dependente de paciente. Por fim, nós pretendemos criar uma base de dados própria para rotulagem manual de Pré-FOG.

6. Agradecimentos

O presente trabalho é decorrente do projeto de Pesquisa e Desenvolvimento (P&D) 001/2020, firmado entre a Fundação da Universidade do Amazonas e FAEPI, que conta com financiamento da Samsung, usando recursos da Lei de Informática para a Amazônia Ocidental (Lei Federal nº 8.387/1991), estando sua divulgação de acordo com o previsto no artigo 39.º do Decreto nº 10.521/2020.

Referências

- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- Filtjens, B., Ginis, P., Nieuwboer, A., Afzal, M. R., Spildooren, J., Vanrumste, B., and Slaets, P. (2021). Modelling and identification of characteristic kinematic features preceding freezing of gait with convolutional neural networks and layer-wise relevance propagation. *BMC medical informatics and decision making*, 21(1):1–11.
- Hu, K., Wang, Z., Wang, W., Martens, K. A. E., Wang, L., Tan, T., Lewis, S. J., and Feng, D. D. (2019). Graph sequence recurrent neural network for vision-based freezing of gait detection. *IEEE Transactions on Image Processing*, 29:1890–1901.
- Li, B., Yao, Z., Wang, J., Wang, S., Yang, X., and Sun, Y. (2020). Improved deep learning technique to detect freezing of gait in parkinson’s disease based on wearable sensors. *Electronics*, 9(11):1919.
- Mazilu, S., Calatroni, A., Gazit, E., Mirelman, A., Hausdorff, J. M., and Tröster, G. (2015). Prediction of freezing of gait in parkinson’s from physiological wearables: an exploratory study. *IEEE journal of biomedical and health informatics*, 19(6):1843–1854.
- Mazilu, S., Calatroni, A., Gazit, E., Roggen, D., Hausdorff, J. M., and Tröster, G. (2013). Feature learning for detection and prediction of freezing of gait in parkinson’s disease. In *Machine Learning and Data Mining in Pattern Recognition: 9th International Conference, MLDM 2013, New York, NY, USA, July 19-25, 2013. Proceedings 9*, pages 144–158. Springer.
- Naghavi, N., Miller, A., and Wade, E. (2019). Towards real-time prediction of freezing of gait in patients with parkinson’s disease: addressing the class imbalance problem. *Sensors*, 19(18):3898.
- Ordóñez, F. J. and Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1).
- Pardoel, S., Kofman, J., Nantel, J., and Lemaire, E. D. (2019). Wearable-sensor-based detection and prediction of freezing of gait in parkinson’s disease: a review. *Sensors*, 19(23):5141.

- Pardoel, S., Nantel, J., Kofman, J., and Lemaire, E. D. (2022). Prediction of freezing of gait in parkinson's disease using unilateral and bilateral plantar-pressure data. *Frontiers in Neurology*, 13.
- Pardoel, S., Shalin, G., Nantel, J., Lemaire, E. D., and Kofman, J. (2021). Early detection of freezing of gait during walking using inertial measurement unit and plantar pressure distribution data. *Sensors*, 21(6):2246.
- Polat, K. (2019). Freezing of gait (fog) detection using logistic regression in parkinson's disease from acceleration signals. In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, pages 1–4. Ieee.
- Reches, T., Dagan, M., Herman, T., Gazit, E., Gouskova, N. A., Giladi, N., Manor, B., and Hausdorff, J. M. (2020). Using wearable sensors and machine learning to automatically detect freezing of gait during a fog-provoking test. *Sensors*, 20(16):4474.
- Rezvanian, S. and Lockhart, T. E. (2016). Towards real-time detection of freezing of gait using wavelet transform on wireless accelerometer data. *Sensors*, 16(4):475.
- Ribeiro De Souza, C., Miao, R., Ávila De Oliveira, J., Cristina De Lima-Pardini, A., Frago De Campos, D., Silva-Batista, C., Teixeira, L., Shokur, S., Mohamed, B., and Coelho, D. B. (2022). A public data set of videos, inertial measurement unit, and clinical scales of freezing of gait in individuals with parkinson's disease during a turning-in-place task. *Frontiers in Neuroscience*, 16:832463.
- Shalin, G., Pardoel, S., Lemaire, E. D., Nantel, J., and Kofman, J. (2021). Prediction and detection of freezing of gait in parkinson's disease from plantar pressure data using long short-term memory neural-networks. *Journal of neuroengineering and rehabilitation*, 18(1):1–15.
- Sigcha, L., Costa, N., Pavón, I., Costa, S., Arezes, P., López, J. M., and De Arcas, G. (2020). Deep learning approaches for detecting freezing of gait in parkinson's disease patients through on-body acceleration sensors. *Sensors*, 20(7):1895.
- Tahafchi, P., Molina, R., Roper, J. A., Sowalsky, K., Hass, C. J., Gunduz, A., Okun, M. S., and Judy, J. W. (2017). Freezing-of-gait detection using temporal, spatial, and physiological features with a support-vector-machine classifier. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2867–2870. IEEE.
- Zhang, W., Huang, D., Li, H., Wang, L., Wei, Y., Pan, K., Ma, L., Feng, H., Pan, J., and Guo, Y. (2021). Sensoring and application of multimodal data for the detection of freezing of gait in parkinson's disease. *arXiv preprint arXiv:2110.04444*.
- Zhang, Y. and Gu, D. (2019). A deep convolutional-recurrent neural network for freezing of gait detection in patients with parkinson's disease. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6. IEEE.
- Zhao, Y., Tonn, K., Niazmand, K., Fietzek, U. M., D'Angelo, L. T., Ceballos-Baumann, A., and Lueth, T. C. (2012). Online fog identification in parkinson's disease with a time-frequency combined algorithm. In *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, pages 192–195. IEEE.