Modelo de avaliação de dados baseado em sinal eletrocardiograma para autenticação e identificação de usuários

João Marcelo Freitas de Almeida[1] **e Eduardo Coelho Cerqueira**[1]

[1]Universidade Federal do Pará (UFPA)
Instituto de Ciências Exatas e Naturais
Faculdade de computação
Belém – PA – Brasil

Abstract. *The rapid spread of wearable technologies has enabled the collection of a variety of bio-signals more easily than ever. This spread allowed wearables devices to meet the new security requirements of authentication and identification applications, being used as a supplier of input data for security systems. Existing eletrocardiogram signal works for user authentication does not consider a population size that resembles that of a real application. Finding real data with a large number of people in ECG data is a challenge. This work proposes a set data quality assessment steps over an electrocardiogram signal dataset, which improves the performance metrics of a biometric system in classification and identification tasks in a configuration closer to a real-world security system. Steps are such as data augmentation, removal of data outliers, and feature generation, which was proven to increase the performance in the PhysioNet Computing in Cardiology 2018 dataset.*

Resumo. *A rápida disseminação de tecnologias vestíveis permitiu a coleta de uma variedade de sinais biológicos com mais facilidade do que nunca. Isto permitiu a sua utilização para cumprir os novos requisitos de segurança de aplicações de autenticação e identificação, sendo utilizado como fornecedor de dados de entrada para estes sistemas de segurança. Os trabalhos existentes sobre ECG para autenticação do usuário não consideram um tamanho de população que se assemelhe ao de uma aplicação real. Encontrar conjuntos de dados de eletrocardiograma com um grande número de classes observadas é um desafio. Este trabalho propõe um conjunto de etapas de avaliação de qualidade de um conjunto de dados de sinal eletrocardiograma, que melhora as métricas de desempenho de um sistema biométrico em tarefas de classificação e identificação em uma configuração próxima de um sistema de segurança no mundo real. As etapas incluem aumento de dados, remoção de outliers e geração de novos recursos (features), que comprovaram aumentar o desempenho no conjunto de dados PhysioNet Computing in Cardiology 2018.*

1. Introdução

Nos últimos anos, a tecnologia de dispositivos vestíveis vem atingindo o uso em larga escala, em uma variedade de aplicações [Seneviratne et al. 2017]. Para os usuários, a conectividade se torna fácil e elementar. Do lado do gerenciador de aplicações em vestíveis, o acesso aos dados e extração de conhecimento destes dados tornaram-se possibilidades

acessíveis, gerenciáveis e convenientes. Conjuntamente, a preocupação com a privacidade de dados do usuário ganhou muita atenção, pois esses dispositivos são uma porta de acesso a dados sigilosos. Neste contexto, os dispositivos vestíveis permitem a coleta de dados biométricos ou de interação ambiental, tais como frequência de pulso e sinais biométricos de eletrocardiograma (ECG), eletroencefalograma (EEG), foto-pletismograma (PPG) e saturação de O2 [Hale et al. 2019, Resque et al. 2019].

Vários trabalhos tem explorado a possibilidade da utilização de sinais biométricos como recurso de entrada para autenticação e identificação. Os sistemas de autenticação podem ser baseados em algo que é pertencente de maneira natural (sinal biométrico) ao indivíduo, diferentemente do que é proposto até então, que se baseia em algo que o usuário possui, como senhas, cartões ou chaves. Uma que vez esses pertences podem ser perdidos, roubados, descobertos ou copiados [Jain et al. 2019], a possibilidade de uso dos sinais biométricos coletados pelos dispositivos vestíveis, focados nas características intrínsecas e únicas do indivíduo, sustentam ser uma alternativa factível para tarefas de autenticação e identificação [Pinto et al. 2018]. Apesar de ter algumas vantagens em comparação com senhas e medidas de segurança tradicionais, impressão digital, voz e rosto também já se mostraram vulneráveis anos atrás [O'Donnell 2019].

Sobre a diferenciação entre as tarefas de identificação e autenticação, evidencia-se que um sistema de autenticação de usuário verifica um conjunto de recursos de entrada em relação a um perfil que já está vinculado às credenciais daquele indivíduo, conhecido como sistema de correspondência *um para um*. Já para identificação, o sistema verifica os recursos coletados em relação a todos os outros modelos, isso que é descrito como um sistema de correspondência *um para muitos*. Assim, o processo de autenticação atesta se o individuo é quem afirma ser, enquanto a identificação atesta se alguém já foi previamente cadastrado em um sistema de correspondência.

Assim, os sistemas de autenticação baseados em características fisiológicas como ECG, PPG e EEG têm a vantagem de garantir identificação do usuário destacando-se por fornecer universalidade, singularidade, natureza oculta, permanente e aquisição simplificada [Barros et al. 2019], [Zhang et al. 2019a]. Assim, o ECG coletado por dispositivos vestíveis tem sido utilizado frequentemente para identificação biométrica, uma vez que possui características morfológicas, estatísticas e leves ondulações (wavelets) únicas para cada indivíduo. Diferentes algoritmos de aprendizado de máquina apresentam desempenho razoável, havendo dados suficientes, para identificação do usuário baseada em dados de ECG [Halevy et al. 2009]. Nesse sentido, para validar o ECG como sinal biométrico para autenticação do usuário, os modelos de aprendizado de máquina devem ser treinados com um número grande de classes observadas (pessoas diferentes), simulando o cenário apresentado em aplicações do mundo real.

A maioria dos trabalhos existentes sobre ECG para autenticação e identificação do usuário, [Barros et al. 2019], [Zhang et al. 2019a], [Zhang et al. 2017], [Barros et al. 2018] [Camara et al. 2018]. analisaram conjuntos de dados com registros de 20 a 290 pessoas. Essa quantidade de dados, apesar de representativa, talvez não seja suficiente para fornecer escalabilidade ao sistema de autenticação e identificação de usuários em cenários reais (milhares de usuários). Considerando, por exemplo, um sistema de controle de entrada de edifício comercial de grande porte, facilmente tem-se centenas de usuários sendo autenticados e identificados todos os dias. Considerando uma

aplicação relacionada à mobilidade em transporte público, o número de usuários pode chegar rapidamente a centenas de milhares em uma cidade grande.

Com base na avaliação preliminar realizada neste trabalho, para identificação do usuário com base no sinal ECG, o desempenho diminui assim que o número de classes alvo aumenta. Por exemplo, a métrica de *precision* começa em torno de 78% para o modelo treinado com dados de 100 classes e termina perto de 60% no modelo treinado com dados de 1500 classes. Isso ocorre porque o modelo de aprendizado de máquina precisa de mais dados e recursos de dados (*features*) por classe para poder aprender a classificar com mais exatidão esse grande número de classes.

É preciso notar também que mais dados de treino disponíveis não necessariamente resolveriam o problema anterior já que além disso se tornariam necessários também recursos de tempo e computação maiores já que o processo de treino do modelo seria mais complexo e lento. Assim sendo, uma avaliação de qualidade de dados (nesse caso, aumentar o número de exemplos somente com os dados já obtidos, remoção de dados anômalos, e geração de novas *features*) devem ser aplicadas a fim de alcançar bons resultados no cenário onde há um maior número de classes. Por último, a implicação prática seria que o desenvolvedor de aplicações de autenticação e identificação biométrica empregaria mais tempo avaliando a qualidade do conjunto de dados do que treinando e os modelos de aprendizado de máquina, uma vez que, nesse cenário, a qualidade dos dados importa tanto ou mais que o algoritmo de classificação. [Halevy et al. 2009].

Encontrar conjuntos de dados suficientemente grandes tornou-se um dos maiores desafios para o desenvolvimento de trabalhos nessa área. Não ter dados rotulados de qualidade suficiente irá gerar sobreajuste (*overfitting*) do modelo, o que tem significado prático de que o sistema é enviesado em certo nível para os dados que lhe foram passados durante o processo de treino e assim portanto, não será capaz de fazer generalizações satisfatórias em cenários de teste[Lemley et al. 2017]. Sobre esta preocupação, na família de aplicações de reconhecimento de imagem, técnicas de espelhamento, redimensionamento, corte, rotação são utilizadas como recurso de aumento de dados legítimos, pois pequenas alterações realizadas por essas técnicas no dados de treino originais não alteram o rótulo da imagem, ao mesmo tempo que fornecem ao modelo novas possibilidades de aprendizado com os mesmos dados de treino, consistindo em uma alternativa viável de diminuição de *overfitting* do modelo quando dados adicionais não podem ser obtidos. Um cuidado essencial deve ser tomado para que essas transformações não alterem a natureza dos dados e as *features* que serão obtidas a partir deles. No cenário de sinais biométricos, mais especificamente de sinal de eletrocardiograma, essas abordagens podem ser semelhantemente propostas para aumentar o tamanho do conjunto de dados de treino.

Neste trabalho, é proposto um esquema de avaliação de qualidade de dados usado para identificação do usuário com base em sinal de ECG, denominado DETECT. Cada etapa do DETECT contribui para o aumento das métricas de classificação e estas devem ser utilizadas conjuntamente para elevar a qualidade dos dados e evitar vieses. Sobre possíveis transformações que o modelo DETECT aplica aos dados de entrada, foram propostas as etapas de aumento do número de exemplos (usando uma técnica de aumento de dados própria), filtragem (que remove ruído de captação do sinal), remoção de *outliers* (para descartar os dados com valores fora do alcance interquartil que podem afetar a tarefa de classificação negativamente), e por último, a etapa de geração de novas *features*.

Para validar o modelo DETECT, escolheu-se o conjunto de dados *PhysioNet Computing in Cardiology 2018*, devido ao grande número de indivíduos diferentes que tiveram seus dados coletados (aproximadamente 2 mil indivíduos) conjuntamente com um dos algoritmos de aprendizado de máquina mais amplamente usados e conhecidos para tarefas de classificação, *Random Forest* (RF). Com base nisso, foi analisada a precisão do modelo DETECT para classificar as pessoas no cenário de autenticação e identificação contínua. Os resultados da avaliação mostram o potencial do modelo DETECT em comparação ao mesmo processo de classificação sem qualquer avaliação de qualidade de dados, aumentando a métrica *precision* em 20% para um modelo treinado com 100 classes, e 16% considerando um modelo treinado com 1500 classes.

Este trabalho está estruturado da seguinte forma: A Seção 2 apresenta os trabalhos relacionados, a seção 3 descreve uma visão geral das diferentes técnicas com ECG para identificação do usuário utilizadas no modelo DETECT. A seção 4 detalha a avaliação do modelo DETECT, incluindo a metodologia avaliativa e os resultados alcançados. Por fim, a seção 5 apresenta as considerações finais e trabalhos futuros.

2. Trabalhos relacionados

[Zhang and Wu 2016] considera uma autenticação usando sinal ECG coletado de eletrodos de dois dedos juntamente com um aplicativo móvel. Eles selecionaram 85 registros de ECG provenientes de dados públicos da plataforma *PHYSIONET* [Goldberger et al. 2000]. Para identificação, o sistema utiliza a técnica de *support vector machines* (SVM) e redes neurais. Um mecanismo de votação exige votos iguais de mais da metade dos classificadores eleitores para validar a pessoa de teste. Alcançou 97,55% de acurácia.

[Camara et al. 2018] teve como foco a autenticação contínua, onde o usuário é autenticado periodicamente, garantindo a presença contínua do usuário. Os registros são classificados usando a técnica de árvore de decisão, SVMs, entre outros. Eles alcançaram *precision* entre 97,4% e 97,9%, usando gravações de 10 indivíduos da base de dados de ritmo sinusal normal do *MIT-BIH (NSRDB)* [Goldberger et al. 2000].

[Zhang et al. 2017] introduziram uma nova abordagem chamada *multi-level wavelet convolutional neural networks* (MCNN) para identificação biométrica de ECG. Eles alcançaram uma taxa de identificação média de 93,5%. Eles usaram uma técnica de obtenção de janelas aleatórias de segmentos de ECG para aumentar a representatividade dos dados para lidar com um conjunto de dados com um pequeno número de amostras.

[Cao et al. 2020] destacaram que o sucesso do modelo de aprendizado de máquina está relacionado à disponibilidade de um conjunto de dados representativo. Eles propuseram uma nova estratégia de aumento de dados projetada especificamente para análise de ECG. Basearam-se na duplicação, concatenação e re-amostragem de segmentos de ECG. Essa estratégia pode aumentar a diversidade das amostras, bem como equilibrar o número de amostras de cada classe, o que facilita a extração das *features* do conjunto de dados pelos modelos de aprendizado profundo. Embora este trabalho não esteja relacionado à autenticação biométrica, ele demonstrou que técnicas de aumento de dados podem ser aplicadas para avaliar a qualidade de conjuntos de dados.

[Palma et al. 2019] usaram estatísticas simples para extração de características do sinal, incluindo a média, desvio padrão, mediana, valor máximo e mínimo, intervalo in-

terquartil, primeiro e terceiro quartis (Q1 e Q3), curtose e assimetria do Sinal de ECG. Alcançou-se um *precision* médio de 99,61% usando sua abordagem com o classificador de *random forests*. Eles usaram um comprimento de segmento de dados de 7s, o que garante um bom número de batimentos cardíacos por segmento, mas que na prática leva muito tempo para ser adquirido e pode incomodar o usuário.

[Pouryayevali et al. 2014] propuseram um conjunto de padrões para registro de sinal de ECG e apresentaram o banco de dados UofT ECG (UofTDB) para avaliar o desempenho de vários métodos biométricos de ECG. Eles gravaram sinais de ECG de 1020 indivíduos em diferentes posturas e condições de exercício. Os sensores Vernier ECG foram usados com uma taxa de amostragem de 200 Hz. Eles tentaram demonstrar que existem fatores que afetam a precisão biométrica ao longo do tempo. Seu banco de dados é promissor e, infelizmente, privado.

Com base na análise do estado da arte, foi concluído que a maioria dos trabalhos existentes sobre identificação e autenticação de usuários com base em ECG considera bancos de dados *PHYSIONET*, já que a maioria deles foram coletados de estudos médicos. No entanto, os conjuntos de dados *PHYSIONET* não têm um grande número de pessoas quando comparados com os grandes bancos de dados biométricos governamentais já em execução em alguns países. Dessa forma, busca-se um conjunto de dados maior para investigar como uma identificação/autenticação de usuário baseada em ECG funcionaria com um número maior de classes.

Foi observado que um ajuste fino dos parâmetros do algoritmo em uso já dá bons resultados com poucas classes alvo, mas, na prática, ao trabalhar com um conjunto de dados maior, não seria suficiente. Uma avaliação na qualidade dos dados, como aumentar o número de exemplos, remover dados anômalos e incluir *features* adicionais, se torna necessário para alcançar os resultados. Na verdade, às vezes os dados coletados não terão a qualidade ou quantidade mínima de amostras e o usuário será solicitado a coletá-los novamente. A etapa de filtragem é o único consenso quando se fala sobre tratamento de dados biométricos, [Zhang et al. 2019a], [Biel et al. 2001], [Zhang et al. 2017], [Zhang and Wu 2016], pois o ruído é comumente atrelado no processo de aquisição do sinal. A normalização do z-score foi usada em [Zhang et al. 2019a] sob conjuntos de dados variados. [Zhang et al. 2019a] também aplicou uma técnica especial de segmentação, mas os outliers não foram removidos, provavelmente afetando os resultados. Em [Zhang et al. 2017], uma divisão de dados cega foi aplicada em segmentos de sinal em oito conjuntos de dados diversos. Em seguida, todas as gravações foram uniformizadas e subtraídas da média para equilibrar sua contribuição na fase de treinamento do modelo.

Em conclusão, constata-se uma falta de técnicas bem definidas para avaliar a qualidade dos dados. A identificação com base em impressões digitais e com base em rosto está mais desenvolvida nesses aspectos. Portanto, a avaliação da qualidade dos dados padronizada é necessária para fornecer uma linha de base para usar o ECG como um sistema biométrico confiável. A Tabela 1 resume alguns dos trabalhos analisados relacionados aos sistemas biométricos de ECG.

Tabela 1. Alguns dos trabalhos relacionados a modelos para autenticação e identificação de usuários baseado em sinais biométricos.

Trabalho	Conjuntos de dados utilizados	Classificadores	Observações
[Camara et al. 2018]	IMDs	DT, SVM e NN	Features não-fiduciais
[Zhang and Wu 2016]	PTB, MITDB, NSRDB	SVM, NN, LIBSVM	Usou mecanismo de votação
[Zhang et al. 2019a]	MIT-BH	PCA + LDA	Abordagem híbrida
[Cao et al. 2020]	Computing in Cardiology Challenge 2017	RNN, LSTM	Duplicação, junção e re-amostragem de sinal ECG
[Pouryayevali et al. 2014]	UofTDB	LDA	Coletou dados em várias posições
[Palma et al. 2019]	PTB	RF	Usou métricas estátisticas para entender o sinal ECG

3. O modelo DETECT

Nesta seção, descreve-se o modelo DETECT para autenticação e identificação biométrica com sinal ECG. Aqui introduz-se um conjunto de etapas para avaliar a qualidade do conjunto de dados de ECG, resultando em melhorias de desempenho nas tarefas de classificação.

3.1. Cenário de avaliação

O ECG tem sido utilizado por muitos pesquisadores em sistemas de identificação biométrica, uma vez que é praticamente único para cada indivíduo, contando com características estatísticas, morfológicas e ondulares [Barros et al. 2019] combinadas de maneiras particular para cada individuo. Neste trabalho, o processo de identificação do usuário com sinal ECG é dividido em cinco etapas: (*i*) Aquisição do sinal de ECG bruto, (*ii*) remoção de ruído, (*iii*) segmentação (e possivelmente aumento de dados), (*iv*) extração de *features* (e possivelmente remoção de dados anômalos) e (*v*) classificação. Estes cinco passos são ilustrados na Figura 1.

Inicialmente, o sinal proveniente de dispositivos vestíveis é obtido e o ruído intrínseco a coleta do sinal é removido. Para a segmentação do sinal, é aplicado um método de repartição às cegas em segmentos de sinal com duração igual de 3s, sendo que a repartição desconsidera se está repartindo o sinal no início, meio ou fim de uma pulsação. A ideia é projetar uma solução que não gaste recursos computacionais encontrando o complexo QRS durante a segmentação dos dados, somente na etapa de extração de *features*. Sendo a etapa de aumento de dados uma alternativa para aumentar o número de segmentos possíveis de utilização. Para finalizar o preparo dos dados tem-se as etapas de extração de *features* para a obtenção das características do ECG do usuário assim

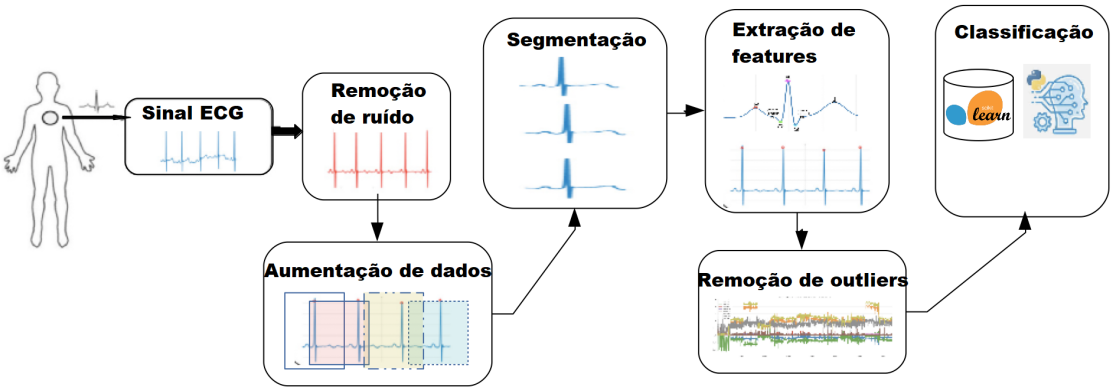

Figura 1. As etapas de identificação/autenticação com sinal ECG.

como a remoção de dados anômalos causados por ruído e outras adversidades na etapa de aquisição dos dados. As *features* resultantes desse processamento são usadas para treinar modelos inteligentes que serão as referências em tarefas de classificação realizadas para distinguir acessos genuínos e ilegítimos.

3.2. Base de dados: PhysioNet Computing in Cardiology 2018

O ECG registra a atividade elétrica gerada pelo coração de forma não invasiva, onde os batimentos cardíacos geram ondas de polarização e despolarização nas fibras musculares, assim registrando a atividade cardíaca com base nas diferenças de potencial elétrico resultantes. Um sinal de ECG típico tem seis pontos notáveis de um único batimento cardíaco saudável: P, Q, R, S, T e U, como visto na *Figura 2*. Estes pontos podem ser usados para identificação do usuário [Su et al. 2019]. O estímulo produzido durante a despolarização e repolarização gera tal comportamento [Marieb and Hoehn 2013]. As fases de despolarização correspondem à onda P (despolarização atrial) e onda QRS (despolarização dos ventrículos). As fases de repolarização correspondem à onda T e onda U (repolarização ventricular) [Biel et al. 2001].

Em trabalhos com sinal ECG, às vezes o maior desafio não é desenvolver um modelo ou escolher o classificador; o principal problema é encontrar um conjunto de dados próximo das situações do mundo real. Existem muitos conjuntos de dados disponíveis e eles diferem no número de indivíduos, estado de saúde dos indivíduos, tipo de sensor de aquisição de sinal utilizado, localização da coleta, e assim por diante. Ao analisar esses bancos de dados de ECG, escolheu-se o banco de dados PhysioNet Computing in Cardiology 2018 [Goldberger et al. 2000], esse banco de dados possui uma captura de uma variedade de sinais fisiológicos registrados durante o sono do usuário durante a noite, incluindo EEG, eletro-oculografia (EOG), eletromiografia (EMG), ECG e saturação de oxigênio (SaO2).

O conjunto de dados inclui 1985 indivíduos únicos, monitorados em um laboratório para o diagnóstico de distúrbios do sono. Desta forma, o conjunto de dados PhysioNet oferece a oportunidade de avaliar como o sistema biométrico baseado em ECG se comporta quando o número de pessoas aumenta. Até onde o se sabe, esta é a primeira vez que o banco de dados PhysioNet Computing in Cardiology 2018 é usado em pesquisas de autenticação ou identificação.

A maioria dos conjuntos de dados disponíveis tem algumas restrições que limitam sua aplicação em simulações práticas, como a persistência das características do ECG de um indivíduo ao longo do tempo. Sobre isso, a forma das ondas de ECG é determinada principalmente por características anatômicas humanas, cuja variabilidade natural ocorre ao longo dos anos. Em um curto espaço de tempo, os ciclos de ECG apresentam pequenas variações qualitativas ou quantitativas, provavelmente causadas por variações no procedimento de aquisição e distorções do filtro. As variações dentro de uma hora são quase idênticas às variações dentro de seis meses [Lugovaya 2005], por exemplo. Certamente existem muitas outras causas naturais e artificiais, intencionais e não intencionais da variabilidade do ECG. Algumas delas estão relacionadas à idade e outras relacionadas a medicamentos que podem alterar temporariamente a configuração do ciclo cardíaco.

Diante disso, é provável que as atualizações periódicas dos registros do conjunto de treinamento e o retreinamento do classificador sejam necessários como componentes de um sistema de identificação. Todos esses conjuntos de dados são muito úteis para avaliar uma operação inicial do sistema biométrico, mas ele terá que ser reconstruído várias vezes durante seu ciclo de vida numa aplicação prática. Assim como um sistema de identificação de ECG enfrentará uma variedade de desafios semelhantes aos apresentados por vários ataques já praticados a outros tipos de sistemas biométricos [Lugovaya 2005].

3.3. Remoção de ruído do sinal ECG

Os sinais de ECG são vulneráveis a diferentes fontes de ruído, como interferência eletromagnética e movimentos repentinos do usuário em regiões próximas ao sensor [Bastos et al. 2019]. Semelhante a outros trabalhos como [Zhang et al. 2019b] e [Pouryayevali et al. 2014], os sinais de ECG brutos foram filtrados usando um filtro *Butterworth* de quarta ordem com frequências de corte de 1 Hz (limite inferior) e 40 Hz (limite superior). Visto que abaixo de 0.5 Hz, o sinal é corrompido pelo desvio da linha de base, e acima de 40 Hz há distorção devido ao movimento muscular e ruído da alimentação elétrica [Pouryayevali et al. 2014].

Como esses dados foram coletados em um ambiente controlado, espera-se que o sinal tenha menos ruído do que o ECG coletado enquanto a pessoa estava dirigindo, fazendo alguns movimentos, e assim por diante. Para sinais coletados em outras circunstâncias, provavelmente processos de filtragem mais sofisticados devams ser aplicados. A Figura 2 mostra o sinal bruto com ruído e o resultado da aplicação do filtro *Butterworth* de quarta ordem. Nesse caso, uma vez que você sabe como deve ser a forma de onda do ECG, é fácil ver que a forma de onda estava ruidosa. A escolha do filtro depende de cada sinal, e outras configurações podem funcionar melhor considerando o ruído no processo de aquisição.

3.4. Aumento de dados por segmentação

ECG é um sinal contínuo capturado de um dispositivo vestível em contato com o corpo. Geralmente é segmentado com base nos picos do sinal a ser usado como entrada no sistema biométrico. Nesse sentido, uma divisão por janelas aleatórias foi utilizada para gerar segmentos de sinal com duração de 3s, sem dispor de nenhuma informação de localização dos batimentos cardíacos nos segmentos, semelhantemente à técnica usada em [Zhang et al. 2017]. Considerando os aspectos práticos do processo de registro, decidiu-se por usar apenas 1 minuto de cada pessoa do banco de dados *PhysioNet Computing*

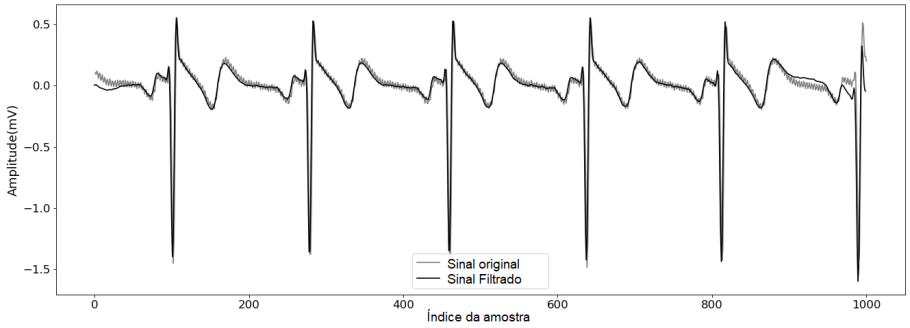

Figura 2. Sinal antes e após o uso do filtro *Butterworth*.

in Cardiology 2018. Este banco de dados tem frequência de 200 Hz e, portanto, cada segmento de 3 s teria pelo menos 600 amostras do sinal. Com esse tamanho de janela seria possível obter pelo menos um batimento cardíaco, permitindo a extração de mais *features* com menos esforço, uma vez que a frequência cardíaca típica varia de 40 a 208 batimentos por minuto.

Considerando o processo de segmentação tradicional, seria possível obter apenas 20 segmentos de 3s com os 60s de dados. Portanto, é pertinente adicionar uma etapa de geração de dados adicionais para melhor aproveitar os recursos dos dados originais. O aumento de dados é o processo de incrementar a quantidade e a diversidade dos dados utilizando para isso somente os dados previamente disponíveis. Este processo é oportuno porque as vezes as amostras colhidas na prática são insuficientes e uma coleta adicional não é possível, portanto, assim como é um processo que mitiga o problema aumentando a quantidade de dados e introduzindo alguma variabilidade. Para cada registro de 60s, coletou-se um exemplo de janela de ECG de 3s escolhida aleatoriamente, que pode incluir um conjunto diferente de batimentos cardíacos e apresentar algumas morfologias de sinal diferentes, dependendo de cada janela de sinal. Foi obtido 60 segmentos de 3s adicionais a partir dos dados iniciais de um individuo pela seleção aleatória de segmentos com a técnica descrita acima. Especificamente, há 80 segmentos para cada individuo, somando os segmentos originais e os aleatoriamente gerados. Isso ajuda o modelo de classificação ao fornecer dados de treino mais variados ao custo de talvez introduzir *overfitting* ao treino.

3.5. Extração de *features* e remoção de outliers

Um sinal de ECG tem muitas *features* que podem ser usadas, contudo, utilizar um número menor tem vantagem para evitar complexidade excessiva do modelo, o que acarretará em alto uso de recursos computacionais [Biel et al. 2001]. Nesse contexto, o ECG possui três características predominantemente usadas para autenticação do usuário: a onda P, o complexo QRS e a onda T, conforme mostrado na Figura 3.

O algoritmo de detecção do complexo QRS deve encontrar todos os pontos máximos locais a cada 3 s, e assim comparou-se os valores de amplitude para armazenar os picos. Inicialmente, o algoritmo encontra o pico máximo de R, e usando um limiar de 66% do valor máximo elimina os pontos abaixo deste valor [Barros et al. 2019]. Os pontos restantes são classificados como picos R. Os primeiros máximos locais anteriores e posteriores ao pico R são classificados como pontos Q e S, respectivamente

[Barros et al. 2018]. Dessa forma, o algoritmo encontra as amplitudes Q, R e S para cada batimento. Considerando que mais de um pico é apresentado em 3s, calculou-se a média e o desvio padrão da amplitude dos picos Q, R e S, que são nossas primeiras 6 *features* extraídas. A amplitude do pico QRS é calculada pela subtração do valor do pico R pela amplitude do pico S. A duração da onda R é calculada subtraindo o tempo do pico S menos o valor do pico Q. E o intervalo R-R é calculado quando mais de um pico R está dentro da janela de 3s.

Além do complexo QRS, também considerou-se os tempos de inicio e fim do intervalo QRS, bem como os pontos de pico P e T. O ponto inicial de cada onda de ECG é considerado o ponto em que a onda inicia a inclinação e o deslocamento é o final de uma onda. Especificamente, o início e o fim do complexo QRS são calculados após os pontos Q e S. Usando Q como referência, o extrator deve voltar ao sinal e calcular a maior inclinação entre o novo ponto e o ponto Q, usando uma janela de tempo de 40 ms. Um procedimento semelhante é feito para encontrar o deslocamento, mas considerando o ponto S como referência, e o extrator caminha até o final do segmento. Finalmente, considerou-se uma janela começando do complexo QRS para encontrar P e T. Os picos mais relevantes encontrados nesta janela de pesquisa são declarados P e T em cada lado do complexo QRS. Depois de coletar essas *features*, é possível calcular a amplitude média das seguintes *features* adicionais: início e fim dos intervalos P e T e QRS, as distâncias QS, QT e recursos relacionados ao tempo a partir desses pontos relevantes.

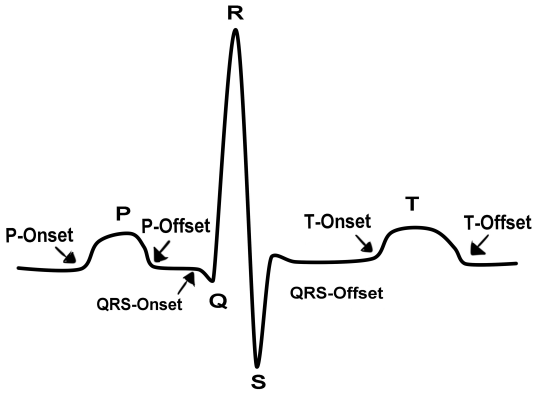

Figura 3. Exemplo de ciclo P-QRS-T, com os picos de todas as formas de onda relevantes.

Essas *features* relacionadas ao tempo são as médias das seguintes medidas: intervalos QT, ST e PP, onda T, e segmentos PQ, ST e TP. Sendo estas simplesmente medições da diferença de tempo entre as ondas principais do sinal de ECG ou a duração de uma única onda. Todas as *features* discutidas são exibidos na Tabela 2.

Este método de extração de *features* obtém um grande número de recursos do sinal de ECG, pretendendo fornecer informações adicionais ao modelo de classificação para aumentar o fator de distinção entre diferentes classes (indivíduos). Há também uma etapa de remoção de medições anômalas, ou *outliers*, que visa limpar dados estatisticamente dispersos para garantir que o modelo não seja alimentado com eles. Erros durante a aquisição de sinais são a principal causa de dados anormais, e isso pode ocorrer devido a fatores como movimento excessivo dos sensores, interferência eletrostática durante a

Tabela 2. *Features* capturadas diretamente do sinal ECG.

N°	*Features*
1	Amplitude média do pico Q (μV)
2	Amplitude média do pico R (μV)
3	Amplitude média do pico S (μV)
4	Desvio padrão do pico Q
5	Desvio padrão do pico R
6	Desvio padrão do pico S
7	Amplitude do complexo QRS (μV)
8	Duração da onda R (ms)
9	Intervalo R-R (ms)
10	Amplitude média do pico P (μV)
11	Amplitude média do pico T (μV)
12	Distância média do intervalo QS
13	Distância média do intervalo QT
14	Amplitude média do início do complexo QRS (μV)
15	Amplitude média do fim do complexo QRS (μV)
16	Intervalo QT médio
17	Intervalo ST médio
18	Média da onda T
19	Média do segmento PQ
20	Média do segmento ST
21	Média do segmento TP
22	Intervalo médio PP

medição, remoção indevida de sensores durante a medição, entre outros. A técnica de remoção foi o intervalo interquartil. Os limites superior e inferior dos quartis foram de 0,5% para ambos os limites. Todos os dados acima ou abaixo desses limites foram descartados. O processo foi aplicado a cada característica do conjunto de dados. Esta abordagem é semelhante ao que a maioria dos sistemas biométricos fazem na prática, isto é, requirir boas amostras para criar um modelo, e se as amostras não atenderem a qualidade mínima, elas são descartadas e a coleta é realizada novamente.

3.6. As tarefas de classificação

A classificação por *random forests* foi escolhida por apresentar bons resultados sem a necessidade de um ajuste fino extenso de parâmetros [Barros et al. 2018]. O algoritmo *random forests* busca reduzir o erro geral de classificação pela predição por decisão da maioria, em um esquema de votação onde uma coleção de arvores de decisão treinadas com dados selecionados aleatoriamente, votam suas próprias predições individuais de um dado de entrada. *random forests* apresenta bom desempenho em dois aspectos críticos: detecção de anomalias e sobre-ajuste do modelo. Durante o processo de treinamento, os dados anômalos estarão presentes em algumas árvores, mas não em todas, e assim o sistema de votação garante que as anomalias sejam minimizadas. O sistema de votação também minimiza o efeito do *overfitting* em relação a árvores de decisão individuais.

4. Avaliação

Nesta seção, é descrita a metodologia, as métricas e os resultados usados para avaliar o modelo DETECT para identificação do usuário com base no sinal de ECG.

4.1. Metodologia

Especificamente foi avaliado o cenário de identificação, onde um único modelo é treinado para diferenciar muitos usuários de acordo com os recursos extraídos do sinal de ECG como entrada. Este modelo único foi implementado como um classificador um contra o resto *(One-vs-Rest)*. Este classificador é construído em cima de instâncias de classificadores de floresta aleatórias como seus estimadores de base. Esses estimadores base são responsáveis por aprender as características do sinal de ECG de cada pessoa. Usar um classificador um contra o resto significa dividir o problema inicial de identificação de 1985 classes em 1985 problemas de classificação binária individuais. Então, quando o modelo classificador de um contra o resto é solicitado para predizer uma determinada classe, cada um dos 1985 estimadores de base calcula as probabilidades dessa classe pertencente à sua própria classe e então o modelo que gerou a maior probabilidade é usado para realmente prever a classe. Essa abordagem é comumente utilizada em problemas de classificação multi-classes e é conhecida por ser computacionalmente eficiente em relação a construção de um único classificador. Além de permitir interpretabilidade dos modelos construídos, já que é possível inspecionar modelos particulares de cada classe.

A implementação foi feita principalmente usando o pacote scikit-learn [Pedregosa et al. 2011] e para cada cenário, os testes consistiram no mesmo modelo: Um classificador *OneVsRest* constituído de classificadores de floresta aleatórios como seus estimadores de base. Para cada cenário, o conjunto de dados foi dividido em conjuntos de treinamento e teste na proporção de 80% para treinamento e 20% para teste (com estado para controle da aleatoriedade da divisão). Em seguida, uma busca em grade (*Grid Search*) foi realizada para ajustar os parâmetros do classificador de florestas aleatórios em um esquema de validação cruzada de 3 subconjuntos. Todos os testes foram executados no servidor do laboratório (CPU Intel (R) Xeon (R) Silver 4112 a 2,60 GHz e 64 Gb de RAM) ou na plataforma Google Colab.

Os parâmetros e valores avaliados foram *max-depth* (60, 80, 100), *min-samples-leaf* (3, 4, 5), *min-samples-split* (8, 10, 12) e *n-estimators* (80, 100, 120). Especificamente, a profundidade máxima da árvore (*max-depth*) significa que os nós são expandidos até que todas as folhas estejam puras. O *min-samples-leaf* é o número mínimo de amostras necessárias para estar em um nó folha. Um ponto de divisão em qualquer profundidade só será considerado se deixar pelo menos *min-samples-leaf* amostras de treinamento em cada um dos ramos esquerdo e direito. O *min-samples-split* é o número mínimo de amostras necessárias para dividir um nó interno. *n-estimators* refere-se ao número de árvores de decisão na floresta aleatória. Definir um valor de estado aleatório também foi necessário para inicializar as amostras de treinamento usadas ao construir árvores e ao amostrar *features* a serem consideradas ao procurar a melhor divisão em nós das árvores de decisão. Com base na pesquisa em grade, encontrou-se os melhores parâmetros para o classificador base, sendo eles 100 para *max-depth*, 3 para *min-samples-leaf*, 10 para *min-samples-split* e 120 para *n-estimators*. A Figura 4 mostra um exemplo de uma árvore de decisão única de um classificador base de florestas aleatórias ao interpretar os dados

recebidos durante a tarefa de identificação. Foi concluído também que valores maiores que os testados aqui para esses parâmetros indicam classificadores mais especializados na previsão das classes ao custo de um aumento considerável nos recursos computacionais e um potencial *overfitting* do modelo.

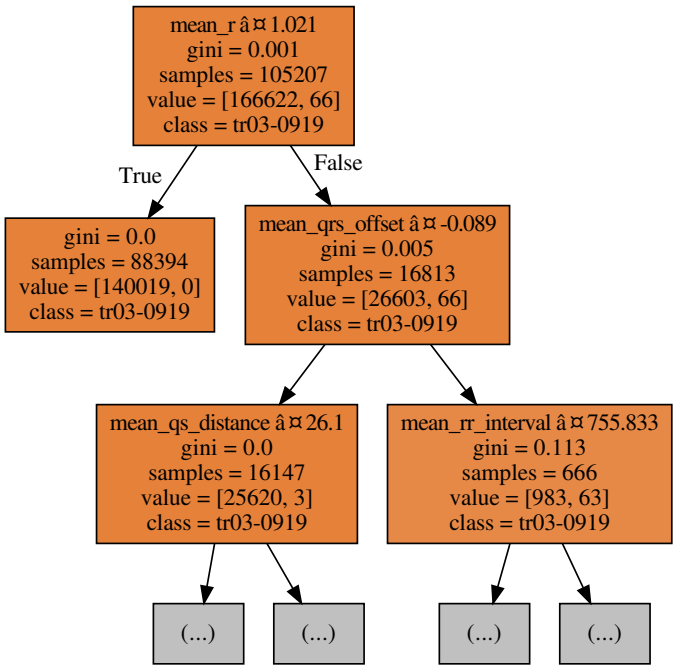

Figura 4. Parte de uma árvore de decisão de uma floresta aleatória *(random forest)*.

Foram analisados quatro modelos de identificação de usuários baseados em ECG, são eles: Cenário 1: Dados sem nenhuma avaliação de qualidade; Cenário 2: Dados com um número maior de exemplos usando aumento de dados; Cenário 3: Dados com um número maior de exemplos e remoção de dados anômalos usando a etapa de remoção de valores fora do intervalo quartil; E por último, o Cenário 4: O Modelo DETECT - usando as todas as etapas de avaliação de qualidade de dados apresentado na Seção 3.

Foram selecionadas as métricas de acordo com os resultados apresentados em uma matriz de confusão para avaliar o modelo DETECT. A matriz de confusão tabula os resultados previstos versus as observações reais. Partindo da matriz de confusão, tem-se as seguintes métricas base: *Positivo verdadeiro* (TP), significando instâncias corretamente identificadas. *Falso positivo* (FP) representando instâncias identificadas incorretamente. *True Negative* (TN) significando instâncias corretamente rejeitadas e *False Negative* (FN) significando instâncias incorretamente rejeitadas.

Com base na matriz de confusão, também foram derivados alguns indicadores de desempenho adicionais para avaliar o modelo em termos de identificação biométrica. Usou-se métricas do campo da identificação biométrica, duas seguintes métricas: *Taxa de falsa aceitação* (FAR) e *Taxa de falsa rejeição* (FRR) para avaliar a proporção de identificações incorretas. FAR avalia instâncias que foram aceitas disfarçados como classes diferentes da pretendida, enquanto o FRR avalia as instâncias que foram rejeitadas incorretamente como membros de uma determinada classe. FAR e FRR também são

conhecidos no campo de métricas de avaliação de aprendizado de máquina como *Taxa de falsos positivos* (FPR) e *Taxa de falsos negativos* (FNR), respectivamente (conforme demonstrado nas Equações (1) e (2).

$$FAR = \frac{FP}{FP + TN} \tag{1}$$

$$FRR = \frac{FN}{TP + FN} \tag{2}$$

Além disso, usou-se métricas de propósito de avaliação geral: *Precision*, *Recall* e *F1-Score*. *Precision* leva em consideração o número de atributos classificados corretamente para uma determinada classe em comparação com o número de características classificadas corretamente e incorretamente para aquela classe, demonstrado na equação (3). *Precision* mede a exatidão do classificador e a relevância das classificações positivas. Maiores valores de *precision* significam maiores números de verdadeiros positivos e um menor número de falsos positivos.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Para complementar o entendimento de precisão, alguns trabalhos na literatura consideram a métrica *recall* como mostrado na Equação (4). *Recall* é a taxa de verdadeiros positivos, ou seja, a proporção de positivo predito corretamente em relação ao número total de observações realmente positivas.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

O F1-score é a média harmônica entre as métricas *precision* e *recall*, conforme mostrado na Equação (5).

$$F1 = 2 * \frac{precision * recall}{precision + recall} \tag{5}$$

4.2. Resultados

Para fins de identificação, alguns segundos do sinal de ECG coletado são passados para o classificador para que ele possa decidir a qual usuário o sinal de ECG pertence. Nesse sentido, realizou-se uma análise quantitativa para avaliar o FAR e o FRR alcançados para cada cenário. A Figura 5 mostra os resultados do FAR para os diferentes modelos de identificação de usuário, considerando o conjunto de dados de desafio PHYSIONET com 1985 indivíduos. Os individuos no cenário 1 tinham valores FAR contidos no maior intervalo interquartil e com os valores mais altos do que qualquer outro cenário. À medida que se aumentou o tratamento de dados, o intervalo interquartil estreitou, e os valores FAR tornaram-se menores (exceto para o cenário 3 que tinha valores FAR ligeiramente mais altos do que o cenário 2). Nos resultados de DETECT, nota-se o intervalo interquartil

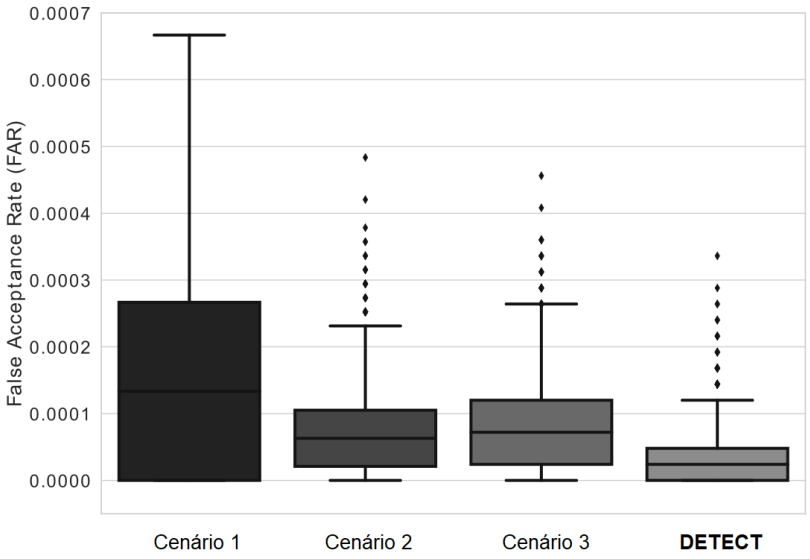

Figura 5. Resultados de FAR para cada cenário.

mais estreito e o menor valores FAR, que representam uma baixa variação nos valores FAR e uma baixa taxa de identificações incorretas.

Em relação à FRR, conforme demonstrado na Figura 6, pode-se analisar um comportamento muito semelhante ao que aconteceu com as taxas de falsa aceitação: O tratamento de dados estreita e diminui o intervalo interquartil das taxas. Os indivíduos no cenário 1 tinham valores de FRR contidos no maior intervalo interquartil e com os valores mais altos do que qualquer outro cenário. No modelo DETECT, há o intervalo interquartil mais estreito e os menores valores FAR, que representam uma variação baixa nos valores FRR e uma baixa taxa de identificações incorretas. Em preocupação com o elevado número de indivíduos, também foi avaliado o desvio padrão dos valores FAR e FRR para cada etapa, conforme visto na Tabela 3.

Tabela 3. Média e desvio padrão de FAR e FRR para cada cenário.

Cenário	**FAR**	**STD**	**FRR**	**STD**
Cenário 1	0.000194	0.000207	0.383813	0.312530
Cenário 2	0.000079	0.000069	0.156959	0.134858
Cenário 3	0.000080	0.000071	0.153801	0.143330
Cenário 4 (DETECT)	**0.000033**	**0.000042**	**0.064003**	**0.084876**

No cenário de autenticação contínua, o ECG seria obtido como um fluxo e as métricas seriam calculadas considerando as tentativas bem-sucedidas contra todas as tentativas durante um determinado tempo para cada individuo. Nesse sentido, foi realizado uma análise quantitativa para avaliar o *precision* alcançado para cada individuo. A Figura 7 mostra o histograma de *precision* para diferentes modelos de identificação do usuário com base em ECG. Considerando para este cenário também o banco de dado *PhysioNet Computing in Cardiology 2018* com 1985 indivíduos. Ao analisar os resultados, pode-se

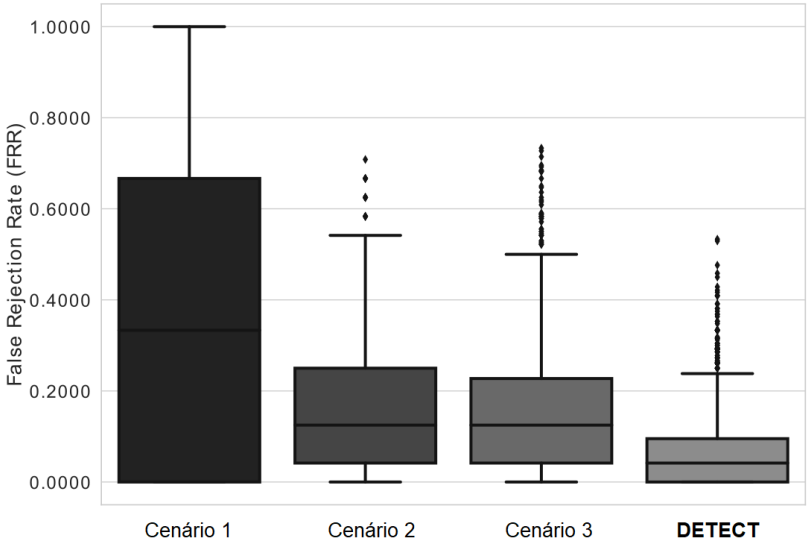

Figura 6. Resultados de FRR para cada cenário.

concluir que o modelo DETECT fornece um *precision* maior do que outros modelos. Especificamente, mais de 1200 indivíduos tiveram *precision* superior a 90%, o que é cerca de 60% da população. Isso porque nosso modelo considera apenas dados com qualidade mínima, evitando falsos positivos causados por outliers.

A métrica de precisão para dados sem qualquer avaliação de qualidade (Cenário 1) apresenta um resultado esparso, onde cada um dos 5 *bins* tem algumas centenas de indivíduos cada, o que representa que o modelo tem dificuldade em identificar indivíduos com precisão devido a um pequeno número de exemplos. Ao aumentar o número de exemplos com aumento de dados (Cenário 2) e remover outliers (Cenário 3), pode-se ver que a altura das barras associadas a uma pontuação de alta *precision* crescem, ao passo que a altura das barras associadas a pontuações de *precision* mais baixa diminuem e quase se esvaziam. No Cenário 2, mais de 1200 indivíduos alcançaram uma pontuação de *precision* de pelo menos 70%), enquanto no Cenário 3 esse número é ainda maior (perto de 1400 indivíduos alcançaram uma pontuação de precisão de pelo menos 70%). E, finalmente, usando DETECT há cerca de 1700 indivíduos com uma pontuação de *precision* de pelo menos 70%.

A Figura 8 mostra os resultados de *precision* para cada um dos diferentes cenários de avaliação de dados. Ao analisar os resultados, conclui-se que o DETECT obteve maior *precision* em relação ao outros cenários analisados. Isso ocorre porque segmentos são extraídos aleatoriamente do fluxo de ECG, o que nos permite aumentar o tamanho do conjunto de dados. O modelo DETECT também realizou a remoção de *outliers* e foi aumentado o número de *features* extraídas do sinal ECG para avaliar o aumento das métricas de segurança para identificação do usuário. Assim então, o modelo DETECT aumenta o *precision* em 25%, 12% e 7% em comparação com o Cenário 1, Cenário 2 e Cenário 3, respectivamente. Isso significa que o modelo DETECT é capaz de fornecer uma proba-

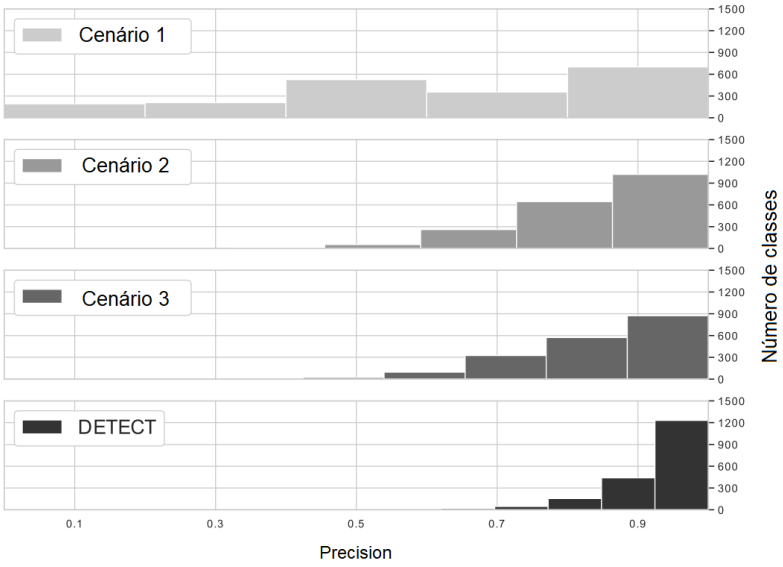

Figura 7. Resultados de distribuição da métrica *precision* para os diferentes cenários.

bilidade maior de que um usuário selecionado aleatoriamente seja previamente identificado, ou seja, quantos das tentativas de identificação retornaram positivos verdadeiros ou quantos das tentativas de identificação não encontraram um individuo, ou seja, negativos verdadeiros.

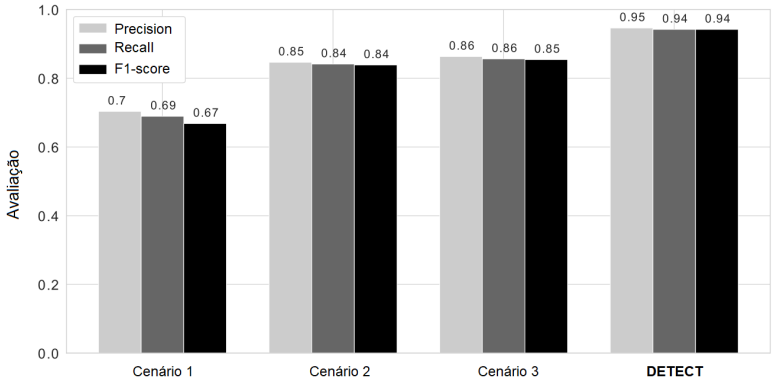

Figura 8. *Precision*, *recall* e *f1-score* para cada cenário.

A Figura 8 também mostra os resultados de *recall* para cada um dos diferentes cenários de avaliação de dados. conclui-se que os resultados de *recall* seguem quase a mesma proporção em comparação com os resultados de *precision*. Cada etapa da proposta de avaliação de dados contribuiu para o aumento geral das métricas avaliadas. É possível perceber que o processo de aumento de dados foi a etapa que mais contribuiu. Considerando que a quantidade de dados é importantes no treinamento do modelo de identificação [Halevy et al. 2009], esse resultado era esperado. A adição de novas *featu-*

res foi a segunda maior contribuinte. Especificamente, no primeiro cenário (dados sem qualquer avaliação), obteve-se uma um *recall* médio de 66%. O segundo cenário (aumentando o número de exemplos) alcançou um recall médio de 78%. O terceiro cenário (aumentando o número de exemplos e removendo outliers) obteve um recall médio de 82%. Finalmente, considerando o modelo DETECT (aumentando o número de exemplos e removendo outliers, e também considerando as *features* início e fim do intervalo QRS e pontos P e S), alcançou-se um *recall* médio de 89%. Isso significa que DETECT tem uma probabilidade maior de que um usuário selecionado aleatoriamente seja recuperado em uma pesquisa, ou seja, quantos dos verdadeiros positivos foram lembrados (encontrados) ou quantos dos acertos corretos foram encontrados.

Por último, a *Figura 9* mostra os resultados da pontuação F1 para cada um dos diferentes cenários de avaliação de dados. Cenário 1, Cenário 2, Cenário 3 e o DETECT alcançaram uma pontuação F1 de 62,28%, 72,45%, 76,4%, 82,49%, respectivamente. Portanto, DETECT tem uma proporção mais alta de positivos preditos corretamente em função do número total de observações realmente positivas, ou seja, indica uma média harmônica alta para ambos, *precision* e *recall*.

Para avaliar como o modelo DETECT funcionaria com conjunto de dados com um número de classes-alvo crescente, dividiu-se o conjunto original em subconjuntos de centenas de indivíduos e construiu-se um modelo com os mesmos parâmetros dos quais foram usados no modelo para todo o conjunto de dados, conforme mostrado na Figura 9. Analisando os resultados, é notável que *precision* e o F1-score começam a reduzir assim que o número de classes alvo aumenta. Por exemplo, a *precision* começa em torno de 78% com um modelo para 100 indivíduos e termina perto de 60% usando um modelo para 1500 indivíduos no cenário com um número maior de exemplos e remoção de outliers. Isso ocorre porque o modelo de aprendizado precisa de mais recursos de dados para generalizar um grande número de classes. Por outro lado, o modelo DETECT manteve o *precision* 20% superior no modelo para 100 indivíduos e 16% superior no modelo para 1500 indivíduos em comparação com o modelo treinado com dados sem avaliação de qualidade. Comportamentos semelhantes acontecem para os resultados de F1-score.

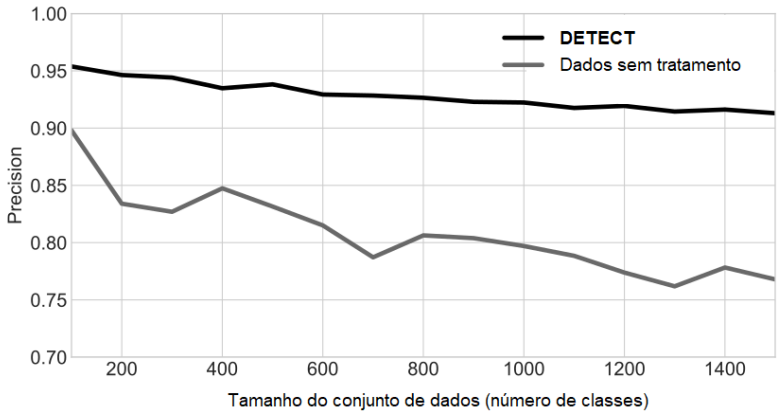

Figura 9. *Precision* para modelos treinados com crescentes números de classes.

5. Conclusão e trabalhos futuros

Este trabalho apresentou um modelo de avaliação de dados para elevar as métricas de FAR, FRR, *precision*, *recall* e f1-score de sistemas de identificação baseados em sinal de eletrocardiograma. Ele permite avaliar o desempenho do sistema biométrico com cerca de 2000 indivíduos, simulando ssitemas em cenários da realidade. Observou-se que, ao propor um conjunto de tarefas de avaliação de dados, as métricas podem melhorar consideravelmente usando o classificador um *One-vs-Rest* baseado no classificador *random forest*. Sem qualquer avaliação nos dados, as taxas médias de falsa aceitação foram 0,0194%, as taxas médias de falsa rejeição foram 38,38% e aproximadamente 600 indivíduos puderam ser classificados com pelo menos 80% de precisão. Depois de aplicar os dados na abordagem proposta, as taxas médias de falsa aceitação foram 83% menores (0,0033%), as taxas médias de falsa rejeição foram 83% menores (6,40%) e aproximadamente 1200 indivíduos puderam ser classificados com pelo menos 80% de *precision*. Mesmo quando se trabalha com um número pequeno de indivíduos, um conjunto de avaliação de dados deve ser utilizado, e a proposta alcançou bons resultados.

Para um subconjunto de 100 indivíduos, alcançou-se um *precision* de 95% considerando o modelo proposto contra 89% de *precision* sem estágios de processamento. Entendeu-se que o conjunto de estágios de processamento pode ser aplicado a todo o processo de construção de conjunto de dados para melhorar os resultados de tarefas de classificação. A proposta é simples e fácil de replicar em outros bancos de dados e uma comparação direta não seria justa, pois a maioria dos trabalhos considera bancos de dados pequenos. Assim, a comparação deu-se em termos de como os resultados melhoraram a cada etapa da avaliação dos dados. Como trabalhos futuros, uma abordagem considerando outros sinais e informações de contexto por meio de serviços e outros biomarcadores seria um caminho a seguir.*

6. Publicações

Esse trabalho de conclusão foi publicado em 22 de maio de 2020 no *journal sensors*, sob o título *Data Improvement Model Based on ECG Biometric for User Authentication and Identification*, [Barros et al. 2020].

7. Agradecimentos

A publicação do artigo original e o desenvolvimento deste trabalho de conclusão de curso foram ambos apoiados financeiramente pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Referências

Barros, A. et al. (2018). Ecg-based user authentication and identification method on vanetss. *In Proceedings of the 10th Latin America Networking Conference (LANC), São Paulo, Brazil*.

Barros, A. et al. (2020). Data improvement model based on ecg biometric for user authentication and identification. *Sensors*.

Barros, A., Rosário, D., Resque, P., and Cerqueira, E. (2019). Heart of iot: Ecg as biometric sign for authentication and identification. *Proceedings of the 15th International Wireless Communications Mobile Computing Conference (IWCMC), Tangier, Morocco*.

Bastos, L., Rosario, D., Cerqueira, E., Santos, A., and Nogueira, M. (2019). Filtering parameters selection method and peaks extraction for ecg and ppg signals. *In Proceedings of the IEEE Latin-American Conference on Communications (LATINCOM), Salvador, Brazil.*

Biel, L., Pettersson, O., Philipson, L., and Wide, P. (2001). Ecg analysis: A new approach in human identification. *IEEE Transactions on Instrumentation and Measurement*.

Camara, C., Peris-Lopez, P., Gonzalez-Manzano, L., and Tapiador, J. (2018). Real-time electrocardiogram streams for continuous authentication. *Applied Soft Compuing Journal*.

Cao, P. et al. (2020). A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation. *Biomedical Signal Processing Control*.

Goldberger, A. et al. (2000). Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiological signals.

Hale, M., Lotfy, K., Gamble, R., Walter, C., and Lin, J. (2019). Developing a platform to evaluate and assess the security of wearable devices. *Digital Communications and Networks*.

Halevy, A., Norvig, P., and Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*.

Jain, A., Nandakumar, K., and Ross, A. (2019). 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*.

Lemley, J., Bazrafkan, S., and Corcoran, P. (2017). Smart augmentation learning an optimal data augmentation strategy. *IEEE Acess 2017*.

Lugovaya, T. (2005). Biometric human identification based on electrocardiogram. master's thesis. *Faculty of Computing Technologies and Informatics, Electrotechnical University 'LETI', Saint-Petersburg, Russia*.

Marieb, E. and Hoehn, K. (2013). Human anatomy and physiology. *9th ed.; Pearson: Glenview, IL, USA*.

O'Donnell (2019). Lindsey researchers bypass apple faceid using biometrics 'achilles heel'. threatpost website.

Palma, A., Alotaiby, T., Alrshoud, S., Alshebeili, S., and Aljafar, L. (2019). Ecg-based subject identification using statistical features and random forests. *Sensors*.

Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pinto, J. R., Cardoso, J., and Lourenço, A. (2018). Evolution, current challenges, and future possibilities in ecg biometrics. *IEEE Access 2018*.

Pouryayevali, S., Wahabi, S., Hari, S., and Hatzinakos, D. (2014). On establishing evaluation standards for ecg biometrics. *In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy*.

Resque, P., Barros, A., Rosário, D., and Cerqueira, E. (2019). An investigation of different machine learning approaches for epileptic seizure detection. *In Proceedings of the*

2019 15th International Wireless Communications and Mobile Computing Conference (IWCMC), Tangier, Morocco.

Seneviratne, S. et al. (2017). A survey of wearable devices and challenges. *IEEE Communications Surveys and Tutorials*.

Su, K. et al. (2019). Human identification using finger vein and ecg signals. *Neurocomputing*.

Zhang, Q., Zhou, D., and Zeng, X. (2017). Heartid: A multiresolution convolutional neural network for ecg-based biometric human identification in smart health applications. *IEEE Access 2017*.

Zhang, Y., Gravina, R., Lu, H., Villari, M., and Fortino, G. (2019a). Pea: Parallel electrocardiogram-based authentication for smart healthcare systems. *Journal of Network and Computer Applications*.

Zhang, Y. and Wu, J. (2016). Practical human authentication method based on piecewise corrected electrocardiogram. *Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China*.

Zhang, Y., Xiao, Z., Guo, Z., and Wang, Z. (2019b). Ecg-based personal recognition using a convolutional neural network. *Pattern Recognition Letters*.