

Seleção de característica via Algoritmos Evolucionários para Reconhecimento Biométrico baseado em Sinais de ECG

Felipe G. S. Teodoro, Daniel M. M. da Costa, Sarajane Marques Peres, Clodoaldo A. M. Lima

Programa de Pós-Graduação em Sistemas de Informação

Escola de Artes, Ciências e Humanidades

Universidade de São Paulo

São Paulo, SP, Brasil

E-mail: {fteodoro,daniel.moura.costa,sarajane,c.lima}@usp.br

Resumo—Atualmente a necessidade de identificação de pessoas de forma precisa vem recebendo atenção da comunidade científica e da indústria, sendo a área de Biometria uma dessas novas estratégias. Biometria é o nome dado ao emprego de características físicas ou comportamentais de uma pessoa para fazer a sua identificação. Modalidades biométricas tradicionais (Face, Íris e Impressão Digital) vêm sofrendo ataques por circunvenção e repetição. De forma a superar esses problemas os sinais biomédicos, tais como sinais de eletrocardiograma (ECG) e sinais de eletroencefalograma (EEG), têm sido estudados para a identificação pessoal. Os sinais de ECG, responsáveis por auxiliar na detecção de cardiopatias, vêm obtendo resultados satisfatórios no reconhecimento biométrico, com o uso de características fiduciais, não-fiduciais e híbridas. Visando selecionar um subconjunto de características mais discriminativas várias técnicas têm sido propostas na literatura. Este trabalho analisa o impacto de algumas estratégias de seleção de característica tais como busca gulosa, algoritmo genético e algoritmo memético sobre desempenho dos Sistemas Biométricos baseados em ECG ao empregar k -Vizinhos mais Próximos e Análise de Discriminante Linear como classificador. Os resultados demonstram que existe um subconjunto de características extraídas do sinal de ECG capaz de fornecer altas taxas de reconhecimento.

Palavras-Chave—Reconhecimento Biométrico, Sinal de Eletrocardiograma, Seleção de característica, Algoritmo Genético, Algoritmo Memético.

I. INTRODUÇÃO

Um sistema biométrico pode ser classificado como um sistema de reconhecimento de padrões, o qual utiliza características físicas ou comportamentais do usuário para realizar o seu reconhecimento [1]. Qualquer característica física ou comportamental humana pode ser usada como modalidade biométrica, desde que esta satisfaça alguns requisitos básicos como universalidade, unicidade, permanência, mensurabilidade e aceitabilidade [2].

Recentemente, esforços têm sido realizados para a pesquisa de novas modalidades biométricas que possam ser utilizadas para reconhecimento de indivíduos. As modalidades biométricas, que englobam sinais que são utilizados em diagnósticos clínicos, como os sinais de eletrocardiograma (ECG), eletroencefalograma (EEG) e eletromiograma (EMG), têm sido considerados modalidades biométricas emergentes [3], já que esses sinais respeitam os requisitos levantados anteriormente. Neste trabalho utilizaremos sinais ECG para identificação de indivíduos.

O primeiro trabalho que utilizou ECG como modalidade biométrica foi proposto por [4], desde então diversos trabalhos vêm sendo propostos com o objetivo de garantir o reconhecimento de indivíduos. Reconhecimento baseado em sinais de ECG pode ser realizado utilizando três métodos diferentes de extração de características [2]: fiducial, não-fiducial ou híbridos. Os métodos fiduciais extraem características temporais, amplitude, área, ângulos e inclinações sobre o complexo QRS detectado no sinal de ECG [4], [5], [6]. Os métodos não-fiduciais não utilizam pontos como características, estes fazem uso de estratégias tais como Transformada Wavelet [7] e coeficientes de Autocorrelação [8]. Os métodos híbridos fazem a combinação de ambas as abordagens [9].

Os Algoritmos Genéticos (GAs) [10] são conhecidos por produzirem soluções de alta qualidade para problemas complexos [11], podendo ser utilizado no processo de seleção de características [12]. Infelizmente, devido à natureza inerente do GA, este frequentemente leva muito tempo para localizar um ótimo local na região de convergência e pode algumas vezes não encontrar o ótimo com uma precisão satisfatória. De forma a solucionar este problema foram propostos os Algoritmos Meméticos (MA) [13][14], o qual combina um operador de busca local ao GA de forma a encontrar soluções melhores. Este trabalho analisa o impacto de algumas estratégias de seleção de característica tais como busca gulosa, GA e MA sobre desempenho dos Sistemas Biométricos baseados em ECG ao empregar k -Vizinhos mais Próximos (kNN) e Análise de Discriminante Linear (LDA) como classificador.

Este artigo está organizado nas seguintes seções: uma descrição formal do sinal de ECG, considerando o contexto de reconhecimento biométrico, é apresentada na Seção II; as características biométricas utilizadas são descritas na seção III; uma revisão sobre os classificadores utilizados é descrita na Seção IV; os experimentos computacionais são apresentados na Seção VI e finalmente na Seção VII as conclusões são descritas.

II. O SINAL DE ELETROCARDIOGRAMA

O eletrocardiograma registra as oscilações do potencial elétrico gerado pelo músculo cardíaco, capturadas por eletrodos colocados sobre a pele em pontos específicos do corpo

e medidas em unidades de milivolts (mV). Um eletrocardiograma normal é composto pelas ondas P, Q, R, S e T. A onda P ocorre devido à despolarização do átrio, logo antes da sístole atrial. Após a onda P, aparecem as ondas Q, R e S, também chamadas de *complexo QRS*, que indicam a despolarização dos ventrículos, antes da sístole ventricular e a onda T que representa a fase de repolarização dos ventrículos, indicando a diástole ventricular. A diástole atrial ocorre no mesmo momento do complexo QRS dificultando o registro da onda T atrial.

III. EXTRAÇÃO DE CARACTERÍSTICAS DOS SINAIS DE ECG

Em [4] foi demonstrado que um único canal do sinal de ECG contém informação suficiente para permitir o reconhecimento biométrico, este trabalho seguiu esta abordagem.

A fim de realizar a extração das características fiduciais é necessário realizar a segmentação e detecção do complexo QRS nos batimentos cardíacos, para isso utilizamos o algoritmo de Pan-Tompkins [15]. Já em relação às características não-fiduciais não é necessário a detecção do complexo QRS, porém neste trabalho fizemos a análise das características fiduciais sobre cada complexo QRS, totalizando 545 características distintas, sendo 40 características fiduciais e 505 características não-fiduciais.

A. Extração de características fiduciais

O algoritmo Pan-Tompkins [15] pode ser resumido nos seguintes passos: inicialmente, dois filtros (um filtro passa baixa e um filtro passa alta) são aplicados a fim de enfatizar as características dos complexos QRS e diminuir as informações irrelevantes. A combinação desses filtros forma um filtro passa-banda de largura de banda que varia de 5 a 11 Hz. Em seguida, um filtro derivativo é aplicado ao sinal, com o objetivo de reduzir o ruído e enfatizar os complexos QRS. Na próxima etapa, os valores resultantes do filtro derivativo são elevados ao quadrado, enfatizando mais o complexo QRS. Posteriormente, um integrador de janela móvel é aplicado, o qual calcula a média das últimas 30 amostras do sinal, possibilitando a detecção de alguns picos R. Em seguida, alguns limiares são ajustados automaticamente para detectar os picos das ondas Q e S. Quando o intervalo entre dois picos R é menor que 360 ms, uma análise é realizada para determinar se o complexo QRS foi identificado corretamente ou se ele corresponde a uma onda T. Se a inclinação máxima que ocorre nessa onda é menor que metade da onda QRS anterior, então essa onda é identificada como uma onda T, caso contrário classificamos o intervalo como um novo complexo QRS.

O algoritmo de Pan-Tompkins não menciona nada a respeito da detecção da onda P [16]. Neste trabalho, realizamos uma busca nos m pontos antes da onda Q de modo a detectar o máximo local, representando a onda P. De modo a melhorar a detecção do pico da onda T foi realizada uma análise de n pontos após a onda S, de forma localizar o máximo local, que é o pico da onda T.

Com o intuito de melhorar a precisão dos valores obtidos para os picos das ondas P, Q, R, S e T, empregamos uma janela deslizante sobre o sinal de ECG contendo 5 batimentos, sendo estes centralizados em relação ao pico R e gerando um novo batimento. De forma geral, para cada complexo $QRS(i)$ é calculado um novo QRS usando a média dos pontos do $QRS(i)$ até o $QRS(i+k-1)$, sendo k o número de QRS que foram utilizados na janela.

A Tabela I apresenta as características extraídas do sinal de ECG. A título de ilustração algumas características são mostradas na Figura 1.

Tabela I
CONJUNTO DE CARACTERÍSTICAS FIDUCIAIS.

Características	
Intervalos	QR, RS, PQ, PR, RT, PS, QT, PT
Amplitudes	QRS, PT, SP, QP, PR, TR, QS, QT, ST, SR
Ângulo dos Segmentos	PQR, QRS, RST
Área dos polígonos formados a partir das coordenadas	$\{x(P), y(P)\}, \{x(P), y(R)\}, \{x(R), y(R)\}$
	$\{x(P), y(P)\}, \{x(Q), y(Q)\}, \{x(P), y(Q)\}$
	$\{x(P), y(P)\}, \{x(R), y(R)\}, \{x(Q), y(Q)\}$
	$\{x(R), y(R)\}, \{x(S), y(S)\}, \{x(T), y(T)\}$
	$\{x(S), y(S)\}, \{x(T), y(S)\}, \{x(T), y(T)\}$
Área do polígono formado a partir das coordenadas do complexo QRS	$\{x(Q), y(Q)\}, \{x(R), y(R)\}, \{x(S), y(S)\}$
	$\{x(Q), y(Q)\}, \{x(R), y(R)\}, \{x(S), y(S)\}$
Soma das características	Soma de todas as áreas
Distância Euclidiana	$\{x(T), y(R)\}, \{x(Q), y(Q)\}$
	$\{x(P), y(R)\}, \{x(S), y(S)\}$
Inclinações	Inclinações obtidas das ondas P, R e T

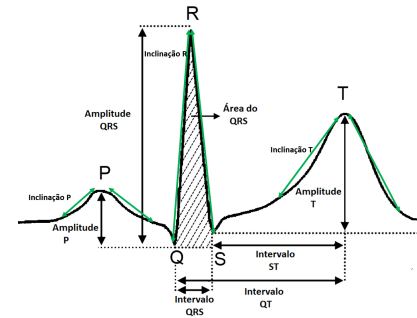


Fig. 1. Exemplo de características fiduciais extraídas do complexo QRS.

O cálculo da inclinação das ondas P, R e T foi realizado seguindo a abordagem de [17] onde para cada onda são extraídas duas inclinações a partir do máximo local de cada onda e o seu mínimo local a esquerda e a direita.

B. Extração de Características não-fiduciais

Para extração de características não-fiduciais utilizamos a Transformada Discreta Cosseno (DCT), características temporais (TD), função de autocorrelação (AC), modelos autoregressivos (AR), Transformada Rápida de Fourier (FFT), Transformada Wavelet (WT), métodos de estimativa da dimensão fractal (FD) e Decomposição de Modo Empírico (EMD).

1) *Transformada Discreta Cosseno*: A DCT é utilizada na compressão de imagens, como no reconhecimento facial [18] e por impressão digital [19]. Os seus coeficientes

podem ser obtidos usando a seguinte equação: $y(k) = w(k) \sum_{n=1}^N x(n) \cos(\frac{\pi(2n-1)(k-1)}{2N})$, onde $k = 1, 2, \dots, N$, $x(n)$ é um sinal de entrada e $w(k) = 1/\sqrt{N}$ para $k = 1$ e $w(k) = \sqrt{2/N}$ para $2 \leq k \leq N$.

2) *Características extraídas do domínio do tempo:* Em [20] foi mostrado que algumas características podem ser facilmente extraídas e calculadas a partir do sinal de ECG filtrado utilizando uma janela de tempo do sinal: energia, média, potência instantânea, máximo, mínimo, diferença entre máximo e mínimo, e índices estatísticos de alta ordem como curtose e medidas de assimetria. Neste trabalho a janela utilizada foi o próprio complexo QRS, da onda P até a onda T.

3) *Função de Autocorrelação:* A função de AC mede o grau de correlação de uma variável, em um dado instante, consigo mesma e em um instante de tempo posterior. Esta permite encontrar padrões repetitivos e irregularidades em um sinal. De acordo com [8] a função de autocorrelação pode ser utilizada como característica não-fiducial do sinal de ECG, utilizando uma janela de tamanho N . A AC pode ser

calculada como: $\hat{R}_{X,X}[m] = \frac{\sum_{i=0}^{N-|m|-1} x[i]x[i+m]}{\hat{R}_{X,X}[0]}$, onde $x[i]$ é a i -ésima amostra da janela definida para o sinal de ECG e $x[i+m]$ é uma amostra posterior também da janela do sinal de ECG com um intervalo de tempo $m = 0, 1, \dots, (M - 1)$ e com a restrição $M \ll N$. Neste trabalho a janela considerada para a autocorrelação corresponde ao complexo QRS extraído.

4) *Transformada de Fourier:* A Transformada de Fourier é muito conhecida e utilizada na análise de reconhecimento de padrões [21], compressão e reconstrução de imagens. Por meio da Transformada de Fourier é possível obter as componentes de frequência de um determinado sinal. Um dos métodos comumente utilizados para cálculo da Transformada Discreta de Fourier é a Transformada Rápida de Fourier. Neste trabalho foram empregados como característica os 10 coeficientes de frequência de maior amplitude de cada complexo QRS.

5) *Transformada Wavelet:* A Transformada Wavelet é uma técnica usada para analisar sinais no domínio do tempo-frequência, ao contrário da Transformada de Fourier que realiza a análise apenas no domínio da frequência. Os coeficientes de WT são definidos como produto interno do sinal com alguma função base [22], sendo esta função base é obtida a partir de uma única função wavelet chamada de Função Mãe.

O cálculo dos coeficientes de Wavelet de um sinal $x(n)$ pode ser realizado usando dois filtros de resposta impulsiva finita, sendo um filtro passa-baixa h e um outro filtro passa-alta g . Isto é conhecido como o Algoritmo Mallat [23]. Ao final a WT gera dois novos sinais, um contendo os detalhes do sinal original (altas frequências) e o outro representa uma aproximação do sinal original (baixas frequências).

6) *Modelos Autoregressivo:* Neste modelo o valor do sinal no instante t é expresso com uma combinação linear dos

valores anteriores e um termo relacionado ao ruído. Os pesos da combinação linear são chamados de coeficiente do modelo AR. De acordo com [24] os coeficientes do modelo AR podem ser utilizados como fontes de características não-fiduciais do sinal de ECG. Seja $x(n)$ uma sequência do sinal de ECG, este pode ser representado como: $\hat{x}(n) = \sum_{i=1}^p \hat{a}_i x(n-i) + \varepsilon(n)$, onde $a_k (k = 1, 2, \dots, p)$ são os coeficientes do modelo AR, $\varepsilon(n)$ é o ruído branco, média zero e desvio padrão σ^2 . Neste trabalho utilizamos uma abordagem semelhante aos trabalhos de [25], no qual foi utilizado o valor de p igual a 2 e 3 sobre cada complexo QRS normalizado.

7) *Dimensão Fractal:* A dimensão fractal representa um índice estatístico de complexidade indicando como os detalhes em um objeto físico muda com a escala na qual são mensurados [26]. FD é geralmente um número fracionário, não inteiro. Neste trabalho foram utilizados os seguintes métodos para a estimação da dimensão fractal: Box Counting, Higuchi, Katz, Sevcik e Hurst. Uma descrição detalhada destes métodos pode ser encontrada em [27].

8) *Modo de Decomposição Empírico:* Modo de Decomposição Empírico (EMD) decompõe um sinal de entrada em um conjunto de funções de modo intrínseco (IMF). As IMFs representam os diferentes modos oscilatórios intrínsecos presentes no sinal de entrada, que podem combinar estruturas ou padrões relacionados a efeitos embutidos no sinal [28].

IV. DESCRIÇÃO DOS CLASSIFICADORES

A. K - Vizinhos mais Próximos

O classificador k -Vizinhos mais Próximos foi proposto em [29] e é amplamente utilizado em várias tarefas de classificação. Um novo exemplo x é atribuído para a classe mais comum entre seus k vizinhos mais próximos medidos por uma função de distância. Neste trabalho a distancia euclidiana foi utilizada como função de distância.

B. Análise de Discriminante Linear

Análise de Discriminante Linear (LDA) busca localizar as direções de projeção nos dados de forma a gerar um conjunto de características com maior poder de separação e com menor quantidade de características. O LDA faz o uso da informação de classe para gerar uma projeção que aumente a separação entre-classes e a similaridade intra-classes [30]. Sobre os dados projetados foi utilizado o kNN.

V. ALGORITMOS GENÉTICOS

Algoritmo Genético [31] trata-se de uma classe de algoritmos que encontram soluções subótimas para problemas combinatórios. Já os Algoritmos Meméticos combinam Algoritmos Genéticos com um procedimento de busca local, inserindo conhecimento a priori no processo de evolução, de forma que as melhores soluções possam ser encontradas usando as informações incorporadas nas soluções atuais. Em [32] foi proposto uma representação binária para tratar a presença ou não de uma característica no problema de

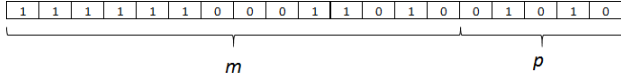


Fig. 2. Codificação utilizada no GA e MA

otimização. Este artigo utilizou a mesma abordagem na codificação da solução para o GA e MA.

Os GAs e os MAs são compostos por uma população de cromossomos (P_{pop}), operadores de seleção, cruzamento, mutação e função de *fitness*. Neste trabalho, a busca local do MA proposta é baseada em um conjunto de trocas sucessivas de 1 até Φ_{it} , onde Φ_{it} é a quantidade de máxima de trocas permitidas. O processo de seleção de características utilizando GA e MA ocorre apenas durante o treinamento do classificador e a evolução da população. Uma vez feito este processo o teste é realizado a partir das características selecionadas. Com um número reduzido de características será possível realizar o reconhecimento num tempo muito pequeno, visto que não será necessário realizar o processo de seleção novamente.

A. Codificação da Solução

Neste trabalho, cada cromossomo da população do GA e MA foi codificado por um vetor C de tamanho $m + p$ bits, onde os m primeiros bits representam a presença ou não de uma característica extraída, m é a quantidade total de característica extraída. Os últimos p bits são utilizados para determinar os parâmetros relacionados com o classificador que será utilizado na função *fitness* como ilustrado na Figura 2.

B. Função Fitness

A função de *fitness* realiza a avaliação de todos os cromossomos presentes na população ao longo da execução do GA e MA. Este processo é importante, pois auxilia na busca pelo máximo global. Neste trabalho, o objetivo do GA e MA é selecionar as melhores características de forma a obter a menor taxa de erro de reconhecimento. Visando solucionar o menor subconjunto de características, foi adicionado um fator de punição P_f que pune os cromossomos de acordo com a quantidade de características codificadas.

Seja T o conjunto total de característica fornecido ao MA para ser utilizado na função de *fitness*. A função de *fitness* pode ser definida de acordo com o Algoritmo 1:

Algoritmo 1 Cálculo da função Fitness

```

for all  $C_i \in P_{pop}$  do
  Extraia as características selecionadas de  $C_i$ , a partir de  $T$  e um subconjunto de características  $T'$ 
  Obtenha o parâmetro  $p$  em  $C_i$ 
  Execute o classificador utilizando  $T'$  e  $p$ 
  Obtenha o erro médio  $e$  utilizando validação cruzada 5-fold
  Obtenha a quantidade de características codificadas de  $C_i$  em  $Qtd_f$ 
  Calcule o fitness  $f_{C_i} = (100 - e)(1 - P_f) + 1 - \frac{Qtd_f}{|C_i|} P_f$ 
end for

```

Podemos observar que a cada avaliação de um indivíduo i da população P_{pop} um novo subconjunto de características

T' é gerada a partir das características selecionadas (bits de $C_i = 1$).

VI. EXPERIMENTOS COMPUTACIONAIS

A. Descrição da Base de Dados

De forma a avaliar o desempenho obtido com a metodologia proposta, foi utilizada uma base de ECG pública, chamada *PTB ECG Database*. Esta base contém o sinal de ECG obtido de 290 pessoas. Os sinais foram obtidos de 209 homens e 81 mulheres com idade entre 17 e 87 anos. O número de gravações por pessoa varia de 2 (coletados em um único dia) a 20 (coletados periodicamente durante 6 meses) [33]. Cada sinal inclui os sinais brutos e filtrados. Apenas os sinais filtrados foram usados nos experimentos realizados neste trabalho.

B. Pré processamento e Parâmetros

A etapa de pré-processamento foi realizada em três fases: filtragem, normalização e segmentação.

Na fase de filtragem diversos componentes de ruído nos sinais de ECG, tais como interferência na linha de alimentação e ruídos de alta frequência, foram filtrados utilizando um filtro passa banda *Butterworth* com frequência de corte de 1 Hz a 40 Hz [9]. Para cada indivíduo, o valor médio do sinal do ECG é subtraído de cada amostra do sinal de ECG e então dividido pelo desvio padrão. Este procedimento resulta em um sinal de ECG normalizado para cada pessoa, com média zero e desvio padrão unitário. Em seguida, os complexos QRS foram detectados com o Algoritmo de Pan-Tompkins.

C. Resultados

Para a avaliação das características extraídas foram realizados dois experimentos. O primeiro experimento foi realizado usando uma busca gulosa sobre a combinação das características extraídas. Neste contexto, as características fiduciais foram divididas em 2 grupos, Amplitudes e Intervalos (AI) e Ângulos e Áreas (AA), as características obtidas a partir da WT foram divididas em 13 grupos de acordo com as funções wavelet: Haar, Symlet, Daubechies e Coiflet de ordem 1 a 4, sendo que foi fixado 4 níveis de decomposição e as outras características formaram um grupo cada uma. Ao final deste processo havia 22 grupos de características. Os grupos foram avaliados individualmente, depois estes grupos foram avaliados dois a dois, em seguida três a três e finalmente quatro a quatro. Os valores de k adotado no classificador k -Vizinhos mais Próximos foram os seguintes: 40, 60, 80 e 100 vizinhos. Cada simulação foi avaliada utilizando validação cruzada com 5 folds. A Tabela II apresenta os resultados dos grupos de características avaliados individualmente. Neste caso, o grupo de característica fiducial AI atingiu a melhor taxa de reconhecimento seguido pela WT usando Coiflet de ordem 4.

O desempenho dos grupos de características, quando avaliados individualmente não produziram resultados satisfatórios. No entanto alguns grupos de características fiduciais

Tabela II
SIMULAÇÕES DE CADA GRUPO DE CARACTERÍSTICAS

Taxa de Acerto	Característica ordem (ord)	k	Classificador	# Carac.
89,7931	AI	100	LDA	19
82,2069	WT coiflet ord 4	100	kNN	33
80,4828	WT coiflet ord 3	100	kNN	33
80	WT symlet ord 4	100	LDA + kNN	33
79,3103	WT symlet ord 3	100	LDA + kNN	33
79,3103	WT daubechies ord 3	100	LDA + kNN	33
78,3448	WT haar	100	LDA + kNN	33
78,3448	WT symlet ord 1	100	LDA + kNN	33
78,3448	WT daubechies ord 1	100	LDA + kNN	33
77,3103	WT coiflet ord1	100	kNN	33
76,6897	AC	100	kNN	7
75,6552	WT coiflet ord 2	100	LDA + kNN	33
74,4138	TD	100	kNN	7
72,7586	WT daubechies ord 4	100	LDA + kNN	33
70,8966	FD	100	LDA + kNN	11
70,8966	WT symlet ord 2	100	LDA + kNN	33
70,8966	WT daubechies ord 2	100	kNN	33
69,3793	AA	100	LDA + kNN	22
58,6897	DCT	80	kNN	7
24,4138	FFT	100	LDA + kNN	15
16,8276	EMD	100	kNN	12
14,2759	AR	40	kNN	5

e não-fiduciais em conjunto produziram uma taxa de reconhecimento superior quando comparadas às características individuais como descrito na Tabela III.

Tabela III
10 MELHORES SIMULAÇÕES - BUSCA GULOSA

Taxa de Acerto	Característica	k	Classificador
94,2069	AI - AC	100	kNN
94,2069	AI - AC	100	LDA + kNN
94,1379	AI - DCT - AC	80	kNN
94,1379	AI - DCT - AC	100	kNN
94,1379	AI - DCT - AC	100	LDA + kNN
94	AI - AC	80	LDA + kNN
93,8621	AI - AC	80	kNN
93,7241	AI - DCT - AC	80	LDA + kNN
93,4483	AI - AC	60	LDA + kNN

Ao analisarmos o desempenho em grupos de 2 conjuntos de características, a combinação AI - AC produziu melhores resultados atingindo uma taxa de acerto de 94,2069, quando avaliamos o desempenho dos grupos de 3 conjuntos de características a combinação AI - DCT - AC produziu o melhor resultado com uma taxa de acerto de 94,1379. Nenhum grupo composto por 4 conjuntos de características produziu um desempenho que chegasse a ser uma 10 melhores nas simulações. Neste cenário o grupo que alcançou melhor desempenho foi à combinação AI - TD - DCT - AC que produziu uma taxa de acerto igual a 91,2414, isto sugere que a adição aleatória de características submetidas ao classificador pode prejudicar o desempenho do mesmo.

O segundo experimento foi realizado usando GA e MA como estratégia de seleção de característica. Para cada grupo de característica foram realizadas 20 simulações, 10 para o GA com classificadores kNN e LDA e outras 10 para o MA utilizando os mesmos classificadores. Para realização

destas simulações foram adotados os seguintes parâmetros: P_{pop} igual a 50, Φ_{it} igual a 3, P_f igual a 0,30 utilizando o operador de cruzamento uniforme e roleta viciada para seleção dos indivíduos para reprodução. O critério de parada dos GAs e MAs foi definido em função do número de gerações (neste caso foi adotado valor igual a 100). Para estas simulações 70% do sinal de ECG de cada individuo foi utilizado durante a evolução do MA e 30% do sinal foi utilizado para teste após a execução do MA. Validação cruzada 5-fold foi aplicada sobre 70% dos dados. Neste caso, 4 folds foram usados para treinamento e a função de *fitness* foi avaliada sobre o fold restante. Este processo é repetido 5 vezes e o *fitness* médio foi usado como medida de desempenho de cada grupo de característica. O melhor grupo de característica foi avaliado sobre o conjunto de teste.

Tabela IV
MÉDIA DOS RESULTADOS DA EXECUÇÕES DO MA

	Treinamento	Teste	Qtd. Características	Fitness
MA-kNN	96,3348%	94,8371%	121	90,206%
MA-LDA	95,652%	94,3276%	110	89,826%
GA-kNN	96,4138%	93,7931%	150	89,1331%
GA-LDA	96,1379%	93,4483%	147	89,6776%

A Tabela IV apresenta os resultados obtidos sobre o conjunto de teste usando GA e MA. As simulações que utilizaram o kNN como classificador atingiram uma melhor taxa de reconhecimento. No entanto, os experimentos envolvendo GA e MA atingiram uma taxa de reconhecimento muito próxima para o conjunto de teste, sendo que as simulações com kNN atingiram uma taxa de reconhecimento igual a 94,8371% com 110 características. Nas simulações envolvendo GA e MA algumas características não-fiduciais, tais como AR e a EMD, estavam presentes em menos de 10% da população. As características que se destacaram, estando presentes em pelo menos 4 simulações são respectivamente AI, AC, DCT, WT com função wavelet Daubechies de ordem 3 e 4, Symlet de ordem 1 e 4 e Coiflet de ordem 3.

VII. CONCLUSÕES

Este trabalho analisou o impacto de algumas estratégias de seleção de característica tais como busca gulosa, algoritmo genético e algoritmo memético sobre desempenho dos Sistemas Biométricos baseados em ECG ao empregar k -Vizinhos mais Próximos e Análise de Discriminante Linear como classificador.

Analisando os resultados obtidos, pode se observar que algumas características extraídas do sinal de ECG para identificação de usuários consegue fornecer bons resultados. Na simulações realizadas foi obtido uma taxa de reconhecimento superior a 92%. Os melhores resultados foram obtidos usando MA, neste caso, foi alcançada uma taxa de 94.32% de reconhecimento utilizando a base de dados PTB ECG Database. Resultado satisfatório tratando-se de um experimento utilizando todos os indivíduos deste banco de dados. Além disso, é possível notar que algumas funções mães, tais

como Coiflet, apresentaram resultados bastante interessantes. Para todos os cenários a combinação de características fiduciais e não-fiduciais produziu melhores resultados.

Ao compararmos estudos publicados com o mesmo banco de dados e características biométricas selecionadas pelo MA, observamos que a taxa de acerto obtida foi satisfatória, visto que nos trabalhos de [8] por exemplo foi alcançada uma taxa de acerto de 100% utilizando apenas 14 indivíduos e características de DTC e AC. Já neste artigo foi alcançada uma taxa de acerto um pouco inferior como citada anteriormente, mas utilizando 290 indivíduos deste mesmo banco dados. No trabalho de [34] foi alcançada uma taxa de acerto de 96,00% utilizando 25 indivíduos com características extraídas via EMD, com metodologia semelhante a utilizada neste artigo. A taxa de acerto obtida pelo MA também é interessante ao compararmos resultados de estudos que utilizaram bancos de dados privado. Por exemplo, [17] alcançou uma taxa de acerto de 96.44% com 100 e utilizando características de AI.

Como trabalhos futuros, serão realizados novos experimentos utilizando outras bases de dados, além do emprego de novas estratégias de extração de características e outros classificadores tais como Máquinas de Vetores Suporte.

REFERÊNCIAS

- [1] R. R. O. e. J. S. F. Luciano R. Costa, "Introdução á biometria," in *VI Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais*, Santos, SP, 2004, pp. 103–151.
- [2] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbach, "Ecg biometric recognition: A comparative analysis," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1812–1824, 2012.
- [3] A. Riera, S. Dunne, I. Cester, and G. Ruffini, "Starfast: a wire-less wearable eeg/ecg biometric system based on the enobio sensor," in *Proceedings of the International Workshop on Wearable Micro and Nanosystems for Personalised Health*, 2008.
- [4] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "Ecg analysis: a new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, 2001.
- [5] J. M. Irvine and S. A. Israel, "A sequential procedure for individual identity verification using ecg," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 3:1–3:13, Jan. 2009.
- [6] M. Tantawi, A. Salem, and M. Tolba, "Ecg signals analysis for biometric recognition," in *Hybrid Intelligent Systems (HIS), 2014 14th International Conference on*, Dec 2014, pp. 169–175.
- [7] A. D. C. Chan, M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 2, pp. 248–253, 2008.
- [8] K. Plataniotis, D. Hatzinakos, and J. Lee, "Ecg biometric recognition without fiducial detection," in *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*, Sept 2006, pp. 1–6.
- [9] Y. Wang, F. Agraftioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP J. Adv. Signal Process.*, vol. 2008, Jan. 2008.
- [10] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989.
- [11] M.-H. Lim, Y. Yuan, and S. Omatu, "Extensive testing of a hybrid genetic algorithm for solving quadratic assignment problems," *Computational Optimization and Applications*, vol. 23, no. 1, pp. 47–64, 2002.
- [12] Y. Ong and A. Keane, "A domain knowledge based search advisor for design problem solving environments," *Engineering Applications of Artificial Intelligence*, vol. 15, no. 1, pp. 105 – 116, 2002.
- [13] Y.-S. Ong and A. Keane, "Meta-lamarckian learning in memetic algorithms," *Evolutionary Computation, IEEE Transactions on*, vol. 8, no. 2, pp. 99–110, April 2004.
- [14] Y.-S. Ong, M.-H. Lim, N. Zhu, and K.-W. Wong, "Classification of adaptive memetic algorithms: a comparative study," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 36, no. 1, pp. 141–152, Feb 2006.
- [15] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.
- [16] M. B. Messaoud, B. Khelil, and A. Kachouri, "Analysis and parameter extraction of p wave using correlation method," 2007.
- [17] T. Hamdi, A. Ben Slimane, and A. Ben Khalifa, "A novel feature extraction method in ecg biometrics," in *Image Processing, Applications and Systems Conference (IPAS), 2014 First International*, Nov 2014, pp. 1–5.
- [18] M. R. M. Isa and S. Aljareh, "Biometric image protection based on discrete cosine transform watermarking technique," in *Engineering and Technology (ICET), 2012 International Conference on*, Oct 2012, pp. 1–5.
- [19] M. S. Al-ani and W. M. Al-aloosi, "Biometrics fingerprint recognition using discrete cosine transform (dct)," *International Journal of Computer Applications*, vol. 69, no. 6, pp. 44–48, May 2013.
- [20] S. Ergin, A. Uysal, E. Gunal, S. Gunal, and M. Gulmezoglu, "Ecg based biometric authentication using ensemble of features," in *Information Systems and Technologies (CISTI), 2014 9th Iberian Conference on*, June 2014, pp. 1–6.
- [21] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2001.
- [22] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, Jul 1989.
- [23] ———, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Academic Press, 2008, ch. 7, pp. 410–418.
- [24] B. Vuksanovic and M. Alhamdi, "Ecg based system for arrhythmia detection and patient identification," in *Information Technology Interfaces (ITI), Proceedings of the ITI 2013 35th International Conference on*, June 2013, pp. 315–320.
- [25] ———, "Analysis of human electrocardiogram for biometric recognition using analytic and ar modeling extracted parameters," November 2014, pp. 428–433.
- [26] M. Barnsley, *Fractals Everywhere*. San Diego, CA, USA: Academic Press Professional, Inc., 1988.
- [27] C. Lima, A. Coelho, R. Madeo, and S. Peres, "Classification of electromyography signals using relevance vector machines and fractal dimension," *Neural Computing and Applications*, pp. 1–14, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s00521-015-1953-5>
- [28] G. Rilling, P. Flandrin, and P. Gonzi, $\frac{1}{2}$ alv $\frac{1}{2}$ s, "On empirical mode decomposition and its algorithms," 2003.
- [29] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theor.*, vol. 13, no. 1, pp. 21–27, Sep. 2006. [Online]. Available: <http://dx.doi.org/10.1109/TIT.1967.1053964>
- [30] O. Kursun and E. Alpaydin, "Canonical correlation analysis for multiview semisupervised feature extraction," in *Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, L. Rutkowski, R. Scherer, R. Tadeusiewicz, Z. Zadeh, and J. Zurada, Eds. Springer Berlin Heidelberg, 2010, vol. 6113, pp. 430–436.
- [31] P. Moscato, "On evolution, search, optimization, genetic algorithms and martial arts - towards memetic algorithms," 1989.
- [32] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, "Feature selection for svms," in *Advances in Neural Information Processing Systems 13*, Max-Planck-Gesellschaft. Cambridge, MA, USA: MIT Press, 2001.
- [33] A. Goldberger, L. Amaral, and I. P. M. R. M. J. M. G. P. C.-K. S. H. Glass, L. Hausdorff JM, "A real-time qrs detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.
- [34] Z. Zhao, L. Yang, D. Chen, and Y. Luo, "A human ecg identification system based on ensemble empirical mode decomposition," *Sensors*, vol. 13, no. 5, pp. 6832–6864, 2013. [Online]. Available: <http://www.mdpi.com/1424-8220/13/5/6832>