

**Título:** Predição da gravidade de lesões em atletas via Programação Genética

**Autor:** Thiago Tavares Magalhães (estudante de graduação FAETERJ)

**Orientador:** Eduardo Krempser da Silva (professor FAETERJ, estudante de doutorado do LNCC)

**Instit. de Vínculo:** CPTI - Instituto Superior de Tecnologia em Ciências da Computação de Petrópolis (FAETERJ) / Endereço: Av. Getúlio Vargas, 335 Bairro Quitandinha, Petrópolis, RJ  
CEP: 25651-075 / Telefone: (24)2235-1079 / E-mail: cepti@faetec.rj.gov.br

**Instit. Onde se desenvolveu a Pesquisa:** Laboratório Nacional de Computação Científica (LNCC) / Endereço: Av. Getúlio Vargas, 333 Bairro Quitandinha, Petrópolis, RJ  
CEP: 25651-075 / Telefone: (24) 2233-6000 / E-mail: krempser@lncc.br

## **PREDIÇÃO DA GRAVIDADE DE LESÕES EM ATLETAS VIA PROGRAMAÇÃO GENÉTICA**

### **1. INTRODUÇÃO**

Atualmente é evidente que a prática esportiva, seja por questões de lazer, de manutenção de condicionamento físico ou de saúde é interesse comum a um número cada vez mais expressivo e diverso de pessoas por todo o mundo (Fong, 2007). Essa crescente democratização do esporte é resultado, em grande parte, do interesse advindo dos meios de comunicação e da mídia de massa em geral em torná-lo um vigoroso braço da indústria do entretenimento (Timpka *et al.*, 2006). O fato real e visível é que se hoje o esporte perdeu muito do seu ideal interligado ao naturalismo, à valorização do homem e ao lazer, característico dos séculos XVIII e XIX para dar lugar a sentidos especialmente econômicos (Betti, 1997), hoje o desporto influencia uma grande parcela da humanidade, não apenas geograficamente, mas atingindo direta ou indiretamente as mais distintas idades e culturas ao redor

do mundo, inclusive alcançando o patamar de ferramenta eficiente e fiel para análises históricas, sociais, políticas, culturais e econômicas em geral.

Há ainda o fato de que, com o advento da literatura na área, a prática esportiva passa a ser cada vez mais vista como um fator indispensável à qualidade de vida por estar associada a um número expressivo de benefícios, tanto físicos quanto psicológicos, não apenas para a população em geral, mas até mesmo para diversos grupos específicos. Há estudos, por exemplo, que demonstram que um programa de exercícios regulares é eficaz em otimizar o perfil lipídico de pacientes com diabetes tipo 1, independentemente do controle glicêmico (Khawali *et al.*, 2003); que confirmam que a atividade física também é tida como um agente capaz de minimizar o sofrimento psíquico de idosos deprimidos e oferecer-lhes a possibilidade de envolvimento psicossocial, elevação da autoestima e implementação das funções cognitivas, culminando possivelmente na saída do quadro depressivo (Stella *et al.*, 2002) e há também os estudos que mostram que as vantagens da atividade física adequada e supervisionada são perceptíveis inclusive em gestantes, independentemente das divergências quanto ao período de início ideal para sua prática (Skinner, 1991).

Entretanto, a despeito de todos esses benefícios citados, é importante ressaltar a íntima relação existente entre o esporte e a possibilidade de lesões, sendo essas presentes tanto a níveis competitivos quanto a níveis recreativos (Olsen *et al.*, 2005; Cohen e Abdalla, 2002). Para ser mais exato, o desporto é, atualmente, uma das principais causas de lesões, comparando-se inclusive com acidentes domésticos, acidentes de viação ou com a violência (Fong, 2007).

À medida que se fala da prática esportiva de alto-rendimento, no entanto, as lesões são ainda mais comuns e significativas. No meio profissional, a elevada exigência física, bem como a repetição constante do gesto técnico e o desgaste de uma atividade intensiva aumentam a probabilidade da incidência de lesões (Atalaia *et al.*, 2008) e essas estão ainda interligadas a fatores exteriores à saúde e ao atleta acometido. Se para os praticantes amadores, uma lesão pode apenas comprometer em parte suas

tarefas diárias (Atalaia *et al.*, 2008), no caso de atletas de elite as lesões apresentam tratamento difícil, consomem tempo, motivação, recursos econômicos, desagradam a patrocinadores e torcedores, podendo ainda desvalorizar eventos dos mais diversos âmbitos e/ou prejudicar (e por vezes findar) a carreira dos próprios atletas (Kujala *et al.*, 2003).

Nesse contexto em que o mero surgimento da lesão acompanha tantos problemas, torna-se visível que os aspectos preventivos devem ser encarados como prioridades pelos profissionais da área esportiva (Fontana, 1999), já que a despeito dos tratamentos existentes, os prejuízos decorrentes de uma lesão são sempre presentes apesar de indesejados, podendo, em alguns casos, ser irreparáveis e altamente comprometedores em vários âmbitos.

O entendimento dos fatores que levam aos danos tem despertado, portanto, um grande e crescente interesse de pesquisadores em biomecânica do esporte, como Valiant (Valiant e Cavanagh, 1985), Stacoff (Stacoff *et al.*, 1988), Pastre (Pastre *et al.*, 2007), entre muitos outros. Isso se deve, especialmente, ao fato de que uma fonte estrategicamente importante para os tão indispensáveis meios de prevenção de lesões é perceber quais e a que nível fatores distintos de risco podem influenciar na ocorrência e/ou na gravidade das avarias. (Vanderlei, 2011). Esses fatores são muitos, podendo dividir-se ainda em diversos grupos (Caine *et al.*, 2008), e não existe um consenso acerca de quais influenciam de fato nas ocorrências das lesões, o que dificulta os diagnósticos, a prevenção, a dosagem de variáveis como níveis de carga e horas de treino e as recomendações da saúde pública para atividades físicas e esportivas a níveis amadores.

Em meio a todo esse contexto mencionado surgem trabalhos de diversos pesquisadores como Marques (Vanderlei, 2011), Arena (Arena e Carazzato, 2007), Guedes (Guedes, 2010) e muitos outros com o intuito de coletar informações de entrada sobre atletas ou situações de treinamento para, valendo-se de análises matemáticas, estatísticas descritivas e comparativas, descobrirem suas relações com informações acerca do diagnóstico das mais diversas lesões desportivas.

Esclarecidas a relevância e a dificuldade do tema, cada vez mais são importantes as propostas científicas que visem à análise classificatória de dados provenientes do desporto. No entanto, à mesma medida em que o esporte se torna cada vez mais interessante à sociedade em geral, mais se destaca a dificuldade da obtenção de análises relacionadas com esse assunto que alcancem além das linhas da estatística e da matemática descritiva.

## **2. OBJETIVOS**

Dado esse complexo cenário abordado, a proposta do presente trabalho é demonstrar a relevante descoberta das relações existentes entre distintos fatores, extrínsecos e intrínsecos, que podem ser determinantes na ocorrência das lesões, não estatisticamente, como já feito antes, mas utilizando-se uma técnica importante de inteligência computacional baseada na teoria da evolução de Darwin denominada Programação Genética.

Acredita-se que a partir do sucesso na determinação de qualquer variável específica (como o nível de gravidade de lesões, por exemplo), será possível demonstrar a eficiência da técnica para inúmeras variáveis (como a possibilidade do surgimento de lesões), mesmo que a partir de entradas de dados diferentes ou mais detalhistas. Vista a imensa adaptabilidade da programação genética a diferentes situações, a mesma técnica é aplicável, ainda, a todo o tipo de desportista, seja amador ou de elite, seja recreativo ou profissional.

No decorrer dos próximos capítulos, abordaremos os materiais utilizados no presente trabalho, desde o algoritmo evolucionista escolhido até a base de dados obtida. Explicaremos, também, o processo prático de aplicação da técnica e analisaremos, finalmente, os diversos resultados obtidos, com o intuito de confirmar a possibilidade do uso eficaz da programação genética como importante ferramenta auxiliar no tão desejado e necessário reconhecimento dos fatores específicos que levam às lesões desportivas em atletas dos mais distintos níveis esportivos e objetivando, ainda, a discussão acerca das relações obtidas entre diversos fatores variáveis e a integridade física dos atletas.

### **3. MATERIAL E MÉTODOS**

Neste capítulo serão descritos os métodos empregados durante a realização da pesquisa e os materiais e as fontes de informação utilizados. Passaremos, portanto, por uma explicação da técnica de maneira geral, por esclarecimentos referentes ao algoritmo específico empregado, citaremos a base de dados utilizada e justificaremos a sua escolha e findaremos o capítulo abordando partes relevantes do processo prático de aplicação da programação genética à mencionada base de dados.

#### **3.1 A Programação Genética**

Para uma abordagem satisfatória e didática dessa importante técnica de programação são necessárias algumas considerações. Assim, no decorrer desta seção seguem-se inicialmente um breve comentário acerca da Teoria da evolução por seleção natural de Darwin, uma abordagem geral da Programação Genética e explicações aprofundadas acerca de cada uma das etapas que a compõem.

##### **3.1.1 A Teoria Darwiniana**

A principal base da programação genética, a Teoria da evolução das espécies por meio da seleção natural foi proposta por Charles Darwin a partir do livro "*A origem das espécies através da seleção natural*", publicado em Londres no ano de 1859.

Filho do médico Robert Darwin, Charles Darwin decidiu não seguir a mesma carreira do pai, mas tornar-se teólogo e pastor anglicano. Após se matricular no bacharelado em Artes, conforme exigido (Pinheiro, 2009), Darwin se interessou em alto grau pelos livros de teologia do pastor William Paley, que apresentava como prova da existência de Deus a complexidade dos seres vivos. Influenciado pelos livros de Paley, Darwin se tornou um estudioso da Biologia e, tempos depois, abandonou a Teologia (Pinheiro, 2009) e se tornou um aplicado naturalista, estudioso de zoologia, botânica e geologia.

Finalmente, inspirado por um conjunto de viagens pelo globo terrestre que tiveram como objetivo estudar a relação das espécies com o meio-

ambiente, Darwin concluiu que a competição comumente encontrada na natureza é uma das responsáveis diretas por selecionar organismos mais adaptados que outros em meio à população total existente que disputa determinados recursos.

Segundo a teoria proposta por Darwin, em um ambiente competitivo em que a fonte predominante de alimentação herbívora seria representada pelas copas das árvores mais altas, por exemplo, as girafas com pescoços longos teriam mais chances de conseguir acesso à alimentação abundante, sobreviver e, finalmente reproduzir, perpetuando assim seus genes e contribuindo para a disseminação das características que foram responsáveis pela sobrevivência no seu habitat, nesse caso transmitindo a característica de pescoços longos.

Assim, com o passar do tempo a seleção natural tenderia a criar indivíduos cada vez mais adaptados aos seus ambientes, sendo as gerações mais recentes probabilisticamente mais adaptadas que as gerações passadas de genitores. Segundo Darwin, a existência do homem e de todas as demais formas de vidas encontradas nos dias atuais é resultado de uma série de acidentes que ocorreram no curso da existência da vida na Terra e que foram responsáveis por, aleatoriamente, oferecer atributos cada vez mais úteis à sobrevivência de cada um dos seres vivos que formam a gigantesca teia evolutiva que prossegue em crescente ampliação.

A teoria evolucionista de Darwin, especialmente fortalecida a partir dos resultados de estudos com ervilhas relativos à genética dos seres vivos divulgados pelo monge agostiniano Gregory Mendel em 1868, alcançou a ampla aceitação e atualmente passa a exercer influência em vários âmbitos científicos e sociais, desde a psicologia e a sociologia até a biologia e a programação de computadores. A partir de suas ideias centrais de seleção natural e adaptação ao longo de gerações desenvolveu-se a técnica utilizada no presente trabalho, denominada programação genética.

### **3.1.2 Definições gerais e aplicações**

A programação genética, cujo desenvolvimento é atribuído a Koza (1992) e à qual nos referiremos apenas por PG, é uma meta-heurística

estocástica de otimização global baseada no aprendizado supervisionado (*supervised learning*) e no princípio darwiniano da seleção natural, sendo, portanto, uma abordagem da computação evolucionária.

De fato a PG é uma técnica bastante ambiciosa por ser, dentre inúmeros fatores, a implementação plena da sonhada capacidade de programar dizendo ao computador, frente à tarefa alvo, apenas o resultado final que se quer alcançar, sem, no entanto, especificar detalhadamente cada um dos passos necessários para que a máquina atinja seus objetivos (Augusto, 2009), sonho que fortemente teve como maior barreira a natureza altamente determinística das pesquisas científicas, que geralmente obedecem a sete princípios comuns: o uso de abordagens corretas, consistentes, justificáveis, determinísticas, ordenáveis, parcimoniosas e decisivas (Koza, 2000). A relação da PG com cada um desses fatores é explicitada a seguir:

- Correto: na ciência geralmente é objetivada a solução exata para o problema proposto, algo que não é aplicável à PG. É comum, entretanto, a percepção de muitas situações em que a utilidade prática real de um resultado não está ligada necessariamente a uma precisão exata, mas à uma proximidade satisfatória do exato (Carneiro, 2009). Como exemplos, podemos citar as pequenas imprecisões em máquinas de cálculo e as aproximações em equações da engenharia e análises de equações matemáticas;
- Consistente: ao contrário do que se espera na ciência convencional, a PG utiliza abordagens inconsistentes e/ou contraditórias como partes da solução dos problemas;
- Justificável: este é o princípio também conhecido como “motivação lógica”. Apesar desse princípio a PG não produz soluções a partir de um fluxo lógico de raciocínio;
- Determinístico: a PG é predominantemente probabilística e não determinística;
- Ordenamento: ao contrário do que se observa nos métodos científicos tradicionais, a desordem é um fator altamente presente na natureza. Semelhantemente, a PG também não está comprometida com a ordem;

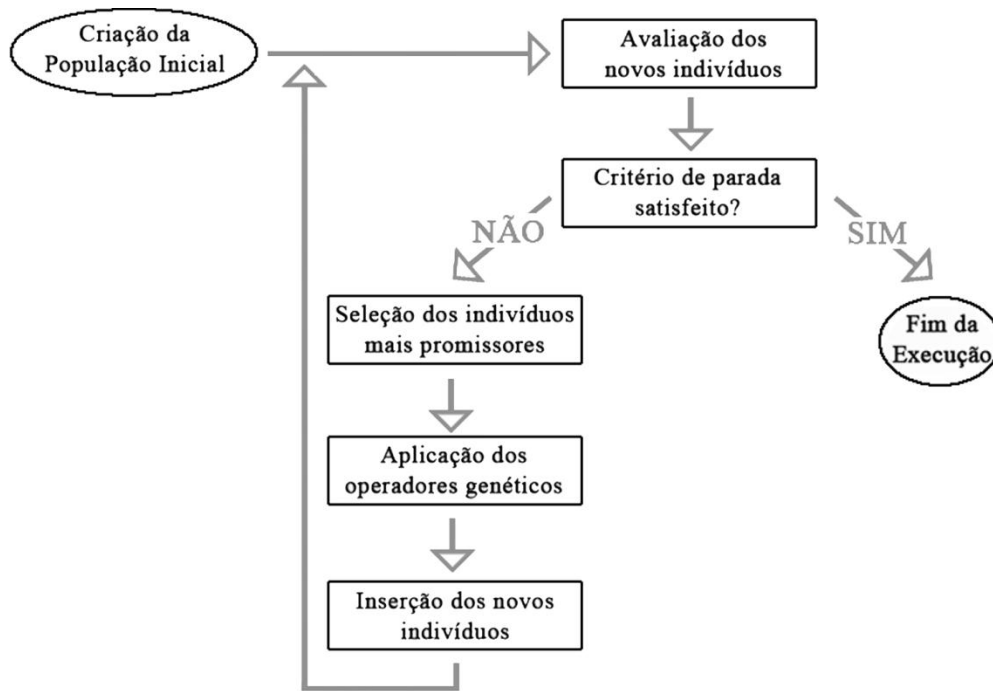
- Parcimônia: a PG não está necessariamente comprometida com soluções simples;
- Decisivo: os processos biológicos naturais não apresentam a definição clara de um critério de fim. A PG segue, portanto, esse mesmo princípio.

Como demonstrado, a técnica não segue nenhum dos sete quesitos tradicionais descritos, já que a evolução em seu estado natural também não está comprometida com nenhum deles. Tanto para essa meta-heurística quanto para a natureza, o importante é a aptidão do indivíduo (Carneiro, 2009), que no caso da PG é definida de acordo com a proximidade da solução ao resultado esperado, através de uma função de avaliação.

Ao ignorar esses princípios tradicionais, a técnica consiste na construção automática de programas de computador que, otimizando ao longo de várias gerações estruturas computacionais capazes de realizar operações lógicas, aritméticas, condicionais e de desvios, são capazes de, a partir de um conjunto de entradas, determinar o conjunto correspondente e correto de saídas. Primeiramente gera-se uma população inicial contendo um número determinado de programas (indivíduos) candidatos criados aleatoriamente. A partir daí surge um processo contínuo e repetitivo que é responsável por selecionar os indivíduos mais promissores (que são capazes de chegar o mais próximo possível do resultado esperado) dentro da população, efetuar a troca de material genético entre eles, através de operadores genéticos como os diferentes tipos de cruzamentos e mutações, e finalmente inserir os seus descendentes dentro da população dando início à próxima geração e reiniciando o processo, não mais a partir da população inicial, porém partindo da geração mais recente. Ao longo do processo, conforme ocorre o surgimento automático de mais gerações, probabilisticamente a PG tende a produzir soluções cada vez mais adaptadas, ou seja, cada vez mais próximas de atingir os objetivos exatos almejados.

As etapas que, basicamente, compõem a estrutura de funcionamento da técnica estão representadas na figura 1 e serão brevemente abordadas abaixo, sendo, porém, discutidas mais profundamente no decorrer deste capítulo:





**Figura 1: Fluxograma das etapas da PG**

- Criação da População Inicial: cria de forma basicamente aleatória os primeiros indivíduos candidatos (população inicial) sobre os quais atuará o processo evolutivo. Para isso, devem estar definidas a representação dos programas (estrutura funcional que será alvo da atuação do processo evolutivo) e os conjuntos primitivos de funções e terminais (instruções de programação básica que podem ser utilizadas para o alcance de uma solução);
- Avaliação dos novos Indivíduos: avaliação da aptidão dos indivíduos através da função de avaliação, que é a função responsável por medir a qualidade com a qual cada indivíduo executa a tarefa do problema;
- Critério de parada: ocorre um teste que verifica se a condição de parada pré-estabelecida foi satisfeita. Normalmente o critério de parada é um número específico máximo de gerações, um tempo máximo de execução ou a obtenção de um resultado satisfatório;
- Seleção dos indivíduos mais promissores: faz-se a seleção dos indivíduos mais adaptados em meio à população total a partir de esquemas de seleção que definem como esses indivíduos são favorecidos na disputa pela procriação;

- Aplicação dos operadores genéticos: criam-se novos indivíduos com base no material genético de seus genitores. Para isso são utilizados, probabilisticamente, dois tipos de operadores genéticos: cruzamento e mutação;
- Inserção dos novos indivíduos: os descendentes gerados a partir da etapa anterior são inseridos na população.

Estruturada dessa forma, a PG mostra-se perfeitamente aplicável à situação das lesões esportivas, por ser baseada no aprendizado supervisionado, definido como o exercício de encontrar saídas satisfatórias a partir de um conjunto de entradas que podem influenciar uma ou mais saídas (Augusto, 2009), e por uma de suas características fundamentais: a pouca exigência de conhecimento sobre o domínio, já que ao programador não é requerido um alto conhecimento sobre o assunto que será alvo de aplicação da PG, uma vez que a função de classificação que, por fim, encontrará os resultados corretos, será encontrada automaticamente por meio da inteligência computacional e não através de conhecimentos específicos do profissional que utilizou a técnica.

Essa característica da pouca exigência de conhecimento sobre o domínio é essencial quando nos referimos às lesões esportivas, devido à dificuldade existente até mesmo entre os especialistas em determinar quais fatores e a que nível esses fatores influenciam de fato na ocorrência e na gravidade dos danos decorrentes da prática do desporto, conforme já mencionado. A PG oferece, nesse sentido, uma boa alternativa tanto para a classificação dos dados quanto para a descoberta dos níveis de influência das variáveis específicas no resultado final em qualquer aplicação em que essa percepção gere dificuldade e desacordo entre os estudiosos da área.

Entretanto é importante ressaltar ainda outras características positivas típicas da programação genética: a técnica é robusta, já que o modelo populacional e o processo estocástico existentes por trás da programação são responsáveis pela solidificação e pela tolerância a ruídos; as soluções obtidas a partir da PG são imediatamente legíveis e interpretáveis (“código-fonte disponível”), devido à atuação direta do processo evolutivo no nível simbólico

das estruturas; a PG possui paralelismo natural, o que significa dizer que as altas demandas computacionais podem ser facilmente particionados em vários níveis, possibilitando assim a redução no tempo de execução e/ou a execução sobre problemas muito mais complexos; a técnica é, ainda, facilmente extensível e modificável por ser uma meta-heurística extremamente versátil, admitindo hibridizações (funcionamento junto a outras técnicas), interação de diversos modelos evolucionistas e as mais diversas representações e linguagens para os programas finais.

Por tudo isso, para que a PG seja aplicável e possa ser candidata a encontrar, por si mesma, a solução para um problema, esse deve atender a apenas duas condições básicas (Augusto, 2009):

- Que a solução possa ser expressa por um programa de computador em uma dada linguagem;
- Que seja possível definir entre duas soluções candidatas qual delas é que melhor se aproxima do objetivo proposto (“função de avaliação”).

Sendo assim, a técnica possui um leque muito abrangente de usabilidade, tendo sido aplicada e bem sucedida em diversas situações. Temos por exemplos:

- Reconhecimento de Padrões: procura-se a evolução de programas capazes de identificar dentro de complexas massas de dados a existência de padrões específicos. A aplicação no diagnóstico de lesões esportivas é um bom exemplo deste caso. Entre outros, temos o diagnóstico de doenças com base em exames médicos, como a classificação de um tumor em maligno ou benigno (Augusto, 2004), por exemplo; o reconhecimento de caracteres ópticos e digitais para diversos fins e a análise de certas possibilidades com base em fatores históricos;
- Regressão Simbólica: Descoberta de expressões matemáticas. Aplica-se em ajustes de curvas, indução de sequências, derivação e integração simbólica, descoberta de identidades matemáticas, previsão de séries temporais, entre outras (Bernardino e Barbosa, 2009);

- Robótica: Desenvolvimento de programas que controlem precisamente ações de robôs em diversas tarefas, como automação industrial, competições, entre outras (Maia e Bianchi, 2000);
- Simulação de Estratégias Inteligentes: Evoluem-se estratégias para jogos, por exemplo, que tendem a decidir dado um conjunto extenso de variáveis qual é a melhor jogada para um momento específico. Também é aplicável a quaisquer tipos de dispositivos que necessitem tomar decisões tendo em conta fatores variáveis. Como exemplo de aplicabilidade da técnica há, ainda, os variados tipos de pilotos-automáticos que necessitam de algum tipo de inteligência computacional para desempenharem suas tarefas (Thorpe *et al.*, 1991);
- Processamento de Imagens e Áudios: Envolve a compactação de imagens e/ou áudios, ou seja, a descoberta de programas capazes de recriar a imagem ou o áudio original da forma mais compacta possível;
- Arte: Evolução de programas responsáveis por codificar artificialmente pinturas, por exemplo.

Sendo, portanto, uma técnica comprovadamente eficaz e de fato muito útil em vários âmbitos, devido às suas características principais e à implementação artificial de conceitos da genética e da evolução por seleção natural, a PG é aplicável a praticamente quaisquer problemas que requerem uma solução, sem, no entanto, deixarem minimamente claro um caminho para que a solução esperada seja encontrada.

### **3.1.3 Representação**

Inicialmente, merece apreço realçar a variedade de estruturas capazes de representar o algoritmo de um programa de computador construído a partir da PG. Há três principais e mais usuais formas de representação da estrutura, normalmente chamada de código genético ou cromossomo, de cada um desses programas, normalmente chamados de indivíduos: linear (Oltean *et al.*, 2009), por árvore (Koza, 1992) ou por grafos (Poli, 1999; Teller, 1996; Miller e Smith, 2006).

Dentre todas essas, a representação por árvores é a mais popular e tradicionalmente a mais utilizada na PG, possivelmente devido à facilidade na

expressão de programas e à facilidade na leitura dos respectivos códigos genéticos, sendo ainda, uma representação naturalmente hierárquica. Essa representação, no entanto, é meramente conceitual, diferindo da representação computacional, em que as árvores podem ser demonstradas de maneira linear sem perderem suas propriedades (Pelikan *et al.*, 1997; Augusto e Barbosa, 2000).

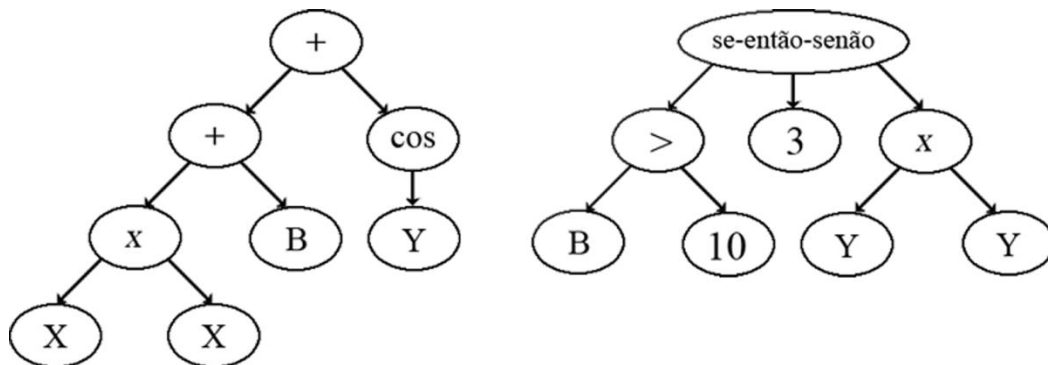
Para a explicação de conceitos específicos da técnica, no presente trabalho os indivíduos candidatos têm o cromossomo demonstrado, portanto, a partir de árvores genéricas formadas por conexões, nomeadas nós, contendo funções ou terminais e apresentando tamanhos e complexidades diferentes, visto que a massa de indivíduos é necessariamente disforme para que possa ocorrer o processo evolutivo.

Para explicitar melhor o conceito de árvores em PG, seguem-se as representações de dois indivíduos. O primeiro apresenta a seguinte equação:

$$x^2 + b + \cos(y)$$

O segundo representa a seguinte expressão lógica:

*Se  $b > 10$  então 3, senão  $y^2$*



**Figura 2: Representação de cromossomos por árvores**

Pode-se perceber através da figura 2 as propriedades hierárquicas e recursivas da representação através de árvores. A árvore à esquerda demonstra o indivíduo cujo cromossomo é a equação citada anteriormente e tem como raiz o nó que contém a função de soma. Já a árvore à direita

representa o indivíduo cujo cromossomo é a expressão lógica, também citada, tendo como sua raiz a função decisiva de teste *se-então-senão*.

Essas estruturas são capazes, enfim, de permitir a visualização simples de qualquer algoritmo de programação obtido a partir da programação genética, conseqüentemente sendo muito favoráveis para explicitar vários âmbitos e tópicos da técnica.

Os conceitos de funções e terminais relacionados com as representações por árvores serão abordados a seguir.

### **3.1.4 Conjuntos primitivos de funções e terminais**

Uma vez definida o tipo de representação utilizada, deve-se, antes da criação da população inicial, definir-se o conjunto de funções e terminais que podem ser explorados pelo processo evolutivo.

Como visto anteriormente, a PG é utilizada com sucesso em problemas com as mais diversas características e das mais distintas áreas de conhecimento. Desse modo, a técnica pode, potencialmente, alcançar a solução desde problemas na área de regressão, que envolvem equações matemáticas para representar determinados tipos de curvas até problemas classificatórios, como os que requerem o reconhecimento de padrões em uma massa determinada de dados distintos.

A resposta encontrada pela inteligência computacional deve, claramente, apresentar características correspondentes ao tipo de problema abordado. Para um problema determinantemente lógico, por exemplo, faz sentido que a solução contenha operadores lógicos como *e*, *ou*, *negação*, *ou exclusivo* e ainda *verdadeiro* ou *falso*, por exemplo. Para um problema puramente matemático e regressivo, por sua vez, faz-se necessária a inclusão de operadores aritméticos como *multiplicação*, *divisão*, *subtração*, *adição* e *resto*.

Por esse motivo, a PG possui dois conjuntos fundamentais distintos: o conjunto de terminais e o conjunto de funções. São esses dois grupos que devem ser definidos com base na natureza do problema e que determinam que ferramentas a PG poderá explorar dentro de todas as possibilidades existentes,

ou seja, definem o espaço de busca para que o processo evolutivo encontre a solução correspondente ao tipo de problema proposto.

O conjunto de terminais deve conter variáveis, constantes e funções que não requeiram argumentos, que não causem, por si só, a necessidade de ligações com outros nós, em outras palavras, que sejam finais. Como exemplos de possíveis conjuntos de terminais poderíamos citar:

$$T = \{1, 5, 10, x, y, z, \text{salário}, \text{idade}, \text{verdadeiro}, \text{falso}\}$$

$$T = \{\text{virar à esquerda}, \text{virar à direita}, \text{avançar}, \text{recuar}\}$$

Já o conjunto de funções deve conter operações que requeiram argumentos. Pode-se dizer, portanto, que um nó que possui uma função necessariamente gera de imediato a necessidade da existência de determinado número de outros nós na árvore, que podem conter, por sua vez, um terminal ou outra função. Como exemplos de conjuntos de funções poderíamos citar:

$$F = \{+, -, \text{seno}, \text{cosseno}, \text{tangente}\}$$

$$F = \{\text{faça-enquanto}, \text{se-então-senão}, <, >\}$$

Os conjuntos primitivos de funções e terminais podem ser combinados livremente, desde que sejam obedecidas as aridades específicas, isto é, desde que a cada função seja atribuída exatamente a quantidade de argumentos necessários. Se determinado operador requer  $n$  argumentos, o nó que o contém terá exatamente  $n$  nós filhos, cada um constituindo um argumento para o nó pai.

A partir desses conceitos, necessariamente qualquer algoritmo obtido a partir da PG, independentemente da sua área de domínio, é uma combinação entre funções e terminais obtida a partir do processo evolutivo simulado. Na figura 2 temos, ilustrativamente, à esquerda a combinação entre um conjunto de terminais  $T = \{Y, B, X, \dots\}$  e um conjunto de funções  $F = \{+, x, \text{cosseno}, \dots\}$ . Ainda na figura 2, à direita há a combinação entre um conjunto de terminais  $T = \{B, 10, 3, Y, \dots\}$  e um conjunto de funções  $F = \{\text{se-então-senão}, >, x, \dots\}$ .

No entanto, é importante para que esses dois conjuntos possam ser manipulados e combinados pelo computador, que sejam obedecidas duas condições:

- **Suficiência:** visto que qualquer solução em PG é uma combinação entre os conjuntos F e T, há a necessidade de que a solução esperada possa, de fato, ser representada a partir desses conjuntos F e T determinados. Essa constatação, aparentemente simples e óbvia, deve ser analisada cuidadosamente na prática, já que em problemas reais ou mais complexos, especialmente, nem sempre é possível a absoluta certeza a respeito de quais funções e terminais específicos serão realmente utilizados ou não na solução do problema. Nessas condições, uma alternativa é aumentar o espaço de busca cuidadosamente, para não degradar o processo evolucionário.
- **Consistência:** definida também como condição de integralidade, é a condição que determina que qualquer função deve ser flexível o suficiente para aceitar como argumento qualquer elemento existente dentro do espaço de busca, ou seja, qualquer elemento existente dentro do conjunto de funções ou dentro do conjunto de terminais. No entanto, em casos reais essa condição deve ser relaxada através do uso de restrições na possibilidade das relações. Deve-se, por exemplo, impossibilitar que o segundo argumento de uma função de divisão seja a constante zero e deve-se evitar a multiplicação entre terminais booleanos, o que não faria sentido, entre outras restrições possivelmente necessárias.

### **3.1.5 Criação da população inicial**

Logicamente, para que o processo evolutivo possa atuar é necessária a existência de uma população inicial de indivíduos que serão comparados e evoluídos através da seleção natural simulada. Essa população inicial não deve apresentar tendenciosidades e deve ser formada por indivíduos diferentes, ainda que esses indivíduos sejam simples e não especializados.

Apesar dos diferentes métodos criados com o objetivo de garantir uma população inicial adequada e que possibilite um espaço considerável de busca



para a atuação da PG, a criação dessa população é basicamente aleatória, desde que se respeitem as aridades dos operadores.

### **3.1.6 Avaliação dos Indivíduos**

A função de avaliação dos indivíduos é a função responsável por definir a aptidão de cada programa e, portanto, é indispensável ao processo evolutivo, visto que esse é totalmente baseado no conceito de seleção de indivíduos melhor adaptados em meio à população.

Uma vez que a PG atua em diversos domínios dos mais variados tipos e a função de avaliação é sempre específica para o problema em questão seria impossível descrever todas as funções de avaliação existentes.

No entanto, um aspecto comum ao processo de formação das funções de avaliação é a existência de uma base de dados para treinamento. Essa base de dados, relativa necessariamente ao problema em questão, contém resultados ideais previamente conhecidos. Para um problema classificatório, por exemplo, em que se deseja diagnosticar uma determinada doença com base em dados clínicos, inicialmente a PG atuaria sobre os dados médicos de determinado número de pacientes, cujos diagnósticos já foram finalizados. A função de avaliação, nesse caso, definiria à que distância os resultados obtidos pela utilização de cada cromossomo produzido pela técnica estão dos resultados previamente conhecidos, com o objetivo de eliminar aqueles cromossomos mais distantes das respostas esperadas e, a partir do processo evolutivo, alcançar um cromossomo satisfatório para classificação da massa de dados de treino. Somente a partir daí, o programa encontrado através da PG seria responsável por receber dados clínicos de pacientes com diagnósticos não conhecidos e, valendo-se da árvore mais satisfatória encontrada, classificá-los em um dos tipos de diagnósticos possíveis.

Vale ainda salientar que em determinados casos pode-se incluir dentro da função de avaliação penalidades objetivando especificidades. Um exemplo seria a inclusão de um coeficiente de penalização por complexidade, em que uma solução seria considerada menos apta conforme sua extensão crescesse além do desejado.

### **3.1.7 Critério de Parada**

O critério de parada define em que momento o ciclo da evolução simulada será terminado.

A princípio, existe o ideal de que a evolução somente deve terminar no momento em que é obtido a partir do processo evolutivo o indivíduo ideal, ou seja, aquele indivíduo cujo cromossoma atinge exatamente o resultado esperado. No, entanto, para muitos casos essa condição deve ser relaxada. Problemas muito complexos ou recursos computacionais limitados são exemplos desses casos. Ainda, no âmbito da classificação de dados há o fato de que caso uma solução tenha uma taxa percentual de acertos excessivamente alta para uma dada base de treino, isso pode significar um ajuste demasiado à base específica de dados utilizada para treinamento, o que pode comprometer futuras classificações.

Assim, visto que a busca pelo cromossomo ideal nem sempre é viável, adota-se a busca pelo cromossomo satisfatório e a partir desse conceito surgem sugestões acerca de diferentes critérios de parada: aptidão dentro da faixa aceitável, limite pré-estabelecido de número de gerações, limite pré-estabelecido de tempo de execução ou, ainda, a estagnação do processo evolutivo – quando a partir de certo momento o processo evolutivo simulado não é capaz de encontrar soluções mais aptas.

### **3.1.8 Esquemas de seleção**

Os esquemas de seleção tratam da maneira como ocorre, dentro da população, a seleção de determinados indivíduos mais promissores para a troca de material genético. Uma vez aplicada a função de avaliação os diferentes programas candidatos estão ranqueados de acordo com a aptidão de seus cromossomos. A partir daí, o esquema de seleção é responsável por conduzir o processo evolutivo para espaços mais promissores, garantindo a conservação da diversidade e garantindo que os códigos genéticos mais adaptados consigam perpetuar descendentes, já que a transmissão de bom material genético tende a produzir indivíduos cada vez mais aptos.

A probabilidade de seleção de indivíduos mal adaptados, bem como a percentagem de indivíduos selecionados dentro da população são definidas pela pressão de seleção. Assim, uma pressão demasiadamente alta, em que apenas um seleto grupo de indivíduos mais aptos é escolhido, garante que, de fato, apenas os melhores indivíduos alcancem a reprodução, mas pode prejudicar a diversidade. Em contrapartida, uma pressão exageradamente baixa garante uma boa diversidade, com grande espaço de busca, mas pode degradar o processo evolutivo por selecionar muitos indivíduos pouco promissores.

### **3.1.9 Operadores genéticos**

Os operadores genéticos tratam da criação de novos indivíduos derivados de indivíduos anteriores. É através dos processos de recombinação e modificação de árvores ou de fragmentos de árvores que novos cromossomos surgem e o espaço de busca é explorado com a finalidade da obtenção de soluções cada vez mais satisfatórias. Dois exemplos de operadores genéticos mais utilizados são o cruzamento e a mutação.

O operador de cruzamento, cuja probabilidade de ocorrência é comumente a maior, forma indivíduos diferentes a partir de dois indivíduos previamente existentes. Esse operador seleciona aleatoriamente sub-árvores de cada um dos indivíduos e as recombina, explorando, assim, um novo espaço de busca, conforme ilustrado na figura 3, em que os indivíduos A e B passam por cruzamento, gerando os indivíduos A' e B'.

Já o operador de mutação modifica a estrutura de um único indivíduo, gerando, assim, outro indivíduo. Geralmente essas mudanças tendem à sutileza, podendo ser definidas, por isso, apenas como perturbações. Na mutação padrão, uma sub-árvore de um indivíduo é escolhida aleatoriamente e substituída por uma nova sub-árvore criada também aleatoriamente, conforme exemplificado na figura 4.

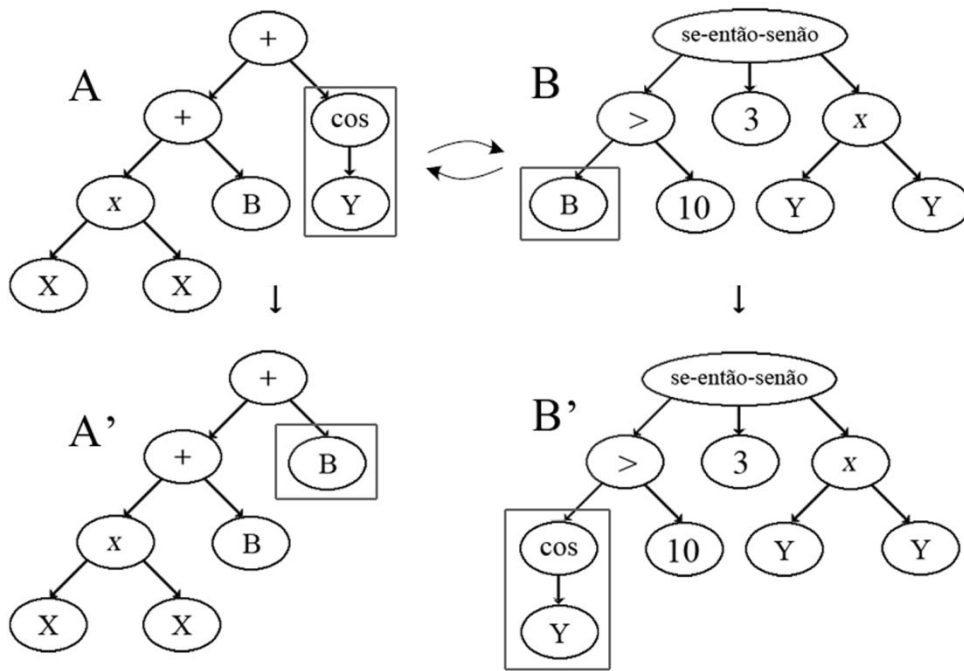


Figura 3: Cruzamento dos indivíduos A e B

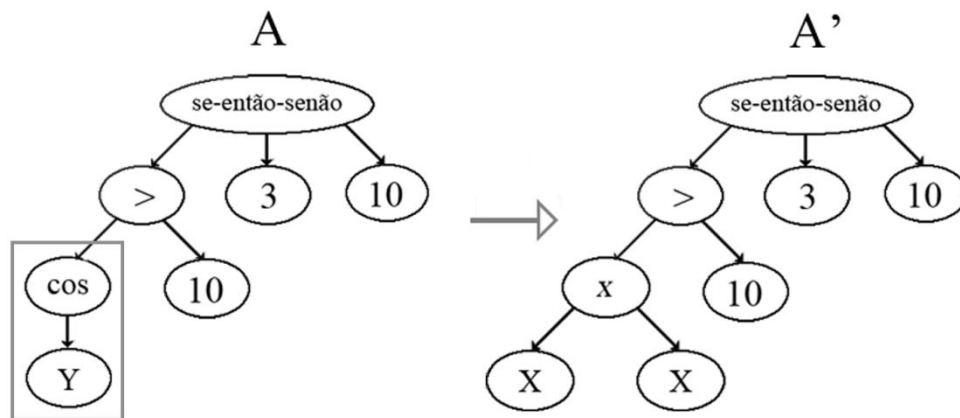


Figura 4: Mutação padrão do indivíduo A

### 3.2 A Implementação da Programação Genética (gpclassifier)

No processo prático de aplicação da programação genética na base de dados que será posteriormente abordada foi utilizado, com o apoio direto do autor, o *gpclassifier* (*Genetic Programming Classifier*). Desenvolvido por Douglas Augusto (Augusto, 2009), o *gpclassifier* pode ser definido como um classificador de dados em constante desenvolvimento criado para a implementação da PG em computação de alto-desempenho, explorando-a

através de conceitos multi-populacionais e da computação massivamente paralela, decorrente da evolução simultânea de diversas populações semi-isoladas. Para atender aos elevados, ainda que distribuídos, custos computacionais da técnica de aprendizado supervisionado foi utilizada a infraestrutura tecnológica do Laboratório Nacional de Computação Científica (LNCC), localizado em Petrópolis, no Rio de Janeiro.

O *gpclassifier* está disponibilizado livremente, sendo, portanto, classificado como Software Livre e pode ser obtido sem quaisquer restrições de uso no seguinte endereço eletrônico:

<http://sourceforge.net/projects/gpclassifier>

Inicialmente, foram submetidos ao software os conjuntos de funções e terminais referentes ao problema sugerido. A partir deste momento começa o processo evolutivo empregando-se a mutação e o cruzamento em suas versões tradicionais. Também foi utilizada a chamada topologia em ilhas. Neste conceito várias populações menores e independentes evoluem simultaneamente umas às outras, cruzando apenas entre si. Eventualmente há a migração de indivíduos entre as ilhas e, portanto, a mistura de materiais genéticos provenientes de populações diferentes e separadas.

Durante os testes necessários ao trabalho proposto, foram realizados experimentos com cerca de 10.000 indivíduos em cada ilha e a quantidade de ilhas variando entre 16 e 64. Assim, em determinados experimentos trabalhou-se com uma população inicial contabilizada em torno de 640.000 indivíduos. O número de gerações criadas até um resultado satisfatório foi variável entre 1.000 e 10.000 gerações.

Comparando-se os resultados obtidos por cada árvore classificadora com os previamente conhecidos a partir da base de dados utilizada, foi definida a eficiência de cada cromossomo obtido pela PG e, finalmente, os mais satisfatórios já puderam ser impressos, fato que demonstra uma das características da PG mais interessantes à proposta do presente trabalho: a leitura e a interpretação imediatas e diretas, possibilitando o aprendizado humano relativo ao problema sugerido.

### 3.3 A base de dados

Uma das tarefas iniciais e mais complexas de trabalhos dependentes de bases de dados acerca de atletas é a obtenção dos dados propriamente ditos. Dada a impossibilidade da criação de uma base de dados própria, específica e ideal, recorreu-se à uma base de dados previamente existente e razoavelmente capaz de atender aos objetivos finais almejados.

Assim, o processo evolutivo foi aplicado à base de dados obtida por Franciele Marques Vanderlei (Vanderlei, 2011) como parte da dissertação apresentada à Faculdade de Ciências e Tecnologia (FCT/UNESP), campus de Presidente Prudente, como requisito para a obtenção do título de Mestre no Programa de Pós-Graduação em Fisioterapia. A amostra utilizada contém dados de 1.311 crianças e adolescentes, de ambos os sexos, com idades variáveis entre 4 e 21 anos, praticantes de esportes oferecidos pela Secretaria Municipal de Esportes de Presidente Prudente. Cada praticante do esporte se relaciona com, além do número de identificação, as seguintes variáveis individuais: modalidade, categoria (variando entre iniciação, aperfeiçoamento ou treinamento), posicionamento (com valores coerentes com a modalidade praticada), dominância (destro, canhoto ou bi destro), gênero, idade, peso, estatura, IMC, tempo de prática, horas semanais de treino e ocorrência ou não de lesão. Aqueles atletas que contém como valor da última variável citada uma resposta positiva ainda preenchem as seguintes variáveis: local anatômico da lesão, mecanismo (contato direto, sem contato ou *overuse*), momento da lesão (variando entre treinamento ou competição), gravidade da lesão, retorno as atividades normais (variando entre assintomático, sintomático ou não retorno) e, finalmente, recidivas (informa se houve repetição da lesão). As modalidades representadas na base de dados são: atletismo, basquetebol, futebol, futsal, ginástica artística, karatê, kung-fu, tênis de mesa, natação e voleibol.

A variável “Gravidade da lesão” foi definida como a incógnita, ou seja, como a resposta almejada, devido à sua intensa ligação com a maioria das variáveis disponíveis na base de dados e devido à grande importância econômica, esportiva e medicinal do entendimento das relações entre fatores capazes de agravar ou não uma lesão. Portanto, foram excluídos do processo

os atletas que não sofreram nenhum tipo de lesão ao longo de sua trajetória esportiva, sendo submetidos ao processo evolucionista os registros de 261 praticantes, com idades entre 8 e 18 anos.

O mais importante, ainda, é o fato de que exatamente a mesma técnica e exatamente os mesmos procedimentos tecnológicos são aplicáveis a bases de dados envolvendo atletas de todos os níveis esportivos e de todas as idades. O presente trabalho, portanto, não trata da aplicabilidade da PG como auxiliar a categoria infanto-juvenil ou amadora, mas, trata da aplicabilidade da técnica a quaisquer bases de dados de lesões esportivas, independentemente de modalidade, idade ou nível competitivo e, adicionalmente, demonstra essa eficácia diante de uma base de dados exemplo composta por atletas de determinado perfil. Os mesmos procedimentos aqui descritos são válidos e eficazes, conseqüentemente, tanto a atletas jovens quanto a atletas idosos e tanto a atletas amadores quanto a atletas de elite.

## **4. RESULTADOS DA PESQUISA E DISCUSSÃO**

Este capítulo tem o objetivo de apresentar algumas das mais aptas árvores finais obtidas a partir dos métodos descritos no capítulo anterior e discutir as relações propostas por essas estruturas. Nesse sentido, é importante realçar novamente que cada árvore foi obtida a partir da aplicação da PG, utilizando-se como alvo de aprendizado a base de dados já especificada e que, portanto, todas as relações descritas a seguir foram descobertas através da implementação de operadores genéticos, funções e terminais, seguindo exclusivamente as técnicas do processo evolucionista descritas na sessão 3.1, independentemente de qualquer conhecimento científico aprofundado sobre qualquer nicho da medicina esportiva.

### **4.1 Cromossomos obtidos e pseudocódigos**

Os algoritmos a seguir representam três árvores de classificação formadas por constantes, variáveis, operadores aritméticos, operadores lógicos e operadores condicionais (SE e SENÃO) que, por sua vez, conduzem a leitura rumo a determinados campos. Na tabela 1, por exemplo, caso a condição da linha 1 não seja satisfeita, será executado o código da linha 10, que conclui,

finalmente, que a lesão é leve. Na mesma tabela, se a condição da linha 1 for verdadeira, será testada a condição presente na linha 2, que se verdadeira levará a execução da linha 3 e se falsa levará a execução da linha 5. Nestes casos é necessária a interpretação de acordo com a indentação do texto, que oferece a possibilidade de uma visão hierárquica simples.

```

1 | SE peso >= 101,429 OU modalidade = atletismo ENTÃO
2 |     SE tempo de treinamento x horas semanais de treino < 9,99966 ENTÃO
3 |         lesão leve
4 |     SENÃO
5 |         SE imc > 21,2409 E local anatômico ≠ tornozelo ENTÃO
6 |             lesão moderada
7 |         SENÃO
8 |             lesão grave
9 | SENÃO
10 |     lesão leve

```

**Tabela 1 representando algoritmo classificador obtido**

```

1 | SE posicionamento = central OU peso/idade = horas semanais de treino OU
   |                                     peso > 101,325 ENTÃO
2 |     lesão moderada
3 | SENÃO
4 |     SE 9,11939 / tempo de treinamento >= horas semanais de treino OU
   |                                     modalidade ≠ atletismo ENTÃO
5 |         lesão leve
6 |     SENÃO
7 |         SE imc < 21,1994 ENTÃO
8 |             lesão grave
9 |         SENÃO
10 |             lesão moderada

```

**Tabela 2 representando algoritmo classificador obtido**

```

1 | SE peso > 101,429 OU imc <= 14,2488 OU modalidade = atletismo ENTÃO
2 |     SE tempo de treinamento x horas semanais de treino < 10,6107 ENTÃO
3 |         lesão leve
4 |     SENÃO
5 |         SE local anatômico ≠ tornozelo E imc >= 21,2409 ENTÃO
6 |             lesão moderada
7 |         SENÃO
8 |             lesão grave
9 | SENÃO
10 |     lesão leve

```

**Tabela 3 representando algoritmo classificador obtido**



## 4.2 Análise dos cromossomos

É importante ressaltar, inicialmente, que cada uma das três árvores representadas por algoritmos nas tabelas 1, 2 e 3 obtiveram taxas de acertos respectivamente iguais a 88,1%, 88,5% e 88,5% quando testadas suas aptidões para a definição das gravidades das lesões explicitadas na base de dados utilizada no presente trabalho. A partir dessa alta acurácia e da robustez da técnica foi possível a extração de confiáveis, interessantes e, por vezes, inesperadas correlações entre os diversos fatores que potencialmente estão ligados à gravidade de lesões em atletas de todos os níveis competitivos, a despeito de qualquer carência de conhecimento prévio sobre a medicina esportiva.

### 4.2.1 Peso e IMC

A primeira relação que fica evidente nos três indivíduos, por exemplo, é a forte influência do peso no que se refere à gravidade das lesões. Em todos os três algoritmos a comparação do peso do atleta com uma constante em quilogramas é determinante para o resultado final. Constante essa que, apesar de não pré-definida em momento algum, apresenta valores muito próximos nos três algoritmos, chegando inclusive à igualdade nas três casas decimais nas tabelas 1 e 3 (linha 1 em ambas as tabelas). As três tabelas sugerem, portanto, que a probabilidade da ocorrência de uma lesão moderada ou grave é consideravelmente maior em atletas com alto peso corporal, o que pode ser mais facilmente visualizado na tabela 1, em que, exceto em praticantes de atletismo, para que o atleta possa ser acometido por lesões moderadas ou graves, seu peso deve ser necessariamente maior do que a constante (na tabela 1, só é possível obter os resultados das linhas 6 ou 8, caso a condição da linha 1 seja verdadeira). Existem, ainda, estudos provenientes de especialistas que abordam a relação entre altos pesos corporais e lesões graves em atletas, embasando a possibilidade de relações entre esses dois fatores (Domingues *et al.*, 2005; Camargo *et al.*, 2009). Dando prosseguimento as relações envolvendo exclusivamente o peso, as linhas 7 e 8 da tabela 2 sugerem, inesperadamente, que um Índice de Massa Corporal relativamente

baixo, ainda que sensivelmente, também pode estar relacionado com lesões de alta gravidade. É possível que essa sugestão se justifique quando se analisam as questões referentes ao desenvolvimento muscular na base de dados específica utilizada. Em crianças ou adolescentes, o IMC sensivelmente baixo pode estar relacionado com um desenvolvimento muscular incompleto ou pouco satisfatório, o que pode ser responsável, em alguns casos, por gerar um Índice de Massa Corporal menos elevado e ao mesmo tempo, aumentar a exposição às lesões, algo que justifica os reforços musculares a que são submetidos inclusive atletas de nível profissional. Ao mesmo tempo, as linhas 5, 7 e 8 da tabela 1 ainda relacionam taxas de IMC sensivelmente superiores com lesões moderadas no tornozelo, o que pode demonstrar que quanto maior é o IMC de um atleta, maiores são o impacto e a pressão exercidos sobre esta parte específica do corpo.

#### **4.2.2 Atletismo**

Quanto à especificação de modalidades, entretanto, as únicas menções encontradas nas árvores são referentes ao atletismo. Nas primeiras linhas das tabelas 1 e 3, o praticante de atletismo tem a possibilidade de sofrer lesões de maior gravidade unicamente por ser praticante de tal modalidade. Mais radicalmente, na primeira linha da tabela 1, por exemplo, o praticante de atletismo é equivalido a um desportista com peso corporal superior a 100 quilogramas, para a possibilidade de lesões de alta gravidade: “Se peso maior ou igual a 101,429 ou modalidade igual a atletismo”. Nas linhas 4 e 5 da tabela 2, pode-se perceber que atletas praticantes de quaisquer modalidades diferentes de atletismo tem mais possibilidades de obter lesões leves (visto uma das condições que devem ser satisfeitas para que o algoritmo siga para a linha 5 e classifique a lesão como leve: “Se [...] ou modalidade é diferente de atletismo”). Essa relação aparentemente mais próxima entre o atletismo e as lesões de maior gravidade é explicitada estatisticamente no trabalho de Marques Vanderlei (Vanderlei, 2011), o que colabora para a credibilidade dos resultados obtidos através da PG.

### **4.2.3 Gênero**

Um dado igualmente interessante é a ausência de qualquer tipo de relação explícita entre as lesões e um determinado gênero em qualquer um dos três algoritmos, algo que concorda perfeitamente com diversos trabalhos de especialistas em medicina esportiva (Domingues *et al.*, 2005, Conte *et al.*, 2002, Hootman *et al.*, 2002).

### **4.2.4 Tempo de treinamento e horas semanais de treino**

Presente nos três cromossomos apresentados, a relação entre o tempo de treinamento (quantidade de anos em que um atleta pratica determinada modalidade), as horas semanais de treino e a gravidade de lesões está definida de três formas diferentes nas linhas 2, 4 e 2 das tabelas 1, 2 e 3, respectivamente. Inicialmente, nota-se de maneira clara que esses dois fatores estão sempre associados a uma constante cujos valores variam minimamente, sempre em torno de 9 ou 10. Partindo de uma interpretação mais atenta, ainda é possível perceber que em todos os três indivíduos, caso a condição entre tempo de treinamento, horas semanais de treino e a mencionada constante seja satisfeita, isto é, seja dita verdadeira, qualquer possibilidade de lesão moderada ou grave é imediatamente descartada e o algoritmo classificatório define o resultado final como “lesão leve”. Partindo da interpretação das multiplicações sugeridas pelas tabelas 1 e 3, também se verifica que quanto maior qualquer uma das duas variáveis, menor é a possibilidade de lesões leves (o que, por sua vez abre possibilidade para lesões moderadas ou graves) e verifica-se, ainda, que as duas grandezas são igualmente importantes, visto que uma variável não influencia mais o resultado final do que a outra. Igualmente, na tabela 2, apesar da representação de uma divisão, ainda é notória a relação inversa que deve existir entre as duas grandezas para que o atleta em questão seja acometido por lesões apenas leves. Segundo os três cromossomos, novamente unânimes, se o atleta é praticante do desporto por mais tempo, a quantidade de treinos semanal deve ser menor, e vice-versa. Para avaliarmos a constante, entretanto, é necessário considerar o fato de que a base de dados específica utilizada para treinamento pelo processo evolucionista compreende atletas jovens, com uma média de tempos de

treinamento em torno de apenas 3 anos. Isso porque, para atletas mais experientes, se o limite entre a multiplicação das duas variáveis deve ser menor que um valor próximo de 10 (tabela 1, linha 2), a satisfação dessa condição seria inviável. No entanto, para um atleta menos experiente que pratica certa modalidade a 3 anos, por exemplo, o tempo semanal de treino não deveria exceder pouco mais de 3 horas, um valor moderadamente plausível. Deve-se considerar, além de tudo, que atletas que praticam o esporte por pouco tempo, especialmente no primeiro ano de atividade, tendem a ser submetidos a níveis e cargas de treino mais leves e iniciais, o que contribui muito para que quanto menor o tempo de prática de um atleta, menor o risco do surgimento de lesões graves. No entanto, na medida em que um atleta se torna mais experiente, seu treinamento tende a atingir um nível mais intenso e exigir maior tempo de recuperação física, o que poderia justificar a recomendação de poucas horas semanais de treino a atletas com maior tempo de prática esportiva. Finalmente, é importante observar mais uma vez que a satisfação dessa condição elimina totalmente (segundo os cromossomos obtidos) qualquer possibilidade de lesões moderadas ou graves, e que por isso a condição deve ser naturalmente mais rígida e difícil de alcançar, o que justifica bem o valor relativamente baixo da constante a ela relacionada.

## **5. Conclusões**

Após a abordagem da relevância do descobrimento das relações entre fatores de influência na gravidade de possíveis lesões de atletas, esse trabalho conclui, primeiramente, que a meta-heurística denominada Programação Genética é uma aliada potencialmente eficaz no estudo do tema especialmente devido a uma das suas principais características: a pouca exigência de conhecimento sobre o domínio estudado.

Para tanto, além de demonstrar teoricamente a técnica e justificar a sua escolha, foram feitos experimentos sobre uma base de dados exemplo, destacando-se, porém, a usabilidade da mesma técnica para outras bases de dados, algo que resultará em saídas adaptadas à base utilizada. A Programação Genética, futuramente poderá ser utilizada com dados mais abrangentes e mais próximos do ideal, com uma única modalidade alvo de

estudo, ou ainda, almejando o melhor entendimento acerca de uma relação específica.

Dentre as relações encontradas nos algoritmos classificadores mais eficazes, algumas foram destacadas. Primeiramente, percebeu-se que processo evolucionista descobriu automaticamente uma intensa relação entre o peso corporal e o IMC dos atletas e as diferentes gravidades de suas lesões, algo com o qual concordam diversos pesquisadores especializados no assunto, conforme demonstrado. Os algoritmos foram responsáveis, ainda, por dedicar atenção especial ao atletismo, dentre todas as 10 modalidades representadas, sugerindo a maior probabilidade de praticantes dessa modalidade específica sofrerem lesões de diferentes gravidades, apoiando, também, o que já foi constatado estatisticamente por pesquisadores anteriormente citados. Destaca-se, ainda, que não houve nenhum tipo de relações entre gêneros e lesões, concordando novamente com outros estudos. Algo que merece mais apreço, entretanto, foi a descoberta das relações entre tempo de treinamento, horas semanais de treino e a gravidade de lesões. Segundo os indivíduos obtidos a partir da Programação Genética, para que o atleta tenha menos chances de sofrer lesões graves ou moderadas, os valores da quantidade de anos de esporte praticados por um atleta e o número de horas de treino devem seguir rígidas relações e, passando por operações matemáticas simples, não devem atingir certas constantes pré-definidas.

Por tudo isso, o presente trabalho contribuiu para explicitar ainda mais a técnica conhecida como Programação Genética e demonstrar a sua eficácia com relação ao tema proposto, confirmando diversas relações científicas entre fatores de influências na gravidade de lesões de atletas e, ainda, sugerindo outras; algo de elevada importância para vários âmbitos do esporte em geral. Ressalta-se, finalmente, que não foram encontrados outros trabalhos com o intuito de demonstrar a predição da gravidade de lesões em atletas utilizando-se a Programação Genética e que, portanto, o presente artigo abre possibilidades e sugestões para novas pesquisas nessa área.

## **6. Referências Bibliográficas**

Arena, S.S., Carazzato, J.G. A relação entre o acompanhamento médico e a incidência de lesões esportivas em atletas jovens de São Paulo. *Rev. Bras. Med. Esporte.* 13(4). (2007).

Atalaia, T., Pedro, R., Santos, C. Definição de Lesão Desportiva – Uma Revisão da Literatura. *Rev. Portuguesa de Fisioterapia no Desporto*, 3(2), 13-21. (2008).

Augusto, D.A. Co-evolução amostra-classificador integrada à programação genética gramatical para a classificação de dados. Dissertação de Mestrado em Ciências em Engenharia Civil – Universidade Federal do Rio de Janeiro (2004).

Augusto, D.A., Barbosa, H.J.C. Programação Genética. em: A. Gaspar-Cunha; R.H.C. Takahashi; C.H. Antunes, *Manual de Computação Evolutiva e Metaheurística.* 67-85. (2009).

Augusto, D.A. Programação Genética multi-populacional e co-evolucionária para classificação de dados. Tese de Doutorado em Engenharia Civil – Universidade Federal do Rio de Janeiro (2009).

Augusto, D.A., Barbosa, H.J.C. Symbolic regression via genetic programming. *SBRN '00: Proceedings of the VI Brazilian Symposium on Neural Networks (SBRN'00)*, IEEE Computer Society, 173. (2000).

Bernardino, H.S., Barbosa, H.J.C. Grammar-Based Immune Programming for Symbolic Regression. *8th International Conference on Artificial Immune System - ICARIS*, 274-287. (2009).

Betti, M. A janela de vidro: esporte, televisão e educação física. Tese de Doutorado em Educação - Faculdade de Educação da Universidade Estadual de Campinas (1997).

Caine, D., Maffulli, N., Caine, C. Epidemiology of injury in child and adolescent sports: injury rates, risk factors, and prevention. *Clin. Sports Med.* 27(1), 19-50. (2008).

de Camargo, P., Leme, G., Eugênio, L. Revisão de artigos de ortopedia e medicina esportiva publicados em periódicos brasileiros. *Acta Ortopédica Brasileira*, 17(6), 344-349. (2009).

Carneiro, M.L. Programação Genética aplicada à programação de controladores lógico programáveis. Dissertação de Mestrado em Engenharia Elétrica e de Computação – Universidade Federal de Goiás (2009).

Cohen, M.; Abdalla, R.J. Lesões nos esportes: diagnóstico, prevenção e tratamento. *Revinter*, 139-152. (2002).

Conte, M., Matiello Jr., E., Chalit, L.V.A.S., Gonçalves A. Exploração de Fatores de Risco de Lesões Desportivas entre Universitários de Educação Física: estudo a partir de estudantes de Sorocaba/SP. *Rev. Bras. Med. Esporte*, 8(4) 151-156. (2002).

Domingues, A.P.T., Marcelo, C., Más, E.F., Ramalho, L.C.B., Godoi, V.J., Teixeira, L.F.M., Riyis, M.T., Fermi, J.P. Implicações do nível de aptidão física na gênese de lesões esportivas. *Rev. Bras. Cineantropom. Desempenho. Hum.* 7(2), 29-35. (2005).

Fong, D., Hong, Y., Chan, L., Yung, P., Chan, K. A Systematic Review on Ankle Injury and Ankle Sprain in Sports. *Sports Med.*, 37(1), 73-94. (2007).

Fontana, R.F. O papel da fisioterapia na performance do atleta. *Rev. de Fisioterapia da Universidade de São Paulo*, 6. (1999).

Guedes, J.M. Lesões em Tenistas Competitivos. *Rev. Bras. Cienc. Esporte*, Campinas, 31(3), 217-229. (2010).

Hootman, J.M., Macera, C.A., Ainsworth, B.E., Addy, C.L., Martin, M.M., Blair S.N. Epidemiology of musculoskeletal injuries among sedentary and physically active adults. *Med. Sci. Sports Exerc.*, 34(5), 838-844. (2002).

Khawali, C., Andriolo A., Ferreira, S.R.G. Benefícios da Atividade Física no Perfil Lipídico de Pacientes Com Diabetes Tipo 1. *Arq. Bras. Endocrinol. Metab.* 47(1). (2003).

Koza, J.R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press. (1992).

Koza, J.R. Genetic Programming, on the programming of computers by means of natural selection. The MIT press. (2000).

Kujala, U., Orava, S., Parkkari, J., Kaprio, J., Sarna, S. Sports Career-Related Musculoskeletal Injuries: Long-Term Health Effects on Former Athletes. *Sports Medicine*. 33(12), 869-875. (2003).

Maia, L., Bianchi, R.A.C. Evoluindo Agentes Jogadores de Futebol. *Workcomp'2000. CTA/ITA*, 113-119. (2000).

Miller, J.F. and Smith, S.L. Redundancy and computational efficiency in cartesian genetic programming. *IEEE Transactions on Evolutionary Computation* 10(2), 167-174. (2006).

Olsen, O., Myklebust, G., Engebretsen, L., Holme, I., Bahr, R. Exercises to prevent lower limb injuries in youth sports: cluster randomised controlled trial. *BMJ*, 330-449. (2005).

- Oltean, M., Grosan, C., Diosan, L., Mihaila, C. Genetic programming with linear representation: a survey. *International Journal on Artificial Intelligence Tools* 18(2), 197-238. (2009).
- Pastre, C.M., Carvalho Filho, G., Monteiro, H.L., Netto Júnior, J., Padovani, C.R., García B.A. Exploração de fatores de risco para lesões no atletismo de alta performance. *Rev. Bras. Med. Esporte* 13(3), 200-204. (2007).
- Pelikan, M., Kvasnicka, V., Pospichal, J. Read's linear codes and genetic programming. J.R. Koza, K. Deb, M. Dorigo, D.B. Fogel, M. Garzon, H. Iba, R. L. Riolo, *Genetic Programming Proceedings of the Second Annual Conference*, Morgan Kaufmann, Stanford University, 268. (1997).
- Pinheiro, J. Darwin e sua presença na teologia. *Revista Espaço Acadêmico*, 95. (2009).
- Poli, R. Parallel distributed genetic programming. em D. Corne, M. Dorigo and F. Glover, *New Ideas in Optimization, Advanced Topics in Computer Science*, 27, 403-431. (1999).
- Skinner, J.S. Prova de Esforço e Prescrição de Exercício Físico para casos específicos - Bases Teóricas e Aplicações Clínicas. *Revinter*. (1991).
- Stacoff, A., Denoth, J., Kaelin, X., Stuessi, E. Running injuries and shoe construction: some possible relationships. *Int. J. Sport Biom.* 4, 342-357 (1988).
- Stella, F., Gobbi, S., Icassatti, D.C., Riani J.L.C. Depressão no Idoso: Diagnóstico, Tratamento e Benefícios da Atividade Física. III Congresso Internacional de Educação Física e Motricidade Humana e IX Simpósio Paulista de Educação Física, Rio Claro, 8(3), 91-98. (2003).
- Teller, A. Evolving programmers: The co-evolution of intelligent recombination operators. em P.J., Angeline, K.E., Kinnear, Jr. *Advances in Genetic Programming* 2, MIT Press, 3, 45-68. (1996).
- Thorpe, C., Herbert, M., Kanade, T., Shafer. Toward autonomous driving: the CMU Navlab. I. Perception. *IEEE Expert*, 6(4), 31-42. (1991).
- Timpka, T., Ekstrand, J., Svanström, L. From Sports Injury Prevention to Safety Promotion in Sports. *Sports Med.*, 36(9), 733-745. (2006).
- Valiant, G.A., Cavanagh, P.R. A study of landing from a jump: implications for the design of a basketball shoe. *Biomechanics IX-B*, Champaign, Illinois: Human Kinetics (1985).
- Vanderlei, F.M. Lesões em crianças e adolescentes praticantes de diferentes modalidades esportivas. *Dissertação de Mestrado em Fisioterapia – Faculdade de Ciências e Tecnologia (FCT-UNESP)* (2011).