

Um Sistema Computacional integrando Suporte à Decisão na Área de Reprodução Humana

Erica Akemi Tanaka¹, Cristina Moraes Junta², Laura Vagnini³, José Augusto Baranauskas⁴, Silvana Giuliatti⁵

^{1,5}Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Brasil. ^{1,4}Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Brasil. ^{2,3}Centro Paulista de Diagnóstico, Pesquisa e Treinamento, Brasil.

Resumo: Clínicas médicas e laboratórios de diagnóstico possuem uma série de dados de pacientes e exames laboratoriais que, em conjunto, auxiliam o médico a confirmar ou não a hipótese diagnóstica, além de direcionar o tratamento adequado. Este estudo tem como objetivo propor uma arquitetura de *software* bem como reportar aspectos do desenvolvimento de um sistema computacional que a implementa. Tais elementos permitem informatizar os processos e dados referentes aos pacientes de reprodução assistida, facilitando, assim, o acesso à informação e à manipulação a fim de indicar o melhor tratamento e prever o sucesso ou não da gravidez após realizar o tratamento. A arquitetura de *software* é composta por três camadas, sendo que uma delas, a camada de classificação é composta por quatro módulos de classificação, voltados ao suporte à decisão. Isso permite aos especialistas médicos não só indicar um tratamento e prever o sucesso do mesmo como também uma melhor compreensão dos padrões de sucesso ou não do processo de fertilização.

Palavras-chaves: Reprodução Assistida, Sistemas de Suporte à Decisão, Aprendizado de Máquina

Abstract: Medical clinics and diagnostic laboratories have a large number of patient data and laboratory tests that together help physicians to confirm or refute diagnosis, besides directing the appropriated treatment. In this research we propose a software architecture to model patients' data and processes in assisted reproduction as well as we describe some aspects of a computer system implementing the proposed architecture. These components allow easy access to information, indicate the best treatment and predict the success or failure of pregnancy after performing some specific treatment. The software architecture is composed by three layers. One of them, the classification layer contains four supervised learning algorithms, directed to decision support. This enables medical experts to indicate and to predict the success of treatments as well as to better understand patterns of success or failure of the fertilization process.

Keywords: Assisted Reproduction, Decision Support Systems, Machine Learning

1 Introdução

Infertilidade é um distúrbio ou condição do sistema reprodutivo masculino ou feminino que reduz a capacidade de ter filhos. Um casal é considerado infértil quando não consegue conceber, num período de 12 meses, sem uso de métodos anticoncepcionais, mantendo relações sexuais frequentes. Aproximadamente 40% dos casos de infertilidade são de causas femininas, 40% são de causas masculinas e 20% são de causas indeterminadas. Atualmente, entre 10 a 15% dos casais têm problemas com a infertilidade. Em consequência disso tem-

se observado um aumento na procura de tratamentos utilizando técnicas de reprodução assistida, que é uma área que engloba uma variedade de técnicas e terapias para o auxílio a casais que desejam ter filhos. O sucesso ou não de uma gestação é determinado, em grande parte, pelos aspectos imunológicos envolvidos na interação feto-gestante, na qual o embrião representa um antigênico estranho ao organismo materno. Portanto, é necessário que seja reconhecido imunologicamente evitando, assim, ser atacado pelas defesas do corpo, garantindo um ambiente favorável para o desenvolvimento do

embrião até que ele tenha a capacidade de vida extra-uterina¹. A reprodução assistida pode ser dividida em dois grandes grupos de técnicas: inseminação artificial e fertilização *in vitro* (FIV) e suas variantes². A inseminação artificial tem a finalidade de facilitar a chegada dos espermatozoides nas trompas em um número adequado para que ocorra a fertilização dos oócitos. Já a fertilização *in vitro*, ocorre no exterior do corpo feminino e possui várias etapas: indução da ovulação e o recolhimento dos oócitos, coleta dos espermatozoides, fertilização em meio de cultura por 48 horas e a transferência do embrião para o útero. Outra técnica utilizada é a injeção intracitoplasmática de espermatozoides (ICSI) que consiste na injeção de um único espermatozoide no citoplasma do ovócito, evitando assim as dificuldades do processo natural em que um espermatozoide tem que passar a 'barreira' do ovócito para nele penetrar.

O presente estudo tem como objetivo descrever uma arquitetura de *software* em três camadas, as quais permitem a manipulação e a recuperação das informações consideradas relevantes pelos especialistas médicos na área de reprodução humana. Em uma das camadas encontra-se o módulo de suporte à decisão que faz uso de algoritmos de aprendizado supervisionados, cujo objetivo consiste em identificar um subconjunto de informações que podem ser utilizadas para fins de prever o sucesso ou não da gravidez bem como indicar o melhor tratamento. Adicionalmente, também é descrito o desenvolvimento de um sistema computacional para *web* implementando a arquitetura proposta.

Este estudo está organizado da seguinte forma: Na Seção 2 são descritos trabalhos relacionados. Na Seção 3 a metodologia utilizada neste estudo é explicada, mostrando a estrutura em três camadas do *framework* desenvolvido; cada uma das camadas é descrita separadamente. As conclusões encontram-se na Seção 4.

2 Trabalhos Relacionados

Um dos problemas da análise dos dados gerados pelo tratamento em reprodução assistida é o grande volume de informação, tornando difícil distinguir quais dados estão realmente envolvidos no sucesso ou não do tratamento. Devido a isso, o uso de técnicas de aprendizado de máquina podem contribuir para extração de conhecimento e suporte à decisão^{3,4}. Recentemente

foi proposta a aplicação de classificadores Bayesianos para seleção de embriões que irão formar o lote a ser transferido para o útero da mulher que realizará a fertilização *in vitro*, a fim de prover um sistema de apoio à decisão⁵. Outro trabalho propõe a aplicação do classificador de árvores de decisão C5.0 (sucessor do C4.5) para aumentar a taxa de sucesso da fertilização *in vitro* levando em conta a fisiologia de um paciente, bem como os resultados obtidos de diferentes estágios de uma FIV⁶. Um estudo refere-se à arquitetura de um sistema de apoio à decisão para doença cardiovascular AptaCDSS-E⁷ que contém quatro componentes principais: interface do *scanner*, módulo de gestão de protocolo, módulo de configuração do serviço e interface de suporte ao diagnóstico. A interface do *scanner* lê os dados brutos gerados pelo 'chip scanner' e os converte para um formato adequado. O módulo de gestão de protocolo controla a comunicação entre todos os componentes. O módulo de configuração inclui quatro classificadores de quatro diferentes paradigmas de classificação para tomada de decisão clínica. A interface de suporte à decisão combina o resultado da decisão de cada classificador e fornece uma descrição simplificada do nível atual da doença do paciente. Outro trabalho aborda a arquitetura de um sistema para tratamento de aneurismas cerebrais @neuIST⁸, estruturado em três camadas: aplicação, *middleware* e recursos. A camada de aplicação consiste em aplicações que fornecem aos usuários finais funcionalidades para pesquisa e tratamento clínico dos aneurismas. A camada de *middleware* medeia entre a camada de aplicação e a camada de recursos, fornecendo acesso aos dados, preparação de dados, mediação semântica e serviços de computação em grade. A camada de recursos é responsável pela manipulação dos recursos e bases de dados.

3 Arquitetura do Framework

Inicialmente foi realizado um levantamento de requisitos do sistema, a partir do qual foram elaborados os diagramas de caso de uso para descrever um cenário que mostra as funcionalidades gerais do sistema do ponto de vista do usuário. Com base nisso, foi proposta uma arquitetura com três camadas, mostradas na Figura 1: (i) *Camada de Apresentação*, composta pela interface com o usuário (GUI —

Graphical User Interface) e responsável por toda interação com os especialistas médicos; (ii) *Camada de Classificação*, composta por quatro módulos de classificação utilizando aprendizado supervisionado, responsável pelo suporte à decisão aos profissionais e (iii) *Camada de Armazenamento* que contém as classes que manipulam a informação do ponto de vista médico, permitindo inclusive a inclusão de novos pacientes no sistema e responsável pelo armazenamento persistente das informações. Nas seções seguintes cada uma das camadas é descrita com mais detalhes.

O diagrama de classes da arquitetura proposta encontra-se ilustrado na Figura 2. As classes **VizinhosProximos**, **RedesNeurais**, **ArvoreDecisao** e **Probabilistico** são responsáveis pela configuração, aprendizado e classificação das pacientes utilizando os algoritmos IBK, J48, MLP e NB (vide Seção 3.2) e têm dependência com a classe **ConfigurarClassificador**, responsável por intermediar as classes dos classificadores com as classes dos dados médicos. A classe **Paciente** contém informações sobre as pacientes que têm problema de infertilidade. A classe **Funcionario** é responsável em modelar o comportamento dos membros da equipe de profissionais. As classes **Funcionario** e **Paciente** são especializações da classe **Pessoa**; esta contém atributos e métodos comuns a qualquer indivíduo. As classes **ExameImunologico** e **ExameGenetico** descrevem os tipos de exames que são feitos para o tratamento e métodos para manipulação dessas informações e são uma especialização da classe **Exame**. A classe **Anamnese** contém as informações referentes ao tratamento, como fator de infertilidade, número de embriões, entre outros. Por último, a classe **Usuario** controla o acesso ao sistema, por meio de perfis de usuário.

3.1 Camada de Apresentação

A camada de apresentação representa a interface com usuário que contém as seguintes funcionalidades: (i) cadastro das informações de pacientes (informações pessoais, de exames e de anamnese); (ii) recuperação das informações de pacientes; (iii) busca avançada, onde ocorre o cruzamento e relacionamento de todos os dados e (iv) classificação. A interface foi implementada na linguagem de programação Java utilizando a plataforma de desenvolvimento Net-



Figura 1: Arquitetura em três camadas

Beans¹. Nesta camada é possível modificar os parâmetros dos algoritmos de aprendizado, com o objetivo de encontrar os melhores resultados dentre os paradigmas de aprendizados utilizados, por exemplo, como ilustrado na Figura 3, foi escolhido o classificador *Multi Layer Perceptron* pelo usuário e o sistema preencheu as configurações com em seus valores *default*, mas que podem ser alteradas de acordo com a necessidade do especialista. Dessa forma, o usuário pode comparar os resultados obtidos a partir de classificadores diferentes e escolher o que melhor se aplica para a tarefa em questão. Uma funcionalidade que não é vista de forma visual pelos usuários mas que foi implementada na camada de apresentação é o controle de acesso de acordo com perfis de usuários. Cada usuário no sistema tem um papel distinto, o que implica que diferentes níveis de acesso estão associados ao usuário, de forma a garantir que somente usuários com o devido acesso possam ter acesso e/ou modificar as informações armazenadas.

3.2 Camada de Classificação

Na camada de classificação, depois que as informações médicas das pacientes são introduzidas no sistema, o mesmo executa os algoritmos de classificação, a pedido do usuário, para aprender, por exemplo, o sucesso ou não da gravidez e para indicar o melhor tratamento, dentre outros. Uma vez gerado um classificador⁹, a camada de classificação pode prever, a partir de informações médicas do paciente, o tratamento adequado e o sucesso ou não do tratamento, dentre outras informações que o usuário pode selecionar. Este é um aspecto muito im-

¹<http://netbeans.org/>

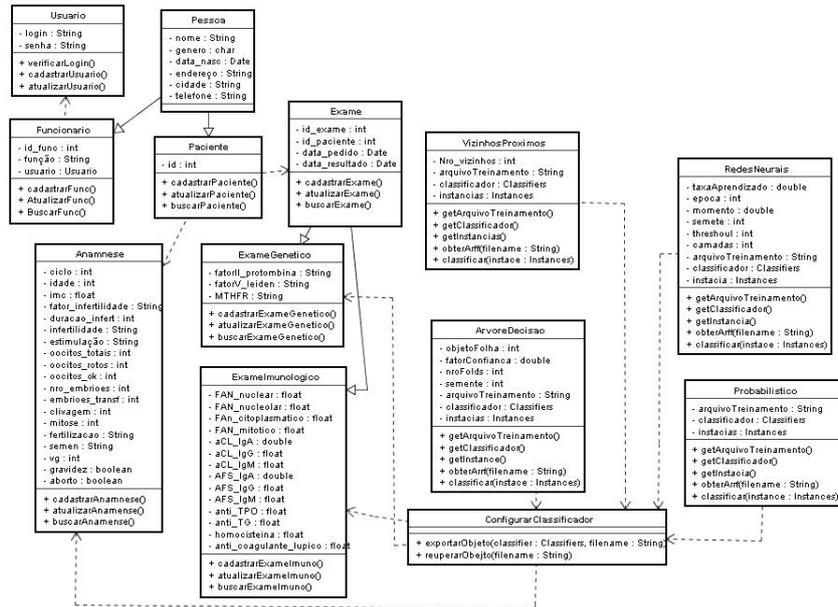


Figura 2: Diagrama de Classes



Figura 3: Interface para configuração dos classificadores

portante, principalmente na área médica, pois os especialistas podem testar hipóteses: selecionando um (sub)conjunto dos atributos contidos na camada de armazenamento e um conceito a ser aprendido (por exemplo, sucesso ou não do tratamento de fertilização) é possível verificar se o (sub)conjunto selecionado gera bons classificadores. Além de permitir testar hipóteses, essa funcionalidade do sistema pode permitir a redução dos custos da reprodução assistida, já que podem ser descartados (não realizados) exames que não se mostrem importantes para o tratamento de uma paciente específica. Esta camada contempla quatro algoritmos de aprendizado de diferentes paradigmas da biblio-

teca Weka¹⁰: IBK (*Instance-Based*) é um indutor preguiçoso (*lazy*), J48 é um indutor de árvores de decisão, o MLP é um indutor de redes neurais e o NB (*naïve Bayes*) é um indutor estatístico muito simples.

No indutor preguiçoso IBK (*Instance-Based*), também conhecido como algoritmo K vizinhos mais próximos (K -Nearest Neighbors ou K -NN)¹¹ a idéia geral é postergar a compilação do conjunto de treinamento, armazenando os exemplos. A classificação de um novo exemplo é efetuada com base no voto dos K exemplos mais próximos utilizando uma métrica de distância¹².

MLP (*Multi Layer Perceptron*) é um modelo de redes neurais artificial que mapeia conjuntos de dados de entrada para um conjunto de saída apropriada, contendo duas ou mais camadas de neurônios. O treinamento é efetuado usando o algoritmo *Backpropagation* e é efetuado em duas fases, cada uma percorrendo a rede em um sentido. Na primeira fase a entrada é apresentada à primeira camada da rede; após os neurônios de uma camada calcularem seus valores de saída, os neurônios da camada posterior utilizam estes valores para definir suas saídas. As saídas produzidas pelos neurônios da última camada são comparadas às saídas desejadas e o erro para cada neurônio da ca-

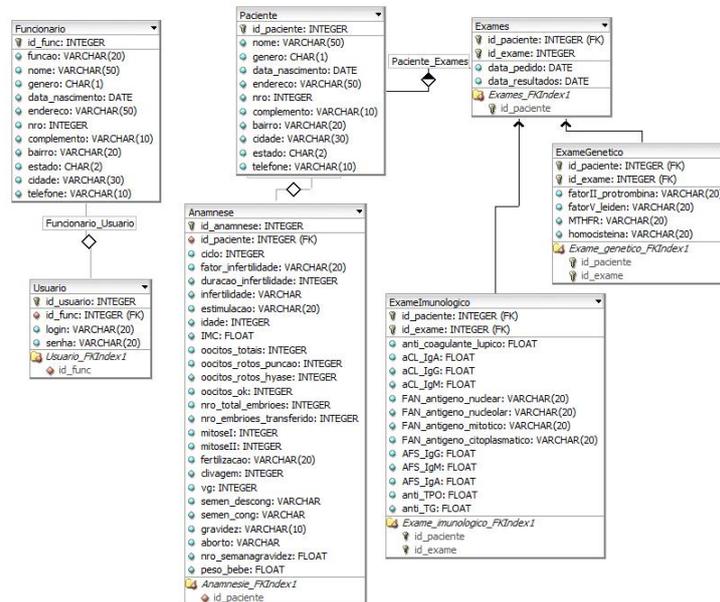


Figura 4: Modelo entidade-relacionamento referente à camada de armazenamento

mada de saída é calculado. Na segunda fase cada neurônio ajusta seus pesos de modo a reduzir seu erro¹³.

Um modelo de grande importância principalmente em domínios nos quais os especialistas humanos precisam confrontar e entender o modelo obtido, tais como aqueles ligados à área médica ou biológica são os classificadores simbólicos. Neste sentido, um classificador simbólico é um modelo cuja linguagem de descrição é equivalente a um conjunto de regras. Um dos classificadores mais populares são, provavelmente, as árvores de decisão¹⁴. Uma árvore de decisão é constituída por dois tipos de nós: (i) nós de decisão, que contêm um teste sobre o valor do atributo ou característica de interesse, que leva a uma sub-árvore; (ii) nós-folha, que indicam a classe correspondente. Novos exemplos são classificados a partir da raiz (o primeiro teste da árvore), descendo através dos nós de decisão até chegar a um nó-folha, que corresponde à classe deste novo exemplo¹⁵. Árvores de decisão de tamanho moderado oferecem uma fácil interpretabilidade de seus resultados para o usuário, característica muito relevante quando se trata de um sistema para auxílio à tomada de decisão médica^{14,16}.

O indutor NB usa a regra de Bayes para calcular a probabilidade de cada classe dado um

exemplo, assumindo que os atributos são independentes¹⁷. Mesmo que, em domínios reais, os atributos não sejam independentes, o algoritmo é bem robusto a violações da condição de independência.

3.3 Camada de Armazenamento

A camada de armazenamento é responsável pela manipulação da informação de forma persistente. Para tanto, um banco de dados foi projetado segundo o Modelo Entidade-Relacionamento conforme a Figura 4; sua modelagem foi feita usando o DBDesigner² e o banco de dados PostgreSQL³. No modelo, a tabela **Paciente** armazena os dados pessoais e residenciais do paciente e tem relacionamento com as tabelas **Anamnese** e **Exames** que armazenam as informações da anamnese e dos exames respectivamente. A tabela **Exames** é uma generalização das tabelas **ExameImunologico** e **ExameGenetico** que armazenam os resultados dos exames. A tabela **Funcionario** armazena os dados pessoais e os profissionais e tem um relacionamento com a tabela **Usuario**, que correspondem aos usuários do sistema. Para a manipulação das informações médicas, o sistema utiliza várias

²<http://www.fabforce.net/dbdesigner4/>

³<http://www.postgresql.org.br/>

classes. As classes **VizinhosProximos**, **RedesNeurais**, **ArvoreDecisao** e **Probabilistico** são responsáveis pela classificação. A classe **Anamese** é responsável pela manipulação dos dados referentes à anamese; por meio dela são feitos o armazenamento, a atualização e a busca dessas informações. A classe **Examelmunologico** é responsável pelo armazenamento, atualização e busca dos dados referentes aos resultados dos exames imunológicos. A classe **ExameGenetico** é responsável pelo resultados dos exames genéticos, podendo armazenar, atualizar e buscar esses dados. As classes **Paciente** e **Funcionario** armazenam dados pessoais de cada paciente e funcionário, respectivamente. A classe **Usuario** é responsável pelo armazenamento das informações para prover o controle de acesso.

4 Conclusões

Neste estudo foi descrita uma arquitetura bem como o sistema computacional que a implemente na área de reprodução assistida, contendo três camadas com funcionalidades distintas. Com isso, os especialistas podem indicar o tratamento adequado para cada paciente bem como o sucesso ou não da gravidez por meio dos resultados dos exames realizados. O módulo de suporte à decisão, implementado na camada de classificação, pode ser utilizado para tentar compreender os padrões de sucesso ou não do processo de fertilização, a partir dos dados de exames, anamnese e terapia. Este é um aspecto muito importante, pois especialistas podem testar hipóteses clínicas, o que pode resultar na redução dos custos da reprodução assistida. Esse estudo está sendo estendido com experimentos para avaliar o desempenho da camada de classificação.

Referências

- 1 Franco JG, Baruffi LR, Mauri AL, Petersen CG. Reprodução Assistida. Revinter; 1997.
- 2 Meola J, Veiga-Castelli LC, Martelli LR. Falando sobre infertilidade; 2009.
- 3 Pollettini JT, Nicolas FP, Panico SRG, Danelluzzi JC, Tinós R, Baranauskas JA, et al. A software architecture-based framework supporting suggestion of medical surveillance level from classification of electronic patient records. In: The 12th IEEE International Conference on Computational Science and Engineering; 2009. p. 166–173.
- 4 Park E, Shin H, Nam HS. Collective Management of Medical Information in a Decision Support System. In: 2009 Int'l Asia Conference on Informatics in Control, Automation and Robotics; 2009. p. 358–360.
- 5 Morales DA, Bengoetxea E, Larrañaga P, García M, Franco Y, Fresnada M, et al. Bayesian classification for the selection of in vitro human embryos using morphological and clinical data. Computer Methods and Programs in Biomedicine. 2008;90:104–116.
- 6 Passmore L, Goodside J, Hamel L, Gonzalez L, Silberstein T, Trimarchi J. Assessing Decision Tree Models for Clinical In-Vitro Fertilization Data;. Available from: <http://homepage.cs.uri.edu/faculty/hamel/pubs/ivf-tr296.pdf>.
- 7 hong Eom A J, chun Kim B S, tak Zhang A B. Available online at www.sciencedirect.com/Expert_Systems_with_Applications; 2007.
- 8 Rajasekaran H, Iacono LL, Hasselmeyer P, Fingberg J, Summers P, Benkner S, et al. @neurIST - Towards a System Architecture for Advanced Disease Management through Integration of Heterogeneous Data, Computing, and Complex Processing Services. Computer-Based Medical Systems, IEEE Symposium on. 2008;0:361–366.
- 9 Monard MC, Baranauskas JA. 4. In: Rezende SO, editor. Conceitos sobre Aprendizado de Máquina. Manole; 2003. p. 89–114.
- 10 Witten IH, Frank E. 2nd ed. Morgan Kaufmann;.
- 11 Mitchell TM. Machine Learning. McGraw-Hill; 1998.
- 12 Aha DW. Lazy Learning. Artificial Intelligence Review. 1997;11:7–10.
- 13 Haykin SS. Redes neurais: princípios e prática. vol. 2; 2001.
- 14 Kingsford C, L SS. What are decision trees? Nat Biotech. 2008;26(9):1011–1013.
- 15 Quinlan JR. C4.5: Programs for Machine Learning. Morgan Kaufmann; 1993.
- 16 Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, et al.. MicroRNAs accurately identify cancer tissue origin, Nature Biotechnology; 2008.
- 17 Heckerman D. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R, editors. Bayesian Networks for Knowledge Discovery. Menlo Park, CA: AAAI; 1996. p. 273–306.