

Identificação de Pacientes Diabéticos com Potencial para Desenvolver o Pé Diabético

Rebeca N. Silva, Vinicius A. Carvalho, Ana Paula R. Borges, Danton D. Ferreira, Bruno H. G. Barbosa
Departamento de Engenharia/Programa de Pós-graduação em Engenharia de Sistemas
Universidade Federal de Lavras
Lavras, MG, Brasil
rebecanonato@posgrad.ufla.br
danton@deg.ufla.br

Ana Cláudia B. H. Ferreira
Programa de Pós-graduação em Enfermagem
Universidade Federal de Juiz de Fora
Juiz de Fora, MG, Brasil

Abstract- Foot complications (diabetic foot) are among the most serious and costly complications of diabetes mellitus. Amputation of all or part of a lower extremity is usually preceded by a foot ulcer. In order to prevent diabetic foot, this paper proposes an automatic system based on the K-means and K-nearest neighbors algorithms to identify patients with diabetes who have a high risk of developing diabetic foot. To design the method, information concerning social scope and self-care of patients was taking into account. Both real and simulated data were used to evaluate the method. Accuracy higher than 90% was achieved showing that the method can be useful at Basic Health Units to triage of diabetic patients helping the healthcare team to reduce the number of cases of diabetic foot.

Keywords — K-means; KNN; Diabetic Foot; Diabetes Mellitus;

I. INTRODUÇÃO

A área da saúde vem se destacando na busca de alternativas para o auxílio ao diagnóstico médico. Para suprir essas alternativas, vê-se necessário o desenvolvimento de sistemas rápidos e precisos no processamento de informações, ou seja, na Mineração de dados. A Mineração de dados vem conquistando seu espaço nos diagnósticos médicos, auxiliando na tomada de decisões dos profissionais de saúde [1]. No Brasil, destacam-se os trabalhos de Santos et al [2] e Souza et al [3]. Santos et al [2] propuseram um sistema para predição da soroprevalência da Hepatite A utilizando modelos de regressão logística e redes neurais artificiais. Os resultados mostram que o modelo neural, aplicado sobre a informação relevante extraída do modelo de regressão logística, apresenta um bom desempenho, alcançando uma eficiência de classificação geral acima de 88%. Souza et al [3] utilizaram processamento neural para auxiliar o diagnóstico médico na detecção de Tuberculose Pulmonar. Este sistema, com base num questionário de sintomas, identifica qual é a chance do paciente ter contraído tuberculose, assim como o classifica em um dentre 3 grupos de risco. Resultados expressivos são

obtidos, atingindo-se uma identificação de pacientes doentes de 100%; e de não-doentes, de 80%.

Na área de Diabetes Mellitus, destacam-se os trabalhos Patil et al em 2010 [4], que propuseram um modelo híbrido de predição, utilizando o algoritmos *K-means*, afim de validar rótulos, e posteriormente aplica o algoritmo de classificação C4.5 [5] (algoritmo utilizado para gerar uma árvore de decisão) que constrói um modelo final que classifica a base de dados dos índios Pima, referentes a índios no Arizona, EUA [6], onde mais da metade da população adulta são portadores de diabetes. O objetivo do trabalho foi investigar como os incidentes de diabetes são afetados pelas características e medidas dos pacientes. Na Coreia, Lee et al [7] realizou um trabalho que desenvolve um Sistema de Monitoramento e Assessoria de Gestão em Pacientes Diabéticos utilizando o método baseado em regras e o algoritmo *K-Nearest Neighbor* (KNN). Esse sistema fornece um tratamento adequado para os pacientes diabéticos, de acordo com seu nível de açúcar no sangue. Kargowda et al [8] apresenta um modelo híbrido que classifica o banco de dados de diabetes dos índios PIMA, Arizona, referente a índios portadores de diabetes, como discutido anteriormente. Esse trabalho utiliza o algoritmo *k-means* para identificar e eliminar casos classificados incorretamente. Em seguida, Algoritmos Genéticos são empregados para selecionar variáveis e, por fim, o KNN é utilizado. Antonelli et al [9] apresenta uma análise de dados que identifica uma determinada doença através de exames realizados pelos pacientes. É utilizado o DBSCAN (*Density-based spatial clustering of applications with noise*) um algoritmo baseado em densidade, além da representação dos dados de exames de pacientes no espaço do modelo vetorial (VSM) utilizando o método de TF-IDF [10], conhecido como frequência de documento de frequência inversa. Esse estudo utilizou a base de dados de pacientes diabéticos fornecidos pelo Centro Nacional de Saúde (CNS) da província de Asti, Itália.

Os trabalhos anteriormente citados utilizam sistemas de reconhecimento de padrões e evidenciam o sucesso da aplicação de métodos inteligentes em parceria com os

profissionais de saúde, de forma a apoiá-los, dando assistência nas tomadas de decisão e fazendo a triagem de pacientes.

De acordo com a Organização Mundial de Saúde (OMS) o Diabetes Mellitus (DM) é considerado a nona causa de morte no mundo obtendo um percentual de 2,2% [11]. O DM é um grupo de doenças metabólicas caracterizadas por hiperglicemias e associadas a complicações e disfunções de vários órgãos [12]. Dentre essas complicações sérias e dispendiosas que afetam os indivíduos com diabetes, destacam-se as complicações com os pés. As úlceras nos pés, que são representadas por um estado fisiopatológico multifacetadas, assimiladas pelo aparecimento de lesões que, conseqüentemente, levam à neuropatia em 80-90% dos casos [13]. Essas úlceras caracterizam-se por lesões cutâneas com perda do epitélio, estendendo até a derme, atravessando-a até chegar aos tecidos profundos, assim podendo atingir o músculo precedendo em 85% das amputações [14]. As úlceras nos pés além de acarretar sérios problemas na saúde física do paciente impactam na qualidade de vida do mesmo. Estudos realizados demonstraram que pacientes com diabetes e pés com úlcera apresentavam quadros mais deprimidos e qualidade de vida baixa [15]. Este quadro clínico pode ser evitado, mediante alguns cuidados. O reconhecimento precoce do pé em risco às úlceras e a prontidão do cuidado destas são primordiais para diminuir o impacto da doença. É responsabilidade dos profissionais de saúde realizar esse reconhecimento precoce, porém nem sempre é cumprida, ocasionando assim amputações em pacientes que não realizaram exames completos ou nunca haviam recebido informações necessárias [16].

A preocupação com o pé diabético vem aumentando no Brasil. Especialistas ressaltam a atenção especial com os pacientes diabéticos, que nem sempre a neuropatia diabética manifesta sintomas e um diagnóstico precoce pode ser a chave para evitar danos aos pés. Ainda é enfatizado que o quadro clínico do paciente vai piorando com o tempo: quando a um diagnóstico precoce, o controle glicêmico pode reverter a situação. Caso não seja tomada nenhuma atitude, pode ser necessária a amputação de dedos ou dos pés [17].

Diante desse contexto, emerge a possibilidade de um sistema para o auxílio ao diagnóstico precoce do pé diabético, bem como mensurar o nível de conhecimento que este paciente possui a respeito do autocuidado com os pés, torna-se importante o desenvolvimento de metodologias que possam atuar mais precisamente neste sentido, permitindo assim a prevenção do pé diabético.

O algoritmo de classificação baseado no vizinho mais próximo (*Nearest Neighbor* – NN) é uma técnica amplamente empregada para reconhecimento de padrões [18]. Introduzido por Aha, D. em 1991 [19], a ideia geral do algoritmo *k-Nearest Neighbor* (KNN) [20] consiste em encontrar os *k* eventos de clusters classificados mais próximos dos eventos não classificados em clusters [21].

Neste contexto, o objetivo do presente trabalho é desenvolver um sistema automático utilizando o algoritmo KNN para a identificação precoce de pacientes portadores de diabetes com potencial para desenvolver o pé diabético. O sistema proposto processará as informações coletadas e

apresentará como saída o enquadramento do paciente portador do diabetes em, basicamente, dois grupos de risco de desenvolver o pé diabético: i) alto risco e ii) baixo risco. Este sistema permitirá o controle automático e sistemático do problema do pé diabético por parte dos agentes de saúde e a atuação dos mesmos diretamente na comunidade de pacientes portadores do diabetes, fazendo a triagem dos pacientes com alto risco de desenvolver o pé diabético.

Na próxima seção a base de dados é apresentada. Na Seção III o projeto e a operação do método proposto são detalhados. A Seção IV apresenta os resultados alcançados pelo método proposto e discussões. As conclusões e propostas de continuidade do trabalho são apresentadas na Seção V.

II. BASE DE DADOS

Realizou-se um levantamento de dados de pacientes diabéticos em uma cidade do interior de Minas Gerais, Brasil. Esses dados foram colhidos através de um formulário contendo 31 perguntas a respeito do âmbito social e principalmente sobre os cuidados com os pés dos pacientes. A Tabela I mostra, resumidamente, as informações coletadas com as 31 perguntas. Este formulário foi previamente aprovado pelo Comitê de Ética e autorizado pela Secretaria de Saúde da cidade. As informações foram coletadas durante visitas domiciliares por um profissional devidamente treinado para tal e um representante do PSF (Posto de Saúde Familiar). Foram entrevistados um total de 153 pacientes portadores de diabetes. Esses dados foram usados no projeto do método.

TABLE I. INFORMAÇÕES COLETADAS

Sexo	O que usa pra remover calos
Idade	Usa bolsa de água quente
Estado Civil	Retira cutículas dos pés
Ocupação (trabalha ou não trabalha)	Verifica o calçado por dentro antes de usar
Escolaridade	Tipo de calçado que usa
Tempo do diagnóstico de Diabetes	Atitude quando percebe alteração nos pés
Diabetes Tipo 1 ou 2	Aspecto interno do calçado
Tempo de tratamento do Diabetes	Hora do dia que sai pra comprar sapatos novos
Tipo de tratamento	Costuma andar descalço
Hábito de examinar os pés	Tipo de meia que usa (material)
Como corta as unhas dos pés	Tipo de meia que usa (aspecto e costura)
O que usa pra lavar os pés	Posição ao assistir televisão
Hábito de lavar os pés	Material do calçado que usa
O que usa pra enxugar os pés	Já percebeu alteração nos pés
Enxuga entre os dedos dos pés	Já foi diagnosticado com diabetes
Usa creme hidratante nos pés	

Após a coleta dos dados via formulários, as informações foram analisadas por especialista da saúde a fim de identificar informações incoerentes, que quando encontradas foram corrigidas. As informações foram então codificadas seguindo um padrão lógico definido pela equipe da área de saúde.

Com o objetivo de validar o método proposto, foram simuladas informações relativas a 30 pacientes. Estas informações foram geradas de acordo com os especialistas da área de saúde. Na geração destas informações, buscou-se garantir que 15 pacientes apresentassem alto risco de desenvolver o pé diabético e 15 apresentassem baixo risco.

III. MÉTODO PROPOSTO

O projeto do método proposto pode ser dividido em etapas, conforme ilustra o diagrama em blocos da Fig. 1 (a). A primeira etapa é a coleta das informações e formação do banco de dados (codificação das informações) que é realizada conforme descrito na Seção II.

A segunda etapa, de pré-processamento, é responsável por normalizar as variáveis de entrada referentes a idade dos pacientes, tempo (em anos) em que o paciente é diabético e tempo (em anos) em que o paciente trata a diabetes. Essa normalização foi realizada de acordo com (1). Não foi aplicado normalização nas demais variáveis uma vez que a codificação das informações foi feita garantindo-se valores entre -1 e 1.

$$\mathbf{z}_i = \frac{\mathbf{x}_i - \mu_i}{\max(\mathbf{x}_i) - \mu_i}, \quad (1)$$

em que \mathbf{x}_i e μ_i são, respectivamente, o paciente e a média do paciente.

Ainda na etapa de pré-processamento, a correlação linear entre as 31 variáveis de entrada foi aplicada com o objetivo de identificar redundância entre as mesmas. Observou-se que a variável referente ao tempo em que o paciente foi diagnosticado como paciente diabético apresentou uma correlação de 0,95 com a variável referente ao tempo em que o paciente trata o problema de diabetes. Dessa forma, eliminou-se a variável referente ao tempo em que o paciente trata o problema de diabetes, e passou-se a trabalhar com 30 variáveis. As demais correlações foram inferiores a 0,40 e, portanto, as demais variáveis foram mantidas.

Após a etapa de pré-processamento os dados são agrupados em dois e três grupos (clusters). A clusterização é realizada pelo método *K-means* [22], escolhido por ser prático e eficiente [23]. O método *k-means* visa à partição de N_i observações em k clusters. Cada observação é classificada em determinado cluster, através da menor distância euclidiana para um dos centroides. Os centroides se posicionam no centro dos clusters baseados na menor distância entre si e na maior distância entre os dados, e são fundamentais na formação dos grupos [24].

A próxima etapa do projeto do método proposto consiste em avaliar os clusters obtidos. Para isso utilizou-se um indicador de qualidade entre homogeneidade intra-cluster e inter-cluster, conhecido por *Silhouettes* [25]. O *silhouettes* é um método de interpretação e validação de conjunto de dados através de um gráfico que representa como cada objeto encontra-se dentro de seu cluster. O valor de cada *silhouettes* é calculado conforme (2).

$$S_i = \frac{b_i - a_i}{\max(b_i - a_i)}, \quad (2)$$

em que, a_i é a distância média dos pacientes a partir dos outros pacientes dentro do mesmo cluster, e b_i é a menor distância média de seus vizinhos. A equação (2) pode ser descrita como

$$S_i = \begin{cases} 1 - a_i / b_i & \text{se } a_i < b_i, \\ 0 & \text{se } a_i = b_i, \\ a_i / b_i - 1 & \text{se } a_i > b_i, \end{cases} \quad (3)$$

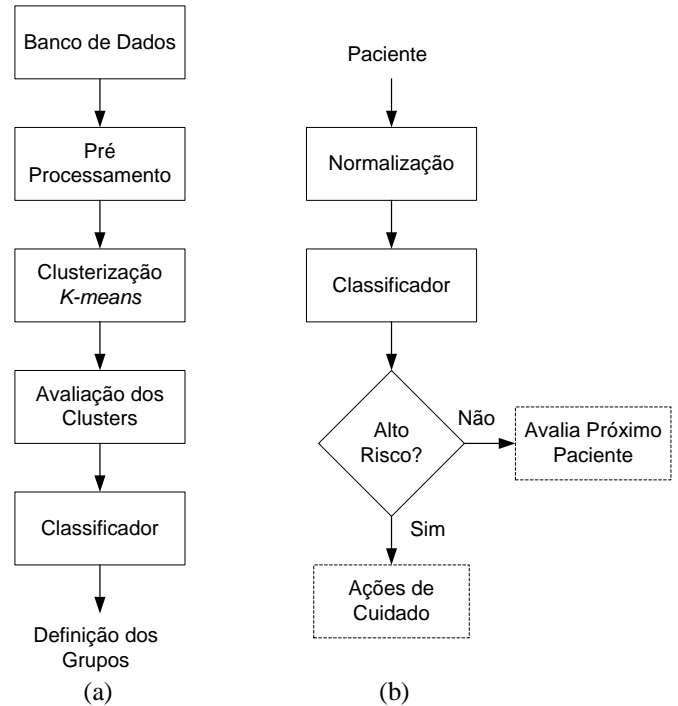


Fig. 1. Método proposto: (a) Etapas do projeto e (b) Operação.

Após a avaliação dos clusters, utilizam-se os centroides construídos pelo algoritmo *k-means* para a classificação dos pacientes do conjunto de dados de validação a partir da distância euclidiana do vetor de informações do paciente até os centroides dos clusters. A distância menor indica a qual cluster o paciente pertence. A utilização dos centroides na classificação leva a uma simplificação na implementação do classificador, visto que apenas duas distâncias euclidianas serão calculadas neste caso.

A fim de testar outros classificadores, o algoritmo KNN também foi empregado para realizar a classificação dos pacientes do conjunto de dados de validação e, de acordo com o enquadramento destes pacientes nos clusters pré-definidos. Para a aplicação do algoritmo KNN foram utilizados 5 ($k = 5$) vizinhos mais próximos e a distância Euclidiana como métrica.

A fase operacional do método proposto está ilustrada na Fig. 1 (b) e consiste em normalizar os dados e classificá-los de acordo com os clusters definidos pelo algoritmo *k-means* na fase de projeto (Fig. 1 (a)). O resultado da classificação enquadra o paciente em um dos dois grupos de risco: (i) alto ou (ii) baixo. Se o paciente for classificado como baixo risco, avalia-se um novo paciente. Se o paciente for classificado como alto risco, devem-se tomar ações de cuidados para a prevenção do pé diabético. Estas ações são definidas pela equipe de saúde local.

IV. RESULTADOS E DISCUSSÕES

As Fig. 2 e 3 mostram, respectivamente, os gráficos *silhouettes* obtidos para o agrupamento realizado pelo *k-means* considerando-se $k=2$ e $k=3$ clusters. O gráfico apresenta os clusters (eixo horizontal) e a quantidade de eventos que cada cluster possui, seguido pelas disposições do grupo representado pelo valor da *silhouettes* (eixo vertical), conforme (3). Maiores valores de *silhouettes* apontam para clusters mais ‘fortes’. Ambos os clusters da Fig. 2 apresentam valores mais altos de *silhouettes*, enquanto que os clusters da Fig. 3 (principalmente os clusters 1 e 2) apresentam poucos valores altos de *silhouettes*. Isso mostra que os clusters da Fig. 2 estão mais separados entre seus vizinhos do que os da Fig. 3, o que é um indicativo de que 2 clusters é melhor do que 3, neste caso. Por outro lado, o cluster 2 da Fig. 2 apresenta valores negativos de *silhouettes*, o que indica que este cluster não está bem separado.

Uma forma quantitativa de avaliar os clusters, ou seja, escolher o número de clusters, usando o método *silhouettes*, é utilizar a comparação entre as médias das *silhouettes* (média de S_i em (2)) obtidas para cada caso. A Tab. II mostra essa comparação para $k = 2, 3, 4$ e 5 clusters. Observa-se que o valor da *silhouettes* diminui com o aumento do número de clusters. O gráfico para $k=4$ e 5 foi não é apresentado por simplificação. A partir destes resultados, apenas as abordagens para $k=2$ e 3 foram consideradas.

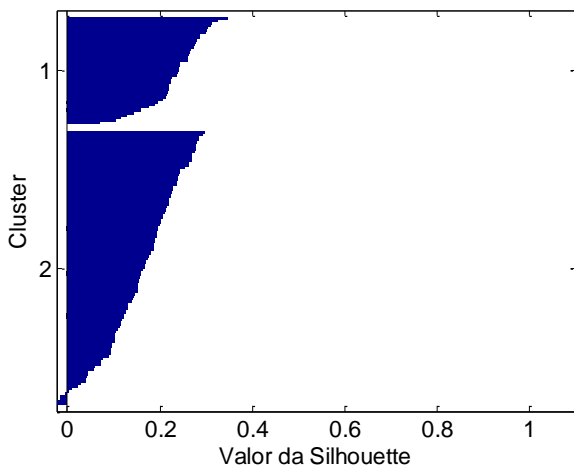


Fig. 2. Gráfico *Silhouettes* para $k=2$ clusters.

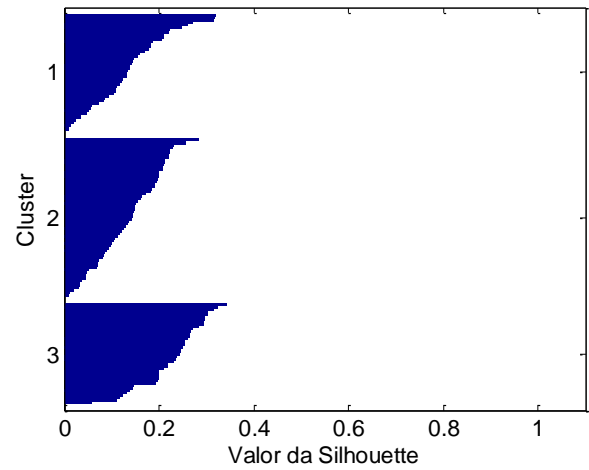


Fig. 3. Gráfico *Silhouettes* para $k=3$ clusters.

TABLE II. ESCOLHA DO NÚMERO DE CLUSTERS

Número de Clusters k	Média das <i>Silhouettes</i>
2	$0,18 \pm 0,01$
3	$0,16 \pm 0,01$
4	$0,14 \pm 0,01$
5	$0,12 \pm 0,02$

Após a definição do número de clusters, os classificadores (baseado nos centroides definidos pelo *k-means* e o algoritmo KNN) foram aplicados aos dados simulados. Para o algoritmo KNN, foram utilizados os 153 dados reais agrupados pelo *k-means* como base de treinamento. As Tabelas III e IV apresentam os resultados da classificação dos dados simulados, considerando-se, respectivamente, 2 e 3 clusters.

A Tabela III mostra que, considerando-se apenas 2 clusters, ambos os classificadores utilizados não foram capazes de separar o grupo de pacientes de alto risco do grupo de pacientes de baixo risco (dados simulados). Já na abordagem com 3 clusters, ambos os classificadores alcançaram um bom desempenho, em que 14 dos 15 pacientes de alto risco foram classificados como pertencentes ao Grupo 1 utilizando-se os centroides como classificadores, e 13 dos 15 pacientes de alto risco foram classificados como pertencentes ao Grupo 1 pelo KNN. Os pacientes de baixo risco foram divididos entre os Grupos 2 e 3 para ambos os classificadores.

Com essa classificação, assume-se que os pacientes (do banco de dados reais) enquadrados no Grupo 1 são de alto risco e os pacientes enquadrados nos Grupos 2 e 3 são de baixo risco de desenvolver o pé diabético. Assim, com esta abordagem detectou-se que 48 pacientes dos 153 foram enquadrados no Grupo de alto risco e, portanto, merecem maior atenção da equipe de saúde de forma a evitar o desenvolvimento do pé diabético.

TABLE III. CLASSIFICAÇÃO (2 CLUSTERS)

Classificação Dados Simulados para 2 Grupos		
CENTRÓIDES		
	Grupo 1	Grupo 2
Alto Risco	9	6
Baixo Risco	7	8
KNN		
	Grupo 1	Grupo 2
Alto Risco	11	4
Baixo Risco	6	9

TABLE IV. CLASSIFICAÇÃO (3 CLUSTERS)

Classificação Dados Simulados para 3 Grupos			
CENTRÓIDES			
	Grupo 1	Grupo 2	Grupo 3
Alto Risco	14	0	1
Baixo Risco	0	7	8
KNN			
	Grupo 1	Grupo 2	Grupo 3
Alto Risco	13	0	2
Baixo Risco	0	7	8

Computacionalmente, o emprego dos centroides leva a um classificador mais simples, pois requer apenas o cálculo de duas distâncias Euclidianas para dar o resultado da classificação de um paciente. Por outro lado, o algoritmo KNN requer o cálculo da distância Euclidiana do paciente analisado até todos os dados do conjunto de treinamento para definir qual classe está mais frequente nos vizinhos mais próximos, tornando-o computacionalmente mais complexo.

A Tabela V mostra os resultados de classificação alcançados pelas duas abordagens (2 clusters e 3 clusters) em termos de sensibilidade, especificidade e acurácia dos classificadores empregados. Observa-se, claramente, que a abordagem com 3 clusters leva a resultados mais expressivos com uma acurácia de 97% e 93% para os classificadores baseado em centroides e KNN, respectivamente.

TABLE V. DESEMPENHO DO CLASSIFICADOR

Classificador	Sensibilidade	Especificidade	Acurácia
CENTROIDES			
2 Clusters	57%	56%	57%
3 Clusters	93%	100%	97%
KNN			
2 Clusters	73%	53%	63%
3 Clusters	87%	100%	93%

V. CONCLUSÃO

A identificação oportuna de casos do pé diabético favorece o estabelecimento de vínculos entre os pacientes e as Unidades

Básicas de Saúde, o que é imprescindível para o sucesso do controle da diabetes.

Entre as vantagens do acompanhamento e controle do diabetes no âmbito da atenção básica estão a possibilidade de evitar o surgimento de complicações, assim como de evitar o agravamento destas ocorrências, reduzindo tanto o número de internações hospitalares quanto a mortalidade relacionada às doenças cardiovasculares, entre outras complicações. É consenso geral entre endocrinologistas e infectologistas que a principal medida no tratamento do pé diabético é a detecção precoce, alcançando-se mais de 90% de sucesso para as úlceras que recebem manejo adequado, incluindo alívio da pressão local, tratamento das infecções e revascularização quando indicada.

Este trabalho propôs um sistema automático para identificar pacientes diabéticos que possuem alto risco de desenvolver o pé diabético. Foram apresentados dois classificadores: i) baseado nos centroides fornecidos pelo algoritmo *K-means* e ii) o algoritmo KNN. Foram utilizados dados reais coletados em Unidades Básicas de Saúde e dados simulados. Os métodos alcançaram resultados satisfatórios, apresentando acurácia superior a 90%. Para a obtenção destes resultados foram levadas em consideração apenas informações sobre o âmbito social e de autocuidado do paciente.

Como propostas futuras os autores pretendem incorporar às variáveis de entrada do sistema, informações sobre o nível de glicose do paciente e histórico de exames realizados ao longo do ano. O sistema será então implementado em Unidades Básicas de Saúde para ser testado de forma mais efetiva. Nesta fase, um ponto importante será verificar se um paciente uma vez classificado com alto risco migrará para o grupo de baixo risco após o acompanhamento sistemático da equipe de saúde.

AGRADECIMENTOS

Os autores gostariam de agradecer às Unidades Básicas de Saúde da cidade onde foram coletados os dados e à FAPEMIG e CAPES pelo suporte.

REFERÊNCIAS

- [1] F. Reynol, "Mineração de dados para diagnósticos médicos", Disponível em < <http://agencia.fapesp.br/11928> > Acessado em 24 de março de 2013.
- [2] M. Santos, M. Seixas, B. Pereira, R. Medronho, "Usando Redes Neurais Artificiais e Regressão Logística na predição da hepatite A", *Rev Bras Epidemiol*, Maranhão, vol.8, pp.117-126, 2005.
- [3] M. Santos, M. Seixas, B. Pereira., Q. Mello, A. Kritski, "Arvore de Classificação e Redes Neurais Artificiais: Uma Aplicação a Predição de Tuberculose Pulmonar", VI Congresso Brasileiro de Redes Neurais; 2003 junho 2-5; Centro Universitário da FEI. São Paulo: 2003.
- [4] B. Patil , R.Joshi, D. Toshniwal, "Hybrid prediction model for Type-2 diabetic patients", *Expert Systems with Applications*. India, vol.37 pp.8102-8108, 2010.
- [5] J. Quinlan, "C4.5 programas de aprendizagem de máquina. San Mateo", CA: Morgan Kaufmann Publishers, 1993.
- [6] J. Ekoé, M. Rewers, R. Williams, P. Zimmet, "The epidemiology of diabetes mellitus", 2.ed. Oxford:Wiley-Blackwell, 2008.
- [7] M. Lee, T. Gatton, K. Kwang, "A Monitoring and Advisory System for Diabetes Patient Management Using a Rule-Based Method and KNN", *Sensors*, vol. 20, pp. 3934-53, Janeiro de 2010.

- [8] A. Karegowda, M. Jayaram, A. Manjunath, "Cascading K-means Clustering and K-Nearest Neighbor Classifier for Categorization of Diabetic Patients", *International Journal of Engineering and Advanced Technology*, vol.1, pp.147-151, fevereiro de 2012.
- [9] D. Antonelli, E. Baralis, G. Bruno, T. Cerquitelli, S. Chiusano, N. Mahoto, "Analysis of diabetic patients through their examination history", *Expert Systems with Applications, Italia*, 2013. Não Publicado
- [10] T. Pang-Ning, M. Steinbach, V. Kumar, "Introdução à mineração de dados", Addison-Wesley, 2006.
- [11] NEWS.MED.BR, 2011. "OMS divulga as dez principais causas de morte no mundo". Disponível em: <<http://www.news.med.br/p/saude/222530/oms-divulga-as-dez-principais-causas-de-morte-no-mundo.htm>>. Acesso em: 22 fev. 2013.
- [12] Brasil. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica, 2006. Disponível em <http://bvsms.saude.gov.br/bvs/publicacoes/diabetes_mellitus.PDF>. Acessado em 22 fev 2013.
- [13] M. Laurindo, C. Recco, B. Roberti, C. Rodrigues, "Conhecimento das pessoas diabéticas acerca dos cuidados com os pés", *Arq. Ciência Saúde. São Paulo*, vol.12, pp.80-84, abril de 2005.
- [14] S. Martins, A. Beraldo, M. Passeri, M. Freitas, A. Pace, "Causas referidas para o desenvolvimento de úlceras em pés de pessoas com diabetes mellitus", *Acta Paul Enferm. São Paulo*, vol.25, pp.218-24, setembro de 2012.
- [15] M. Salomé, L. Blanes, M. Ferreira, "Avaliação de sintomas depressivos em pessoas com diabetes mellitus e pé ulcerado", *Rev. Col. Bras. Cir. São Paulo*, vol.38, pp.327-333, 2011.
- [16] C. Farjado, "A importância do cuidado com o pé diabético: ações de prevenção e abordagem clínica", *Revista Brasileira Medicina da Família e Comunidade, Rio de Janeiro*, vol.2, pp.43-58, abril de 2006.
- [17] C. Dissat, P. Rodrigues, "Especial: Pé Diabético", *Diabetes A Revista da SBD*, vol.20, pp.11-14, Janeiro de 2013
- [18] C. Ferrero, M. Monard, "Algoritmo KNN para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia", *Dissertação de Mestrado, São Paulo*, janeiro de 2009.
- [19] D. Ahah, D. Kibler, M. Albert, "Instance-based learning algorithms", *Machine Learning* 6, pp. 37-66, 1991.
- [20] "Classification using nearest neighbors", Disponível em <<http://www.mathworks.com/help/stats/classification-using-nearest-neighbors.html?searchHighlight=knn>>, Acessado em: 24 de março de 2013.
- [21] F. Gagliardi, "Instance-based classifiers applied to medical databases: Diagnosis and knowledge extraction", *Inteligência Artificial em Medicina, Itália*, vol.52, pp. 123-139, Julho de 2011.
- [22] J. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations". *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, pp. 281-297, 1967.
- [23] P. Lloyd, "Least square quantization in PCM". *Bell Telephone Laboratories Paper*, 1957. Published in journal much later: Lloyd., S. P. "Least squares quantization in PCM". *IEEE Transactions on Information Theory*, 1982.
- [24] K-Means Clustering, Disponível em <http://www.mathworks.com/help/stats/k-means-clustering.html#bq_679x-19>, Acessado em : 24 de março de 2013.
- [25] P. Rousseeuw, " *Silhouettes*: A graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics*, vol.20, pp.53-65, 1987.