Non-invasive method to analyse the risk of developing diabetic foot

Rebeca N. Silva¹, Ana C.B.H. Ferreira², Danton D. Ferreira¹, Bruno H.G. Barbosa¹

¹Engineering Department, Federal University of Lavras, Lavras, Minas Gerais 3030/37200-000, Brazil

²School of Nursing, Federal University of Juiz de Fora, Juiz de Fora, MG 36036-900, Brazil

E-mail: danton@deg.ufla.br

Published in Healthcare Technology Letters; Received on 20th June 2014; Revised on 14th October 2014; Accepted on 14th October 2014

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. To prevent diabetic foot, an automatic non-invasive method to identify patients with diabetes who have a high risk of developing diabetic foot is proposed. To design the proposed method, information concerning social scope and self-care of 153 diabetic patients was presented to the K-means clustering algorithm, which divided the data into two groups: high risk and low risk of developing diabetic foot. In the operational stage, the Euclidian distance from the information vector to the centroids of each group of risk is used as criterion for classification. Both real and simulated data were used to evaluate the method in which promising results were achieved with accuracy of 0.97 ± 0.06 for simulated data and 0.68 ± 0.16 considering the classification of specialists as the gold standard for real data. The method requires a simple computational processing and can be useful for basic health units to triage diabetic patients helping the health-care team to reduce the number of cases of diabetic foot.

1. Introduction: Many studies have been recently developed to seek alternatives to help make medical diagnoses. To address these alternatives, developing fast and accurate algorithms for information processing (data mining) becomes necessary. Data mining is conquering its space in medical diagnostics, aiding health professionals in decision making [1].

In terms of research in diabetes mellitus, it is important to mention the works described in [2–5]. Lee *et al.* [2] conducted a study to develop a system for monitoring and management advisory in diabetic patients using the rule-based method and the K-nearest neighbour algorithm. Antonelli *et al.* [3] presented an analysis of data that identifies a particular disease through tests performed by diabetic patients (examination history). Appuhamy *et al.* [4] proposed a mathematical framework for predicting diabetes prevalence using incidence rates estimated within the model using body mass index data. Espeland *et al.* [5] exploited principal component analysis to describe patterns of weight changes in response to intensive lifestyle intervention in individuals with type-2 diabetes.

The studies aforementioned used pattern recognition systems and demonstrated the successful application of intelligent methods in partnership with health professionals to support them, assisting in decision making and classifying the patients.

According to the World Health Organization, diabetes mellitus is considered to be the eighth cause of death in the world, with a percentage of 2.6% in 2011 [6]. Foot complications are among the most serious and costly complications that affect individuals with diabetes, representing a major health-care burden with significant morbidity [7]. According to Alexiadou and Doupis [8], diabetic foot ulcers are defined as: neuropathic in the presence of peripheral diabetic neuropathy and absence of ischaemia; ischaemic if the patient presents peripheral artery disease but no diabetic peripheral neuropathy; and neuroischaemic if neuropathy and ischaemia coexist. These ulcers are characterised by skin lesions with loss of epithelium, extending into the dermis, passing through it till reaching deeply into tissues, and thus reaching muscle proceeding in 85% of amputations [9].

Studies have shown that patients with diabetes and foot ulcer are more likely to develop depression and have low quality of life [10, 11]. Early detection of foot ulcers and readiness of caring them are paramount to lessen the impact of the disease.

On the other hand, simple actions can be effective to prevent diabetic foot complications. In [12], a reduction in the foot

amputation rate from 0.8 to 0.5% is shown when appropriate diabetic foot care and education are implemented. According to Shearman and Pal [13], the right foot care, good blood glucose control and diabetes education can prevent up to 85% of diabetic foot amputations. It is evident from the current literature that foot care knowledge as well as practice are the key points to prevent diabetic foot ulcers [11, 14–16].

According to Jirkovská *et al.* [17] and Monteiro-Soares *et al.* [18], prevention through identification of high-risk individuals represents the most effective way of reducing the rate of ulcer formation and leg amputation in diabetic patients.

In this Letter, we focused on patient information about the social characteristics, habits and mainly self-care to develop an automatic non-invasive method for the early identification of diabetic patients with high risk to develop diabetic foot. For this, we have exploited hidden similarities between the data by using an unsupervised approach. The proposed method processes the collected information and provides an output of the framework of diabetic patients in two groups of risk in developing diabetic foot: (i) high risk and (ii) low risk.

Regarding risk stratification systems for diabetic foot ulcers, the work reported in [18] presents a good systematic review, in which five stratification systems were discussed and compared. The authors have concluded that the quality of evidence for these systems is low, as little validation of their predictive ability has been done. Actually, among the five stratification systems discussed in [18], only the system proposed in [19] was validated over time. Unfortunately, the validation of such systems is considered as an obstacle by the authors, as it demands a large number of patients who have developed or not developed diabetic foot in a certain period of time. Therefore the problem of how to select the best system for widespread implementation is still open [18].

The innovation of the proposed method is the usage of non-invasive features (avoiding clinical exams) which produces the final classification pointing out the developing diabetic foot risk of a given patient according to his or her self-care actions and habits. In addition, the application of unsupervised-based machine-learning techniques for developing the method is, to the best of our knowledge, innovative for this field. For instance, methods in [19, 20] used statistical models like univariate and multivariate analysis and cumulative risk. The method in [21] used the 10 g Semmes-Weinstein monofilament test, tuning fork and cotton

Table 1 Collected information and population characterisation

Questions	Statistics
sex	68.5% female
age (years)	64.6 ± 13.8
marital status	62.5% married
occupation (working or not working)	86.9% retirees
education	95.0% illiterate or <8 years of schooling
time from diagnosis of diabetes (years)	11.7 ± 10.5
diabetes type 1 or 2	type 2 (91.5%)
time of treatment of diabetes (years)	11.6 ± 9.7
type of treatment	insulin (5.9%) and oral medication (30.7%)
habit of examining your feet	67.5%
how to cut toenails	into the corners and too short (47.5%)
what do you use to wash the feet	ordinary soap (97.4%)
habit of washing your feet	daily (83.0%)
what do you use to wipe your feet	common towel (83.6%)
wipe between the toes	76.0%
use moisturiser on feet	42.0%
what do you use to remove callus	chemical (16.4%)
use a hot water bottle	34.0%
removes cuticles feet	45.0%
checks the footwear inside before using	65.0%
type of shoe used	open shoe (65.0%)
attitude change when he realises the feet	65.0% seek medical or nurse
inner aspect of the footwear	seamless (52.5%)
time of day out to buy new shoes	any time (84.3%)
usually barefoot	75.0% never barefoot
type of socks that uses (material)	cotton (86.3%)
type of socks that uses (appearance and sewing)	seamless and clear (37.9%)
position to watch television	lying (45.0%)
material of the shoe that uses	synthetic leather (40.0%)
noticed changes in the feet	30.0%
already been diagnosed with diabetic foot	7.5%

Data are mean ± standard deviation or %.

wisp. The methods in [22, 23] used a number of risk factors in the patient to define the level of risk based on the Scottish Intercollegiate Guideline Network and American Diabetes Association, respectively.

The proposed approach has three advantages: (i) it is a non-invasive method; (ii) the collected information of the patients (as required by the method) may provide an overview of the diabetic population which could improve how primary health-care teams administer health-care actions, and guide diabetic patients; (iii) this facilitates the work of the health-care team in monitoring the diabetic patients by forming a triage, which is very useful, especially for large diabetic populations, in particular in rural and remote areas.

2. Database: We conducted a survey of data from diabetic patients in a city in Minas Gerais, Brazil. These data were collected through a questionnaire containing 31 questions about the social condition, habits and foot care among patients (Table 1). This form was approved by the ethics committee (no. 51950). Data were collected during home visits by a nurse from October to December 2012. We interviewed a total of 153 randomly selected patients with diabetes, who are 18 years old or older, from four health basic units.

Figure 1 Proposed method

a Design stage

b Operation stage

After collecting the data, the information was analysed to identify inconsistencies or outliers. The information was then coded, in which the variables ranged between -1 and 1 , except those related to age or time (years). The information was coded following a logical pattern so that answers to the questions in Table 1 that present risk factors were coded as 1 and other answers as -1 . For questions with more than two possible answers, values between -1 and 1 were allocated.

3. Method: The design stages of the proposed method are shown in Fig. 1a. The first step is collecting the information and forming a database. The second stage (preprocessing) is responsible for normalising the input variables related to time (years). This normalisation was performed according to

$$z_i = \frac{x_i - \mu_i}{\max(x_i) - \mu_i} \quad (1)$$

in which x_i and μ_i are the vector of variable i and its average, respectively.

The linear correlation between the 31 input variables was applied to identify redundancy between them. It was observed that the variable referring to the time that the patient was diagnosed as diabetic showed a correlation of 0.95 , with the variable referring to the time the patient treats the diabetes problem. Thus, the latter was eliminated, with 30 variables remaining. The other correlations were less than 0.40 , and therefore, the other variables were kept.

After the preprocessing stage, the data were grouped into two and three groups (clusters). The clustering was performed through the K-means method [24] because of practicality and efficiency. The K-means aims at partitioning N observations into k clusters. Each observation is classified into the cluster that achieved the smallest Euclidean distance between its centroid and the observation vector.

The next step is the cluster evaluation. For this, we used an indicator of quality between intracluster homogeneity and intercluster separation, known as silhouette index [25]. Silhouette is a method that represents how each object is within your cluster and in relation to another. The silhouette value for a given patient (p_i) in a cluster C is computed as [3]

$$s(p_i) = \frac{b(p_i) - a(p_i)}{\max\{a(p_i), b(p_i)\}}, \quad s(p_i) \in [-1, 1] \quad (2)$$

where $a(p_i)$ is the average distance of patient p_i from all other patients in the cluster C , and $b(p_i)$ is the smallest of average distances from its neighbour clusters. Negative silhouette values represent wrong patient placements, whereas positive silhouette values represent better patient assignments [25].

After cluster evaluation, we used the centroids constructed by K-means algorithm for final classification.

For the pattern recognition problem investigated in this Letter, class labelling of the training patterns is not available and, therefore, we have to deal with the unsupervised classification problem. For this, we have to make a prediction based on groups by applying cluster analysis to the available dataset, and the resulting clusters are characterised based on the characteristics of the patterns by which they are formed. In the sequel, if we are given an unknown pattern, we can determine the cluster to which it is more likely to belong, and we characterise it based on the characterisation of the respective cluster [24].

To characterise the derived clusters, information related to 30 fictitious patients were used. These patients were simulated by a diabetes expert ensuring that 15 patients had high risk of developing diabetic foot, and 15 had low risk. For this, the high-risk fictitious patients were generated, answering the questions in Table 1 in favour of developing diabetic foot (risk factors present), and the low-risk fictitious patients were generated, answering the questions without the presence of risk factors.

The final classification of the simulated dataset was used to specify the clusters, achieved by K-means, as either being of high or low risk. The cluster to which the majority of the simulated high-risk patients belong is assigned as the high-risk cluster.

The operational phase of the proposed method is illustrated in Fig. 1b and consists of normalising the data and classifying them according to the clusters defined in the design phase. The result of the classification fits the patient into the high- or low-risk groups. If the patient is classified as low risk, a new patient is evaluated. If the patient is classified as high risk, care-taking actions defined by the local health team should be made to prevent diabetic foot.

4. Results and discussion: Table 1 presents the population characterisation as based on responses gathered from the

Figure 2 Silhouette values
a Two clusters
b Three clusters

Figure 3 Silhouette values for the artificial dataset

questionnaire. Notably, the primary population consists of lower literacy elderly patients. Accordingly, this characteristic compromises self-care tasks as their literacy abilities would hinder them in accomplishing care measures independently. Therefore primary attention (care) is a crucial preventative measure for diabetic complications.

The collected data were then grouped using the K-means algorithm, and Figs. 2a and b show the silhouette values obtained considering two and three clusters, respectively. Most points in the second cluster of both cases (two and three clusters) have the largest silhouette values, indicating that the cluster is somewhat separated from neighbouring clusters. However, the other clusters show few high values of silhouette, and the first cluster of Fig. 2a contains a few points with negative values, indicating that this cluster is not well separated. Four or more clusters led to smaller silhouette values, and therefore, they were not exploited.

Fig. 3 shows the silhouette values obtained for the artificial dataset groups. The simulated high- and low-risk patients are represented by clusters 1 and 2, respectively. All silhouette values of cluster 1 were higher than 0.6, whereas those achieved for cluster 2 were higher than 0.7, indicating that the clusters are very well separated from each other. This means that the groups were

Table 2 Results of classification for two clusters (artificial dataset)

Classifier	Level of risk	Number of patients		Accuracy
		Cluster 1	Cluster 2	
K-means	high risk	9	6	0.57 ± 0.18
	low risk	7	8	

Accuracy $\pm$ 95% confidence interval.

Table 3 Results of classification for three clusters (artificial dataset)

Classifiers	Level of risk	Number of patients			Accuracy
		Cluster 1	Cluster 2	Cluster 3	
K-means	high risk	14	0	1	0.97 ± 0.06
	low risk	0	7	8	

Accuracy $\pm$ 95% confidence interval.

Table 4 Classification of the specialist (real dataset)

K-means	Specialist			Accuracy
	High risk	Low risk	Not assigned	
high risk	11	4	5	0.69 ± 0.16
low risk	6	11	3	

Accuracy ± 95% confidence interval.

properly generated by the specialist, ensuring the creation of distinct groups.

Tables 2 and 3 show the classifier performance on the simulated data, considering two and three clusters, respectively. Considering two clusters, the classifier was not able to separate high-risk patients from low-risk ones. However, using three clusters, the classifier achieved good performance, in which 14 out of 15 high-risk patients were classified as pertaining to cluster 1 (accuracy of 0.97 ± 0.06). The low-risk patients were divided into clusters 2 and 3. It is important to note that the three groups were defined by the K-means algorithm only using the collected data. The artificial data were used to find out the high- and low-risk groups, since the artificial data were labelled by a diabetes specialist. Besides, the results show that the implemented classifier generalised well on unseen data.

As far as the classification of the real data is concerned, 48 of the 153 patients were classified in the group of high risk, and therefore, they require more attention from the health-care team to prevent the development of diabetic foot.

To compare these results with the opinion of a diabetes expert, 20 patients were randomly chosen from each group (high and low risk) and analysed by the specialist. According to the specialist, since the number of variables is high the classification of these 40 patients is a difficult task. Table 4 shows the classification of the specialist, where it can be noted that 8 of 40 patients were not assigned to a class, supporting the statement that this is a difficult task. Moreover, comparing the specialist's classification with that of the proposed method, we observed that 22 patients were classified in agreement with the proposed method, and 10 were classified in disagreement, leading to an accuracy of 0.69 ± 0.16 .

Analysing the high-risk group responses to the questionnaire, most of the patients, for instance, do not check the footwear inside before using it, nor have the habit of examining their feet. This behaviour did not occur with most of the assigned low-risk patients. This result is consonant with the Summary of Diabetes Self-care Activities measure [26], considered as a reliable self-report instrument. Furthermore, the proposed method groups the patients based on their behavioural similarities. In this way, the proposed approach takes into account many aspects to classify a patient.

Computationally, the use of centroids leads to a simple classifier because it only requires the calculation of three Euclidean distances to classify a patient.

The dataset formed by the collection of information from retired lower literacy elderly patients on the whole portrays part of the diabetic population which may limit the applicability of the obtained classifier in terms of its generalisation capability. However, it is important to note that these are the predominant characteristics of the population seeking medical help in Basic Health Units in Brazil. Moreover, the methodology can be used in different places and environments even if redesigned.

5. Conclusion: This work proposed an automatic non-invasive method to identify diabetic patients who are at high risk of developing diabetic foot. Using both real and simulated data, the method achieved satisfactory results with accuracy of 0.97 ± 0.06

for simulated data and 0.68 ± 0.16 considering the classification of specialists as the gold standard for real data.

The method requires simple computational processing, and therefore, it may be implemented in basic computers in Basic Health Units, allowing automatic and systematic control of diabetic patients regarding diabetic foot complication. In addition, the method can support health agents making the screening of patients with high risk of developing diabetic foot according to their habits, self-care and social condition.

The monitoring of patients through the method is important to see whether a patient classified as high risk will migrate to low-risk group after the systematic monitoring of the health team. Therewith, the health-care team may have an idea of the impact of its actions over the diabetic population in favour of self-care knowledge and practice.

6. Acknowledgments: The authors would like to thank the health team that works on the studied Basic Health Units, as well as the Federal University of Lavras (UFLA), Brazil, for support.

7 References

- [1] Kaur H., Wasan S.K.: 'Empirical study on applications of data mining techniques in healthcare', *J. Comput. Sci.*, 2006, **2**, (2), pp. 194–200. (doi: 10.3844/jcssp.2006.194.200)
- [2] Lee M., Gatton T., Kwang K.: 'A monitoring and advisory system for diabetes patient management using a rule-based method and KNN', *Sensors*, 2010, **10**, (4), pp. 3934–3953. (doi: 10.3390/s100403934)
- [3] Antonelli D., Baralis E., Bruno G., Cerquitelli T., Chiusano S., Mahoto N.: 'Analysis of diabetic patients through their examination history', *Expert. Syst. Appl.*, 2013, **40**, (11), pp. 4672–4678. (doi: 10.1016/j.eswa.2013.02.006)
- [4] Appuhamy J.A., Kebreab E., France J.: 'A mathematical model for determining age-specific diabetes incidence and prevalence using body mass index', *Ann. Epidemiol.*, 2013, **23**, (5), pp. 248–254. (doi: 10.1016/j.annepidem.2013.03.011)
- [5] Espeland M.A., Bray G.A., Neiberg R., *ET AL.*: 'Describing patterns of weight changes using principal components analysis: results from the Action for Health in Diabetes (Look AHEAD) Research Group', *Ann. Epidemiol.*, 2009, **19**, (10), pp. 701–710. (doi: 10.1016/j.annepidem.2009.06.001)
- [6] World Health Organization: 'The top 10 causes of death, 2013. The 10 leading causes of death in the world, 2000 and 2011', <http://www.who.int/mediacentre/factsheets/fs310/en/index.html>, accessed 27 November 2013
- [7] Bakker K., Apelqvist J., Schaper N.C. and on behalf of the International Working Group on the Diabetic Foot Editorial Board: 'Practical guidelines on the management and prevention of diabetic foot 2011', *Diabetes Metab. Res. Rev.*, 2012, **28**, (Suppl 1), pp. 225–231. (doi: 10.1002/dmrr.2253)
- [8] Alexiadou K., Doupis J.: 'Management of diabetic foot ulcers', *Diabetes Ther.*, 2012, **3**, (1), article 4. (doi: 10.1007/s13300-012-0004-9)
- [9] Jeffcoate W.J.: 'The incidence of amputation in diabetes', *Acta Chir. Belg.*, 2005, **105**, pp. 140–144
- [10] Tennvall G.R., Apelqvist J.: 'Health-related quality of life in patients with diabetes mellitus and foot ulcers', *J. Diabetes Complicat.*, 2000, **14**, (5), pp. 235–241. (doi: 10.1016/S1056-8727(00)00133-1)
- [11] Raspovic K.M., Wukich D.K.: 'Self-reported quality of life and diabetic foot infections', *J. Foot Ankle Surg.*, 2014, **53**, (6), pp. 716–719.
- [12] Leese G.P., Stang D., Pearson D.W.: 'A national approach to diabetes foot risk stratification and foot care', *Scott. Med. J.*, 2011, **56**, (3), pp. 151–155. (doi: 10.1258/smj.2011.011113)
- [13] Shearman C.P., Pal N.: 'Foot complications in patients with diabetes', *Surg. Oxf.*, 2013, **31**, (6), pp. 240–245. (doi: 10.1016/j.mpsur.2010.02.002)
- [14] Chellan G., Srikumar S., Varma A.K., *ET AL.*: 'Foot care practice – the key to prevent diabetic foot ulcers in India', *Foot*, 2012, **22**, (4), pp. 298–302. (doi: 10.1016/j.foot.2012.08.007)
- [15] Vedhara K., Dawe K., Wetherell M.A., *ET AL.*: 'Illness beliefs predict self-care behaviours in patients with diabetic foot ulcers: a prospective study', *Diabetes Res. Clin. Pract.*, 2014, **106**, (1), pp. 67–72. (doi: 10.1015/j.diabres.2014.07.018)
- [16] Boulton A.J.M.: 'The diabetic foot', *Medicine*, 2010, **38**, (12), pp. 644–648. (doi: 10.1016/j.mpmed.2010.08.011)

- [17] Jirkovská A., Boucek P., Wosková V., Barbos V., Skibová J.: 'Identification of patients at risk for diabetic foot. A comparison of standardized noninvasive testing with routine practice at community diabetes clinics', *J. Diabetes Complicat.*, 2001, **15**, (2), pp. 63–68. (doi: 10.1016/S1056-8727(00)00141-0)
- [18] Monteiro-Soares M., Boyko E.J., Ribeiro J., Ribeiro I., Dinis-Ribeiro M.: 'Risk stratification systems for diabetic foot ulcers: a systematic review', *Diabetologia*, 2011, **54**, (5), pp. 1190–1199. (doi: 10.1007/s00125-010-2030-3)
- [19] Boyko D.J., Ahroni J.H., Cohen V., Nelson K.M., Heagerty P.J.: 'Prediction of diabetic foot ulcer occurrence using commonly available clinical information: the Seattle diabetic foot study', *Diabetes Care*, 2006, **29**, (6), pp. 1202–1207. (doi: 10.2337/dc05-2031)
- [20] Lavery L.A., Armstrong D.G., Vela S.A., Quebedeaux T.L., Fleischli J.G.: 'Practical criteria for screening patients at high risk for diabetic foot ulceration', *Arch. Intern. Med.*, 1998, **158**, (2), pp. 157–162. (doi: 10.1001/archinte.158.2.157)
- [21] Schaper N.C., Apelqvist J., Bakker K.: 'International consensus and practical guidelines on the management and the prevention of the diabetic foot', *Curr. Diab. Rep.*, 2003, **3**, (6), pp. 475–479. (doi: 10.1007/s11892-003-0010-4)
- [22] Leese G.P., Reed F., Green V., ET AL.: 'Stratification of foot ulcer risk in patients with diabetes: a population based study', *Int. J. Clin. Pract.*, 2006, **60**, (5), pp. 541–545. (doi: 10.1111/j.1368-5031.2006.00899.x)
- [23] Mayfield J.A., Reiber G.E., Sanders L.J., Janisse D., Pagach L.M.: 'Preventive foot care in diabetes: American Diabetes Association', *Diabetes Care*, 2004, **27**, (Suppl 1), pp. S63–S64. (doi: 10.2337/diacare.27.2007.S63)
- [24] Theodoridis S., Koutroumbas K.: 'Pattern recognition' (Academic Press – Elsevier, 2009, 4th edn.)
- [25] Rousseeuw P.: 'Silhouettes: a graphical aid to the interpretation and validation of cluster analysis', *J. Computat. Appl. Math.*, 1987, **20**, pp. 53–65. (doi: 10.1016/0377-0427(87)90125-7)
- [26] Toobert D.J., Hampson S.E., Glasgow R.E.: 'The summary of diabetes self-care activities measure', *Diabetes Care*, 2000, **23**, (7), pp. 943–50. (doi: 10.2337/diacare.23.7.943)