

Associating Risks of Getting Strokes with Data from Health Checkup Records using Dempster-Shafer Theory

Sergio Peñafiel*, Nelson Baloian*, Jose A. Pino*, Jorge Quinteros*,
Álvaro Riquelme*, Horacio Sanson**, Douglas Teoh**

*Department of Computer Science, Universidad de Chile, Santiago, Chile

**Allm Inc., Tokyo, Japan

spenafiel@dcc.uchile.cl, nbaloian@dcc.uchile.cl, jpino@dcc.uchile.cl, jorge.quinteros@ug.uchile.cl,
alvaroriquelme@uchile.cl, horacio@allm.net, d.teoh@allm.net

Abstract—Prediction of future diseases from historical data of medical patients is a topic that has gained increasing interest given the growing availability of such data in electronic format. Most of the developed systems are based on machine learning techniques, which are good to find relations between data but do not help explaining causalities. In particular, it would be difficult to get a meaningful medical explanation for the relationship between a patient's health checkup data and the risk of developing a certain disease. On the other hand, expert system approaches, like Bayesian networks, are based on medical knowledge but have trouble dealing with high levels of uncertainty, which is crucial in this kind of scenario. In this work we present a prediction system for the risk of a patient having a (heart or brain) stroke based on past medical checkup data. The system is based on the Dempster-Shafer Theory of plausibility which is good for handling uncertainty. The data used belongs to a rural hospital in Okayama, Japan, where people are compelled to undergo annual health checkups by law. The model also produces rules that are able to relate data from exam results with the aforementioned risk, thus proposing a cause from the medical point of view. Experiments comparing the results of the Dempster-Shafer method with other machine learning methods like Multilayer perceptron, Quadratic discriminant analysis and Naive Bayes show that our approach performed the best in general, with an overall prediction accuracy of 61% and with the best precision value on true positive cases of stroke.

I. INTRODUCTION

Cardiovascular disease (CD), a non-communicable disease (ND) [1], [2], is the leading global cause of death, accounting for more than 17.3 million deaths per year in 2013, a number that is expected to grow to more than 23.6 million by 2030. In 2015, globally an estimated 7.4 million deaths were due to coronary heart disease and another 6.7 million deaths were due to stroke. In 2010 the estimated global cost of CD was calculated to be \$863 billion, and it is estimated to rise to \$1044 billion by 2030. Only in Japan, in 2012 around 55000 under 70 years old died because of CD [3].

The World Health Organization's Global Strategy for the Prevention and Control of non-communicable Diseases (2013) recommends the evaluation of various strategies for early detection and screening of ND to promote research on prevention and control of these diseases [3]. Being aware of the risk of developing these illnesses in early stages has proved to be

beneficial not only for prevention, but also for treatment of patients [2].

It is known that many countries gather electronic records of patients' medical checkups from over the years. These records usually include symptoms, examination results and the history of diseases of each patient. In the case of Japan, all active workers have to undergo every year a medical checkup to monitor their health, data that is also electronically recorded. This results in large databases of health information that can potentially be used in many ways to train classification or prediction models.

The problem of disease risk prediction has been studied through different approaches including statistical analysis [4], if-then rules [5], data mining [6] and machine learning [7]–[9]. In recent years, with the rise of new machine learning techniques and the growing amounts of medical data available, the problem has gained renewed interest from the data science community.

Generally speaking, there are mainly two different approaches for designing predictive algorithms in a variety of domains: the *expert system* approach, which consists in “encoding” the knowledge of experts in the required field in an algorithm, like if-then rules, Bayesian networks or association rules (like the ones implemented in the Prolog programming language), and the “machine learning” approach which consists in algorithms that handle large amounts of data, searching for relations among them, like neuronal networks, decision trees, genetic algorithms, particle swarm optimization and multilinear regressions.

Both approaches have pros and cons. Expert systems need the collaboration of experts, who very frequently cannot structure their knowledge, thus making it difficult to code it into an algorithm. It is also difficult to model some uncertainties experts have into this kind of systems [10]. A very important advantage of expert systems is that they allow to somehow comprehend and check why a certain prediction is made, which in the medical scenario is at least important, if not mandatory [11]. On the other hand, machine learning methods have the advantage that sometimes they can help discover relationships and causalities that were not previously known by experts.

In this work we present a classification model for the stroke risk prediction problem. Our goal is to build a system that is not only good in accuracy but its results can also be interpreted from the medical point of view, thus allowing medical experts to be part of the classification process and be able to learn from it. To accomplish this, the particular expert system approach we used is based on the Dempster-Shafer Theory of evidence. The system was developed with an eclectic approach, in which the rules for the predictive algorithm developed under the expert system approach are inferred with machine learning techniques, that can be later validated by experts.

The proposed model is based on rules called hypotheses; each of these hypotheses have a mass assignment function and a condition to be applicable. The model fits the values of the mass assignment function according to the data using both statistical analysis and uncertainty measurement. After training, the model has a clear medical interpretation of the values obtained and its relation with the embedding of the patients in feature vectors.

The developed model is applied to a large data set of Japanese electronic medical records from the Hospital of Tsuyoyama, Japan, which include exam results and patient attributes. To evaluate the performance of the model, we also present results of some experiments that compare our method against some classic machine learning algorithms.

II. RELATED WORK

This section first reviews previous works on data mining for disease prediction, then briefly introduces the principles of the Dempster-Shafer theory and finally, for illustration, the application of this theory to general classification problems is discussed.

A. Data mining in disease prediction

Many data mining models have been applied to the problem of diseases risk prediction, including both classic machine learning algorithms, like including Support Vector Machines [12], Multilayer perceptrons or Logistic regressions [8], as well as other non-conventional machine learning methods such as deep artificial neural networks [13].

In the particular case of stroke risk prediction, according to Khosla et al. (2010) [7] the use of machine learning methods combined with a specific pre-processing of the raw data and a feature extraction method can give results as good in accuracy as traditional indicators and methods to predict stroke risk.

Recalling that the area of interest is the medical one, it is important that the results obtained can be interpreted in some way by medical experts [11], so they can corroborate and also learn about which factors, possibly unknown to medical literature so far, increment the risk of having a stroke. Most of the works that can be found in the literature do not consider this point when modeling. A possible cause for this might be that the operations to fit the model do not have a clear meaning in the feature vector space, therefore it is not possible to given an interpretation to results obtained by this process if they are restricted to classical classification models.

Bayesian approaches to disease prediction with interpretable results have also been developed [9], but being a classic probability model that assigns probabilities to single outcomes, deterministic rules are obtained at the cost of losing uncertainty information or hiding it in complex conditional probability rules. Uncertainty is a first-class entity of Dempster-Shafer theory and it is considered from the beginning in the proposed model.

B. Dempster-Shafer Theory

The Dempster-Shafer Theory [14], also called the theory of belief functions, is a generalization of the Bayesian theory that does not require to establish prior probabilities to only single outcomes, it allows to assign “probabilities” to multiple outcomes measuring the degree of uncertainty of the process.

Let X be the set of all states or events of a system. A mass assignment function m is a function that satisfies:

$$m : 2^X \rightarrow [0, 1], \quad m(\phi) = 0, \quad \sum_{A \subseteq X} m(A) = 1 \quad (1)$$

The term $m(A)$ can be interpreted as the probability of getting exactly the outcome A , and not a subset of A .

The belief indicator is presented as the total evidence to support an outcome, and it is given by the following expression:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

The plausibility indicator is defined as the total amount of evidence that can support an outcome. This can be computed using a formula similar to the one for belief:

$$Pl(A) = \sum_{B \cap A \neq \phi} m(B) \quad (3)$$

It has been proved that belief and plausibility define lower and upper bounds for the real probability of the event [15], so then the following inequality is always true:

$$Bel(A) \leq P(A) \leq Pl(A) \quad (4)$$

Multiple evidence sources expressed by their mass assignment functions of the same frame of discernment can be combined using the Dempster Rule (DR) [16]. Given two mass assignment functions m_1 and m_2 , a new mass assignment function m_c can be constructed by the combination of the other two using the following formula:

$$\begin{aligned} m_c(A) &= m_1(A) \oplus m_2(A) \\ &= \frac{1}{1-K} \sum_{B \cap C = A \neq \phi} m_1(B)m_2(C) \end{aligned} \quad (5)$$

Where K is a constant representing the degree of conflict between m_1 and m_2 and it is given by the following expression:

$$K = \sum_{B \cap C = \phi} m_1(B)m_2(C). \quad (6)$$

The Dempster Rule has useful mathematical properties like associativity, commutativity, and being absorbent to certainty and neutral to uncertainty.

C. Dempster-Shafer Applications

Works applying Dempster-Shafer to classification and regression problems show promising results in performance. These can be grouped into two categories: Dempster-Shafer used to improve another technique, and using the theory directly to build a classifier.

For the first group, Dempster-Shafer theory is applied to a specific part of the process. For example, in [17] a variation of an artificial neural network is presented, which behaves similar to a Multilayer Perceptron but uses the weights of neuron links as evidence input for the Dempster-Shafer theory as well as the Dempster Rule for pooling. Other example of this is [18], where a variation of k-nearest neighbors is presented. In this work the distances of the neighbours are used as evidence sources for the Dempster-Shafer theory and the outcome class of a record from the input dataset is the maximum belief value according to the sources.

On the other hand, there are works focused on building classifiers by expanding or specializing the theory. For example, [19] presents a modified version of the Dempster-Shafer theory that allows to include prior knowledge about the possible outcomes. A process to classify data using this framework is also presented.

III. PROBLEM DEFINITION

The problem to be addressed can be described as: given a patient's historic medical information, what is the risk of he/she having a stroke.

This problem can be reduced to one of discrete supervised learning, i.e., a classification problem. Under this approach the problem can be defined as follows: First, we consider two classes; the patient has a stroke or does not have it. The outcome of the algorithm is then the predicted class for each patient. However, this is not completely defined because outcome data is also time dependant; it is different for the classifier and for the further decision making process to get stroke within the next year or in the next 10 years. To handle this in a convenient discrete manner, three future time intervals were established to evaluate the model: having a Stroke within one year, having a stroke within three years and finally, having a stroke within ten years. We believe the choice of these intervals suitably weighs the level of "emergency" of the patient's condition.

Under this redefinition of the disease risk prediction problem, the developed solution is presented in the next section.

IV. PROPOSED MODEL

Now we present the model implemented to predict stroke risk in patients. First we discuss the embedding for the patients' data and then the classifier itself is explained.

A. Embedding

Embedding is the process of representing a study object as a feature vector. In this case, study objects are patients with their respective medical information. There are three main data sources containing useful patient information that were used:

- 1) **Patients' checkups.** This data source contains all the information about results of exams Japanese workers have to undergo annually. Each exam is indexed by a code, and the result is also included. Examples of exams are LChO exam and blood pressure. There are 46 different types of exams, and the date when they were collected is also registered.
- 2) **Patients' attributes.** It contains general information about patients such as height, weight, body fat, age, gender, and waist measurement.
- 3) **Patients' disease history.** It contains information about medical receipts and diagnoses of patients. The diseases are indexed by ICD-10 codes ([20]) and the date of each diagnosis is also included.

These data sources are not completely robust, there are many missing values and in many cases joining the data sources for a patient is not possible.

Because of this, in this work we tried using 3 different embedding structures for patients' information; from simple examination results to more complex feature vectors that consider more data about each patient.

The first proposed embedding takes into account the most recent result of each kind of exam a patient can have. A feature vector for a patient then is represented as:

$$x = (e_1, e_2, \dots, e_N) \quad (7)$$

where e_i represents the most recent result of exam i . In the Tsuyoyama data set there are 46 different types of exam, but not all patients have taken every exam so there are many missing values even in this simple representation. Figure 1 shows the distribution of the patients and the respective number of different exams taken in the whole data set. The chart shows that most of the patients have less than 24 different exam results.

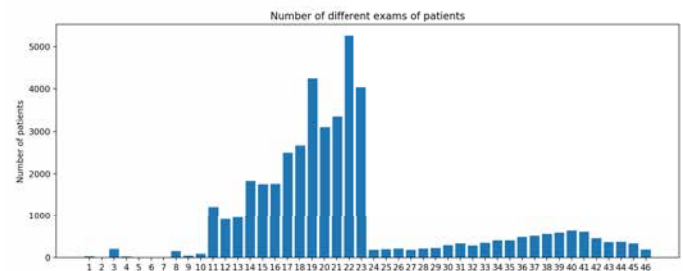


Fig. 1. Distribution of patients and number of different exam taken.

Recalling the other data sources, the second embedding tested was to include, alongside the exam results, the patients' attributes. Including this, a patient has the following feature vector associated:

$$x = (e_1, e_2, \dots, e_N, a_1, a_2, \dots, a_M) \quad (8)$$

where a_j represents the value of the patient attribute j . As mentioned before, in many cases executing joins between data sources is not possible, but at least now it is possible to obtain age and gender for all patients.

The last embedding that we tried is adding temporal information about the examinations and recent history. In this embedding, the gap in days between the evaluation time and the examination date of each type of exam is included, alongside the number of exams taken in the last year. With this, the feature vector becomes:

$$x = (e_1, e_2, \dots, e_N, a_1, a_2, \dots, a_M, t_1, t_2, \dots, t_N, K) \quad (9)$$

where t_i represents the difference in days between the examination i date and the evaluation date, and where K represents the number of exams of any kind taken in the last year.

The temporal information is used as a measurement of the relevance of the exam i to the prediction.

B. Classifier

Upon establishing these models for the data input, the proposed classifier is constructed on top of the Dempster-Shafer theory, trying to keep the properties of uncertainty and combinations in the model as well as providing a framework to deal with the stroke prediction problem.

Since the problem was defined as having 2 classes, the mapping to Dempster-Shafer theory requires a frame of discernment $X = \{\text{Stroke}, \text{No Stroke}\}$. Using this, a mass assignment function will operate from the power set of X , i.e.:

$$2^X = \{\phi, \{\text{Stroke}\}, \{\text{No Stroke}\}, \{\text{Stroke}, \text{No Stroke}\}\} \quad (10)$$

For simplicity, let's name S the Stroke singleton, NS the No stroke singleton and E the Either group (which accounts for the uncertainty).

A rule is defined as a pair containing a predicate $p : X \rightarrow \mathbb{B}$, which is a Boolean function to be evaluated using the feature vectors, and a mass assignment function m . A rule can be interpreted as the following: if p is true for a patient then it is known that m is applicable. Interpretability is straightforward from the mass assignment function because each component of the function of a rule can be interpreted as: S measures the certainty of having stroke, NS measures the certainty of not having stroke and E measures the uncertainty of both outcomes.

Rules can be established directly by experts or they can be inferred from data. For this work we developed a method to build rules considering certainty and uncertainty sources from data.

For each attribute of the feature vector, its domain is divided in N bins. For attributes having a continuous data range the division in bins is done by statistic breaks. For each bin a rule

is created. Let $D \subseteq X$ be the patients of X that satisfy the rule and let D_0 and D_1 the patients of D that do not have stroke (NS class) and have stroke (S class), respectively. The certainty indicator for each class is the relative frequency of the patient of each class. The uncertainty indicator is given by 2 metrics:

- **Rule representativeness (SR).** This measures the portion of patients that satisfy the rule from the total and is used to avoid outliers generating rules that are not generalizable to the rest of the patients. It is given by the following formula:

$$SR = 1 - \frac{|D|}{|X|} \quad (11)$$

- **Class distinction (Diff).** This measures the difference between both classes: if the difference is low this indicator takes a high value and if the difference is clear, this indicator takes a small value. This is used to measure the real contribution of the rule to the final prediction and is given by the following formula:

$$Diff = 1 - \frac{\max(D_0, D_1) - \min(D_0, D_1)}{\max(D_0, D_1)} \quad (12)$$

To summarize, table I shows the final formulas used to build a mass assignment function, given a predicate p obtained from the division of the domain of a characteristic.

TABLE I
FORMULAS USED TO BUILD A MASS ASSIGNMENT FUNCTIONS.

A	m(A)
ϕ	0
NS	$\frac{1}{1 + SR + Diff} \times \frac{D_0}{D}$
S	$\frac{1}{1 + SR + Diff} \times \frac{D_1}{D}$
E	$\frac{SR + Diff}{1 + SR + Diff}$

After the model was calibrated, an evaluation process was performed as follows: for a patient represented by the vector x_i , search all rules that x_i satisfy and extract their mass assignment functions, that we will call them $m_1, m_2, \dots, m_k$. In order to consider all these rules, the combination of all mass assignment functions is applied using Dempster Rule, obtaining a final mass assignment function m_f .

$$m_f = DR(m_1, DR(m_2, \dots)) = m_1 \oplus m_2 \oplus \dots \oplus m_k \quad (13)$$

After this process, m_f contains the information of the outcome of the model. The predicted class is the one with higher belief.

$$\tilde{y}_i = \operatorname{argmax} \operatorname{Belief}(m_f) \quad (14)$$

V. RESULTS

In order to test the model various experiments were performed using the available data from the Hospital of Tsuyoyama. For the experiments a subset of the patients' data set was used. Patients with few different exam results (12 or lower) were not accounted. The evaluation date was set to July 2015 and consequently all data before this date was used to build the corresponding embedding for each patient. The data up to one year after July 2015 (i.e. July 2016) was used to query the history and determine the class (Stroke, No Stroke). The resulting set was separated into two sets: a training set (70%) for the classifier to learn the rules and a testing set (30%) for performance evaluation. In addition to our model, three other machine learning method were tested: Multilayer perceptron, Quadratic Discriminant Analysis and Naive Bayes. All four methods used the same data set and embedding Model.

In this section results of the experiments are presented. Each result is presented as a Receiver Operating Characteristic (ROC) curve and/or a Precision-Recall curve. These are two commonly used methods to evaluate performance in binary classification: the ROC curve corresponds to a chart that illustrates the rate of true positives to false positives when moving the acceptance threshold and the PR curve shows the performance of the classifier over certain class, showing the variation of the true positive rate with the total number of positive predictions when varying the acceptance threshold.

The first experiment consisted in using the proposed model with the simple embedding 7, setting the number of bins for parsing the continuous data to 6. The result obtained can be visualized in the ROC curve of figure 2.

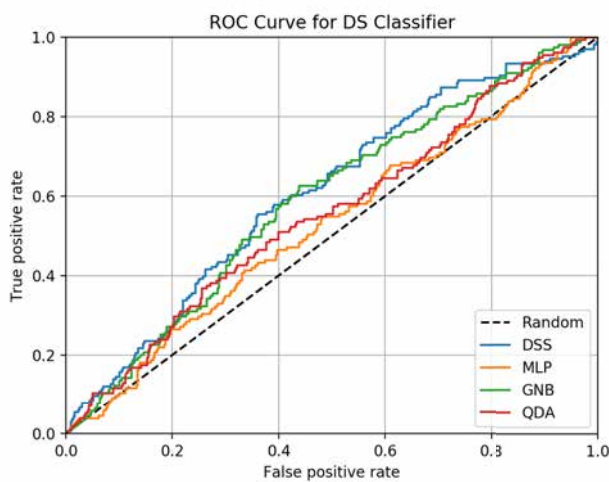


Fig. 2. ROC curve for stroke prediction using exam embedding.

The next experiment consisted in including the patients' attributes to feature vectors described in 8 obtaining the ROC curve shown in figure 3.

The next experiment was performed only varying the number of bins for the continuous data, the same training and testing sets for each number of breaks were used. The result is shown in figure 4.

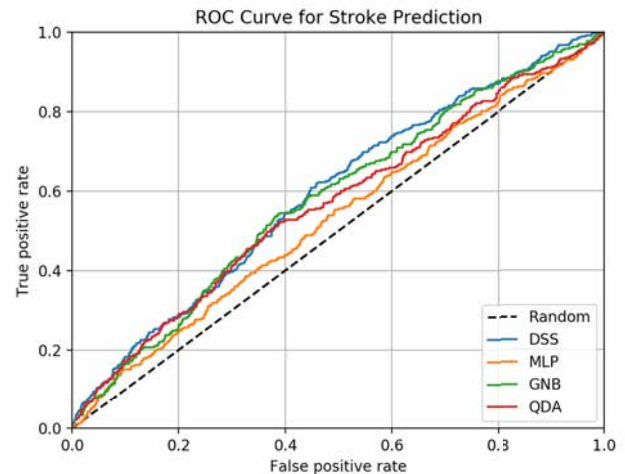


Fig. 3. ROC curve for stroke prediction using exam and attributes embedding.

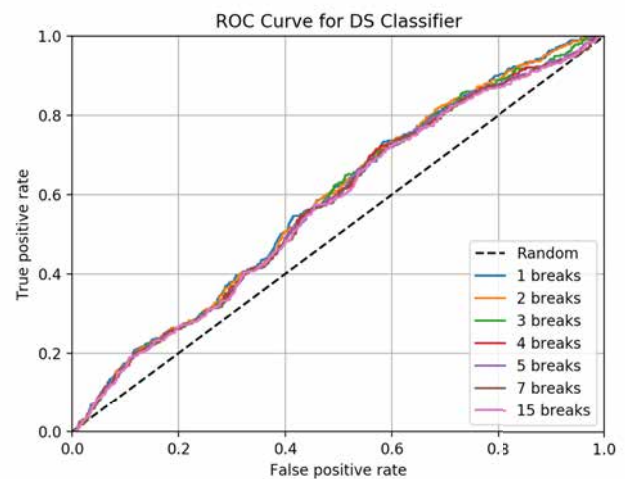


Fig. 4. ROC curve for stroke prediction varying the number of bins.

The next experiment was to compare the strategies used to deal with missing values. Two of these strategies were tested: Mean imputation, which means to complete missing values with the mean value of each attribute, and skip missing values, which is the method used in the previous experiments. The result is presented in figure 5.

The final experiment was including temporal information of the exams and the exam frequency using the embedding described in 9. It uses 3 bins for continuous data and skips missing values. The results are shown in figures 6 and 7.

For this experiment some of the most relevant rules for predicting stroke class and the values of the mass assignment function (NS, S, E) are shown in the table II.

VI. DISCUSSION

The results presented show that the proposed model performs better than the other classic machine learning algorithms in most of the cases. Overall it gets an accuracy of 60%. Results of figure 2 show that even in non-complex embedding

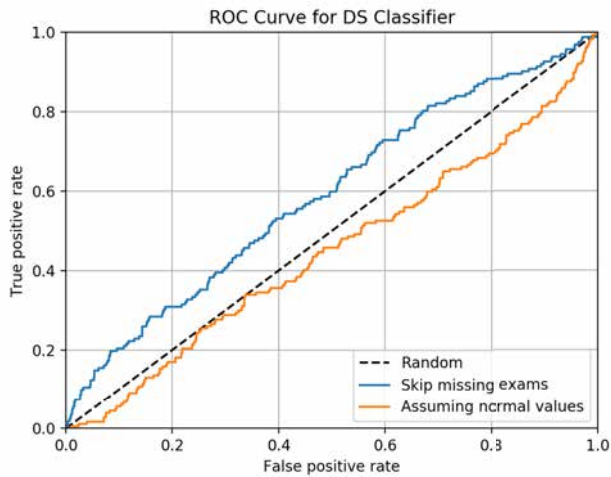


Fig. 5. ROC curve for stroke prediction using different strategies for handling missing values.

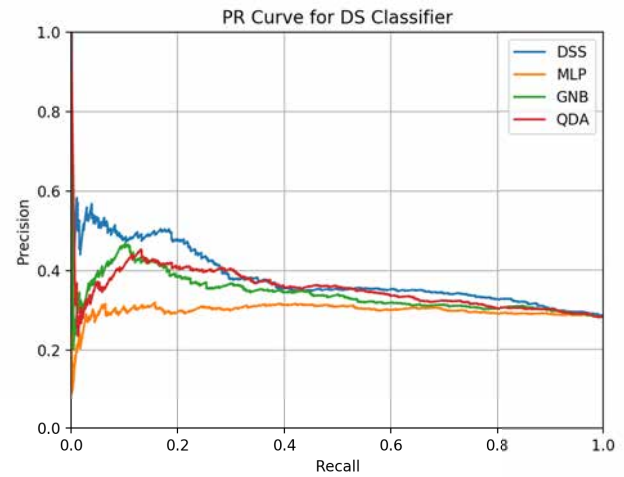


Fig. 7. PR curve for stroke class using exams, patients' attributes and temporal information.

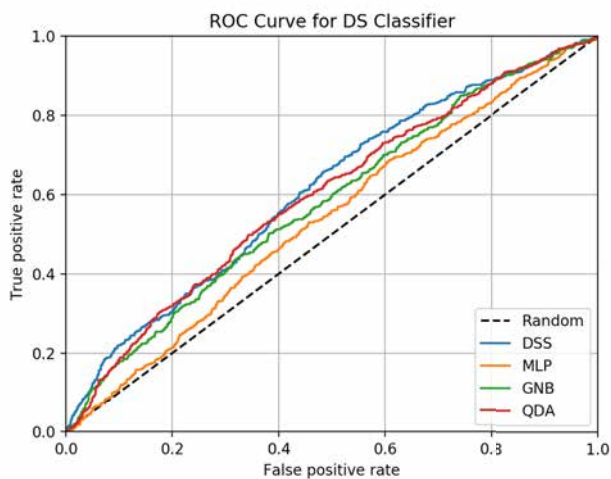


Fig. 6. ROC curve for stroke prediction using exams, patients' attributes and temporal information.

TABLE II
MOST RELEVANT RULES FOR CEREBRAL STROKE RISK.

Exam / Indicator	Range	NS	S	E
Body fat	Greater than 14.0	0.481	0.288	0.231
HbA1c	Greater than 5.17	0.504	0.279	0.217
WBC	Lower than 2865	0.544	0.262	0.194
PLT	Greater than 59.4	0.550	0.259	0.191
HDL-C	Greater than 40.7	0.644	0.214	0.142

the proposed model performs good in accuracy. The addition of extra attributes to feature vectors shown in figures 3 and 6 helps to improve the accuracy of all methods, but the classic machine learning approaches tend to increase more in accuracy and tend to get closer to the ROC curve of the DS method. However, figure 7 show that the proposed method performs better in the stroke class which is the interesting class to be

predicted, and then it gives a better true positive rate than the others.

Regarding to the parameters of the model, the number of breaks for continuous data was varied in order to find out the optimal value. The results of figure 4 show that the curves tend to have the same behavior, which means that the method of breaking statistically the domain of continuous attributes does not introduce a big change in the prediction. In other words, rules considering large ranges for variables summarize the information of shorter ranges and they perform the same. In fact, it is also possible to see that when increasing the number of breaks the accuracy decreases a little. This is because shorter ranges have less data to build their rules and then they are more susceptible to learn wrong relations.

Two different strategies for dealing with missing values were performed in order to get the best outcome. The results of figure 5 clearly show the method of skipping the missing values is better than the method of imputing or assuming the normal values given by the mean of the attribute. The reason why this happens is that the number of missing values is, in many cases, higher than the actual values, and then the rules discovered by this strategy do not reflect a real pattern and it causes the model to decrease accuracy.

Finally, the table II shows the final values of the mass assignment function for a certain rule. The rules were selected according to the mass assignment function values. Rules having a higher value in stroke class and a lower value in uncertainty were prioritized. The table shows the expressiveness power of the model. In fact, each row of the table gives meaningful causes about the risk of getting stroke for a patient.

This causality can be contrasted with actual medical evidence. The study of body fat index done in Japan presented in [21] shows that a higher value of body fat increase stroke risk which coincide with the first rule. Similarly, in [22] authors present a work which relates higher values of HbA_{1c} with a higher risk of getting stroke. A lower value of HDL-C is usually associated to risk of stroke, but in [23] it is shown

that in Asian population the relation is inverse which coincide with the rule obtained by our method.

VII. CONCLUSION

In this work we presented a method based in the Dempster Shafer Theory to predict stroke occurrences. The model consists either in the classifier and the embedding for patients' data, taking care of the interpretability of the model. Results show that the method is better than other machine learning approaches using the same data sources and embedding model. The contribution of the model is that discovered rules have a clear interpretation and give hints about causality of outcomes allowing to medic experts to know exactly what the model is doing and compare it with medical evidence.

As a future work, the model can be extended by proving to incorporate new data, and different rules generation. For example, data about patient history is only used to know if a patient got stroke or not, but it can be also used as part of the feature vectors. Expert knowledge can be incorporated by expressing them as new rules for the system. Rules extracted also indirectly imply independence of the variables, that may be false in some cases, a better rule generator should also discover dependencies and express them as rules.

REFERENCES

- [1] W. H. Organization *et al.*, "2008-2013 action plan for the global strategy for the prevention and control of noncommunicable diseases: prevent and control cardiovascular diseases, cancers, chronic respiratory diseases and diabetes," 2009.
- [2] —, "Noncommunicable diseases country profiles 2014," 2014.
- [3] E. J. Benjamin, M. J. Blaha, S. E. Chiuve, M. Cushman, S. R. Das, R. Deo, S. D. de Ferranti, J. Floyd, M. Fornage, C. Gillespie *et al.*, "Heart disease and stroke statistics 2017 update: a report from the American Heart Association," *Circulation*, vol. 135, no. 10, pp. e146–e603, 2017.
- [4] E. Faught, S. D. Trader, and G. R. Hanna, "Cerebral complications of angiography for transient ischemia and stroke prediction of risk," *Neurology*, vol. 29, no. 1, pp. 4–4, 1979.
- [5] C. Ordóñez, "Comparing association rules and decision trees for disease prediction," in *Proceedings of the international workshop on Healthcare information and knowledge management*. ACM, 2006, pp. 17–24.
- [6] K. Srinivas, B. K. Rani, and A. Govrdhan, "Applications of data mining techniques in healthcare and prediction of heart attacks," *International Journal on Computer Science and Engineering (IJCSSE)*, vol. 2, no. 02, pp. 250–255, 2010.
- [7] A. Khosla, Y. Cao, C. C.-Y. Lin, H.-K. Chiu, J. Hu, and H. Lee, "An integrated machine learning approach to stroke prediction," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 183–192.
- [8] I. Colombet, A. Ruelland, G. Chatellier, F. Gueyffier, P. Degoulet, and M.-C. Jaulent, "Models to predict cardiovascular risk: comparison of cart, multilayer perceptron and logistic regression," in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2000, p. 156.
- [9] B. Letham, C. Rudin, T. H. McCormick, D. Madigan *et al.*, "Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model," *The Annals of Applied Statistics*, vol. 9, no. 3, pp. 1350–1371, 2015.
- [10] D. Heckerman, E. Horvitz, and B. N. Nathwani, "Toward normative expert systems part i," *Methods of information in medicine*, vol. 31, 2016.
- [11] I. Bratko, "Machine learning: Between accuracy and interpretability," in *Learning, networks and statistics*. Springer, 1997, pp. 163–177.
- [12] F. Mittag, F. Büchel, M. Saad, A. Jahn, C. Schulte, Z. Bochdanovits, J. Simón-Sánchez, M. A. Nalls, M. Keller, D. G. Hernandez *et al.*, "Use of support vector machines for disease risk prediction in genome-wide association studies: Concerns and opportunities," *Human mutation*, vol. 33, no. 12, pp. 1708–1718, 2012.
- [13] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of alzheimer's disease with deep learning," in *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*. IEEE, 2014, pp. 1015–1018.
- [14] G. Shafer *et al.*, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.
- [15] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The annals of mathematical statistics*, pp. 325–339, 1967.
- [16] G. Shafer, "Dempster's rule of combination," *International Journal of Approximate Reasoning*, vol. 79, pp. 26–40, 2016.
- [17] T. Denoeux, "A neural network classifier based on dempster-shafer theory," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 30, no. 2, pp. 131–150, 2000.
- [18] —, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE transactions on systems, man, and cybernetics*, vol. 25, no. 5, pp. 804–813, 1995.
- [19] D. Fixsen and R. P. Mahler, "The modified dempster-shafer approach to classification," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 27, no. 1, pp. 96–104, 1997.
- [20] W. H. Organization, *International Classification of Functioning, Disability and Health: ICF*. World Health Organization, 2001.
- [21] N. Kawate, K. Kayaba, M. Hara, K. Kotani, S. Ishikawa, J. M. S. C. S. Group *et al.*, "Body mass index and stroke incidence in japanese community residents: The jichi medical school (jms) cohort study," *Journal of Epidemiology*, 2017.
- [22] M. Chonchol, R. Katz, L. F. Fried, M. J. Sarnak, D. S. Siscovick, A. B. Newman, E. S. Strotmeyer, A. Bertoni, and M. G. Shlipak, "Glycosylated hemoglobin and the risk of death and cardiovascular mortality in the elderly," *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 20, no. 1, pp. 15–21, 2010.
- [23] I. Saito, K. Yamagishi, Y. Kokubo, H. Yatsuya, H. Iso, N. Sawada, M. Inoue, and S. Tsugane, "Association of high-density lipoprotein cholesterol concentration with different types of stroke and coronary heart disease: The japan public health center-based prospective (jphc) study," *Atherosclerosis*, vol. 265, pp. 147–154, 2017.



working in his thesis about interpretable methods for classification and regression and its applications in real scenarios.



(Armenia), University of Waseda (Japan) and American University of Armenia (Armenia).



the Journal of Educational Technology and Society.

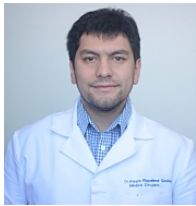
Sergio Peñaflor was born in Coquimbo, Chile in 1994. He started studying at the Faculty of Physical and Mathematical Sciences of Universidad de Chile in 2013. He obtained the Bachelor's degree in Computer Science in early 2017. He is now a Master student in computer science from Universidad de Chile. He worked as Research Assistant in CEAMOS Santiago Chile in urban crime prediction to optimize police patrolling. He did an internship to Allm Inc. Tokyo Japan to collaborate in disease prediction using machine learning. He is currently

Nelson Baloian is an Associate Professor of Computer Science at the University of Chile. He obtained his Engineering title from the University of Chile and his Ph.D. from the University of Duisburg, Germany. He has been Chairman of the Computer Science Dept. of the University of Chile. His research interests are collaborative computer supported learning, mobile computing and geo-collaboration. He is author of about 80 conference articles and 20 journal papers in these areas. He has been Visiting Professor at the University of Duisburg-Essen (Germany), University of Waseda (Japan) and American University of Armenia (Armenia).

José A. Pino is a full professor of Computer Science at the University of Chile. His research interests include ubiquitous computing, human-computer interaction and socio-technical systems. He has served as President of the Chilean Computer Science Society (SCCC) and President of CLEI (the Latin American Association of Universities Concerning Information Technology). He has co-authored six books and published research papers in international conferences and journals, including Journal of the ACM, Communications of the ACM, ACM Computing Surveys, Decision Support Systems, Group Decision and Negotiation, and Future Generation Computer Systems. He is currently a member of the Editorial Boards of the Journal of Network and Computer Applications and the Journal of Educational Technology and Society.



Jorge Quinteros was born in Santiago, Chile in 1992. He joined University of Chile's Faculty of Physical and Mathematical Sciences in 2011. He obtained a Bachelors degree in electrical engineering and a Bachelors degree in computer science engineering in 2016, both from University of Chile, and is now on his way to obtaining a Masters degree in computer science from the same institution. He has experience as Research Assistant in the fields of football robotics, data mining for astronomy and recently did an internship on early 2017 as a Research Assistant in the field of health informatics in Allm Inc. in Tokyo, Japan. Research interests include robotics, machine learning and more recently blockchain technology and e-Health.



Alvaro Riquelme is a Physician from University of Concepción, Chile. He is currently coursing a Medical Informatics Master at University of Chile. Until now has worked as a physician on various Clinics and Emergency Departments in Santiago, Chile. His master thesis is related to usability of health information systems on clinical environments.

Horacio Sanson obtained his Masters Degree at the Global Information and Telecommunication Institute of Waseda University in Tokyo, Japan, in 2005, where he researched distributed localization on emergency wireless networks. From there, he worked on several engineering positions until he became lead research engineer at Allm Inc. on 2014, where he spends his time applying engineering and machine learning to solve practical medical and health-care issues.

Douglas Teoh graduated from The Univ. of Queensland (Australia) in 2010, receiving a B.Eng (First Class Honours) in software engineering. He joined Allm Inc. as an engineer in 2015, and is currently involved with applying machine learning research in the medical domain.