IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 17, 2020, accepted December 4, 2020, date of publication December 24, 2020, date of current version January 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047195

Predicting Stroke Risk With an Interpretable Classifier

SERGIO PEÑAFIEL[1], NELSON BALOIAN[1], HORACIO SANSON[2], AND JOSÉ A. PINO[1]
[1]Department of Computer Science, Universidad de Chile, Santiago 8370456, Chile
[2]Allm Inc., Tokyo 150-0002, Japan

Corresponding author: Sergio Peñafiel (spenafie@dcc.uchile.cl)

The work of Sergio Peñafiel was supported by the Conicyt (Chile) Master Scholarship under Grant 22180506.

ABSTRACT Predicting an individual's risk of getting a stroke has been a research subject for many authors worldwide since it is a frequent illness and there is strong evidence that early awareness of having that risk can be beneficial for prevention and treatment. Many Governments have been collecting medical data about their own population with the purpose of using artificial intelligence methods for making those predictions. The most accurate ones are based on so called black-box methods which give little or no information about why they make a certain prediction. However, in the medical field the explanations are sometimes more important than the accuracy since they allow specialists to gain insight about the factors that influence the risk level. It is also frequent to find medical information records with some missing data. In this work, we present the development of a prediction method which not only outperforms some other existing ones but it also gives information about the most probable causes of a high stroke risk and can deal with incomplete data records. It is based on the Dempster-Shafer theory of plausibility. For the testing we used data provided by the regional hospital in Okayama, Japan, a country in which people are compelled to undergo annual health checkups by law. This article presents experiments comparing the results of the Dempster-Shafer method with the ones obtained using other well-known machine learning methods like Multilayer perceptron, Support Vector Machines and Naive Bayes. Our approach performed the best in these experiments with some missing data. It also presents an analysis of the interpretation of rules produced by the method for doing the classification. The rules were validated by both medical literature and human specialists.

INDEX TERMS Dempster-Shafer theory, stroke, expert systems, interpretable classification.

I. INTRODUCTION

Cerebrovascular accidents, also known as strokes, are the second leading cause of death worldwide and the third leading cause of disability [1]. Stroke is defined as the sudden death of some brain cells due to lack of oxygen and, in many cases, it is asymptomatic. Strokes have an enormous impact on countries' socio-economic development. For example, according to the American Heart Association, in the United States more than 140 thousand people died due to stroke as underlying cause in 2016. This means that stroke accounted for one of every 19 deaths. The estimation for the sum of direct and indirect costs of strokes in 2015 was \$45.5 billion. This cost is projected to double in the next 20 years, reaching a total expenditure of \$94.3 billion by 2035 [2]. The World Health Organization (WHO) declares strokes as one of the growing crises having received very little attention to date [3]. Early awareness of having this illness has proved to be beneficial for prevention and treatment [4].

Middle- and high-income countries have developed health-care systems that have been collecting patients' data for years. Some of them follow international standards for information structure such as ICD codes or FHIR standard [5]. In addition, countries like Japan have strict policies requiring active workers to have an annual checkup regardless if they are healthy or not [6]. These protocols produce large datasets for both healthy and sick cases, and they show the evolution of the patient medical status through the years. This valuable information can be applied to create a non-invasive method to monitor patients. For example, we could use it to estimate the risk of getting a stroke in the next few years. However, it is common that these datasets contain missing or erroneous data. The literature shows that there have been some efforts in the past to tackle this problem developing models based

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente García-Díaz.

on machine learning that use this data to assess the risk of a patient of getting a stroke in the future (e.g. [7]). The literature shows that the most accurate models belong to the so called "black-box" type [8] which do not give much information about the reasons why they classify a person as risky or not. This explanation is sometimes even more valuable information than the accuracy of the prediction itself since this leads to new knowledge discovery. This characteristic is often named interpretability.

The objective of this work is to develop a new method for automatic and individual stroke risk prediction within a year using data from electronic health records (EHR) and exam results. To accomplish this, we propose to use an interpretable rule-based method like the one presented in our previous work [9] but using rule mass optimization based on Gradient Descent proposed by Peñafiel *et al.* [10] instead of statistical analysis. The highlights of the new model are the ability of handling missing data, the use of actual medical knowledge encoded as rules for the model and the automatic optimization of rule values.

The developed model was applied to a dataset of Japanese Electronic Health Records (EHR) from the Tsuyoyama Hospital, Japan, which includes exam results and patient attributes. We compare our method with others for screening stroke and we discuss the most important attributes to assess this risk. Results show that it outperforms other machine learning models and stroke screening methods. In addition, we validated the rules that our model found as the most important ones for detecting stroke. We compared them to medical studies and we applied an experts survey. Experts agreed with most of the rules found by our model.

II. RELATED WORK

This section first reviews previous works on stroke risk prediction which define baselines for comparison. Then interpretable machine learning techniques are presented, which can be suitable candidates as models for stroke risk prediction.

A. METHODS FOR STROKE RISK PREDICTION

CHADS_2 is a score metric to evaluate stroke risk prediction proposed by Gage *et al.* [11]. The name CHADS_2 is an acronym for the method itself, the letter "C" assigns a point to patients who have congestive heart failure, "H" assigns a point to patients with hypertension, "A" is for patients aged 75 or older, "D" is for patients with diabetes mellitus and finally "S_2" assigns 2 points to patients who have had stroke, ischemic attack or thromboembolism in the past. CHADS_2 score is widely used in the medical field to predict stroke occurrences because of its simple formulation and evaluation. A drawback of this method is that it only uses five variables for decision making.

Letham *et al.* [12] used Bayesian Rule Lists (BRL) to develop an interpretable model to predict stroke risk within a year for patients diagnosed with atrial fibrillation. The method uses decision lists, which consist of a series of "if . . . then . . ." statements, then it uses a generative model called Bayesian Rule Lists which automatically produces a posterior distribution over possible decision lists allowing inferences to be made about stroke risk.

Another approach to predict stroke risk was proposed by Peñafiel *et al.* [9]. In their work, a model based on Dempster-Shafer Theory was proposed; this model operates using rules that were built using training data statistics; we will call this model as DS-Stat. When evaluating a new case, the corresponding rules are combined using the Dempster Rule to provide a final mass assignment function, then the belief for this assignment is computed and given as the model response. This model achieves an accuracy of 61%, An important feature of the model is that it is interpretable. The authors showed the essential rules associated to stroke and how they were verified with medical literature.

Teoh [13] also proposed a method to predict stroke risk. In his work, a Recurrent Neural Network (RNN) [14] was used in combination with a custom loss function. The model uses all available data for a patient, such as exam results and diagnosis structured like a time series for the RNN. Then, fully-connected layers are applied to predict the class. The custom loss function was tested and compared with classical ones. This function was proven to behave better since it takes into account the context of data more accurately. Interpretability is also covered in this work by deleting specific attributes from the feature vector and observing the change in the accuracy indicator. The best result of the model achieves an area under the receiver operator curve (AUC ROC) score of 0.669.

Weng *et al.* [15] present another work on stroke risk prediction using routine clinical data and machine learning techniques. In their work, they analyze data from more than 380,000 patients who have 30 variables that could affect this risk. They analyze four traditional machine learning methods: Logistic Regression, Random Forest, Gradient Boosting and Neural Networks. The results show that all the methods exceed the baseline that the authors propose, the best of which are Neural Networks with an AUC ROC of 0.764. This work also shows a typical example of the trade-off between interpretability and performance in machine learning, because random forest or logistic regression achieve less accurate results but their results can be explained while Neural Networks, which are those that obtain the least error, behave like black-boxes.

B. INTERPRETABLE CLASSIFICATION

Classification is the process of assigning a label to a study object given its features or defining it as being of a certain class. In machine learning, classification is usually called as discrete supervised learning. Many methods to solve classification tasks, known as classifiers, have been proposed, such as Support Vector Machine [16], Artificial Neural Networks [17], and Bayesian networks [18].

Classifiers can be divided into two groups: the ones that behave like black-boxes and the ones that are interpretable.

A black-box classifier is a model that uses the input to make complex non-linear operations in order to get the output, in many cases these computations are even data-dependant thus it is nearly impossible to know exactly how a decision was made by the model. The most remarkable black-box classifiers are Artificial Neural Networks and specially Deep Networks [19]. On the other hand, interpretable classifiers give clear information about why a certain decision was made and which is the importance of the features in the process. The most known interpretable classifier is the Decision Tree. Empirically, it is observed that there is a trade-off between interpretability and accuracy, meaning that the most interpretable methods are usually the ones with high error in prediction.

The importance of having an interpretable decision support system is critical in many areas including the medical one [20]. For a physician or an insurer it is much more important to know why a patient is classified as risky for a certain disease instead of having a machine telling his/her class even if this machine is more accurate. Given explanations alongside with decisions is richer for decision making and prevention of future diseases. To illustrate this observation, consider the work Deep EHR by Shickel *et al.* [21]. In that work, a review of proposed Electronic Health Records (EHR) representations mainly based on deep learning encoding techniques are compared. One of the limitations of all proposed representations is the lack of interpretability; the authors emphasize that this is an important aspect in the medical field but currently its coverage is very low among existing proposed solutions.

In this work we opted for the Dempster-Shafer using Gradient Descent Classifier (DSGD). This is an interpretable tabular classifier that can handle missing values and introduce expert knowledge for prediction. In order to understand the model we will briefly explain the Dempster-Shafer Theory (DST) and then we will explain the model itself.

C. DEMPSTER-SHAFER THEORY

The Dempster-Shafer Theory (DST) [22] is a mathematical framework to reason with uncertainty and incomplete data. It is often called a generalization of the Bayesian theory because is more expressive than classical Bayesian models due of the inclusion of uncertainty.

Let X be the set of all states of a system called frame of discernment. A mass assignment function, or simply mass, $\mathbf{m}$ is a function that satisfies:

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

where A is a subset of X and ϕ is the empty set. The term $\mathbf{m}(A)$ can be interpreted as the likelihood of getting exactly the outcomes of the set A, and not a subset of A. Masses are the elements that encode knowledge about the process.

The plausibility metric is defined as the total amount of evidence that can support an outcome. This formulation is the following:

$$Pl_m(A) = \sum_{B \cap A \neq \phi} \mathbf{m}(B) \tag{2}$$

Multiple evidence sources expressed by their mass assignment functions of the same frame of discernment can be combined using the Dempster Rule (DR) [23]. Given two mass assignment functions $\mathbf{m_1}$ and $\mathbf{m_2}$, a new mass assignment function $\mathbf{m_c}$ can be constructed by the combination of the other two using the following formula:

$$\begin{aligned} \mathbf{m_c}(A) &= \mathbf{m_1}(A) \oplus \mathbf{m_2}(A) \\ &= \frac{1}{1-K} \sum_{B \cap C = A \neq \phi} \mathbf{m_1}(B)\mathbf{m_2}(C) \end{aligned} \tag{3}$$

where K is a constant representing the degree of conflict between $\mathbf{m_1}$ and $\mathbf{m_2}$ and it is given by the following expression:

$$K = \sum_{B \cap C = \phi} \mathbf{m_1}(B)\mathbf{m_2}(C). \tag{4}$$

D. DEMPSTER-SHAFER USING GRADIENT DESCENT CLASSIFIER

The DSGD model uses DST principles in their computations which allows to express more complex scenarios while remaining simple and interpretable. Like any other classifier, given an input feature vector X, this model predicts the class associated to this record y' from a known set of possible classes. To apply DST our frame of discernment is the set of all classes. Then the DSGD model operates using masses from this domain and combining them using the Dempster Rule.

The model is based on rules that are composed by statements that can be verified with data and a mass assignment function which encodes the evidence for the records that satisfy the statement. We denote these rules as pairs (m, s) where m is the mass and s is the statement. An example of a statement of a rule is "The patient has diabetes". Rules can be defined by the user using his/her expert knowledge or can be generated automatically using statistics from the training data, allowing totally automatic classification or expert assisted classification. Just the statement must be provided to create a rule; the values of the mass assignment functions are initially set at random but with a high uncertainty.

To classify a record, the model looks for the rules in the rule set RS whose statement is satisfied by the values record. The mass assignment functions of these rules are combined using the Dempster Rule [23], then the model computes the plausibility of the combined mass. Finally, the predicted class is the one with highest plausibility.

$$\begin{aligned} M_x &= \{m \mid (m, s) \in \mathrm{RS} \wedge s(x)\} \\ m_f &= \bigoplus_{m \in M_x} m \\ y' &= \underset{class}{argmax}\ \mathrm{Pl}(m_f) \end{aligned} \tag{5}$$

The model also provides a training algorithm in which it "learns" the optimal values of these mass assignment functions producing the minimum error in classification. In the training phase, the model computes the loss of the predictions with the actual classes for all records in the training set. Then, the model uses this error as a target function for optimization and it applies Gradient Descent to update the values of the masses. This process is very similar to the training process of Neural Networks models; however, the model remains simple by the fact of relying on DST rather than using complex non-linear operations like those of Neural Networks.

Figure 1 summarizes the classification and training process. The input is a set of rules, a feature vector X and an actual class y. The model selects the rules which X satisfies; these rules are combined using Dempster Rule and then the plausibility for each class is computed, and the predicted class y' is the one with the highest value. Finally, the loss is computed using the actual class y and the mass values are updated by applying Gradient Descent.

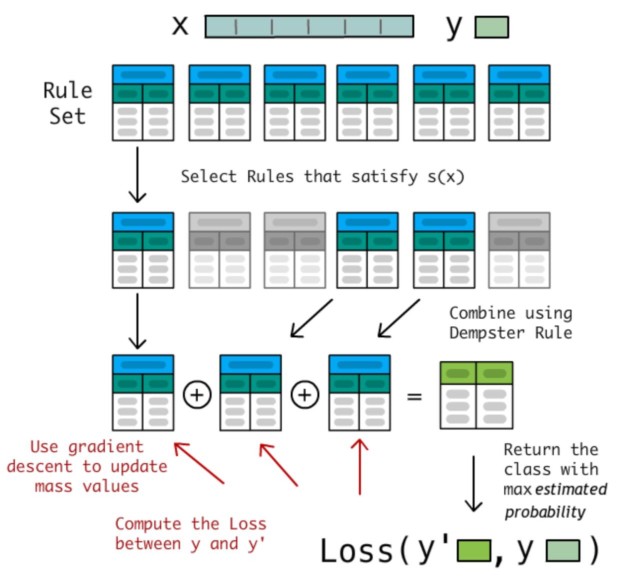

FIGURE 1. Training process in DSGD model.

The model is also interpretable since it is built using the meaningful rule statements which are not altered during training phase. The importance for prediction of each rule can be computed once the optimal values for the masses are set. In this work [10], the indicator γ measures rule importance and it is defined as the geometric mean between the mass m of the class k and the complement of the uncertainty.

$$\gamma(m, k) = \sqrt{m_k(1 - m_U)} \tag{6}$$

If γ takes a value 0 for a rule, it means that this rule does not contribute to the prediction of that class, while a value of 1 means that the rule is essential for the prediction of this class. After the γ values are computed we can sort the rules and identify the most important rules for the prediction of a particular class.

Finally, note that the model can handle missing data because it considers only the rules for which their statements can be verified. If an attribute of the statement is not present in a data record, the model skips this rule for this record.

III. DATA DESCRIPTION

There are three main data sources containing useful patient information which were used in our study:

1) **Patient Checkups**. This data source contains information about results of exams all Japanese workers have to undergo annually. Each exam record contains an exam identifier code, the date when it was collected and the result obtained. Examples of exams are LChO exam and blood pressure. There are 23 different types of exams.
2) **Patient demographics.** It contains general information about patients such as height, weight, body fat, age, gender, and waist measurement.
3) **Patient disease history.** This data source contains information about medical receipts and patients' diagnoses. The diseases are indexed according to ICD-10 codes [24] 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 all the data sources for a patient is not possible. If we completely exclude incomplete data, then around 95% of the dataset will be lost. If we decide to eliminate all this information then the method will have scarce data for training, therefore a method to handle missing values is required. However, data that cannot be joined is useless, because we will miss almost all information about the patient and because one of the goals of the study is to relate exam result to stroke risk.

The impact of missing values can be observed in Figure 2. It shows the 23 types of exams and the number of patients in the dataset with at least one result for the corresponding exam.

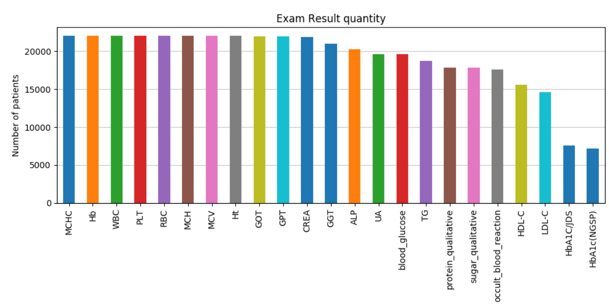

FIGURE 2. Types of exams and number of patient that have results for that exam.

After applying the described filters we selected 27876 patients' records, 22140 of them did not have strokes (79.4%), and 5736 had (20.6%). All patients have at least one checkup per year and patients who have more diseases tend to have also more exam results as expected.

IV. PROBLEM STATEMENT

The problem can be defined as follows: Determine whether or not a patient will have a stroke in the next year automatically using her/his historical medical information. This problem can be seen as a binary classification problem, where our input is the historical information defined in the previous section and the outcome are two classes, namely "The patient will have stroke within the next year", and "The patient will not have stroke within the next year".

The range of one year was chosen because it is a reasonable period of time for taking preventive action in the case there is high stroke risk. Concerning the data, there is a trade-off between the number of years ahead for the prediction (the higher, the better) and the amount of data available for training (the higher, the better). For example, if the model predicts a stroke within two years, at least these two years of data should be used for validation, and then less data will be available for learning the patterns and training.

V. PROPOSED MODEL

In this section the proposed model for addressing the stroke risk prediction problem is presented.

A. EMBEDDING

Embedding is the process of representing an observation or record as a structured data, typically, a vector. Considering the data sources already described, a straightforward embedding is to consider each patient as an individual row vector in a matrix. The column names are the attributes of all sources, i.e. the exam results, the past diseases and the demographics.

The problem with this representation is that for one patient there are many records of exam results, diseases and demographics within the data history. In order to solve this problem for the case of exam results and demographics, we propose to use the most recent results only, since most recent results should be most representative of the current condition of the patient.

For the case of disease history, the data is presented like a log showing the time when a patient was diagnosed with certain disease. This representation cannot be inputted directly into the vector; instead we can consider to have a column for each disease and mark it as 1 if the patient had the disease in the past and 0 if not. A drawback of this approach comes from the fact that ICD-10 defines more than 14000 different codes, and then the dimensionality of the feature vector will be very high thus slowing down the computations of any method. Moreover, this representation is very sparse because certain specific diseases occur only in few patients. To solve this problem, we used expert knowledge to select a group of diseases which are likely to be related with stroke occurrence; these diseases are: Type 2 diabetes mellitus (ICD-10: E11), Cerebrovascular diseases (ICD-10: I60-I69), Ischemic heart diseases (ICD-10: I20-I25) and Diseases of arteries, arterioles and capillaries (ICD-10: I70-I79).

B. CLASSIFIER AND RULES

The main features of the Dempster-Shafer Classifier using Gradient Descent presented by Peñafiel *et al.* [10] are: it can achieve similar accuracy compared to other classification methods such as Support Vector Machines or Random Forest; the model is rule-based and allows to generate rules automatically or to define custom rules according to data; the model can handle missing information which is a requirement according to the chosen embedding; and the model is interpretable which means that it can give an explanation of any prediction.

The rules of the model were defined according to the following process:

- For exams that are qualitative or semi-quantitative such as protein qualitative, sugar qualitative and occult blood reaction; also for past history of diseases such as diabetes, cerebrovascular diseases, cardiovascular diseases, and arteries-related diseases; and for gender; we create a rule for each possible outcome of these attributes.
- For quantitative exams and demographics, we used expert knowledge to define the cutoff values between normal values and abnormal values. These cutoff values are also called reference interval and they are normally used in the medical field for example the ones used internationally by Medscape [25]. The values we used are presented in Table 1.

TABLE 1. Cutoff values for attributes based on normal medical ranges.

Attribute	Cutoff values
WBC	4000, 10000
Ht	38
MCV	88, 102
MCH	27, 32
MCHC	30, 35
PLT	10, 40
UA	2, 7
LDL-C	120, 160
CREA	0.5, 1.2
TG	100, 200
HDL-C	40, 60
GOT	30, 100
GPT	30, 100
ALP	200, 1400
GGT	30, 100
Blood Glucose	100, 126, 200
Hb	7, 10, 13
BMI	18, 25, 30, 35, 40
Age	80
Body Fat	11, 22
Waist Measurement	95

Using these cutoff values we created a rule for each interval defined by these values, for example, for LDL-C there is a rule activated (this means, we apply the rule) when LDL-C is smaller than 120, another rule is activated when LDL-C is between 120 and 160, and finally another rule is activated when LDL-C is greater than 120.

- Finally, for quantitative measures we also created rules to be activated when pairs of attributes are outside their normal ranges in order to search for more complex relationships in data. An example of this kind of rules is a rule activated when UA is greater than 7 (above normal range) and PLT is smaller than 10 (below normal range).

This process generated a total of 558 different rules. The classifier was trained using Mean Squared Error (MSE) as loss function, and using ADAM optimizer [26] with a learning rate of 0.001.

VI. RESULTS

Results are divided in three parts. In the first part, the result of the model predictions and their performance metrics are shown. In the second part, the model is compared to the other stroke risk prediction methods presented in the related work and finally, an analysis of the interpretability of the obtained result is presented in the third part.

A. MODEL PERFORMANCE

After defining the embedding and configuration for our model in this problem, we tested the model performance using a 5-fold experiment over the dataset. In other words, we split the dataset into a training set (80%) and validation set (20%) randomly. We repeated the process five times, varying the data records belonging to the training and validation sets. For each fold, the following indicators were calculated: accuracy, sensitivity, specificity, F_1 macro, and the area under the ROC curve. Table 2 shows the average of these indicators among all folds.

TABLE 2. Performance results for stroke prediction problem.

Indicator	Value
Accuracy	0.854
Sensitivity	0.595
Specificity	0.878
F_1 macro	0.451
AUC ROC	0.875

These results will be compared to other ones in two cases in order to validate them. In the first case, we will compare our results to those of other machine learning techniques using the same embedding. In the second case, we will compare our model to current methods for stroke risk prediction from different sources such as clinical procedures and other data science proposed solutions.

B. COMPARISON WITH OTHER MACHINE LEARNING MODELS

We will compare our method to the following machine learning models: Random Forest with 100 trees, Naive Bayes, K-Nearest Neighbors with $K = 5$, Multilayer Perceptron with a single hidden layer of size 100, and SVM with RBF Kernel.

These other benchmark methods do not support missing values on data. This could be mentioned as an advantage of our model over the others. We opted for filling the missing values since removing data records changes the dataset, and then results become incomparable. We will use the mean imputation strategy for this task. This process consists of assuming that the best estimation for a missing value is the mean of the available data. Therefore within an attribute, we compute the average for the available information and use that value to fill the missing ones.

Table 3 shows the results of the different models and our model for the validation set. Furthermore, Figure 3 shows the ROC curve for all methods in the same chart to compare them easily.

TABLE 3. Results of various methods for the stroke risk problem.

Model	Accuracy	AUC ROC
DSGD	**0.854**	**0.875**
RF	0.849	0.861
NB	0.618	0.838
KNN	0.794	0.717
MLP	0.821	0.813
SVM	0.820	0.574
Majority	0.793	0.500

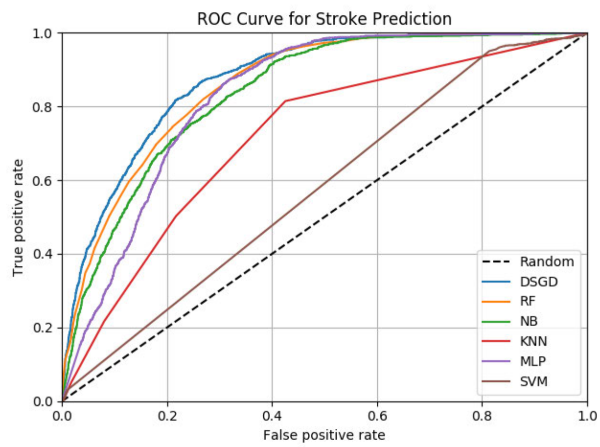

FIGURE 3. ROC curve for various methods in the stroke risk problem.

From the analysis of these results, we can observe that our model is the one that achieves the best performance for this problem in both accuracy and area under the ROC curve. In the ROC curve chart, we can see that our model (blue line) is always over the other model curves showing that it is the most accurate in the prediction of both classes.

We identified two main reasons why our model outperforms the other models. First, our model can handle missing values directly, which avoids the errors introduced by the mean imputation. In fact, we can see that the models that perform worst are KNN and SVM, which are based on geometry. In these cases, mean imputation causes many records to have a distance of 0 or colinearities in some dimensions, which introduces difficulties to the methods for performing correctly.

The second reason why we believe our method works better is that the rules we used to construct the model are meaningful and validated by medical studies. Every rule statement represents an actual condition that patients can have. This property helps the model to distinguish different patients better than the rest of the models.

C. COMPARISON WITH OTHER STROKE RISK ASSESSMENT METHODS

As explained before, many models have been proposed to evaluate the risk a patient has of getting a stroke. These methods offer baselines for our model since they are created with the same purpose. Some of them are very basic, but they are widely applied by clinicians, while others are complex and use state-of-the-art data science techniques.

In particular we tested the methods presented in section 2.1: CHADS_2, Bayesian Rule Lists (BRL), Multi-layer percetron (MLP) from Weng et al, Dempster-Shafer classifier using statistic estimation (DS-Statistical) and a RNN with a custom loss function. Table 4 shows the area under the receiver operator curve AUC score for each method from the results reported in their works. Due to the differences on the data used as input, these results may not be directly comparable but still they offer an approximation of the performance. From this table, we observe that our method outperformed other methods.

TABLE 4. Comparison of results for various stroke risk prediction methods.

Method	AUC
CHADS_2 [11]	0.721
BRL [12]	0.756
MLP [15]	0.764
DS-Statistical [9]	0.612
RNN with custom loss function [13]	0.669
DS-Gradient Descent	**0.838**

VII. INTERPRETABILITY

Besides the model performance and the reached accuracy, our model is able to extract the most important rules while predicting whether a patient will have a stroke.

In order to do this, all masses have been adjusted after the training phase to find the values that minimize the error when predicting stroke. Those values, according to Dempster-Shafer Theory, are assignments to every subset of possible outcomes; in our case, they will be the null set, the singleton for "No stroke" class, the singleton for "Stroke" class and the complete set. The mass of a singleton measures the contribution to a particular outcome, whereas the mass of the complete set measures the uncertainty of any outcome.

Having all these mass values, the procedure to extract the most contributory rules for a specific class is to compute the contribution score γ defined in Equation 6 as the geometric mean between the singleton of the class and the complement of the uncertainty.

Table 5 and Table 6 show the top 7 most important rules for the prediction of the classes "Stroke" and "No Stroke" respectively according to γ indicator.

From these tables, we can observe that the model can produce meaningful explanations for the whole problem and clearly distinguish and sort the contributory rules from the useless rules for each class.

TABLE 5. Most important rules for class "Stroke".

#	Rule	MNS	MS	Unc	γ value
1	HD cerebrovascullar = 1	0.000	0.755	0.245	0.755
2	$10 < \text{PLT} < 40$	0.002	0.410	0.589	0.411
3	$\text{Hb} > 13$	0.007	0.318	0.674	0.322
4	HD diabetes = 1	0.003	0.253	0.743	0.255
5	Body fat > 24	0.002	0.236	0.762	0.237
6	HDL-C > 60	0.000	0.213	0.787	0.213
7	Waist measurement > 95	0.007	0.207	0.786	0.210

MNS: Mass of "No Stroke" singleton, MS: Mass of "Stroke" singleton, Unc: Mass of Uncertainty.

TABLE 6. Most important rules for class "No Stroke".

#	Rule	MNS	MS	Unc	γ value
1	PLT > 40	0.853	0.000	0.147	0.853
2	GOT > 100	0.733	0.000	0.267	0.733
3	Hb < 7	0.676	0.003	0.321	0.677
4	GPT > 100	0.560	0.000	0.440	0.560
5	ALP > 1400	0.495	0.003	0.502	0.496
6	$7 < \text{Hb} < 10$	0.465	0.002	0.533	0.465
7	WBC > 10000	0.457	0.002	0.541	0.458

MNS: Mass of "No Stroke" singleton, MS: Mass of "Stroke" singleton, Unc: Mass of Uncertainty.

For the case of the "Stroke" class, the rule "HD cerebrovascular = 1", which means "the patient had a cerebrovascular disease in the past" is by far the most important rule while predicting this class with a γ score of 0.755. The second most contributory rule which relates the count of platelets in the blood (PLT) to low range of values has a contribution of 0.411, which is almost half of the contribution of the first rule. The following rules decrease their contribution more slowly. For the case of "No Stroke" rules, there is not a clear rule that defines that class. Alternatively, many rules achieve high contribution values, meaning that there exist many more distinct configurations for these variables that determine a healthy patient.

Furthermore, many of the rules that appear for the "Stroke" class also appear in their contrary form for the "No Stroke" class. This is the case of the count of platelets (PLT), which for lower values relates to the occurrence of strokes and for higher values related to healthy patients. Also, the higher values of Hemoglobin concentration (Hb) are related to stroke, whereas lower and average values are related to healthy patients.

Although the obtained rules that explain the decisions made by the model seem to be meaningful and coherent, we need to prove that. These interpretability results will be tested concerning their applicability and correct meaning. For that purpose, we will contrast the rules with current medical knowledge in order to see whether they are in accordance by presenting these rules to physicians and experts and ask them about their correctness.

VIII. VALIDATION

In this section, the interpretability results presented in the previous section will be compared to explanation models,

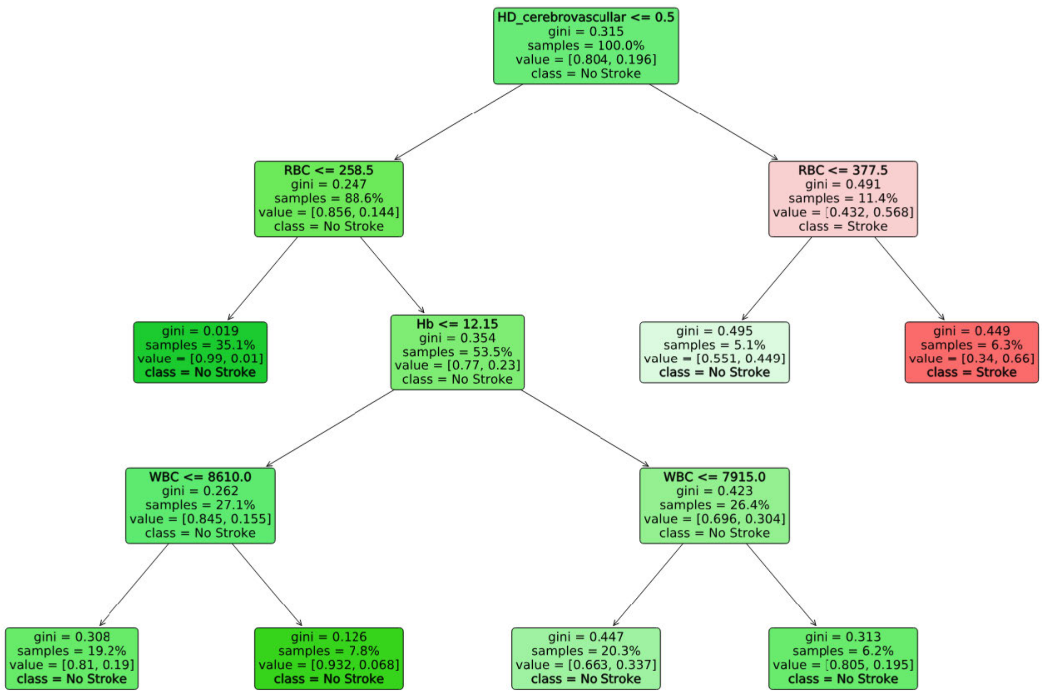

FIGURE 4. Decision Tree for stroke risk problem.

medical literature and an expert survey to validate whether these results are correct explanations for the problem.

A. DECISION TREE

The first benchmark model to compare our interpretability results is a Decision Tree. As we mentioned before, Decision Trees is one of the simplest and more interpretable models to make inferences and predictions. Also, they are widely used in non-computer science fields such as medicine, where many of the procedures to apply in certain cases are obtained from the result of a Decision Tree.

We build a Decision Tree using the CART strategy to find the splits [27]. We impose the condition that all nodes must have at least 5% of the samples of the dataset, and the maximum depth is 4.

The Figure 4 shows the resulting Decision Tree. In this figure, each inner node present the following information: the decision rule of the node, the impurity of the split, the proportion of the samples that were used to generate the split, the proportion of samples of each class (in the form $[p_{\text{No Stroke}}, p_{\text{Stroke}}]$), and the predicted class if the process is terminated up to this node. Leave nodes present the same information except for the decision rule. The color of the nodes are mapped to the classes. Green is for "No Stroke" class and red is for "Stroke" class, darker colors mean more certainty in the prediction.

Comparing the interpretability, we can see that the first node of the Decision Tree has a decision rule concerning the "HD cerebrovascular" attribute. This rule coincides with the most important rule to predict the stroke class according to our model (see Table 5). This result shows that our method can find the same most important attribute as the CART algorithm. Then, the next level of the tree shows two nodes with the "RBC" attribute for the rule. According to the Decision Tree, this is the second most important attribute for prediction, but in this case this attribute does not appear in our top 7 most important rules of either "Stroke" and "No Stroke" classes. Thus, this result reveals that our model and the Decision Tree differ. One explanation for this difference is that our model does not separate the data, it applies all rules evenly to patients that had stroke before or not. Then it may be more difficult to find a rule regarding the "RBC" attribute because of that restriction. Finally, analysing the deeper levels, we see that the features that appear are "Hb" and "WBC". The "Hb" also appears as the third most important rule for both classes in our method. The "WBC"' appears as the seventh most important rule for the "No Stroke" class.

B. LIME

Second we will compare the explanations of our model with the ones obtained using LIME [28]. LIME is a model able to extract local explanations of any classified instance regardless of the classification method. LIME has become widely used for interpretability purposes because it works even for models which behave like black-boxes such as complex Artificial Neural Networks.

Since LIME gives an explanation for single instances instead of the entire system, some data records (i.e., patients) will be selected to produce the explanations. Then they will be compared to our interpretability results. Also, as these

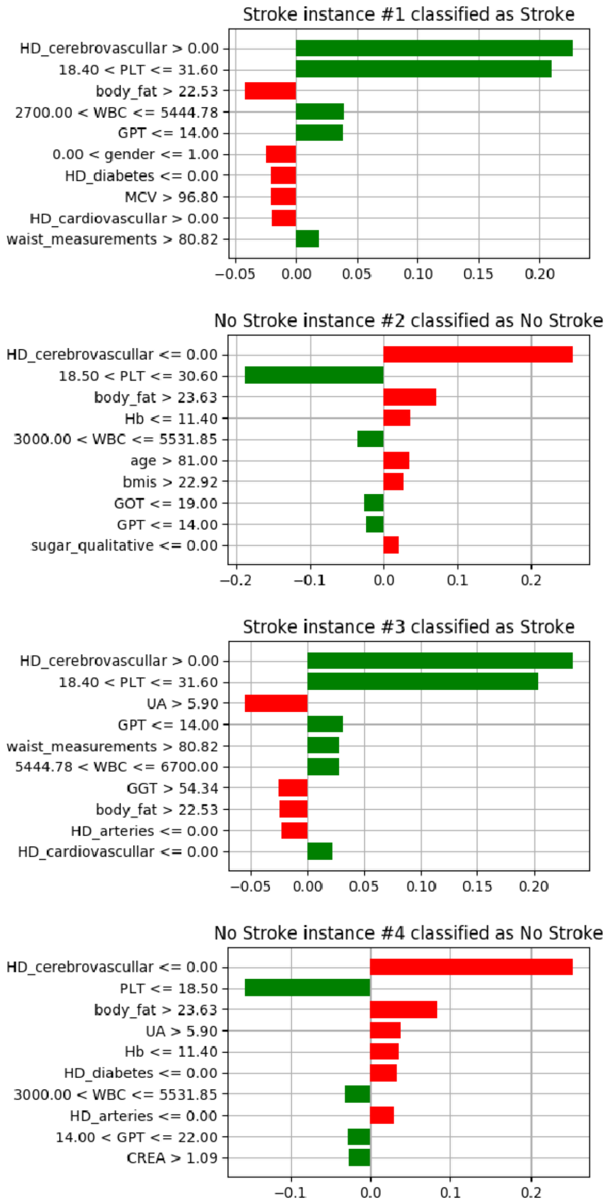

FIGURE 5. **LIME explanations for "Stroke" (first and second) and "No Stroke" instances (third and fourth).**

data records are part of our validation set, we know the class they belong to, thus several experiments can be performed. For example, explaining a true-positive or true-negative or failure cases, i.e., the false-positive and false-negative cases.

From the analysis of these results, we can observe that in the "Stroke" cases, the most important attribute is "HD cerebrovascular" which is the attribute that indicates whether or not the patient had a stroke in the past. When this attribute has a value of 1, LIME assigns a high contribution to the prediction of the Stroke class. The second most contributory attribute in all cases was "PLT", which is the count of platelets in the blood. In these cases, lower values of PLT support the prediction of the Stroke class.

For the "No Stroke" case, the most useful feature for prediction is "HD cerebrovascular" similar to the previous case. However, in this case, a value of 0 indicates a low stroke risk, being the opposite statement compared to the last example. Although this is not a rule our model obtains from Table 6 for the "No Stroke" class, it can be explained by the fact that it is the opposite of the most important rule for the other class. Moreover, we can observe in the explanations that high values for PLT, GOT, and GPT, are related to the prediction of the "No Stroke" class, similar to the rules obtained by our model.

C. CONTRASTING THE MODEL WITH MEDICAL LITERATURE

The rules of our model describe medical statements or hypotheses that can be verified or refuted based on real medical works reflecting the current knowledge of stroke risk in the medical field.

We will analyze from the medical point of view each one of the rules presented in Table 5 reported as the most important ones for predicting the "Stroke" class.

1) CEREBROVASCULAR DISEASE IN THE PAST

The first and most influential rule for classifying stroke is "HD cerebrovascular $= 1$". This rule states that patients who had suffered a cerebrovascular disease, including a stroke before, are more likely to have another one.

In the medical field, a recurrence of a disease happens when a patient had a disease, and then he or she has it again. Clinicians and experts study recurrence for most lethal diseases, including stroke. They also define the recurrence rate as the fraction of patients who suffered the illness again within a specified period of time.

Stroke recurrence is a well-known and documented clinical effect [29], [30]. Within two years, the recurrence rate among these patients is about 14% [31], and within five years, it increases to approximately 25%, which is at least five times more likely compared to patients who had not experienced a stroke before. Mortality also increases after a first stroke episode. In ten years, patients with stroke have a relative risk of death of 1.7 compared to healthy population [32].

All these facts point that knowing a patient has suffered a stroke is a significant piece of evidence for increasing her/his risk and then for the classification itself.

2) LOW VALUES OF PLATELETS

Rule 2 considers that a low platelet counting is related to the stroke risk. Platelets play an important role after a stroke occurs because experts believe that they participate in the thromboembolic creation that may initiate stroke falls. Although the relationship between the count of platelets in blood and stroke is still unclear, several studies have found that after the stroke occurred, the platelet count tends to decrease significantly. For example, D'erasmo *et al.* find an inverse correlation of -0.41 between platelets count and Hyperfibrinogenemia, which is related directly to the stroke event [33].

3) HIGH VALUES OF HEMOGLOBIN CONCENTRATION

Rule 3 states that high values of Glycosylated Hemoglobin concentration are associated with a high risk of stroke.

Like previous cases, there are prospective studies that validate this conclusion. For example, Rocco *et al.* [34] present a work which declares that a high concentration of Glycosylated Hemoglobin (HbA_{1c}) is an important factor for prediction of symptomatic intracerebral hemorrhage and acute stroke. The study also reports that Glycosylated Hemoglobin is a better predictor for stroke than blood glucose or a history of diabetes mellitus.

4) DIABETES CONDITION

Rule 4 is the next most influential rule for screening stroke according to our model. This rule states that patients who have been diagnosed with diabetes are more likely to have a stroke. Like other rules, several studies validate it.

In particular, Abbott *et al.* [35] present a 12-year follow-up study of 690 patients with diabetes and 6908 nondiabetic subjects. They conclude that the relative risk of thromboembolic stroke for those with diabetes compared with those without diabetes was 2.0 (95% confidence limits, 1.4 to 3.0).

5) HIGH LEVEL OF BODY FAT

Rule 5 relates a high level of body fat with a high risk of stroke. High levels of fat in blood and its accumulation is often designated as a strong predictor for acute stroke. For example, Walker *et al.* present a study relating obesity to stroke risk among US men [36]. They found that the relative risk of obesity patients is 1.29 relative to healthy patients.

As another example, Folsom *et al.* [37] performed a prospective study with 191 patients in order to determine associations of several factors, including fat distribution to ischemic stroke. They conclude that there is a relative risk of 1.74 when incrementing the body fat distribution. These findings validate our rule.

6) HIGH RATES OF HDL-C

Rule 6 states that high values of HDL-C are related to the risk of getting a stroke. This statement is the most controversial of the presented ones since most medical studies conclude that there is no evidence that high levels of HDL-C are related to cerebrovascular diseases. Instead, several prospective studies found that low concentration of HDL-C correlates with a high risk of atherosclerotic diseases, even when the mechanism of how this low concentration eases the disease occurrence is still unclear [38]. Therefore, these findings invalidate our rule.

However, let us recall that the patients we are analyzing are Japanese workers. Many medical findings are subject to the population where the study takes place. Thus some of these conclusions may not apply directly to other populations. The two studies that we presented before are about western people, which have different genetics and habits to eastern people. This fact of no representativeness could explain why our model finds this rule as necessary. For example, Saito *et al.* [39] say that despite low HDL-C is an established risk factor, evidence regarding stroke and stroke subtypes is very limited for Asian population. They hint that the relationship is inverse, which coincides with the rule obtained by our method.

Even when we can explain why this contradiction occurred, we still need to perform further studies to understand this causality correctly.

Finally, this example shows one of the most significant characteristics of the interpretability proposed by our model because any of the statements that do not have a clear interpretation in the expert field of the problem can guide future new research on that topic. This feature makes our model a helpful tool for knowledge discovery.

7) HIGH VALUES OF WAIST MEASUREMENTS

The last of the top 7 rules for predicting stroke class, relates high values of waist measurements with a high risk of stroke.

Unlike the previous cases, waist measurement is a metric that helps to build or find problems with other indicators such as body fat, cholesterol, and triglycerides. Hence, waist measurement is not a condition that directly affects the risk of cerebrovascular diseases.

For example, adiposity is a condition derived from the abnormality of several indicators such as waist measurements, and body mass index. Studies have confirmed that adiposity was associated with a high risk of total and ischemic stroke in men [40].

Moreover, it is valid to assume that high values of waist measurements are related to other conditions such as high weight, obesity, and high levels of body fat [41]. Then, as we explained in the fifth rule, high levels of body fat are indeed related to high stroke risk.

D. EXPERT SURVEY

Another test to assess the interpretability of the results of our model is an expert survey. In this survey, we presented medical experts, mainly neurologists, the rules that increase stroke risk according to the model results. For each of these rules, we asked whether they consider their statements are true, false, or if there is no correlation (NA) to stroke risk, according to their experience and knowledge. Some of the conditions were intentionally inverted to prevent biased answers to one option (e.g., an expert that responds all questions as true).

Table 7 shows the questions of the test and our expected results according to the interpretability results obtained. The questionnaire was built using Google Forms, which is widely used for this purpose and allows us to share it easily. The survey is in Spanish because we requested Chilean physicians and neurologists to respond to it, given our possibilities to conduct this test.

Moreover, we added two other rules that are less contributory to the estimation of stroke risk to check the "no correlation" option and to verify that the most important rules discovered by the model are also the most accepted statements among experts.

TABLE 7. Expert survey questions and results.

#	Rule	Question	Expected Result	True Responses	False Responses	NA Responses
1	HD cerebrovascullar = 1	A patient who had a stroke in the past is more likely to have a stroke again.	True	**100%**	0%	0%
3	Hb > 13	A patient who has a percentage of glycosylated hemoglobin (HbA1c) greater than 13% is more likely to have a stroke.	True	**94%**	0%	6%
4	HD diabetes = 1	A patient who has type 2 diabetes mellitus is less likely to have a stroke.	False	0%	**100%**	0%
5	Body fat > 24	A patient with a body fat index greater than 24% is more likely to have a stroke.	True	**88%**	0%	12%
6	HDL-C > 60	A patient who has a high-density cholesterol (HDL-C) level greater than 60 mg/dL is more likely to have a stroke.	True	31%	**63%**	6%
10	BMI < 18	A patient with malnutrition (BMI under 18) is less likely to have a stroke.	False and NA	12%	**63%**	25%
11	WBC < 4000	A patient who has a white blood cell count (WBC) less than 4000 per microliter is less likely to have a stroke.	False and NA	6%	31%	**63%**

A total of 16 participants correctly answered the survey. We appreciate the interest of these experts in answering the survey. The last three columns of Table 7 present the results of the relative frequency of each option for every rule. For each row we highlight the highest value which is the most frequent choice for this rule.

From these results, we can observe that all experts verify Rule 1 regarding stroke recurrence and Rule 4 about diabetes condition, since they all agree that these factors increase the stroke risk.

Almost all experts also confirm that Rule 3 concerning high levels of Hemoglobin raises the risk of getting a stroke, only one of them chose the ‘‘no correlation’’ option.

Likewise, almost all specialists consider Rule 5 concerning high values of body fat, as significant for the prediction of stroke. In this case, two of them selected the ‘‘no correlation’’ choice.

The results about Rule 6 are the most controversial since only one expert said that there is ‘‘no correlation’’ between high levels of HDL-C and stroke occurrence. However, the other experts had divided opinions about whether this variable increases or decreases the risk of getting a stroke. 62.5% of them said that the statement is false, thus invalidating our result, and 31.25% said that the result obtained by our method is valid. As we anticipated in the previous section, HDL-C is usually known as the ‘‘good cholesterol’’, then many clinicians consider that having a high value is healthy, which explains why most of them invalidate our finding.

Finally, Rules 10 and 11 concerning malnutrition and low levels of white blood cells are included for validation. These are rules that our model considers that are less important to the classification, and our expected result for these cases is that most experts choose the NA or False option. Looking at the results for these rules, we can verify our expected results. These are the rules with more NA choices, and in both cases, the number of False answers is higher than the True option.

IX. CONCLUSION

In this work we presented a novel model for predicting stroke occurrences within a year using data from electronic health records. This article presents the continuation of the work on stroke risk by Peñafiel *et al.* [9] and a use case for the model proposed by the same authors [10]. The proposed model achieves better performance metrics than our previous model and even outperforms many of the proposed and most used methods for stroke risk assessing in the case of some missing data. Since absence of some data occurs frequently in practice,the study asserts this model is a helpful tool for predicting stroke risk based on EHRs. We also showed that including expert knowledge for rule definition and using Gradient Descent for rule mass optimization is a powerful combination which may explain the model outstanding performance.

Another important contribution of our model is that it can extract the most contributory rules used for making predictions. This feature has several advantages like increasing the validity of the model, early detection of wrong inferences of the model and it also helps us validate the model clinically. We contrasted the 7 most important rules to medical literature. For the most influential rules, there were studies that directly correlate these statements to a high risk of getting a stroke. For the other rules there were studies that correlate them indirectly by an intermediate variable. We also asked experts through a survey whether they consider the rules found by our model raise stroke risk. In this survey most of the rules were verified.

Although the model achieves an outstanding performance, it can be improved in several ways, for example, incorporating additional data sources like patient conditions such as smoking habits or impairments, which are factors that other studies consider in their models.

Also, we can achieve better performance by improving the underlying model. One of the most important drawbacks of the model is that it discretizes the data into groups. We can get rid of that discretization by defining a score that expresses the degree in which a patient belongs to certain rule. This improvement could help the model differentiate more accurately between two similar patients which, in the current model, both patients may satisfy the same rules thus being assigned to the same class.

REFERENCES

[1] *Global Health Estimates 2016: Disease Burden by Cause, Age, Sex, by Country and by Region, 2000–2016*, WHO, World Health Organization, Geneva, Switzerland, 2018.

[2] E. J. Benjamin, P. Muntner, and M. S. Bittencourt, ''Heart disease and stroke statistics-2019 update: A report from the American heart association,'' *Circulation*, vol. 139, no. 10, pp. e56–e528, 2019.

[3] W. Johnson, O. Onuma, M. Owolabi, and S. Sachdev, ''Stroke: A global response is needed,'' *Bull. World Health Org.*, vol. 94, no. 9, pp. 634A–635A, 2016.

[4] *Noncommunicable Diseases Country Profiles 2014*, World Health Organization, Geneva, Switzerland, 2014.

[5] D. Bender and K. Sartipi, ''HL7 FHIR: An agile and RESTful approach to healthcare information exchange,'' in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 326–331.

[6] Y. Kudo, T. Satoh, S. Kido, M. Ishibashi, E. Miyajima, M. Watanabe, T. Miki, M. Tsunoda, and Y. Aizawa, ''The degree of Workers' use of annual health checkup results among japanese workers,'' *Ind. Health*, vol. 46, no. 3, pp. 223–232, 2008.

[7] J.-W. Lee, H.-S. Lim, D.-W. Kim, S.-A. Shin, J. Kim, B. Yoo, and K.-H. Cho, ''The development and implementation of stroke risk prediction model in national health insurance Service's personal health record,'' *Comput. Methods Programs Biomed.*, vol. 153, pp. 253–257, Jan. 2018.

[8] F. Doshi-Velez and B. Kim, ''Towards a rigorous science of interpretable machine learning,'' 2017, *arXiv:1702.08608*. [Online]. Available: http://arxiv.org/abs/1702.08608

[9] S. Penafiel, N. Baloian, J. A. Pino, J. Quinteros, A. Riquelme, H. Sanson, and D. Teoh, ''Associating risks of getting strokes with data from health checkup records using dempster-shafer theory,'' in *Proc. 20th Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2018, pp. 239–246.

[10] S. Peñafiel, N. Baloian, H. Sanson, and J. A. Pino, ''Applying dempster–shafer theory for developing a flexible, accurate and interpretable classifier,'' *Expert Syst. Appl.*, vol. 148, Jun. 2020, Art. no. 113262.

[11] B. F. Gage, A. D. Waterman, W. Shannon, M. Boechler, M. W. Rich, and M. J. Radford, ''Validation of clinical classification schemes for predicting stroke: Results from the national registry of atrial fibrillation,'' *Jama*, vol. 285, no. 22, pp. 2864–2870, 2001.

[12] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, ''Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model,'' *Ann. Appl. Statist.*, vol. 9, no. 3, pp. 1350–1371, Sep. 2015.

[13] D. Teoh, ''Towards stroke prediction using electronic health records,'' *BMC Med. Informat. Decis. Making*, vol. 18, no. 1, p. 127, Dec. 2018.

[14] S. Hochreiter and J. Schmidhuber, ''Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[15] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, ''Can machine-learning improve cardiovascular risk prediction using routine clinical data?'' *PLoS ONE*, vol. 12, no. 4, Apr. 2017, Art. no. e0174944.

[16] F. Mittag *et al.*, ''Use of support vector machines for disease risk prediction in genome-wide association studies: Concerns and opportunities,'' *Hum. Mutation*, vol. 33, no. 12, pp. 1708–1718, Dec. 2012.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ''Learning representations by back-propagating errors,'' *Nature*, vol. 323, no. 6088, p. 533, 1986.

[18] N. Friedman, D. Geiger, and M. Goldszmidt, ''Bayesian network classifiers,'' *Mach. Learn.*, vol. 29, no. 2, pp. 131–163, Nov. 1997.

[19] J. D. Olden and D. A. Jackson, ''Illuminating the 'black box': A randomization approach for understanding variable contributions in artificial neural networks,'' *Ecolog. Model.*, vol. 154, nos. 1–2, pp. 135–150, Aug. 2002.

[20] I. Bratko, ''Machine learning: Between accuracy and interpretability,'' in *Learning, Networks and Statistics*. Vienna, Austria: Springer, 1997, pp. 163–177.

[21] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, ''Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis,'' *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1589–1604, Sep. 2018.

[22] G. Shafer, *A Mathematical Theory of Evidence*, vol. 1. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[23] G. Shafer, ''Dempster's rule of combination,'' *Int. J. Approx. Reasoning*, vol. 79, pp. 26–40, Dec. 2016.

[24] *International Classification of Functioning, Disability and Health: ICF*, World Health Organization, Geneva, Switzerland, 2001.

[25] A. Farinde, ''Lab values, normal adult: Laboratory reference ranges in healthy adults,'' Medscape, New York, NY, USA, Tech. Rep., 2019, Art. no. 2172316. [Online]. Available: https://emedicine.medscape.com/article/2172316-overview

[26] D. P. Kingma and J. Ba, ''Adam: A method for stochastic optimization,'' 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[27] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification Regression Trees*. Boca Raton, FL, USA: CRC Press, 1984.

[28] M. Ribeiro, S. Singh, and C. Guestrin, '''Why should i trust you?': Explaining the predictions of any classifier,'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Demonstrations*, 2016, pp. 1135–1144.

[29] H. S. Jerrgensen, H. Nakayama, J. Reith, H. O. Raaschou, and T. S. Olsen, ''Stroke recurrence: Predictors, severity, and prognosis. The copenhagen stroke study,'' *Neurology*, vol. 48, no. 4, pp. 891–895, Apr. 1997.

[30] R. L. Sacco, P. A. Wolf, W. B. Kannel, and P. M. McNamara, ''Survival and recurrence following stroke. The framingham study,'' *Stroke*, vol. 13, no. 3, pp. 290–295, May 1982.

[31] D. B. Hier, M. A. Foulkes, M. Swiontoniowski, R. L. Sacco, P. B. Gorelick, J. P. Mohr, T. R. Price, and P. A. Wolf, ''Stroke recurrence within 2 years after ischemic infarction,'' *Stroke*, vol. 22, no. 2, pp. 155–161, Feb. 1991.

[32] M. Prencipe, F. Culasso, M. Rasura, A. Anzini, M. Beccia, M. Cao, F. Giubilei, and C. Fieschi, ''Long-term prognosis after a minor stroke: 10-year mortality and major stroke recurrence rates in a hospital-based cohort,'' *Stroke*, vol. 29, no. 1, pp. 126–132, Jan. 1998.

[33] E. D'erasmo, M. Acca, F. Celi, F. Medici, T. Palmerini, and D. Pisani, ''Plasma fibrinogen and platelet count in stroke,'' *J. Med.*, vol. 24, nos. 2–3, pp. 185–191, 1993.

[34] A. Rocco, P. U. Heuschmann, P. D. Schellinger, M. Kohrmann, J. Diedler, M. Sykora, C. H. Nolte, P. Ringleb, W. Hacke, and E. Juttler, ''Glycosylated hemoglobin a1 predicts risk for symptomatic hemorrhage after thrombolysis for acute stroke,'' *Stroke*, vol. 44, no. 8, pp. 2134–2138, Aug. 2013.

[35] R. D. Abbott, R. P. Donahue, S. W. MacMahon, D. M. Reed, and K. Yano, ''Diabetes and the risk of stroke: The honolulu heart program,'' *Jama*, vol. 257, no. 7, pp. 949–952, 1987.

[36] S. P. Walker, E. B. Rimm, A. Ascherio, I. Kawachi, M. J. Stampfer, and W. C. Willett, ''Body size and fat distribution as predictors of stroke among US men,'' *Amer. J. Epidemiology*, vol. 144, no. 12, pp. 1143–1150, Dec. 1996.

[37] A. R. Folsom, M. L. Rasmussen, L. E. Chambless, G. Howard, L. S. Cooper, M. I. Schmidt, and G. Heiss, ''Prospective associations of fasting insulin, body fat distribution, and diabetes with risk of ischemic stroke. The atherosclerosis risk in communities (ARIC) study investigators,'' *Diabetes Care*, vol. 22, no. 7, pp. 1077–1083, Jul. 1999.

[38] A. M. Gotto and E. A. Brinton, ''Assessing low levels of high-density lipoprotein cholesterol as a risk factor in coronary heart disease: A working group report and update,'' *J. Amer. College Cardiol.*, vol. 43, no. 5, pp. 717–724, 2004.

[39] 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, Oct. 2017.

[40] G. Hu, J. Tuomilehto, K. Silventoinen, C. Sarti, S. Männistö, and P. Jousilahti, ''Body mass index, waist circumference, and waist-hip ratio on the risk of total and type-specific stroke,'' *Arch. Internal Med.*, vol. 167, no. 13, pp. 1420–1427, 2007.

[41] T. B. Harris, M. Visser, J. Everhart, J. Cauley, F. Tylavsky, T. Fuerst, M. Zamboni, D. Taaffe, H. E. Resnick, A. Scherzinger, and M. Nevitt, ''Waist circumference and sagittal diameter reflect total body fat better than visceral fat in older men and women: The health, aging and body composition study,'' *Ann. New York Acad. Sci.*, vol. 904, no. 1, pp. 462–473, Jan. 2006.

SERGIO PEÑAFIEL was born in Coquimbo, Chile, in 1994. He received the degree from the Faculty of Physical and Mathematical Sciences, Universidad de Chile, in 2013, and the master's degree in computer science in 2020. He worked as a Research Assistant eieth CEAMOS Santiago Chile in urban crime prediction to optimize police patrolling. He was an Intern at Allm Inc., Tokyo, Japan, collaborating in disease prediction using machine learning. He is the co-founder of Create a Chilean startup aiming to bring machine learning solutions to industry. His thesis work and current research is about interpretable methods for classification and regression and its applications in real scenarios.

NELSON BALOIAN received the degree in engineering from the University of Chile and the Ph.D. degree from the University of Duisburg, Germany. He has been the Chairman of the Computer Science Department, University of Chile, where he is currently an Associate Professor of computer science with the University of Chile. He has been a Visiting Professor with the University of Duisburg-Essen, Germany, Waseda University, Japan, and the American University of Armenia, Armenia. His research interests are collaborative computer supported learning, mobile computing, and geo-collaboration. He has authored about 80 conference articles and 20 journal articles in these areas.

HORACIO SANSON received the master's degree from the Global Information and Telecommunication Institute, Waseda University, Tokyo, Japan, in 2005. He was with Waseda University, where he researched distributed localization on emergency wireless networks. He also held several engineering positions. He became a Lead Research Engineer with Allm Inc., in 2014, where he applies engineering and machine learning to solve practical medical and health-care issues.

JOSÉ A. PINO is currently a Full Professor of computer science with the University of Chile. He has coauthored six books and published research articles 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*. His research interests include ubiquitous computing, human–computer interaction, and socio-technical systems. He is currently a member of the Editorial Board of the *Journal of Network and Computer Applications*. He has served as the President for the Chilean Computer Science Society (SCCC) and the Latin American Association of Universities Concerning Information Technology (CLEI).