IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Explainable Prediction of Chronic Renal Disease in the Colombian Population using Neural Networks and Case-Based Reasoning

Gabriel R. Vásquez-Morales[1], Sergio M. Martínez-Monterrubio[2], Pablo Moreno-Ger[3] and Juan A. Recio-García[4]

[1]Office of Information and Communications Technology. Ministry of Health and Social Protection. Carrera 13 # 32-76, Bogotá, Colombia. gvasquez@minsalud.gov.co

[2]Group of Artificial Intelligence Applications. Department of Software Engineering and Artificial Intelligence. Faculty of Computer Science, Office 420, Universidad Complutense de Madrid. Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, Spain. sergim13@ucm.es

[3]School of Engineering, Universidad Internacional de La Rioja (UNIR), Avenida de la Paz 137, 26006 Logroño, La Rioja, Spain. pablo.moreno@unir.net

[4]Group of Artificial Intelligence Applications. Department of Software Engineering and Artificial Intelligence. Faculty of Computer Science, Office 420, Universidad Complutense de Madrid. Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, Spain. jareciog@ucm.es

Corresponding author: Juan A. Recio-Garcia (e-mail: jareciog@fdi.ucm.es).

This work was supported by the Spanish Committee on Economy and Competitiveness (TIN2017-87330-R).

ABSTRACT This paper presents a neural network-based classifier to predict whether a person is at risk of developing chronic kidney disease (CKD). The model is trained with the demographic data and medical care information of two population groups: on the one hand, people diagnosed with CKD in Colombia during 2018, and on the other, a sample of people without a diagnosis of this disease. Once the model is trained and evaluation metrics for classification algorithms are applied, the model achieves 95% accuracy in the test data set, making its application for disease prognosis feasible. However, despite the demonstrated efficiency of the neural networks to predict CKD, this machine-learning paradigm is opaque to the expert regarding the explanation of the outcome. Current research on eXplainable AI proposes the use of twin systems, where a black-box machine-learning method is complemented by another white-box method that provides explanations about the predicted values. Case-Based Reasoning (CBR) has proved to be an ideal complement as this paradigm is able to find explanatory cases for an explanation-by-example justification of a neural network's prediction. In this paper, we apply and validate a NN-CBR twin system for the explanation of CKD predictions. As a result of this research, 3,494,516 people were identified as being at risk of developing CKD in Colombia, or 7% of the total population.

INDEX TERMS Chronic kidney disease prediction, neural networks, case-based reasoning, twin systems, explainable AI, Support Vector Machines, Random Forest.

I. INTRODUCTION

One of the fields of application of artificial intelligence today is the health sector. Machine learning algorithms have been widely used in disease prediction and classification tasks. Some algorithms such as SVM, Random Forest and neural networks have been used to predict and classify patients with diabetes [1] [2], Alzheimer [3], heart disease [4], cancer [5] [6] and liver cirrhosis [7], among others. Neural networks [8] [9] [10] and, in general, machine learning algorithms [11] [12] have now begun to be used successfully as support tools for the diagnosis and early detection of diseases [13] [14] [15] [16] [17]. These algorithms are trained with large volumes of data, including diagnostic images, laboratory tests and medical records. This study proposed the use of a neuronal network for the prognosis of chronic renal disease in the Colombian population. In Colombia, by 2017, there were 1,406,364 people with Chronic Kidney Disease, a prevalence of 2.9 cases per 100 inhabitants [18]. Figure 1 shows the increase in the prevalence of people with CKD in recent years.

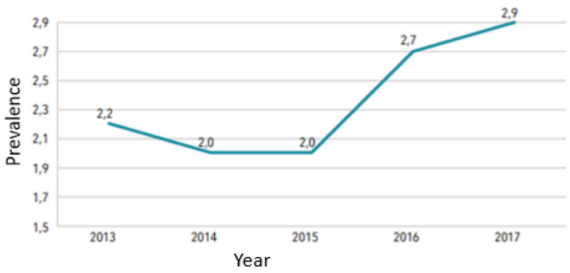

FIGURE 1. Prevalence of CKD in Colombia. The y axis shows the number of people with CKD per 100 inhabitants.

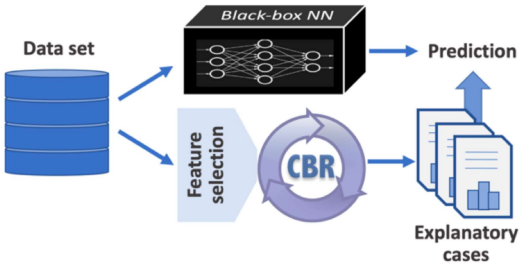

FIGURE 2. NN-CBR Twin system for the explanation of CKD predictions

Neural networks have been used for the detection of CKD and other pathologies related to the urinary system. Some models have made it possible to detect kidney stones [19] [20] from the results of laboratory tests such as creatinine, uric acid, glucose, lymphocytes and other blood components. Other work has focused on predicting the survival of patients with CKD [21]. Neural networks and other machine learning techniques have also been applied to identify a patient's stage of chronic kidney disease [22] [23] [24]. Other studies propose a neural network model for detecting CKD from patient laboratory data [25] [26] [27], as well as comparisons with other machine learning models [28] [29]. The results of the evaluation metrics have demonstrated the effectiveness of models trained with neural networks, obtaining values of up to 97% accuracy [30]. One study that comes very close to the objective of this paper is [31], in which a hybrid neural network was developed in order to predict CKD in patients with hypertension from their medical records. The results of the experiment showed that the neural network was able to correctly predict CKD cases with an accuracy of 89.7%.

Despite the tremendous performance of neural networks, they work as black-box systems and their effectiveness is limited by their inability to explain their predictions to the experts. The problem of explainability in Artificial Intelligence is not new but the rise of the deep-learning as a very successful classification technique has created the necessity to understand how these systems make a prediction in order to increase users' reliability and trust.

Complementarily to the development of an accurate neural network for the prediction of CKD, the second main contribution of this paper is the use of Case-based Reasoning (CBR) [32] for the generation of explanations associated to the prediction of the neural network. CBR systems are claimed to have a "natural" transparency as they are based on the reuse of previous experiences or examples.

Therefore, this paper proposes a particular solution for the explanation of the outcomes of the neural network to the experts, where this opaque, black-box machine learning system is explained by a more interpretable, white-box CBR system; following the so-called twin-systems approach [33]. This approach is illustrated by Figure 2, where the dataset is used as the input of the NN and to create the cases of the CBR system after a feature selection preprocessing. Next, the prediction of the NN is explained by means of the most similar explanatory cases to the patient description, following the explanation-by-example paradigm.

The paper runs as follows. Section II describes in detail the proposed twin system, including the dataset collection, the NN implementation and the CBR explanation system. Section III presents the evaluation of the approach and the associated results. Section IV contains the discussion of the impact of the research, and Section V concludes the paper and outlines the lines of future work.

II. METHODS

To train the classifier, information was taken from two population groups: 20,000 patients diagnosed with CKD in 2018, and an equal sample of healthy people. Demographic information was collected for both population groups, including sex, age, place of residence and ethnicity, along with the history of disease diagnoses, obtained from the RIPS database [34] (Individual record of health service delivery) of the Colombian Ministry of Health and Social Protection. Based on this data set, training and optimization of the neural network was carried out, as well as two other machine learning models, one with vector support machine (SVM) and the other with Random Forest.

A. CLEANING AND PREPARATION OF THE DATA SET

The data set required for neural network training was obtained from the RIPS database of the Colombian Ministry of Health and Social Protection. This database contains information on health care provided to all persons affiliated with the country's health system since 2009. In order to obtain the sample of individuals with CKD, the persons whose diagnoses in the RIPS database corresponded to codes within subgroup N18 of the International Classification of Diseases (ICD) [35] were identified. This subgroup corresponds to diagnoses of "chronic renal insufficiency". Those persons whose diagnosis date corresponded to the year 2018 were then selected. The number of people obtained in this consultation was 203,015. Considering that model training can be costly with such a large data set, a random sample of 20,000 people was taken.

The set of people in the control group was selected by filtering persons that required any health care during 2018, and a random sample was taken with the same number of elements existing in the group of people with CKD, i.e. 20,000

individuals. In this group of people, a filter was carried out to discard those who had previously been diagnosed with CKD. Once the persons had been identified, the two sets of data were integrated and the demographic variables corresponding to the person's sex, age, ethnicity and department of residence were added. All the diagnoses identified in the RIPS database for each person were also aggregated, as well as the number of medical attentions carried out as a result of this diagnosis. These data were used to select characteristics, create categorical variables, treat null values and other cleaning operations required to obtain the final data set of 40,000 records and 7,493 variables, including the class or output variable. Figure 3 illustrates this process.

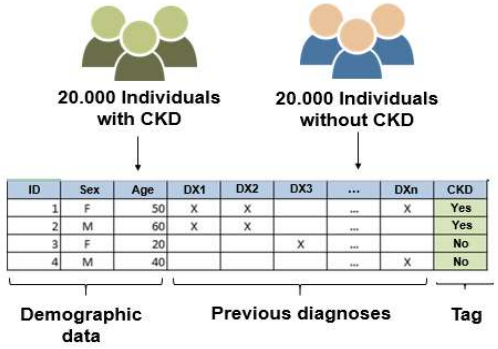

FIGURE 3. Data sets for model training.

B. DEFINITION OF THE NEURAL NETWORK TOPOLOGY

The neural network was designed with a topology composed of 5 layers according to the diagram in Figure 4. The input layer corresponds to the characteristics or input variables of the network, with 7,492 nodes or neurons. The next 3 layers correspond to the hidden layers of the model and contain 500, 100 and 50 neurons respectively. The last layer corresponds to the neuron that represents the only class of the binary classification problem. The objective of the training is to obtain the optimal values for the parameters of the network, composed by the weights (W) of each layer and by the values of bias [36].

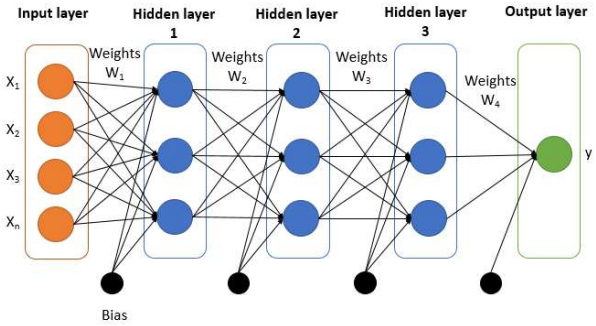

FIGURE 4. Topology of the neural network with 5 layers.

Layer (type)	Output Shape	Param #
dense_44 (Dense)	(None, 500)	3746500
dense_45 (Dense)	(None, 100)	50100
dense_46 (Dense)	(None, 50)	5050
dense_47 (Dense)	(None, 1)	51

Total params: 3,801,701

FIGURE 5. Summary of the design of the neural network.

Once the model was defined, training was performed using backpropagation [37] together with the gradient descent algorithm, although in practice a variation of this algorithm known as stochastic gradient descent (SGD) [38] is used, which is less expensive computationally. Figure 5 presents the summary of the neural network developed with the Keras library [39] and the TensorFlow framework1.13.1 [40], along with the number of parameters to train for each layer.

The number of parameters in each layer corresponds to the number of nodes in the previous layer, multiplied by the number of nodes in the current layer, plus a bias value for each node in the current layer. For example, for the first hidden layer the number of parameters is given by the number of nodes of the input layer (7.492) multiplied by the number of nodes of the first hidden layer (500), plus 500 bias values corresponding to each node of the current layer. This adds up to a total of 3,746,500 parameters for the first hidden layer. For the remaining layers the number of parameters is 50,100, 5,050 and 51, giving a total value of 3,801,701 parameters for the entire network. For the training of the network 70% of the data was used and 10 iterations or epochs were performed with a lot size of 1,000 elements applying cross validation. Once the model was trained, the remaining 30% of the data was used as a set of tests to identify the performance obtained with the initial model. As a result, an accuracy value equal to 49.91% was obtained. Considering that this is a very low value, it was necessary to adjust in some hyperparameters used in the training of the model. This optimization included changes in the activation function, the training algorithm, as well as normalization and early stopping functions.

C. OPTIMIZATION OF HYPERPARAMETERS

During the training of the initial network model, the accuracy value was too low, so the sigmoid activation function was changed to a ReLU function [41]. The accuracy value obtained with this setting in the model was 59.96%. Figure 6 compares the value of accuracy obtained for each model. In green the values obtained with the sigmoid activation function and in blue the values obtained with the ReLU activation function.

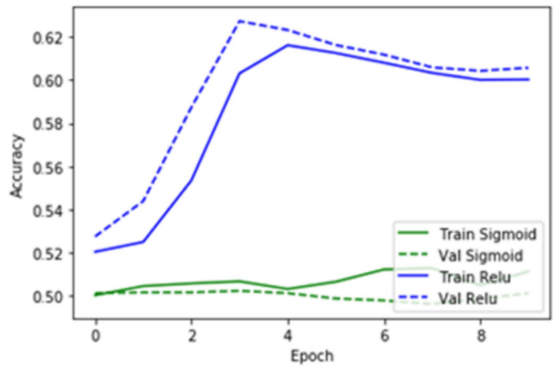

FIGURE 6. Comparison of accuracy obtained with different activation functions.

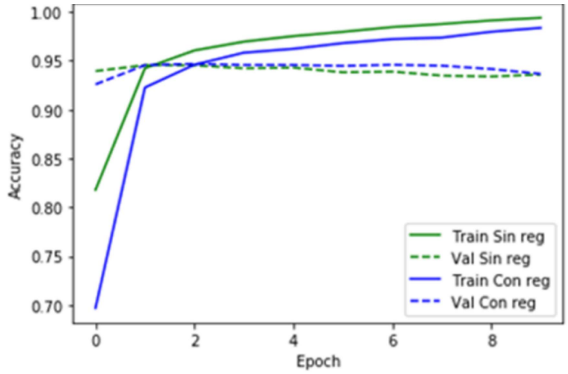

FIGURE 8. Comparison of accuracy obtained with different optimizers.

As can be seen in the graph above, the ReLU activation function allows the network to improve the accuracy measurement, even though this value is still below the desired threshold. In order to improve network performance, the optimization algorithm was changed from SGD to Adam [42] [43]. This optimization resulted in an accuracy value of 92.65%. This value considerably improves the metric obtained previously with SGD equal to 59.96%. It also exceeds the 90% threshold proposed in the project objectives. Figure 7 shows the accuracy measurement in both the training set and the validation set using Adam optimization (in blue), compared to that obtained previously with SGD (in green).

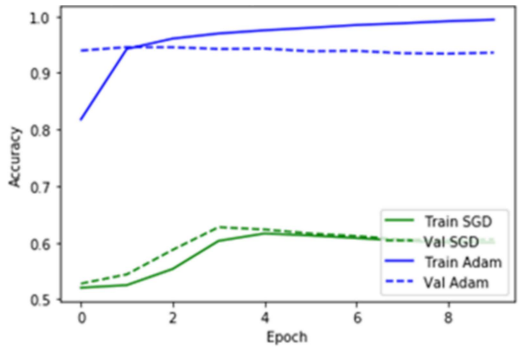

FIGURE 7. Comparison of accuracy obtained with different optimizers.

The figure above shows how the accuracy value reaches values above 90%, although the model tends to overfitting. This is because the algorithm is memorizing the training data and is not able to generalize what has been learned to the validation set. To counteract the effect of overfitting, a normalization technique known as dropout [44] was applied. Once the dropout technique is applied to the output of each layer of the network, with a hyperparameter value equal to 0.5, an accuracy value equal to 93.71% is obtained, improving as compared to that obtained previously without applying regularization. The following figure shows the comparison between error and accuracy values for models with and without regularization. Figure 8 shows that the model trained by applying dropout (in blue) reduces the presence of overfitting, although it does not eliminate it completely.

Another technique widely used in practice to prevent over-adjustment is early stopping [45]. This technique makes it possible to identify the iteration or period in which the net begins to fall into overfitting and stops training at that moment. The accuracy obtained after optimizing the model corresponds to a value of 95%. Figure 9 presents the evolution of the accuracy in the final neural network model. It can be observed that the number of training times decreased to 3 and that the accuracy of the training set and the set of tests converge at one point in the graph.

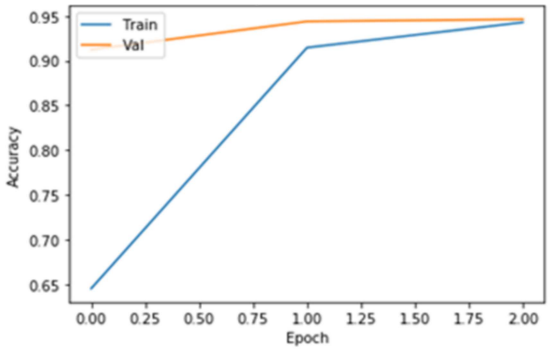

FIGURE 9. Accuracy in the final neural network model.

D. EXPLANATION OF THE NEURAL NETWORK PREDICTIONS USING CASE-BASED REASONING

Case-based reasoning is a paradigm for combining problem-solving and learning that has become one of the most successful applied subfields of artificial intelligence in recent years [46]. CBR is based on the intuition that problems tend to recur, so that new problems are often similar to previously encountered problems and, therefore, past solutions may be of use in the current situation [47]. CBR basically consists on a retrieval stage where the most similar cases to the given

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2948430, IEEE Access

		Age	E119	M255	N390	I10X	J449	M542	E109	R104	R51X	I839	R520	H400	M791	M545	CKD
Example	2727	67	0.0	1.0	1.0	0.0	0.0	0.0	0.0	3.0	7.0	0.0	1.0	6.0	0.0	4.0	1

		Age	E119	M255	N390	I10X	J449	M542	E109	R104	R51X	I839	R520	H400	M791	M545	CKD
Neighbor 1	6837	66	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	6.0	0.0	2.0	1
Neighbor 2	24540	70	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0	1.0	0.0	6.0	0.0	4.0	1
Neighbor 3	5855	69	0.0	1.0	6.0	13.0	0.0	0.0	0.0	1.0	5.0	0.0	0.0	7.0	1.0	0.0	1

FIGURE 10. Example of the explanation-by-example approach.

query are retrieved, and an adaptation stage that combines the retrieved cases to build a solution.

When applied to classification problems, CBR is considered a lazy learning approach where instead of generating some kind of abstract representation of the set of training examples, it uses those training examples in the neighborhood of the problem example (k Nearest Neighbours) [48] to determine its class.

As CBR is considered a transparent technique, it is being applied for the explanation of opaque machine-learning techniques such as neural networks [33] [49] [50]. There are several proposals to create NN-CBR twins, where the CBR system is not used to solve the problem, but to find explanatory cases for the user. Although CBR can be also applied to solve the classification problem, the performance of this paradigm is not comparable to neural networks. This way, these twin systems follow an explanation-by-example approach where the justification given to the user is based on the comparison of the input of the NN to similar examples obtained by the CBR system. As Figure 2 shows, these NN-CBR twins require the dataset to be available for both systems. This way, every instance in the dataset can be considered as an explanation case. However, the performance of the CBR process is influenced by the number of features/variables that represent each case due to the underlying nearest neighbor method. Therefore, the CBR system requires a preliminary stage of feature selection where the most significant variables are selected to represent an explanatory example.

To obtain the most relevant variables to represent explanatory cases we have reused the information given by the random forest classifier built to evaluate the performance of the NN and presented in the following section. We have selected the 15 most discriminant variables according to this classifier. These variables are shown in Table I, where the age of the patient is most relevant feature, followed by common diseases associated to CKD.

The complete representation of explanatory cases includes these variables and the actual classification value. By using a nearest neighbor retrieval method, we obtain the most similar cases to the given query. In case an example retrieved by the CBR does not correspond to the classification provided by the NN, it is not taken into consideration.

Then the adaptation stage is in charge of combining the most similar explanatory cases to generate the explanation through the pairwise comparison of the variables. Age is compared by using a threshold (10 years[1]), whereas the remaining variables are considered as binary values where any value different to 0 is a positive feature. As an illustrative example, Figure 10 shows how our system highlights the features in common between the given query and the 3 most similar explanatory cases. This way, the user is provided with a justification of the outcome based on the most significant medical variables.

TABLE I
VARIABLES SELECTED TO DESCRIBE EXPLANATION CASES

Variable	Description	Relevance
Age	Patient age in years	0.055603
E119	Type 2 diabetes mellitus without complications	0.035742
M255	Pain in joint	0.033023
N390	Urinary tract infection, site not specified	0.030263
I10X	Essential (primary) hypertension	0.028844
J449	Chronic obstructive pulmonary disease, unspecified	0.021672
M542	Cervicalgia	0.020566
E109	Type 1 diabetes mellitus without complications	0.019875
R104	Pain localized to other parts of lower abdomen	0.018675
R51X	Headache	0.016306
I839	Asymptomatic varicose veins of lower extremities	0.010781
R520	Acute pain	0.010668
H400	Glaucoma suspect	0.010570
M791	Myalgia	0.010283
M545	Low back pain	0.010094

Additionally, the system can generate a description of the explanation in natural language using text templates. It is based on the identification of the common features between the patient description, the most similar explanatory case and the other similar cases. Following the example in Figure 10, the corresponding explanation is: "The diagnosis is *positive* because there are similar patients with CKD. All of them have a *similar age* and were diagnosed with *headache* and *glaucoma suspect*. Most of them were also diagnosed with *urinary tract infection* and *low back pain*". As this

[1] This 10 years threshold was defined by consulted CKD experts.

explanation is generated by collecting the common features between similar examples, it can be considered as a constructive adaptation scheme regarding the CBR terminology [51]. As Figure 10 illustrates, age variable is within the 10 years threshold for the three retrieved examples and they all share in common the R51X (headache) and H400 (urinary tract infection) diseases. Additionally, the text includes those diseases that the first neighbor has in common with the patient description and the other retrieved examples. In this case, N390 (urinary tract infection) and M545 (low back pain).

Once we have presented our explainable classification method, following section presents its evaluation and the corresponding results.

III. RESULTS

We have evaluated both subsystems: neural network and case-based reasoner. Firstly, the accuracy of the NN classification is compared to state-of-the-art classifiers in order to validate its performance. Next, the explanation system has been also evaluated by analyzing the explanatory cases provided by the CBR system.

A. NEURAL NETWORK EVALUATION

The objective of metrics for the evaluation of classification algorithms is to identify the predictive capability of the model. To achieve this objective, the classes predicted by the algorithm for each example of the test set are compared with the real value of the class and in this way, it is identified if the model managed to classify the data correctly. A widely-used technique is the dispersion matrix, in which the examples classified correctly and incorrectly are counted, grouping them into true positives, true negatives, false positives and false negatives [52]. Figure 11 graphically shows the confusion matrix for the final neural network model. As can be seen from the graph, the network correctly classified 95% of cases and only failed in 5%.

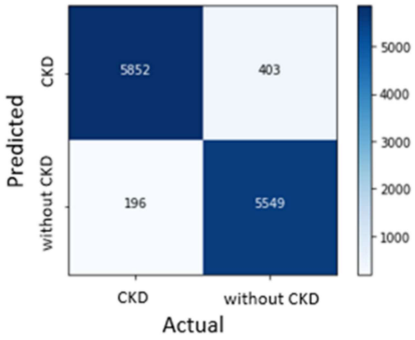

FIGURE 11. Confusion matrix of the neural network model.

From the confusion matrix, the evaluation metrics of classifiers indicated in Table II were obtained. These metrics confirm, on one hand, the predictive capability of the classifier regarding the accuracy metric, and on the other hand, a tendency of the model to predict with greater precision the positive rather than the negative examples. This can be observed bearing in mind that the sensitivity value, which measures the proportion of positive examples correctly classified, is higher than the specificity value, which measures the proportion of negative examples correctly classified.

TABLE II
NEURAL NETWORK EVALUATION METRICS

Metric	Value
Sensibility	97%
Specificity	93%
Precision	94%
Recall	97%
F-Measure	95%
Accuracy	**95%**
AUC	98%

Another useful metric to know the performance of binary classification models is the area under the curve (AUC) [53]. This measure establishes the algorithm's ability to identify the largest number of truly positive cases without falling into false positives. Figure 12 shows the ROC curve obtained along with its AUC value. The AUC value equal to 98% demonstrates an equilibrium in the behavior of the neural network, which strives to predict the greatest number of true positive cases, avoiding falling into false positives.

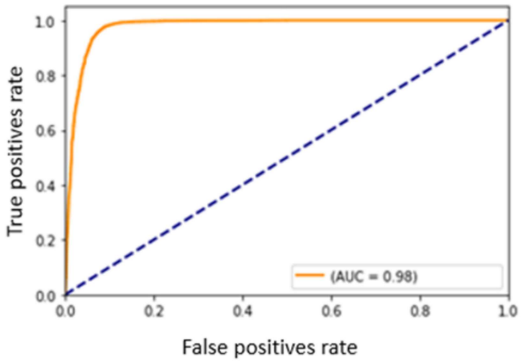

FIGURE 12. ROC/AUC curve.

B. COMPARISON WITH RANDOM FOREST AND SVM MODELS

Table III presents a comparison of the main metrics obtained after applying the neural network models, SVM [54] [55] [56] and Random Forest [57] [58]. The table shows that the best classifier is the neural network, considering an accuracy value of 95%. This is followed by Random Forest with an accuracy of 92%, and last is SVM with a value of 61%.

The sensitivity metric, which measures the proportion of positive examples correctly classified, shows a better performance of the neural network against Random Forest, and of these two against SVM. In terms of specificity, the model that best classifies negative examples is Random Forest, although the value obtained by the neural network is quite close. The same happens with the precision metric that indicates the proportion of examples that are truly positive.

TABLE III
MODEL COMPARISON METRICS

Metric	Neural Network	SVM	Random Forest
Sensibility	97%	69%	90%
Specificity	93%	53%	94%
Precision	94%	60%	93%
Recall	97%	69%	90%
F-Measure	95%	64%	91%
Accuracy	**95%**	**61%**	**92%**
AUC	98%	61%	92%

Exhaustiveness or recall values favor the neural network by a wide margin, which obtains a value of 97, as opposed to 90 for Random Forest and 69 for SVM. The F-value corresponds to a balance between precision and recall and simplifies the performance of a classification algorithm into a single metric. The neural network gets an F-value of 95, Random Forest of 91 and SVM of 64. Finally, the area under the curve (AUC) of the neural network model again outperforms the other algorithms because of its tendency to correctly identify true positives and avoid false positives. The set of applied metrics shows that the model trained with neural networks achieved an excellent performance, surpassing the threshold proposed in the project objectives. On the other hand, the model trained with Random Forest also obtains good results, very close to those of the neural network and is seen as a good alternative to the latter model. The performance of the SVM model for this case is considered very low and its implementation is discarded.

C. COMPARISON WITH CBR APPROACH

Although CBR can be also applied to provide a prediction for the CKD, in our approach we have used it as a twin system to provide explanations because the machine learning techniques usually achieve a better performance. However, this assumption must be validated. In case the CBR achieves a performance similar to NN or SVM these systems could be discarded and the CBR can be used for both the prediction and explanation.

For the evaluation of the CBR system a case base with 12.000 explanatory cases was randomly selected from the global dataset. These cases were described by the 15 most discriminant variables obtained by the random forest model as described in Section II.D and enumerated in Table I. The resulting confusion matrix of the cross-validation shows an average accuracy of 86% (Table IV shows additional metrics).

TABLE IV
CBR EVALUATION METRICS

Metric	Value
Accuracy	86%
Sensibility	84%
Specificity	88%
Precision	88%
Recall	84%
F-Measure	86%
Area under curve (AUC)	86%

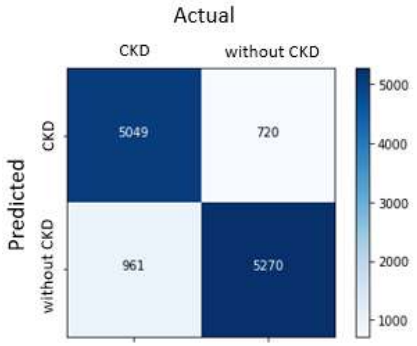

FIGURE 13. Confusion matrix of the case-based classifier.

As expected, the performance of CBR for the prediction of CKD is significantly worse than NN. Therefore, our system will use this technique to find explanatory examples instead of performing the classification task. Following subsection presents its evaluation.

D. EVALUATION OF THE CASE-BASED EXPLANATION METHOD

To evaluate the explanation method, we have analyzed the common features between the query –patient description- and the explanatory cases, using this metric as an estimation of their quality. The intuition behind this metric is that an explanatory case should be as similar as possible to the query. This way, the explanation given to the user will contain more feature-level justifications as we described in Section II.D. Concretely, we have obtained the three most suitable explanations for every case in the case base and analyzed the percentage of common features between them. Table V shows an example of this comparative analysis. For each query, the identifier of three most similar neighbors is provided together with the percentage of common features. As we can observe in this example, the most suitable explanatory case has a higher percentage of common features (around 60% to 66%) whereas the second and third nearest neighbors have lower feature overlap (46% to 53% and 26% to 53% respectively).

In order to understand the global efficiency of the CBR explanatory system we have computed the average common features between query and explanatory cases for the whole case base. Surprisingly, the average common features was similar for the three nearest explanatory cases: 29.,55% for

TABLE V
MOST SIMILAR NEIGHBORS COMPARATIVE ANALYSIS

Person	Neighbor1		Neighbor2		Neighbor3	
ID	**ID**	**Percent**	**ID**	**Percent**	**ID**	**Percent**
17051	19720	66.67	22343	46.67	19078	26.67
2950	21544	66.67	10981	46.67	21681	53.33
24627	5112	66.67	12425	46.67	12991	46.67
337	17200	66.67	2540	46.67	11435	53.33
23019	2692	60.00	17684	53.33	28854	53.33

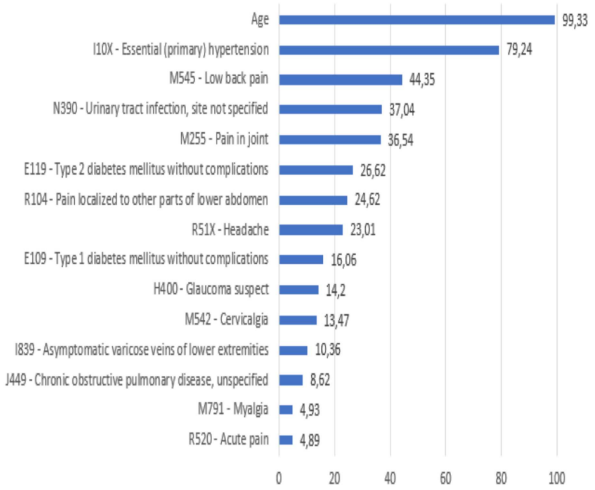

FIGURE 14. Average relevance of the features used to represent the explanatory cases.

1NN, 28.88% for 2NN and 28.38 for 3NN. We assume that by increasing the number of retrieved examples these values will decrease. However, a similar percentage of common features in the three most similar cases implies a homogeneity in the case base and the ability of the CBR system to obtain representative examples. It confirms that the reduction of the number of cases from the 40.000 available instances to the randomly selected 12.000 cases has no impact in the performance of the explanation system as the CBR method is able to find several explanatory examples with an equivalent quality. Additionally, although a percentage around 30% may seem a low value it has to be considered as a positive result if we take into account that every case is described by 15 variables and, therefore, the explanation system can provide explanations by using up to 4 or 5 features.

Finally, we have analyzed the relevance of each variable by computing the percentage of cases that have that concrete variable in common with the most similar explanatory case. For example, the probability that a case and its most similar explanatory case have the same age (within the 10 years threshold) is 93.33%.

This way, the five diseases that are, on average, used to explain the outcome of the neural network are: age (93.33%), essential hypertension (79.24%), low back pain (44.35%), urinary tract infection (37.04%), and pain in joint (36.54%). Figure 14 shows the corresponding value for each feature. We can observe that the relevance of variables decreases smoothly, being 15 a good cut-off value when choosing the number of features to describe an explanatory case.

Once we have evaluated the performance of our explainable prediction system, next section presents a discussion about the impact of the achieved results.

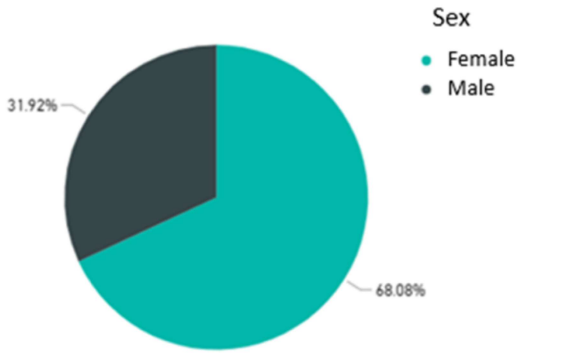

FIGURE 15. Distribution of people at risk of developing CKD by sex.

IV. DISCUSSION

Once the prognosis was made for the group of 39,277,086 people, the model identified that 3,494,516 people are at risk of developing chronic renal disease in Colombia. This figure corresponds to 7% of the total Colombian population, estimated at 49,834,240 by 2018 [59]. Figure 15 shows the gender distribution of people who were predicted to be at risk of developing CKD. In this graph, it can be identified that 68.08% correspond to women, and the remaining 31.92% correspond to men.

Figure 16 shows the distribution of these people by age. Small isolated groups can be observed in ages less than 25 years, a smooth growth between 25 and 45 years and a significant increase from 45 years, whose peak is centered in the 60 years. From this age the number of people begins to decrease.

Finally, Figure 17 presents the distribution of the percentage of persons according to their place of residence, in this case at departmental level. The percentage is obtained taking as numerator the number of persons identified in each department, and as denominator the total population of the department. The departments with the highest percentage of people with CKD are Bogotá D.C., Antioquia, Risaralda, Caldas, Santander, Atlántico, Boyacá and Quindío.

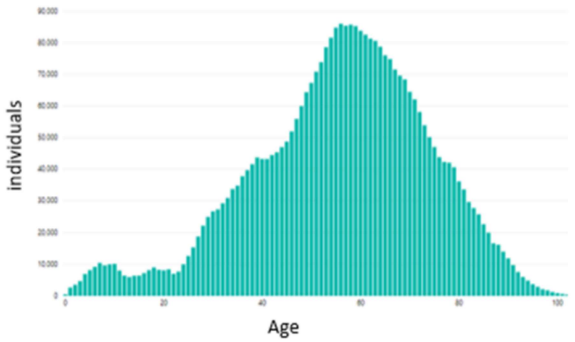

FIGURE 16. Distribution of people at risk of developing CKD by age.

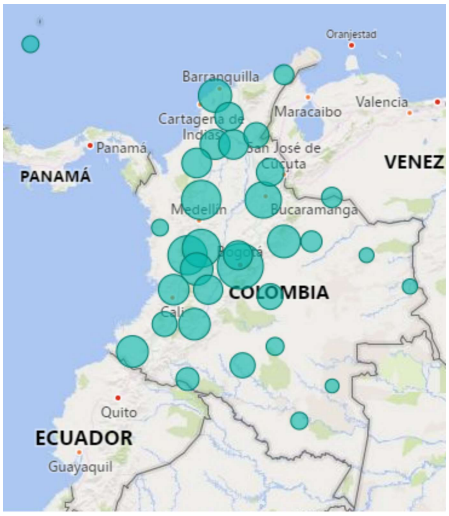

FIGURE 17. Map of coverage of the population at risk of developing CKD.

V. CONCLUSION

All previous studies for CKD prediction use laboratory test information as input variables for model training and a relatively small pool of patients is available. However, in this work we have collected a training data set corresponding to 40,000 people, and for the implementation of the model the information from health care performed on 39,277,086 people during the years 2009 to 2018 was used. These include demographic information, such as sex, age, ethnicity and place of residence, as well as the history of the pathologies that have been diagnosed to the person, coded according to the International Classification of Diseases (ICD). From this novel dataset, we have developed a neural network approach that can predict the risk of developing chronic kidney disease with an accuracy of 95%. This is a remarkable result as the most similar study achieves an accuracy of 89.7% [31],. This result was obtained by training a neuronal network with 5 layers: an input layer with 7,492 neurons, corresponding to the variables or characteristics of the model, 3 hidden layers with 500, 100 and 50 neurons respectively, and an output layer with a single neuron representing the class of the binary classification problem. For network training, a ReLU activation function was used in the hidden layers and a sigmoid function to find the probability in the output layer. Adam was used as the training algorithm of the model, which showed a higher convergence speed compared to other more traditional algorithms such as gradient descent and stochastic gradient descent (SGD). To combat the effect of overfitting in the network, regularization was used using the dropout technique, in conjunction with early stopping.

In addition to the neural network, other machine learning models such as random forest and vector support machines (SVM) were trained and accuracy values of 92% and 61% respectively were obtained. The random forest model obtains good results, very close to those of the neural network and is considered as an alternative to the latter model. The performance of the SVM model for this case is considered very low and its implementation is discarded to predict new cases of CKD.

The neural network classifier demonstrated its ability to learn how to identify disease risk factors and then apply this knowledge to the prognosis of new cases. The number of people predicted by the model to be at risk of developing chronic renal disease in Colombia is 3,494,516, or 7% of the total population. Based on the person's unique identifier, it is possible to monitor the person and implement preventive activities that minimize the risk of developing CKD.

However, despite the demonstrated efficiency of neural networks to predict CKD, this machine-learning paradigm is opaque to the expert regarding the explanation of the outcome. This transparency is necessary in order to increase the expert's trust in the provided diagnosis and the acceptability of the system. This way, current research on eXplainable AI focus on promoting transparency in this type of systems. One of the most successful approaches to this challenge is the use of twin systems, where a black-box machine-learning method is complemented by another white-box method that provides explanations about the predicted values. Here, case-based reasoning has proved to be an ideal complement as this paradigm is able to find explanatory cases for an explanation-by-example justification of a neural network's prediction. In this paper, we have applied and validated a NN-CBR twin for the explanation of CKD predictions. Firstly, we select the most relevant features for the CKD diagnosis and create a case-base of explanatory examples. These examples are later used to generate text explanations that justify the outcomes of the neural network.

As future work, we would like to explore further ways to combine the explanatory cases and generate the explanations. In this work, we have binarized the features to compare the most similar explanatory cases to the given patient description. However, we could apply more advanced and knowledge-rich strategies that consider the magnitude of the associated values. For example, generating explanations such as "The diagnosis is *positive* because your hypertension is high and there are several similar CKD cases where hypertension is moderate or high". Another possible future line of work is the use of the latent features of the neural networks to select the most significant variables that represent the cases. Complementary, as the choice of 12.000 cases for the case base was made based on preliminary experimentations, it should be possible to analyze the behavior of the system when decreasing that value to reduce the size of the case base and analyze its impact in the performance of the CBR system.

Finally, we would like to conduct a user based evaluation to validate the impact of the generated explanations in the acceptance by the expert of the predictions given by the neural network.

ACKNOWLEDGMENT

This work was supported by the Spanish Committee on Economy and Competitiveness (TIN2017-87330-R) and SECTEI (Subsecretaría de Ciencia, Tecnología e Innovación de la Ciudad de México).

REFERENCES

[1] W. Yu, T. Liu, R. Valdez, M. Gwinn and M. J. Khoury, "Application of support vector Machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes.," *BMC medical informatics and decision making,* vol. 10, no. 1, p. 16, 2010.

[2] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in genetics,* vol. 9, 2018.

[3] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin and H. Benali, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," *Neuroradiology,* vol. 51, no. 2, pp. 73-83, 2009.

[4] I. S. F. Dessai, "Intelligent heart disease prediction system using probabilistic neural network.," *International Journal on Advanced Computer Theory and Engineering (IJACTE),* vol. 2, no. 3, pp. 2319-2526, 2013.

[5] Z. H. Zhou, Y. Jiang, Y. B. Yang and S. F. Chen, "Lung cancer cell identification based on artificial neural network ensembles," *Artificial Intelligence in Medicine,* vol. 24, no. 1, pp. 25-36, 2002.

[6] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. W. M. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol and others, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama,* vol. 318, pp. 2199-2210, 2017.

[7] Y. Cao, Z. D. Hu, X. F. Liu, A. M. Deng and C. J. Hu, "An MLP classifier for prediction of HBV-induced liver cirrhosis using routinely available clinical parameters," *Disease markers,* vol. 35, no. 6, pp. 653-660, 2013.

[8] W. S. McCulloch and W. Pitts, "A logical calculus of ideas immanent in nervous," *Bulletin of Mathematical Biophysics,* vol. 5, p. 115–133, 1943.

[9] K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks,* vol. 2, pp. 359-366, 1989.

[10] G. E. Hinton, S. Osindero and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation,* vol. 18, p. 1527–1554, 2006.

[11] A. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Developmen,* vol. 3, no. 3, pp. 210-29, 1959.

[12] K. Tretyakov, "Machine learning techniques in spam filtering," *Data Mining Problem-oriented Seminar, MTAT,* vol. 3, no. 177, pp. 60-79, 2004.

[13] D. West and V. West, "Improving diagnostic accuracy using a hierarchical neural network to model decision subtasks," *International journal of medical informatics,* vol. 57, no. 1, pp. 41-55, 2000.

[14] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology,* vol. 284, pp. 574-582, 2017.

[15] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan and U. R. Acharya, "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Computing and Applications,* pp. 1-7, 2018.

[16] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan and others, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell,* vol. 172, pp. 1122-1131, 2018.

[17] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger and G. Thoma, "Image analysis and machine learning for detecting malaria," *Translational Research,* vol. 194, pp. 36-55, 2018.

[18] Fondo Colombiano de Enfermedades de Alto Costo, Situación de la enfermedad renal crónica, la hipertensión arterial y la diabetes mellitus en Colombia, Bogotá, 2017.

[19] K. Kumar and B. Abhishek, "Artificial neural networks for diagnosis of kidney stones disease," *Information Technology and Computer Science,* vol. 1, pp. 41-48, 2009.

[20] Y. Kazemi and S. A. Mirroshandel, "A novel method for predicting kidney stone type using ensemble learning," *Artificial intelligence in medicine,* vol. 84, pp. 117-126, 2018.

[21] H. Zhang, C. L. Hung, W. C. C. Chu, P. F. Chiu and C. Y. Tang, "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks," *2018 IEEE International Conference on Bioinformatics and Biomedicine,* pp. 1351-1356, 2018.

[22] E. H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms," *Informatics in Medicine Unlocked,* p. 100178, 2019.

[23] J. Xiao, R. Ding, X. Xu, H. Guan, X. Feng, T. Sun and Z. Ye, "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression," *Journal of translational medicine,* vol. 17, no. 1, p. 119, 2019.

[24] A. Soldevila, Análisis de la progresión de la enfermedad renal crónica avanzada mediante técnicas de aprendizaje máquina, Valencia, 2017.

[25] S. Hore, S. Chatterjee, R. K. Shaw, N. Dey and J. Virmani, "Detection of chronic kidney disease: A NN-GA-based approach," *Nature Inspired Computing,* pp. 109-115, 2018.

[26] A. Al Imran, M. N. Amin and F. T. Johora, "Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide & Deep Learning," *2018 International Conference on Innovation in Engineering and Technology,* pp. 1-6, 2018.

[27] S. Chatterjee, S. Banerjee, P. Basu, M. Debnath and S. Sen, "Cuckoo search coupled artificial neural network in detection of chronic kidney disease," in *2017 1st International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech)*, 2017.

[28] A. Subasi, E. Alickovic and J. Kevric, "Diagnosis of chronic kidney disease by using random forest," in *CMBEBIH 2017*, Springer, 2017, pp. 589-594.

[29] H. Polat, H. D. Mehr and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," *Journal of medical systems,* vol. 41, p. 55, 2017.

[30] H. Kriplani, B. Patel and S. Roy, "Prediction of Chronic Kidney Diseases Using Deep Artificial Neural Network Technique," *Computer Aided Intervention and Diagnostics in Clinical and Medical Images,* pp. 179-187, 2019.

[31] Y. Ren, H. Fei, X. Liang, D. Ji and M. Cheng, "A hybrid neural network model for predicting kidney disease in hypertension patients based on electronic health records," *BMC Medical Informatics and Decision Making,* vol. 19, no. 2, p. 51, 2019.

[32] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI communications,* vol. 7, pp. 39-59, 1994.

[33] M. T. Keane and E. M. Kenny, "How case-based reasoning explains neural networks: A theoretical analysis of XAI using

post-hoc explanation-by-example from a survey of ANN-CBR twin-systems," in *International Conference on Case-Based Reasoning*, 2019.

[34] Ministerio de Salud y Protección Social, "Resolución 3374 de 2000," 13 mayo 2019. [Online]. Available: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/Resoluci%C3%B3n_3374_de_2000.pdf.

[35] World Health Organization, "International Classification of Diseases," 23 01 2019. [Online]. Available: https://www.who.int/classifications/icd. [Accessed 12 09 2019].

[36] M. Nielsen, Neural Networks and Deep Learning, San Francisco, CA, USA: Determination press, 2015.

[37] D. Rumelhart, G. Hinton and R. Williams, "Learning representations by back-propagating errors.," *Nature,* vol. 323, p. 533–536, 1986.

[38] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.

[39] A. Gulli and S. Pal, Deep Learning with Keras, Packt Publishing Ltd, 2017.

[40] S. Abrahams, D. Hafner, E. Erwitt and A. Scarpinelli, TensorFlow for Machine Intelligence: A Hands-on Introduction to Learning Algorithms, Bleeding Edge Press, 2016.

[41] K. Jarrett, K. Kavukcuoglu and Y. LeCun, "What is the best multi-stage architecture for object recognition?," *2009 IEEE 12th international conference on computer vision,* pp. 2146-2153, 2009.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[43] G. Hinton, "Neural networks for machine learning," *Coursera, video lectures,* 2012.

[44] N. Srivastava, "Improving neural networks with dropout," *Universidad de Toronto,* 2013.

[45] C. M. Bishop, "Regularization and complexity control in feed-forward networks," *Proceedings International Conference on Artificial Neural Networks,* p. 141–148, 1995.

[46] Association for the Advancement of Artificial Intelligence, "Case-Based Reasoning (CBR) - Using Similar Situations from the Past to Solve a Present Problem," 2011. [Online]. Available: http://aaai.org/AITopics/CaseBasedReasoning. [Accessed 06 02 2012].

[47] D. B. Leake, Case-Based Reasoning: Experiences, lessons and future directions, MIT press, 1996.

[48] E. Fix and J. Hodges, "An important contribution to nonparametric discriminant analysis and density estimation," *International Statistical Review,* vol. 3, pp. 233-238, 1951.

[49] L. Gates, C. Kisby and D. Leake, "CBR Confidence as a Basis for Confidence in Black Box Systems," in *International Conference on Case-Based Reasoning*, 2019.

[50] M. Caro-Martinez, J. A. Recio-Garcia and G. Jimenez-Diaz, "An Algorithm Independent Case-Based Explanation Approach for Recommender Systems Using Interaction Graphs," in *International Conference on Case-Based Reasoning*, 2019.

[51] B. Díaz-Agudo, E. Plaza, J. A. Recio-García and J.-L. Arcos, "Noticeably new: Case reuse in originality-driven tasks," in *European Conference on Case-Based Reasoning*, 2008.

[52] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters,* vol. 27, no. 8, p. 861–874, 2006.

[53] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern recognition,* vol. 30, pp. 1145-1159, 1997.

[54] V. Vapnik, Estimation of Dependences Based on Empirical Data, New York: Springer-Verlag, 1982.

[55] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995.

[56] G. James, D. Witten, T. Hastie and R. Tibshirani, An Introduction to Statistical Learning with Applications in R, New York: Springer, 2013.

[57] L. Breiman, "Random forests," *Machine Learning,* vol. 45, pp. 5-32, 2001.

[58] L. Breiman, "Bagging predictors," *Machine Learning,* vol. 24, no. 2, pp. 123-140, 1994.

[59] DANE, "Proyecciones de población," 11 mayo 2019. [Online]. Available: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion.

GABRIEL R. VÁSQUEZ-MORALES was born in Colombia. He received the B.S. degree in systems engineering from Universidad de Cundinamarca, Colombia in 2006 and the specialization degree in software engineering from Universidad Incca de Colombia in 2011. He obtained the M.S. degree in artificial intelligence at Universidad Internacional de La Rioja (UNIR) in Spain. Since 2012 is data scientist in the Technology Office and Department of Epidemiology and Demography of the Ministry of Health and Social Protection, Bogotá, Colombia. His research interest includes data science, artificial intelligence, artificial vision, machine learning and robotics applied to medicine and biology.

SERGIO M. MARTÍNEZ-MONTERRUBIO was born in Mexico City. He received a Computer Science degree in 1998 and a master's in international business administration in 2003, both at the Universidad Nacional Autonoma de Mexico, UNAM. Sergio received a Ph.D. in Computer Science in 2016 from the Instituto Tecnologico y de Estudios Superiores de Monterrey, ITESM (*Mexico*). Since 2017 is an investigator making a postdoctoral research in the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (*UCM*) in GAIA *(Group of Artificial Intelligence Applications)* located in the Computer Sciences Faculty and Engineering at the UCM Campus. His professional experience includes working in companies as McAfee antivirus, Oracle, Entrust Technologies, Colgate Palmolive, ABC Medical Center and Continental Automotive Systems. Mr. Martinez's awards and honors include the Mexico´s first national prize for his thesis in computer science by ANFECA in 1999, doctoral scholarship with Conacyt and postdoctoral scholarship with SECTEI. For his doctoral studies in the ITESM receives the doctorate medal. His research interests are artificial intelligence, data mining, big data, computer science applied to medicine, machine learning and its applications in cyber security.

PABLO MORENO-GER was born in Madrid, Spain, in 1981 and got his computer engineering degree from Universidad Complutense de Madrid in 2004. He received his PhD. in computer science in 2007 in the Department of Software Engineering and Artificial Intelligence at Universidad Complutense de Madrid. He is the Director of the School of Engineering and Technology at UNIR. Previously he worked at Universidad Complutense de Madrid, where he was a member of the e-UCM research group and Vicedean for Innovation at the School of Computer Engineering. He has a strong research record in Technology-Enhanced Learning, AI, Machine Learning and Learning Analytics, and has published over 150 academic articles in these topics. He currently holds the IBM Chair on Data Science in Education (http://research.unir.net/ibmchair/) and is a member of the IAR Research Group in AI and Robotics (http://gruposinvestigacion.unir.net/ia/).

JUAN A. RECIO-GARCIA was born in Guadalajara, Spain, in 1980 and got his computer engineering degree from Universidad Complutense de Madrid in 2003. Currently he is Head of Department of Software Engineering and Artificial Intelligence at Universidad Complutense of Madrid, where he obtained a PhD in Computer Science in 2008.

His research has focused on the confluence of Software Engineering and Case-Based Reasoning (CBR), being the author of the COLIBRI platform for the development of CBR systems within the Group of Applications of Artificial Intelligence (GAIA) of the University Complutense of Madrid. This platform is a reference in the CBR community with more than 30.000 downloads. He also leads a research project founded by the Spanish Committee on Economy and Competitiveness on eXplainable AI by applying Case-Based Reasoning as the explanation paradigm. Previously, he leaded another project on the personalization and integration of social interactions in context-aware group recommender systems. He has also a strong connection with companies for the application of AI solutions to industry projects.