ISSN 1870-4069

Elderly Mortality from COVID-19 in Mexico City: A Computational Intelligence Approach Based on Random Forests

Sinuhe Mazuti Osorio-Rivero, Guillermo Molero-Castillo,
Everardo Bárcenas, Rocío Aldeco-Pérez

Universidad Nacional Autónoma de México,
Facultad de Ingeniería,
Mexico

mazuti.96@gmail.com, guillermo.molero@ingenieria.unam.edu,
{ebarcenas, raldeco}@unam.mx

Abstract. Computational intelligence encompasses a wide variety of techniques and algorithms that are applied to address complex real-world problems. It is in the field of health where its use becomes significant to understand the behavior of a given disease, such as COVID-19. In this sense, there are sectors of the population that can easily develop a complication or die from such a condition, such as people over 60 years of age. As this is a growing and vulnerable population, it is important to make efforts to analyze the risks and effects that the elderly may present. This paper presents the implementation of a computational intelligence method for the prediction of mortality in older adults infected with SARS-CoV-2 in Mexico City. Open data, published and distributed by the Government of Mexico City, were used for this analysis. The results show that the variables with the greatest contribution of information for classification were: intubated, patient care, pneumonia and intensive care unit (ICU). This concludes that a hospitalized patient, who is admitted to the intensive care unit and requires intubation, has a high probability of being classified as 'dead'. In addition, the results show that variables related to the patient's age and sex are more important than variables associated with comorbidities.

Keywords: COVID-19, computational intelligence, elderly, random forests.

1 Introduction

Today, Artificial Intelligence (AI) encompasses a wide variety of subfields, ranging from general purpose areas, such as learning and perception [1], to more specific ones, such as applications based on computational intelligence, machine learning, deep learning, reinforcement, or mixed [2].

Specifically, Computational Intelligence (CI) concentrates a wide variety of techniques and algorithms that are applied to mimic human reasoning power in order to cope with complex real-world problems [3].

Today, it is evident the momentum that CI has taken in its application in different fields of human activity, such as health, security, education, biology, among others. Undoubtedly, at present, it is in the field of health where its use becomes significant to

pp. 203–216; rec. 2023-04-17; acc. 2023-05-01

understand the behavior of certain diseases, such as COVID-19, which is currently considered a pandemic affecting humanity.

The COVID-19 pandemic was caused by a new type of coronavirus, known as SARS-CoV-2. The first cases of infected people date back to December 2019 in the city of Wuhan, China.

Thus, from the first infections until October 2022, there are more than 615 million confirmed cases and more than 6.5 million deaths worldwide [4]. In the specific case of Mexico City, the object of study in this research, more than 1.74 million confirmed cases and more than 57 thousand deaths have been reported [5].

In this sense, there are sectors of the population that can easily develop a COVID-19 complication and even die. These sectors of the population are called vulnerable or at-risk groups. Among them are people 60 years of age or older, considered to be the elderly.

According to the Government of Mexico City, this vulnerable group of older adults is classified into two categories [6]: i) with comorbidity, which is characterized as people over 60 years of age, who have one or more diseases considered as factors of vulnerability; and ii) without comorbidity, which are identified as older adults without any disease or disorder that is considered a vulnerability.

There are certain characteristics, diseases and conditions in older adults that considerably affect their health status. For these reasons, they are considered one of the groups with the greatest vulnerability to COVID-19 disease [7]. This adult population can easily develop complications and even die from the disease.

Therefore, it is important to identify the patterns that condition their health status. The purpose of this is to provide useful information to understand and make better decisions about the management of pandemic disease. In addition, to provide a reflective analysis of the vulnerable group mentioned.

In addition to the above, the infection of older adults with the SARS-CoV-2 virus has a direct influence on society due to the way in which the epidemic risk traffic light COVID-19 operates, established by the Government of Mexico City through which the level of population risk and the increase or decrease of local activity is announced through colors, as well as the appropriate health safety measures for the reopening of work and educational activities and the use of public spaces [8].

On the other side, it is important to highlight the increase in the population of older adults in the last decade, where it went from 9.1% in 2010 to 12.0% in 2020. While the young population aged 0 to 17 years decreased from 35.4% in 2010 to 30.4% in 2020 [9]. This means that the population of older adults in Mexico is increasing.

For this reason, being a growing and vulnerable population, it is important to make efforts, from different perspectives, as is the case of computational intelligence, to analyze the risks and affectations that the elderly may present. This type of analysis is useful for identifying patterns in the form of trends in the population under analysis.

The aim of this research work was to implement a computational intelligence method, specifically random forests, for the classification of mortality in older adults infected with SARS-CoV-2 in Mexico City. For this, open data was used, published by the Government of Mexico City.

This paper is organized as follows: Section 2 presents the background of artificial and computational intelligence, COVID-19 in the adult population and the main related works; Section 3 describes the method established as a proposed solution; Section 4 presents the results obtained, based on a use case, such as the adult population; and Section 5 summarizes the main conclusions and future work.

2 Background

Artificial intelligence as an area of knowledge, proposed by John McCarthy in 1956, which refers to the science and engineering for the construction of intelligent machines, has faced multiple challenges during the last decades, due to the transition of states with emerging technologies, methods and algorithms [10, 11].

This makes traditional artificial intelligence incompatible with the increasing demands in search, optimization and resolution that problems require. The path from traditional to modern has enabled the emergence of better computational tools such as computational intelligence [11].

Through computational intelligence it is possible to build models, reasoning, machines and processes, based on structured and intelligent behaviors [11]. This type of intelligence adopts methods that tolerate incomplete, imprecise and uncertain knowledge in complex environments. In this way, they allow approximate, flexible, robust and efficient solutions [12]. Therefore, computational intelligence can be implemented to address problems that affect today's society [10].

Undoubtedly, to build inductive learning models, which base their function on the discovery of patterns from examples, one of the most used algorithms in computational intelligence are decision trees (DTs), through which prognosis and classification problems can be solved, aiming to build a hierarchical, efficient and scalable structure based on the conditions (variables) established in the data. The divide and conquer strategy are used for this purpose.

2.1 Random Forests

A tree is graphically represented by a set of nodes, leaves and branches. The main node or root is the attribute (variable) from which the classification process starts. The internal nodes correspond to each of the attribute conditions associated with a given problem. While each possible answer to the conditions is represented by a child node.

The branches coming out of each of these nodes are labeled with the possible values of the attribute. The final nodes or leaf nodes correspond to a decision, which coincides with some class (label) of the variable to be classified [13].

It is important to mention that sometimes decision trees are susceptible to overfitting, which means that they tend to learn very well from the training data, but their generalization may not be as good. One way to improve the generalization of decision trees is to combine several trees, known as random forests (RFs).

Random forests are widely used today. They aim to build an ensemble of decision trees, which when put together, what is actually happening is that they see different portions of the data. No tree uses all the training data, but each one is trained with different samples for the same problem.

By combining the results, the errors are compared with each other, and it has a prediction (forecast or classification) that generalizes better to the problem. Figure 1

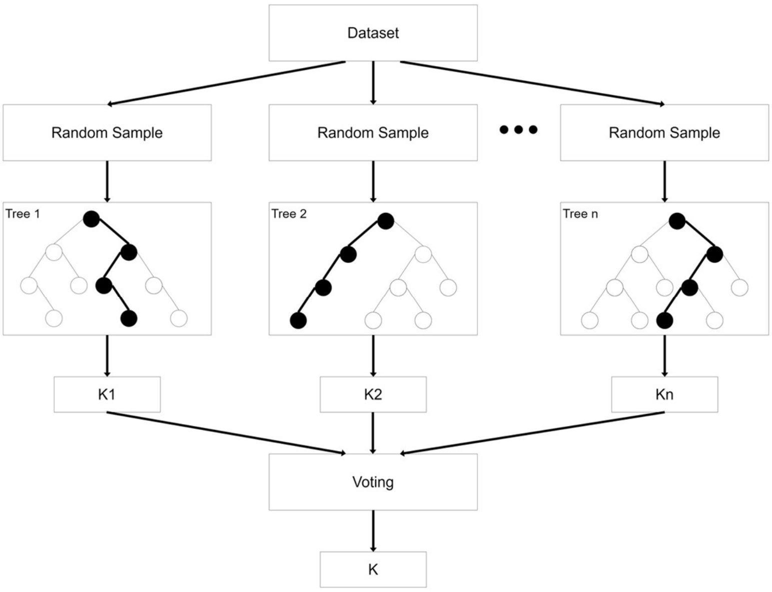


Fig. 1. Random forest general scheme.

shows the general scheme of the operation of random forests for classification, which consists of four steps:

1. Selection of random samples from the data set.
2. Construction of a decision tree for each sample and its respective result.
3. Voting (classification) based on the results obtained.
4. Selection of the result with the most votes (ranking).

2.2 Related Work

At present, one of the significant applications of random forests, due to the COVID-19 pandemic, is the classification of mortality in patients infected by the SARS-CoV-2 virus.

The objective is to classify characteristics (variables) of patients at risk of mortality due to this disease [14], as is the case of vulnerable groups, for example, the elderly.

In [15] it is stated that older people are more likely to contract COVID-19 and develop complications. These same authors mention that in the United States, through the Center for Disease Control and Prevalence (CDC) [16], it was identified that people over 65 years of age, accounted for 31% of SARS-CoV-2 infections, 45% of hospitalizations, 53% of admissions to intensive care units, and 80% of deaths caused by this infection.

Table 1. Related work.

Author	Description	Algorithm used	Limitations
Rami *et al.* (2022) [17]	Three experiments were performed using a data set of patients with COVID-19. Seven classification models were tested. The best performance was with the Bagging algorithm, with an accuracy of 83.55%.	Bagging, J48, Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and Threshold Selector.	Records from 582 patients were used, of which 15 features were used for the first experiment, 6 for the second and 11 for the third.
Alves *et al.* (2021) [18]	Cases of Italian older adults hospitalized for COVID-19 were analyzed. Subsequently, the comorbidities of each group were analyzed. Dementia, diabetes, chronic kidney disease and high blood pressure were the main diseases involved in mortality.	Statistical analysis (Stata software).	The number of registered cases used ranged from 18 to 1591 patients.
Khan *et al.* (2021) [14]	The mortality rate of patients with COVID-19 was analyzed. Sociodemographic and clinical data from patients from different countries were used, and the models were evaluated for accuracy, precision, sensitivity and specificity. Deep Neural Networks model achieved a better prediction with 97% accuracy.	Deep Neural Network (DNN), Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbor (KNN).	It was used 103888 patient records from 45 countries, with the largest number from India (98632), and Philippines (4493). Nevertheless, there were less than 200 records for remaining countries
Cardoso *et al.* (2021) [19]	Algorithms were used to predict COVID-19 positive cases and find patterns in the databases of six districts (municipalities) in Argentina.	Fuzzy relationships and Artificial Neural Networks.	The artificial neural network model obtained an average error of 20%.
Akinnuwesi *et al.* (2021) [20]	Computational intelligence methods for the diagnosis of people with COVID-19 were analyzed. The performance of each algorithm was measured in terms of accuracy, precision, recall, balanced and accuracy. The methods with the best performance were MLP, FCM and DNN.	Logistic Regression (LR), Support Vector Machine (SVM), Naïve Byes (NB), Multilayer Perceptron (MLP), Fuzzy Cognitive Map (FCM) and Deep Neural Network (DNN).	Dataset limited to 600 records, of which 80% were used in training and 20% in testing.

Today, some researches have been identified that have provided knowledge about the COVID-19 disease by means of implementations of computational intelligence algorithms. These works have different approaches and objects of study. Table 1 summarizes five of these works, where the work performed, algorithms used, and limitations identified are briefly described.

Several related works use computational intelligence algorithms to address problems arising from the COVID-19 pandemic. It is important to highlight the use of the methods and tools provided by computational intelligence to understand the risks and affectations that certain vulnerable groups may suffer, as is the case of the elderly.

In relation to the related works identified, random forests were used as a classification algorithm for this research, with the purpose of taking advantage of the benefits and advantages of this computational intelligence approach, based on the divide and conquer strategy, with which explanatory rules are extracted, an advantage that other algorithms do not have by providing solutions without justification or explanation [10]. In addition, the random forest algorithm has shown high accuracy for classifying records of people with COVID-19 [17].

On the other hand, most of the identified related works perform the investigations with general population information sources. In contrast, in this research, the study focuses on the vulnerable group of older adults in Mexico City. This allows to learn about factors that condition the health status of this population, which is currently increasing and has a higher mortality rate due to COVID-19 [21].

In addition to the above, a data source containing 591352 records was used, corresponding to real cases captured during two years of the population under study. While the related works identified carried out their research with data sources of smaller period and size.

2.3 Motivation

Random forests are ideal for working with a large amount of data and multiple variables, due to the fact that it selects random samples to train classification or prognostic (regression) models, as the case may be [22].

In this sense, it is important to analyze the group of older adults, because it is one of the vulnerable groups that have been severely affected by COVID-19 disease. Increasing age conditions, a decrease in immune response and regenerative capacities, as well as a decrease in body mass index, functionality and an increase in comorbidities.

Given these situations, there is evidence of an increased risk of hospitalization and mortality compared to the general population [23]. Therefore, through this research, specialized technology is used to analyze the vulnerable group of older adults in Mexico City affected by COVID-19 disease.

3 Method

The solution method for the analysis of mortality of elderly in Mexico City, as a result of COVID-19, was divided into four stages (Figure 2): i) acquisition of data source, ii) variables selection, iii) classification, and iv) validation.

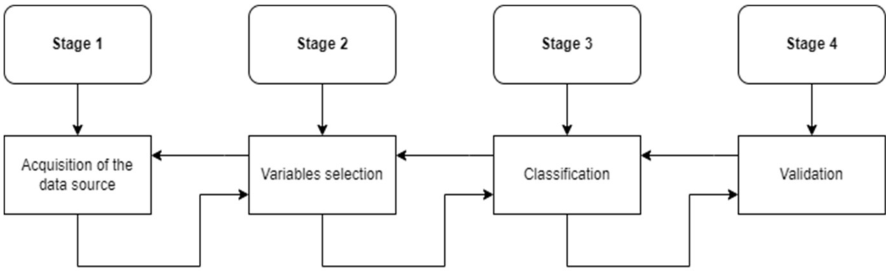


Fig. 2. Scheme of the method used as a proposed solution.

3.1 Data Source

In accordance with the decree published in the Official Journal of the Federation on February 20, 2015, which establishes the regulation on Open Data, the General Directorate of Epidemiology, made available to the general population the historical bases published since April 14, 2020 on cases associated with COVID-19 [24].

Thus, at present, there are sources of data on COVID-19 produced by different state, national and international entities, covering different populations. For example, for Mexico City alone, 24 data sources related to COVID-19 were found, such as historical hospital capacity, preliminary affluence in public transportation, inventories of contingency measures, among others.

In particular, for this research, data from the General Directorate of Epidemiology of the Ministry of Health of Mexico [24] were used, which are published periodically to facilitate all users access, use, reuse and redistribution of the same.

The purpose is to monitor possible cases of COVID-19 at the federal level, and specifically in Mexico City. Therefore, the period analyzed in this research comprises from April 14, 2020 to April 14, 2022, which represents 591352 real cases of COVID-19 in older adults in Mexico City, that is, records of two consecutive years.

3.2 Variables Selection

The original data source contains 40 variables, which provide information about the patient's case. However, not all the variables provide significant information for this research. So, an exploratory data analysis (EDA) was performed in order to make a careful selection of these variables.

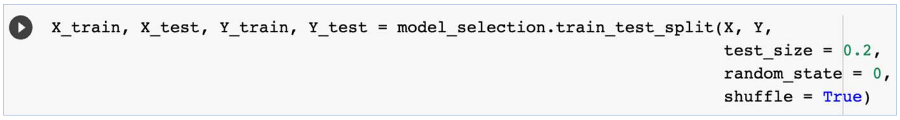
Thus, from a selection of significant variables from the medical point of view and data analysis, a data source consisting of 20 variables was obtained, which are listed in Table 2. All the selected variables contain relevant information about the patient, characteristics of the medical treatment received and the evolution of the disease.

Through which patterns can be identified that allow an accurate classification of the mortality of older adults infected with SARS-CoV-2.

While other variables were discarded because they included redundant or irrelevant information, as is the case of sample taking (SAMPLING), result of the laboratory test applied to the patient to confirm the disease (LAB_RESULT), SARS-CoV-2 antigen sample (ANTINGEN_SAMPLING), to mention a few of these.

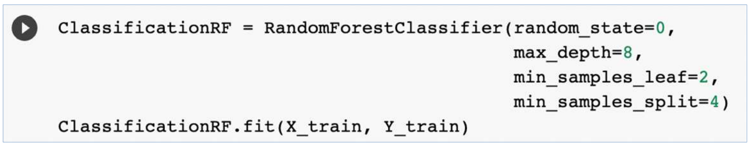
Table 2. Selected variables.

Item	Name	Description	Values
1	SEX	Identifies the sex of the patient.	1-Female, 2-Male, 99-Not Specified
2	PATIENT_CARE	Identifies the care type that patient received.	1-Ambulatory, 2-Hospitalized, 99-Not Specified
3	STATE	Identifies the situation (alive or dead) of the patient.	1-Alive 2-Dead
4	INTUBATED	Identifies whether the patient required intubation.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
5	PNEUMONIA	Identifies whether the patient was diagnosed with pneumonia.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
6	AGE	Identifies the patient's age.	Numeric
7	DIABETES	Identifies if the patient has a diagnosis of diabetes.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
8	COPD	Identifies if the patient has a diagnosis of Chronic Obstructive Pulmonary Disease (COPD).	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
9	ASTHMA	Identifies if the patient has a diagnosis of asthma.	1-Yes, 2-No, 97-Does not apply, 98-Ignore, 99-Not Specified
10	INMUSUPPR	Identifies if the patient has a diagnosis of immunosuppression.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
11	HYPERTENSION	Identifies if the patient has a diagnosis of hypertension	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
12	OTHER_DISEASES	Identifies whether the patient has a diagnosis of other diseases.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
13	CARDIOVASCULAR	Identifies whether the patient has a diagnosis of cardiovascular disease.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
14	OBESITY	Identifies if the patient has a diagnosis of obesity.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
15	CHRONIC_RENAL	Identifies if the patient has a diagnosis of chronic renal insufficiency.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
16	SMOKING	Identifies if the patient has a smoking habit.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
17	OTHER_CASE	Identifies if the patient had contact with any other case diagnosed with SARS-CoV-2.	1-Yes, 2-No, 97-Not Applicable, 98-Ignore, 99-Not Specified
18	ANTIGEN_RESULT	Identifies the result of the SARS-CoV-2 antigen sample analysis.	1-Positive SARS-CoV-2, 2-Negative SARS-CoV-2, 97- Not Applicable (Case without sample).
19	FINAL CLASSIFICATION	Identifies the classification of the Covid-19 test result: confirmed, invalid, not performed, suspect, and negative.	1-Confirmed by the Clinical Epidemiological Association, 2-Confirmed by the Ruling Committee, 3-Confirmed case, 4-Invalid by laboratory, 5-Not laboratory performed, 6- Suspect case, 7-Negative to SARS-COV-2
20	ICU	Identifies whether the patient required admission to an Intensive Care Unit (ICU).	1-Yes, 2-No, 97-Not applicable, 98-Unknown, 99-Not Specified



```
X_train, X_test, Y_train, Y_test = model_selection.train_test_split(X, Y,
                                                                    test_size = 0.2,
                                                                    random_state = 0,
                                                                    shuffle = True)
```

Fig. 3. Separation of the data vectors for the training and testing of the algorithm.



```
ClassificationRF = RandomForestClassifier(random_state=0,
                                          max_depth=8,
                                          min_samples_leaf=2,
                                          min_samples_split=4)
ClassificationRF.fit(X_train, Y_train)
```

Fig. 4. Configuration of parameters for the operation of the algorithm.

Which are redundant due to the existence of another variable that indicates the final result of the antigen test for SARS-CoV-2 (ANTIGEN_RESULT).

3.3 Classification

The first criterion for classification using random forests is to calculate the entropy for all classes and attributes. Entropy is a measure of uncertainty (information) that is represented as follows:

$$Entropy(S) = l(S) = lnf(S) = \sum_{i=1}^{n} -p_i log_2 p_i, \qquad (1)$$

where: S is a collection of elements (objects), and p_i is the probability of possible values. Subsequently, the best attribute is selected based on the information gain of each variable, which is represented as:

$$(S, A) = Entropy(S) - \sum_{v \in V(A)} \frac{|Sv|}{|S|} Entropy(Sv), \qquad (2)$$

where: S is a collection of elements, A are the variables, Sv is a subset of elements, and $V(A)$ is the set of values that A can take.

Based on the above, for the classification of COVID-19 mortality in older adults in Mexico City, the cases were selected based on the following conditions: a) that the age was greater than 59 years, this because in Mexico City people from that age are considered older adults; b) that the medical unit and residence of the patient were Mexico City; and c) there is a class variable that allows identifying the state of life or death of people infected with SARS-CoV-2, that is, 'Alive' or 'Dead' cases, respectively.

Once the preparation and selection of variables was completed, a structure made up of 19 independent variables and one class variable (STATE), described in Table 2, was established as an input matrix for the operation of the algorithm. Subsequently, for the classification and validation process, the input matrix was divided into training and test data vectors, as shown in the code segment written in Python (Figure 3).

Finally, the maximum depth parameter that random forest estimators can reach (maximum depth of 8 levels) has been adjusted. The criteria of the minimum number of samples required before splitting a node (at least 4 elements) and the minimum number of samples in a leaf node (at least 2 elements) were also adjusted, as shown in Figure 4.

Table 3. Classification matrix.

		Classification	
		Dead	**Alive**
Real	**Dead**	4,047	2,783
	Alive	1,889	109,552

These parameters were adjusted in order to avoid overfitting the random forest estimators. In this sense, with the established configuration and the training and test data vectors, the algorithm was applied.

3.4 Validation

For the validation of the random forest, 113599 test cases were used (20% of records, new cases, which were not used in the training process), of which 4672 were misclassified. Table 3 shows the classification matrix obtained for the case study.

The classification matrix shows information about the performance of the algorithm, that it, it allows measuring the accuracy of the results obtained [25]. The classification matrix contains four types of results, which are: true positives, true negatives, false positives and false negatives.

4 Results

The variables with the greatest gain of information were identified, which are shown in Table 4. The variable with the greatest relevance for classifying cases was INTUBATED, which refers to whether it was necessary to intubate the patient.

Another important variable was PATIENT_CARE, which describes the care received by the patient (outpatient - inpatient). The third variable was ICU, whose objective is to identify whether the patient was admitted to an intensive care unit. The fourth variable was PNEUMONIA, which to identify whether the patient was diagnosed with pneumonia.

The following variables with lower percentages refer to the classification of the COVID-19 test result (FINAL_CLASSIFICATION), the result of the antigen sample analysis (ANTIGEN_RESULT) and whether the patient had contact with any other case diagnosed with the disease (OTHER_CASE). Subsequent variables refer to patient characteristics (age and sex) and comorbidities (renal failure, diabetes, cardiovascular disease and obesity).

It was observed in one of the estimators of the random forest that the main node was the variable UCI, which corroborates an important gain of information (22.53%). Other variables at the next levels (child nodes) of the estimators were the variables INTUBATED (27.38%), FINAL_CLASSIFICATION (7.1%), PNEUMONIA (13.32%) and ANTIGEN_RESULT (2.37%). These variables belong to the group that provides more information to classify the data.

Regarding validation, it was observed that the positive cases were the records correctly assigned by the algorithm in the 'Dead' category, as can be seen in Table 3, of

Table 4. Information gain.

Variable	Importance
INTUBATED	27.38 %
PATIENT_CARE	24.60 %
ICU	22.53 %
PNEUMONIA	13.32 %
FINAL_CLASSIFICATION	7.10 %
ANTIGEN_RESULT	2.37 %
OTHER_CASE	0.77 %
AGE	0.74 %
SEX	0.24 %
CHRONIC_RENAL	0.22 %
DIABETES	0.14 %
CARDIOVASCULAR	0.10 %
OBESITY	0.09 %
OTHER_DISEASES	0.09 %
COPD	0.08 %
SMOKING	0.08 %
HYPERTENSION	0.07 %
ASTHMA	0.05 %
INMUSUPPR	0.05 %

which 4047 correct classifications were obtained, considered as true positives. On the other side, true negatives were the cases classified as 'Alive', when they really belong to that category. In this case, 109552 correct classifications were obtained, considered true negatives. The true positive and negative results represent a successful classification of the analyzed cases in the category to which they belong.

The opposite happens with the results of false negatives and false positives, which are cases misclassified by the algorithm. False negatives represent cases of type 'Dead' that were classified in the 'Alive' category, in this case 2783 false negatives were obtained. On the other side, the false positive results are those cases that belong to the 'Alive' category and were classified as 'Dead'. In the classification matrix, 1889 false positives were observed.

As a result of the application and validation of the algorithm, 96.04% accuracy and 98% precision were obtained. On the other side, the average error was 3.96%, demonstrating a remarkable classification of survival and mortality of COVID-19 cases in older adults in Mexico City.

As part of the validation, the classification was also tested through a decision tree, whereby which an outstanding accuracy of 95.3% and an average precision of 97% were obtained. Nevertheless, through random forest, as observed, a better result was obtained both in accuracy (96.04%), precision (98%) and less error of misclassified cases (3.96%). Which represents that the solution through the random forest was better. Furthermore, it significantly reduces decision tree weaknesses such as overfitting.

5 Conclusions

A hospitalized-type patient is one who was admitted to the intensive care unit, required intubation, and was diagnosed with pneumonia. This type of patient has a high probability of being classified as 'Dead', since the variables with the greatest information gain for the classification were INTUBATED, PATIENT_CARE, ICU (Intensive Care Unit), and PNEUMONIA.

Variables related to the patient's age and sex have a higher degree of importance than variables associated with comorbidities, such as: obesity, hypertension, asthma, diabetes and others. For the case of persons classified as deceased, the variable AGE was an important separating condition, where deaths were mainly between 67.5 and 76.5 years old.

Based on the results obtained, the comorbidities with the highest degree of importance in the classification were chronic kidney disease, diabetes, cardiovascular disease and obesity. On the other side, the comorbidities with the lowest degree of importance were: hypertension, asthma and immunosuppression. The variables related to chronic obstructive pulmonary disease and smoking provided a low percentage of information gain.

The random forest algorithm obtained an average accuracy of 96.04% and a precision of 98%, which means that the mortality classification of older adults infected with SARS-CoV-2 in Mexico City was remarkable, whose cases were registered in two years, that is, from April 14, 2020 to April 14, 2022.

There are different vulnerable groups that are severely affected by the SARS-CoV-2 virus; in this case, efforts and attention were focused on the elderly. However, there are other sectors of society that need to be analyzed, such as indigenous groups, migrants, people with disabilities, among others.

As future work, to enrich the results obtained, it is intended to make a new analysis with updated information and new algorithms used in computational intelligence, such as support vector machines (SVM) and deep neural networks (DNN).

This may be important and of great interest due to the behavior of the pandemic due to the disease caused by COVID-19 and its impact on the population, especially on certain vulnerable groups.

References

1. Russell, S., Norvig, P.: Inteligencia artificial: Un enfoque moderno. Pearson Education, Prentice Hall (2004)
2. Kaplan, A., Haenlein, M.: Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, vol. 62, no. 1, pp. 15–25 (2019) doi: 10.1016/j.bushor.2018.08.004
3. Kumar, G., Jain, S., Singh, U.: Stock market forecasting using computational intelligence: A survey. Archives of Computational Methods in Engineering, vol. 28, no. 3, pp. 1069–1101 (2021) doi: 10.1007/s11831-020-09413-5
4. World Health Organization: Weekly epidemiological update on COVID-19 -5 October 2022. Emergency Situational Updates (2022) https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---5-october-2022

5. Secretaría de Salud: Informe técnico semanal COVID-19 México, octubre 4, 2022. Subsecretaría de Prevención y Promoción de la Salud (2022) https://www.gob.mx/cms/uploads/attachment/file/765717/Informe_Tecnico_Semanal_COVID19_2022.10.04_1_.pdf
6. Gobierno de México: Criterios para las poblaciones en situación de vulnerabilidad que pueden desarrollar una complicación o morir por COVID-19 en la reapertura de actividades económicas en los centros de trabajo (2021)
7. Vega-Rivero, J. A., Ruvalcaba-Ledezma, J. C., Hernández-Pacheco, I., Acuña-Gurrola, M. F., López, L.: La salud de las personas adultas mayores durante la pandemia de COVID-19. Journal of Negative and No Positive Results, vol. 5, no. 7, pp. 726–739 (2020) doi: 10.19230/jonnpr.3772
8. Cortés, R., Dyer, D.: Lineamiento para la estimación de riesgos del semáforo por regiones COVID-19. Secretaria de Salud (2021)
9. INEGI: En México somos 126’014,024 habitantes: Censo de población y vivienda 2020 (2021)
10. Fulcher, J.: Computational intelligence: An introduction. In: Fulcher, J., Jain, L.C. (eds) Computational Intelligence: A Compendium. Studies in Computational Intelligence, Springer, Berlin, Heidelberg, vol. 115 (2008) doi: 10.1007/978-3-540-78293-3_1
11. Raj, J. S.: A comprehensive survey on the computational intelligence techniques and its applications. Journal of ISMAC, vol. 1. No. 3, pp. 147–159 (2019) doi: 10.36548/jismac.2019.3.002
12. Kruse, R., Borgelt, C., Braune, C., Mostaghim, S., Steinbrecher, M.: Computational Intelligence: A methodological introduction. Springer, London (2016)
13. Martínez, R., Ramírez, N., Mesa, H., Suárez, I., Trejo, M., León, P., Morales, S.: Árboles de decisión como herramienta en el diagnóstico médico. Revista Médica de la Universidad Veracruzana, vol. 9, no. 2, pp.19–24 (2009)
14. Khan, I., Aslam, N., Aljabri, M., Aljameel, S., Kamaleldin, M., Alshamrani, F., Chrouf, S.: Computational intelligence-based model for mortality rate prediction in COVID-19 patients. International Journal of Environmental Research and Public Health, vol. 18, no. 12, pp. 6429 (2021)
15. Leandro-Astorga, G., Calvo, I. B.: Infección por COVID-19 en población adulta mayor: recomendaciones para profesionales. Revista Médica de Costa Rica y Centroamérica, vol. 86, no. 629, pp. 44–50 (2021)
16. Centers for Disease Control and Prevention: Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States (2020) https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm
17. Mohammad, R. M. A., Aljabri, M., Aboulnour, M., Mirza, S., Alshobaik, A.: Classifying the mortality of people with underlying health conditions affected by COVID-19 using machine learning techniques. Applied Computational Intelligence and Soft Computing, vol. 2022, no. 3783058, pp. 1–12 (2022) doi: 10.1155/2022/3783058
18. Paulo-Alves, V., Golghetto-Casemiro, F., Gedeon de Araujo, B., de Souza-Lima, M. A., Silva de Oliveira, R., Tamires de Souza-Fernandes, F., Campos-Gomes, A. V., Gregori, D.: Factors associated with mortality among elderly people in the COVID-19 pandemic (SARS-CoV-2): A systematic review and meta-analysis. International Journal of Environmental Research and Public Health, vol. 18, no. 15 (2021) doi: 10.3390/ijerph18158008
19. Cardoso, A., Talame, L., Amor, M.: Aprendizaje automático aplicado a la pandemia del virus Covid-19 en Argentina. In: XXIII Workshop de Investigadores en Ciencias de la Computación, pp. 78–81(2021)
20. Akinnuwesi, B. A., Fashoto, S. G., Mbunge, E., Odumabo, A., Metfula, A. S., Mashwama, P. Uzoka, F. M., |Owolabi, O., Okpeku, M., Amusa, O. O.: Application of intelligence-based computational techniques for classification and early differential diagnosis of COVID-19

disease. Data Science and Management, vol. 4, pp. 10–18 (2021) doi: 10.1016/j.dsm.2021.12.001

21. Wang, X., Song, G., Yang, Z., Chen, R., Zheng, Y., Hu, H., Su, X., Chen, P.: Association between ageing population, median age, life expectancy and mortality in coronavirus disease (COVID-19). Aging (Albany NY), vol. 12, no. 24, pp. 24570–24578 (2020) doi: 10.18632/aging.104193
22. Merino, R. F., Chacón, C. I.: Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. Interfases, no. 10, pp. 165–189 (2017)
23. Rodríguez, Y. L., López, L. A.: A propósito del artículo "COVID-19. De la patogenia a la elevada mortalidad en el adulto mayor y con comorbilidades". Revista Habanera de Ciencias Médicas, vol. 19, no. 4 (2020)
24. Gobierno de México: Datos abiertos de bases históricas. Dirección General de Epidemiología de la Secretaría de Salud, www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-direccion-general-de-epidemiologia
25. Inca-Balseca, C. L., Paredes-Proaño, A. M., Cornejo-Reyes, P. J., Mena-Reinoso, Á. P.: Eficiencia de modelos de predicción de COVID-19 usando curvas ROC y matriz de confusión. Dominio de las Ciencias, vol. 8, no. 2, pp. 1442–1460 (2022)