# Exploring COVID-19 Trends in Mexico During the Winter Season with Explainable Artificial Intelligence (XAI)

Angélica Guzmán-Ponce, Rosa María Valdovinos-Rosas, Jacobo Leonardo González-Ruiz, Iván Francisco-Valencia, and J. Raymundo Marcial-Romero

***Abstract*—COVID-19 has become the most significant pandemic in recent years. Today, Mexico has recorded millions of infections and deaths since the pandemic started. Around the world, machine learning methods have been used to understand, predict or develop strategies to manage the virus and the pandemic. Although algorithms provide good results, it is necessary to understand why a model makes specific predictions with a particular data set. To explain this question, we apply Explainable Artificial Intelligence (XAI) in this paper. With this, it is possible to understand the characteristics that influence the model decisions when denoting between deaths and survivors.**

**As a case of study, the positive cases detected during the winter season of 2020-2021 and 2021-2022 were considered. In this season, respiratory diseases increased considerably, and in the study period, they influenced the increase in positive cases and the spread of COVID-19. Preliminary results suggest that age is essential when using a Random Forest model. Preliminary results suggest that age is essential when determining the prognosis of a patient infected by COVID-19 in winter seasons.**

**Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/8595**

***Index Terms*—XAI, Interpretable Random Forest, COVID-19, winter season, Mexico.**

## I. Introduction

The severe acute respiratory syndrome, SARS-CoV-2, responsible for the COVID-19 pandemic, has been the precursor of immense global issues. Since its initial identification in Wuhan, China, in December 2019, it has claimed over 6 million deaths worldwide[1]. Until now, Mexico has recorded over 7 million confirmed cases; unfortunately, 334,336 of these cases have culminated in death[2]. The repercussions of COVID-19 have a significant impact on society. Of this, the deep vulnerabilities and interdependencies of various sectors, such as education, transportation and politics, stand out. The latter results in the urgent need for cooperation and preparation to face the challenges derived from COVID-19.

A. Guzmán-Ponce is with the Department of Computer Languages and Systems, Institute of New Imaging Technologies, Universitat Jaume I, 12071 Castellón de la Plana, Spain (e-mail: aguzman@uji.es).

R. M. Valdovinos-Rosas, J. L. González-Ruiz, I. F.-Valencia and J. R. Marcial-Romero are with the Faculty of Engineering, Autonomous University of the State of Mexico, Toluca, Mexico (e-mails: rvaldovinosr@uaemex.mx, jlgonzalezru@uaemex.mx, ifranciscov@uaemex.mx, and jrmarcialr@uaemex.mx).

[1]https://coronavirus.jhu.edu/map.html

[2]https://datos.covid-19.conacyt.mx/

In recent years, Machine Learning (ML) has become one of the main disciplines that supports the challenges represented by the study of viruses, diseases and other topics related to the diagnosis and management of diseases. [1]–[3]. The state-of-the-art delineates potential solutions, encompassing medicine, political science, and social sciences, between others. [4]–[6].

The efficacy of ML algorithms depends on the quality and the data intrinsic characteristics [7]. In the context of the COVID-19, diverse data sources have been used with ML methods. For instance, from medical imaging modalities like X-ray and tomography scans, which aid in directly detecting the virus [8], [9], to big datasets acquired from the health sector. This last category of data is particularly notable because it enables understanding patterns related to patient survival, death causes or provides insights into the dynamics of transmission of the virus and spread trajectory [10]–[12].

ML models have complex architectures and algorithms, showing considerable promise in forecasting behavioural patterns within datasets. Nevertheless, these models often remain enigmatic, because it works like a *black-box* [13]. Such opacity has challenges, especially when it is imperative to understand the underlying mechanisms driving the predictions of the model.

With the need for transparency and intelligibility, there has been growing emphasis on Explainable Artificial Intelligence (XAI) [14]. XAI aims to bridge the gap between complex model outputs and human interpretability, i.e. XAI offers a means to interpret models, providing users with insights into how and why a given model makes specific decisions [15].

The research community has exhaustively studied COVID-19 due to its fast spread and unique medical symptoms. A particular focus has been on understanding its behaviour in seasons where respiratory virus occur more. During the winter season, medical practices face challenges due to respiratory illnesses. To manage these challenges, medical practices must focus on preventing the transmission of respiratory viruses, promoting and facilitating inmunization, and ensuring timely and accurate diagnosis and treatment of respiratory infections [16]. In fact, although it is not the only symptom, pneumonia was characteristic of COVID-19 and according to statistics worldwide, it is in the winter season that both cases and their severity increased.

Derived from the aforementioned, it is possible to note the need to further explore the patterns that allow models based on artificial intelligence to perform prediction or classification

tasks in patients diagnosed with COVID-19.

Up to now, there has been a notable absence of research focused on the incidence of COVID-19 during winter seasons and the application of Explainable Artificial Intelligence (XAI) methods. Consequently, the study detailed in this paper introduces a well-known XAI methodology that elucidates the predictions derived from patient data, enabling the identification of key attributes during the winter months, a period marked by a significant rise in respiratory diseases. Specifically, we analyze the cases reported in Mexico during the winter seasons of 2020-2021 and 2021-2022.

Although the use of machine learning algorithms have been and are being used in a large number of applications, one of the unknowns that is of interest in this article is to provide greater transparency and understanding of the models used with the help of XAI strategies. The case study in this paper is focused on identifying essential information about factors that influence patient outcomes, to support decision making. In this way, this research promotes the integration of machine learning in public health and highlights the need to have transparent methodologies in the forecasting and management of diseases that worsen in winter periods such as COVID-19.

From now on, the paper is structured as follows: Section 2 presents a series of preliminaries that include the general aspects of XAI. Section 3 describes the methodological strategy used in this study. The results obtained are shown in Section 4 and finally, section 5 includes the main concluding remarks.

## II. Preliminaries and Overview

The adoption of Artificial Intelligence (AI) techniques, especially ML methods, has had a significant clinical impact on diagnosis and disease prediction. However, in real medical practice, these models must be explained to users so that they can better understand how they work and the results they provide. To do this, XAI helps to the end user making decisions using AI algorithms. The remainder of this section describes XAI and provides an overview of work related to its medical applications.

### A. Explainable Artificial Intelligence

In the medical domain, the methodologies of IA through which knowledge is garnered often remain enigmatic for most health practitioners. The aim in this sector extends beyond making class predictions. Some questions arise, such as *What makes this prediction reliable?* and *How did this model get a conclusion?*. To address these inquiries, XAI act as a post-hoc analytical method, which elucidates the features that contribute to the model performance [17]. Notably, SHAP (SHapley Additive exPlanation) values emerge as a prevalent technique in XAI methods. SHAP is based on game theory to provide an interpretation of the predictions of ML models through the Shapley values. This approach applies game theory to allocate the output of a predictive model among its input features equitably. These features are compared to players in game theory, with the allocation based on their respective contributions to the overall prediction [18].

Consider a set of instances denoted by $DS$ named a dataset with $n$ instances, where an instance is composed by a pair $(x, y)$, wherein $x$ is an array of features described as $[x_1, x_2, \cdots, x_m]$ with $m$ features. Assume that $M$ is the set of all features, and the class is denoted by $y$. The Shapley value is defined by the following equation:

$$\phi_i = \sum_{S \subseteq M \setminus \{i\}} \frac{|S|!(|M| - |S| - 1)!}{|M|!} [f(S \cup \{i\}) - f(S)] \quad (1)$$

Where $S$ is a subset of $M$ excluding the $i$-th feature; the terms $f(S \cup \{i\})$ denotes the prediction made by the model with it includes the $i$-th feature, while $f(S)$ represents the prediction made by the model without this feature [18].

**Algorithm 1** SHapley Additive exPlanations

**Require:** $model, DS$
**Ensure:** Shapley values for each feature
  $M \leftarrow$ getAllFeatures$(DS)$
  $shapValues \leftarrow$ initializeArray$(|M|, 0)$
  **for** $i = 0$ **to** $|M| - 1$ **do**
    $feature \leftarrow M[i]$
    **for all** $S \subseteq M \setminus feature$ **do**
      $s1 \leftarrow S \cup \{feature\}$ {subset With Feature}
      $s2 \leftarrow S$ {subset Without Feature}
      $weight \leftarrow \frac{|S|! \times (|M| - |S| - 1)!}{|M|!}$
      $contribution \leftarrow$ modelPredict$(model, s1)$ − modelPredict$(model, s2)$
      $shapValues[i] \leftarrow shapValues[i] + weight \times contribution$
    **end for**
  **end for**
  **return** $shapValues$

The SHAP algorithm describe in 1, initiates by creating a group with no features, setting all initial SHAP values to zero. Features are then randomly added to this group. The impact of each feature is evaluated by comparing model predictions with and without the inclusion of that feature across all possible sequences of feature inclusion. The average of the features effect is calculated by dividing its total impact by the number of possible feature combinations.

SHAP values highlight the importance of each feature in predictions made by the model. Features that significantly increase predictions have high positive SHAP values, while those that decrease predictions show negative values. These values provide a granular view of how each feature influences the model, ensuring that the sum of all SHAP values equals the difference between a specific prediction and the average of all predictions in the dataset [15].

### B. Medical XAI

State-of-the-art has been employing XAI in the context of the COVID-19 pandemic. Predominantly, the research are focuses on methodologies and datasets suitable to images, such as X-rays and on non-image datasets from hospitals. An example of the latter are hospital records that include personal data of patients, laboratory results and others related to the medical history. But in no case they correspond to a diagnostic image, for example, ultrasound, tomography, magnetic resonance, etc.

Ong et al. [19] applied deep learning to interpret chest X-ray scans to diagnose COVID-19. While early research indicated that COVID-19 impacts could be seen in chest X-rays. They used the SHAP method to enhance the understanding of

COVID diagnosis through X-ray scans. The results effectively highlight significant areas in the images, thereby increasing the transparency and interpretability of the model.However, the trade-off between model complexity and interpretability, the absence of rigorous clinical validation, and the impact of imbalanced data on model performance highlights the need for further research to enhance the robustness and clinical relevance of XAI applications in medical diagnostics.

Given that doctors often need help deciphering why certain AI-based systems make specific predictions, Chauhan et al. [20] presented an approach to differentiate between Covid-19 positive and negative chest X-ray images using XAI. The results suggest that the model is interpretable and ensures that the outcomes are easily understood, promoting trust among healthcare professionals and patients. Nonetheless, notable constraints encompass the potential of variability in accuracy when applied to diverse real-world settings and the requirement for a certain level of technical proficiency to fully understand the XAI explanations, which may not be accessible to all healthcare practitioners.

On non-image datasets, Kırboğa et al. [17] used data from the Erzurum Training and Research Hospital and implemented the SHAP strategy to an interpretable explanation of Troponin data was provided in the COVID-19 process. For reference, Troponin is a protein located in heart muscles. Through the high-performing XGBoost model, they identified the most important features through the SHAP method. Notably, D-Dimer (an indicator of potential blood clotting conditions), mortality rate, CKMB (an enzyme termed Creatine kinase, which accelerates specific chemical reactions), and Glucose emerged as the most significant features. These findings suggest that XAI can forecast trends using big historical datasets. Despite employing explainable artificial intelligence to enhance transparency, these models still require a high level of expertise for interpretation.

Choudary et al. [21] classified whether the patient is affected with COVID or not and elucidated the significance of every attribute on the output using SHAP. Their results suggested that “Abroad travel” is the attribute that impacts the model and is followed by “Attended large gathering” y “Sore Throat”. These are the features that influence a certain person to get tested positive for COVID in June of 2020, which is when the dataset has been prepared. However, the study uses a synthetic dataset based on early WHO guidelines, limiting its real-world applicability and generalization to new data.

Additionally, to uncover potentially overlooked bio-markers indicative of COVID-19 infection severity, Wu et al. [22] analysed data from confirmed COVID-19 patients from Zhuhai, China. The explanation by SHAP reveals specific increases and decreases in certain bio-markers linked to severe infection and higher mortality risk. For instance, a patient with elevated N-terminal pro-brain natriuretic peptide levels might be classified as normal by a decision tree due to a lack of symptoms. However, there is a possibility that such a patient may show symptoms later on and progress to a more severe condition. The small sample size, complex interpretation methods like LIME and potential overfitting can be limitations. Dependency on detailed feature engineering and challenges in generalizing findings underscore the need for further research with more significant, diverse datasets to enhance clinical applicability.

The works mentioned above, despite being proposals that involve machine learning algorithms, mostly identify two limitations: the lack of seasonal representation in the data sets and detailed understanding of the behavior of the algorithms.

The latter limits the ability of AI models to capture variations in the presentation and severity of COVID-19 during winter periods, in addition to affecting the diagnostic accuracy and clinical generalization of the developed models. Both aspects are a central part of the study presented in this paper.

## III. Methodology

With the continuous generation of data, the extraction of knowledge from datasets has become a required mechanism in the medical domain. In the literature there are several useful methodologies for these purposes. The study presented here summarises the knowledge extraction process in three phases (Fig. 1). Each of these phases are explained in the following sections.

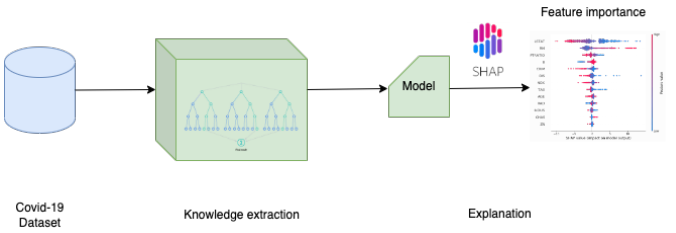



Fig. 1. Methodological Framework of XAI for identifying key features of COVID-19 during the winter season.

### A. COVID-19 Dataset

The COVID-19 database from Mexico is an open-access resource available on the official website of the General Directorate of Epidemiology[3]. Consolidated from health centres across Mexico, it has been updated since April 14, 2020 until today. The database includes patient details like nationality, age, gender, location, treatment facility, pre-existing medical conditions, and COVID-19 test results between others.

Our study was focused on data about patients who received a positive COVID-19 diagnosis during two specific winter intervals: from November 1, 2020, to February 28, 2021, and subsequently, from November 1, 2021, to February 28, 2022. The analytical focus was prominently placed on two critical locations in Mexico: Mexico City and Mexico State, due to they having shown a historical high infections and mortality rates in Mexico [23].

The General Epidemiology Department provides a data directory with categorized attributes within the dataset. For the purpose of this study, the focus was narrowed to confirmed SARS-COV2 cases from Mexico City and Mexico State, as identified by the Clinical Epidemiological Association and the Determination Committee based on positive results from laboratory and antigen tests conducted by health personnel.

[3]https://www.gob.mx/salud/documentos/datos-abiertos-152127

TABLE I
DATA DISTRIBUTION

| Winter-season | Survived | Died | IR |
|---|---|---|---|
| 2020-2021 | 210,695 | 16,410 | 12.84 |
| 2021-2022 | 47,403 | 674 | 70.33 |

The selection of the dataset for COVID-19 was guided by clinical relevance and the availability of epidemiological records, concentrating on crucial attributes such as age, gender, and pre-existing health conditions, including pneumonia, diabetes, and hypertension. These factors are known to influence the severity and mortality of COVID-19 [24], [25].

The final dataset considers two categories: those records without a death date were labelled as *Survived* (class 1), and the rest as *Died* (class 0). The study consider specific attributes due to their importance, which includes Sex (female/male), Intubation, Pneumonia, Pregnancy, Diabetes, Chronic Obstructive Pulmonary Disease (COPD), Asthma, Immune disorders, Hypertension, Heart disease, Obesity, Chronic kidney issues, Smoking habits, medical intensive care unit (MICU). Except sex attribute, all other attributes take binary values of "yes" or "no". The Age attribute was categorised based on the next range:

- Over 60 (age $>= 60$): *adult60*
- Adult in their 50s (age between 50 to 59): *adult50*
- Adult in their 40s (age between 40 to 49): *adult40*
- Adult in their 30s (age between 30 to 39): *adult30*
- Adult in their 20s (age between 18 to 29): *adult20*
- Youth (age between 12 to 17): *Youth*
- Child (age $<= 11$): *child*

The data were pre-processed using one-hot encoding, giving as result a dataset with 35 features. Table I shows the details of the data distribution of the subsets considered during the winter seasons in two central Mexican locations. The final column presents the imbalance ratio (IR), defined as the proportion of instances in the most representative class to those in the least representative classes. This metric offers a quantitative perspective on the disparity between classes [26].

As can be seen, in the data distribution shown in Table I, a severe class imbalance is evident, mainly in the winter period of 2020-2021. This is represented by a substantial imbalance ratio of 12.84. It is considered that there is a class imbalance when some of the classes that make up the study problem are more represented than others. This situation represents a problem because most ML algorithms tend to better recognise the most represented class (majority), ignoring the least represented class (minority) [26].

### B. Knowledge Extraction

In this study, a ML model was built to analyse the relationship that may exist between COVID-19 positive patients and its prevalence in winter seasons. For evaluation purposes, each dataset was assessed using the 10-fold cross-validation method. This method evaluates the performance of ML models by partitioning the dataset into ten subsets. For each iteration, nine subsets perform as the training set, while the remaining subset functions as the test set. We incorporate Scikit-learn, a ML API in Python [27].

For classification purposes, we employ the Random Forest (RF) classifier, an ensemble learning method. We use all default hyper-parameters except for the random seed, which we set to 42. This technique combines multiple decision trees, linking the Bagging algorithm and the random subspace approach (Fig. 2). Herein, each tree is built from a bootstrap sample of the original dataset and is not pruned after construction, which can lead to partial over-fitting of the data.

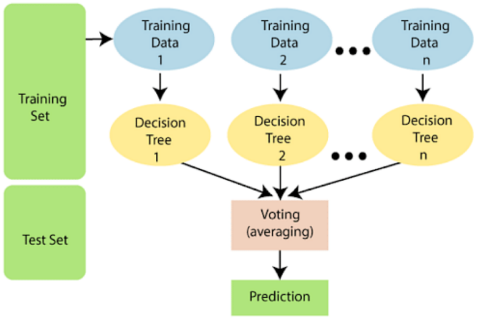



Fig. 2. Random forest algorithm strategy[4].

The feature to split at each branch in the tree is chosen from a random subset of size $m_s$ from the complete set of features to increase diversity. This subset is different for each branching point. Notably, this subset varies for each branching instance. [28]. The Scikit-learn API uses averaging to improve the predictive accuracy and control over-fitting.

### C. Performance Evaluation

To assess the performance of the model, we consider the confusion matrix. It is commonly used a $2\times2$ confusion matrix as that given in Table II, where each entry $(i, j)$ contains the number of correct/incorrect classifications [29]. The next simple measures can be derived: TP and TN are the number of true positives (actually positive, and classified as positive or minority class) and the true negatives (actually negative, and classified as negative or majority class) respectively. FP and FN refer to the number of instances misclassified: FP actually negative, but classified as positive, and FNactually positive, but classified as negative.

TABLE II
CONFUSION MATRIX

| | Classified as positive | Classified as negative |
|---|---|---|
| Actual positive | True Positive (TP) | False Negative (FN) |
| Actual negative | False Positive (FP) | True Negative (TN) |

As evidenced in Table I, the datasets exhibit a class imbalance problem, making the geometric mean a well-established measure within this context [30].

$$g = \sqrt{a^+ \cdot a^-} \tag{2}$$

[4] https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm

where $a^+$ is the sensitivity or true-positive rate (i.e., the accuracy on the minority class):

$$a^+ = \frac{TP}{TP + FN} \tag{3}$$

and $a^-$ is the specificity or true-negative rate (i.e., the accuracy on the majority class):

$$a^- = \frac{TN}{TN + FP} \tag{4}$$

In addition to existing metrics, we added the F1-Score macro to evaluate our model. The F1-Score balances precision and recall, treating all classes equally. This metric provides unbiased insights into the accuracy of the model for each class [26].

### D. Explainability Model

In this study, the interpretability and explainability of the random forest model were explored utilising the SHapley Additive exPlanations (SHAP) methodology [15].

The SHAP method has shown its efficacy in the medical field, shedding light on clinical decision-making derived from image and non-image datasets. Furthermore, its usage has been widely studied in COVID-19 cases. Consider that SHAP has been developed with Python[5], the last step of the Fig. 1 illustrates the process to get an overview of which features are most important for the random forest model we can plot the SHAP values of every feature.

## IV. Results and Discussion

This study aims to mitigate the impact of COVID-19 by exploring correlations between dataset characteristics and patient outcomes, with a particular focus on identifying predictors of mortality. Figs. 3a and 3b show initial data analysis that demonstrates these relationships.

From the figures, we can see that, for example, Pneumonia_NO has a very high correlation with the class, indicating that not having pneumonia is strongly related to being alive. Similarly, Intubation_NO also shows a significant correlation with survival, meaning that patients who do not require intubation tend to survive more. Conversely, their presence is related to the probability of not surviving. As observed, all comorbidities are correlated with this fact in both winter periods.

In this study, a comprehensive analysis was performed to evaluate the performance of the Random Forest classifier when it is applied to data from COVID-19 positive patients. As we can see in Table III, the classifier shown suboptimal performance on geometric mean, which can mainly be attributed to the presence of class imbalance problem. Specifically, survivals occurred significantly more frequently than deaths.

From results in Table III, the classifier demonstrates better performance during the first winter season (2020-2021) compared to the subsequent winter (2021-2022). This difference in performance is associated with the less evident class imbalance observed in the first winter period (see in Table I). The latter could be due to the fact that mass vaccination campaigns began in 2021, which could have caused a decrease in infections, as well as deaths due to the disease, as documented by Camacho et al. [31]. Both metrics (geometric mean and macro F1-score) show a consistent and proportional decrease, reinforcing the observation that the model performed worse in the second season due to fewer cases of died. The following tables IV and V show the average confusion matrix across the 10-fold cross-validation for 2020-2021 and 2021-2022 winter seasons, illustrating predictive model performance.

[5]https://shap.readthedocs.io/en/latest/

TABLE III
Outcomes in Random Forest Evaluation

| Winter-season | Geometric Mean | F1-score macro |
|---|---|---|
| 2020-2021 | 67.02% | 75.09% |
| 2021-2022 | 42.79% | 63.15% |

TABLE IV
Average Confusion Matrix Across Folds on 2020-2021

| | Classified as Survived | Classified as Died |
|---|---|---|
| Actual Survived | 20636 | 433 |
| Actual Died | 887 | 754 |

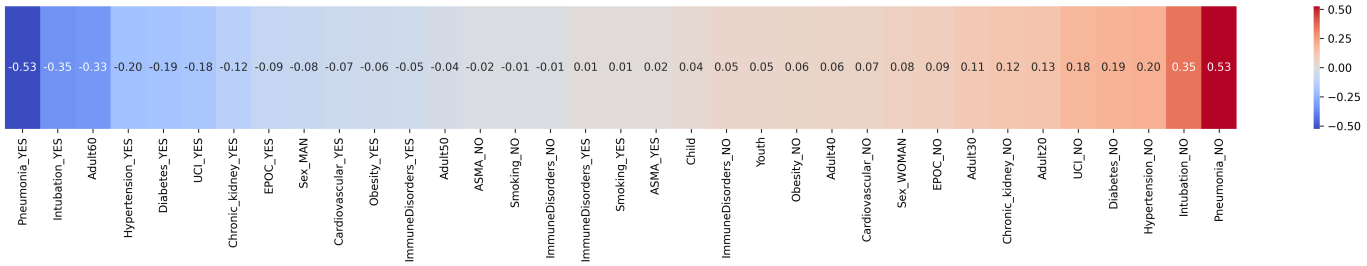
TABLE V
Average Confusion Matrix Across Folds on 2021-2022

| | Classified as Survived | Classified as Died |
|---|---|---|
| Actual Survived | 4727 | 13 |
| Actual Died | 54 | 13 |

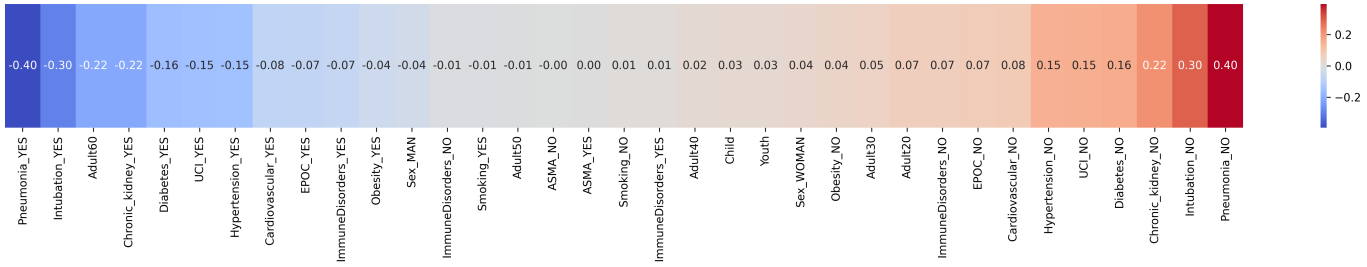
From Table IV, the sensitivity and specificity values are approximately 97.94% and 45.95%, respectively. These suggest that the performance of identifying as survived is high. However, the performance on identifying as dead is significantly lower, at only about 46%. Therefore, while the model is highly effective in identifying Survived, its ability to identify as dead correctly is relatively poor. This pattern is consistent with Table V, where the sensitivity reaches approximately 99.73% and specificity stands at roughly 19.40%. This suggests that, although the model is effective at identifying survivors, its performance in detecting death is significantly inadequate.

Explainable Artificial Intelligence (XAI) facilitates the interpretability of classification models, and in the context of this study, it provides the facilities to identify the determining characteristics of presenting COVID-19 disease. Historically, winter seasons have demonstrated a strong correlation with respiratory ailments. Nonetheless, in the era of COVID-19, discerning specific attributes that signal survival among the infected people becomes imperative.

For this reason, by using the SHAP method it was possible to identify the 20 most prominent characteristics that contribute to the differentiation of classes within the model.




(a) Winter 2020-2021




(b) Winter 2021-2022

Fig. 3. Analysis of Feature-Class Correlations.

Fig. 4 illustrates the most consequential features for the winter season spanning 2020-2021.

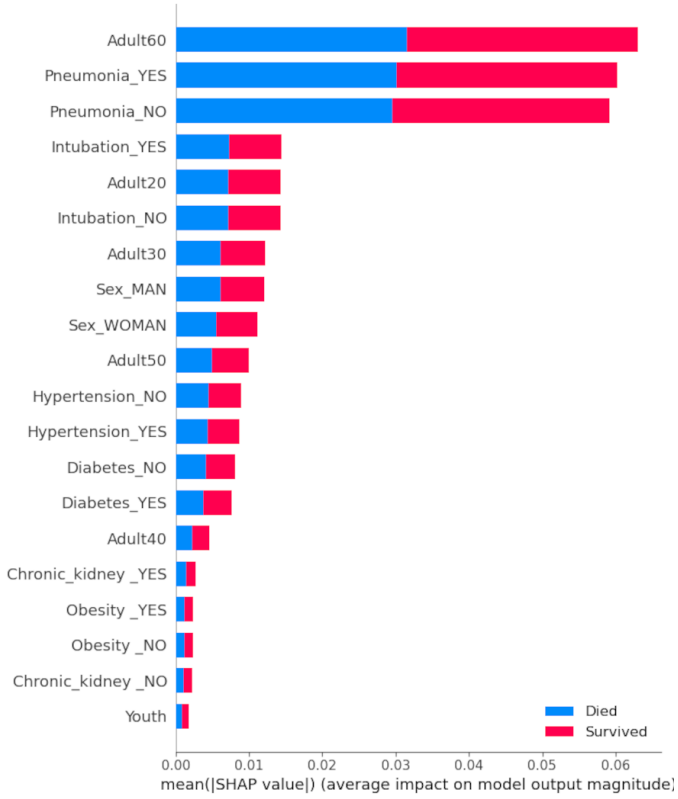



Fig. 4. SHAP Summary Plot for Survival Predictions, Winter 2020-2021.

From Fig. 4 some interesting aspects can be drawn. The findings underscore that age emerges as a critical determinant, particularly for individuals aged 60 and above. Furthermore, pneumonia is identified as a significant feature influencing predictions of survival and mortality. Additionally, the use of intubation during this particular season is noted as an informative characteristic for predicting survival outcomes.

For the winter season spanning 2021-2022, Fig. 5 reveals that age persists as the paramount feature analogous to observations from the prior winter period, particularly for individuals exceeding 60 years. In addition, the manifestation or absence of pneumonia also persists as a significant attribute in determining the likelihood of surviving a COVID-19 infection. Contrarily, compared to the preceding winter, the factor of intubation did not emerge as a distinctive characteristic predictor of survival. The latter could be due to the fact that by this period, a large part of the population in Mexico had already received at least one dose of vaccine.

Something important to note in this winter period is the role of underlying health conditions, which has increased its predictive relevance. In particular, the presence of chronic kidney complications has been identified as a relevant characteristic within the model, suggesting that their presence correlates with a higher probability of survival.

Throughout both winter seasons, the prevalence of hypertension, diabetes and obesity can be observed. These conditions, commonly are categorised as comorbidity, have long been identified as determining factors that can aggravate many diseases. Age groups such as 20, 30, 50, and 40 years old emerge as significant features for the model during both winter periods. However, interestingly, the Adult40 age group is ranked towards the lower end of the feature ranking, implying that it has a relatively lesser impact on the model.

Segments of this age cohort had received vaccination against COVID-19 or had acquired immunity naturally due to previous infection, as reported by Bello-Chavolla et al. [32]. Moreover, they highlight the effectiveness of the vaccines decreased in

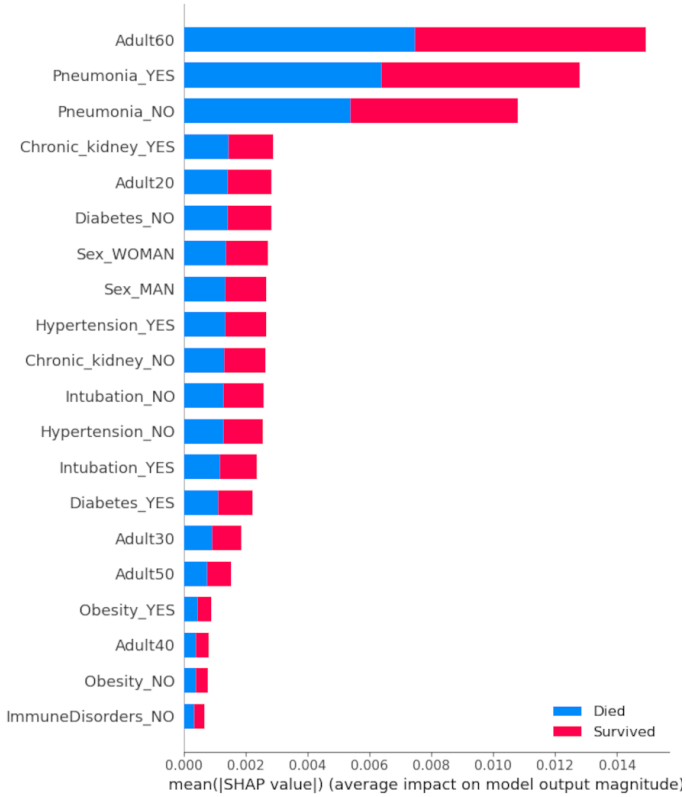



Fig. 5. SHAP Summary Plot for Survival Predictions, Winter 2021-2022.

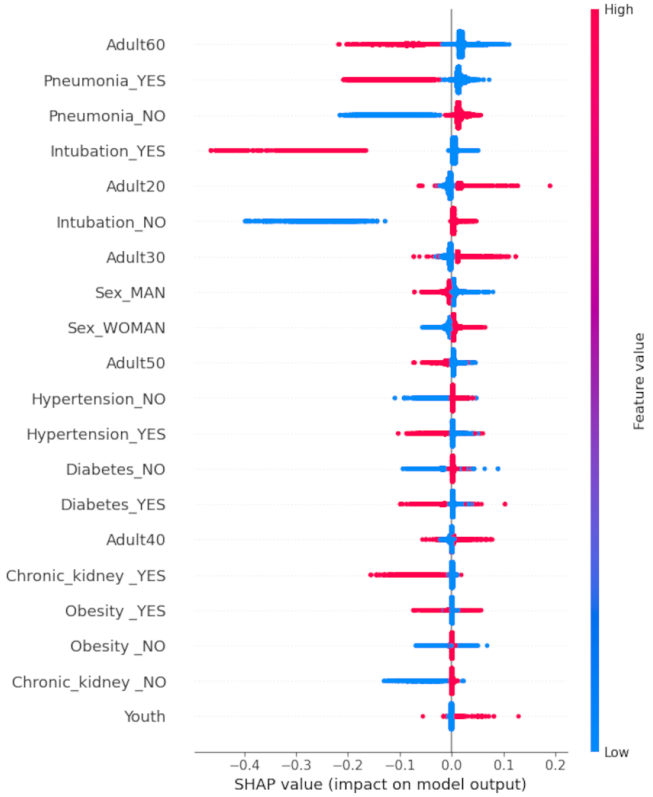



Fig. 6. SHAP values on Survied from the impact of the random forest model for winter 2020-2021.

adults over 60 years of age. The connection between this age group and health outcomes suggests that vaccination campaigns targeting these age groups were more effective or that the group was more receptive to the immunity produced by them. Notably, most vaccinated cases were within the 30 to 59 age group [32]. While the dataset used in this study does not allow for precise tracking of vaccinated or unvaccinated individuals, this could be another reason age ranges influence the model.

When decision tree models are applied to data from various winter periods, there are noticeable differences in the maximum magnitude of SHAP values. For instance, the maximum SHAP value for the 2020-2021 model is around 0.06, whereas for the 2021-2022 model, it is 0.014. These changes can be attributed to seasonal fluctuations in patient populations, advancements in public health interventions, natural variability, and data biases, particularly due to class imbalance in the dataset.

Figs. 6 and 7 illustrate the importance and impact of different features on predicting the Survived class. In this case, a high and positive SHAP value for a feature would indicate that the presence of that feature significantly increases the probability of survival. Conversely, a high and negative SHAP value would indicate that the presence of that feature significantly decreases the probability of the class of interest.

In both winter seasons, the model demonstrates a clear difference in the influence of pneumonia. These features decrease the probability of survival, as indicated by red points tending towards the left. At the same time, the absence increases the probability of being alive, as evidenced by blue points shifting towards the right. However, starting from the fourth impactful characteristic, it is observed that for the first period, intubation was determined to decrease the probability of survival compared to the second period, where chronic kidney disease is a characteristic that has a more significant impact. For the age of 20 years, the probability of survival is higher. After this feature, we can see that, in general, the presence of comorbidities decreases the probability of survival; this aligns with medical understanding [25].

## V. Conclusions

The COVID-19 pandemic caused by SARS-CoV-2 has affected the entire world, emphasising the need for rapid action and global cooperation. Machine learning has become an indispensable tool to address related challenges, assist detection, tracking, and create solutions in multiple fields. This study was focused on data from two particular winter seasons, concentrating on Mexico City and Mexico State due to the high mortality rates reported during the study periods, mainly in 2021.

Our findings highlighted the performance of the random forest classifier in the two winter seasons studied. Although there was a significant class imbalance issue, the classifier performed better in the first winter season. The difference in performance is due to the result of the vaccination campaigns

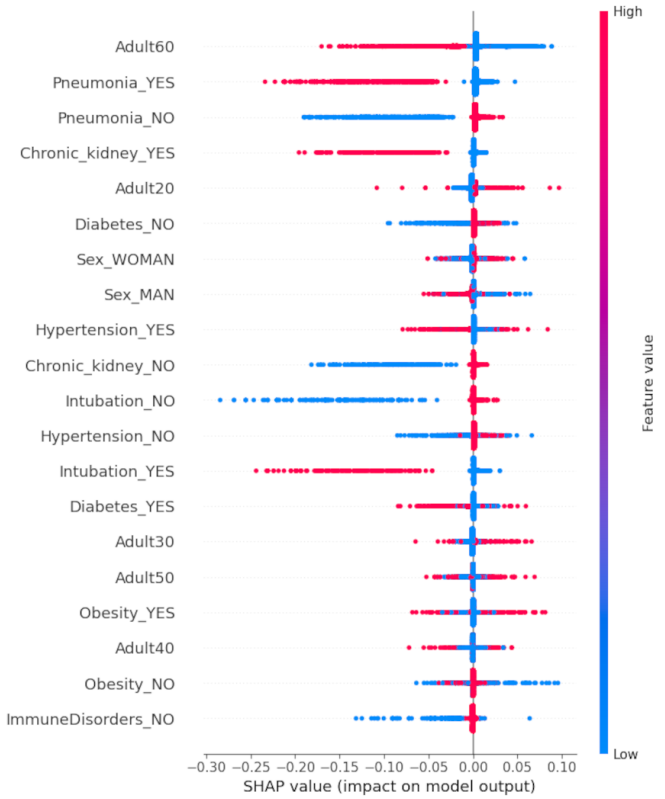



Fig. 7. SHAP values on Survived from the impact of the random forest model for winter 2021-2022.

carried out during the years 2021 and 2022, since the number of survivors increased.

With XAI, it was possible to discern the characteristics that most influence the appearance of COVID-19 during the winter seasons. The results suggest that age, specifically for people aged 60 years and older, and the manifestation or absence of pneumonia were dominant determinants of survival outcomes. However, the role of intubation diminished last winter, and chronic kidney complications emerged as a higher factor. Hypertension, diabetes and obesity constantly appeared as co-morbidities that aggravate the disease. One observation was the distinctive susceptibility or resilience to COVID-19 in the 30-50 age group during the winter of 2021-2022, suggesting the potential effectiveness of vaccination campaigns in this area. It is essential to highlight the importance of vaccination tracking in COVID-19 during winter seasons. However, the current database does not permit adequate tracking. We acknowledge this limitation and the need for improved data systems for future research.

In summary, this research highlights the critical role of machine learning in understanding the complexities of the COVID-19 pandemic. It also offers insights into the potential protective effect of vaccination during the studied winter seasons in Mexico. The findings from this study could be instrumental in guiding public health decisions and providing valuable insights for future pandemic alerts. Future research will address the class imbalance using data-level preprocessing methods and refine our model using insights from highlighted features. It will involve a comprehensive tuning process using grid search techniques to explore various parameter settings. Additionally, comparisons with models like Support Vector Machines and Neural Networks will be conducted to expand insights and validate findings across different algorithmic approaches.

## Acknowledgments

This work was partially supported by the COMECyT with UAEMex register 6847/2023E; Angélica Guzmán-Ponce had the support of the Margarita Salas postdoctoral contract MGS/2021/23 (UP2021-021), funded by the European Union-NextGenerationEU.

## References

[1] R. F. Albuquerque Paiva de Oliveira, C. J. A. Bastos Filho, A. C. A. M. V. F. de Medeiros, P. J. Buarque Lins dos Santos, and D. Lopes Freire, "Machine learning applied in sars-cov-2 covid 19 screening using clinical analysis parameters," *IEEE Lat. Am. Trans.*, vol. 19, no. 6, pp. 978—-985, Jun. 2021. doi: 10.1109/TLA.2021.9451243

[2] C. Noak and S. Wang, "Analysis of covid-19 data using advanced machine learning models," in *2022 3rd PRML*, Chengdu, China, 2022, pp. 430–437. doi: 10.1109/PRML56267.2022.9882191

[3] S. Tiwari, P. Chanak, and S. K. Singh, "A Review of the Machine Learning algorithms for Covid-19 Case Analysis," *IEEE Trans. Artif. Intell.*, vol. 4, no. 1, pp. 44–59, 2023. doi: 10.1109/TAI.2022.3142241

[4] N. Alballa and I. Al-Turaiki, "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review," *Informatics in Medicine Unlocked*, vol. 24, p. 100564, 2021. doi: 10.1016/j.imu.2021.100564

[5] Y. V. Kistenev, D. A. Vrazhnov, E. E. Shnaider, and H. Zuhayri, "Predictive models for COVID-19 detection using routine blood tests and machine learning," *Heliyon*, vol. 8, no. 10, p. e11185, 2022. doi: 10.1016/j.heliyon.2022.e11185

[6] V. Paramarta, A. Mailangkay, H. Amalia, and D. Chrismas, "Public sentiment analysis of indonesian tweets about covid-19 vaccination using different machine learning approaches," in *2022 Seventh ICIC*, Bali, Indonesia, 2022, pp. 1–5. doi: 10.1109/ICIC56845.2022.10006896

[7] D. García-Gil, F. Luque-Sánchez, J. Luengo, S. García, and F. Herrera, "From Big to Smart Data: Iterative ensemble filter for noise filtering in big data classification," *Int. J. Intell. Syst.*, vol. 34, no. 12, pp. 3260–3274, 2019. doi: 10.1002/int.22193

[8] I. Rodrigues, G. L. Santos, D. F. Sadok, and P. T. Endo, "Classifying COVID-19 positive X-ray using deep learning models," *IEEE Lat. Am. Trans.*, vol. 19, pp. 884—-892, Jun. 2021.

[9] E. Cortés and S. Sánchez, "Deep Learning Transfer with AlexNet for chest X-ray COVID-19 recognition," *IEEE Lat. Am. Trans.*, vol. 19, no. 6, pp. 944—-951, Jun. 2021. [Online]. Available: https://latamt.ieeer9.org/index.php/transactions/article/view/4336

[10] A. D. Orjuela-Cañon and O. Perdomo, "Clustering Proposal Support for the COVID-19 Making Decision Process in a Data Demanding Scenario," *IEEE Lat. Am. Trans.*, vol. 19, no. 6, pp. 1041—-1049, Jun. 2021. [Online]. Available: https://latamt.ieeer9.org/index.php/transactions/article/view/4403

[11] D. C. d. S. Gomes and G. L. d. O. Serra, "Machine Learning Model for Computational Tracking and Forecasting the COVID-19 Dynamic Propagation," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 3, pp. 615–622, 2021. doi: 10.1109/JBHI.2021.3052134

[12] S. Ghafouri-Fard, H. Mohammad-Rahimi, P. Motie, M. A. Minabi, M. Taheri, and S. Nateghinia, "Application of machine learning in the prediction of COVID-19 daily new cases: A scoping review," *Heliyon*, vol. 7, no. 10, p. e08143, 2021. doi: 10.1016/j.heliyon.2021.e08143

[13] F. Giuste, W. Shi, Y. Zhu, T. Naren, M. Isgut, Y. Sha, L. Tong, M. Gupte, and M. D. Wang, "Explainable Artificial Intelligence Methods in Combating Pandemics: A Systematic Review," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 5–21, 2023. doi: 10.1109/RBME.2022.3185953

[14] I. Ahmed, G. Jeon, and F. Piccialli, "From artificial intelligence to explainable artificial intelligence in industry 4.0: A survey on what, how, and where," *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5031–5042, 2022. doi: 10.1109/TII.2022.3146552

[15] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, 2020. doi: 10.1038/s42256-019-0138-9

[16] C. E. Kistler, R. L. Jump, P. D. Sloane, and S. Zimmerman, "The Winter Respiratory Viral Season During the COVID-19 Pandemic," *J. Am. Med. Dir. Assoc.*, vol. 21, no. 12, pp. 1741–1745, 2020. doi: 10.1016/j.jamda.2020.10.030

[17] K. K. Kırboğa, E. U. Küçüksille, M. E. Naldan, M. Işık, O. Gülcü, and E. Aksakal, "CVD22: Explainable artificial intelligence determination of the relationship of troponin to D-Dimer, mortality, and CK-MB in COVID-19 patients," *Comput. Methods Programs Biomed.*, vol. 233, p. 107492, 2023. doi: 10.1016/j.cmpb.2023.107492

[18] Y. Chen, R. Calabrese, and B. Martin-Barragan, "Interpretable machine learning for imbalanced credit scoring datasets," *Eur. J. Oper. Res.*, 2023. doi: 10.1016/j.ejor.2023.06.036

[19] J. H. Ong, K. M. Goh, and L. L. Lim, "Comparative analysis of explainable artificial intelligence for covid-19 diagnosis on cxr image," in *2021 IEEE ICSIPA*, Kuala Terengganu, Malaysia, 2021, pp. 185–190. doi: 10.1109/ICSIPA52582.2021.9576766

[20] T. Chauhan and S. Sonawane, "Explicable ai for surveillance and interpretation of coronavirus using x-ray imaging," in *2023 ESCI*, Pune, India, 2023, pp. 1–6. doi: 10.1109/ESCI56872.2023.10099633

[21] M. N. S. Choudary, V. B. Bommineni, G. Tarun, G. P. Reddy, and G. Gopakumar, "Predicting Covid-19 Positive Cases and Analysis on the Relevance of Features using SHAP (SHapley Additive exPlanation)," in *2021 Second ICESC*, Coimbatore, India, 2021, pp. 1892–1896. doi: 10.1109/ICESC51422.2021.9532829

[22] H. Wu, W. Ruan, J. Wang, D. Zheng, B. Liu, Y. Geng, X. Chai, J. Chen, K. Li, S. Li, and S. Helal, "Interpretable Machine Learning for COVID-19: An Empirical Study on Severity Prediction Task," *IEEE Trans. Artif. Intell.*, vol. 4, no. 4, pp. 764–777, 2023. doi: 10.1109/TAI.2021.3092698

[23] S. Dahal, R. Luo, M. H. Swahn, and G. Chowell, "Geospatial Variability in excess death rates during the COVID-19 pandemic in Mexico: Examining Socio Demographic, Climate and Population Health Characteristics," *Int. Jour. Infec. Diseases*, vol. 113, pp. 347–354, 2021. doi: 10.1016/j.ijid.2021.10.024

[24] P. Liu, A. Blet, D. Smyth, and H. Li, "The Science Underlying COVID-19 Implications for the Cardiovascular System," *Circulation*, vol. 142, no. 1, pp. 68–78, 2020. doi: 10.1161/CIRCULATIONAHA.120.047549

[25] J. Tian, X. Yuan, J. Xiao, Q. Zhong, C. Yang, B. Liu, Y. Cai, Z. Lu, J. Wang, Y. Wang, S. Liu, B. Cheng, J. Wang, M. Zhang, L. Wang, S. Niu, Z. Yao, X. Deng, F. Zhou, W. Wei, Q. Li, X. Chen, W. Chen, Q. Yang, S. Wu, J. Fan, B. Shu, Z. Hu, S. Wang, X. Yang, W. Liu, X. Miao, and Z. Wang, "Clinical characteristics and risk factors associated with COVID-19 disease severity in patients with cancer in Wuhan, China: a multicentre, retrospective, cohort study," *Lancet Oncol.*, vol. 21, no. 7, pp. 893–903, 2020. doi: 10.1016/S1470-2045(20)30309-0

[26] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 2019. doi: 10.1016/j.patcog.2019.02.023

[27] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, Praga,Czech Republic, 2013, pp. 108–122.

[28] C. Sammut and G. I. Webb, Eds., *Random Forests*. Boston, MA: Springer US, 2017, pp. 1054–1054. ISBN 978-1-4899-7687-1

[29] Q. Kang, X. Chen, S. Li, and M. Zhou, "A noise-filtered under-sampling scheme for imbalanced classification," *IEEE Trans. on Cybernetics*, vol. 47, no. 12, pp. 4263–4274, 2017.

[30] J. Deng, J. Sun, W. Peng, D. Zhang, and V. Vyatkin, "Imbalanced multiclass classification with active learning in strip rolling process," *Knowl.-Based Syst.*, vol. 255, p. 109754, 2022. doi: 10.1016/j.knosys.2022.109754

[31] M. E. Camacho Moll, A. M. Salinas Martínez, B. Tovar Cisneros, J. I. García Onofre, G. Navarrete Floriano, and M. Bermúdez de León, "Extension and Severity of Self-Reported Side Effects of Seven COVID-19 Vaccines in Mexican Population," *Front. Public Health*, vol. 10, 2022. doi: 10.3389/fpubh.2022.834744

[32] O. Y. Bello-Chavolla, N. Antonio-Villa, S. Valdés-Ferrer, C. Fermín-Martínez, L. Fernández-Chirino, D. Ramírez-García, J. Mancilla-Galindo, A. Kammar-García, J. Ávila Funes, C. H. Zuniga-Gil, M. García-Grimshaw, S. E. Ceballos-Liceaga, G. Carbajal-Sandoval, J. A. Montes-González, C. Zaragoza-Jiménez, G. Garcia-Rodriguez, R. Cortes-Alcala, G. Reyes-Terán, H. Lopez-Gatell, and L. Gutiérrez-Robledo, "Effectiveness of a nationwide covid-19 vaccination program in mexico against symptomatic covid-19, hospitalizations, and death: a retrospective analysis of national surveillance data," *International Journal of Infectious Diseases*, vol. 129, pp. 188–196, 2022. doi: 10.1016/j.ijid.2023.01.040

**Angélica Guzmán-Ponce** she was conferred a Ph.D. degree in Computer Science, from the Autonomous University of the State of Mexico in 2021. She is a member of the National System of Researchers (CONACYT) Level I. She is currently a postdoctoral researcher in the University Jaume I (Castellon de la Plana, Spain) and the Universitat Politècnica de València. Her research interests lie in Machine Learning and Graph Theory.

**Rosa María Valdovinos-Rosas** she was conferred a Ph.D. degree in Computer Science. She is a member of the National System of Researchers (CONACYT) Nivel II and the AMEXCOMP. She has been a Full-Time Lecturer-Researcher at the Autonomous University of the State of Mexico (UAEMex). She contributes to the strengthening and consolidation of the scientific community through the training of quality human resources and disseminating knowledge and science in academic-scientific events at a national and international level.

**Jacobo Leonardo González-Ruiz** full-time Professor at the Autonomous University of the State of Mexico. Member of the CONACYT, Level C. In 2008 he obtained the title of Computer Engineer and completed the master's degree in engineering sciences in 2014. In 2018, he obtained the degree of PhD in Engineering Sciences from the Autonomous University of the State of Mexico.

**Iván Francisco-Valencia** he obtained a Ph.D. degree in Engineering Sciences from the Autonomous University of the State of Mexico in 2021 and is a Level C member of the National System of Researchers (CONACYT). Currently, he is a postdoctoral researcher at the Toluca Institute of Technology, Mexico. His research interests include Combinatorial Game Theory, Optimization, and Artificial Intelligence

**J. Raymundo Marcial-Romero** he received his PhD in Computational Science from Birmingham University in 2005. He has been a Full-Time Lecturer-Researcher at the Department of Engineering of the Autonomous University of the State of Mexico. He is a member of the National System of Researchers (CONACYT) and Nivel I. His research interest includes approximation theory, computational complexity and graph theory.