Health information mining: a ranking of analytical methods for determining which factors influence the efficacy of cosmetic surgery

Mohammed Khaleel Almadhoun[1], Salman Asghar[2], Dr. Ayesha Billoo[3], Dr. Saneea Saeed[4], Dr. Usama Rizwan[5], Dr Rani Hajra Batool[6]

[1]Mutah University, Karak, Jordan. Email: Pceecpp@gmail.com
[2]Department of Product and Industrial Design (PID), University of Engineering and Technology (UET), Lahore, Pakistan. Email: salman.asghar@uet.edu.pk
[3]Jinnah Medical and Dental College, Karachi, Pakistan. Email: ayeshabilloo1d@gmail.com
[4]Department of Internal Medicine, University Hospitals of Morecambe Bay NHS Foundation Trust, United Kingdom. Email: saneea2020@gmail.com
[5]BDS, Dentistry, Ziauddin University, Karachi, Pakistan. Email: usamarizwan106@gmail.com
Ripha International University, Islamabad, Pakistan. Email: Hajra8850@gmail.com

ABSTRACT:

The evaluation session is a crucial first step in the increasingly common practice of cosmetic surgery worldwide, but not all patients who schedule one have the procedure. To identify the factors influencing the effectiveness of cosmetic surgery, this article proposes an analytical model based on a grading system employing several methods, including correlation analysis, principal component analysis, decision trees, logistic regression, and association rules. Data from a surgical clinic in Medellin is used to verify the accuracy of the analytical model. Patients who opt for cosmetic surgery are those who meet the following criteria: 1) they want more than one procedure; 2) their primary procedure of choice is one of the following: facial enhancement, breast augmentation, discectomy, abdominoplasty, implantology, or buttock augmentation; 3) they have an average body mass index; 4) they don't smoke; 5) they don't have a history of diabetes in their family.

Keywords: data mining; data analysis; health care. Cosmetic surgery.

DOI: 10.48047/ecb/2023.12.9.236

INTRODUCTION:

A plethora of research has been conducted on the increasing popularity of cosmetic surgery in recent years. Individuals in Pakistan interested in learning more about plastic surgery's benefits have increased their consultation appointment scheduling. Clinics see more patients due to experts' increased expertise, and active people need regular checkups, healing, and follow-up care from the specialist. Due to the many patients, evaluation appointments must be prioritized to reduce the wait times for new patients (Morera de la Vall González, 2018).

Due to the essential nature of the appraisal appointment in the context of cosmetic surgery, it is important to determine or set minimum criteria for their assignment priority. Personal information, medical history, medications, family history, habits, and a physical evaluation are all part of the patient's medical record, which can be used to establish a priority list. The decision to get cosmetic surgery is influenced by factors which may be determined by analyzing the history, which contains data on roughly 85 different variables. The solution will help a cosmetic clinic prioritize which patients should be seen for evaluations based on those features (Espinosa-Neisa, 2021; Hernandez Celis et al., 2021).

Since most evaluation jobs are delegated over the phone, the person in charge of this procedure will likely need to collect data on the indicated factors by telephone. Data Science is utilized as an interdisciplinary study to evaluate the 85 factors; this includes applying mathematical methods, statistics, data visualization, computing algorithms for data mining, and cloud processing. "Big Data" and "Data Analytics" have recently been used to delve into data science. The analysis is where the solution to this problem of interest is supposed to be found. The article uses data mining methods to retrieve valuable data from patients' medical records who have undergone cosmetic surgery (Rojas-Rojas, 2020).

Prediction and description are the two primary focuses of data mining approaches. Applying supervised algorithms such as neural networks, decision trees, support vector machines, Bayesians, k nearest neighbors, and others, prediction enables the forecasting of future events on an objective variable. However, unsupervised algorithms like k-means, cobweb, Expectation-Maximization, and association rules can be used to explore the dataset thanks to the description. Data mining has been significantly used in medicine, allowing researchers to make accurate prognoses and detailed descriptions of the effects of life-threatening illnesses, including cancer, diabetes, and leukemia (Serafín Muñoz, 2022).

Predictive work includes using supervised algorithms as decision trees to extrapolate the success of brachytherapy in treating prostate cancer. In addition, association rules were formulated based on the variables used in the prediction, and Bayesian algorithms are suggested to be used to compare or evaluate the findings acquired thus far. It also significantly contributes to the prediction field by analyzing survival rates among female cervical cancer patients using decision tree methodology. We also discovered classification rules derived from the WEKA software (Serafín Muñoz, 2022; Varela Lorenzo & Delgado Gutierrez, 2020).

Application descriptions for health data mining are shown below. Their study used a Kohonen neural network to identify the shared features of diabetic patients and discovered three distinct patient profiles. Two clustering algorithms, the hierarchical algorithm (CHAMELEON) and the algorithm k partitional means were used to uncover the comorbidity related to type 2 diabetes mellitus (DM2) in Pakistan through data mining. The model's emphasis was on determining the kind of comorbid diagnosis, and the second step was determining the specific comorbid diagnosis.

Finally, efficient and visually straightforward Bayesian networks and decision trees were put to use, describing patients with this diagnosis. Some of the factors that may have a role in the diagnosis of certain diseases in the future have been identified through studies of health data mining, which are available in the review of the relevant literature. Particular emphasis is placed on breast cancer, the use of mammography, and the analysis of those images

in the study, demonstrating the significance of data mining and its application in the detection and diagnosis of cancer patients. Independent applications of techniques, including principal components analysis, decision trees, and association rules, were used to uncover hidden connections (PRIETO AMPARÁN, 2020).

The CRISP-DM 1.0 technique used data mining to study diabetic and pre-diabetic individuals, resulting in factor analysis. Attributes pertinent to the survey were determined, and a description was provided for the data that will be utilized to categorize patients in a particular region of Cuba. According to another factor analysis article, data mining successfully uncovered risk factors associated with fetal death via classification and clustering algorithms, yielding the most correlated variables.

These studies detail the use of both predictive and descriptive methods to zero in on the most important variables affecting the various diagnoses. In contrast to prior works, this paper provides an analytical model based on a voting system, which integrates the results of several methodologies. The proposed analytic model helps assess what aspects affect the success of aesthetic surgery. This article is structured as follows. The procedure, the proposed analytical model, and the examined data are all outlined in Section 2. The collected results are reported in Section 3, and their analysis is provided in Section 4. The results and directions for further study are presented in Section 5.

MATERIALS AND METHODS

Below, we detail the data mining approach that served as the basis for our investigation. Research, the analytical model proposed, and the data analysis description (López et al., 2020).

Methodology for Mining Data

You can find a wide variety of data mining techniques, such as CRISP-DM, SEMMA, KDD, etc., in published works. CRISP-DM is unique since it provides a comprehensive and detailed approach to the mining process. Business understanding, data understanding, data preparation, modeling, assessment, and distribution are the proposed six stages of sequential development in CRISP-DM. Interviews with relevant experts are conducted while learning the business; in this example.

It is possible to count the total number of variables and records. In the data preparation stage, it is proposed to create procedures for selecting, cleaning, and transforming data to obtain high-quality mining data. Similar to the previous step, this one is quantitative and features a few variables that are very sensitive to data quality. In the modeling stage, picking the proper analytic methods for the job is a top priority. This is the quantitative stage, concluding with each model's results. Models built in the preceding step are evaluated for quality using these metrics.

Since the justification for the selected method will be articulated emphatically, this stage is distinguished by its quantitative nature. The implementation stage follows an analysis of the mined data and the establishment of the factors that contributed to the final decision. Apt choices. At this stage, justification and recommendations for the specialist's next steps are made (González et al., 2018).

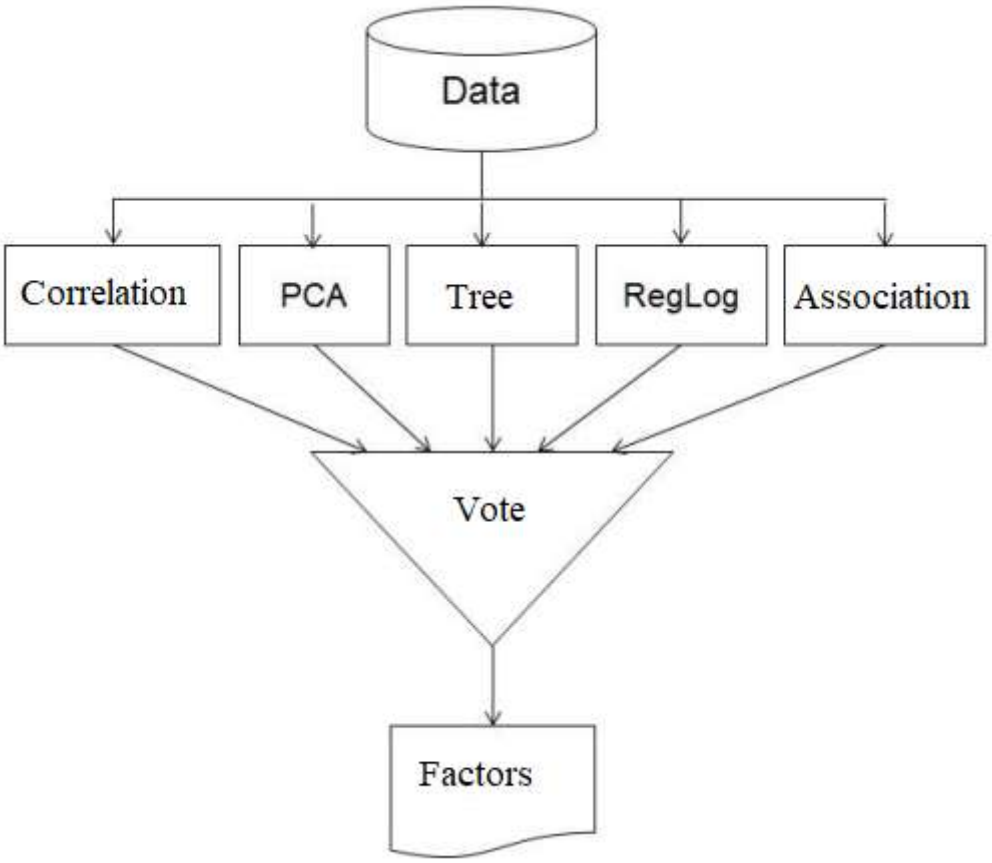


Figure 1: Voting Methodology Illustration

All the methods used look at the same data to discover what factors in a patient's medical history influence their choice of cosmetic surgery. By employing each technique, one can compile a report of the elements most responsible for shaping the surgical intervention. We use a vote on the criteria most prioritized by the methods we've employed to conclude (Freixa-Pamias et al., 2019).

The Analyzed Data:

Medical records of 2448 patients who saw Clinic plastic surgeon John Jairo Valencia Quintero for an examination were reviewed for this study. Following the removal of non-essential demographic information, 57 variables were chosen and organized as follows (Tables 1 through 7): demographic information (Table 1), medical history (Table 2), drug history (Table 3), gynecological history (Table 4), family history (Table 5), habits (Table 6), and physical assessment (Table 7).

Table 1: Individually identifiable data are listed.

VARIABLE	TYPE OF VARIABLE
Gender	Categorical
Age	Categorical
Marital Status	Categorical

Occupation	Categorical
Desired Surgery 1	Categorical
Desired Surgery 2	Categorical

Table 2: Pathological history variables.

VARIABLE	TYPE OF VARIABLE
Thyroid	Categorical
Heart	Categorical
AIDS HIV	Categorical
Kidneys	Categorical
Vesicle	Categorical
Arthritis	Categorical
Coagulation Problems	Categorical
Blood Pressure	Categorical
Lungs	Categorical
Nervous System	Categorical
Stomach	Categorical
Diabetes	Categorical
Asthma	Categorical
Lupus	Categorical
Cancer	Categorical
Fibromyalgia	Categorical
Liver	Categorical
Scleroderma	Categorical
Other	Categorical

RESULT:

Specifically, the 3.8 release of the data mining program WEKA was used for the experiments. One of the most popular data mining tools is WEKA (short for "Waikato Environment for Knowledge Analysis"), which was developed by a team at the University of Waikato in New Zealand (Ramos-Rincón, 2021). The procedures used in the voting system have yielded the following findings, which are detailed in detail below.

Table 3: Variables of drug use.

VARIABLE	TYPE OF VARIABLE
Aspirin	Categorical
Ibuprofen	Categorical
Arthritis	Categorical
Sleeping	Categorical
Cough	Categorical

Thyroid	Categorical
Hormones	Categorical
Tretinoin	Categorical
antibiotics	Categorical
Pressure	Categorical
Diabetes	Categorical
Diuretics	Categorical
phenobarbital	Categorical
Digital	Categorical
Cortisone	Categorical
To lose weight	Categorical

Table 4: Variables of the gynecological history.

VARIABLE	TYPE OF VARIABLE
Menstrual age	Categorical
Planning method	Categorical

Table 5: Family history variables.

VARIABLE	TYPE OF VARIABLE
arterial hypertension	Categorical
coronary artery disease	Categorical
Arthritis	Categorical
disorders of coagulation	Categorical
Asthma	Categorical
Diabetes	Categorical
breast cancer	Categorical
Another Categorical background	Categorical

Table 6: Variables of habits.

VARIABLE	TYPE OF VARIABLE
cigarette	Categorical
Liqueurs	Categorical
Coffee	Categorical
sleeping pills	Categorical
Other Categorical habits	Categorical

Table 7: Physical evaluation variables.	
VARIABLE	TYPE OF VARIABLE
BMI (body mass index)	Numeric

Indicator Analysis and Factor Selection: The correlation analysis shows that the ten variables in Table 8 play a significant role in the model, with the highest correlations to the dependent variable.

Table 8: Ranking with correlation value.		
RANKED	VARIABLE	VALUE
1	desired surgery	2 0.82628
2	family_history_hyperten	0.14981
3	liquor_habit	0.10978
4	coffee_habit	0.10828
5	BMI	0.08274
6	family_history_diabetes	0.07954
7	family_history_others	0.07823
8	habit_cigarette	0.07627
9	planning	0.07005
10	desired surgery	1 0.06545

With a correlation of 0.82628 on the studied variable, the "Desired surgery 2" variable was the most influential in determining whether or not cosmetic surgery would be carried out on a given patient. This indicates that the second intervention, a patient wanting to perform consultation in the evaluation, is crucial in determining whether or not a surgery would be completed. One of the ten most significant associations is the variable "Desired surgery 1." High blood pressure, diabetes, and other hereditary conditions were revealed to be among the top 10 predictors of the dependent variable.

Considerable weight was also given to vices, including heavy alcohol, caffeine, and tobacco use. Many people who seek out or have liposuction and abdominoplasty do so because they are overweight or obese, making the role of the variable "B.M.I." (Body Mass Index) undeniable. Last but not least, "Planning" was where the investigated variable was placed (Cardoso et al., 2018).

The P.C.A. Factor Selection Procedure

Table 9 ranks the factors using the P.C.A. (Principal Component Analysis) method.

Table 9: Ranking of variables with P.C.A.	
RANKED	VARIABLE
1	B.M.I.

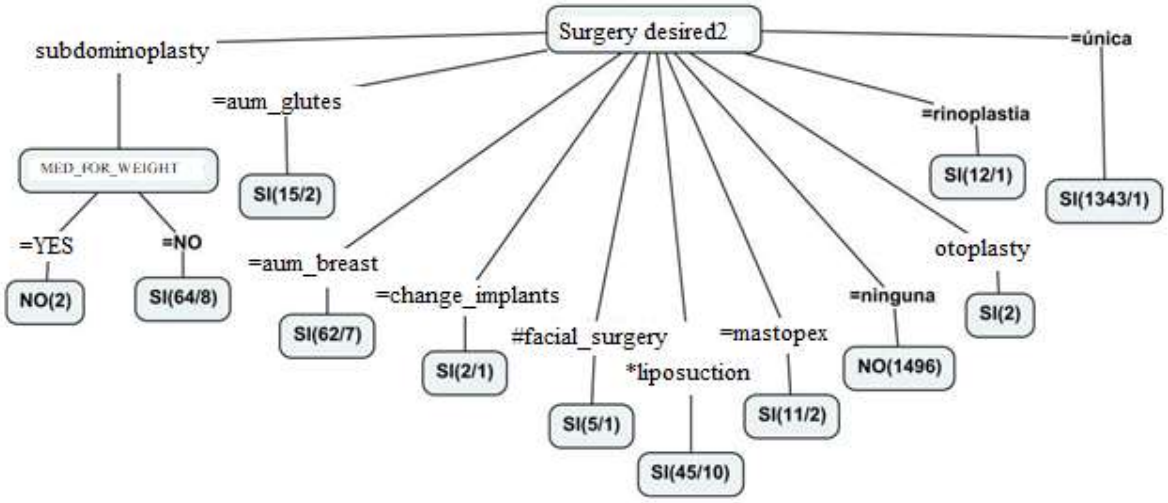
2	desired surgery1=facial_surgery
3	desired surgery1=breast_increase
4	desired surgery1=bichectomy
5	desired surgery1=brachioplasty
6	desired surgery1=increase_implants
7	desired surgery1=increase_buttocks
8	desired surgery1=abdominoplasty
9	occupation=domestic_service
10	occupation=public_server

It is clear from the tree that "desired surgery 2" is the most critical factor in determining the outcome. Patients who claim they do not want additional surgery and who desire a stomach tuck but also take weight-loss medicines are not operated on (Al-subhi et al., 2020). It has been observed that the most crucial variable in this type of analysis is the "B.M.I.," and that patients who have surgery are those who, as a second surgery, indicate that they want a tummy tuck and do not take weight loss drugs, an increase in buttocks, an increase in outcomes (Body Mass Index).

Facial surgery, breast augmentation, discectomy, Brachioplasty, abdominoplasty, and the other six most important factors are all part of the patient's initial list of desired surgeries. The third and fourth categories reflect the patients' respective professions, such as housekeepers and government workers.

Using Decision Trees to Select Relevant Factors

Predicting which patients will get cosmetic surgery can be done with 98.7 percent precision using a decision tree-based approach. Figure 2 shows the resulting tree of kn



knowledge acquisition. Breast augmentation, implant revision, facelift, mastopexy, ear and nose reshaping, and other cosmetic procedures (Miche et al., 2020).

Figure 2: Decision-making structure, illustrated.

Logistic Regression for Factor Analysis

We can use a logistic regression equation to determine which patients will get cosmetic surgery and which will not. The success of a procedure may be predicted with 98.5% precision using this strategy. This method's regression equations are summarised in Table 9, which lists the factors that favorably affect the choice to have a second cosmetic operation (Delgado-Enciso et al., 2021). As shown by the generated regression equations, the patient's desire to undergo cosmetic surgery, refusal to take certain medications, sharing specific anecdotes from his childhood, or engaging in other potentially detrimental behaviors are essential factors. Relevant vocations and marital statuses are also provided for each choice.

VARIABLES	DATA MINING EXPERIMENTS					
	1.Corre/	2 PCA	3. dec arb	4. Return	5. Assoc	No, TIMES
sex					X	1
age				X	X	2
marital status				X		1
occupation		X		X		2
desired surgery1	X	x		X		3
desired surgery2	X		X	X	X	4
aspirin medicine					X	1
medicine ibuprofen				X	X	1
medicine_ for arthritis				X	X	2
sleep medicine					X	1
cough medicine					X	2
thyroid medicine					X	2
1hormone medicine				X	X	1
medicine tretinoin					X	1
antibiotic medicine					X	1
medicine for pressure					x	2

Fig 3. Voting system result

DISCUSSION:

To accomplish the goal set out in this article, we will now cast our votes on the methods used to determine what elements affect the actual performance of cosmetic surgery. Figure 3 shows the results of the voting. Following a vote, the following criteria were specified to be the most important: the type of surgery chosen for the first and second preferences, the respondent's body mass index, their smoking habits, and their family history of diabetes. Those who fall under the category of surgical patients are people who:

They openly request a second operation.

- Facial, breast, biceps, implant, buttock, and abdominoplasty procedures rank highest on the list of most-wanted operations.
- They have a healthy weight and don't smoke regularly.
- No one in your family has ever been diagnosed with diabetes.

One is the patient's desire for a second surgical procedure: This is the most crucial consideration when deciding whether or not to get cosmetic surgery. Many patients want to

avoid having two operations and instead have one that combines two procedures because of the financial benefits (operating room, anesthesia, instrumentation, and medical expenses).

It is also essential to know what kind of surgery the patient wants before their first one, which is why data mining was used to determine that evaluation appointments for facial surgery, breast augmentation, discectomy, breast augmentation, and abdominoplasty are the most likely to result in surgery.

Thirdly, a patient's body mass index must be within a healthy range before receiving cosmetic surgery. This means that the patient must neither be morbidly obese nor dangerously underweight in height.

Fourth, the effects of smoking regularly. For the best results from cosmetic surgery, you should break the habit of smoking cigarettes. Plastic surgeon John Jairo Valencia requests that smokers quit the practice at least two months before the aesthetic procedure. This is because tobacco use impairs the oxygenation of tissues and, thus, the healing process, resulting in deformity to level at the skin level and an increased risk of contracting infections.

Factor 5 - Diabetes in the family: A patient's susceptibility to a disease is often influenced by hereditary factors. Since a diabetic patient is more likely to experience difficulties after surgery due to a lack of nutrients and oxygen reaching the tissues that need mending, developing this antecedent is particularly noteworthy. The sickness may have spread to other organs in this patient's body, causing damage there as well. Thus, it is crucial to investigate the value of a patient without a diabetes family history for cosmetic surgery.

CONCLUSION:

This article presents the results of an analytical model developed to study the factors affecting the efficacy of cosmetic surgery procedures; the model was then applied to a dataset containing medical records from an aesthetic study conducted at the Las Américas clinic in Medellin, Pakistan. The suggested analytical model is a voting system that allows factor analysis to incorporate the outcomes of other investigations, such as correlation analysis, principal component analysis, decision trees, logistic regression, and association rules.

With the analytical model, we can see how much each factor affects the choice to operate by using correlation and principal component analysis. Determine the impact of elements with the help of the decision tree, logistic regression, or a priori techniques. According to the findings, the primary and secondary procedures that a patient displays interest in having done at the office of Dr. John Jairo Valencia Quintero are the most important considerations for determining whether or not to proceed with cosmetic surgery. An additional concern is the patient's body mass index, which, despite typically being high for individuals seeking aesthetic body surgery, must not exceed morally accepted limitations.

Lastly, the patient's smoking habit is a significant determinant because of the adverse effects tobacco use has on the recovery process. As a last consideration, the family history of diabetes is being researched as a new finding. Studies have been proposed to investigate how such a history affects the decision to undergo aesthetic surgery. Finally, some recommendations for future research in the development of analytical models that facilitate the functioning of the study are offered, including 1) patient profiling or the use of a clustering

algorithm to define, by grouping, characteristics of John Jairo Valencia's plastic surgery patients. 2) Create a foresight model of potential postoperative problems.
Furthermore, it is suggested that the proposed voting system be used to conduct factor analysis in other domains and to incorporate additional voting techniques.

REFERENCES:

Al-subhi, S. H., Rubio, P. A. R., Pérez, P. P., Mahdi, G. S. S., & Leyva-Vázquez, M. (2020). New support tool to decision making in diagnosis, treatment and prognosis for cardiovascular diseases during pregnancy. *Revista Cubana de Obstetricia y Ginecología*, *46*(1), 1-16.

Cardoso, A. L., Figueiredo, A. G., Sales, L. G., Neta, A. M. S., Barreto, Í. D., & Trindade, L. M. (2018). Gastroesophageal reflux disease: prevalence and quality of life of health science undergraduate students. *Acta Gastroenterológica Latinoamericana*, *48*(3), 197-205.

Delgado-Enciso, I., Paz-Garcia, J., Barajas-Saucedo, C. E., Mokay-Ramírez, K. A., Meza-Robles, C., Lopez-Flores, R., Delgado-Machuca, M., Murillo-Zamora, E., Toscano-Velazquez, J. A., & Delgado-Enciso, J. (2021). Safety and efficacy of a COVID-19 treatment with nebulized and/or intravenous neutral electrolyzed saline combined with usual medical care vs. usual medical care alone: A randomized, open-label, controlled trial. *Experimental and Therapeutic Medicine*, *22*(3), 1-16.

Espinosa-Neisa, Y. M. (2021). Model to evaluate customer satisfaction in e-commerce through multi-criteria and semantic analysis.

Freixa-Pamias, R., Gràcia, P. B., Latre, M. L. R., Basile, L., Chamero, P. S., & Lupón, J. (2019). Improving medical attention through an integral care model for patients with ischemic heart disease or atrial fibrillation. *Revista espanola de cardiologia (English ed.)*, *72*(9), 779-781.

González, I. R., Ramos, A. A. N., Pérez, E. E., Barrera, M. E. R., Hernández, T. M. I., & Lienhard, R. (2018). About Lyme disease in Cuba and its knowledge by medical personnel. *Revista Cubana de Salud Pública*, *44*(1), 3-13.

Hernandez Celis, A. A., Ayobi Mendoza, N., & Correa Garcia, T. (2021). BioCheck: A Smart Checklist for Preoperative Evaluation.

López, J. M. D., González, B. G., & Alcázar-Navarrete, B. (2020). Personalized medicine in chronic obstructive pulmonary disease: how close are we? *Archivos de Bronconeumologia*, *56*(7), 420-421.

Miche, M. E. V., Batista, D. M., & Beltrandes, A. G. (2020). Knowledge of the harmful effects of ICT of students of the Latin American School of Medicine. *Panorama Cuba y Salud*, *15*(1), 6-10.

Morera de la Vall González, G. (2018). *Dust production in mining supression measures in quarry blasting* ETSI_Energia].

PRIETO AMPARÁN, J. A. (2020). *MODELACIÓN Y ANÁLISIS GEOESPACIAL DE FACTORES DETERMINANTES EN LA PRODUCTIVIDAD DE LOS ECOSISTEMAS, EN EL ESTADO DE CHIHUAHUA, MÉXICO* Universidad Autónoma de Chihuahua].

Ramos-Rincón, J. M. (2021). Imported diseases and travel medicine. In (Vol. 156, pp. 558-560).

Rojas-Rojas, A. F. (2020). Intelligent system to support micro injection process through artificial intelligent techniques and cae model integration.

Serafín Muñoz, A. H. (2022). Bioindicators and bio monitoring: Water quality control and sample management from La Purisima dam in the state of Guanajuato, Mexico, involving safety measures against covid 19.

Varela Lorenzo, A., & Delgado Gutierrez, A. (2020). Glucose classification and prediction system with Neural Networks.