Chapter 7 Identification and classification of mature corn cob species using Artificial Intelligence algorithms

Capítulo 7 Identificación y clasificación de especies de mazorcas de maíz maduro mediante algoritmos de Inteligencia Artificial

REYES-NAVA, Adriana^{†*}, LÓPEZ-GONZÁLEZ, Erika and GIL-ANTONIO, Leopoldo

Tecnológico Nacional de México-Tecnológico de Estudios Superiores de Jocotitlán, Division of Computer Systems Engineering.

ID 1st Author: Adriana, Reyes-Nava / ORC ID: 0000-0002-4440-909X

ID 1st Co-author: Erika, López-González / CVU CONAHCYT ID: 289386

ID 2nd Co-author: *Leopoldo*, *Gil-Antonio* / **ORC ID**: 0000-0002-7445-9426

DOI: 10.35429/H.2023.13.91.102

A. Reyes, E. López and L. Gil

*adriana.reyes@tesjo.edu.mx

A. Reyes, E. López and B. Hernández (AA. VV.) Computer Technology and Innovation. Handbooks-TI-©ECORFAN-Mexico, Mexico City, 2023

Abstract

This work presents the identification of different species of mature corn cobs produced in the northern area of the State of Mexico, specifically white, yellow, black and pink cobs by using a Convolutional Neural Network associated to a classification algorithm, the analysis is performed by supervised learning since the labeling of the cobs for training and validation is previously performed, so that in the prediction stage the algorithm is able to identify unlabeled species to check the effectiveness of the algorithm. This work is divided into two phases: the first one is to determine the performance of the algorithm for a sample of images and analysis through real time video and the second one is the implementation of the physical mechanism that is in charge of the classification. It is worth mentioning that this work focuses on the development of the first phase, which is based on the development and training of the algorithm based on Convolutional Neural Networks for the correct identification and classification of the cob.

Artificial Intelligence algorithms, Supervised Learning, Mature Corn Cob

Resumen

Este trabajo presenta la identificación de diferentes especies de mazorcas de maíz maduro producidas en la zona norte del Estado de México, específicamente mazorcas blancas, amarillas, negras y rosas mediante el uso de una Red Neuronal Convolucional asociada a un algoritmo de clasificación, el análisis se realiza mediante aprendizaje supervisado ya que previamente se realiza el etiquetado de las mazorcas para el entrenamiento y validación, de tal forma que en la etapa de predicción el algoritmo es capaz de identificar especies no etiquetadas para comprobar la efectividad del algoritmo. Este trabajo se divide en dos fases: la primera es determinar el rendimiento del algoritmo para una muestra de imágenes y análisis a través de video en tiempo real y la segunda es la implementación del mecanismo físico que se encarga de la clasificación. Cabe mencionar que este trabajo se enfoca en el desarrollo de la primera fase, la cual se basa en el desarrollo y entrenamiento del algoritmo basado en Redes Neuronales Convolucionales para la correcta identificación y clasificación de la mazorca.

Algoritmos de Inteligencia Artificial, Aprendizaje Supervisado, Mazorca de Maíz Madura

1. Introduction

Nowadays, information analysis using artificial intelligence algorithms is one of the most important mechanisms for decision-making and automating complex tasks for human beings, such is the case of text, image, video and audio analysis. The applications of Artificial Intelligence are very varied and we can currently apply this type of technological implementation in practically any sector. In this sense, image classification is considered by many to be one of the most outstanding developments since it is the basis of training for the development of more complex systems (Ponce Cruz, P, 2011).

Currently, the evolution of algorithms has advanced by leaps and bounds and there are numerous investigations that focus on the analysis of images and video as case studies, having as their main objective, to analyze and solve a problem, provide knowledge and provide tools to different development sectors, in the same way they provide technological help to non-specialized people in search of functional applications for a specific task.

The case study of this work focuses on finding an efficient solution for the classification of mature corn cobs to determine the species to which it belongs, whether it is white, yellow, black or pink, also considering the identification of 4 common diseases that can occur and thus hinder their correct classification given the characteristics of these diseases. all this work is carried out through the application of Artificial Intelligence algorithms, in order to make the classification process in the northern region of the State of Mexico more efficient. Currently, the sorting process is done manually by people who sort by color and also by disease, but it is a time-consuming job that requires several people since normally more than a ton of corn is collected per hectare. The problem for which this classification arises is because the corn that is planted is criollo, that is, the best ears that are harvested from the previous year are chosen in order to obtain the seed, however, as it is already several generations old, alterations can occur and sometimes the seed not only belongs to the predominant color, but it has grains of a different color, which causes mixtures to begin to be made at ripening.

The second problem: once the type of cob has been identified, it must be verified if it is healthy or has some disease, normally this work is done by people over 50 years of age, while the younger generations do not know how to determine the type of disease and how to stop its spread. Currently there are many works focused on the analysis of images of people, wounds, fruits, vegetables, objects and plants, but there are not fully identified works of cob analysis, so the implementations of this type of development in this sector in the northern area of the State of Mexico are new.

A case where AI is applied is, for example, in the area of medicine, given that the use of AI in the classification of medical information is becoming more and more common and has proven to be a valuable tool in the diagnosis and treatment of diseases. In particular, deep learning has proven to be effective in the classification and segmentation of medical images, as exemplified by the article titled "Artificial Intelligence: Development of Classification and Segmentation Algorithms in Chest X-Ray." Two deep learning models were developed for the classification of thoracic structures and segmentation of posterior costal arches in chest x-rays. Model 1, based on a pre-trained Bayesian approach, was used for the classification of thoracic structures and the identification of cardiomegaly. Model 2 focused on autonomic segmentation of the posterior costal arches on PA chest radiographs. Manual labeling of the rib arches was required for model 2 in a set of images of subjects without diagnosed pathologies. Model 2 used a convolutional neural network architecture with three convolutional layers and was evaluated by comparing its accuracy with the results of human radiologists.

The results showed that Model 2 achieved an accuracy of 95.6%, slightly exceeding the accuracy of human radiologists (94.5%). These results underscore the high accuracy of AI in the classification of thoracic structures and the identification of cardiomegaly, as well as in the segmentation of posterior costal arches (Raschio, E., Contreras, C., Allende, F., & Maturana, P., 2021). This suggests that AI can be a valuable tool for automating some of the processes involved in medical imaging, which could have a significant impact on healthcare in the future.

On the other hand, in the field of environmental conservation, the AmazonCRIME dataset represents an innovative solution for detecting areas related to transnational environmental crimes in the Amazon rainforest. The AmazonCRIME dataset, consisting of 30,000 labeled and georeferenced multispectral imagery, generated using Sentinel-2 imagery and the Google Earth Engine platform. The dataset is used to feed advanced Geospatial Artificial Intelligence models for the detection of areas linked to transnational environmental crimes in the Amazon rainforest. Experiments have been carried out using convolutional neural networks (CNNs) to classify images and detect areas of deforestation, illegal mining, illicit crops, among others (Pinto-Hidalgo, J. J., & Silva-Centeno, J. A., 2022).

The results mentioned in the article under the title "AmazonCRIME: A Geospatial Artificial Intelligence Dataset and Benchmark for the Classification of Potential Areas Linked to Transnational Environmental Crimes in the Amazon Rainforest" are promising, as the CNNs models achieved a classification accuracy of over 90% in most classes (runways, deforestation, forestry, illegal mining, illicit crops/potential area of coca crops and water). In addition, the models also managed to identify areas affected by the mining of natural resources and airstrips with rudimentary characteristics, demonstrating their important contribution to the field of environmental crime detection in the Amazon rainforest (Raschio, E., Contreras, C., Allende, F., & Maturana, P., 2021).

This work is organized in 5 sections, where the points related to the research topic are discussed, organized as follows:

- Section 1. Introduction, the focus of the work is presented and the state of the art on works related to image analysis using different artificial intelligence algorithms is presented.
- Section 2. Theoretical foundation, here are presented the necessary concepts on image analysis, development methodology and classification using Artificial Intelligence algorithms.
- Section 3. Development, this section shows the steps to follow to carry out the classification of the different types of ears of mature corn and the identification of their diseases or if they are healthy.

- Section 4. Results: The results obtained based on the experimentation carried out to identify the color of the cob and the type of disease are shown.
- Section 5. Conclusions, the conclusion reached based on the results obtained is presented and the future work that is planned to be carried out is shown.

2. Theoretical background

2.1. Artificial intelligence

Artificial Intelligence (AI) has emerged as a technology with a significant impact on the classification and processing of information in various disciplines. Its machine learning and data processing capabilities have enabled a wide range of applications ranging from image and text categorization to segmentation of complex information in areas such as medicine, fraud detection, customer service automation, and much more.

Artificial intelligence (AI) is the field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, problem-solving, and pattern recognition (What Is Artificial Intelligence? - Artificial Intelligence (AI) Explained - AWS, 2023). It refers to systems or machines that mimic human intelligence to perform tasks and can iteratively improve from the information they collect. For example, chatbots to understand problems faster and provide more efficient answers to users are typically hosted within web pages, intelligent assistants are used to analyze critical information from large data sets, and recommendation engines that can provide automated recommendations for TV shows based on users' viewing habits (Oracle, 2022).

There are currently 3 types of artificial intelligence, the first called narrow or weak artificial intelligence which refers to the ability of a computer system to perform a task with better accuracy than a person, the second known as strong general or human-level artificial intelligence which refers to the ability of a computer system to outperform people in any intellectual task and finally artificial super intelligence which is a computer system that has managed to outperform people in almost every field, including scientific creativity and social skills (What Is Artificial Intelligence? | Microsoft Azure, 2023).

Currently, the applications of artificial intelligence are present in many areas of daily life, for example, in the facial detection of people in certain areas, in virtual voice assistants such as Siri or Alexa and in industries such as transport, education, culture, medical services, etc.

Artificial intelligence applications seek to make people's lives easier, for example, with personal assistants we can make use of chatbots, with which we interact to be able to more quickly find products and services within our reach and these choices are stored so that we can later use the chatbot's recommendations and have a better service. The second example can be found in agricultural use, where platforms can be created that carry out a historical analysis of the information stored and can determine what the best yield of the fields will be and it is possible to warn of adverse environmental impacts that may harm production (Iberdrola, 2017).

We also find applications of artificial intelligence within finance, since they can help banks detect fraud, predict market patterns and advise what operations customers should carry out within their platforms, for example, apply for loans, credit cards and other transactions, another use is found in education since it allows us to know if a student may be a cause of abandonment, By analyzing the information, we can determine if their learning is optimal or requires some kind of attention so that they can be guided by the corresponding authorities in decision-making.

How these we can list different applications in industries and companies where artificial intelligence has given good answers within the market, however, we must also consider the risk represented by the development of this type of systems, because by automating tasks and improving the performance of operations, jobs are at risk since fewer and fewer personnel are needed to be able to carry out the activities. This means a potential risk for the population despite the benefits that artificial intelligence gives us.

Within the area of artificial intelligence there are several types of learning, these will help us determine how the algorithm works and the way in which we can train the system that is being generated, this classification will depend on the use of each of the applications and the mechanisms of implementation of these, as shown in Figure 1.

In this case, we must make a distinction in terms of the type of learning, since machine learning and deep learning are going to refer more to the application of the different artificial intelligence algorithms that we can find, while supervised, unsupervised, semi-supervised and reinforcement learning are going to refer to the way in which the algorithm is going to be trained.

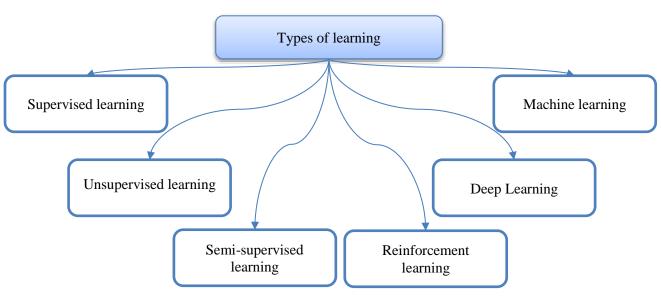


Figure 1 Types of learning

Source: Own Work

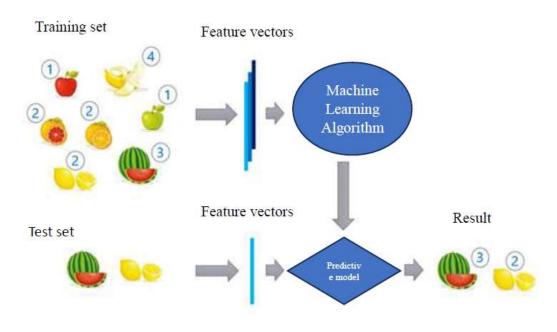
2. 2. Supervised learning

The branches of artificial intelligence are mainly divided into natural language processing, expert systems, computer vision, automatic speech recognition, planning, robotics and machine learning, it is in the latter where we find a sub-classification which is divided into supervised, unsupervised, reinforcement and deep machine learning, for the case study of this work supervised machine learning is used since the label of the class to which the species being tested during training belongs.

Supervised machine learning is where the data is previously labeled to carry out the training of the algorithms that will be in charge of classifying data and predicting results accurately (see figure 1), as the data enters the model, the weights are adjusted until the models perform a classification close to 100% using different types of validation such as hold-out or cross validation (Calvo, 2019).

On the other hand, unsupervised learning learns from the data since it enters without labeling and seeks that the data is understood by itself, in this case the model must learn to adjust the results and the groupings that are made until it finds a relationship of the data and can make a correct classification (What is unsupervised learning? 2020).

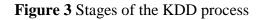
Figure 2 Supervised learning

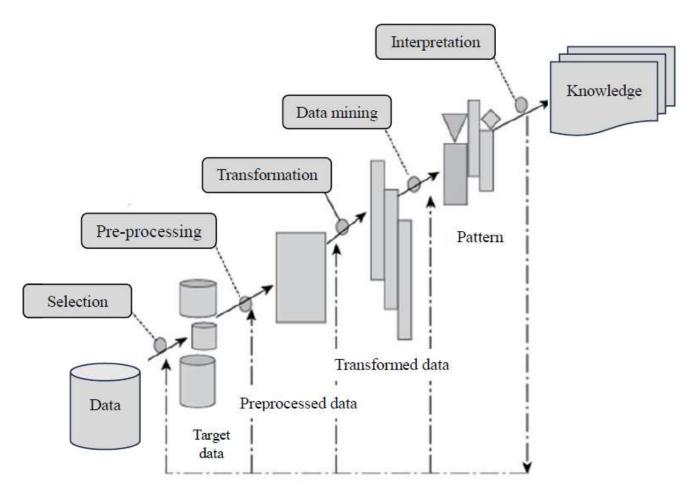


Source: obtained from (Calvo, 2019)

2. 3. KDD (Knowledge Discovery in Databases)

The process that is followed to obtain knowledge of a database through the application of data mining algorithms is shown in figure 1.1, this process is known as KDD (Knowledge Discovery in Databases), it refers to the analysis of large databases using different algorithms to obtain useful information for the organization (Tangarife Morales, 2016).





Source: Obtained from (Timarán-Pereira, S. R., et al, 2016)

The description of the stages of the knowledge extraction process according to (Timarán-Pereira, S. R., *et al*, 2016) are:

- Selection: Once the relevant and priority knowledge has been identified and the goals of the kdd process have been defined, from the point of view of the end user, an objective dataset is created, selecting the entire dataset or a representative sample of it, on which the discovery process is carried out.
- Pre-processing: data quality is analyzed, basic operations such as noisy data removal are applied, strategies are selected for handling unknown data (missing and empty), null data, duplicate data, and statistical techniques for their replacement.
- Transformation: useful features are sought to represent the data depending on the goal of the process. Dimension reduction or transformation methods are used to decrease the effective number of variables under consideration or to find invariant representations of the data.
- Data mining: it is the search and discovery of unsuspected and interesting patterns, applying discovery tasks such as classification.
- Interpretation: the patterns discovered are interpreted and the classification is made.

3. Development

In this stage, the application of the KDD process is presented, through which the analysis and classification of ears is carried out in real time through video analysis, considering the training stage of the algorithm based on a deep neural network with previously identified images according to the type of ear to which they belong.

3.1. Data selection and pre-processing.

The first stage of the work consisted of collecting the dataset of images of different cob samples of the 4 types of corn that will be analyzed, having a total of 200 images of ears between white (see figure 1), black (see figure 2), yellow (see figure 3) and pink (see figure 4), the latter is one of the least common to find so it becomes more complicated to identify its variations and In addition to having 50 images with ears of the 4 diseases that will be identified among the 4 colors of cobs (see Figure 5), it is important to mention that diseases are usually present in white and yellow cobs since they are the most predominant species.

Figure 4 Specimens of white maize

Figure 5 Black maize specimens



Source: Own Work

Figure 6 Specimens of yellow maize



Source: Own Work

Figure 7 Specimens of pink corn

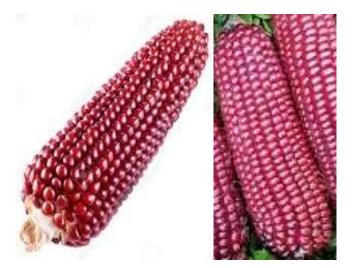


Figure 8 Specimens of diseased maize



Source: Own Work

To determine what kind of disease the ear has, the following aspects should be considered according to (Taba, S., 2004 and Varón de Agudelo, F., & Sarria Villa, G. A., 2007):

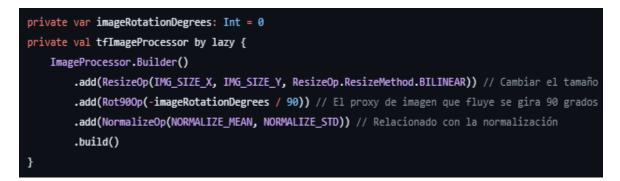
- Aspergillus: occurs when infected ears are stored with high moisture content, it may contain yellow-green, ivy-green or black masses on the grain or on the olote.
- Gibberella: This is most common in cold, moist areas. The first signs of the infection are the formation of white mycelia, which descend from the tip of the cob and give a reddish and pink coloration to the infected grains, this type of infection can be poisonous to some species of animals.
- Fusarium: This is the most common disease, causes infected grains to develop a cottony mold and can be toxic to animals.
- Common Carbon: this disease can be detected in the young plant from the germination of the corn when it is produced in the "huitlacoche" fungus consumed by some people, when it reaches the state of maturity it produces a black color similar to coal dust, hence its name.

3.2. Transformation and Data Mining

In this stage, the application of the first phase of the convolutional network is carried out, since the characteristics of the images are extracted, resized and scaled to be able to make simpler images and later convert them to vectors, so that the multilayer perceptron that is in charge of the analysis to determine to which class they belong can obtain the necessary data and make the correct classification of them.

For example, Figure 9 shows the code snippet where the image is readjusted to be analyzed, after this step comes the whole process of extracting data and features.

Figure 9 Resizing the Image



At this stage, in order to determine how the algorithm works to evaluate the result of the image analysis, the process of converting the image to a YUV format is carried out, then to an RGB bitmap, it is converted into tensorflowImage and finally it is passed to the tensorflowBuffer, so that the result of the analysis can be interpreted in the form of a list. This allows the detection to be more accurate since the model's inferences are carried out in this step.



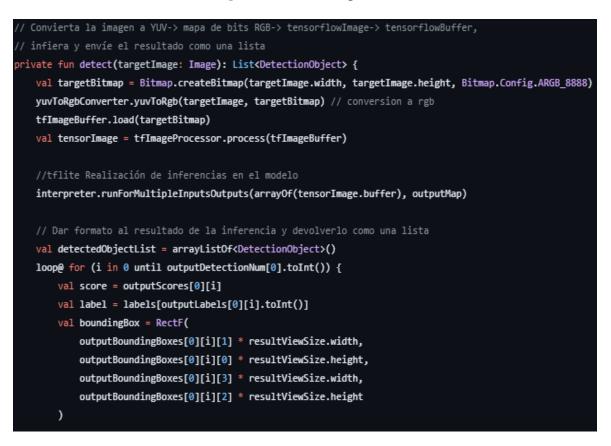
```
class YuvToRgbConverter(context: Context) {
    private val rs = RenderScript.create(context)
    private val scriptYuvToRgb = ScriptIntrinsicYuvToRGB.create(rs, Element.U8_4(rs))
    private var pixelCount: Int = -1
    private lateinit var yuvBuffer: ByteArray
    private lateinit var inputAllocation: Allocation
    private lateinit var outputAllocation: Allocation
```

Source: Own Work

3.3. Interpretation and knowledge

Once the entire process of transformation and processing of the images is completed, the validation stage of the system is carried out, for this the result of the analysis is arranged in the form of a list to determine what the results will be from the training carried out as shown in image 11.

Figure 11 Detection process



4. Results

Once the entire process of algorithm design, training and validation for each of the cob species was carried out, the results were obtained as shown in the table below. In the analysis of these results it is important to mention the type of validation that was used for the evaluation of the algorithm, in this case a hold out validation of 70% of samples for training and 30% of samples for testing was used, this data partition was performed 5 times and the accuracy shown is the average of these data. It is important to clarify that the experimentation is in an initial phase, so it is considered that in order to obtain better results it is necessary to increase the number of samples and the number of times of training to apply a cross-validation type.

In the same way, the result of 5 classes of cob species is considered, since the diseases that were previously explained are all included in a single category so that a simpler analysis of each of them is carried out.

Species of cob	Precision
White Corn	0.87
Black Corn	0.89
Yellow corn	0.76
Pink Corn	0.72
Diseased corn	0.70

 Table 1 Accuracy of analysis

5. Conclusions

Based on the results presented, it can be concluded that the development of this system to classify mature corn cobs of different species is a complex issue, since the characteristics of each of the elements have to be evaluated and the conditions in which the photographs were taken and the different factors that influence them must also be considered. for example, the brightness, the background of the image, and the quality of the image.

The accuracy achieved by the algorithm in the best of cases reached 87% of correct classification, that is, at least 8 out of 10 ears will be classified according to the class to which they belong, however, this percentage is not enough so it is sought to have an accuracy of at least 90% classification, in order to reach this percentage, it is necessary to take the following actions:

- Expand the data set with which you are working to be able to do a more exhaustive training and thus make the system learn in a more effective way.
- Have better control of the environmental conditions in which photographs of the samples to be analyzed are taken.
- Try another type of validation to compare the results.
- Experiment with another algorithm and do an additional feature extraction process to compare the results and determine which algorithm works more accurately.

As a future work, it is planned to carry out the implementation of the physical mechanism that is responsible for classifying the cob species in real time through video analysis, thus achieving a better quality in the product and improving the separation times of the producers, since currently this task is carried out manually by a group of people who are hired only for this purpose.

6. References

Calvo, D. (2019). Aprendizaje supervisado. Diego Calvo. Consultado 01 de octubre, 2023, desde https://www.diegocalvo.es/aprendizaje-supervisado/

Iberdrola. (2017). ¿Somos conscientes de los retos y principales aplicaciones de la Inteligencia Artificial? Iberdrola. Consultado 01 de octubre, 2023, desde https://www.iberdrola.com/innovacion/que-esinteligencia-artificial Pinto-Hidalgo, J. J., & Silva-Centeno, J. A. (2022). AmazonCRIME: un conjunto de datos y punto de referencia de Inteligencia Artificial Geoespacial para la clasificación de áreas potenciales vinculadas a Crímenes Ambientales Transnacionales en la Selva Amazónica. Revista de Teledetección, (59), 1-21. DOI: https://doi.org/10.4995/raet.2022.15710

Ponce Cruz, P. (2011). Inteligencia artificial con aplicaciones a la ingeniería (2nd ed.). Barcelona: Marcombo. URL:

 $\label{eq:https://books.google.es/books?hl=es&lr=&id=9ENzEAAAQBAJ&oi=fnd&pg=PR1&dq=Inteligencia+artificial+con+aplicaciones+a+la+ingenieri%CC%81a+&ots=3Ir0ThVg31&sig=p71WGBM6KOjiXpTRgbHXj0HPZN0#v=onepage&q=Inteligencia%20artificial%20con%20aplicaciones%20a%20la%20ingenieri%CC%81a&f=false$

¿Qué es la inteligencia artificial? - Explicación de la inteligencia artificial (IA) - AWS. (2023). Amazon Web Services, Inc. Consultado 01 de octubre, 2023, desde https://aws.amazon.com/es/what-is/artificial-intelligence/

Oracle. (2022). ¿Qué es la inteligencia artificial (IA)? | Oracle. Www.oracle.com. Consultado 01 de octubre, 2023, desde https://www.oracle.com/mx/artificial-intelligence/what-is-ai/

¿Qué es la inteligencia artificial? | Microsoft Azure. (2023). Azure.microsoft.com. Consultado 01 de octubre, 2023, desde https://azure.microsoft.com/es-mx/resources/cloud-computing-dictionary/what-is-artificial-intelligence/#how

¿Qué es el aprendizaje no supervisado? (2020). TIBCO Software. Consultado 01 de octubre, 2023, desde https://www.tibco.com/es/reference-center/what-is-unsupervised-learning

Raschio, E., Contreras, C., Allende, F., & Maturana, P. (2021). Inteligencia artificial: Desarrollo de algoritmos de clasificación y segmentación en radiografía de tórax. Revista chilena de radiología, 27(1), 8-16. DOI: http://dx.doi.org/10.4067/S0717-93082021000100008

Tangarife Morales, C. (2016). Estudio comparativo de metodologías para la detección de áreas de control de tensión. Colombia: Universidad Tecnológica de Pereira. Facultad de Ingenierías Eléctrica, Electrónica, Física y Ciencias de la Computación. Ingeniería Eléctrica. URI: https://hdl.handle.net/11059/7078

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y AlvaradoPérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: http://dx.doi.org/10.16925/9789587600490

Taba, S. (2004). Enfermedades del maíz: una guía para su identificación en el campo. Cimmyt. URL: https://books.google.es/books?hl=es&lr=&id=x4wC_JOmYL8C&oi=fnd&pg=PA1&dq=Enfermedades +del+ma%C3%ADz:+una+gu%C3%ADa+para+su+identificaci%C3%B3n+en+el+campo&ots=KorFL 3inSV&sig=iDdSNUP_NvP1TcS0s55aY2l7yUc#v=onepage&q=Enfermedades%20del%20ma%C3% ADz%3A%20una%20gu%C3%ADa%20para%20su%20identificaci%C3%B3n%20en%20el%20camp o&f=false

Varón de Agudelo, F., & Sarria Villa, G. A. (2007). Enfermedades del maíz y su manejo: compendio ilustrado. Instituto Colombiano Agropecuario-ICA. URL: http://bibliotecadigital.agronet.gov.co/bitstream/11348/6414/1/200753110056_DocumentoMaiz_ICA.p df